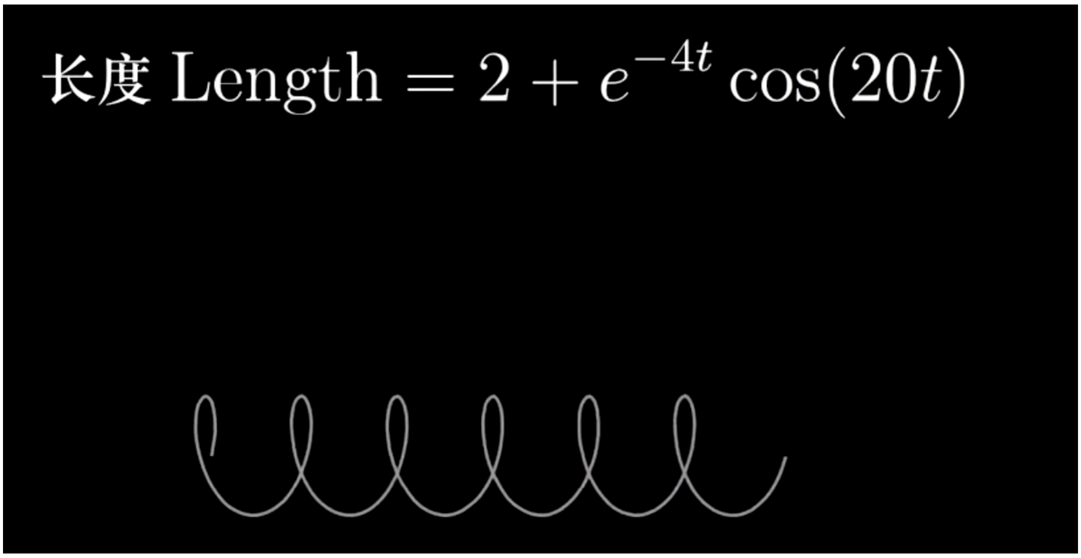
03-乘积法则与链式法则

> “Using the chain rule is like peeling an onion: you have to deal with each layer at a time, and if it is too big you will start crying.”
>
> **“运用链式法则就好比剥洋葱：你得一层一层地剥开它的心，要是它的个头太大，你还会鼻酸留泪。”**

