**Derivatives.**

It would be hard to do applied mathematics without derivatives. Newton invented calculus (along with Leibniz) to formulate his three laws of motion such as force = mass times acceleration (1666). See E.T. Bell, *Men of Mathematics*, for some of this story. The derivative can be used to represent all sorts of rates of change - not just distance with respect to time. Thus one finds differential equations in physics, chemistry, economics, ecology, weather. For example, the predator-prey equations describe the evolution of two interacting species such as cats and mice:

where \( x_1(t) \) = the number of mice at time \( t \) and \( x_2(t) \) = the number of cats at time \( t \),

\[
\begin{align*}
  x_1' &= (a - bx_2 - nx_1)x_1 \\
  x_2' &= (cx_1 - d - nx_2)x_2
\end{align*}
\]


The derivative \( f'(x) \) can be thought of geometrically as the slope of the tangent line to the curve \( y = f(x) \) at the point \( (x, f(x)) \).

\[
f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x} = \frac{dy}{dx} \quad \text{(Leibniz notation)}.
\]

More precisely, if \( f \) is defined on an interval \( I \ni \{c\} \), we make the following definition. Here \( I \ni \{c\} \) means \( I \) need not be open, but it must contain an open interval \( (c-\delta, c+\delta) \), for some \( \delta > 0 \), to get a 2-sided derivative.

**Definition.** \( f'(c) = \text{the derivative of } f \text{ at } c \) is defined by

\[
f'(c) = \lim_{h \to 0} \frac{f(c+h)-f(c)}{h}.
\]  

(1)

If this limit exists we say that \( f \) is differentiable at \( c \). When \( f \) is differentiable for
all $x \in S < 1$, we say $f$ is differentiable on the set $S$.

**Note.** Normally we will assume that $I$ is an open interval containing the point $c$. If instead $I = [c,b]$, then we would say the derivative is a right-hand derivative. Similarly if $I = (a,c]$, we say the derivative is a left-hand derivative.

**Example.** $y = |x|$ has only one-sided derivatives at $x=0$.

![Graph of $y = |x|$ with no 2-sided derivative at x=0](image)

By #2 on p. 69 of Lang, the function $y = |x|$ does not have a 2-sided derivative at $x=0$. It does have both one-sided derivatives though.

**Example.** $f(x) = \begin{cases} 0, x \in \mathbb{Q} \\ 1, x \notin \mathbb{Q} \end{cases}$

**Claim.** This function is nowhere differentiable.

It is basically impossible to draw this graph since there are points with $f(x)=0$ or 1 in any open interval no matter how small. Thus the function is not continuous at any point. By a result, which we will prove soon, this implies the function is not differentiable at any point. A more interesting example is Lang #1 on p. 69.

![Graph of $f(x)$ with $1$ on irrationals and $0$ on rationals](image)

Both lines really look solid even though the function is well defined and $f(x)$ has a unique value for each $x \in \mathbb{R}$. 52
Alternative Definition of the Derivative: The Linear Approximation Property.

Suppose I is an interval (containing more than one point) and suppose \( f: I \rightarrow \mathbb{R} \). If \( c \in I \), then \( f \) is differentiable at \( c \) if and only if there exists a real number \( L = f'(c) \) so that we have the following formula for some function \( \phi \) defined in an interval containing 0:

\[
f(c + h) = f(c) + Lh + \underbrace{\phi(h)}_{\text{linear in } h} \quad \text{where} \quad \lim_{h \to 0} \frac{\phi(h)}{h} = 0.
\]

Note that this means \( \phi(h) \) is a second order term. For if we set \( \psi(h) = \frac{\phi(h)}{h} \), for \( h \neq 0 \), and \( \psi(0) = 0 \), we have

\[
\phi(h) = h \cdot \psi(h)
\]

which approaches 0 faster than \( h \) does as \( h \to 0 \). This 2nd definition of derivative is much like saying \( f \) has the beginning terms of a Taylor expansion at \( x = c \).

We are saying here that a differentiable function \( f \) at \( c \) should look like a straight line (the tangent line) up to second order terms.

Proof of the Equivalence of the 2 Definitions of Derivative.

(1) implies (2).

Suppose there exists \( f'(c) \). Then set \( L = f'(c) \) and

\[
\phi(h) = f(c + h) - f(c) - Lh, \quad \text{for } h \neq 0, \quad \text{and} \quad \phi(0) = 0.
\]

Here we hold \( c \) fixed and let \( h \) vary near 0. We must investigate

\[
\frac{\phi(h)}{h} = \frac{f(c + h) - f(c) - Lh}{h} = \frac{f(c + h) - f(c)}{h} - L, \quad \text{for } h \neq 0,
\]

and by (1) this approaches 0 as \( h \) goes to 0, since \( L = f'(c) \).

(2) implies (1).

Suppose there is a real number \( L \) and a function \( \phi \) as in formula (2). Subtract \( f(c) \) from the first equality in (2) and then divide by \( h \) to obtain

\[
\frac{f(c + h) - f(c)}{h} = L + \frac{\phi(h)}{h} \quad \text{which approaches } L \text{ as } h \to 0.
\]

This means that \( f'(c) = L \) and \( f \) is differentiable at \( c \).

Properties of the Derivative.

1) If \( f \) is differentiable at \( c \), then \( f \) is continuous at \( c \).

