Chapter 4: Basic differentiation rules **Section 4**

The chain rule

What you need to know already:	What you can learn here:
➤ The concept and definition of derivative, basic differentiation rules.	➤ How to differentiate a composite function by using the chain rule.

The differentiation rules we have seen so far can be used only to differentiate very simple functions or combinations of them that involve only the arithmetic operation of addition, subtraction, multiplication and division. But there is another way to combine functions that is common and important, yet cannot be handled by such basic rules, namely composition. For your convenience, here is a reminder of what composing functions entails.

Knot on your finger

Given two functions $y = f(x)$ and $y = g(x)$, their ***composition***, $y = f\left[g(x)\right]$, is obtained by applying first the function g to the original input x, and then applying the function f to the output $g(x)$ of the first one. In symbols:

$$f \circ g(x) = f\left(g(x)\right): \; x \underset{g}{\rightarrow} g(x) \underset{f}{\rightarrow} f\left[g(x)\right]$$

Example: $f(x) = x^3$; $g(x) = x^2 + 1$

If we compose these two functions we obtain:

$$f\left(g(x)\right) = f\left(x^2+1\right) = \left(x^2+1\right)^3$$

Example: $f(x) = e^x$; $g(x) = x^2$

If we compose these two functions we obtain:

$$f\left(g(x)\right) = f\left(x^2\right) = e^{x^2}$$

One property of the composition is very important to remember and will play an important role when we start using the chain rule.

Technical fact

In general, the composition of two functions is ***NOT commutative***, meaning that in general:

$$f\left(g\left(x\right)\right) \neq g\left(f\left(x\right)\right)$$

Example: $f\left(x\right)=x^3;\ g\left(x\right)=x^2+1$

If we compose these two functions in the other order, we obtain:

$$g\left(f\left(x\right)\right)=g\left(x^3\right)=\left(x^3\right)^2+1=x^6+1$$

Clearly, this is not the same function as $f\left(g\left(x\right)\right)=\left(x^2+1\right)^3$

Example: $f\left(x\right)=e^x;\ g\left(x\right)=x^2$

If we compose these two functions in the other order, we obtain:

$$g\left(f(x)\right)=g\left(e^x\right)=\left(e^x\right)^2=e^{2x}$$

Again we can see that this is not the same as $f\left(g(x)\right)=e^{x^2}$.

How can we differentiate a composite function? Notice that none of the rules we have seen so far applies here. Now, just like for the product rule, there are some methods that are enticing in their simplicity, but are wrong: beware of them. Since I have seen several such methods and I do not want to place wrong ideas in your head, I will only show you the correct method, which turns out to also be quite simple. To see what this method is, we use Leibniz notation and a basic substitution.

If we let $y=f(x);\ u=g(x)$, so that $y=f\left(g(x)\right)=f(u)$ and use Leibniz notation, we can see that:

$$\frac{d}{dx}f\left(g(x)\right)=\frac{df\left(u\right)}{dx}$$

The last expression is a strange derivative, since it has different variables in the top and bottom, but remember that *u* is itself a function of *x*, so we are still computing the correct item.

Notice that this is NOT a fraction, but the limit of a fraction. Yet, if it were a fraction we could multiply and divide it by *du*, thus getting:

$$\frac{d}{dx}f\left(g(x)\right) = \frac{df\left(u\right)}{dx} = \frac{df\left(u\right)}{du}\frac{du}{dx} = \frac{df\left(u\right)}{du}\frac{dg\left(x\right)}{dx}$$

But this last expression makes sense as a product of derivatives, each having the same variable in the top and bottom. This all sounds reasonable, but a bit sloppy, since we are treating a derivative as if it were a fraction. It turns out, however, that all needed details are valid and it can be proved formally that this formula is correct.

Therefore it gives us the correct rule.

Technical fact: the chain rule

If $y=f\left(x\right)$ and $y=g\left(x\right)$ are differentiable functions, the derivative of their composition $y=f\left(g(x)\right)$ is given by the product:

$$y'=f'\left(g(x)\right)\times g'(x)$$

$$=\frac{df\left(g(x)\right)}{d\left(g(x)\right)}\times\frac{dg(x)}{dx}$$

In this formula, the first derivative is computed by considering $g\left(x\right)$ as a single variable.