将世界模型化的时候，一把需要 将简单函数混合、组合，我们现在的目标是理解这些复杂的组合如何求导



函数的混合、组合可以归结为三种组合函数：

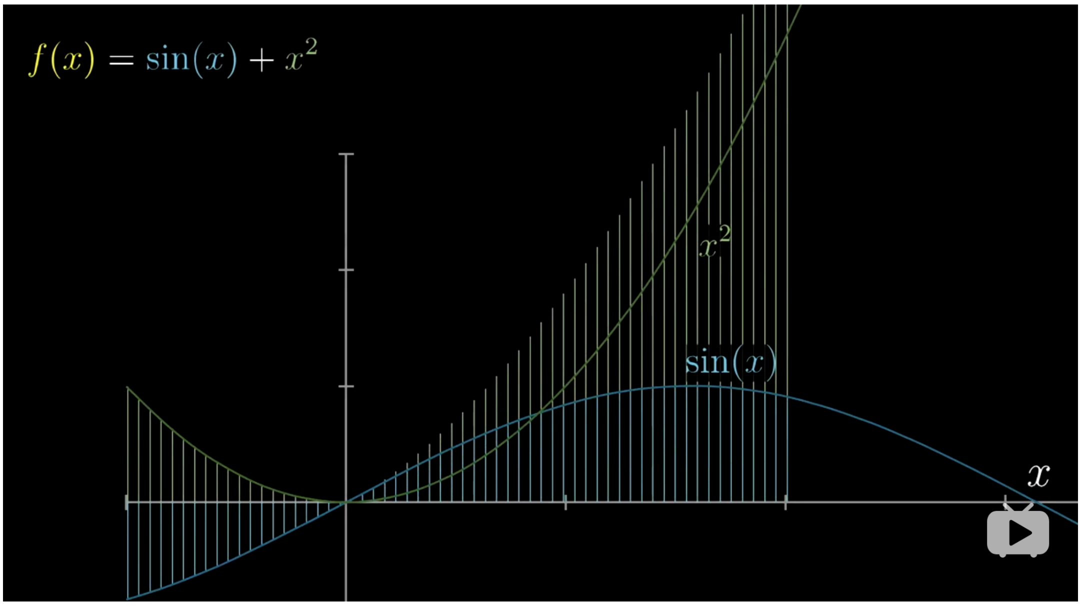
- 函数相加
- 函数相乘
- 函数套函数，也就是复合

3.1 加法法则（sum rule）

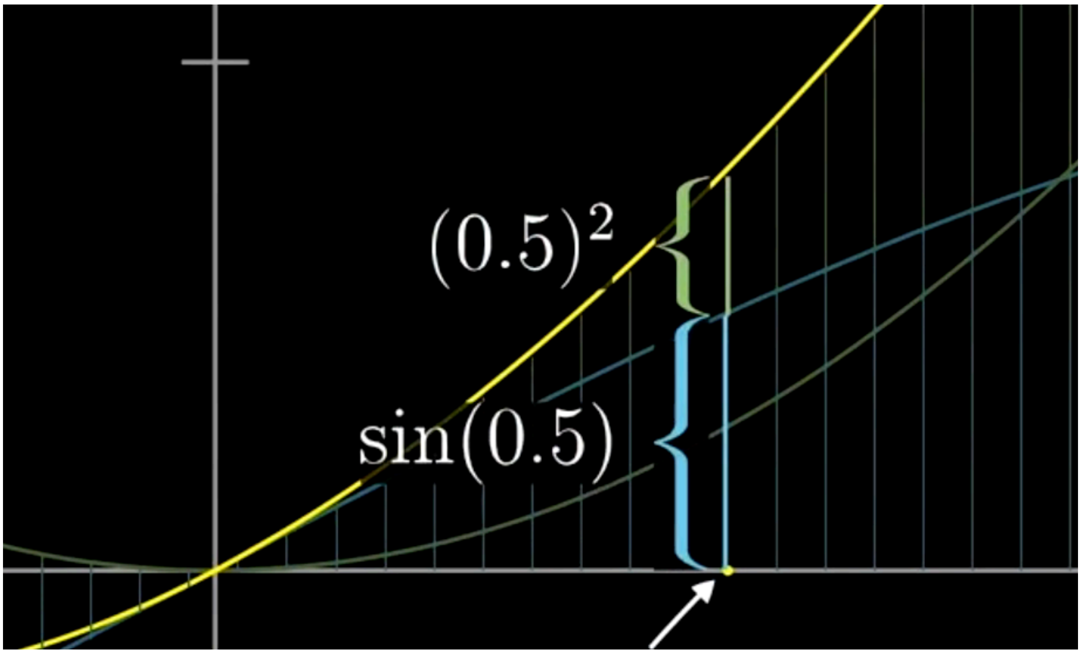
$$\frac{d}{dx}(g(x)+h(x))=\frac{dg}{dx}+\frac{dh}{dx}$$

加法求导就是对两个函数分别求导，再求和。这是什么意义呢?

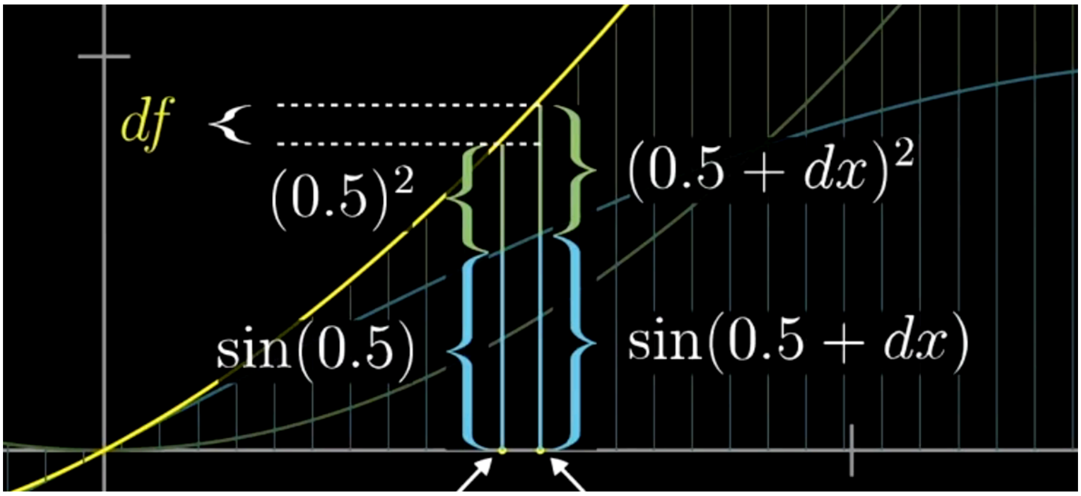
例如，我们来考虑f(x)=sin(x)+x^2。这个f(x)函数，你输入任意的变量值，就相当于把sin(x)和x^2对应的函数值相加（it's a function where，for every input，you add together the values of sin(x) and x^2 at that point）。