2) Suppose that \( f, g: I \rightarrow \mathbb{R} \) and both \( f \) and \( g \) are differentiable at \( c \in I \). Then the sum and product of \( f \) and \( g \) are both differentiable:

\[
(f + g)'(c) = f'(c) + g'(c)
\]

\[
(f \cdot g)'(c) = f(c) \cdot g'(c) + f'(c) \cdot g(c).
\]

If moreover, we assume that \( g(c) \neq 0 \), then \( \frac{f}{g} \) is differentiable at \( c \) and
\[
\left( \frac{f}{g} \right)' = \frac{f'(c)g(c) - g'(c)f(c)}{g^2(c)}
\]

3) **Chain Rule.** Given intervals I, J (each with > 1 point),

Suppose \( f: I \rightarrow J \) and \( g: J \rightarrow \mathbb{R} \) with \( f \) differentiable at \( c \) and \( g \) differentiable at \( f(c) \).

Then the composite function \( (g \circ f)(x) = g(f(x)) \) is differentiable at \( c \) with

\[
(g \circ f)'(c) = g'(f(c)) \cdot f'(c).
\]

The Leibniz notation makes this easy to remember, setting \( u = f(x) \) and \( y = g(f(x)) = g(u) \), then

\[
\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.
\]

Just cancel the differentials \( du \).

**Proof of 1.**

We need to show that the existence of \( f'(c) \) implies that \( f(c+h) \) approaches \( f(c) \) as \( h \) approaches 0. Look at definition (2) to see this:

\[
f(c+h) = f(c) + Lh + \phi(h)
\]

approaches \( f(c) \) as \( h \) approaches 0, since \( \phi(h) \rightarrow 0 \) as \( h \rightarrow 0 \). We noted this just after (2).

Note. Differentiable at \( c \) implies continuous at \( c \). But the converse is false. By Exercise 2, p. 69 of Lang, you know that \( y = |x| \) is continuous at \( x = 0 \) but not differentiable there. A more interesting example is the Weierstrass example of a continuous nowhere differentiable function mentioned on pp. 49-50 of the notes. We will study that example in detail later.

**2) Proofs of the Product Rule and Quotient Rule.**

**Product Rule.**

As in the proof of the rule for limit of a product, we add and subtract a term in between the two things subtracted in the numerator of the difference quotient defining the derivative:

\[
\frac{f(c+h)g(c+h) - f(c)g(c)}{h} = \frac{f(c+h)g(c+h) - f(c+h)g(c) + f(c+h)g(c) - f(c)g(c)}{h}
\]

\[
= \frac{f(c+h)g(c+h) - f(c+h)g(c)}{h} + \frac{f(c+h)g(c) - f(c)g(c)}{h}
\]

\[
= f(c+h) \frac{g(c+h) - g(c)}{h} + g(c) \frac{f(c+h) - f(c)}{h}.
\]

This clearly approaches \( f(c)g'(c) + g(c)f'(c) \), to prove the product rule. Here we have used the rules for limit of sum and product.

**Quotient Rule.**

By the product rule, it suffices to do

\[
\left( \frac{1}{g} \right)' = -\frac{g'(c)}{g^2(c)}.
\]
To see this, look at
\[ \frac{1}{g(c+h)} - \frac{1}{g(c)} = \frac{1}{h} \left( \frac{g(c) - g(c+h)}{g(c+h)} \right) = \frac{1}{g(c) \cdot g(c+h)} \frac{g(c)-g(c+h)}{h}. \]
This clearly approaches the desired limit.

3) Proof of the Chain Rule.

It is easy to convince ourselves of the formula by writing
\[ \frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \frac{\Delta u}{\Delta x}. \]
The problem with this proof is that \( \Delta u = f(x+\Delta x)-f(x) \) might vanish at random places. Thus we are writing \( \lim_{\Delta x \to 0} \Delta u = 1. \)

In order to write a legal proof of the chain rule, we will use the second definition of derivative (2). So we use (2) for the function \( f \) and obtain:
\[ f(c+h) = f(c) + f'(c)h + \phi_1(h) \text{ with } \phi_1 \text{ as in (3) above.} \]
Set \( k = f(c+h)-f(c) \) and \( d = f(c). \)
Then using (2) for the function \( g \), we have
\[ g(d+k) = g(d) + g'(d)k + \phi_2(k), \text{ where } \phi_2(k) = k\psi_2(k), \psi_2(k) \to 0, \text{ as } k \to 0. \]
Plug (4) into (5) to see that if \( k = f(c+h)-f(c) \),
\[ g(f(c+h)) = g(f(c)+k) = g(f(c)) + g'(f(c)) \left( f'(c)h + \phi_1(h) \right) + \phi_2(k) \]
\[ = g(f(c)) + g'(f(c)) f'(c)h + \left\{ g'(f(c)) \phi_1(h) + k \psi_2(k) \right\}. \]
By the second definition of the derivative, to finish our proof of the chain rule, all we have to show is that the last term in brackets is a 2nd order term; i.e., still approaches 0 after being divided by \( h \). Since \( \frac{k}{h} \to f'(c) \) as \( h \to 0 \), and \( k \to 0 \) as \( h \to 0 \), this term in brackets is indeed 2nd order.

This completes the proof of the chain rule. See Bryant, *Yet Another Introduction to Analysis*, for an alternative discussion.

**The Usual Examples.**

1) \( f(x) = \alpha = \text{constant} \) implies \( f'(x) = 0 \) for all \( x \).
\[ f'(c) = \lim_{h \to 0} \frac{f(c+h)-f(c)}{h} = \lim_{h \to 0} \frac{\alpha - \alpha}{h} = 0. \]

2) \( f(x) = \alpha x + \beta, \) for constants \( \alpha, \beta \in \mathbb{R} \) implies \( f'(x) = \alpha \), for all \( x \).
\[ f'(c) = \lim_{h \to 0} \frac{f(c+h)-f(c)}{h} = \lim_{h \to 0} \frac{\alpha(c+h) + \beta - \alpha c - \beta}{h} \]
\[ = \lim_{h \to 0} \frac{\alpha h}{h} = \alpha. \]
3) \( \frac{d(x^n)}{dx} = n x^{n-1} , \quad n=0,1,2,3, \ldots \).