Example: $f\left(g\left(x\right)\right)=\left(x^2+1\right)^3$

By using the chain rule, with $u=x^2+1$, we see that:

$$\frac{d}{dx}f\left(g\left(x\right)\right)=\frac{df}{du}\frac{du}{dx}=3\left(x^2+1\right)^2\times\left(2x\right)=6x\left(x^2+1\right)^2$$

Wait a minute: if composition is not commutative, how did you figure out which function to use first?

Excellent point, since, as I mentioned earlier, this is very important in the application of the rule. Here is the key:

Knot on your finger

Since composition is implemented by using brackets, a composite function can be recognized and decomposed as follows:

- ***Brackets*** are used to divide the first from the second function. When brackets are not present because other notation is used, they can be ***inserted*** appropriately.
- The function that is applied ***first*** is the one ***inside*** the brackets; the function that is applied ***second*** is the one ***outside*** the brackets.

Example: $y=\sqrt{e^x+x^2}$

Although there are no brackets, we can see that this function has an inside and an outside portion. In fact, we can write it as:

$$y=\left(e^x+x^2\right)^{1/2}$$

This clearly shows that the inside function – the one applied first in the composition – is $u=g\left(x\right)=e^x+x^2$, while the outside function – used second in the composition – is $y=f\left(u\right)=u^{1/2}$.

I am confused: didn't you say that in the chain rule we use the outside function first, the one that is inside second?

Yes, and that is a common reason for students being confused at the beginning. Let me clarify.

Knot on your finger

In a composite function the ***inside*** function is the one applied ***first to evaluate*** the function, but the ***outside*** function is the one ***differentiated first*** in the chain rule.

That can become tricky!

It can, but here is a different approach, much more visual and practical, that can help you understand both situations. I am confident that you are familiar with the ***Matryoshka*** dolls that originate in the Russian and Ukrainian culture; wooden dolls nested inside each other, as shown here. A composite function is structured like one of these dolls, with one function nested inside the other. This analogy extends to how the chain rule works.

But the Matryoshka can have several layers inside!

So can a composite function.

Knot on your finger

Several functions can be composed successively, so as to generate a multi-layered composite function.

Example: $f(x)=\sqrt{\sin\left(x^2+e^x\right)^2}$

By inserting brackets appropriately, we can see that this function is a composition of four layers:

$$f(x)=\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)^{\frac{1}{2}}$$

The first function used – and the last differentiated – is x^2+e^x. Next we have x^2, since the previous layer is squared; then $\sin x$; and finally $x^{\frac{1}{2}}$.

Notice that I am now eliminating the extra variable u – and other ones we may need for multiple layers – in order to keep the notation light. It is still there behind the scenes, in the proof of the rule, but we need not use it on a regular basis.

You kept me waiting long enough: how does this doll analogy work?

Strategy for applying the chain rule

When applying the chain rule to a composite function

1. Write the function with ***enough brackets*** to see its layered structure analogous to a Matryoshka doll.
2. ***Differentiate the outermost layer***, by using the appropriate rule, just as if you were opening the outermost layer of the doll. While doing this, consider the entire ***inside*** portion ***as a single variable*** and leave it as is, just as you would not touch any inside layers as you open the outside one in the doll.
3. ***Write down*** the derivative so obtained and leave it on the page, just as you would leave the outer layer of the doll on the table.
4. Move on to the ***next layer***, just as you would move to the next layer of the doll, and repeat the same procedure.
5. ***Keep repeating*** the procedure until all layers have been differentiated and placed next to each other, connected by multiplication.

Are you saying that applying the chain rule is like opening a Matryoshka doll?

Exactly!

So, why is it called "chain" rule?

Because this Matryoshka analogy is not commonly known, while mathematician have consistently seen the rule as a chain of differentiations, done one after the other on the successive layers of a composite function.

Example: $y=\sqrt{e^x+x^2}$

To differentiate this function, first we look at it clearly as a composite function by inserting brackets as we did before:

$$y=\left(e^x+x^2\right)^{1/2}$$

Now we start from the outermost doll… I mean layer, which is a power, and use the power rule on it, leaving the inside untouched:

$$y'=\frac{1}{2}\left(e^x+x^2\right)^{1/2-1}\times\left(e^x+x^2\right)'$$

Notice that we have only done the first step of the procedure, so we still need to differentiate the inside layer, which we do with the addition, exponential and power rules:

$$y'=\frac{1}{2}\left(e^x+x^2\right)^{-1/2}\times\left(e^x+2x\right)$$

And we are done! We can, if we want or need, rearrange this derivative as:

$$y'=\frac{e^x+2x}{2\sqrt{e^x+x^2}}$$

But this is not a required step to compute the derivative, only to work with it for later uses.

