Mathcamp: Basic Material

Avidit Acharya

June 5, 2020

Contents

1 Preliminaries

1.1 Mathematical Statements

The table at the bottom of the page lists some common mathematical symbols and their abbreviations. Mathematical statements in this course will seldom involve abbreviations or symbols other than the ones listed in the table (except ones that you surely already know such as $=$, $\leq$, etc.). When new symbols arise, I will explain them. As an example, a typical statement is

$$\forall x \in X \text{ and } \forall y \in Y, \exists z \in Z \text{ s.t. } x + y = z,$$

which you will read "For every x in the set X and every y in the set Y, there is an element z in the set Z such that x plus y equals z." Thus a mathematical statement is nothing more than a statement in the English language (or any other language for that matter), where the vocabulary is limited to words like "for all," "there is," and "such that." The objective is to determine which statements are true, and which are not.

It will be important to note the difference between "if" and "only if." S_1 **if** S_2 means that you can derive S_1 from S_2, but it may not be the case that you can derive S_2 from S_1. On the other hand, S_1 **only if** S_2 means that you can derive S_2 from S_1 but that you may not be able to derive S_1 from S_2. S_1 if and only if S_2 means that S_1 can be derived from S_2 and S_2 can be derived from S_1. When this happens, S_1 and S_2 are equivalent: one statement does not say any more or any less than the other statement.

Every mathematical statement has a negation. The **negation** of statement S_1 is written $\neg S_1$ (read "not S_1"). For example, you can negate the statement

"Every left-handed man in Palo Alto has a beard,"

by presenting a left-handed man in Palo Alto who does not have a beard. You cannot negate this statement by presenting a right-handed man in Palo Alto who has a beard; or by presenting left-handed woman in Palo Alto without a beard. You may find it hard to believe that people make these errors, but they do.

Symbol	How to read it
$\in$	"in the set," or "is in the set" depending on context
$\exists$	"there is a(n)"
$\forall$	"for all" or "for every"
s.t.	"such that"
w.l.o.g.	"without loss of generality"

The statement "S_1 is true if S_2 is true" is equivalent to the statement "$\neg S_2$ is true if $\neg S_1$ is true," or "S_2 is not true if S_1 is not true." These are both the **contrapositive** of the first statement. A statement is always equivalent to its contrapositive.

The **converse** of the statement "S_1 is true if S_2 is true" is the statement "S_2 is true if S_1 is true." A statement is not equivalent to its converse. To see this, let S_1 be the statement "U2 rocks," and S_2 be the statement "All Irish bands rock."

Proofs We will occasionally do proofs. Often we will prove a statement directly from a set of other statements. This is called "**direct proof**." Sometimes we will prove a statement by constructing an object that is claimed to exist. This kind of proof is **constructive**.

Sometimes it will be convenient to prove a statement by proving its contrapositive. If we assume that S_1 is true and would like to prove that S_2 is true, then we can prove this by assuming that S_2 is not true and then showing that this implies that S_1 cannot be true. This is proof by **contradiction**. This method is closely related to the method of proof called **reductio ad absurdum**, which allows us to conclude that the statement S_1 is false if S_1 implies a statement S_2 and its negation $\neg S_2$. Both S_2 and $\neg S_2$ cannot simultaneously be true, so S_1 must be false.

Another common method of proof is proof by **induction**. Suppose you wanted to prove that a sequence of statements S_1, S_2, S_3 ... are all true if S_0 is true, but the sequence never terminates. To prove this by induction, you first show that S_1 is true if S_0 is true. Then you show that for any positive integer k, S_k implies S_{k+1} when S_0 is true. That completes the proof. Under **strong induction** you show that S_1 is true. Then you show that for every positive integer k, $S_1, \ldots, S_k$ together imply S_{k+1}.

There are also other methods of proof. Do a google search.

Sets, numbers, notation etc. A **set** is a "well-defined" collection of elements. This means that I can describe to you what kinds of things are in the set, and you will be able to know exactly whether something is in the set or not. Sometimes, we can describe a set by simply listing out its elements: $A = \{a_1, a_2, a_3, \ldots\}$. Whenever we use curly brackets, that is { }, those are sets and inside the brackets is a list of elements in the set or a mathematical statement describing the common property satisfied by the elements belonging to the set. That is, we may write a set as $\{x \in X : P\}$ which you can read as "the set of $x \in X$ such that P holds," where P is some property.

When I say "number" I always mean a real number, except when I mean a natural number/positive integer. The set of real numbers will be denoted $\mathbb{R}$, which consists of all the numbers you know, except the imaginary numbers (e.g. 32, -0.73, 0, 19/7, 4π, and e^{23} are all in $\mathbb{R}$, but $5i$ is not). When I say that a is **weakly** larger than b, then $a \geq b$; when I say that a is **strictly** larger than b, then $a > b$. The concepts of "weakness" and

"strictness" will extend to other settings where equality is and is not allowed (as when a set may be either a strict or weak subset of another set).

If r is a number and we write $|r|$ (read "absolute value of r") then that means r if r is nonnegative and $-r$ if it is negative. The **triangle inequality**, $|a+b| \leq |a| + |b|$, follows from this definition; here, a and b are numbers. The two sides of the equal sign, $=$, may have numbers or other kinds of objects such as sets, as above. The context will make that clear. The notation ":=" means the left hand side is being defined as the right hand side; vice versa for "=:". Sometimes I will use "$\equiv$" instead of "$=$" to denote that to two sides of the equality are equivalent, or identically equal to each other. One common shorthand that I will use is "$\forall i = 1, 2, 3 \ldots$" which you read as "for every positive integer i."

Exercise 1. A number is **rational** if and only if it can be expressed as the ratio of two integers. Prove, by contradiction, that $\sqrt{2}$ is irrational. *Hint:* If $\sqrt{2}$ is rational, then it can be expressed as the ratio of two integers that are not both even. Express it as such, take squares, etc. and remember that only even numbers have even squares.

The quantity **infinity**, denoted by ∞, is not a number, but I will sometimes treat it informally as one—for example, when I say a set has an "infinite number" of elements. It is a quantity bigger than every number, just as $-\infty$ is smaller than every number. The set of natural numbers is $\mathbb{N} := \{1, 2, 3, ...\}$. The set of integers is $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$. Positive numbers are larger than 0 and negative numbers are smaller. Subscripts such as $+$ or - on a set of numbers indicate nonnegative and nonpositive subsets; when written twice like $++$ or $--$, they indicate positive and negative subsets; so, e.g., $\mathbb{N} = \mathbb{Z}_{++}$.

For any number $n \in \mathbb{N}$, the notation $n!$ (read "n factorial") denotes $n \times (n-1) \times (n-2) \times ... \times 1$, and we define $0! := 1$. The notation $\binom{n}{k}$ (read "n choose k") is

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}$$

and gives the number of ways in which k balls can be chosen from an urn of n balls without repetition and where the order in which they are chosen does not matter.

Exercise 2. Prove **Pascal's rule**; i.e., for all $n \in \mathbb{N}$,

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}, \qquad \forall k = 1, ..., n$$

Theorem 1. (Binomial Theorem) *For all $n \in \mathbb{N}$,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k.$$

Proof. The proof is by induction. The **basis case** is $(x+y)^1 = x+y$, which holds. Now for the **inductive step**: we assume that the result holds for n and must prove it holds for $n+1$.

Note that $(x+y)^{n+1} = x(x+y)^n + y(x+y)^n$. This implies that the **coefficient** of $x^i y^j$ in the **polynomial** $(x+y)^{n+1}$ is given by the sum of coefficients of $x^{i-1}y^j$ and $x^i y^{j-1}$ in the polynomial $(x+y)^n$. Therefore, by the **inductive hypothesis** the coefficient of $x^{n+1-k}y^k$ in $(x+y)^{n+1}$ is $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$. (This follows from Pascal's rule above.) Coefficients of $x^i y^j$ in the polynomial $(x+y)^{n+1}$ such that $i+j \neq n+1$ are zero. Therefore, we have

$$(x+y)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^{n+1-k} y^k,$$

which proves the inductive step. □

1.2 Sets, Relations and Functions

The set A is a **subset** of the set B, written $A \subset B$, if every element of A is also an element of B. Two sets are equal if they are subsets of each other. If A is a set, then the set of all subsets of A is called the **power set** of A and is denoted $\mathcal{P}(A)$. If A is a set with a finite number of elements then $|A|$ (read "cardinality of A") denotes the number of elements in A. The set B where $|B| = 0$ is unique, it is called the **empty set**, and it is denoted $\emptyset$. You should realize that for every set A, we have $\emptyset \in \mathcal{P}(A)$ and $A \in \mathcal{P}(A)$. If A is a set and B is a subset of A, then the set $A \backslash B$ is the set of all elements that are in A but not in B. It is called the **complement** of B in A.

Exercise 3. Prove by induction that if a set has n elements then it has 2^n subsets.

The **cartesian product** of A and B, denoted $A \times B$, is the set of *all* **pairs** (a, b) such that $a \in A$ and $b \in B$. The product $A \times B \times C$ is the set of all triples (a, b, c), with $a \in A, b \in B, c \in C$, and so on. $A \times A \times ... \times A$ (n times) is often denoted A^n and is the set of all "n**-tuples**" $(a_1, a_2, ..., a_n)$ where each entry in the tuple is an element of A. Familiarize yourself with $\mathbb{R}^n$, the set of n-tuples (also called "vectors") of real numbers.

A (binary) **relation** R, over $A \times B$, is a subset of $A \times B$. We often write aRb to mean the same thing as $(a, b) \in R$.

A **function** f, over $A \times B$ (often denoted $f : A \to B$), is a relation that has the following property: if $(a, b) \in f$ and $(a, b') \in f$ then $b = b'$. We say that f "maps" the set A to the set B, so we will sometimes refer to a function as a "mapping." A is called the **domain** while B is called the **range** of f. The statement $(a, b) \in f$ is often written $f(a) = b$, which you are probably more familiar with. For $A' \subseteq A$, the set $\{b \in B : \exists a \in A' \text{ with } f(a) = b\}$ is called the **image** of A' under f and is denoted $f(A')$. The image of the domain A under f is called the **codomain** of f.

We say f is **surjective** (or **onto**) if for every $b \in B$ there exists $a \in A$ such that $f(a) = b$. We say that f is **injective** if $f(a) = f(a')$ implies $a = a'$. A function that is both injective and surjective is called **bijective**. Bijective functions are **invertible**, that

is, given $b \in B$, there is a unique $a \in A$ such that $f(a) = b$. Functions that are not bijective are not invertible. (Please convince yourself that this is true.) If $f : A \to B$ is invertible, then there is a function $f^{-1} : B \to A$ such that for all $a \in A$, $f^{-1}(f(a)) = a$; f^{-1} is called the **inverse** of f. Notice that $f(f^{-1}(b)) = b$ for all $b \in B$.

Exercise 4. Consider the functions $f : A \to B$ and $g : B \to C$. Verify that the set

$$\{(a, g(f(a))) : a \in A\}$$

is also a function. *Hint:* Is it a relation? Over what? Does it satisfy the property that relations must satisfy to be functions? We call such a function the **composition** of f and g and we denote the function as $g \circ f$. Its domain is A and range is C.

For any two sets, A and B, define their **union** by $A \cup B = \{x : x \in A \text{ or } x \in B\}$ and their **intersection** by $A \cap B = \{x : x \in A \text{ and } x \in B\}$.

Let $y > x$. Then $[x, y]$ is the set of all numbers between x and y, including both x and y. Alternatively, we can write $(x, y]$ to exclude x or $[x, y)$ to exclude y or (x, y) to exclude both x and y. Remember that you should not confuse the interval (x, y) with the pair (x, y): when this notation is used the context will make it clear which of these we refer to. All of these sets, $[x, y]$, $(x, y]$, $[x, y)$ and (x, y), which happen to be subsets of $\mathbb{R}$, are called **intervals**. $[x, y]$ is a **closed interval**, while (x, y) is an **open interval**; $[x, y)$ and $(x, y]$ are **half open** intervals.

Finally, we note two notions of infinite sets. A set S is **countably infinite** if there is a bijective function mapping the set of natural numbers $\mathbb{N}$ to S. A set is **countable** if it is finite or countably infinite.

2 Matrix Algebra

2.1 Vectors & Matrices

An $n \times m$ **matrix** A is an array of numbers with n rows and m columns. A_i denotes the ith row and is itself a $1 \times m$ matrix. A^j denotes the jth column and is an $n \times 1$ matrix. Any $n \times 1$ matrix is also called a **vector** of size n. $\mathbb{R}^n$ denotes the set of all vectors of size n and $\mathbb{R}^{n \times m}$ denotes the set of all matrices that are $n \times m$.

Often we write $[a_{ij}]_{i=1,\ldots,n}^{j=1,\ldots,m}$ (or simply $[a_{ij}]$ when it is clear what n and m are) to denote the matrix A; and $[a_i]_{i=1,\ldots,n}$ (or simply $[a_i]$) to denote the $n \times 1$ matrix (i.e. vector) a. If $A = [a_{ij}]$ and $B = [b_{ij}]$ are both $n \times m$ matrices then $A + B$ is defined as the $n \times m$ matrix $[a_{ij} + b_{ij}]$. The **transpose** of the matrix $A = [a_{ij}]$ is the matrix $A' = [a_{ji}]$. The **dot product** of two vectors $a = [a_i]$ and $b = [b_i]$ is defined as the sum $\sum_{i=1}^n a_i b_i$ and is denoted $a'b$ or $b'a$ or $a \cdot b$. The **length** of a vector a of size n is $(a \cdot a)^{0.5}$ and is denoted $||a||$. If $a_i = 0$ for all $i = 1, \ldots, n$ then the vector a is called the **zero-vector** of size n and is denoted 0_n or just 0 when it is clear what n should be. If $a_i = 1$ for all $i = 1, \ldots, n$ then a is called the **one-vector** of size n and is denoted 1_n.

The product AB of an $n \times m$ matrix A and an $l \times k$ matrix B is not defined unless $l = m$, in which case it is the $n \times k$ matrix $[(A_i \cdot B^j)_{ij}]$. If c is a number then $c[a_{ij}] = [ca_{ij}]$. A **square matrix** is an $n \times n$ matrix, where n is called the **order** of the matrix. A **symmetric matrix** is one that is equal to its transpose.

A **lower triangular matrix** of order n is a square matrix of order n where $a_{ij} = 0$ for all $j > i$. An **upper diagonal matrix** of order n is a square matrix of order n whose transpose is a lower triangular matrix of order n. A **diagonal matrix** of order n is a lower triangular matrix of order n that is also an upper triangular matrix.

The **identity** matrix of order n is a diagonal matrix of order n where $a_{ij} = 1$ for all $i = j$. It is denoted I_n or just I when it is clear what n should be.

Exercise 5. Verify that (i) $A + B = B + A$, (ii) $(A + B) + C = A + (B + C)$, (iii) $(AB)C = A(BC)$, (iv) $A(B + C) = AB + AC$, (v) $(A + B)' = A' + B'$, (vi) $(AB)' = B'A'$, and (vii) $AI = A$ and $BI = B$ for any $n \times m$ matrices A and B (note that I does not denote the same matrix in the two equations: the two Is differ by their order so that the products are defined), and (viii) $I = I^2 = I^3 = \cdots$.

Exercise 6. Prove that if a and b are two vectors each of size n, then $|a \cdot b| \leq ||a||||b||$. This is known as the **Cauchy-Schwartz** inequality. *Hint:* If $b = 0_n$ then the result follows. If $b \neq 0_n$ then you can let $x = \frac{a \cdot b}{b \cdot b}$ and write $a = a - xb + xb$. Then you will have to show that $||a||^2 = ||a - xb||^2 + x^2||b||^2$ to get $x^2||b||^2 \leq ||a||^2$ and from this derive the result.

Exercise 7. Verify that the triangle inequality holds for vectors. That is, if x and y are each vectors of size n, $||x + y|| \leq ||x|| + ||y||$.

2.2 The Rank of a Matrix

Consider m vectors each of size n. Call the set of these vectors $V = \{a_1, \ldots, a_m\}$. A linear combination of V is an expression of the form $x_1a_1 + x_2a_2 + \cdots + x_ma_m$ where $x_1, \ldots, x_m$ are all numbers. V is said to be **linearly independent** if

$$x_1a_1 + x_2a_2 + \cdots + x_ma_m = 0_n \tag{1}$$

implies that $x_i = 0$ for all $i = 1, \ldots, m$. V is said to be **linearly dependent** if there are numbers $x_1, \ldots, x_m$, not all of which are equal to 0, such that

$$x_1a_1 + x_2a_2 + \cdots + x_ma_m = 0_n. \tag{2}$$

Any set like V is either linearly independent or linearly dependent.

Exercise 8. Let $a_1 = 3$, $a_2 = 7$, $b_1 = 2$, $b_2 = 4$, $c_1 = 0$, and $c_2 = 2$. Is the set $\{[a_i], [b_i], [c_i]\}$ linearly dependent or independent?