**Proof.** By mathematical induction.

We just did the cases \( n=0 \) and \( n=1 \). Now we do the induction step.

Assume that \( \frac{d(x^n)}{dx} = n x^{n-1} \). Then prove \( \frac{d(x^{n+1})}{dx} = (n+1) x^n \).

Use the fact that \( x^{n+1} = x \cdot x^n \) and the product rule to see that:

\[
\frac{d(x^{n+1})}{dx} = \frac{d(x \cdot x^n)}{dx} = x \frac{d(x^n)}{dx} + x^n \frac{d(x)}{dx} = x (n x^{n-1}) + x^n \cdot 1 = (n+1) x^n .
\]

**Remarks.** We can use these examples and the rules for derivatives to deduce that all polynomials

\[ p(x) = \alpha_n x^n + \alpha_{n-1} x^{n-1} + \ldots + \alpha_1 x + \alpha_0 , \quad \text{for } \alpha_i \in \mathbb{R}, \]

are differentiable at all \( x \). By the quotient rule we can differentiate rational functions

\[ \frac{p(x)}{q(x)} \]

for polynomials \( p(x), q(x) \) at all points \( x \) such that \( q(x) \neq 0 \).

Any computer knowing Mathematica can grind out lots of derivatives.

Once we know a few more rules for differentiation we can differentiate some more complicated functions; e.g.

\[ x^x = \exp(x \log x); \quad \sqrt{1+x^2}; \quad \sin\left(\frac{1}{x}\right), \quad x \neq 0. \]

Our next goals are to figure out the mean value theorem and the formula for the derivative of the inverse of a function. We shouldn’t forget that there is another important application of derivatives - to the location of maxima and minima of functions.

**The Mean Value Theorem.**

Suppose that \( f: [a,b] \rightarrow \mathbb{R} \) is continuous and differentiable on the open interval \( (a,b) \).

Then there is a point \( c \in (a,b) \) such that

\[ f'(c) = \frac{f(b) - f(a)}{b-a}. \]
This says the tangent to $y = f(x)$ at $x = c$ is parallel to the line through the point $(a, f(a))$ and $(b, f(b))$.

**Proof.**

According the C.H. Edwards, Jr., in *The Historical Development of the Calculus*, p. 314, our proof is due to O. Bonnet (1819-1890). We split the proof into cases.

**Special Case.** $f(a) = f(b)$. (Rolle's Theorem).

![Graph](image)

We need to find $c \in (a, b)$ such that $f'(c) = 0$. If $f(x) = \text{constant}$, any point $c$ will work. So let's assume that $f(x) \neq f(a)$ for some $x \in (a, b)$. Let $f(x) > f(a)$. (Otherwise we will make a similar argument). Define

$$f(c) = \max \left\{ f(x) \mid x \in [a, b] \right\}.$$  

The point $c \in (a, b)$ exists by the Weierstrass Theorem on p. 47 of the notes. It follows that $c = a$ and $c = b$.

**Claim.** The First Derivative Test.

$$f'(c) = 0.$$

**Proof of Claim.**

$$\frac{f(c+h) - f(c)}{h} = \begin{cases} 
\text{number } \geq 0, \text{ for } h \text{ small positive} \\
\text{number } \leq 0, \text{ for } h \text{ small negative}
\end{cases}$$

The only way that this difference quotient can have a two-sided limit as $h$ approaches 0 is that the limit is 0. Why? Limits preserve inequalities.

**General Case.** $f(a) \neq f(b)$.

Look at a new function which Rolle (or Bonnet) pulled out of his magic hat:

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

Then $g$ satisfies the hypothesis of the special case (Rolle's Theorem); i.e.,

*Note: equation of line joining $(a, f(a))$ and $(b, f(b))$ is $y - f(a) = \frac{f(b) - f(a)}{b-a} (x - a)$.*

So we've just subtracted the line $y = f(a) + \frac{f(b) - f(a)}{b-a} (x-a)$ from $f(x) + f(a)$. 

57
To see this, note that
\[ g(a) = g(b) = f(b) - \frac{f(b) - f(a)}{b-a} (b - a) = f(b) - \left\{ f(b) - f(a) \right\} = f(a) = g(a). \]

So by the special case (Rolle's Theorem), we know there is a point \( c \in (a,b) \) so that \( g'(c) = 0 \).

Now we compute \( g'(c) \) remembering that \( a, b \) are constants and \( \frac{d}{dx} (x - a) = 1 \). We obtain:

\[ 0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b-a}. \]

This proves the mean value theorem.

**Corollary.** Suppose that \( f: [a,b] \rightarrow \mathbb{R} \) is a continuous function and suppose that \( f \) is differentiable on the open interval \((a,b)\). If \( f'(x) > 0 \) for all \( x \in (a,b) \), then \( f(x) \) is monotone strictly increasing on \([a,b]\); i.e., for every pair \( u, v \in [a,b] \) with \( u < v \), we have \( f(u) < f(v) \).

**Proof.**

By the mean value theorem

\[ f(v) - f(u) \]

\[ v-u \]

for some \( c \in (u,v) \).

Since \( f'(c) > 0 \) and \( (v-u) > 0 \), it follows that \( f(v) - f(u) \) must also be positive.

**Remarks.** Assume that \( f: [a,b] \rightarrow \mathbb{R} \) is a continuous function and suppose that \( f \) is differentiable on the open interval \((a,b)\). Similarly one can show that if \( f'(x) < 0 \) for all \( x \in (a,b) \), then \( f(x) \) is monotone strictly decreasing on \((a,b)\); i.e., \( u < v \) implies \( f(u) > f(v) \). Here \( u, v \in [a,b] \).

