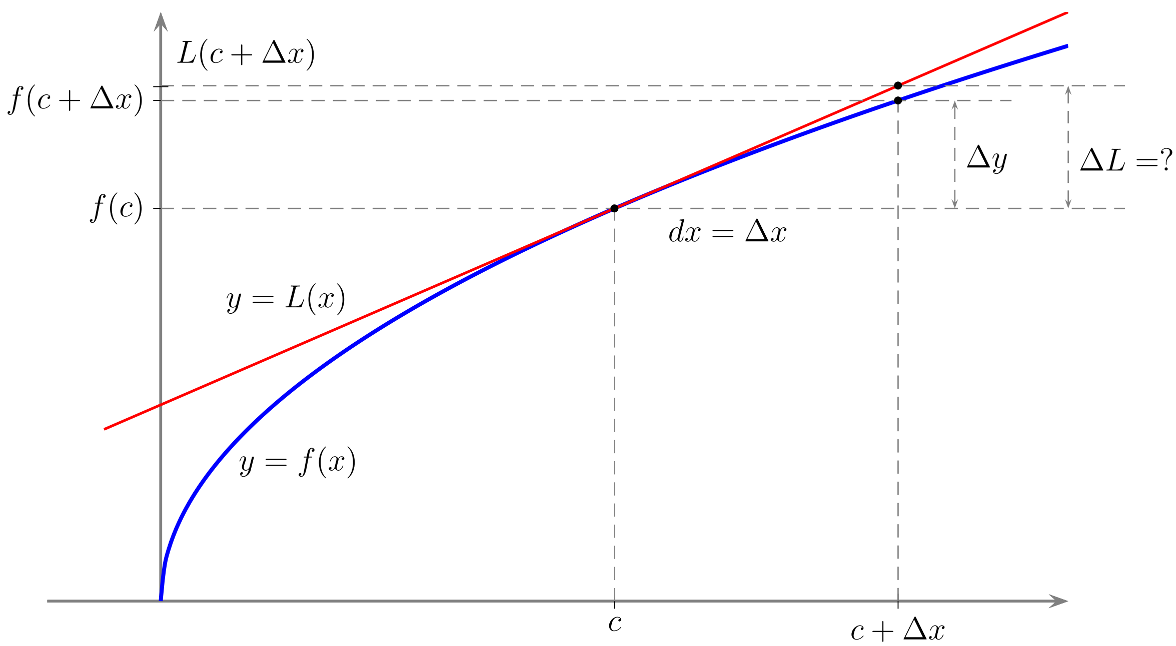
5.28 The Chain Rule

In a first semester calculus class we often say that a differentiable function can be well approximated by its tangent line.
Let us make this more precise. Let f differentiable on an open interval $I = (a, b)$ and let $c \in I$. We define the linearization of f at c by

(1) $$L(x) = f(c) + f'(c)\underbrace{(x - c)}_{\Delta x}$$

We then say

$$f(x) \approx L(x), \qquad \text{for all } x \text{ “near” } c \text{ (see below).}$$



Definition. The Differential
Let f differentiable on an open interval $I = (a, b)$ and let $x \in I$. Now let L be as defined in (1). Then we define the **differential**, df as the change in $L(x)$ from x to $x + dx$. That is,

$$\begin{aligned} df &= \Delta L = L(x + dx) - L(x) \\ &= f(x) + f'(x)((x + dx) - x) - f(x) \\ &= f'(x)dx \text{ (or } f'(x)\Delta x) \end{aligned}$$

Recall that the actual change in $f(x)$ from x to $x + \Delta x$ is $\Delta f = f(x + \delta x) - f(x)$. We then define the **Standard Linear Approximation** of f by

$$\Delta f \approx df$$

It follows that the error (i.e., the difference between the true change and the estimated change) is

$$\begin{aligned} \text{error} &= \Delta f - df \\ &= f(x + \Delta x) - f(x) - f'(x)\Delta x \\ &= \left(\frac{f(x + \Delta x) - f(x)}{\Delta x} - f'(x) \right) \Delta x \\ &= \epsilon \, \Delta x \end{aligned}$$

where

$$\epsilon = \frac{f(x + \Delta x) - f(x)}{\Delta x} - f'(x) \to 0$$

as $\Delta x \to 0$.

It follows that f is differentiable at x if and only if

(2) $$\Delta f = f'(x)\Delta x + \epsilon \Delta x$$

where $\epsilon \to 0$ as $\Delta x \to 0$.

We are now in position to prove the following theorem.

Theorem 1. The Chain Rule
Suppose that g is differentiable at x and f is differentiable at $g(x)$. Then the composite function $(f \circ g)$ is differentiable at x and

(3) $$(f \circ g)'(x) = f'(g(x))\, g'(x)$$

Proof. Let $u = g(x)$ and $y = f(u)$. By (2) there exists ϵ_1 and ϵ_2 such that

$$\Delta u = (g'(x) + \epsilon_1)\Delta x$$

where $\epsilon_1 \to 0$ as $\Delta x \to 0$ and

$$\Delta y = (f'(u) + \epsilon_2)\Delta u$$

where $\epsilon_2 \to 0$ as $\Delta u \to 0$. Combining expressions, we obtain

$$\begin{aligned}\Delta y &= (f'(u) + \epsilon_2)(g'(x) + \epsilon_1)\Delta x \\ &= (f'(u)g'(x) + f'(u)\epsilon_1 + g'(x)\epsilon_2 + \epsilon_1\, \epsilon_2)\Delta x\end{aligned}$$

Now we divide by Δx and let $\Delta x \to 0$ to get

$$\begin{aligned}(f \circ g)'(x) &= \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \\ &= \lim_{\Delta x \to 0} (f'(u)g'(x) + f'(u)\epsilon_1 + g'(x)\epsilon_2 + \epsilon_1\, \epsilon_2) \\ &= f'(u)g'(x)\end{aligned}$$

Now since $u = g(x)$ we are done.

□