Let A be an $n \times m$ matrix. Take $\mathcal{A}^C = \{A^1, \ldots, A^m\}$, which is the set of columns of A, and let $\phi^C : \mathcal{P}(\mathcal{A}^C) \to \mathbb{R}$ be the function defined by

$$\phi^C(Z) = \begin{cases} 0 \text{ if } Z \text{ is linearly dependent} \\ |Z| \text{ if } Z \text{ is linearly independent} \end{cases} \tag{3}$$

Similarly take $\mathcal{A}_R = \{A_1, \ldots, A_n\}$, which is the set of rows of A, and let $\phi_R : \mathcal{P}(\mathcal{A}_R) \to \mathbb{R}$ be the function defined in exactly the same way as ϕ^C. Since n and m are both finite, ϕ^C and ϕ_R both achieve maximums on their domains. The **column rank** and **row rank** of A are then defined, respectively, as

$$c := \max_{Z \in \mathcal{P}(\mathcal{A}^C)} \phi^C(Z) \qquad \text{and} \qquad r := \max_{Z \in \mathcal{P}(\mathcal{A}_R)} \phi_R(Z).$$

The row rank (and column rank) of a matrix does not change when any of the following three operations are applied to the matrix:

1. interchanging any two rows (or columns)
2. multiplying each entry in a given row (or column) by a nonzero number
3. replacing any row (or column) by itself plus a number k times another row (or column)

Exercise 9. Convince yourself that the column and row ranks of a matrix are invariant to row and column operations above.

Theorem 2. *If A is an $n \times m$ matrix with c and r positive, then $r = c$; i.e., row rank equals column rank.*

Proof. Since $r > 0$ the matrix is not one where all of the entries are 0. Pick one nonzero component and through a series of successive row and column operations convert it to a matrix B where $b_{11} \neq 0$. This $b_{11} \neq 0$ is called the pivot entry. Now multiply the first row of this matrix by b_{21}/b_{11} and subtract it from the second row. Then multiply it by b_{31}/b_{11} and subtract it from the third row. Continue doing so down the rows. Then go across the columns doing the same thing until you get a matrix that has 0s in every row except the first, and in every column except the first. If there are any other entries that are nonzero, then you can pick any nonzero entry and after a series of column and row interchanges you can convert it to a matrix C where $c_{22} \neq 0$. Taking c_{22} to be the pivot entry, after a series of operations like those performed on B, you arrive at a matrix, D that has nothing but zeros in the second column and second row except in the the d_{22} position. Continue this process until you run out of candidates for pivot entries or you run out of spaces for pivot entries. Either way, you have a matrix of 0s except along a diagonal. Therefore, the column rank is equal to the row rank since the row and column ranks of this final matrix are equal to that of the matrix you started with. □

In light of this result, the column rank and row rank of a matrix are referred to simply as the **rank** of the matrix. An $n \times m$ matrix A is said to have **full rank** if the rank of the matrix is equal to the smaller of m and n, or $\min\{m, n\}$.

Exercise 10. Use row and column operations to calculate the rank of the matrix:

$$M = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 1 & 0 \\ -2 & -1 & 3 \\ -1 & 4 & 2 \end{bmatrix} \tag{4}$$

2.3 The Determinant & Inverses

Square matrices are special because they are the only kinds of matrices for which we can calculate what is called the **determinant**. Consider the square matrix A of order n. Consider the $(n-1) \times (n-1)$ submatrix of A created by deleting row i and column j. Call that matrix $A(i, j)$. The (i, j)-**cofactor** of A is defined as

$$C_{ij}(A) = (-1)^{i+j} \det A(i, j),$$

where $\det A(i, j)$ is the determinant of the matrix $A(i, j)$. Now the determinant of a 1×1 matrix is the value of the single entry. For an $n \times n$ matrix A, the determinant is defined as

$$\det A = a_{11}C_{11}(A) + \cdots + a_{1n}C_{1n}(A). \tag{5}$$

You may object that this definition is circular since we use the notion determinant to define the cofactor. However, since we defined the determinant of a 1×1 matrix, the above equality helps us to recursively define determinants for any $n \times n$ matrix.

The determinant may seem like a mysterious concept to you. To make it more concrete, suppose that A is a 2×2 matrix and its columns are A^1 and A^2. The determinant of A is the area of the parallelogram spanned by the vectors A^1 and A^2. If A is a 3×3 matrix then the determinant of A is the volume of the parallelepiped spanned by the vectors A^1, A^2 and A^3. If A is a 4×4 matrix, then the determinant of A is ...

Exercise 11. Find a simple formula for the determinant of any 2×2 matrix. Use this formula and equation (5) to calculate the determinant of the matrix M given in (4) with the second row deleted.

Exercise 12. Show that the determinant of any lower- or upper- triangular matrix is simply the product of the diagonal entries.

After having done Exercise 12, and knowing that you can convert a matrix into a lower or upper triangular matrix using row and column operations, the following properties will be useful to you in calculating the determinant of any matrix.

Let A be any square matrix of order n.

1. Let $[A, j, w]$ denote the $n \times n$ matrix in which the vector jth column of A is replaced by the vector w of size n. If u and v are two column vectors each of size n, then

$$\det[A, j, u + v] = \det[A, j, u] + \det[A, j, v].$$

   This decomposition holds for all $j = 1, ..., n$.

2. If the matrix B is obtained from A by interchanging any two rows (or columns) of A then $\det B = -\det A$.

3. If A and B are two matrices of order n then $\det AB = \det A \det B$.

4. If B is obtained from A by multiplying each entry of some given row (or column) of A by a nonzero constant k, then $\det B = k \det A$.

5. If B is obtained from A by replacing any row (or column) of A by itself plus k times some other row (or column), where k is any number, then the determinant remains unchanged.

6. $\det A = \det A'$.

Exercise 13. Convince yourself that the properties of determinants listed above hold. Then show that if a matrix has a row (or column) of zeros then its determinant is 0.

Exercise 14. Show that a square matrix A has full rank if and only if $\det A \neq 0$.

Theorem 3. (**Cramer's Rule**) *Let $A = [A^1, \ldots, A^n]$ be a square matrix of order n where the columns are $A^1, \ldots, A^n$. Suppose that $\det A \neq 0$ and let v be a vector of size n. Then the system of equations $Ax = v$ where x is a vector of n variables has a unique solution where each x_i, $i = 1, ..., n$ is given by*

$$x_i = \frac{\det[A^1, \ldots, A^{i-1}, v, A^{i+1}, \ldots, A^n]}{\det A}.$$

Proof. By Exercise 14, the matrix A has full rank. By row operations of the kind described above the augmented $n \times n+1$ matrix $[A^1, \ldots, A^n, v]$, where v is a vector of size n can be reduced to a matrix with zeros above and below the diagonal, and 1s on the diagonal, as in

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & c_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & c_n \end{bmatrix}$$

Therefore, the system of equations $Ax = v$ where x is a vector of n variables has a unique solution. Call it x^*. Thus

$$\begin{aligned} \det[A^1, \ldots, A^{i-1}, v, A^{i+1}, \ldots, A^n] &= \det[A^1, \ldots, A^{i-1}, Ax^*, A^{i+1}, \ldots, A^n] \\ &= \sum_{j=1}^{n} x_j^* \det[A^1, \ldots, A^{i-1}, A^j, A^{i+1}, \ldots, A^n] \\ &= x_i^* \det A. \end{aligned}$$

which follows from the properties of determinants listed above. Divide both sides by $\det A$ to find the solution given in the theorem. □

Theorem 4. (**inverse of a matrix**) *If A is an $n \times n$ matrix with $\det A \neq 0$ there exists a unique matrix B (called the inverse of A) such that $AB = BA = I_n$. Moreover, this matrix is given by $B = [C_{ij}(A)/\det A]'$, where $C_{ij}(A)$ is the (i, j)-cofactor of A*

Proof. To see why B is unique suppose that there was another matrix C such that $CA = I_n$. Then $CAB = B$, but also $CAB = C(AB) = CI_n = C$. So $B = C$. The analogous argument holds if $AC = I_n$.

Now we prove that B exists. Let e_{jn} be the size n vector such that there is a 1 in the jth position and 0 everywhere else. Then for any $n \times n$ matrix $X = [x_{ij}]$ solving $AX = I_n$ we have $e_{jn} = AX^j$ where X^j is the jth column of X. Since $\det A \neq 0$, the matrix A has full rank (by Exercise 14), and thus the solution exists and is unique (by row reduction). We have left to show that $XA = I_n$. By the properties of matrix multiplication

and determinants, we can find a matrix Y such that $A'Y = I_n$, which is equivalent to $Y'A = I_n$, and we have $I_n = Y'(AX)A = (Y'A)XA = XA$.

Finally, we derive the formula mentioned in the theorem. By Cramer's rule,

$$\begin{aligned} x_{ij} &= \det[A^1, \ldots, A^{i-1}, e_{jn}, A^{i+1}, \ldots, A^n] / \det A \\ &= \det[A^1, \ldots, A^{i-1}, e_{jn}, A^{i+1}, \ldots, A^n]' / \det A \\ &= C_{ji}(A) / \det A \end{aligned}$$

which gives us the formula. □

The unique matrix B that is the inverse of A is typically denoted A^{-1}, and this is how we will denote it from here on.

The following properties are useful. Whenever inverses exist,

1. $(A')^{-1} = (A^{-1})'$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $\det A^{-1} = 1/\det A$
4. The inverse of a lower (or upper) triangular matrix is a lower (or upper) triangular matrix.

Exercise 15. Prove the four properties above and find the inverse of the matrix in (4) with the second row deleted (if it exists).

2.4 Eigenvectors and Eigenvalues

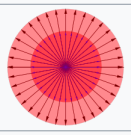
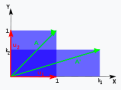
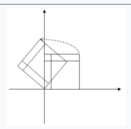
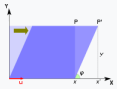
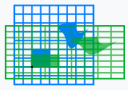
Let A be a square matrix of order n. A vector of size n is an **eigenvector** of A if there is a number λ such that $Av = \lambda v$. If $v \neq 0_n$ then λ is unique because $\lambda_1 v = \lambda_2 v$ implies $\lambda_1 = \lambda_2$. In that case, λ is said to be an **eigenvalue** of A belonging to v.

Theorem 5. *Let A be a square matrix of order n. Then λ is an eigenvalue of A belonging to some nonzero vector if and only if* $\det(A - \lambda I) = 0$.

Proof. Asume that λ is an eigenvalue of A. Then by definition, there is a vector $v \neq 0$ such that $Av = \lambda v$. In other words, $Av - \lambda v = 0_n$. This implies $(A - \lambda I_n)$ is a matrix with linearly dependent columns (since $v \neq 0$), so that the rank of the matrix is less than n. Therefore, by Exercise 14 it must be that $\det(A - \lambda I) = 0$.

Now interpret $(A - \lambda I_n)v = 0_n$ as a set of n equations. If $(A - \lambda I_n)$ does not have full rank then there is at least one equation that is a linear combination of the others. Eliminate one of the redundant equations. Now you are left with a system of equations with more variables than unknowns, which means that at least one variable can be set freely. That is equivalent to setting one entry of v freely. Make that one entry nonzero. Therefore, there is a vector $v \neq 0_n$ such that $Av = \lambda v$. □

Note that a matrix A represents the transformation of a vector space in the following sense: for each vector v, the vector Av is a new vector transformed by A. I reproduce examples of such transformations in the vector space $\mathbb{R}^2$ from Wikipedia, along with form of the matrix form of A that represents the transformation.

	scaling	unequal scaling	rotation	horizontal shear	hyperbolic rotation
illustration					
matrix	$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$	$\begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$	$\begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ $c = \cos\theta$ $s = \sin\theta$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} c & s \\ s & c \end{bmatrix}$ $c = \cosh\varphi$ $s = \sinh\varphi$

With this in mind, an eigenvector of A is simply a vector whose direction does not change, or is completely reversed (which happens when the associated eigenvalue is negative), under the transformation represented by A. Thus under "scaling," all vectors are eigenvectors of the transformation matrix. Under "horizontal shear" only the vectors parallel to the x-axis are eigenvectors, and so on.

Suppose A is an $n \times n$ matrix and the solution to $\det(A - \lambda I) = 0$ yields n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. When plugged back into $Av = \lambda v$, one can find corresponding eigenvectors associated with these eigenvalues. Since $Akv = \lambda kv$ is equivalent to $Av = \lambda v$ for any nonzero k, these eigenvectors are not unique. We may take ones that are normalized, i.e. ones whose lengths are set to 1. These are called the "unit eigenvectors." For each eigenvalue λ_i, we therefore have one eigenvector v_i. Then create the diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} \tag{6}$$

and the matrix P whose first column is v_1, second is v_2, third is v_3 etc. all the way up to v_n. Now it turns out that

$$A = PDP^{-1}.$$

You may look at a proof in any advanced linear algebra textbook or try to come up with one yourself. If you try to come up with your own proof, think in terms of matrices shifting axes; e.g., in $\mathbb{R}^2$ the vectors $(0, 1)$ and $(1, 0)$ define unit movement in the x and y directions. Suppose we wanted to rotate our coordinate system and re-write the vector in the new system. How would we do that?

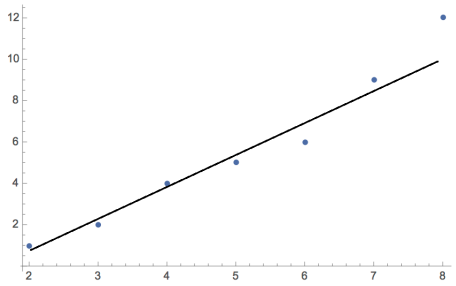
Exercise 16. Let

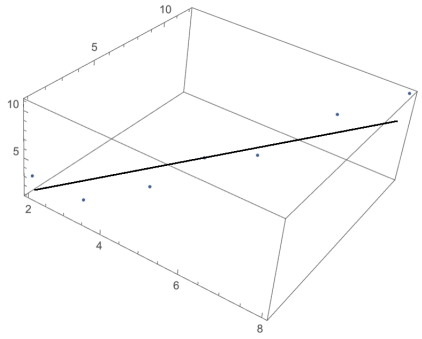
$$A = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix} \tag{7}$$

Find A^{19}.

2.5 Application: The Algebra of Least Squares

Let y be vector of size n and $\bar{X}$ an $n \times k$ matrix. We will refer to $(\bar{X}, y)$ together as the "data." To motivate what we plan to do, consider the case where $k = 1$. Then $\bar{X}$ too is a vector of size n. Suppose we plot values of (x_i, y_i), $i = 1, ..., n$, with the x's on the horizontal axis and y's on the vertical axis. We have a scatter of n points. We want to think of fitting a line through this scatter to "summarize" the relationship, linearly. When $k = 2$, we can imagine doing the same thing, but now we have a scatter of points in three-dimensional space. Again, we will want to think of fitting a line through this scatter. For $k > 2$, imagine doing the same thing even though it is hard to depict. It will be convenient to keep in mind the cases of $k = 1$ and $k = 2$ as we proceed.





A line is defined by its slope or **gradient**. In $(k+1)$-space, it can be written as $\tilde{y} = \tilde{x} \cdot m + b$ where m is a vector of size k, b is a number, $\tilde{x}$ is a vector of variables $(x_1, ..., x_k)$. Fitting the line means choosing values of m and b. How should these values be chosen? One way that we explore here is to "minimize the sum of squared residuals," also called the **ordinary least squares** (OLS). The point (x_i, y_i) from our data, plotted in $k + 1$-space, is off from the fitted line by the amount $\varepsilon_i := y_i - (x_i \cdot m + b)$. Note that

$$\sum_{i=1}^{n} (\varepsilon_i)^2 = \sum_{i=1}^{n} (y_i - (x_i \cdot m + b))^2 = (y - X\beta) \cdot (y - X\beta) =: S(\beta)$$

where X is the $n \times k + 1$ matrix of data with the 1_n vector appended to it (i.e., $\bar{X}$ with the 1_n vector added as a column at the end) and $\beta = (m, b)$ is the $(k + 1) \times 1$ matrix (i.e., vector) $(m_1, ..., m_k, b)$. We refer to each ε_i as the residual for observation i, and $S(\beta)$ as the sum of squared residuals.

We are interested in the value of $\beta = (m, b)$ that minimizes $S(m, b)$. You will solve this problem as an exercise later (after we introduce optimization). For now, note that

$$\begin{aligned} S(\beta) &= (y - X\beta)'(y - X\beta) \\ &= (y' - \beta' X')(y - X\beta) \\ &= y'y - \beta' X' y - y' X\beta + \beta' X' X\beta \\ &= y'y - 2y' X\beta + \beta' X' X\beta \end{aligned} \tag{8}$$

Exercise 17. Give reasons for why each of the equalities in (8) holds.

3 Differential Calculus

3.1 Limits, Continuity, and the Derivative

A real sequence, or simply **sequence**, is a collection of numbers $a_1, a_2, a_3, \ldots$ that can be indexed $1, 2, 3, \ldots$. The sequence in the previous sentence can be abbreviated $\{a_n\}_{n=1}^{\infty}$, and is said to **converge** if there is a number a such that

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ such that } \forall n \geq N, |a_n - a| < \epsilon.$$

The number a, if it exists, is unique and is called the **limit** of the sequence $\{a_n\}_{n=1}^{\infty}$. To see why it is unique, suppose both a and a' were limits of the convergent sequence $\{a_n\}_{n=1}^{\infty}$. Then that would mean that for all $\epsilon > 0$, there are numbers N and N' such that $|a_n - a| < \epsilon$ for all $n \geq N$ and $|a_n - a'| < \epsilon$ for all $n \geq N'$. Then for $n \geq \max\{N, N'\}$,

$$|a - a'| = |(a - a_n) + (a_n - a')| \leq |a_n - a| + |a_n - a'| < \epsilon + \epsilon = 2\epsilon.$$

where the first inequality in the centered statement follows from the triangle inequality. Since you can pick an ϵ arbitrarily small, this concludes the argument that $a = a'$. Therefore, the limit of a convergent sequence is unique. We often abbreviate the statement "the sequence $\{a_n\}_{n=1}^{\infty}$ converges to the limit a" as

$$\lim_{n\to\infty} a_n = a. \tag{9}$$

Exercise 18. Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be convergent sequences with limits a and b respectively, and let c be a number. Convince yourself that the following statements are true: (a) $\lim_{n\to\infty} ca_n = ca$, (b) $\lim_{n\to\infty} a_n + b_n = a + b$, (c) $\lim_{n\to\infty} a_n - b_n = a - b$, (d) $\lim_{n\to\infty} a_n b_n = ab$, and (e) if $\forall n$, $b_n \neq 0$ and $b \neq 0$, then $\lim_{n\to\infty} a_n/b_n = a/b$.