And if \( f'(x) \geq 0 \) for all \( x \in (a,b) \), then \( f(x) \) is increasing but not necessarily strictly increasing; i.e., \( u \leq v \) implies \( f(u) \leq f(v) \).

Similarly if \( f'(x) \leq 0 \), then \( f(x) \) is decreasing; i.e., \( u \leq v \) implies \( f(u) \geq f(v) \).

The following corollary of the mean value theorem is useful when evaluating integrals.

**Important Special Case.** Suppose that \( f'(x) = 0 \) for all \( x \in (a,b) \). Then \( f(x) \) is constant for all \( x \in [a,b] \). For \( f(x) \) is both increasing and decreasing which means \( f(u) = f(v) \) for all \( u, v \in (a,b) \). We can take \( u = a \) and see that \( f(v) = f(a) \) for all \( v \in [a,b] \). Thus \( f \) must be constant.

**Theorem.** \( f: [a,b] \rightarrow \mathbb{R} \) continuous. Suppose \( f''(x) > 0 \) for all \( x \in (a,b) \). Then \( f(x) \) is strictly convex up. This means the graph of \( y = f(x) \) \( u \leq x \leq v \), lies below the secant line connecting \((u, f(u))\) and \((v, f(v))\).
Proof.
The line $L$ is given by
$$L(x) = f(u) + \left\{\frac{f(v) - f(u)}{v-u}\right\}(x-u).$$
Set $g(x) = L(x) - f(x)$. Our goal is to show that $g(x) > 0$ for all $x \in (u,v)$.

Use the mean value theorem on $f$ to see that
$$g'(x) = f'(c) - f'(x)$$
for some $c \in (u,v)$.

Use the mean value theorem on $f'$ to see that
$$g''(x) = f''(d)(c-x),$$
for some $d$ between $c$ and $x$.

If $x \in (a,c)$, then the fact that $f''(d) > 0$, implies that $g$ is strictly increasing on $(a,c]$. Similarly, if $x \in (c,b)$, we find that $g$ is strictly decreasing on $(c,b)$. Now $g(u)=0$ and $g(v)=0$. So we see that $g(x) > 0$ for all $x \in (u,v)$, which proves the theorem.

Next we will discuss inverse functions and their derivatives and apply the result to $\exp(x)$ and $\log(x)$, $x^2$ and $\sqrt{x}$, $x^n$ and $x^{1/n}$.

**Theorem on Differentiation of Inverse Functions.** Suppose $f:[a,b] \to \mathbb{R}$ is continuous and such that $f'(x) > 0$ for all $x \in (a,b)$. Then there is an inverse function $g:[f(a),f(b)] \to [a,b]$ such that $g(f(x)) = x$ for all $x \in [a,b]$ and $f(g(y)) = y$ for all $y \in [f(a),f(b)]$. We write $g = f^{-1}$. The inverse function is differentiable on $(f(a),f(b))$ and
$$g'(y) = \frac{1}{f'(g(y))},$$
for all $y \in [f(a),f(b)]$.

Before proving this theorem, let's give a few examples.
Example 1. 

\[ f(x) = x^n, \quad n=1,2,3,4, \ldots \]
\[ g(y) = y^{1/n}, \text{ defined on } y \geq 0, \text{ if } n \text{ is even, and for all } y \text{ if } n \text{ is odd.} \]

Then \[ g'(y) = \frac{1}{n} y^{\frac{n-1}{n}}. \]

Proof using the Theorem.

Assume \( y > 0 \) if \( n \) is even and that \( y \neq 0 \). Since \( g(y) \) is the inverse function to \( f(x) = x^n \), we have

\[ g'(y) = \frac{1}{f'(g(y))} = \frac{1}{\frac{1}{n} y^{\frac{1}{n}} (n-1)} = \frac{1}{n} y \]

Using this rule, we can show the following formula.

Example 2. Suppose that \( r \in \mathbb{Q} \) and \( r = \frac{p}{q} \), with integers \( p \) and \( q \), \( q > 0 \), with \( p \) and \( q \) having no common factors. Then define

\[ f(x) = x^r = \left( x^\frac{1}{q} \right)^p, \quad \text{for } x \geq 0 \text{ if } q \text{ is even,} \]

with derivative given by the usual formula

\[ f'(x) = r x^{r-1}. \]

Exercise. Prove This!

Proof of the Theorem on Differentiation of Inverse Functions.

1) Definition of \( f^{-1} \).

We are assuming that \( f: [a,b] \to \mathbb{R} \) is continuous with positive derivative \( f'(x) \) for all \( x \in (a,b) \). Thus \( y = f(x) \) is strictly increasing on \([a,b]\). For \( a < u < v < b \) implies (by the mean value theorem) that \( f(a) \leq f(u) < f(v) \leq f(b) \). This shows that \( f \) is one-to-one on \([a,b]\). By the intermediate value theorem, we know that \( f \) maps \([a,b]\) onto the interval \([f(a), f(b)]\). Thus the inverse function \( g: [f(a), f(b)] \to [a,b] \) is well defined, also 1-1 and onto. For we define \( g(y) = f^{-1}(y) = x \) if and only if \( x = f(y) \).

Note that \( x \) is unique since \( f \) is 1-1. And we know that such an \( x \) exists for all \( y \in [f(a), f(b)] \) because \( f \) maps onto. Recall Lang’s Exercise 7, p. 8.

2) The inverse function is also increasing.

Proof by Contradiction.