Example: $f(x)=\sqrt{\sin\left(x^2+e^x\right)^2}$

We have seen that we can write this function as:

$$f(x)=\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)^{\frac{1}{2}}$$

So, we apply the Matryoshka procedure, starting from the outside and going gradually in:

1) Power rule on the outside square:

$$f'(x)=\frac{1}{2}\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)^{-\frac{1}{2}}\times\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)'$$

2) We now move to the next layer, which is the sine function. We have not seen how to compute its derivative yet, but you likely remember it from your high school days, or you can look it up on a reference source and identify it as the cosine function. Therefore:

$$f'(x)=\frac{1}{2}\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)^{-\frac{1}{2}}\times\left(\cos\left[\left(x^2+e^x\right)^2\right]\right)\times\left[\left(x^2+e^x\right)^2\right]'$$

3) Next we differentiate the other square, still with the power rule:

$$f'(x)=\frac{1}{2}\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)^{-\frac{1}{2}}\times\left(\cos\left[\left(x^2+e^x\right)^2\right]\right)\times$$
$$\times\left[2\left(x^2+e^x\right)\right]\left(x^2+e^x\right)'$$

4) Finally we differentiate the last piece as we did before, thus concluding that:

$$f'(x)=\frac{1}{2}\left(\sin\left[\left(x^2+e^x\right)^2\right]\right)^{-\frac{1}{2}}\times\left(\cos\left[\left(x^2+e^x\right)^2\right]\right)\times$$
$$\times\left[2\left(x^2+e^x\right)\right]\times\left(2x+e^x\right)$$

No need to beautify this expression, unless we need to, but keep all the brackets in the right place and drop only the ones that are not really needed.

Does it get any more complicated than that?

It can, but we shall not deal with such mind-boggling examples. It is sufficient for you to understand the procedure and be able to apply it to functions of this order of difficulty.

I am going to conclude this section by showing you a simple and interesting theoretical application of the chain rule.

We have seen how to compute the derivative of the natural exponential function, but that proof relied on the convenient choice of e as the base. What if we need to differentiate a different, non-natural exponential function?

With the chain rule, the task is easy, and the answer goes further in explaining why in calculus we prefer to work with the natural exponential function.

Technical fact:

The general exponential rule

If $y = a^x$, where a is a positive constant, then:

$$y' = a^x \ln a$$

Proof

Since any positive number a can be written as $a = e^{\ln a}$, we can write:

$$y = a^x = \left(e^{\ln a}\right)^x = e^{(x \ln a)}$$

So, our function can be seen as a composite function and we can use the chain rule. First the natural exponential rule:

$$y' = e^{(x \ln a)} \times (x \ln a)'$$

Then the linear rule:

$$y' = e^{(x \ln a)} \times \ln a = a^x \ln a$$

Example: $y = \pi^{3x}$

We can rewrite this function as:

$$y = \left(\pi^3\right)^x$$

We can now apply the general exponential rule, since the base is just a constant:

$$y' = \left(\pi^3\right)^x \times \ln \pi^3 = 3\pi^{3x} \ln \pi$$

Example: $y = 3^{x^2}$

This time we need to be careful, because the exponent is itself a function. However, the base is still constant, so we can start with the general exponential rule. By highlighting that:

$$y = 3^{x^2} = 3^{\left(x^2\right)}$$

we get:

$$y' = 3^{x^2} \times \ln 3 \times \left(x^2\right)' = 2x3^{x^2} \ln 3$$

In the *Learning questions* you will see more theoretical uses of the chain rule. The next section will then discuss a very useful and important method based on the chain rule. But it is now time to practice what you have seen.

Summary

- The chain rule allows us to differentiate composite functions.
- Implementing the chain rule is analogous to opening a Matryoshka doll: one layer at a time, from the outside in, keeping the inside untouched at each step.

Common errors to avoid

- Do NOT differentiate two layers at once as you apply the rule. Instead, take your time and do one layer at a time.

Learning questions for Section D 4-4

Review questions:

1. Describe when and how to use the chain rule.
2. Explain what are the similarities and differences between the chain rule and a Matryoshka doll.