If $\{a_n\}_{n=1}^{\infty}$ is a sequence then let $s_n = \sum_{k=1}^{n} a_k$. This gives rise to the sequence $\{s_n\}_{n=1}^{\infty}$ of **partial sums**. If this sequence converges to a limit s, then we say that the **series** $\sum_{n=1}^{\infty} a_n$ converges to the sum s. If the sequence of partial sums does not converge, then we say that the series diverges.

Exercise 19. If $|r| < 1$, a is a number and n a natural number, then prove that

$$a + ar + ar^2 + \cdots + ar^n = \frac{a(1 - r^{n+1})}{1 - r}.$$

Then use this to show that the series $\sum_{n=0}^{\infty} ar^n$ converges to $a/(1 - r)$.

Exercise 20. This exercise describes the multiplier effect of spending. There are m shopowners in Mali. A tourist enters Mali and spends \$10 at Ms. 1's shop. Ms. 1 takes 80% of her profit and spends it at Ms. 2's shop; Ms. 2 spends 80% of her profit at Ms. 3's; ... and so on; Ms. m spends 80% of her profit at Ms. 1's, and this continues in a loop. For every

dollar transaction at a Malian shop, 70 cents is the cost of the goods sold. What Malians do not spend at each others shops, they save at the Timbuktu Bank. What fraction of the $10 spent by the tourist gets saved at the bank? What is the value of total purchases by Malians resulting from the tourist spending $10 at Ms. 1's shop?

Now, we want to capture the idea that a function $f : S \to \mathbb{R}$ (where S is an interval, could be $(-\infty, \infty)$) is "**continuous** at $x \in S$" if for all sequences $\{x_n\}_{i=1}^{\infty}$ that converge to x, the sequence $\{f(x_i)\}_{i=1}^{\infty}$ converges to $f(x)$. By the definition of convergence, this means that if $\{x_i\}_{i=1}^{\infty}$ converges to x, then

$$\forall \epsilon > 0, \exists N \text{ s.t. } \forall n \geq N, |f(x_n) - f(x)| < \epsilon.$$

But by the definition of $\{x_i\}_{i=1}^{\infty}$ converging to x, this is the same as saying

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } y \in S \text{ and } |x - y| < \delta \text{ implies } |f(x) - f(y)| < \epsilon.$$

We say that f is "a continuous function" if it is continuous at every point in S.

Exercise 21. Note that the sum and product of two continuous functions are also continuous. Prove that the composition of two continuous functions is continuous.

Next, we say that the function $f : S \to \mathbb{R}$ is "**differentiable** at $x \in S$" if S is an open interval and $\exists a \in \mathbb{R}$ such that

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } y \in S \text{ and } |x - y| < \delta \text{ implies } \left| \frac{f(x) - f(y)}{x - y} - a \right| < \epsilon.$$

It is a "differentiable function" if it is differentiable at every point in S.

Typically, the number a will depend on x, so we may as well write $a(x)$. If $a(x)$ is unique (which it is, and you can verify this), then $\{(x, a(x)) : x \in S\}$ is a function over $S \times \mathbb{R}$ whenever f is differentiable. In that case, we define the function $f' : S \to \mathbb{R}$, with $f'(x) = a(x)$, which we call the (first) **derivative** of f. The derivative of f', if it exists, is denoted f'', and is called the second derivative of f, and so on.

It is also important to know that we can define differentiability another way. If for all sequences $\{x_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} x_n = y$ and $x_n \neq y$ for all n we have

$$\lim_{n\to\infty} \frac{f(x_n) - f(y)}{x_n - y} = f'(y) \tag{10}$$

then we say that f is differentiable at y, where its derivative is $f'(y)$. Often, we abuse notation to write this statement as

$$\lim_{x\to y} \frac{f(x) - f(y)}{x - y} = f'(y). \tag{11}$$

and read the limit as "the limit as x approaches y..." Similarly, I'll use the notation $\lim_{x\to y^+}$ to mean that the limit holds for all sequences $\{x_i\}_{i=1}^{\infty} \subset (y, \infty)$ that converge to y,

and $\lim_{x\to y^-}$ means that the limit holds for all sequences $\{x_i\}_{i=1}^{\infty} \subset (-\infty, y)$ that converge to y. Also as an abuse of notation, $\lim_{x\to+\infty}$ means that the sequence in consideration increases without bound while $\lim_{x\to-\infty}$ means it decreases without bound.

Exercise 22. Convince yourself that the two definitions of differentiability are equivalent. That is, derive the second from the first, and the first from the second. *Hint*: Write down the ϵ, δ definition of the limit in (10). Also convince yourself that if a function is differentiable, then it is continuous. *Hint*: Multiply the last expression in the ϵ, δ definition of differentiability by $|x - y|$.

Exp and Log There is a result (that I will not cover) that says that the series

$$\sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... =: f(x)$$

converges for all values of x, and that the derivative of the function f whose domain is the real numbers, can be found by differentiating term by term. Thus,

$$f'(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \tag{12}$$

That's not strange. In fact, $f(x) = f'(x) = f''(x) = \cdots$ for this function, and we have a special name for it. We call such a function $f(x) = \exp(x)$, and it turns out that this function equals a number e (which happens to be irrational) raised to the power x. Thus $f(x) = e^x$. This function is strictly increasing and bijective if its range is defined to be only the positive numbers. (In fact, you can draw a graph of it to verify this.) Recall that bijective functions are invertible. The associated inverse function is called $\log(x)$. You will prove in a later exercise that the derivative of $\log(x)$ exists, and equals $1/x$.

3.2 Properties of the Derivative

If $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ are differentiable at $y \in S$ and c is a number, then cf, $f + g$, $f - g$ and fg are all differentiable at y. Here, cf is the function defined by multiplying $f(x)$ by c at all $x \in S$, $f + g$ is the function defined by adding $f(x)$ to $g(x)$ at all $x \in S$. Instead of adding, we subtract to define $f - g$ and multiply to define fg. If $g(x) \neq 0$ for all $x \in S$, then f/g, which is the function defined by dividing $f(x)$ by $g(x)$, is also differentiable. In fact, it is easy to show that

$$\begin{aligned}
[cf]'(y) &= cf'(y), \\
[f + g]'(y) &= f'(y) + g'(y), \\
[f - g]'(y) &= f'(y) - g'(y)
\end{aligned}$$

Now notice that

$$\frac{f(x)g(x) - f(y)g(y)}{x - y} = f(x)\frac{g(x) - g(y)}{x - y} + \frac{f(x) - f(y)}{x - y}g(y), \tag{13}$$

which is the main step in the proof of the **product rule**:

$$[fg]'(y) = f(y)g'(y) + f'(y)g(y). \tag{14}$$

In fact, all that one has to do is take limits on both sides of (13) and then use the fact that differentiable functions are continuous. Similarly, notice that

$$\frac{1/g(x) - 1/g(y)}{x - y} = -\frac{1}{g(x)g(y)}\frac{g(x) - g(y)}{x - y} \tag{15}$$

helps prove that $\left[\frac{1}{g}\right]'(y) = -\frac{g'(y)}{(g(y))^2}$. Take the limit on both sides of (15) and combine this result with the product rule to get the **quotient rule**:

$$\left[\frac{f}{g}\right]'(y) = \frac{g(y)f'(y) - g'(y)f(y)}{(g(y))^2}. \tag{16}$$

Finally, let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two functions and assume the composition $f \circ g$ is defined on an open interval, S. Suppose that g is differentiable at $x \in S$ and that f is differentiable at $g(x)$. Then $f \circ g$ is differentiable at x, with derivative

$$[f \circ g]'(x) = f'(g(x))g'(x). \tag{17}$$

This fact is known as the **chain rule**, and the following argument shows why it is true. Since f is differentiable at $g(x)$, then there is an error term $r(y)$, implicitly defined for any $y \in S$ by

$$f(g(y)) - f(g(x)) = [f'(g(x)) + r(y)][g(y) - g(x)]; \tag{18}$$

this error term has limit 0 as $g(y) \to g(x)$. But by the definition of continuity, it has limit 0 as $y \to x$ as well. Now divide both sides of (18) by $y - x$ and take the limit on both sides as x approaches y. On the left hand side you will get $[f \circ g]'(x)$. On the right hand side, the $r(y)$ term will vanish.

Exercise 23. Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined by $f(x) = ax^n$ where $a \in \mathbb{R}$ and $n \in \mathbb{R}$. Find its 1st, 2nd and 3rd derivatives using the limits definition of the derivative.

3.3 The Derivative in Multiple Dimensions

Let $f : S \to \mathbb{R}$ be a function and $S \subset \mathbb{R}^n$. Suppose that for any $\epsilon > 0$ there exists $\delta > 0$ such that if $y \in S$, $||x - y|| < \delta$ implies that $|f(x) - f(y)| < \epsilon$ then f is said to be **continuous** at x. If the statement is true for every $x \in S$ then f is said to be a continuous function.

Similarly, let S_1, S_2, ..., S_n be open intervals; we can allow some or all of them to be $(-\infty, \infty)$. Then $S := S_1 \times ... \times S_n$ is a subset of $\mathbb{R}^n$, and we call S an **open box**. A function $f : S \to \mathbb{R}$ is said to be **differentiable** at $x \in S$ if for all $\epsilon > 0$ there is a $\delta > 0$ such that $y \in S$ and $||x - y|| < \delta$ implies

$$|f(x) - f(y) - a(x) \cdot (x - y)| < \epsilon ||x - y||,$$

for some vector $a(x)$ of size n. Akin to the one-dimensional case, the vector $a(x)$ is called the **derivative** of f at $x \in S$ and is unique for each x whenever it exists. If f is differentiable at all points in S then it is a differentiable function, and we can define the derivative of f to be the function $\nabla f : S \to \mathbb{R}^n$ such that $\nabla f(x) = a(x)$.

It is not hard to show that if both $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $x \in \mathbb{R}^n$ then so is $c_1 f + c_2 g$, where c_1 and c_2 are numbers. Fortunately,

$$\nabla(c_1 f + c_2 g)(x) = c_1 \nabla f(x) + c_2 \nabla g(x).$$

In fact, the chain rule also applies: if $h : \mathbb{R} \to \mathbb{R}$, then

$$\nabla[h \circ f](x) = h'(f(x)) \nabla f(x). \tag{19}$$

Let $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is an open box. Let $e_j \in \mathbb{R}^n$ be the vector with 0s in every entry except for the jth, where the entry there is a 1. Then the jth **partial derivative** of f at the point $x \in S$ exists if for all $\epsilon > 0$ there is a $\delta > 0$ such that for any number t for which $x + te_j \in S$, $t < \delta$ implies

$$\left| \frac{f(x + te_j) - f(x)}{t} - a \right| < \epsilon \tag{20}$$

The number a, if it exists, is unique for each x and is the jth partial derivative. It defines the partial derivative function, $\frac{\partial f}{\partial x_j} : S \to \mathbb{R}$, a function defined by $\frac{\partial f(x)}{\partial x_j} = a$.

Similarly, if we replace every occurence of e_j in the definition of partial derivative by μ, where $\mu \in \mathbb{R}^n$ and restrict t to be positive, then we have the definition of "the **directional derivative** of f at x in the direction μ."

Theorem 6. *Consider $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is an open box. Then (i) if f is differentiable then it is continuous; (ii) if f is differentiable at x then $\partial f(x)/\partial x_j$ exist for all j and $\nabla f(x) = [\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_n]'$; (iii) if $\partial f(x)/\partial x_j$ exist for all j and are all continuous on x then $\nabla f(x)$ exists and is given by $\nabla f(x) = [\partial f(x)/\partial x_1, \ldots, \partial f(x)/\partial x_n]'$; (iv) if f is differentiable at x then the directional derivative of f exists for any vector μ and is equal to $\nabla f(x) \cdot \mu$.*

Exercise 24. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function such that $f(0, 0) = 0$, and for $(x, y) \neq 0$,

$$f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}.$$

Is f differentiable at $(0, 0)$?

Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is an open box. Suppose f is differentiable at $x \in S$, and suppose that each partial derivative function of f is differentiable at x. Denote the jth partial of $\partial f(x)/\partial x_i$ (also called the "(i,j)-cross partial") by $\partial^2 f(x)/\partial x_j \partial x_i$ if $j \neq i$ and $\partial^2 f(x)/\partial x_i^2$ if $j = i$. Then the **Hessian** of f at x is the matrix

$$Hf(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \cdots & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{21}$$

If every partial derivative of f is a continuous function, then we say that f is **continuously differentiable** or C^1.

If every (i,j)-cross partial of f is a continuous function then we say that f is C^2, and when f is C^2, it turns out that the Hessian is a symmetric matrix with

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \tag{22}$$

for all $i = 1, \ldots, n$ and $j = 1, \ldots, n$. This fact is called **Young's theorem**, and you will demonstrate it through an example below.

Let $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is an open box. Now let us treat x_j, $j \neq i$ as constants and define the function $g : S_i \to \mathbb{R}$ to be

$$g(x_i) \equiv f(x_i; x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$

where the semicolon simply divides the free and fixed variables (i.e., "parameters"); alternatively, we may sometimes use " | " instead of the semi-colon to separate the free and fixed variables. Then you will be relieved to know that

$$\frac{\partial f}{\partial x_i} \equiv \frac{dg}{dx_i}. \tag{23}$$

So go ahead and use the chain rule, product rule, quotient rule, etc. that we described in the one variable case to calculate partial derivatives.

Exercise 25. Convince yourself that (22) and (23) hold.

Exercise 26. Let $f(x_1, x_2) = \log x_1 (x_2)^2 + x_1 x_2$ and assume that f is C^2. Demonstrate Young's theorem for this function.

4 Real Analysis

Let $\{a_{1k}\}$, $\{a_{2k}\}$, ..., and $\{a_{nk}\}$ be sequences that converge to a_1, a_2, ..., and a_n respectively. Then the sequence of vectors, $\{[a_{1k}, a_{2k}, \ldots, a_{nk}]'\}_{k=1}^{\infty}$ converges to the vector $[a_1, a_2, ..., a_n]'$. This is the **convergence** of vectors.

A **closed set** is a set of vectors $X \subset \mathbb{R}^n$ where the limit of every convergent sequence $\{x_k\} \subset X$ also lies in X. If for all $x \in X \subset \mathbb{R}^n$, there exists an open box $S \subset X$ such that $x \in S$, then X is said to be an **open set**.

A **bounded set** is a set X for which there is an open box $S = S_1 \times S_2 \times \cdots \times S_n$ such that $X \subset S$ and each S_i, $i = 1, ..., n$, is the open interval $(-z, z)$, where $z > 0$.

A **subsequence** $\left\{x_{m(k)}\right\}$ of a sequence $\{x_k\}$ is a sequence of some (or all) of the elements of $\{x_k\}$ appearing in the order in which they appear in $\{x_k\}$.

A **compact set** is a set is a set $X \subset \mathbb{R}^n$ such that every sequence in X has a convergent subsequence whose limit is in X.

A **convex set** is a set $X \subset \mathbb{R}^n$ where if $x \in X$ and $y \in X$ then $\alpha x + (1 - \alpha)y \in X$ for all $\alpha \in (0, 1)$.

The **supremum** of a set $X \subset \mathbb{R}$ is the lowest number sup X such that every number greater than sup X is greater than every number in X. This is also called the "lowest upper bound" of X. The **infimum** is the "greatest lower bound". A fact that we will not prove is that if X is a bounded set (i.e. there is a number $z > 0$ such that $X \subseteq [-z, z]$) then both $\inf X$ and $\sup X$ are elements of $\mathbb{R}$.

Exercise 27. Show that if $X \subset Y \subset [a, b]$ for some interval $[a, b]$ then $\inf Y \leq \inf X$ and $\sup Y \geq \sup X$.

Exercise 28. Show that if A and B are both convex sets, their intersection is convex but not necessarily their union.

Exercise 29. A real-valued function f defined on some interval $[a, b] \subseteq \mathbb{R}$ is said to be **convex** if for all $x \neq y$, $f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)$ for all $\alpha \in (0, 1)$ and **strictly convex** if the inequality holds strictly. f is said to be **concave** if $-f$ is convex, and **strictly concave** if $-f$ is strictly convex. Show that if f is convex, then for all $x_0 \in [a, b]$ there exists a line $y = f(x_0) + m(x - x_0)$ through $(x_0, f(x_0))$ such that $f(x) \geq f(x_0) + m(x - x_0)$ for all $x \in [a, b]$; that is, f is weakly higher than the line. Show that if f is concave, then for all $x_0 \in [a, b]$ there is a line through $(x_0, f(x_0))$ with f is weakly lower than the line.

4.1 Existence of Extreme Values

Our objective in this part is to establish the result that a continuous real-valued function defined on a compact set achieves both a maximum and minimum on that set. This is a central result that underpins optimization theory.