Suppose that \( f(a) \leq s < r \leq f(b) \). Show that \( f^{-1}(s) < f^{-1}(r) \). Well, otherwise we have \( f^{-1}(s) = f^{-1}(r) \). Since \( f \) is increasing and erases \( f^{-1} \), we can apply \( f \) to both sides of this inequality and obtain

\[ s = f(f^{-1}(s)) \leq f(f^{-1}(r)) = r. \]

This contradicts the assumption that \( s < r \). So \( f^{-1} \) must be increasing.
3) **Continuity of the Inverse Function.**

Given $\varepsilon > 0$, look at the picture to find $\delta = \min(\delta_1, \delta_2)$.

To show that $f^{-1}$ is continuous at $y = r = f(c)$, we proceed as follows. Given $\varepsilon > 0$, let 

$$f(r + \delta_2) = c + \varepsilon \quad \text{and} \quad f(r - \delta_1) = c - \varepsilon.$$ 

Set $\delta = \min(\delta_1, \delta_2)$. Then we claim that 

$$|y - f(c)| < \delta \implies |f^{-1}(y) - c| < \varepsilon.$$ 

To see this, note that $|y - f(c)| < \delta$ means that $-\delta < y - f(c) < \delta$.

This implies 

$$r - \delta_1 < f(c) - \delta < y < f(c) + \delta < r + \delta_2.$$

Therefore 

$$c - \varepsilon = f^{-1}(r - \delta_1) < f^{-1}(y) < f^{-1}(r + \delta_2) = c + \varepsilon.$$

So 

$$-\varepsilon < f^{-1}(y) - c < \varepsilon.$$ 

This means that 

$$|f^{-1}(y) - c| < \varepsilon,$$ 

when $|y - f(c)| < \delta$.

The continuity of $f^{-1}$ is thus proved.

4) **Formula for the derivative of the inverse function.**

Let $f(c) = r$ and $f^{-1}(r) = c$. Then 

$$(f^{-1})'(r) = \lim_{y \to r} \frac{f^{-1}(y) - f^{-1}(r)}{y - r} = \lim_{y \to r} \frac{x - c}{f(x) - f(c)}$$ 

$$= \lim_{x \to c} \frac{1}{f'(c) - f'(f^{-1}(r))} = \frac{1}{f'(c)} = \frac{1}{f'(f^{-1}(r))}.$$
We have used the continuity of the inverse function to see that $y$ approaches $r$ implies that $x=f^{-1}(y)$ approaches $c=f^{-1}(r)$.

**Old Favorite Functions and Their Derivatives.**

**Exponential and (Natural) Logarithm.**

**Definition.**

$$e^x = \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots \quad (1)$$

This power series converges for all $x \in \mathbb{R}$. See Lang, p. 237. In fact, this power series even converges for matrix $x$ or complex numbers $x$. Assuming that we can differentiate the power series term-by-term (see Lang, p. 237), we have

$$\frac{d}{dx} e^x = 0 + 1 + \frac{2x}{2} + \frac{3x^2}{6} + \cdots + \frac{nx^{n-1}}{n!} + \cdots$$

$$= 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots = e^x.$$

Also we see that

$$e^0 = 1 + 0 + 0 + 0 + \cdots = 1.$$

**Conclusion.** Assuming a few results on power series, $f(x)e^x$ satisfies

$$f'(x) = f(x), \quad f(0) = 1. \quad (2)$$

This is a first order linear ordinary differential equation with initial condition. If you remember anything from that part of your calculus course which covered differential equations, you should know that the solution to such a problem is unique. Lang defines $\exp(x)$ by (2) and shows that the solution to (2) is unique. See Lang, pages 78-79.

**Facts About the Exponential Function.**

We shall prove these things using our definition (1). Lang gives different proofs using his definition (2).

1) **exp takes addition to multiplication.**

$$e^{x+y} = e^x \cdot e^y.$$

**Proof.**

Multiply out the power series and use the binomial theorem:

$$\sum_{n=0}^{\infty} \frac{a^n}{n!} \sum_{m=0}^{\infty} \frac{b^m}{m!} = \sum_{k=0}^{\infty} \left( \sum_{m+n=k} \frac{a^m b^n}{n! m!} \right), \text{ regrouping terms}$$

62
\[
\sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \frac{a^n b^{k-n}}{n!(k-n)!}, \text{ setting } m = k - n,
\]

\[
= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{\infty} \frac{k!}{n!(k-n)!} a^n b^{k-n}, \text{ multiplying by } 1 = \frac{k!}{k!},
\]

\[
= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{\infty} \frac{k!}{n!(k-n)!} a^n b^{k-n}
\]

\[
= \sum_{k=0}^{\infty} \frac{1}{k!} (a+b)^k, \text{ by the binomial theorem.}
\]

2) **The exponential \( \exp(x) \) never vanishes.**

\[ e^x \neq 0 \quad \text{and} \quad e^{-x} = \frac{1}{e^x}. \]

In fact \( e^x > 0 \) for all \( x \in \mathbb{R} \).

**Proof.**

By the preceding result, we know that \( e^x e^{-x} = e^0 = 1 \). Thus we have proved that \( e^{-x} = \frac{1}{e^x} \). It follows from this that \( e^x \neq 0 \), since

\[ e^x = 0 \text{ would imply } 0 = e^x e^{-x} = 1. \]

Contradiction.

Since the Taylor series is positive if \( x \leq 0 \) and \( e^x = 1/e^x \), it follows that \( e^x > 0 \) for all \( x \).