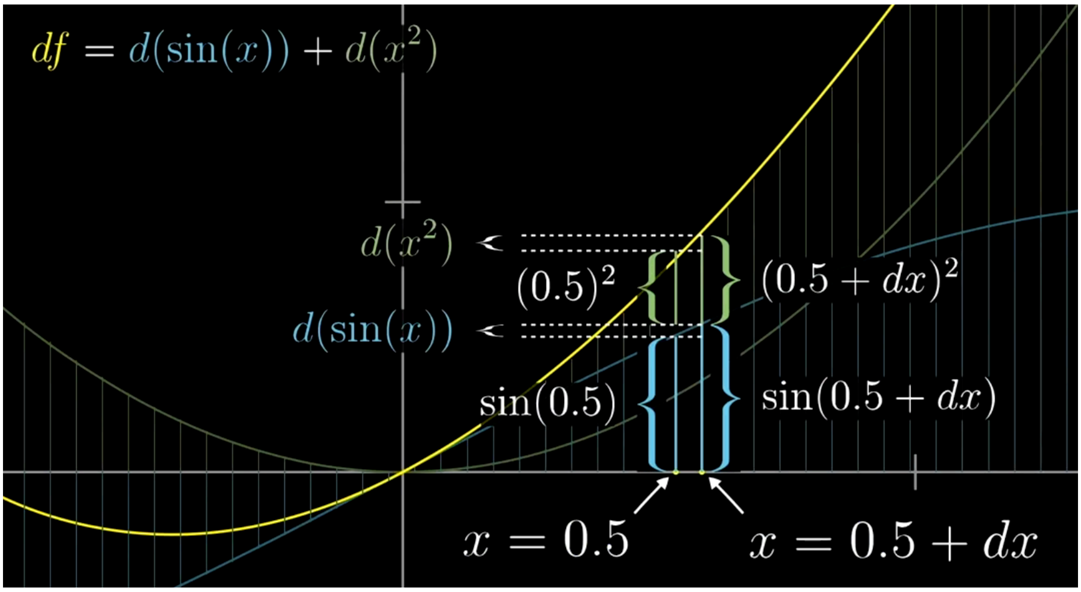
例如，取值0.5时，它们的和就是这两段摞起后的长度



对于导数值，看看x稍微做变化之后会发生什么。df就是dx发生变化后在f(x)上的变化量。



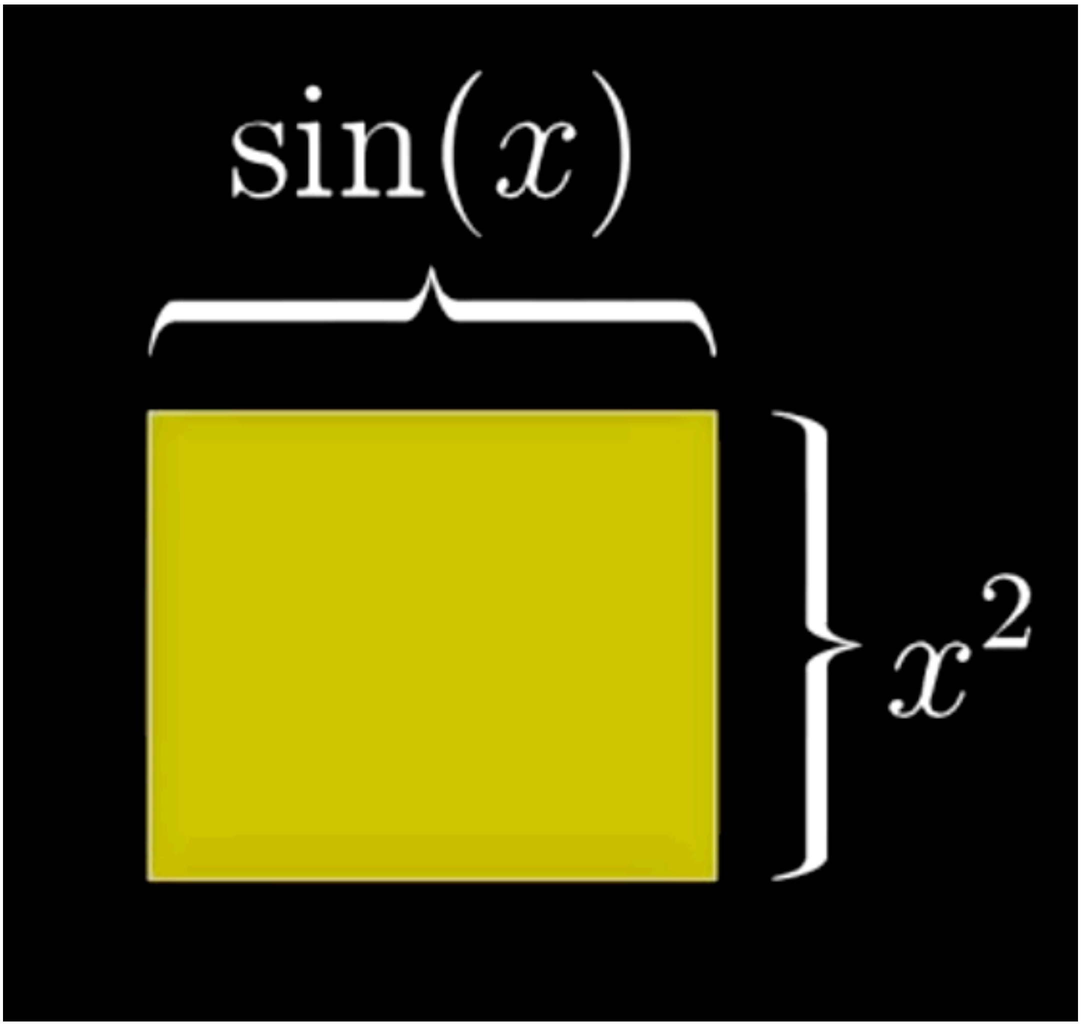