Theorem 7. (Heine-Borel Theorem) *A set $X \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded.*

I have put the proof for you to read in the appendix to this section. The Heine-Borel theorem implies the following result, sometimes called the "extreme value theorem."

Theorem 8. (Weierstrass theorem) *Let $X \subset \mathbb{R}^n$ be a compact set and f be a continuous real-valued function on X. Then f attains a minimum and maximum on X.*

Proof. The first step is to show that the codomain $f(X)$ of f is compact. Let $\{y_k\} \subset f(X)$ be a sequence. For each k pick $x_k \in X$ such that $f(x_k) = y_k$ (which you can do by construction). This gives us a sequence $\{x_k\} \subset X$. Since S is compact you can pick an infinite subsequence $\left\{x_{m(k)}\right\} \subset \{x_k\}$ that converges to some $x \in X$. Let $y = f(x)$ and $y_{m(k)} = f(x_{m(k)})$. Since $\left\{x_{m(k)}\right\}$ converges to x and f is continuous, the infinite sequence $\left\{f(x_{m(k)})\right\}$ converges to $f(x)$. But $f(x) \in f(X)$ so $f(X)$ is compact.

The second step is to show that because $f(X)$ is compact $\sup f(X) \in f(X)$ and $\inf f(X) \in f(X)$, and these are the maximum and minimum we need. First of all, boundedness (from Heine-Borel) tells us that $\sup f(X) < \infty$ and $\inf f(X) > -\infty$. Now, let N_k be the interval $(\sup f(X) - 1/k, \sup f(X)]$ where $k = 1, 2, \ldots$. Let $f(X)_k = f(X) \backslash N_k$. Then $f(X)_k$ is not empty for each k, otherwise we would have an upper bound strictly smaller than $\sup f(X)$. Now for each $f(X)_k$ pick any $y_k \in f(X)_k$. The sequence $\{y_k\}$ must converge to $\sup f(X)$. Since $f(X)$ is closed, $\sup f(X) \in f(X)$. The argument for inf is identical. □

4.2 Intermediate & Mean Value Theorems

We now turn to a set of results known as intermediate and mean-value results, along with an application of these results to the method known as "L'Hopital's rule."

Theorem 9. (intermediate value theorem) *Let f be a continuous real-valued function defined on an interval $[a, b]$. For $f(x) > f(y)$ and any $c \in (f(y), f(x))$, there exists a number $z \in (\min\{x, y\}, \max\{x, y\})$ such that $f(z) = c$.*

Proof. Let $g(x) = f(x) - c$. Construct a sequence of intervals $\{S_i\}_{i=0}^{\infty}$ as follows. Begin with $S_0 = [a, b]$. If $g(x) = 0$ at the midpoint of this interval, then we're done. Otherwise, g changes sign between the endpoints on either the right half or the left. Pick the half that it changes sign on and call the interval S_1. If it is 0 at the midpoint, then again we're done. If not, again pick the half on which it changes sign and call it S_2, and so on. Either we reach a point where $g(x)$ takes a value of 0 at the midpoint of an interval, or we obtain an infinite sequence of intervals. In the latter case, the sequence of left endpoints and the sequence of right endpoints both converge to the same limit z between x and y. By continuity and the change of sign condition, $g(z) = 0$. □

Theorem 10. (generalized intermediate value theorem) *Let $X \subset \mathbb{R}^n$ be a convex set, $f : X \to \mathbb{R}$ a continuous function, and x and y points such that $f(x) < f(y)$. Then for any c such that $f(x) < c < f(y)$ there is an $\alpha \in (0, 1)$ such that $f((1-\alpha)x + \alpha y) = c$.*

Proof. Let $g : [0, 1] \to \mathbb{R}$ be defined by $g(\beta) = f((1-\beta)x + \beta y)$ for $\beta \in [0, 1]$. Since f is continuous, g is continuous and $g(0) = f(x)$, $g(1) = f(y)$, and $g(0) < c < g(1)$. By the intermediate value theorem there is $\alpha \in (0, 1)$ such that $c = g(\alpha) = f((1-\alpha)x + \alpha y)$. □

Theorem 11. (Rolle's theorem) *If f is continuous on the interval $[a, b]$, differentiable everywhere on (a, b), and $f(a) = f(b)$, then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. If f is constant then $f'(x) = 0$ for all x, and we'd be done. So for challenge's sake, let f not be constant. By the Weierstrass Theorem, f attains maximum and minimum values on $[a, b]$. Since f is not constant, either the maximum is greater than $f(a)$ or the minimum is less than $f(a)$ (or both). If the maximum value is greater than $f(a)$ then any point x at which it is attained lies in (a, b) (it can't be b because $f(a) = f(b)$ by assumption). The numerator of

$$\frac{f(y) - f(x)}{y - x} \tag{24}$$

where $x \neq y \in [a, b]$ is always non-positive and the denominator can have either sign. Now take the limit as $x \to y$ on the centered expression above. Due to different signs on different sides, the limit cannot be positive or negative. But it exists by the assumption that f is differentiable on (a, b). So it must be 0. The argument is similar if f's minimum is less than $f(a)$. □

Theorem 12. (mean value theorem) *If f is continuous on the interval $[a, b]$ and differentiable everywhere on (a, b), then there exists $c \in (a, b)$ such that*

$$(b - a)f'(c) = f(b) - f(a).$$

Proof. Note that we have the same assumptions as in Rolle's theorem, except the assumption that $f(a) = f(b)$. We need to show that there is a point $c \in (a, b)$ with

$$\frac{f(b) - f(a)}{b - a} = f'(c). \tag{25}$$

But this is easy. Let g be a function on $[a, b]$ defined by

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a) \tag{26}$$

and notice that $g(a) = g(b) = f(a)$. By Rolle's theorem, there is a point $c \in (a, b)$ such that

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}, \tag{27}$$

and we are done. □

Exercise 30. Suppose that $f : (a, b) \to \mathbb{R}$ is a differentiable function, and so is $f' : (a, b) \to \mathbb{R}$. ($f$ is "twice differentiable" if its derivative is a differentiable function.) Suppose also that $f''(x) < 0$ for all $x \in (a, b)$. Show that f must be strictly concave. If instead f is convex and twice differentiable, show that $f''(x) \geq 0$ for all $x \in (a, b)$.

Exercise 31. Let $X \subset \mathbb{R}^n$ be a convex and open set and let $f : X \to \mathbb{R}$ be a differentiable function. Then the **generalized mean value theorem** states that for any $x \in X$ and $y \in X$ there is an $\alpha \in (0, 1)$ such that

$$f(x) - f(y) = \nabla f((1 - \alpha)x + \alpha y)(x - y)$$

Prove this by defining g exactly the same as in the proof of the generalized intermediate value theorem. *Hint*: Notice that $g'(\alpha) = \nabla f((1 - \alpha)x + \alpha y) \cdot (b - a)$.

Theorem 13. (another mean value theorem) *Suppose that there are continuous functions $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$. Suppose that these functions are differentiable at every point in (a, b) and that $g'(x) \neq 0$ for all $x \in (a, b)$. Then, there exists $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}. \tag{28}$$

Proof. With the assumptions of the theorem, the mean value theorem implies that $g(b) - g(a) \neq 0$. Now define $h : [a, b] \to \mathbb{R}$ by

$$h(x) = f(x)[g(b) - g(a)] - g(x)[f(b) - f(a)]. \tag{29}$$

I'll give you \$100 if $h(a)$ is not equal to $h(b)$. Now again, by the mean value theorem, there is $c \in (a, b)$ such that

$$h'(c) = f'(c)[g(b) - g(a)] - g'(c)[f(b) - f(a)] = 0. \tag{30}$$

Since we agreed that $g(b) - g(a) \neq 0$, you can divide both sides of this final equality by $[g(b) - g(a)]g'(c)$ to prove the stated result. □

Theorem 14. (L'Hopital's Rule) *Let $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$ be differentiable everywhere on (a, b) and that $g(x) \neq 0$ and $g'(x) \neq 0$ for $x \in (a, b)$. Then it is not so unfortunate that $\lim_{x \to a^+} f(x) = \lim_{x \to a^+} g(x) = 0$, for it is the case that:*

$$\lim_{x \to a^+} \frac{f(x)}{g(x)} = \lim_{x \to a^+} \frac{f'(x)}{g'(x)} \tag{31}$$

so long as the limit on the right hand side exists.

Proof. I will try to derive (31) from the assumptions. To make sure f and g are continuous at a, we need $f(a) = g(a) = 0$. This does not have to be the case, but we can just redefine f and g to be so if it isn't. Call the limit on the right side of (31), L. By the properties of limits, for any $\epsilon > 0$, we can find an interval $T = (a, a + \delta)$ such that

$$\left| \frac{f'(c)}{g'(c)} - L \right| < \epsilon \tag{32}$$

for $c \in T$. Invoking the other mean value theorem, we can then argue

$$\left| \frac{f(x) - f(a)}{g(x) - g(a)} - L \right| < \epsilon \tag{33}$$

for all $x \in T$. Then using that $f(a) = g(a) = 0$, we've derived (31). □

Even though we didn't prove it, L'Hopital's rule would still be true if $a = -\infty$ or $a = \infty$, and/or if $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = \infty$ instead of 0. Furthermore, a does not have to be approached from the right (which is not possible in the case of $a = \infty$ anyway).

4.3 The Implicit and Inverse Function Theorems

The implicit and inverse function theorems are important results. For example, the implicit function theorem will be invoked in the proof of Lagrange's theorem, which we use in optimizing a function subject to equality constraints.

Theorem 15. (implicit function theorem) *Given $n \geq 1$, let a typical point of the set $\mathbb{R}^{n+1}$ be denoted by (x, y), where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}$. Let $S \subset \mathbb{R}^{n+1}$ be an open box, and let $F : S \to \mathbb{R}$ be a differentiable function with continuous partial derivatives. Let (x^*, y^*) be a point in S such that*

$$\frac{\partial F(x^*, y^*)}{\partial y} \neq 0 \tag{34}$$

and let $F(x^*, y^*) = 0$. Then there is an open box $B \subset \mathbb{R}^n$ such that $x^* \in B$, and a differentiable function $g : B \to \mathbb{R}$ whose partial derivatives are continuous, such that $g(x^*) = y^*$, and $F(x, g(x)) = 0$ for all $x \in B$. The derivative of g at any $x \in B$ is:

$$\frac{\partial g}{\partial x_j} = -\frac{\partial F / \partial x_j}{\partial F / \partial y}. \tag{35}$$

Proof. We first construct the function g so that $F(x, g(x)) = 0$, and then we show that it has the remaining properties stated in the theorem. It is the function g that we call the "implicit function" in the name of the theorem.

Due to (34) we can assume without loss of generality that $\partial F(x^*, y^*)/\partial y > 0$. By continuity of $\partial F/\partial y$, there is a small open box $A \subset \mathbb{R}^{n+1}$ containing (x^*, y^*) such that

$$\frac{\partial F(x, y)}{\partial y} > 0, \qquad \forall (x, y) \in A \tag{36}$$

Thus $F(x^*, \cdot)$ is increasing in y in a neighborhood of y^*, which means we can find y_1, y_2 satisfying $F(x^*, y_1) < 0 < F(x^*, y_2)$ and $y_1 < y^* < y_2$. Again by the continuity of F we can find an open box $B \in \mathbb{R}^n$ containing x^* such that $B \times [y_1, y_2] \subset A$ and $F(x, y_1) < 0 < F(x, y_2)$ for all $x \in B$. By the intermediate value theorem, for each $x \in B$, there is $y \in (y_1, y_2)$ such that $F(x, y) = 0$. Uniqueness of this y is guaranteed by (36). This uniqueness allows us to define the continuous function $g(x) = y$ having the properties described in the theorem, except that we have yet to show (35) and the fact that that the partial derivatives are continuous.

To that end, fix $x \in B$, and let $y = g(x)$. Then by definition of the derivative (please re-read the definition of the derivative in multiple dimensions), we have

$$F(x + se_j, y + t) - F(x, y) = s\frac{\partial F(x, y)}{\partial x_j} + t\frac{\partial F(x, y)}{\partial y} + \varepsilon\sqrt{s^2 + t^2}$$

where ε is a "correction term" that goes to 0 as $\sqrt{s^2 + t^2}$ goes to 0. Now pick s small enough so that $x + se_j \in B$ and set $t = g(x + se_j) - g(x)$. For this value of t, the left hand side equals of the centered equality above equals 0 given the way g was constructed. Therefore, rearranging the equality we get

$$t\frac{\partial F(x, y)}{\partial y} = -s\frac{\partial F(x, y)}{\partial x_j} - \varepsilon\sqrt{s^2 + t^2}, \tag{37}$$

which, after substituting t on the left and dividing both sides by $\partial F(x, y)/\partial y \neq 0$, in turn rearranges to

$$\frac{g(x + se_j) - g(x)}{s} = -\frac{\partial F(x, y)/\partial x_j}{\partial F(x, y)/\partial y} - \frac{\varepsilon}{\partial F(x, y)/\partial y}\frac{\sqrt{s^2 + t^2}}{s} \tag{38}$$

Keeping in mind that our choice of t goes to 0 as s goes to 0, take the limit as $s \to 0$ on both sides. The question is: Can we eliminate the right hand term by doing this? The answer is yes if $\sqrt{s^2 + t^2}/s$ is bounded for small values of s, in which case its behavior cannot overwhelm the convergence of ε to 0. But in fact this term *is* bounded. I show this in the appendix. Please read the argument, and come back here.

Therefore, taking the limit as $s \to 0$ on both sides of (38) gives us the j'th partial derivative of g on the left side, and and the right side of (35) on the right, as required. We know that these partials are continuous since $\partial F(x, y)/\partial y$ is non-vanishing, and the partials of F are continuous. Thus g is differentiable by Theorem 6 part (iii). □

Exercise 32. Use the implicit function theorem to find dy/dx along the circle, $x^2 + y^2 = 1$. Where does dy/dx not exist?

Theorem 16. (inverse function theorem) *Let f be differentiable at every point on (a, b), and let $f'(x) \neq 0$ for all $x \in (a, b)$. Assume that f is invertible and let its inverse be*

$g = f^{-1}$. *Then g is differentiable at $f(x)$ and*

$$g'(f(x)) = \frac{1}{f'(x)}. \tag{39}$$

Proof. To see why g is differentiable at $f(x)$, let $f(x) = y$ and $f(x') = y' \neq y$ for some x and x' both in S. You can make this assumption because there is an open interval that contains x, and $y' \neq y$ for any $x' \neq x$ that we choose in this interval. This is because $f'(x) \neq 0$. Now since g and f are inverses of each other,

$$\frac{g(y') - g(y)}{y' - y} = \frac{x' - x}{f(x') - f(x)}. \tag{40}$$

Now recall that the inverse of a continuous function is continuous, take the limits $\lim_{x' \to x}$) on both sides and note that the left side is the limit as $f(x') \to f(x)$ (by continuity) which gives us the definition of $g'(y)$ while the right side is $1/f'(x)$. □

Alternatively, if we knew already that g is differentiable at $f(x)$, then this follows from the chain rule. Since $g \circ f(x) = x$, we have $[g \circ f]'(x) = 1$. Then plugging this into (17), we have $1 = g'(f(x))f'(x)$, which rearranges to give (39). But of course we had to first establish that g is differentiable at $f(x)$.

Exercise 33. Show that a continuous strictly increasing function f defined on an interval $[a, b]$ has a continuous strictly increasing inverse.

Exercise 34. Use the inverse function theorem to find the derivative of $\log x$.

Exercise 35. Find the derivative of $f(x) = \frac{\log(3x^2+2)}{e^{6x}+1}$.

Exercise 36. Is the inverse of $f(x) = x^3$ differentiable everywhere?

Exercise 37. Use L'Hopital's rule to calculate $\lim_{x \to 0^+} x^x$. (*Hint*: Do some easier L'Hopital's rule problems from a textbook of your liking first, remember the properties of log and e, and then think about continuity.)

Exercise 38. Prove that for a sequence of numbers $\{x_n\}_{n=1}^{\infty}$ that converges to x,

$$\lim_{n \to \infty} (1 + x_n/n)^n = e^x.$$

Hint: One way to do this is to use the binomial theorem and L'Hopital's rule; but there are other ways.

5 Integral Calculus

5.1 The Riemann Integral

Let $S = [a, b]$ be a closed interval, and let f be a function that is bounded on S, i.e. there is a number, C, such that $|f(x)| \leq C$ for all $x \in S$. A **partition** of $[a, b]$ is a set P of points $x_0, x_1, \ldots, x_n$ such that

$$a = x_0 < x_1 < \cdots < x_n = b.$$

A **refinement** of a partition P is another partition P' such that $P \subset P'$. Define

$$L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1}), \text{ where } m_k = \inf\{f(x) \mid x \in [x_{k-1}, x_k]\},$$
$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1}), \text{ where } M_k = \sup\{f(x) \mid x \in [x_{k-1}, x_k]\}.$$

Now if P and P' are partitions of S and P' is a refinement of P, then

$$L(f, P) \leq L(f, P') \leq U(f, P') \leq U(f, P). \tag{41}$$

This is easy to see geometrically. In fact, the coarsest partition bounds the set of all $L(f, P)$ and $U(f, P)$ with respect to partitions, since

$$m(b - a) \leq L(f, P) \leq U(f, P) \leq M(b - a),$$

where $m = \inf\{f(x) \mid x \in S\}$ and $M = \sup\{f(x) \mid x \in S\}$. Now define

$$\mathcal{L} = \sup_P \{L(f, P)\} \text{ and } \mathcal{U} = \inf_P \{U(f, P)\}.$$

It is easy to see that $\mathcal{L} \leq \mathcal{U}$. The argument is as follows. For a fixed partition P^*, $L(f, P^*)$ is a lower bound for the set of $U(f, P)$, because for any two partitions P and P', there is a refinement of both of them, P'', which by (41), leads to

$$L(f, P) \leq L(f, P'') \leq U(f, P'') \leq U(f, P').$$

Therefore, $L(f, P) \leq \mathcal{U}$, and this is true for every P, completing the argument.

