Differentiation with respect to a coordinate.

A **numerical function** is a function whose domain and range are subsets of $\mathbf{R}$, the set of real numbers. Suppose f is a numerical function. We say f **is differentiable at** a a is an interior point of the domain of f and

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. The **derivative of** f denoted

$$f'$$

is, by definition, the set of ordered pairs (a, b) of real numbers f is differentiable at a and

$$b = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

Evidently, the domain of f' is the set of points in the domain of f at which f is differentiable; this set could be empty, in which case f' is the empty function.

A very important fact about differentiation of numerical functions is the following.

The Chain Rule. Suppose

(1) f is a numerical function and f is differentiable at a;

(2) g is a numerical function and g is differentiable at $f(a)$.
Then $g \circ f$ is differentiable at a and

$$(g \circ f)'(a) = g'(f(a))f'(a).$$

Proof. See any good book on one variable calculus. Let me know when you find one. $\square$

We now extend these notions as follows.

Let S be a set. We say y is a **real variable on** S if $y : S \to \mathbf{R}$. We say x is a **coordinate on** S if x is a real variable on S and x is one to one.

Suppose y is a real variable on S and x is a coordinate on S. Note that $y \circ x^{-1}$ is a numerical function whose domain is the range of x and whose range is the range of y. Evidently,

$$y = (y \circ x^{-1}) \circ x.$$

That is, the numerical function $y \circ x^{-1}$ is what you do to x to get y. We say

$$y = b \quad \textbf{when} \quad x = a$$

and write

$$y|_{x=a} = b$$

if a is in the range of x and $y(x^{-1}(a)) = b$. We set

$$\frac{dy}{dx} = (y \circ x^{-1})' \circ x$$

and call $\frac{dy}{dx}$ the **derivative of** y **with respect to** x; note that $\frac{dy}{dx}$ is a function whose domain is a subset of S and whose range is a subset of $\mathbf{R}$. If the domain of $\frac{dy}{dx}$ is all of S, which amount to saying that $y \circ x^{-1}$ is differentiable at each point of the range of x, we say that y **is differentiable with respect to** x.

We have the following.

The chain rule for real variables. Suppose x and t are coordinates on the set S and y is a real variable on S. Suppose y is differentiable with respect to x and x is differentiable with respect to t.

Then y is differentiable with respect to t and

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}.$$

Proof. Unwrap the definitions and invoke the chain rule for numerical

Corollary. Suppose x and t are coordinates on S. Then

$$\frac{dt}{dx} = \frac{1}{\frac{dx}{dt}}.$$

Proof. Obvious. I hope. □

Important Remark. Here is a good way to think of the chain rule. Given a coordinate x on S, let

$$\frac{d}{dx}$$

be the function which assigns $\frac{dy}{dx}$ to the variable y on S. That is, $\frac{d}{dx}$ is an *operation* you apply to one variable on S to get another variable on S, or at least a subset of S. The chain rule amounts to the statement if t is another coordinate on S then

$$\frac{d}{dt} = \frac{dx}{dt}\frac{d}{dx}.$$

Gee, that was so much fun we'll do it again! We have

$$\frac{d^2}{dt^2} = \left(\frac{dx}{dt}\right)^2\frac{d^2}{dx^2} + \frac{d^2x}{dt^2}\frac{d}{dx}.$$

We leave as an exercise for the reader to define the terms in this equation which need defining and then to derive the equation. You could also write it as

$$\frac{d^2}{dt^2} = \left(\frac{dx}{dt}\right)^2\frac{d}{dx}\frac{d}{dx} + \frac{d}{dt}\left(\frac{dx}{dt}\right)\frac{d}{dx}.$$

These notations are more subtle than you might think; make sure you understand them.

Example. Let $S = (0,\infty)$ and let $x(a) = a$ for $a \in (0,1)$. Note that x^p is a coordinate on S for any nonzero real number p. Let p and q be nonzero real numbers. We have

$$\frac{d(x^p)}{d(x^q)} = \frac{d(x^p)}{dx}\frac{dx}{d(x^q)} = \frac{d(x^p)}{dx}\frac{1}{\frac{d(x^q)}{dx}} = \frac{px^{p-1}}{qx^{q-1}} = \frac{p}{q}x^{p-q}.$$

Here's another way to do it which proceeds straight from the definition. I don't recommend doing it this way, but it illustrates how things work at a low level. The formalism developed above is designed to avoid having to do what we are about to do. What function do you do to x^q to get x^p? You raise x^q to the power $\frac{p}{q}$. That is, if we set $f(a) = a^{\frac{p}{q}}$ for $a \in (0,\infty)$ then

$$x^p = f \circ x^q.$$

Thus, as $f'(a) = \frac{p}{q}a^{\frac{p}{q}-1}$ for $a \in (0,\infty)$

$$\frac{d(x^p)}{d(x^q)} = f' \circ x^q = \frac{p}{q}(x^q)^{\frac{p}{q}-1} = \frac{p}{q}x^{p-q}.$$

Example. Let $C = \{(a,b) \in \mathbf{R}^2 : a^2 + b^2 = 1\}$ and let

$$S = \{(a,b) \in C : a > 0,\ b > 0\}.$$

Thus S is the part of the unit circle in the first quadrant of the Euclidean plane. We define

$$x : S \to \mathbf{R}, \quad y : S \to \mathbf{R}, \quad \theta : S \to \mathbf{R}, \quad u : S \to \mathbf{R}$$

by setting

$$x(a,b) = a, \quad y(a,b) = b, \quad \theta(a,b) = \arctan\frac{b}{a}, \quad u(a,b) = \frac{a}{1-b}$$

for $(a,b) \in S$. Note that each of x, y, θ, u are a coordinate for S. It would be nice if the S were the whole circle C, or more of it than the part of C in the first quadrant but the then the definitions of θ and u would break down and none of these functions would be coordinates. **Each of these four coordinates is a different way of tagging points on S by a number. That's how you should think of coordinates.** The first three should be familiar; maybe the last one isn't.

Exercise. Show that if $(a,b) \in S$ then the point

$$(0, u(a,b))$$

is the point of intersection of the line passing through $(0,1)$ and (a,b) with the line $\{(a,0) : a \in \mathbf{R}\}$.[1] This point of intersection is called the **stereographic projection** of (a,b) on the x-axis.

Note that

$$x = \sqrt{1-y^2}, \quad x = \cos\theta, \quad x = \frac{2u}{1+u^2}.$$

Exercise. Express each of y, θ, u as a function of the each the other three coordinates.

Exercise. Verify that $\frac{dx}{dy}$ is a coordinate for S, that $\frac{dy}{dx}$ is differentiable with respect to $\frac{dx}{dy}$ and compute

$$\frac{d\frac{dy}{dx}}{\frac{dx}{dy}}.$$

[1] Some people would call this line the x-axis. We won't do this because we've already used the identifier x for something else, namely x of a point in S is its "x" coordinate.