等价于这两个函数上dx变化引起的变化量的和。



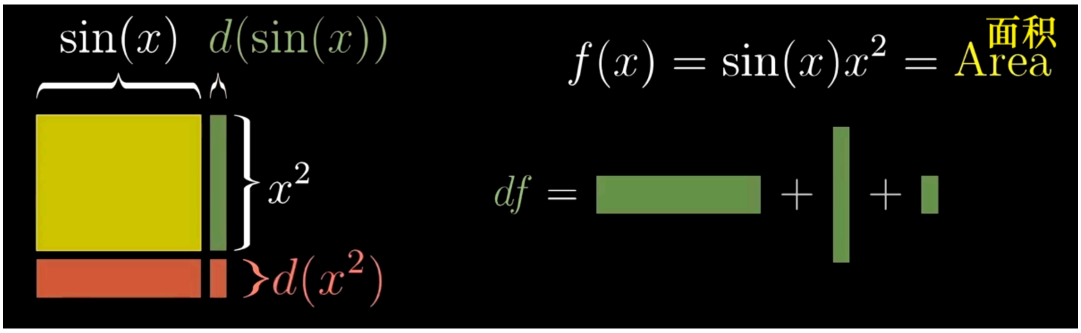
所以df/dx，也就是f(x)函数微小变化与自变量x微小变化的比值是f(x)各部分（parts）导数的和。