A bounded function, $f : S \to \mathbb{R}$, is **Riemann integrable** if $\mathcal{L} = \mathcal{U}$. If this is the case, the common value is denoted $\int_S f(x)dx$ or $\int_a^b f(x)dx$.

Lemma 1. *f is Riemann integrable on $[a, b]$ if and only if for all $\epsilon > 0$ there is a partition P such that*

$$U(f, P) - L(f, P) < \epsilon. \tag{42}$$

Proof. Let us start by assuming that f is Riemann integrable on $[a, b]$ and try to obtain the implication. Let

$$\mathcal{I} = \int_a^b f(x)dx. \tag{43}$$

Take any $\epsilon > 0$. Since $\mathcal{L} = \mathcal{U}$, it follows that there must be partitions P' and P'' such that

$$\mathcal{I} - \frac{\epsilon}{2} < L(f, P') \text{ and } U(f, P'') < \mathcal{I} + \frac{\epsilon}{2},$$

which is just a simple application of (41). From this we get $\mathcal{I} < L(f, P) + \frac{\epsilon}{2}$ and $U(f, P) - \frac{\epsilon}{2} < \mathcal{I}$, which when combined gives us $U(f, P) - L(f, P) < \epsilon$. Check.

We can go the other way now. Assume that (42) is true. By definition, $\mathcal{U} \leq U(f, P)$ for any P and $\mathcal{L} \leq L(f, P)$ for any P. $U(f, P) \geq L(f, P)$ and $\mathcal{U} \geq \mathcal{L}$, so it must be that $0 \leq \mathcal{U} - \mathcal{L} < \epsilon$. But ϵ is arbitrary, so $\mathcal{U} = \mathcal{L}$. Therefore, f is Riemann integrable. □

Exercise 39. We now introduce a new concept. A function $f : [a, b] \to \mathbb{R}$ is **uniformly continuous** if for all $\epsilon > 0$ there is $\delta > 0$ such that if $x, y \in [a, b]$ and $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$. This differs from regular continuity in that here δ does not depend on x, whereas in the regular case it may. Show that if $f : [a, b] \to \mathbb{R}$ is continuous, then it is uniformly continuous.

Theorem 17. *If $f : [a, b] \to \mathbb{R}$ is continuous, then it is Riemann integrable.*

Proof. Now we are going to use the Lemma 1 and Exercise 39 to prove that continuous functions are integrable. Since f is uniformly continuous, by the exercise above it is uniformly continuous; i.e., for any $\epsilon > 0$ there is $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon/(b - a)$, where x and y are elements of $[a, b]$. (It doesn't matter that we divided ϵ by $b - a$; if you don't like that, you could have chosen $(b - a)\epsilon$ instead; it was arbitrary after all.) Now for any ϵ take the associated δ and partition the interval $[a, b]$ into n subintervals, each of size $l < \delta$. By definition, the m_k and M_k differ by at most $\epsilon/(b - a)$. Therefore,

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) \\ &< \sum_{k=1}^{n} \frac{\epsilon}{b - a} l = \frac{\epsilon}{b - a} \sum_{k=1}^{n} l = \epsilon. \end{aligned}$$ □

Now one thing that I'm not going to prove but that you should know is that if a bounded function is continuous except at a finite number of points, then it is Riemann integrable. There are even stronger results, but we won't need to prove them here.

Before proceeding it will be convenient to state definitions and properties (which we will not prove) relating to the Riemann integral.

Theorem 18. *Let f and g be Riemann integrable on $[a, c]$ and let $k \in \mathbb{R}$. Then,*

1. $\int_a^c kf(x)dx = k\int_a^c f(x)dx$
2. $\int_a^c [f(x) + g(x)]dx = \int_a^c f(x)dx + \int_a^c g(x)dx$
3. *The same as (2) but with minus signs.*
4. $|\int_a^c f(x)dx| \leq \int_a^c |f(x)|dx$
5. $\int_a^c f(x)dx \leq \int_a^c g(x)dx$ *if* $f(x) \leq g(x)$ *for all* $x \in [a, c]$, *and*
6. *If* $b \in [a, c]$, *then* $\int_a^c f(x)dx) = \int_a^b f(x)dx + \int_b^c f(x)dx$

In addition, we also have the following two definitions.

7. $\int_b^b f(x)dx := 0$ *for* $b \in [a, c]$, *and*
8. $\int_a^c f(x)dx := \int_c^a f(x)dx$

5.2 The Fundamental Theorem of Calculus

Now we introduce and prove an important theorem that connects integration with differentiation: the fundamental theorem of calculus. This important result says, roughly, that integration is the inverse operation of differentiation. We first have the following lemma that sets the stage for this result.

Lemma 2. (**Leibniz's Rule**) *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous so that it is Riemann integrable. For $x \in [a, b]$ define F by*

$$F(x) = \int_a^x f(t)dt. \tag{44}$$

Then F is differentiable at every point $x \in (a, b)$ and $F'(x) = f(x)$.

Proof. Fix $x \in (a, b)$ and choose $y \in (a, b)$ close to x. Then

$$\frac{F(y) - F(x)}{y - x} = \frac{1}{y - x}\left[\int_a^y f(t)dt - \int_a^x f(t)dt\right] = \frac{1}{y - x}\int_x^y f(t)dt \tag{45}$$

whence

$$\frac{F(y) - F(x)}{y - x} - f(x) = \frac{1}{y - x}\int_x^y [f(t) - f(x)]dt. \tag{46}$$

So given $\epsilon > 0$ we can choose $\delta > 0$ such that $|y - x| < \delta$ implies that the absolute value of the right hand side is less than ϵ. □

Theorem 19. (**Fundamental Theorem of Calculus**) *Suppose that* $f : [a, b] \to \mathbb{R}$ *and* $G : [a, b] \to \mathbb{R}$ *are continuous functions,* G *is differentiable everywhere in* (a, b)*, and* $G'(x) = f(x)$*. Then*

$$\int_a^b f(x)dx = G(b) - G(a)$$

Proof. Take F from the statement of the lemma above, and notice that $F - G$ has derivative 0, so it must be constant. Therefore,

$$F(b) - G(b) = F(a) - G(a) = -G(a) \text{ implies } F(b) = G(b) - G(a).$$

But $F(b) = \int_a^b f(x)dx$, completing the argument. □

5.3 Properties of the Integral

As with the case of the derivative, we now run through a few important properties of the integral that will come handy later on.

Theorem 20. (**integration by parts**) *Suppose that* $f : [a, b] \to \mathbb{R}$ *and* $g : [a, b] \to \mathbb{R}$ *have continuous first derivatives on* (a, b)*. Then if* x *and* y *are elements of* (a, b)*,*

$$\int_x^y f(t)g'(t)dt = \big[f(y)g(y) - f(x)g(x)\big] - \int_x^y f'(t)g(t)dt \tag{47}$$

Proof. Simply integrate the product rule in the form: $fg' = (fg)' - f'(g)$. □

Exercise 40. Find $\int_0^\pi x(\sin x)dx$.

Theorem 21. (**change of variables**) *Suppose that* g *is differentiable on the open interval* (a, b) *and that its derivative is continuous. Let* T *be an open interval such that* $g(x) \in T$ *for all* $x \in (a, b)$*. If a function* f *is continuous on* T *then the composition* $f \circ g$ *is continuous on* (a, b)*, and*

$$\int_a^b [f \circ g](x)g'(x)dx = \int_{g(a)}^{g(b)} f(g)dg \tag{48}$$

Proof. Simply integrate the chain rule. □

Exercise 41. Find $\int_0^1 x\sqrt{(1 - x^2)}dx$.

Improper Integrals You will be pleased to know that if f is Riemann integrable on the interval $[a, b]$ its integral on $(a, b]$, $[a, b)$ and (a, b), are all defined to be the same as its integral on $[a, b]$. You will also be pleased to know that $a = -\infty$ and/or $b = \infty$ are allowed so long as you remember that

$$\int_a^\infty f(x)dx \text{ is really } \lim_{b\to\infty} \int_a^b f(x)dx \text{ and } \int_{-\infty}^b f(x)dx \text{ is really } \lim_{a\to-\infty} \int_a^b f(x)dx.$$

5.4 The Integral in Multiple Dimensions

For $X \subset \mathbb{R}^n$ let $f : X \to \mathbb{R}$ be a continuous function. We would like to develop a notion of the integral

$$\int_A f(x)dx$$

that corresponds to computing the volume under f in a region $A \subseteq X$. Suppose, for example, that $A = [a_1, b_1] \times [a_2, b_2]$ so that it is a rectangle in $\mathbb{R}^2$. It seems natural to think of partitioning this rectangle into small rectangles (just as we partitioned the interval in $\mathbb{R}$ into small intervals), compute the volume of f under each small rectangle, and then take the limit as the partition gets finer. This is exactly what we do, even though graphical intuitions may not carry over into higher dimensions.

Suppose that $A = [a_1, b_1] \times ... \times [a_n, b_n]$. Let $f^1 = f$ and view $f^1(x_1; x_2, \ldots, x_n)$ as a function of only x_1, i.e. we are holding $x_2, \ldots, x_n$ constant. It is good to know that even in the multivariate setting the continuity of f allows us to define the Riemann integral,

$$f^2(x_2; x_3, \ldots, x_n) = \int_{a_1(x_2,\ldots,x_n)}^{b_1(x_2,\ldots,x_n)} f(x_1; x_2, \ldots, x_n)dx_1, \tag{49}$$

which is a continuous function only of $x_2, \ldots, x_n$, but viewed as a function of x_2 while holding $x_3, \ldots, x_n$ constant. Now hold $x_3, \ldots, x_n$ constant and consider

$$f^3(x_3; x_4, \ldots, x_n) = \int_{a_2(x_3,\ldots,x_n)}^{b_2(x_3,\ldots,x_n)} f^2(x_2; x_3, \ldots, x_n)dx_2. \tag{50}$$

Continue this idea until you have integrated with respect to x_n. What you have calculated then is the multiple integral

$$\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n)dx_1 \ldots dx_n. \tag{51}$$

Theorem 22. *Suppose that $A = [a_1, b_1] \times ... \times [a_n, b_n] \subseteq X \subset \mathbb{R}^n$ and f is a continuous real-valued function defined on X. Then,*

$$\int_A f(x)dx = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_n)dx_1 \ldots dx_n.$$

An implication is that the order of the variables you integrate with respect to does not matter; you get the same answer at the end. This fact is called **Fubini's theorem**.

But what happens if A does not have the form $[a_1, b_1] \times ... \times [a_n, b_n]$? For example, what if A is a circle, and we want to compute the volume under f in the region A. Fortunately, the change of variables generalizes in the following way.

Theorem 23. (**generalized change of variables**) *Suppose that f is a continuous real-valued function defined on $\mathbb{R}^n$. Suppose that $U \subset \mathbb{R}^n$ be an open set and $g : U \to \mathbb{R}^n$ an*

injective function with continuous partial derivatives. Consider a bounded open set $A \subset \mathbb{R}^n$ such that there is a compact set B for which $A \subset B \subset U$ and suppose that the ***Jacobian*** *matrix, $J(x) = [\partial g_i/\partial x_j]_{i=1,...,n}^{j=1,...,n}$, is invertible for all $x \in A$. If $[f \circ g](A)$ is bounded, then*

$$\int_{g(A)} f(x)dx = \int_A [f \circ g](x) \left|\det J(x)\right| dx$$

Exercise 42. Let A be the region:

$$A = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 \geq -x_1 - 1, x_2 \leq -x_1 + 1, x_2 \geq x_1 - 1, x_2 \leq x_1 + 1\right\}.$$

Now calculate

$$\int_A \left(\frac{x_1 - x_2}{x_1 + x_2 + 2}\right)^2 d(x_1, x_2). \tag{52}$$

Exercise 43. (**differentiating under the integral**) Suppose that $S \subset \mathbb{R}^n$ is an open box and $f : S \times [a, b] \to \mathbb{R}$ is a continuous function with continuous partial derivatives $\partial f/\partial x_i$ for $i = 1, ..., n$. Then the function $\varphi(x) = \int_a^b f(x, t)dt$ is continuously differentiable. Take this as given and prove that

$$\frac{\partial \varphi(x)}{\partial x_i} = \int_a^b \frac{\partial f(x, t)}{\partial x_i} dt.$$

Hint: Simple proofs use Fubini's theorem and Leibniz's rule. If it is more convenient, feel free to focus on the case of $S \subset \mathbb{R}$ being an open interval (i.e. the $n = 1$ case).

5.5 Taylor's Theorem

We now prove two versions of a remarkably useful theorem, the first for functions of one variable and the second for functions of multiple variables. These results have many applications– for example, in our later proof of the Central Limit Theorem.

Theorem 24. (**Taylor's Theorem**) *Let S be an open interval and $f : S \to \mathbb{R}$ a function whose first, second,..., and mth derivatives exist and are all continuous. Denote the jth derivative of f by $f^{(j)}$. For $a, b \in S$, and $n \leq m$ we have*

$$f(b) = f(a) + \frac{f'(a)}{1!}(b - a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b - a)^{n-1} + R_n \tag{53}$$

where

$$R_n = \int_a^b \frac{(b - t)^{n-1}}{(n-1)!} f^{(n)}(t)dt. \tag{54}$$

Proof. The fundamental theorem of calculus tells us that $f(b) = f(a) + \int_a^b f'(t)dt$. Notice that this is simply equation (53) for $n = 1$. Now assume that (53) is true. Perform the

integration in (54) by parts where the function whose derivative is not taken is $f^{(n)}(t)$, and the derived function is $\frac{(b-t)^{n-1}}{(n-1)!}$. If we do this we get

$$R_n = \frac{(b-a)^n}{n!} f^{(n)}(a) + \int_a^b \frac{(b-t)^n}{n!} f^{(n+1)}(t) dt. \tag{55}$$

Plug this back into (53) and, by induction, we are done. □

Exercise 44. Assume in the above theorem that S contains 0 and remember that because $f^{(n)}(x)$ is continuous and $[a, b]$ is compact, the function is bounded on $[a, b]$. Argue that under these conditions, the **Taylor polynomial**

$$P_{n-1}(x) = f(0) + f'(0)x + \frac{f^{(2)}(0)}{2!}x^2 + \cdots + \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} \tag{56}$$

is a good approximation of f when n is large and $x \in [a, b]$. More formally, prove that for all $\epsilon > 0$, and for all $x \in [a, b]$ there exists δ and N such that if y is a point in $[a, b]$ that is within δ of x (i.e., $|x - y| < \delta$) then $|P_{n-1}(y) - f(x)| < \epsilon$ for all $n \geq N$.

Exercise 45. Recall from the Weierstrass Theorem that because it is continuous, f attains a maximum and a minimum on the interval $[a, b]$. Use the mean value theorems to show that there is a number $c \in [a, b]$ such that

$$R_n = f^{(n)}(c)\frac{(b-a)^n}{n!}. \tag{57}$$

What does this say for (53)?

Theorem 25. (**generalized Taylor's theorem**) *Let $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is an open set. If f is C^1 then for any $x \in S$ and $y \in S$, we can write*

$$f(y) = f(x) + \nabla f(x)(y - x) + R_1(x, y), \tag{58}$$

where R_1 is a function with the property

$$\lim_{y \to x} \left(\frac{R_1(x, y)}{||x - y||} \right) = 0. \tag{59}$$

If f is C^2 then for any $a \in S$ and $b \in S$, we can write

$$f(y) = f(x) + \nabla f(x)(y - x) + \frac{1}{2}(y - x)' Hf(x)(y - x) + R_2(x, y), \tag{60}$$

where $Hf(x)$ is the Hessian matrix and R_2 is a function with the property

$$\lim_{y \to x} \left(\frac{R_2(x, y)}{||x - y||^2} \right) = 0, \tag{61}$$

and where $\lim_{y \to x}$ is the limit for the convergence of vectors (i.e., the limit holds for all sequences of vectors $\{x_n\}$ that converge to vector y such that $x_n \neq y$ for all n).

Exercise 46. Write down three terms of the Taylor series and evaluate the it at $(x, y) = (1, 1)$ for $f(x, y) = \log(xy)$.

6 Optimization

The **interior** of a set $S \subset \mathbb{R}^n$ is the set

$$\text{int } X = \{x \in S : \exists \text{ an open box } B \text{ such that } x \in B \subset S\}$$

Similarly, the **boundary** of S, denoted ∂S, is the set

$$\partial S := \{x \in \mathbb{R}^n : \text{every open box } B \text{ s.t. } x \in B \text{ contains points } y \in S \text{ and } z \notin S\}.$$

The following result is intuitive, and I encourage you to translate your intuition for why it must be true into a proof. In fact, you can do it in an exercise.

