1 Introduction

The material in these notes is assumed to be known by the students taking this course. The intention is to freshen up your memory. Understanding this material is key to a proper understanding of the topics we will cover in the following weeks.

2 Preliminaries

2.1 Notation

$\mathbb{R}$ or equivalently $\mathbb{R}^1$ denotes the *real number line*. $\mathbb{R}^1$ is sometimes referred to as 1-dimensional Euclidean space. The 2-dimensional Euclidean space is represented by $\mathbb{R}^2$, the 3-dimensional space by $\mathbb{R}^3$, and, more generally, k-dimensional Euclidean space is represented by $\mathbb{R}^k$.

Suppose $a \in \mathbb{R}^1$ and $b \in \mathbb{R}^2$ with $a < b$.

A *closed interval* $[a, b]$ is the subset of $\mathbb{R}^1$ whose elements are greater than or equal to a and less than or equal to b.

An *open interval* (a, b) is the subset of $\mathbb{R}^1$ whose elements are greater than a and less than b.

A *half open interval* $[a, b)$ is the subset of $\mathbb{R}^1$ whose elements are greater than or equal to a and less than b.

A *half open interval* $(a, b]$ is the subset of $\mathbb{R}^1$ whose elements are greater than a and less than or equal to b.

A closed interval includes its endpoints while an open interval does not.

2.2 Functions

Suppose we have two sets X and Y. A *function* f defined on X takes each element of X and assigns a unique element of Y to it. The set X is called the *domain* of f and Y is called the *range* of f.

The notation

$$f : X \to Y$$

reads "f maps X to Y". Here the set X is the domain of f and the set Y is the range.

A *real-valued function* f is a function whose domain and range are sets of real numbers.

2.2.1 Polynomials

A *monomial* is a function f that can be written as $f(x) = ax^k$ for some number a and some positive integer k. k is called the *degree* of the monomial. For instance,

$$f(x) = 5.4x^3$$

is a 3rd degree monomial.

A *polynomial* is a function f that can be written as the sum of monomials. The degree of a polynomial is the largest degree of any of the monomials that the polynomial is composed of. For instance,

$$f(x) = 2.5x^8 + 0.38x^5 - 32.4x + 302$$

is an 8th degree polynomial.

A *linear function* f has the following form

$$f(x) = mx + b.$$

The graph of a linear function is a straight line. The slope of the line of is m and its y-intercept is b.

3 Basic Differential Calculus

3.1 First Derivatives

There are multiple types of notation that can be used to represennt a derivative of a function $f(x)$. The notation $f'(x)$ is one way to represent the first derivative of $f(x)$. This notation makes clear that the derivative is a new function of x. The notation

$$\frac{d}{dx} f(x)$$

is another way to represent the first derivative of $f(x)$. This notation emphasizes that the derivative is being calculated from $f(x)$ with respect to x.

The first derivative of f at x is defined as

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

if this limit exists. If this limit exists we say that f is differentiable at x.

3.2 Rules for Calculating Derivatives

In what follows in this subsection f and g are assumed to be differentiable functions of x and k is taken to be an arbitrary constant.

3.2.1 Rule 1

$$\frac{d}{dx}x^k = kx^{k-1}$$

Example 3.1.

$$f(x) = x^3$$
$$\frac{d}{dx}f(x) = 3x^2$$

3.2.2 Rule 2

$$\frac{d}{dx}\left(f(x) \pm g(x)\right) = f'(x) \pm g'(x)$$

Example 3.2.

$$f(x) = x^4 \qquad g(x) = x^2$$
$$\frac{d}{dx}\left(f(x) - g(x)\right) = 4x^3 - 2x$$

3.2.3 Rule 3

$$\frac{d}{dx}(kf(x)) = kf'(x)$$

Example 3.3.

$$k = 5.3 \qquad f(x) = x^2$$
$$\frac{d}{dx}(kf(x)) = 10.6x$$

3.2.4 Rule 4

Rule 4 is called the *product rule*.