3.2 乘积法则（product rule）

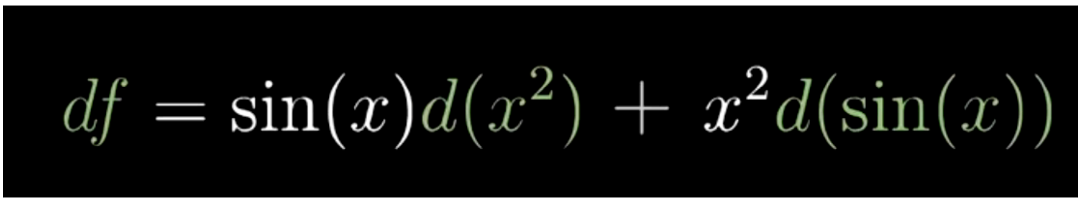
在数学中，如果处理的是两个东西的乘积，通过面积来理解是一种好的方法。例如，f(x)=sin(x)x^2



x变化时，这两个边会变化，引起面积的变化。所以由乘积定义的f(x)就是这个盒子的面积。对于导数，求的就是x的微小变化dx将如何影响面积。



这个增加的面积由三块组成。前两块组合为：



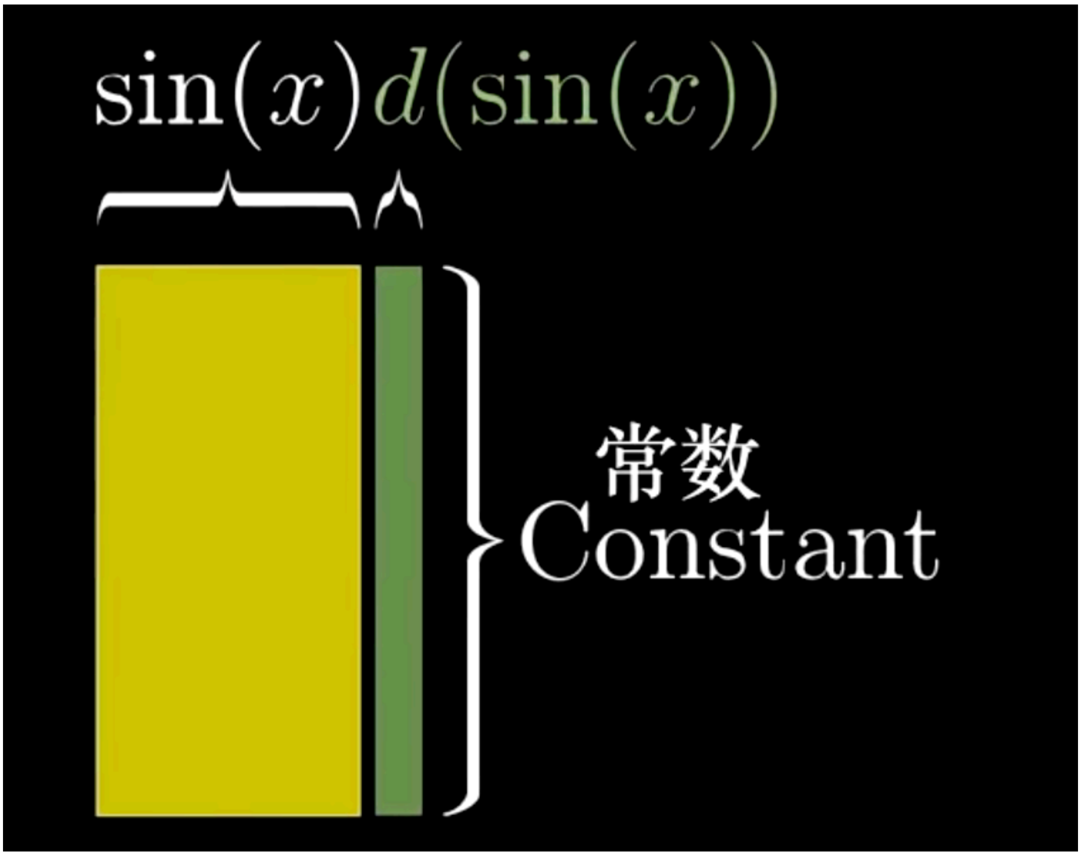
第三块面积（角落那一小块）可以忽略掉，因为它的面积最终会趋近于dx^2。当dx逼近0时，它就可以被忽略掉。

最终，我们可以获得乘法法则的口诀：左乘右导，右乘左导（left d(right)+Right d(left)）

$$(uv)' = uv' + vu'$$

在上面的例子中：左乘右导就是下边矩形的面积，而右乘左导就是右边矩形的面积：

如果函数和一个常数相乘，比如2sin(x),事情就会简单的多。最终只会增加一个面积，2乘d(sinx)