Theorem 26. *Suppose $S \subset \mathbb{R}$ and suppose that $f : S \to \mathbb{R}$ is differentiable function on the interior of S. If f achieves a maximum or minimum in the interior of S then $f'(x) = 0$ at the point x at which it achieves a maximum or minimum. If $x \in$ int S is a maximum then $f''(x) \leq 0$, and if it is a minimum then $f''(x) \geq 0$.*

Note that $f'(x) = 0$ and $f''(x) \leq 0$ are only necessary conditions for $x \in \text{int } S$ being a maximum. This exercise demonstrates how they are insufficient.

Exercise 47. Let $f : [-2, 5] \to \mathbb{R}$ be defined by $f(x) = x^3 - 4.5x + 6x + 1$. By the Weierstrass theorem, f achieves a maximum and minimum on its domain. We can find the values of x where $f'(x) = 0$. Which of these are maximum values of f and minimum values of f? Calculate $f(x)$ at points where $f'(x) = 0$. Then calculate $f(-1)$ and $f(4)$.

What are the issues with sufficiency? Part of the problem is that of "global" versus "local" minima and maxima. A **global** maximum of a real-valued function f defined on $S \in \mathbb{R}^n$ is a point $x \in X$ such that $f(x) \geq f(y)$ for all $y \in S$. A **local** maximum of the same function is a point $x \in S$ such that there exists an open box $B \subset S$ that contains x for which $f(x) \geq f(y)$ for all $y \in B$. Obviously, global maxima are local but not the other way around.

The problem of sufficiency does not arise in the one-dimensional case, $S \subset \mathbb{R}$, when S is a closed and convex set (i.e., closed interval) and f is either concave or convex.

Theorem 27. *If f is a concave (resp. convex) real-valued function on some closed interval $S \subset \mathbb{R}$ and $f'(x) = 0$ for some $x \in$ int S, then x is a global maximum (resp. minimum) of f on S. If f is strictly concave (resp. strictly convex) then it is the unique global minimum (resp. maximum) of f on S. If f is differentiable on int S but there is no x such that $f'(x) = 0$, then the extrema (maxima and minima) of f lie on ∂S.*

Exercise 48. Prove (or convince yourself) that the two theorems above are true.

6.1 Unconstrained Optimization

Let's extend the ideas above to the case where f is a real-valued function defined on a set $S \subset \mathbb{R}^n$ that is closed and convex. As before, we say that f is **concave** if $\forall x, y \in S$ such that $y \neq x$ we have,

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) \quad \forall \alpha \in (0,1); \tag{62}$$

and f is **strictly concave** if in the inequality above, you replace $\geq$ with just $>$. It is **convex** if the function $-f$ is concave and **strictly convex** if $-f$ is strictly concave.

Exercise 49. Continue to let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a closed and convex set. Prove that f is concave if and only if the function $g(t) = f(x + tz)$ is concave on the set $T = \{t \in \mathbb{R} \mid x + tz \in S\}$, where $x \in S$, $z \in S$ and $z \neq 0_n$.

Before proceeding any further, I will define an important concept. An $n \times n$ matrix A is said to be **negative semi-definite** if for all $x \in \mathbb{R}^n$, we have $x'Ax \leq 0$. If the inequality is strict for all $x \neq 0_n$ then A is **negative definite**. A is **positive definite** if $-A$ is negative definite and **positive semi-definite** if $-A$ is negative semi-definite.

Lemma 3. (**Multidimesnional Concavity Lemma**) *Consider the function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is convex and has a nonempty interior. Assume that f has a Hessian where all the entries are continuous. Then the following three statements are equivalent: (i) f is concave, (ii) $Hf(x)$ is negative semi-definite for all $x \in S$, and (iii) $f(x) \leq f(y) + \nabla f(y)(x - y)$ for all $x, y \in S$. In addition, if $Hf(x)$ is negative definite for all $x \in S$ then f is strictly concave.*

Proof. I demonstrate only that (i) and (ii) are equivalent, as well as the additional claim about strict concavity. I first show that (i) implies (ii), and hope that you are convinced that it is enough to establish the claim on the interior of S since continuity takes care of the boundary points. Now let g be the function and T be the set both defined for Exercise 49. Notice that

$$g'(t) = \nabla f(x + tx) \cdot z = \sum_{i=1}^{n} \frac{\partial f(x + tx)}{\partial x_i} z_i \tag{63}$$

because $g(t) = f(x + tz)$ implies that

$$\lim_{h \to 0} \frac{g(t+h) - g(t)}{h} = \lim_{h \to 0} \frac{f(x + tz + hz) - f(x + tz)}{h} \tag{64}$$

and the right hand side is the definition of directional derivative of f at $x+tz$ in the direction of z. (Remember that that is $\nabla f(x) \cdot z$.) Now take the second derivative of g knowing that that is the directional derivative of each of the partial derivatives of f added:

$$g''(t) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f(x + tz)}{\partial x_j x_i} z_i z_j. \tag{65}$$

But this is just the equation $g''(t) = z'Hf(x+tz)z$. Now f is concave, which by Exercise 49 means that g is concave, and because z is arbitrary (as long as we are in the set T), that means that $z'Hf(x+tz)z \leq 0$, i.e. Hf is negative semi-definite since the argument holds for any x and any z such that $x+tz$ is in the domain of f. Proving (ii) implies (i) is simply writing the argument backwards, and the claim about strict concavity is merely replacing the $\leq$ sign by the $<$ sign in that proof. □

Exercise 50. Complete the proof of the lemma above.

Now for the definition of a critical point. A **critical point** of a function $f : S \to \mathbb{R}$ with $S \subset \mathbb{R}^n$ is a vector $x \in \text{int } S$ such that $\nabla f(x) = 0_n$.

Theorem 28. (**First Order Conditions Theorem**) *If the differentiable function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$, reaches a local interior extremum (i.e., maximum or minimum) at x^* then x^* is a critical point of f.*

Proof. Again write $f(x^* + tz)$ in terms of the familiar g of Exercise 52 (yes, this time x^* instead of x). Note that $g(0) = f(x^*)$. Because g coincides with some value of f for every t, then $g(t)$ must reach a local extremum at $t = 0$ if f reaches an extremum at x. This means that $g'(0) = 0$. Now recall from the proof of the multidimensional concavity lemma why it should be that (63) should hold with x replaced by x^*. Combined with $g'(0) = 0$, this implies that $\nabla f(x^*)z = 0_n$. Since this holds for any z, it must hold for all $e_j, j = 1, \ldots, n$. Therefore, each of the partials are 0, and we have $\nabla f(x^*) = 0_n$. □

Theorem 29. (**Second Order Conditions Theorem**) *If the C^2 function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$, reaches a local interior maximum (resp. minimum) at x^*, then $Hf(x^*)$ is negative (resp. positive) semidefinite.*

Proof. Recall that (65) must be true. Then if f is maximized at x it must be that g is maximized at $t = 0$, and thus $g''(0) \leq 0$. That means that the right hand side of (65) is nonpositive, or that the Hessian of f at x is negative semidefinite. (This is again because z is arbitrary.) A similar argument holds for the "minimized" case. □

Theorem 30. (**Global Maximum Theorem**) *If the C^2 function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$, has a negative semidefinite Hessian at every x in the interior of S, then if x^* is a critical point of f, the function f achieves a global maximum at x^*.*

Proof. Because the Hessian is everywhere negative semidefinite, f is concave by the multidimensional concavity lemma. The same lemma tells us that $f(x) \leq f(x^*) + \nabla f(x^*)(x - x^*)$ for all $x \in S$. Now if $\nabla f(x^*) = 0_n$ then $f(x) \leq f(x^*)$ for all $x \in S$. So f reaches a global maximum. □

A similar theorem holds for global minima. However, in most applications if what you are using the first and second order conditions to find interior minima of f, you can do this by finding interior maxima of $-f$.

Exercise 51. Re-read the proof of the global maximum theorem. Then prove that if the Hessian of f is negative definite at every x in its domain, and x^* maximizes f, then x^* is the *unique* global maximizer of f.

Now here is a useful result to check the definiteness of the Hessian. To state the result, I'll define a **submatrix** of a matrix A informally to be a matrix that is a portion of A.

Theorem 31. *Let $D_1(x)$ be the determinant of the top left 1×1 submatrix of the $n \times n$ matrix $Hf(x)$, $D_2(x)$ be the determinant of the top left 2×2 submatrix of $Hf(x)$,..., and $D_k(x)$ be the determinant of the top left $k \times k$ submatrix of $Hf(x)$. If $(-1)^i D_i(x) > 0$ for all $i = 1, \ldots, n$ then $Hf(x)$ is negative definite, and if this is true for all x in the domain of f then f is strictly concave. If $D_i(x) > 0$ for all $i = 1, \ldots, n$ then $Hf(x)$ is positive definite, and if this is true for all x in the domain of f then f is strictly convex.*

Exercise 52. Read and understand the theorem above.

Exercise 53. Refer back to the application called "The Algebra of Least Squares." Review it and prove that the $(k+1)$-dimensional vector of partial derivatives of $S(\beta)$ with respect to the components of the vector of variables $\beta = (m_1, ..., m_k, b)$, which we abbreviate as $\partial S/\partial \beta$, is given by

$$\frac{\partial S(\beta)}{\partial \beta} = -2y'X + 2X'X\beta.$$

Then show that the value of β that minimizes $S(\beta)$ is the value of β that solves

$$0_n = \frac{\partial S(\beta)}{\partial \beta}.$$

Finally, compute the value of β that minimizes $S(\beta)$ as a function of the data only, i.e. as a function of $(\bar{X}, y)$.

6.2 Equality-Constrained Optimization

Let $f, g_1, ..., g_k$ be C^1 functions defined on some set $S \subset \mathbb{R}^n$. The problem we are interested in is the problem of maximizing a function f (or, equivalently, minimizing $-f$) subject to the constraints that $g_i(x) = 0$. That is, we are looking for a point $x \in S$ that maximizes f on the set

$$\mathcal{D} := \{x \in S : g_i(x) = 0, \forall i\}.$$

We write the problem as

$$\max_{x \in S} f(x) \qquad \text{subject to} \qquad g_i(x) = 0, \ \forall i \tag{P1}$$

Alternatively, we could have written

$$\max_{x \in \mathcal{D}} f(x) \tag{66}$$

The following theorem gives us some ideas about how we may search for a solution.

Theorem 32. (**Lagrange's Theorem**) *Let $S \subset \mathbb{R}^n$. Let $f : S \to \mathbb{R}$ and $g_i : S \to \mathbb{R}, i = 1, \ldots, k$, be C^1 functions. Let x^* be a point in the interior of S and suppose that x^* is an optimum (local maximum or local minimum) of f subject to the constraints, $g_i(x) = 0, i = 1, \ldots, k$. If the gradient vectors $\nabla g_i(x^*), i = 1, \ldots, k$, are linearly independent then there exists a vector $\Lambda = (\lambda_1, \ldots, \lambda_k)' \in \mathbb{R}^k$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{k} \lambda_i \nabla g_i(x^*) = 0_n \tag{67}$$

This theorem is kind of important, so let me provide some intuition and warnings before giving the proof. Recall that int S denotes the interior of S and write the problem as

$$\max_{x \in \text{int } S} f(x) \text{ subject to } g_i(x) = 0, i = 1, \ldots, k.$$

Now define the so-called **Lagrangean** for this problem,

$$\mathcal{L}(x, \Lambda) := f(x) + \sum_{i=1}^{k} \lambda_i g_i(x),$$

and apply the first order conditions theorem to $\mathcal{L}$ while deleting the last k equations:

$$\nabla_n \mathcal{L}(x, \Lambda) = \nabla f(x) + \sum_{i=1}^{k} \lambda_i \nabla g_i(x) = 0_n, \tag{68}$$

where $\nabla_n \mathcal{L}$ simply means the vector with only the first n entries of usual gradient of $\mathcal{L}$, since differentiating with respect to Λ gives us back the constraints. Thus the information that we discard was something we already knew.

Now we have to be convinced of two things. First, we must be convinced that the solution to the constrained maximization problem, x, is a vector of the first n entries of a critical point of $\mathcal{L}(x, \Lambda)$. To convince yourself, understand that the directional derivative $\nabla f(x) \cdot h = 0$ at the maximum for all small length vectors h take x to points in which the constraints g_i continue to be satisfied. If the g_i continue to be satisfied after moving a small amount in the direction h, then no change occurs in any g_i, i.e. $\nabla g_i \cdot h = 0$ for all i. But that means that $\nabla_n \mathcal{L}(x, \Lambda) \cdot h = \nabla f(x) \cdot h$ for any movement h that keeps the constraints

satisfied. This says that you cannot increase or decrease the objective function f by making small movements in any "permissible" direction. So we must be at a critical point of $\mathcal{L}$ (except that we don't know the λ_i). Suppose we were at a maximum or minimum at x. Then the constraints would be satisfied and so would (67).

The second thing that we must convince ourselves is that the vector Λ exists. That's a little harder, and for that we will rely on the proof.

Proof of Theorem 32. I prove the statement only for $k = 1$, noting that the general result follows a very similar argument. Let $g(x) = g_1(x)$ so that we can drop the subscript from now on. Let the local optimum be x^*. The rank condition on g tells us that $\nabla g(x^*) \neq 0_n$. Without loss of generality, assume that the first component of this vector is nonzero. Denote the first coordinate of a vector $x \in \mathcal{D}$ by w and the last $n-1$ of them by z. (Recall the definition of $\mathcal{D}$ from above.) Write $x = (w, z)$. Let $x^* = (w^*, z^*)$ denote the local optimum. Let $\nabla f_w(w, z)$ denote the derivative of f with respect to w alone and $\nabla f_z(w, z)$ the derivative with respect to z alone. The derivative of g is partitioned analogously into a number, $\nabla g_w(w, z)$, and a vector $\nabla g_z(w, z)$ of size $n - 1$.

To prove the theorem we must show that there exists $\lambda \in \mathbb{R}$ such that

1. $\nabla f_w(w^*, z^*) + \lambda \nabla g_w(w^*, z^*) = 0$
2. $\nabla f_z(w^*, z^*) + \lambda \nabla g_z(w^*, z^*) = 0_{n-1}$

To show this, we have to use the implicit function theorem (Theorem 15). This says that there is an open box $B \subseteq \mathbb{R}^{n-1}$ containing z^* and a C^1 function $h : B \to \mathbb{R}$ such that $h(z^*) = w^*$ and $g(h(z), z) = 0$ for all $z \in B$. Also,

$$\nabla h(z) = -\frac{\nabla g_z(h(z), z)}{\nabla g_w(h(z), z)} \tag{69}$$

which is none other than the formula in the implicit function theorem.

Define λ now as

$$\lambda = -\frac{\nabla f_w(w^*, z^*)}{\nabla g_w(w^*, z^*)}$$

which rearranges to

$$\nabla f_w(w^*, z^*) + \lambda \nabla g_w(w^*, z^*) = 0.$$

That's the first thing we had to show, which is a bit simpler than the second. Define the function $\phi : B \to \mathbb{R}$ by $\phi(z) = f(h(z), z)$. Since f has a local optimum at $(w^*, z^*) = (h(z^*), z^*)$, then ϕ has a local optimum at z^*. Since B is open, z^* is an unconstrained local optimum of ϕ and the first-order conditions for an unconstrained optimum imply $\nabla \phi(z^*) = 0_{n-1}$, i.e. by the chain rule:

$$\nabla f_w(w^*, z^*) \nabla h(z^*) + \nabla f_z(w^*, z^*) = 0_{n-1}. \tag{70}$$

Substitute (69) in this to get

$$\nabla f_z(w^*, z^*) + \lambda \nabla g_z(w^*, z^*) = 0_{n-1}.$$

and that's it. □

Exercise 54. We did not prove the chain rule appearing in (70) in the multidimensional case. In the case of two variables, let $x(t)$, $y(t)$ be two differentiable functions of t and let $f(x, y)$ be a differentiable function. For the purposes of this demonstration, define $\partial x = x(t + h) - x(t)$ and $\partial y = y(t + h) - y(t)$. Then

$$\begin{aligned} f'(x(t), y(t)) &= \lim_{h \to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h} \\ &= \lim_{h \to 0} \frac{f(x + \partial x, y + \partial y) - f(x, y + \partial y) + f(x, y + \partial y) - f(x, y)}{h} \\ &= \lim_{h \to 0} \frac{f(x + \partial x, y + \partial y) - f(x, y + \partial y)}{h} + \lim_{h \to 0} \frac{f(x, y + \partial y) - f(x, y)}{h} \end{aligned}$$

On the right is the definition of the partial of f with respect to t through y, which by the single variable chain rule is

$$\frac{\partial f}{\partial y} \frac{dy}{dt}$$

Apply the mean value theorem to the limit on the left by picking an $x' \in [x, x + \partial x]$ such that the limit is equal to

$$\lim_{h \to 0} \frac{\partial x}{h} \frac{\partial f(x')}{\partial x} = \frac{\partial f}{\partial x} \frac{dx}{dt}.$$

That gives us

$$f'(x(t), y(t)) = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}.$$

Use an extended argument based on this demonstration to argue that (70) is true.

Caution with Lagrange Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x, y) = -y$ and $g : \mathbb{R}^2 \to \mathbb{R}$ be the function $g(x, y) = y^3 - x^2$. Notice that the maximum of f subject to $g(x, y) = 0$ is at $(0, 0)$, since if y is negative x would have to be the square root of a negative number for the constraint to be met. If you set up the Lagrangean, $\mathcal{L}$ of this problem (defined in the previous lecture note) and take the first order conditions then you get

$$\begin{aligned} -2\lambda x &= 0 \\ -1 + 3\lambda y^2 &= 0 \\ -x^2 + y^3 &= 0 \end{aligned}$$

Now for the first equation to be true, either $\lambda = 0$ or $x = 0$. If $\lambda = 0$ then the second equation is a contradiction. If $x = 0$ then the third implies that $y = 0$. Plug this back into the second and obtain a contradiction again.