$$\frac{d}{dx}(f(x) \cdot g(x)) = f'(x)g(x) + f(x)g'(x)$$

Example 3.4.

$$f(x) = 2x^3 \qquad g(x) = 3x$$

$$\begin{aligned} \frac{d}{dx}(f(x) \cdot g(x)) &= (6x^2)(3x) + (2x^3)3 \\ &= 18x^3 + 6x^3 \\ &= 24x^3 \end{aligned}$$

Note that we would get the same answer if we multiplied $f(x)$ by $g(x)$ to get $6x^4$ and then took the derivative of this.

Note that Rule 3 is a special case of Rule 4.

3.2.5 Rule 5

Rule 5 is called the *quotient rule*.

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$

Example 3.5.

$$f(x) = 2x^3 \qquad g(x) = 3x$$

$$\begin{aligned} \frac{d}{dx}\frac{f(x)}{g(x)} &= \frac{(6x^2)(3x) - (2x^3)3}{(3x)^2} \\ &= \frac{18x^3 - 6x^3}{9x^2} \\ &= \frac{12x^3}{9x^2} \\ &= \frac{4}{3}x \end{aligned}$$

Note that we would get the same result by first dividing $2x^3$ by $3x$ to get $\frac{2}{3}x^2$ and then differentiating.

3.2.6 Rule 6

Rule 6 is called the *power rule.*

$$\frac{d}{dx}\left(f(x)^k\right) = k(f(x))^{k-1}f'(x)$$

Example 3.6.

$$f(x) = (x^2 - 3) \qquad k = 2$$

$$\begin{aligned}\frac{d}{dx}\left(f(x)^k\right) &= 2(x^2 - 3) \cdot 2x \\ &= 4x^3 - 12x\end{aligned}$$

3.2.7 Rule 7

Rule 7 is called the *chain rule.*

$$\frac{d}{dx}\left(g\left(f(x)\right)\right) = g'\left(f(x)\right) \cdot f'(x)$$

Example 3.7. *To make this example a bit more interesting note that*

$$\frac{d}{dx}\exp(x) = \exp(x)$$

(The $\exp(\cdot)$ *function is introduced below in section 4.2).*

$$f(x) = 2x^3 + 1 \qquad g(x) = \exp(x)$$

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

Note that the power rule is a special case of the chain rule.

3.3 Second and Higher Order Derivatives

Suppose f is a differentiable function defined on $\mathbb{R}^1$. In this case we can also think of $f'(x)$ as a function defined on $\mathbb{R}^1$. As such we can ask whether we $f'(x)$ is differentiable. The *second derivative* of f at x is defined as

$$f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}$$

if this limit exists.

Notation for the second derivative of f at x includes the following: $f''(x)$ or $\frac{d^2}{dx}f(x)$.

3.4 Uses of Derivatives

There are many uses of derivatives but in this class we will be primarily interested in using derivatives to a) find maxima or minima of functions and b) approximate functions.

3.4.1 Minimization and Maximization

If f has a local minimum or maximum at x_0 we know that $f'(x_0) = 0$. If $f'(x_0) = 0$ and $f''(x_0) > 0$ we know that x_0 is a min of f. If $f'(x_0) = 0$ and $f''(x_0) < 0$ we know that x_0 is a max of f. If $f'(x_0) = 0$ and $f''(x_0) = 0$ x_0 can be a min, a max, or neither.

To find the maximum of a function f we calculate $f'(x)$, set it equal to 0, and then solve for x. To check to see that the resulting value is a max we need to check to see that f'' is less than 0 at the value we solved for.