3.3函数复合的链式法则

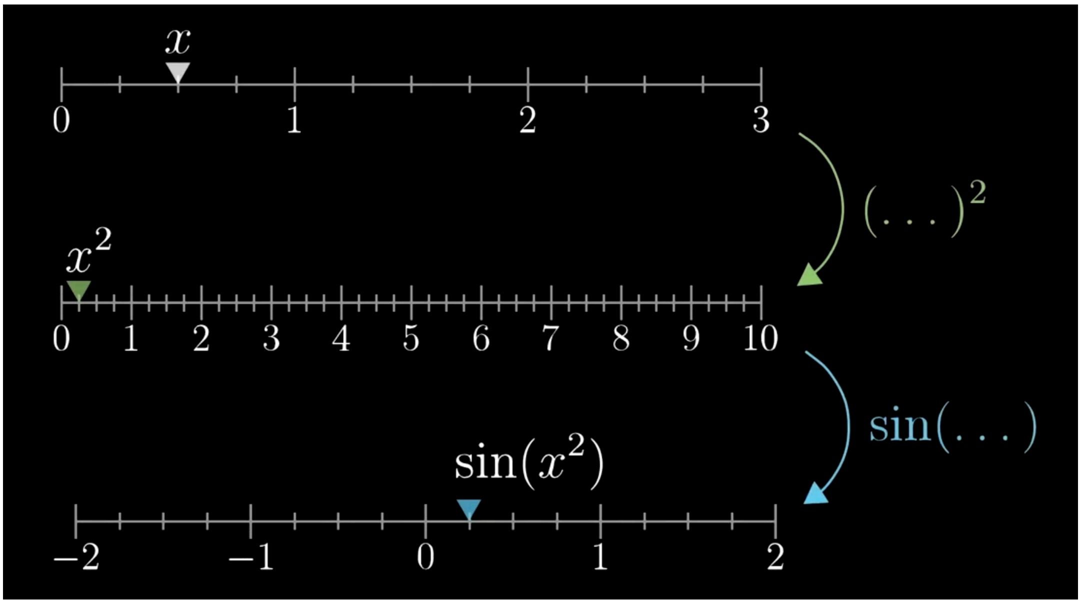
函数复合（function composition）是一种常见的函数组合方式，出现的极为频繁。