Exercise 55. What went wrong in the example above?

Now suppose that f is defined by

$$f(x,y) = \frac{1}{3}x^3 - \frac{3}{2}y^2 + 2x \tag{71}$$

instead, and g was the benign $g(x,y) = x - y$. The constraint qualification is met since $\nabla g(x,y) = (1,-1)$ for all x and y. Now set up the Lagrangean and take the first order conditions:

$$\begin{aligned} x^2 + 2 + \lambda &= 0 \\ -3y - \lambda &= 0 \\ x - y &= 0 \end{aligned}$$

There are two solutions: $(x,y) = (2,2)$ or $(1,1)$, and seeing that $f(2,2) = 2/3$ while $f(1,1) = 5/6$ you could guess that $(2,2)$ is a minimum and $(1,1)$ is a maximum. But in fact, $f(0,0) = 0$ and $f(3,3) = 1.5$.

Exercise 56. What went wrong here?

Second Order Conditions The conditions in Lagrange's theorem don't tell us if we are at a minimum or maximum of the function over the constraint. In the appendix, I go over the second order conditions of the the theorem to address this issue.

6.3 Inequality-Constrained Optimization

In many applications, the constraints in a constrained optimization problem take the form of inequalities rather than equalities. For example, "maximize welfare subject to not spending more than a certain budget." The constraint is that spending must be smaller than or equal to the budget; the optimum may be at a point where the entire budget is spent, or it may not. Then, inheriting the notation of the previous section, we consider the problem

$$\max_{s \in S} f(x) \qquad \text{subject to} \qquad g_i(x) \geq 0, \ \forall i \tag{P2}$$

Theorem 33. (**The Karush-Kuhn-Tucker Theorem**) *Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$ and $g_i : S \to \mathbb{R}$, $i = 1, \ldots, k$, be C^1 functions. Let x^* be a point in the interior of S and suppose that x^* is an optimum (local maximum or local minimum) of f subject to the constraints, $g_i(x) \geq 0$, $i = 1, \ldots, k$. If the gradient vectors $\nabla g_i(x^*)$, $i = 1, \ldots, k$, are linearly independent then there exists a vector $\Lambda = (\lambda_1, \ldots, \lambda_k)' \in \mathbb{R}^k$ such that*

$$\nabla f(x^*) + \sum_{i=1}^{k} \lambda_i \nabla g_i(x^*) = 0_n \text{ and } \lambda_i g_i(x^*) = 0 \text{ for all } i = 1, \ldots, k.$$

In addition, λ_i, $i = 1, ..., k$, are all nonnegative if x^* is a maximum and nonpositive if it is a minimum.

This one I'm not going to prove, but I will give some intuition for it. Begin with the simple problem

$$\max_x f(x) \text{ subject to } x \geq 0$$

and notice that if x^* is the solution then it satisfies one of the following three cases:

1. $x^* = 0$ and $f'(x^*) < 0$,
2. $x^* = 0$ and $f'(x^*) = 0$,
3. $x^* > 0$ and $f'(x^*) = 0$.

These imply that

1. $f'(x^*) \leq 0$,
2. $x^*[f'(x^*)] = 0$,
3. $x^* \geq 0$.

In the world of $\mathbb{R}^n$, these correspond to

1. $\frac{\partial f(x^*)}{\partial x_i} \leq 0$,
2. $x_i^* \left[\frac{\partial f(x^*)}{\partial x_i}\right] = 0$
3. $x_i^* \geq 0$,

which must hold for all $i = 1, \ldots, n$ if x^* maximizes $f(x)$ subject to $x_i \geq 0$ for all i. Now convince yourself that the problem

$$\max_{x_1, x_2} f(x_1, x_2) \text{ subject to } g(x_1, x_2) \geq 0,$$

is equivalent to

$$\max_{x_1, x_2, z} f(x_1, x_2) \text{ subject to } g(x_1, x_2) - z = 0 \text{ and } z \geq 0.$$

Take the first order conditions of the Lagrangian, $\mathcal{L}$ of this latter beast while ignoring the inequality constraint, and arrive at

$$\begin{aligned} \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} &= 0 \\ \frac{\partial f}{\partial x_2} + \lambda \frac{\partial g}{\partial x_2} &= 0 \\ \frac{\partial f}{\partial \lambda} = g(x_1, x_2) - z &= 0 \end{aligned}$$

These give the critical points of $\mathcal{L}$. The only additional necessary conditions come from the inequality constraint $z \geq 0$. The insight from the previous discussion tells us the equivalent of properties (1) – (3) for z, i.e that:

$$\begin{aligned} -\lambda &\leq 0 \\ z(-\lambda) &= 0 \\ z &\geq 0 \end{aligned}$$

Summarizing, we have

$$\begin{aligned} \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} &= 0 \\ \frac{\partial f}{\partial x_2} + \lambda \frac{\partial g}{\partial x_2} &= 0 \\ \lambda g(x_1, x_2) &= 0 \\ \lambda &\geq 0 \\ g(x_1, x_2) &\geq 0 \end{aligned}$$

These are called **Kuhn-Tucker conditions**.

Exercise 57. Re-read the cautionary examples for Lagrange and consider the problem of finding (x, y) to maximize $f(x, y) = -(x^2 + y^2)$ subject to $h(x, y) = (x-1)^3 - y^2 \geq 0$. Find a solution to this problem just by looking at the functions. (You don't have to set up any Lagrangean or Karush-Kuhn-Tucker conditions). Now show that this problem cannot be analyzed using the Karush-Kuhn-Tucker theorem. Which of the assumptions is violated?

The exercise above demonstrates that the same caution should be taken when applying the Karush-Kuhn-Tucker theorem as should be taken when applying Lagrange's theorem. However, we have the following result that gives sufficient conditions for describing a solution to the constrained optimization problem with the Kuhn-Tucker conditions.

Theorem 34. (**Sufficient Conditions Kuhn-Tucker Theorem**) *Let $S \subset \mathbb{R}^n$ be an open set and $f : S \to \mathbb{R}$ be a concave C^1 function and $g_i : S \to \mathbb{R}$, $i = 1, \ldots, k$, be C^1 functions such that*

$$\mathcal{D} = \{x \in S : g_i(x) \geq 0, \ \forall i\}$$

is a convex set. If $x^* \in X$ and there are numbers $\lambda_1, \ldots, \lambda_k \geq 0$ such that

$$\nabla f(x^*) + \sum_{i=1}^{k} \lambda_i \nabla g_i(x^*) = 0 \text{ and } \lambda_i g_i(x) = 0 \text{ for all } i = 1, \ldots, k,$$

then x^* solves the program

$$\max_{x \in \mathcal{D}} f(x).$$

Proof. Suppose the theorem isn't true, i.e. there is an $x \in \mathcal{D}$ such that $f(x) > f(x^*)$. Let $v = x^* - x$ and begin by writing the definition of the directional derivative of f at x^* in the direction $-v$:

$$\begin{aligned} -\nabla f(x^*) \cdot v &= \lim_{t \to 0} \frac{f(x^* - tv) - f(x^*)}{t} \\ &= \lim_{t \to 0} \frac{f((1-t)x^* + tx) - f(x^*)}{t} \\ &\geq \lim_{t \to 0} \frac{(1-t)f(x^*) + tf(x) - f(x^*)}{t} \\ &= f(x) - f(x^*) > 0 \end{aligned} \tag{72}$$

where the weak inequality in (72) is due to the concavity of f and the fact that small $t > 0$ probably means $t < 1$ at least. The last line is due to the assumption.

Now, due to the convexity of $\mathcal{D}$, we have $x^* - tv = (1-t)x^* + tx \in \mathcal{D}$ for small t. So for each g_i such that $g_i(x^*) = 0$ we have

$$-\nabla g_i(x^*) \cdot v = \lim_{t \to 0} \frac{g_i(x^* - tv) - g_i(x^*)}{t} \geq 0. \tag{73}$$

For all the rest, we have $\lambda_i = 0$ as per one of the Karush-Kuhn-Tucker conditions. Gathering all observations, we have the following contradiction:

$$0 > \nabla f(x^*) \cdot v = -\left(\sum_{i=1}^{k} \lambda_i \nabla g_i(x^*) \right) \cdot v = -\left(\sum_{i=1}^{k} \lambda_i \nabla g_i(x^*) \cdot v \right) \geq 0. \tag{74}$$

We are done. □

6.4 The Envelope Theorem

Consider the problem

$$\max_x f(x, a) \text{ subject to } g(x, a) = 0 \text{ and } x \geq 0, \tag{P3}$$

where x is the usual vector of variables (size n), and a is a vector of parameters (size m). Suppose that for each vector a, the solution to this problem is unique, and denote it $x(a)$. Now define the **maximum value function**,

$$M(a) = f(x(a), a);$$

in other words, M is a function of a subject to x having been chosen to solve Problem (P3). Now suppose we would like to analyze how M varies as a varies.

Theorem 35. (**Envelope Theorem**) *Consider Problem* (P3) *and assume that f and g are C^1 in a. For each a let $x(a)_j \geq 0$ for all $j = 1, \ldots, n$ and assume that the $x(a)_j$ are*

also C^1 in a. Let $\mathcal{L}(x, a, \lambda)$ be the Lagrangean for Problem (P3) *and let $(x(a); \lambda(a))$ solve the Kuhn-Tucker conditions for the problem. Let $M(a)$ be the maximum value function for f. Then*

$$\frac{\partial M(a)}{\partial a_j} = \frac{\partial \mathcal{L}}{\partial a_j} \text{ evaluated at } (x(a), \lambda(a)), \text{ for all } j = 1, \ldots, m.$$

Proof. First write the Lagrangian

$$\mathcal{L} = f(x, a) + \lambda g(x, a).$$

If $x(a)$ solves Problem (P3), then Karush-Kuhn-Tucker says

$$\nabla_n f(x(a), a) + \lambda(a) \nabla_n g(x(a), a) = 0 \tag{75}$$

$$g(x(a), a) = 0 \tag{76}$$

which define solutions $(x(a), \lambda(a))$. Then note

$$\frac{\partial \mathcal{L}(x(a), \lambda(a))}{\partial a_j} = \frac{\partial f(x(a), a)}{\partial a_j} + \lambda(a) \frac{\partial g(x(a), a)}{\partial a_j}. \tag{77}$$

Also note,

$$\begin{aligned}
\frac{\partial M(a)}{\partial a_j} &= \sum_{i=1}^{n} \left[\frac{\partial f(x(a), a)}{\partial x_i} \right] \frac{\partial x_i(a)}{\partial a_j} + \frac{\partial f(x(a), a)}{\partial a_j} \\
&= \lambda(a) \sum_{i=1}^{n} \left[-\frac{\partial g(x(a), a)}{\partial x_i} \frac{\partial x_i(a)}{\partial a_j} \right] + \frac{\partial f(x(a), a)}{\partial a_j} \\
&= \lambda(a) \frac{\partial g(x(a), a)}{\partial a_j} + \frac{\partial f(x(a), a)}{\partial a_j} \\
&= \frac{\partial \mathcal{L}(x(a), \lambda(a))}{\partial a_j}
\end{aligned}$$

where the first equality comes from the chain rule you argued in Exercise 54; the second from substituting (75); the third from substituting the derivative of the left hand side of (76) and applying the chain rule; and the fourth from substituting (77). □

Exercise 58. Verify that the envelope theorem is true for the following problem.

$$\max_{x_1, x_2} x_1 x_2 \text{ subject to } a - 2x_1 - 4x_2 = 0 \text{ and } x_i \geq 0 \text{ for } i = 1, 2$$

You can dispense with the nonnegativity constraints since the solution will satisfy the Karush-Kuhn-Tucker conditions.

7 Probability Theory

7.1 Probability Spaces

Let Ω be a set of **outcomes**, also referred to as a **sample space**, e.g. the outcome of a coin toss or the roll of two dice, etc. A set of subsets $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is called a $\sigma-$**algebra** if and only if (i) $\Omega \in \mathcal{F}$, (ii) if $A \in \mathcal{F}$ then $\Omega \backslash A \in \mathcal{F}$ and (iii) if $A_1, A_2, A_3, ... \in \mathcal{F}$ then $A_1 \cup A_2 \cup A_3 \cup ... \in \mathcal{F}$ for any countable sequence $\{A_1, A_2, A_3, ...\}$ of subsets of Ω. The elements of $\mathcal{F}$ are called **events**. Given a sample space Ω and a σ-algebra $\mathcal{F}$, a **probability measure** is a function $P : \mathcal{F} \to \mathbb{R}$ such that (i) $P(A) \geq 0$ for all $A \in \mathcal{F}$, (ii) $P(\Omega) = 1$, and (iii) for any countable sequence of events $A_1, A_2, A_3, ...$ that are **disjoint** (i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$) then $P(A_1 \cup A_2 \cup ...) = \sum_i P(A_i)$. The triple $(\Omega, \mathcal{F}, P)$ is called a **probability space**. Whenever we use the notation $P(\cdot)$ below it will represent the probability measure over a sample space Ω with $\sigma-$algebra $\mathcal{F}$. It will be useful to review the set theory that we did at the start before proceeding.

Exercise 59. Sometimes we will use the notation A^c to denote the complement $\Omega \backslash A$ of A in Ω. First prove **De Morgan's laws**: (i) $A \cap B = (A^c \cup B^c)^c$ and (ii) $A \cup B = (A^c \cap B^c)^c$. Then use these to prove the following four properties of the probability function: (i) $A \in \mathcal{F}$ implies $P(A^c) = 1 - P(A)$, (ii) $A, B \in \mathcal{F}$ with $A \subseteq B$ implies $P(A) \leq P(B)$, (iii) $A, B \in \mathcal{F}$ implies $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, and (iv) $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ with $A_1 \subseteq A_2 \subseteq A_3 \subseteq ...$ implies $P(\bigcup_{i=1}^{\infty} A_i) = \lim_{n \to \infty} P(A_n)$. A consequence of some of these properties, you will notice, is that $P(A) \in [0, 1]$ for all $A \in \mathcal{F}$.

Consider an event $B \in \mathcal{F}$ with $P(B) > 0$. Then the **conditional probability** given event B, which we denote by the function $P(\cdot|B) : \mathcal{F} \to \mathbb{R}$ is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \forall A \in \mathcal{F} \tag{78}$$

It is easy to verify that $P(\cdot|B)$ is also a probability measure, and so it satisfies all of the probability of the probability measure that you proved in the above exercise. In addition, note that

$$A, A' \subseteq B \text{ with } P(A') > 0 \text{ implies } \frac{P(A|B)}{P(A'|B)} = \frac{P(A)}{P(A')}$$

so that the ratio of conditional probabilities of two events equals the ratio of their unconditional probabilities when they are both sub-events of the conditioning event. The definition of a conditional probability gives rise to a new probability space, which we call the **conditional probability space**.

Next, we say that two events A and B are **independent** if $P(A \cap B) = P(A)P(B)$. This means that if $P(A) = 0$ or $P(B) = 0$ then A and B must be independent, and if $P(B) > 0$ then the two events are independent if $P(A|B) = P(A)$. As well, if A and B are

independent then A^c and B are independent, so are A and B^c and A^c and B^c. Similarly, two events A and A' are **conditionally independent**, conditional on a third event B, if $P(A \cap A'|B) = P(A|B)P(A'|B)$. The other properties also carry over.

Exercise 60. Prove that if A and B are two independent events then A^c and B^c are two independent events (where A^c and B^c denote the complements of A and B respectively).

Finally, we derive an important formula known as **Bayes rule**:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

This follows from noting that the numerator in (78) is simply $P(A \cap B)$ while the denominator equals $P(A \cap B) + P(A^c \cap B) = P\left((A \cap B) \cup (A^c \cap B)\right) = P(B)$. Since conditional probabilities require the conditioning event to have positive probability, the formula holds only when these necessary conditions hold. The formula is useful in computing $P(A|B)$ when $P(A)$, $P(B|A)$ and $P(B|A^c)$ are known but $P(A|B)$ is not. The following exercises provide you with examples of this.

Exercise 61. Suppose the probability that a person trying to enter the US is a terrorist is 10^{-5} and a new DHS program correctly classifies a terrorist as being one 99.8% of the time, and correctly classifies an innocent visitor as being innocent 99.99% of the time. Suppose a person is classified as a terrorist. What is the probability that s/he is one?

Exercise 62. Prove that

$$P(X \cap Y|Z) = \frac{P(Y \cap Z|X)P(X)}{P(Z)}.$$

Exercise 63. Consider a gameshow in which there are three opaque boxes, only one of which has a cash prize inside; the others are empty, and the prize is put randomly in one box. The gameshow host asks a contestant to choose a box. Whichever box he chooses, one of the other two will be empty so she will open up one box (among the two that were not chosen, and at random if both are empty) and reveal that it is empty. Then she will ask the contestant whether he wants to stay with the box he picked, or switch to the other closed box. After that, the contestant's decision is final, and wins the prize if it is inside the box he decided to go with. Should the contestant stay with the box he originally picked, or switch to the other closed box after an empty box was revealed to him? Explain your answer in detail using probability theory.