3) **The exponential grows faster than any power of \( x \) as \( x \) goes to infinity.**

\[
\lim_{x \to \infty} \frac{e^x}{x^n} = \infty \text{ for all } n \in \mathbb{Z}^+.\]

**Proof.**

By the Taylor expansion for \( x > 0 \), we have

\[ e^x > \frac{x^{n+1}}{(n+1)!} \text{ (one term of the power series).} \]

Therefore when \( x > 0 \), we have

\[
\frac{e^x}{x^n} > \frac{x^{n+1}}{(n+1)! x^n} = \frac{1}{(n+1)!} x \quad \text{as} \quad x \to \infty, \text{ as } x \to \infty.
\]

Here \( n \) is fixed.

Now we can graph \( y = e^x \). We know that it is strictly increasing and \( y" = y > 0 \) implies it is convex up. The function \( y = e^x \) goes to \( \infty \) as \( x \) goes to \( \infty \). Since \( e^{-x} = 1/e^x \), \( y \) goes to 0 as \( x \) goes to \(-\infty\).
4) The number $e$ is irrational.

Proof by Contradiction.

Suppose instead that $e = \frac{a}{b}$, with $a, b \in \mathbb{Z}^+$. Let $k < b$ and look at the $k$th partial sum in the series for $e$. If we subtract the $k$th partial sum from $e$, we get

$$e - \sum_{n=0}^{k} \frac{1}{n!} = a - \sum_{n=0}^{k} \frac{1}{n!}$$

$$= \sum_{n=k+1}^{\infty} \frac{1}{n!} = \text{the remaining terms in the series for } e.$$

Multiplying by $k!$ we find

$$k! \left| \frac{a}{b} - \sum_{n=0}^{k} \frac{1}{n!} \right| = k! \sum_{n=k+1}^{\infty} \frac{1}{n!}.$$

Now the left hand side is in $\mathbb{Z}^+$ while (by the argument below) the right hand side is less than 1. This is a contradiction, proving that $e$ is irrational.

To see that the right hand side of the last formula is less than 1, proceed as follows.

$$k! \sum_{n=k+1}^{\infty} \frac{1}{n!} = \frac{k!}{(k+1)!} + \frac{k!}{(k+2)!} + \frac{k!}{(k+3)!} + \cdots$$

$$= \frac{1}{k+1} + \frac{1}{(k+1)(k+2)} + \frac{1}{(k+1)(k+2)(k+3)} + \cdots$$

64
\[
< \frac{1}{k+1} + \frac{1}{(k+1)^2} + \frac{1}{(k+1)^3} + \cdots \\
= \sum_{n=1}^{\infty} \frac{1}{(k+1)^n} = \frac{1}{k+1} \sum_{n=1}^{\infty} \frac{1}{k^n} = \frac{1}{k} ,
\]

by the formula for the geometric series.

Lambert proved that e and \(\pi\) are irrational in 1761. Gelfond proved \(e^\pi\) is irrational in 1929. See Hardy and Wright, *Introduction to the Theory of Numbers*, p. 46, for the proof of irrationality of \(\pi\). In fact both e and \(\pi\) are transcendental; i.e., they are not roots of a polynomial with rational coefficients. Note that rational numbers are roots of 1st degree polynomials with rational coefficients.

There is a whole branch of number theory devoted to studying such questions. Recently a French mathematician (Apéry) showed that the following value of the Riemann zeta function

\[
\zeta(3) = \sum_{n=1}^{\infty} n^3 \text{ is irrational.}
\]

Apéry was an unknown older French mathematician when he found his proof and no one believed him at first. See Van Der Poorten, "A proof that Euler missed ...", *Mathematical Intelligencer*, 1, (1979), 195-203. Euler showed that

\[
\zeta(2) = \sum_{n=1}^{\infty} n^2 = \frac{\pi^2}{6} \quad \text{and} \quad \zeta(4) = \sum_{n=1}^{\infty} n^4 = \frac{\pi^4}{90} ,
\]

and similar formulas for \(\zeta(2n), \; n=3,4,5,6, \ldots\) .

5) \((e^x)^y = e^{xy}\).

We postpone the proof until we have discussed the logarithm.

---

**The Logarithm (Natural).**

Here we reverse the treatment of exp and log in most calculus books. Note that \(e^{\cdot}:\mathbb{R} \rightarrow (0,\infty)\) is an onto map by property 3) above, since \(e^x\) certainly goes to \(\infty\) as \(x\) goes to \(\infty\) and \(e^{-x} = \frac{1}{e^x}\) must go to 0 as \(x\) goes to \(-\infty\). Moreover since it is continuous, by the intermediate value theorem, it must hit every point in \((0,\infty)\). Since the derivative of \(e^x\) is positive, the function \(f(x) = e^x\) satisfies the hypotheses of our theorem on inverse functions on p. 59. The function \(e^x\) is strictly increasing (and the graph is convex up).

**(Notes**

So we can define an inverse function
**Definition.** \( g(y) = \log y = f^1(y) = x \) if \( y = f(x) = e^x \).

So \( \log: (0,\infty) \rightarrow \mathbb{R} \) is 1-1, onto, strictly increasing function.

We write \( \log y \) instead of \( \ln y \) or \( \log_e y \) because this is the only \( \log \) for us. So now most of the properties of \( e^x \) translate to properties of \( \log \).

**Properties of the Logarithm.**

0) **The Differential Equation.**

\( g(y) = \log y \) satisfies the differential equation:

\[
g'(y) = \frac{1}{y} \quad \text{and} \quad g(1) = 0.
\]

**Proof.**

We use the theorem on differentiating inverse functions on p. 59 of the Notes to see, if \( f(x) = e^x = y \),

\[
g'(y) = f'(g(y)) = \frac{1}{f(g(y))} = \frac{1}{y}, \quad \text{since} \quad f(g(y)) = y.
\]

Since \( f(0) = 1 \), it follows that \( g(1) = 0 \).