例如：我们将函数h塞（shove in）进g函数中，得到一个新的函数g：

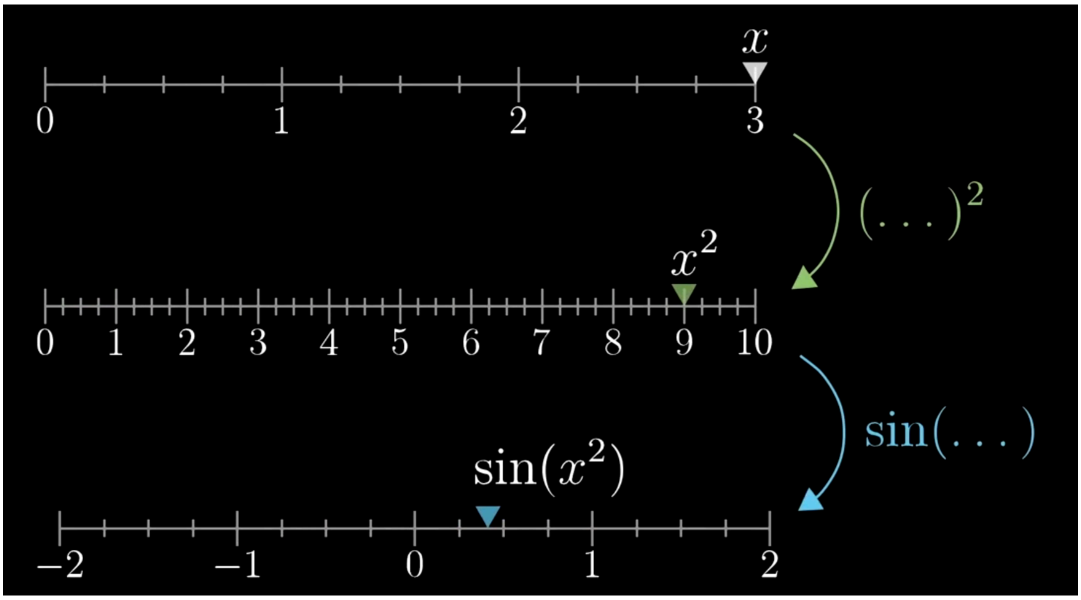
$$g(x) = \sin(x) \qquad h(x) = x^2$$

$$g(h(x)) = \sin(x^2)$$

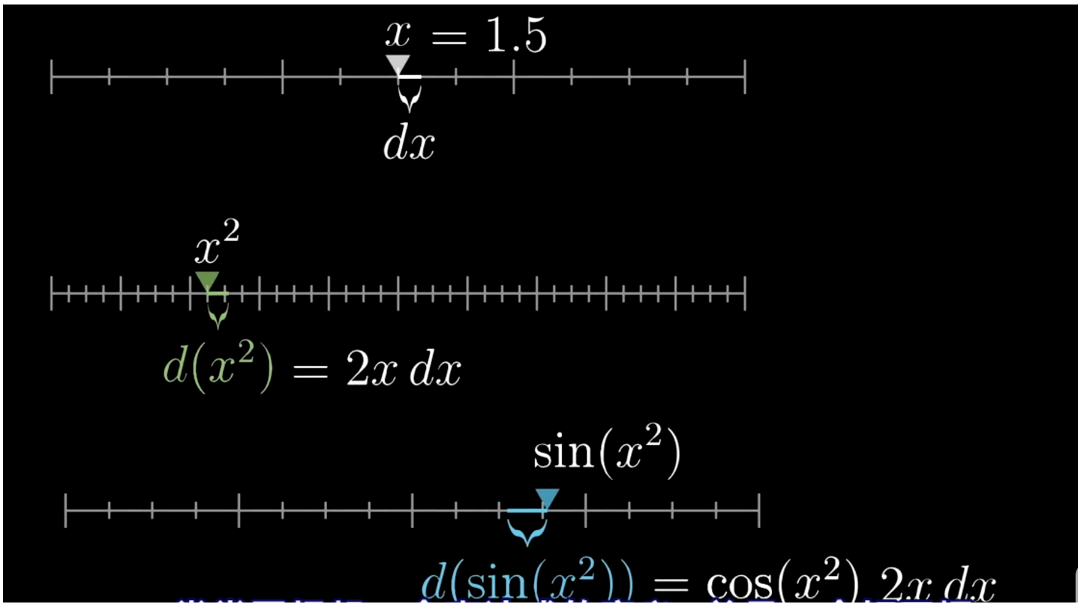
这个新函数sin(x^2)的导数是什么？
我们用一种可视化方法来看看：



第一个数轴上x的移动，第二个数轴上的x^2会相应的移动，siin(x^2)又会在第三个数轴上的移动。例如，x移动到3：



想要计算导数，我们让x的值稍微变化dx的位置：



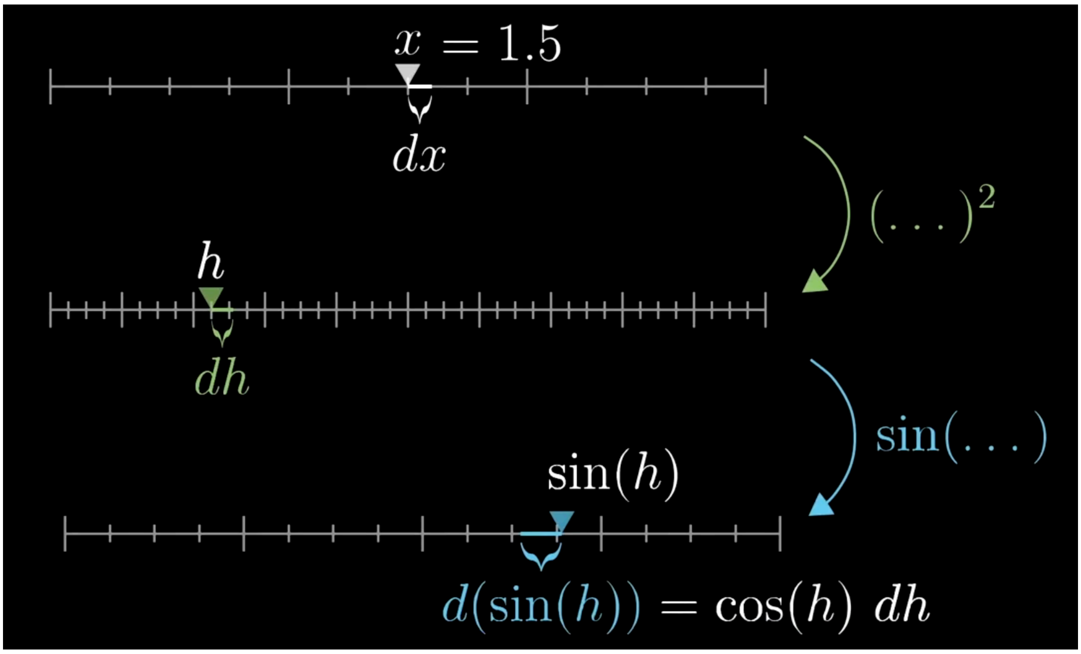
结论：

$$\frac{d}{dx}\underbrace{\sin(\overbrace{x^2}^{\text{内层 Inner}})}_{\text{外层 Outer}} = \underbrace{\cos(x^2)}_{\text{d(Outer) d(外层)}}\,2x$$