Memory questions:

1. For what type of functions is the chain rule used?
2. Which formula describes the chain rule?
3. What is the general exponential rule of differentiation?

Computation questions:

For each of the functions presented in questions 1-24, use the chain and other appropriate rules to compute its derivatives.

1. $y=\left(3x^3-5\sqrt{x}\right)^4$
2. $y=(x+1)(3x-5)^{27}$
3. $y=4e^{x^2-x}$
4. $y=e^{\sqrt{x}}$
5. $y=e^{-x}\sqrt{e^x+x}$
6. $y=\sqrt[3]{3-x^3}$
7. $y=\sqrt{\dfrac{5x-5}{3x^2-6}}$
8. $y=\sqrt[6]{x^3+x}$
9. $y=(x+1)^3\sqrt{3x-5}$
10. $y=(x+1)(3x-5)^{27}$
11. $y=\dfrac{e^{\sqrt{x}}}{\sqrt{e^x}}$
12. $y=\sqrt{\dfrac{x^2+1}{x-1}}$
13. $y=\sqrt[4]{\dfrac{5x-5}{3x^2-6}}$
14. $y=\dfrac{\sqrt{x+1}}{\sqrt{x}+1}$
15. $y=\dfrac{x}{\sqrt[6]{x^3+x}}$
16. $y=\dfrac{x^3\sqrt[3]{x+1}}{(3+x)^{\pi}}$
17. $y=\dfrac{4+\dfrac{2}{x^2}-\sqrt{3x}}{(x+1)^2}$
18. $f(x)=\dfrac{x-1}{\sqrt{x^2+2}}$
19. $y=\dfrac{\sqrt{x^2-2}\sqrt[3]{3-x^3}}{\sqrt[5]{5-x^5}}$
20. $y=\dfrac{\sqrt{x}+\dfrac{1}{\sqrt{x}}}{3x^4\left(\sqrt{x}+5\right)^2}$
21. $f(x)=\dfrac{e^{x^2}}{x}$
22. $f(x)=\dfrac{x}{e^{x^2}}$
23. $f(x)=\dfrac{\sqrt{x^3}-e^{2x}}{x^3-5x}$
24. $f(x)=\dfrac{\sqrt{x}+e^x}{3x^4e^{-x}}$

25. Compute the derivative of the function $f(x)=\dfrac{1}{\sqrt{x^2+9}}$ at x=4 by using two different sequences of differentiation rules.

Theory questions:

1. What is the formula for the derivative of a function of the form $y = \left[f(x) \right]^n$, also called the *generalized power rule*?

2. Which three rules are needed to differentiate the function $y = 2^{-x}$?

Proof questions:

1. By using a counterexample, show that the commonly used (by beginners) formula $\dfrac{df\left(g(x)\right)}{dx} = f'\left(g'(x)\right)$ is wrong.

2. Use the chain rule to prove the quotient rule.

3. Show that the natural exponential rule is a special case of the general exponential rule.

4. Prove that if r is a positive number and we consider the functions $f(x) = \left(\dfrac{x}{r-x}\right)^r$ and $g(x) = x^r$, then $\left(\dfrac{f}{g}\right)' = \dfrac{f'}{g'}$. This means that, while the correct quotient rule is the one given in the previous section, this simpler version, that draws ire from all calculus teachers, turns out to work, but in only in very few and special cases!

Application questions:

1. Construct the equation of the line tangent to $y = e^{x^2 - x}$ at the point $(1, 1)$

2. Determine the values of *a*, *b*, *c* and *d* (if any) for which the curves $f(x) = x^3 + ax^2 + b$ and $g(x) = e^{x^2} + cx + d$ have the same tangent line at $(0, 2)$ and determine the equation of such tangent line.

3. The position of a particle is described by the function $s = (3t^3 - 2)^{\frac{3}{2}}$, where distance is in meters and time in seconds. Find the velocity function and the velocity of the particle when $t = 2$ seconds.

4. An object moves on the *x* axis according to the function $x = \sqrt[3]{t+5}$. Estimate its position at $t = 3.2$ by using differentials.

Templated questions:

1. Construct a simple composite function involving functions whose derivative you know how to compute, and then differentiate it.

2. When computing the derivative of a given function, by using differentiation rules, identify each and every rule you use, in the specific order in which you use it. If a rule is used more than once, identify each instance where it is used.

What questions do you have for your instructor?