This property gives the usual calculus book’s definition of \( \log y \) as

\[
\log y = \int_1^y \frac{1}{t} \, dt.
\]

1) **The log takes multiplication to addition.**

\[ \log(u \cdot v) = \log u + \log v. \]

**Proof.**

To see this use property 1) of \( \exp(x) \). This says \( \exp(x+y) = \exp(x) \cdot \exp(y) \).

Let \( u = e^x \) and \( v = e^y \).

Then take logs of both sides of the equation:

\[ \log(e^{x+y}) = \log(e^xe^y). \]

Since \( \log \) and \( \exp \) are inverse, we have

\[ \log u + \log v = x + y = \log(e^xe^y) = \log(uv). \]

Now we can graph \( y = \log x \). Since \( y' = 1/x \) is positive for \( x > 0 \), we know the function is strictly increasing. And \( y'' = -\frac{1}{x^2} < 0 \), for \( x > 0 \). This means the graph is convex down. We know \( \log 1 = 0 \) and \( \log e = 1 \). Recall that \( e \approx 2.71828 \ 1828 \ 459045 \ ... \) It is fairly easy to remember the first few digits thanks to all the repetition.

66
**Definition.** The **Power Function.**

If \( a > 0 \), define

\[
a^x = \exp(x \log a), \quad \text{for all } x \in \mathbb{R}.
\]

This is another way to define \( a^x \) for \( x \in \mathbb{R} \). It is not hard to see that this definition gives the same answer as that on p. 60 of the notes.

**2) Properties of Powers.**

a) \( \log (a^x) = x \log a \)

b) \( a^x \cdot a^y = a^{x+y} \)

c) \( (a^x)^y = a^{xy} \).

**Proof.**

a) By the definition of \( a^x \) and the fact that \( e^x \) and \( \log \) are inverse functions, we have:

\[
\log(a^x) = \log(e^{x \log a}) = x \log a.
\]

b) Prove this backwards by taking logs of both sides. If we get an identity, we are done, since \( \log \) is 1-1. We can write a more careful proof as follows. We need to use the definition of \( a^x \) as well as the first property of logs to see the following.

\[
\log (a^x \cdot a^y) = \log a^x + \log a^y = x \log a + y \log a = (x + y) \log a.
\]

But this is \( \log(a^{x+y}) \). So we have

\[
\log (a^x \cdot a^y) = \log(a^{x+y}),
\]

which is the log of the desired equality. Exponentiate it to prove b).
c) Again we take logs of both sides and use the definitions to see
\[ \log((a^x)^y) = y \log(a^x) = y \log a = \log (a^y). \]

So we have proved
\[ \log((a^x)^y) = \log (a^{xy}). \]
Exponentiate both sides to get the desired equality.

In particular, c) implies that
\[ (a^{1/n})^n = a \quad \text{for all} \quad n \in \mathbb{Z}^+, \quad a > 0. \]

**Example.** \[ \lim_{h \to 0} (1 + h)^{1/h} = e. \]

**Proof.**
Take logs. Set \( g(x) = \log x. \) Note that
\[ \log (1+h)^{1/h} = \frac{1}{h} \log(1+h) = \frac{\log(1+h) - \log 1}{h} = g'(1) = 1, \]

since \( \frac{d \log x}{dx} = \frac{1}{x} \) and this has the value 1 at \( x=1. \)

So reversing this by exponentiating we find that the limit is \( e^1 = e. \)

Lang does a certain number of examples of this sort, for which it helps to take logs to find limits. You will find a similar problem in #17 on p. 88 of Lang, which gives an alternative definition of \( e^x = \lim_{n \to \infty} (1 + \frac{x}{n})^n. \)

Later, if we are lucky we will have time to do Stirling's formula:
\[ \lim_{n \to \infty} \frac{n!}{\sqrt{2\pi n} \, n^n e^{-n}} = 1, \] which is written as
\[ n! \sim \sqrt{2\pi n} \, n^n e^{-n}, \text{ as } n \to \infty. \]
The symbol \( \sim \) is read as "is asymptotic to".

3) \[ \lim_{y \to \infty} \log \frac{y}{y} = 0. \]

**Proof.**
Set \( y = e^x. \) Note that
\[ y \to \infty \quad \text{if and only if} \quad x \to \infty. \]

So
\[ \frac{\log y}{y} = \frac{x}{e^x} \to 0, \text{ as } x \to \infty. \]
**Complex Numbers and Trigonometric Functions.**

Note that $-1$ is not a square of a real number. So to solve the equation $x^2+1=0$ we adjoin a number $i$ to the real numbers.

Let $i = \sqrt{-1}$. Then we define the set $\mathbb{C}$ of complex numbers to be:

$$\mathbb{C} = \{ z = x + iy \mid x, y \in \mathbb{R} \}.$$ 

If $z = x + iy$ with $x, y \in \mathbb{R}$, we say $x$ is the real part of $z$ and $y$ the imaginary part; $x = \text{Re} \ z$ and $y = \text{Im} \ z$.

We can visualize the complex numbers as points in the plane with $z = x + iy$, $x, y \in \mathbb{R}$, corresponding to the point $(x, y)$.

The set $\mathbb{C}$ forms a field; i.e., it satisfies the 9 field axioms listed on page 21 of these notes. If $z = x + iy$ and $w = u + iv$, then

$z + w = (x + u) + i(y + v)$ (add them like vectors);

$z \cdot w = (xu - yv) + i(xy + yu)$, obtained by multiplying as usual and remembering $i^2 = -1$;

and if $w \neq 0$, you can find

$$\frac{z}{w} = \frac{(x + iy) \cdot (u - iv)}{(u + iv)(u - iv)} = \frac{xu + yv + i(yu - xv)}{u^2 + v^2} = \frac{xu + yv}{u^2 + v^2} + i \frac{yu - xv}{u^2 + v^2}.$$  

To say that $\mathbb{C}$ forms a field means you can add, subtract, multiply and divide by non-0 numbers with the usual laws (such as the distributive and commutative).