Example 3.8. *We want to find the maximum of $f(x) = -(x-1)^2$. We can rewrite $f(x)$ as $f(x) = -x^2 + 2x - 1$. We can calculate $f'(x)$ as $f'(x) = -2x + 2$. Setting $f'(x) = 0$ and solving for x reveals that $\hat{x} = 1$ is the value of x for which $f(\hat{x}) = 0$. The second derivative of f is $f''(x) = -2$. This is negative for all values of x so we now that $\hat{x} = 1$ is a max of f.*

3.4.2 Taylor Series Approximations

Often it will be useful to approximate a function near some point with an easily manipulated polynomial. Taylor's theorem suggests how we might do this.

Theorem 1 (Taylor's Theorem (1 Dimension)). *Suppose the function f is $k+1$ times differentiable on an open interval I. For any points x and $x+h$ in I there exists a point w between x and $x+h$ such that*

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \cdots + \frac{1}{k!}f^{[k]}(x)h^k + \frac{1}{(k+1)!}f^{[k+1]}(w)h^{k+1}. \tag{1}$$

It can be shown that as h goes to 0 the higher order terms in equation 1 go to 0 much faster than h goes to 0. This means that (for small values of h)

$$f(x+h) \approx f(x) + f'(x)h$$

This is referred to as a *first order Taylor approximation of f at x*. A more accurate approximation to $f(x+h)$ can be constructed for small values of h as:

$$f(x+h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2$$

This is known as a *second order Taylor approximiation of f at x*.

Note that the first order Taylor approximation can be rewritten as:

$$f(x+h) \approx a + bh$$

where $a = f(x)$ and $b = f'(x)$. This highlights the fact that the first order Taylor approximation is a linear function in h.

Similarly, the second order Taylor approximation can be rewritten as:

$$f(x+h) \approx a + bh + \frac{1}{2}ch^2$$

where $a = f(x)$, $b = f'(x)$, and $c = f''(x)$. This highlights the fact that the second order Taylor approximation is a second order polynomial in h.

Example 3.9 (The Delta Method). *The delta method is a method to approximate the mean and variance of a nonlinear function g of a random variable X with known mean and variance. The basic idea is to approximate g at the mean of X with a first order Taylor series. We can then use standard rules for calculating the mean and variance of a linear function of a random variable.*

Suppose $E[X] = \mu$ and $V[X] = \sigma^2$. A first order Taylor series approximation to g at μ is

$$\begin{aligned} g(x) &\approx g(\mu) + (x - \mu)g'(\mu) \\ &\approx (g(\mu) - \mu g'(\mu)) + x g'(\mu) \end{aligned}$$

which is a linear function of x. Recall that for constants a and b, $E[a + bX] = a + bE[X]$, and $V[a+bX] = b^2 V[X]$. In the Taylor series approximation above $(g(\mu) - \mu g'(\mu))$ plays the role of the constant a and $g'(\mu)$ plays the role of the constant b. Thus

$$\begin{aligned} E[g(X)] &\approx (g(\mu) - \mu g'(\mu)) + E[X]g'(\mu) \\ &\approx (g(\mu) - \mu g'(\mu)) + \mu g'(\mu) \\ &\approx g(\mu) \end{aligned}$$

and

$$\begin{aligned} V[g(X)] &\approx (g'(\mu))^2 V[X] \\ &\approx (g'(\mu))^2 \sigma^2 \end{aligned}$$

Note that the delta method is only accurate to the extent that the linear approximation to g is accurate over all high probability regions of X.

Example 3.10 (The Newton Raphson Algorithm). *Suppose we want to find the value of x that maximizes some twice continuously differentiable function $f(x)$.*

Recall

$$f(x + h) \approx a + bh + \frac{1}{2}ch^2$$

where $a = f(x)$, $b = f'(x)$, and $c = f''(x)$. This implies

$$f'(x + h) \approx b + ch.$$

The first order condition for the value of h (denoted $\hat{h}$) that maximizes $f(x + h)$ is

$$0 = b + c\hat{h}$$

Which implies $\hat{h} = -\frac{b}{c}$. In other words, the value that maximizes the second order Taylor approximation to f at x is

$$\begin{aligned} x + \hat{h} &= x - \frac{b}{c} \\ &= x - \frac{1}{f''(x)} f'(x) \end{aligned}$$

With this in mind we can specify the Newton Raphson algorithm for 1 dimensional function optimization.