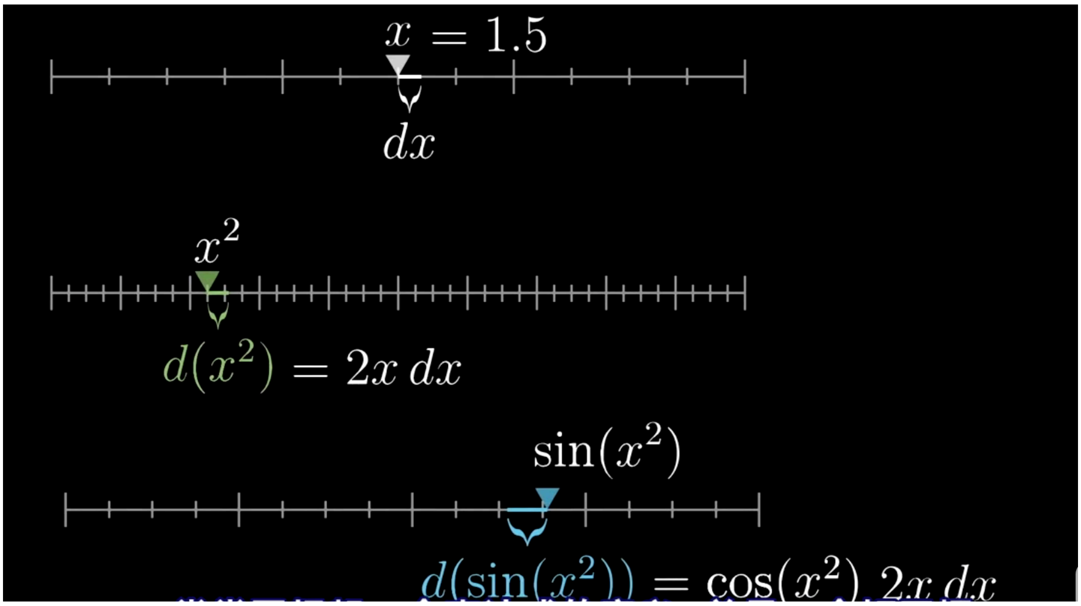
这就是链式法则（chain rule）：

$$\frac{d}{dx}\underbrace{g(\overbrace{h(x)}^{\text{内层 Inner}})}_{\text{外层 Outer}} = \underbrace{\frac{dg}{dh}(h(x))}_{\text{d(Outer) d(外层)}}\overbrace{\frac{dh}{dx}}^{\text{d(Inner) d(内层)}}(x)$$

注意这里我们内层求导用的是dg/dh。这是因为我们不能立刻知道最下面的变化量与x的依赖关系。但是我们可以对中间变量h求导，也就是说把第三行的变化量表示成第二行变化量dh的倍数，然后再进一步求出dh是什么



最后代入h与dh：



所以，链式表达式是在说：g的微小变化与h的微小变化是什么，再乘以h的微小变化与x的微小变化的值,结果就是g的微小变化与x的微小变化的比值