**Definition.** **Absolute Value** $|z| = \sqrt{x^2 + y^2}$ if $z = x + iy$, for $x, y \in \mathbb{R}$.

**Complex Conjugate** $\overline{z} = x - iy$.

Then $|z|^2 = zz$. 

Note: $\mathbb{C}$ is not an ordered field—just a field.
This absolute value has all the usual properties.
1) \( |z| = 0 \) for all \( z \in \mathbb{C} \) and \( |z| = 0 \) if and only if \( z = 0 \).

(A complex number is 0 if and only if both real and imaginary parts are 0).

2) \( |zw| = |z| \cdot |w| \).
3) triangle inequality \( |z + w| \leq |z| + |w| \).

See Lang, p. 37 for the proof.

This time there really is a triangle!

If \( z \in \mathbb{C} \), we can define
\[
\exp(z) = e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad \text{a complex power series.}
\]

Then we have \( e^{z+w} = e^z e^w \) just as in the real case. You multiply complex power series the same way as the real ones. And the binomial theorem is true for complex \((z+w)^n\), \( n=1,2,3,\ldots \).

So \( e^z = e^{x+iy} = e^x e^{iy} \).

Let's look at
\[
e^{iy} = \sum_{n=0}^{\infty} \frac{(iy)^n}{n!} = 1 + iy + \frac{(iy)^2}{2} + \frac{(iy)^3}{6} + \frac{(iy)^4}{24} + \cdots
\]
\[
= 1 + iy - \frac{y^2}{2} - i \frac{y^3}{6} + \frac{y^4}{24} + \cdots
\]
\[ = \left\{ 1 - \frac{y^2}{2} + \frac{y^4}{24} - \cdots \right\} + i \left\{ y - \frac{y^3}{6} + \frac{y^5}{5!} - \cdots \right\} \]

\[ = \cos y + i \sin y. \]

Here we are defining sine and cosine by their Taylor series. You can deduce the standard trigonometric identities from \( e^{a+b} = e^a e^b \).

One finds that \( |e^{ix}| = 1 \), for \( x \in \mathbb{R} \).

**Proof.**

\[ |e^{ix}|^2 = e^{ix} e^{-ix} = e^0 = 1, \text{ for } x \in \mathbb{R}. \]

If we define sine and cosine by their Taylor series, we see that

\[
c(x) = \cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \quad \text{and} \quad s(x) = \sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.
\]

You can check that

\[
c'(x) = -s(x) \quad \text{and} \quad s'(x) = c(x), \quad c(0) = 1 \quad \text{and} \quad s(0) = 0.
\]

Lang starts from here on p. 90 to deduce the usual formulas (1)-(5) of Lang, p. 91, plus the fact that \( \sin x \) and \( \cos x \) are periodic of period \( 2\pi \).

Personally I prefer to deduce (1)-(5) on p. 91 of Lang from the fact that

\[ e^a e^b = e^{a+b}. \]

This gives \( \sin^2 x + \cos^2 x = 1 \) and the picture above. Identities (2) and (3) from Lang, p. 91 are clear from the Taylor series. (4) and (5) are Exercises.

This uses the fact that an angle of \( 2\pi \) radians sweeps through the entire circle. It is perhaps amazing that this can be proved without reference to the picture. Lang does this on pages 92-3. He defines \( \pi/2 \) to be the first positive zero of cosine (rather than defining \( 2\pi \) to be the arc length of a circle of radius 1). Of course he has to
first show that a zero exists.

The treatment of elementary functions by power series can be found, for example, in P. Dienes, *The Taylor Series*, Chapter 4.

Any finite combination via sum, product, composite, difference, quotient of the functions

\[
\begin{align*}
& a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad a_i \in \mathbb{C}, \\
& a^x, \log x, \ e^x, \ \sin x, \ \cos x, \ \sin^{-1} x, \ \cos^{-1} x
\end{align*}
\]

are called **elementary functions**. Of course, they might be complicated such as

\[
\exp(\pi x) - 2001 \log x - 500 \ x^2/3 + \sqrt{x}.
\]

These are the functions considered in calculus.

Still there are other favorite functions which are used by applied mathematicians for various purposes; e.g., in the solution of partial differential equations of mathematical physics. A good reference for these is another Dover book: N.N. Lebedev, *Special Functions*. Mathematica, Maple and Matlab know most special functions. They will even provide beautiful color graphs.

We have already seen the **gamma function**:

\[
\Gamma(s) = \int_{t=0}^{\infty} e^{-t} t^{s-1} \, dt, \quad \text{for } \Re s > 0.
\]

The incomplete gamma function is the chi-squared distribution which is very important in statistics. Legendre polynomials \( P_n(x) \) can be found in Lang, Exercise 2, p. 107. They arise in problems with spherical symmetry; e.g., solution of the Schrödinger equation for the hydrogen atom; study of earthquakes, the sun's magnetic field. See my book or any Math'110 book.

Not every function you can think of is expressible by its Taylor series. There is an example in Exercise 6 on p. 82 of Lang. This is an infinitely differentiable function which is not represented by its Taylor series at the points \( a \) and \( b \):

\[
f(x) = \begin{cases} 
  \exp\left(-\frac{1}{(x-a)(x-b)}\right) & \text{, } a < x < b, \\
  0 & \text{, otherwise.}
\end{cases}
\]

This function is graphed below. It can be used to provide \( C^\infty \)-glue. It glues 0 in an infinitely differentiable way to non-zero values.