Algorithm 3.1: NEWTONRAPHSON1D$(f, x_0, tolerance)$

comment: *Find the value $\hat{x}$ of x that maximizes $f(x)$*

$i \leftarrow 0$
while $|f'(x_i)| > tolerance$
do $\begin{cases} i \leftarrow i+1 \\ x_i \leftarrow x_{i-1} - \frac{1}{f''(x_{i-1})} f'(x_{i-1}) \end{cases}$
$\hat{x} \leftarrow x_i$
return $(\hat{x})$

Caution: *Note that the Newton Raphson Algorithm doesn't check the second order conditions necessary for $\hat{x}$ to be a maximizer. This means that if you give the algorithm a bad starting value for x_0 you may end up with a min rather than a max.*

4 Exponents and Logarithms

4.1 Exponential Functions

Exponential functions have the form

$$f(x) = a^x$$

a is called the *base*.

- if x is a positive integer $f(x) = a^x$ means multiply a by itself x times
- $a^0 = 1$ by definition
- $a^{1/x} = \sqrt[x]{a}$
- $a^{n/x} = (\sqrt[x]{a})^n$
- $a^x = \frac{1}{a^{|x|}}$ when $x < 0$

4.2 The Number e

The number e is defined as

$$e \equiv \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

further

$$e^k \equiv \lim_{n \to \infty} \left(1 + \frac{k}{n}\right)^n$$

To 3 decimal places $e \approx 2.718$.
e^k can also be written $\exp(k)$.

4.3 Logarithms

Inverse function undo the operation of another function. If g is the inverse function of f then

$$g(f(x)) = x$$

A logarithmic function is the inverse function of an exponential function.
The base a logarithm $\log_a(\cdot)$ has the following property

$$y = \log_a(x) \iff a^y = x$$

The base a logarithm of x is the power to which one must raise a to get x.
It follows that

$$a^{\log_a(x)} = x$$

and

$$\log_a(a^x) = x$$

In statistics applications, the base e logarithm is often used. The base e logarithm is called the natural logarithm and can be written as either $\log_e(\cdot)$ or $\ln(\cdot)$.

$$\ln(\exp(x)) = x$$

4.4 Properties of Exponential Functions

- $a^r \cdot a^s = a^{r+s}$
- $a^{-r} = \frac{1}{a^r}$
- $\frac{a^r}{a^s} = a^{r-s}$
- $(a^r)^s = a^{rs}$
- $a^0 = 1$

4.5 Properties of Logarithms

- $\log(r \cdot s) = \log(r) + \log(s)$
- $\log(\frac{1}{s}) = -\log(s)$
- $\log(\frac{r}{s}) = \log(r) - \log(s)$
- $\log(r^s) = s\log(r)$
- $\log(1) = 0$

4.6 Calculating Derivatives of $\exp(\cdot)$ and $\ln(\cdot)$

The base e exponential function and logarithm have particularly simple first derivatives

$$\frac{d}{dx}\exp(x) = \exp(x)$$

$$\frac{d}{dx}\ln(x) = \frac{1}{x}$$

If $f(x)$ is a differentiable function then

$$\frac{d}{dx}\exp(f(x)) = \exp(f(x)) \cdot f'(x)$$

$$\frac{d}{dx}\ln(f(x)) = \frac{f'(x)}{f(x)} \quad \text{if } f(x) > 0$$

References

Apostol, Tom M. 1967. *Calculus, Volume I*. New York: Wiley, second edition.

Simon, Carl P., and Lawrence Blume. 1994. *Mathematics for Economists*. New York: W.W. Norton.