7.2 Random Variables

Given a probability space $(\Omega, \mathcal{F}, P)$, a (one-dimensional) **random variable** X is a function $X : \Omega \to \mathbb{R}$ for which $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$. Note that by this definition, it

is possible to generate new random variables from functions of random variables; e.g., for a function $h : X(\Omega) \to \mathbb{R}$, the composition $h \circ X$ is a random variable on the same probability space if $\{\omega : h(X(\omega)) \leq x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$, etc. Given the probability space and random variable defined on it, we say that a set $A \subseteq \mathbb{R}$ is **measurable** if $\{\omega : X(\omega) \in A\} \in \mathcal{F}$.

When the probability space and random variable that we are considering are clear, we will often abuse notation and write $P(\{\omega : X(\omega) \in A\})$ as $P(X \in A)$ or $P(A)$ or $P(\text{“some description of } A\text{”})$ for any measurable set A.

The cumulative distribution function, or **cdf**, of a random variable X is a function $F : \mathbb{R} \to \mathbb{R}$ defined by $F(x) = P(\{\omega : X(\omega) \leq x\})$. It is easy to verify that F has the following three properties

(i) it is nondecreasing,

(ii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, and

(iii) it is “**right continuous**;” i.e., for all x, we have $\lim_{\tilde{x} \to x^+} F(\tilde{x}) = F(x)$.

In fact, we will refer to any function F that has the above properties as a cumulative distribution function even when there is no underlying probability space and random variable associated with it. For example we might say that the “ideology is distributed according to F.” When we say this we mean that $F(x)$ gives the fraction of individuals whose ideology is at or to the left of the ideology x.

If for all but a countable number of values of $x \in \mathbb{R}$, there exists $\epsilon > 0$ such that F is constant on $[x - \epsilon, x + \epsilon]$ then X is said to be a **discrete** random variable. When this is the case, at each of the countable number of points at which this condition fails, F must have a **jump**; i.e., if x is one of these points, then

$$f(x) = F(x) - \lim_{\tilde{x} \to x^-} F(\tilde{x}) > 0$$

F is then said to be a **step function**. The size of the jump $f(x)$ is the “probability of x,” i.e., $f(x) = P(\{x\}) = P(\{\omega : X(\omega) = x\})$. The function $f(\cdot)$ mapping the jump points to $\mathbb{R}$ is called the probability mass function, or **pmf**.

If F is differentiable at every point $x \in \mathbb{R}$ then X is said to be a **continuous** random variable. The derivative of F, which we denote f will be called the probability distribution function, or **pdf**. Note then that $F(x) = \int_{-\infty}^{x} f(y)dy$ by the fundamental theorem of calculus. For $x_1 < x_2$, we have

$$P(\{x : x_1 < x \leq x_2\}) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx,$$

and note that $P(\{x\}) = \int_{x}^{x} f(y)dy = 0$ if X is continuous.

If X is not discrete and its cdf is differentiable at all but a countable number of points at which there are jumps, then X is said to be **mixed**. In this case, a jump point is also

often called a **mass point** or **atom**. While much (but not all) of what we say in the sequel will also apply to mixed random variables, we will implicitly assume from here on that a given random variable is either discrete or continuous. If $h : \mathbb{R} \to \mathbb{R}$ is a function and A is a measurable set, we sometimes use the notation

$$\int_A h(x)dF(x) := \begin{cases} \sum_{x\in A} h(x)f(x) & \text{if } X \text{ is discrete} \\ \int_A h(x)f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

This notation also has meaning in the case where F is mixed; to learn more about it, look up the "Riemann-Stieltjes integral." Note that $P(A) = \int_A dF(x)$ in both cases.

Finally, the **support** of a discrete/continuous random variable X is the set of points at which there are jumps/the set of points at which the derivative of F takes positive value. We will denote the support of X by $\mathsf{supp}\, X$ or $\mathsf{supp}\, F$ or $\mathsf{supp}\, f$.

Note: The support is usually defined replacing "set of points at which the derivative of F takes a positive value" by "the smallest closed set containing the set of points at which the derivative f of F takes a positive value." Even though this would change the notion of support (by including some extra points) this typically does not pose any problems.

Exercise 64. Suppose that $f(x)$ equals 0 for negative values of x and equals $Ke^{-\alpha x}(1 - e^{-\alpha x})$ for nonnegative values of x, for some $\alpha > 0$. (i) Find K such that $f(x)$ is the pdf of a continuous random variable. (ii) Find the corresponding cdf. (iii) Find the probability that the random variable takes value strictly larger than 1.

7.3 Transformations of Random Variables

Let X be a random variable on a probability space and consider an injective function $h : \mathbb{R} \to \mathbb{R}$. Redefining its range to be the image of $\mathbb{R}$, this function is bijective so it has inverse h^{-1}. In the discrete case, the distribution of the random variable $Y = h \circ X$ is

$$f_Y(y) = P(Y = y) = P(X = h^{-1}(y)) = f_X(h^{-1}(y))$$

where f_X is the pmf of X and f_Y the pmf of Y. From this we can generate the cdf of Y, which is

$$F_Y(y) = \sum_{\tilde{y}\le y} f_X(h^{-1}(\tilde{y})).$$

In the continuous case, suppose that h is increasing and has a differentiable inverse. Then for $y \in h(\mathbb{R})$, we have

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(h^{-1}(y))}{dy} = f_X(h^{-1}(y))\frac{dh^{-1}(y)}{dy}$$

and $f_Y(y) = 0$ for $y \notin h(\mathbb{R})$. If, on the other hand, h is decreasing (but also still has a differentiable inverse) then we would derive the same expression as above but with a

negative sign in front of it. This follows after noting that

$$F_Y(y) = P(Y \leq y) = P(X \geq h^{-1}(y)) = 1 - F_X(h^{-1}(y)),$$

and then taking the derivative of the left and right most sides with respect to y.

Exercise 65. Consider a continuous random variable X with pdf f_X. Let $h(x) = x^2$, and consider the random variable $Y = h \circ X$. Find an expression for the pdf of Y that depends only on y and f_X.

Exercise 66. If X is a continuous random variable with pdf $f(x) = 1/[\pi(1 + x^2)]$, what is the pdf of $1/X$?

7.4 Joint, Marginal and Conditional Distributions

Given a probability space, consider a pair of random variables (X, Y) each defined on this space. Let F_X denote the cdf of X and F_Y the cdf of Y; f_X and f_Y will denote their pmf/pdf respectively. The **joint cdf** of X and Y is the function $F_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ defined for $x = (x, y) \in \mathbb{R}^2$ by

$$F_{X,Y}(x, y) = P(\{\omega : X(\omega) \leq x \text{ and } Y(\omega) \leq y\})$$

If X and Y are discrete then

$$F_{X,Y}(x, y) = \sum_{\tilde{x} \leq x} \sum_{\tilde{y} \leq y} f(\tilde{x}, \tilde{y})$$

where

$$f_{X,Y}(\tilde{x}, \tilde{y}) = P(\{\omega : X(\omega) = \tilde{x} \text{ and } Y(\omega) = \tilde{y}\})$$

is the **joint pmf**, and if they are continuous then

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}$$

where

$$f_{X,Y}(\tilde{x}, \tilde{y}) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

is the **joint pdf**. These definitions generalize the the natural way to larger collections of random variables than just pairs; e.g., a collection of n random variables $(X_1, X_2, ..., X_n)$ defined on the same probability space. What we say below also generalizes.

Note that the cdf of X can be recovered from the joint cdf of (X, Y) as

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

and the joint pdf or pmf can be recovered as

$$f_X(x) = \sum_y f_{X,Y}(x,y) \qquad \text{or} \qquad f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

in the discrete and continuous cases, respectively. Note the summation is over all jump points. These derived functions are called the **marginal** cdf, pmf, or pdf of X given the joint cdf, pmf or pdf of X, Y. The concept of support defined above carries over to these distribution functions.

X and Y are said to be **independent** if their joint pmf in the discrete case, or joint pdf in the continuous case, can be written as the product of their pmfs or pdfs respectively; that is, the joint pmf or pdf is the product of the marginal pmfs or pdfs. This implies that the joint cdf is the product of the marginal cdfs as well. Note that X and Y can be independent only if the support of one does not depend on the other. Also, if X and Y are independent, then for function g and h, if $g(X)$ and $h(Y)$ are also random variables on the same probability space, this pair is also independent.

If X and Y are discrete then $f_{X,Y}$ represents their joint pmf while if they are continuous then it represents their joint pdf. With this in mind, we may be able to construct a new collection of random variables which we will denote $\{X \mid Y = y\}_y$ each member to be read as X "given" (or "conditional on") $Y = y$. If the marginal distribution $f_Y(y) > 0$, then the joint pmf or pdf of this random variable is given by

$$f_{X|Y=y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

which we say is the **conditional distribution** of X given $Y = y$. The conditional joint cdf of this collection can be generated from the joint pmf or pdf by summing or integrating. Please note the connection between these distributions and conditional probability: $X|Y = y$ is a random variable on the conditional probability space. For example, in the discrete case, observe that

$$f_{X|Y=y}(x|y) = P\left(\{\omega : X(\omega) = x\} \mid \{\omega : Y = y\}\right).$$

Note that if X and Y are independent, then

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \text{ so } f_{X|Y=y}(x|y) = f_X(x), \ \forall y$$

i.e., the conditional distribution is always equal to the marginal distribution. Also, define two random variables X and Y to be **conditionally independent** conditional on a third random variable Z if their joint conditional pmf or pdf $f_{X,Y|Z}$ is the product of their marginal conditional pdfs or pmfs $f_{X|Z}$ and $f_{Y|Z}$.

Exercise 67. Let $f_{X,Y}(x,y) = 15x^2y$ for $0 < x < y < 1$ and zero elsewhere be the joint pdf of two continuous random variables X and Y. (i) What are the marginal pdfs and cdf of

X and Y? (ii) Are X and Y independent? (iii) What is the conditional distribution (both pdf and cdf) of Y given X?

Exercise 68. Consider a sample of size 2 drawn without replacement from an urn containing three balls numbered 1, 2, 3. Let X be the number on the first ball drawn and Y the larger of the two numbers drawn. (i) Find the joint pdf of X and Y. (ii) Find the distribution of $X \mid Y = 3$.

Exercise 69. Let $F_{X,Y}$ be a joint cdf of the discrete random variables X and Y with marginal cdfs F_X and F_Y. Prove that $F_X(x) + F_Y(y) - 1 \leq F_{X,Y}(x,y) \leq \sqrt{F_X(x)F_Y(y)}$.

7.5 Expectations and Other Moments

The expectations operator, denoted $E[\cdot]$, is a mapping that works on random variables defined on a probability space. If X is a random variable, then the **expectation** of X, if it exists, is given by

$$E[X] = \int_{-\infty}^{\infty} x dF(x)$$

The expectation exists if $\int_{-\infty}^{\infty} |x| dF(x)$ converges to a finite number. Recall that integrals with bounds of ∞ or $-\infty$ on either end represent limits of sequences, which means that the sequences must converge. The expectation of a random variable is also called its **mean**.

In what follows we will work with expectations and typically drop the qualifier that some property about expectations is true only if the expectations exist. The following are four properties of the expectations operator. You should verify these.

(i) It inherits linearity from the linearity of summation and integration; i.e., for the random variable constructed by linearly combining two random variables X and Y with weights $a, b \in \mathbb{R}$, we have

$$E[aX + bY] = aE[X] + bE[Y].$$

(ii) The expectation of a function g of a random variable is the expectation of the random variable it generates; i.e., for $Y = g(X)$, we have

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x)$$

where F is the cdf of X.

(iii) The expectation of a random variable created as a function of several random variables is computed by integrating with respect to the joint distribution; e.g., for $Z = h(X,Y)$, we have

$$E[Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) dF_{X,Y}(x,y).$$

(iv) If X and Y are independent, then

$$E[XY] = E[X]E[Y].$$

The conditional expectation of X given $Y = y$ is simply the expectation of the random variable $X \mid Y = y$ defined above. Thus,

$$E[X|Y = y] = \int_{-\infty}^{\infty} x dF_{X|Y=y}(x|y),$$

which is a function that depends on y. We will abbreviate it by referring to it as $\mu_X(y)$; sometimes this is called a **regression** function. Note that it is a random variable; therefore, it has an expectation. Conveniently, its expectation is simply the (unconditional) expectation of X. For example, in the continuous case, we have

$$\begin{aligned} E[\mu_X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx f_Y(y)dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\ &= \int_{-\infty}^{\infty} x f_X(x)dx \\ &= E[X] \end{aligned}$$

The first equality is by definition, the second follows because f_Y is independent of x, the third by definition of the conditional distribution, the fourth by Fubini's theorem, the last two by the definitions of marginal distribution and expectation of X. Nevertheless,

$$E[E[X|Y]] = E[X]$$

holds in the discrete and mixed cases as well, so long as the expectations exist, and is known as the **law of iterated expectations**.

Exercise 70. Let X and Y be two random variables. Consider the random variable $Y - h(x)$ where the function h of x is called the **forecast** of Y, and $Y - h(x)$ the forecast error. Suppose we wanted to minimize the expectation of the square of the forecast error, $E[(Y - h(x))^2 \mid X = x]$. Show that the optimal forecast (i.e., the function h that solves this minimization problem) is $h(x) = E[Y \mid X = x]$.

Exercise 71. Let X be a random variable that is nonnegative with probability 1, with cdf F. Show that if the expectation of X exists, then $E[X] = \int_0^{\infty}(1 - F(x))dx$ if X is continuous. (You can show that this is also true if X is discrete.)

Exercise 72. Let X be a continuous random variable with cdf F. The **median** m of X is defined as the value of x such that $F(x) = 1/2$. Suppose that m is unique. Show that $E[|X - a|]$ is minimized by choosing $a = m$.

The expectation of a random variable X is sometimes called its first moment, and often denoted $\mu = E[X]$. In general, the quantity $E[X^k]$, if it exists, is called the kth **moment** of X. The k'th centered moment of X is $E[(X - \mu)^k]$, where μ is the first moment. The second centered moment has a special name: it is called the **variance** of X and is often denoted $Var[X]$ or σ^2. Its positive square root, σ, is called the **standard deviation**. Observe that

$$E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - \mu^2.$$

For a vector of random variables $X = (X_1, ..., X_n)$ each defined on the same probability space, the mean is defined as the vector $E[X] = (E[X_1], ..., E[X_n])$. The $n \times n$ matrix $\Sigma = E[(X - E[X])(X - E[X])]$ is called the **covariance matrix** of X. The (i, j)th element of Σ is $\sigma_{ij} = E[(X_i - E[X_i])(X_j - E[X_j])]$ is the covariance between X_i and X_j. We also write $Cov[X_i, X_j] = \sigma_{ij}$. Since $\sigma_{ij} = \sigma_{ji}$, Σ is a symmetric matrix.

Let α and β be vectors of length n. Then $E[\alpha' X] = \alpha' E[X]$, where α' denotes the transpose of α. As well, $Var[\alpha' X] = \alpha' \Sigma \alpha \geq 0$ so that Σ is positive semi-definite. The **covariance** between $\alpha' X$ and $\beta' X$ is $E[(\alpha' X - \alpha' E[X])(\beta' X - \beta' E[X]) = \alpha' \Sigma \beta$. For an $n \times k$ matrix of scalars A, the expectation of AX is $E[AX] = A' E[X]$ and the covariance matrix is $A' \Sigma A$. Finally, the **correlation** between X_i and X_j is $\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii} \sigma_{jj}}$.

Exercise 73. Prove that $-1 \leq \rho_{ij} \leq 1$. *Hint:* Σ is positive semi-definite. Then show that if X_i and X_j are independent then $\rho_{ij} = 0$ (but note that the converse is not true).

Exercise 74. Let X and Y be random variables. Assuming that all necessary moments exist, prove that the values of a and b that minimize $E[(Y - a - bX)^2]$ are

$$a = E[Y] - \frac{Cov[X, Y]}{Var[X]} E[X] \qquad \text{and} \qquad b = \frac{Cov[X, Y]}{Var[X]}.$$

Exercise 75. Let X and Y be two (jointly) continuous random variables. Define the "conditional variance of Y given $X = x$" as

$$\sigma^2_{Y|X=x}(x) = Var[Y|X = x] = E[(Y - E[Y|X = x])^2 | X = x]$$

and let $Var[Y|X] = \sigma^2_{Y|X}(X)$. Show that $Var[Y] = E[Var[Y|X]] + Var[E[Y|X]]$.

Theorem 36. *Suppose that X is a random variable defined on some probability space.*

1. *(Jensen) If $h : \mathbb{R} \to \mathbb{R}$ is a convex function then $E[h(X)] \geq h(E[X])$.*
2. *(Markov) If $P(X \geq 0) = 1$, then $P(X \geq K) \leq \frac{E[X]}{K}$ for all $K > 0$.*

3. (Chebychev) If $Var[X] < \infty$ *then* $P\big(|X - E[X]| \geq K\big) \leq \frac{Var[X]}{K^2}$ *for all* $K > 0$.

Proof. (i) If h is convex then by Exercise 29 there is a line $y = h(E[X]) + m(x - E[X])$ such that

$$h(x) \geq m(x - E[X]) + h(E[X]), \qquad \forall x$$

Then taking expectations, we have

$$E[h(X)] \geq E[m(x - E[X])] + E[h(E[X])] = h(E[X]).$$

(ii) I prove this in the continuous case, leaving the discrete case to you. Note that

$$E[X] = \int_0^\infty x f(x)dx \geq \int_K^\infty x f(x)dx \geq K \int_K^\infty f(x)dx = KP(X \geq K).$$

(iii) Let $Y = (X - E[X])^2$ so $P(Y \geq 0) = 1$ and $E[Y] = Var[X]$. Then we have

$$P(|X - E[X]| \geq K) = P(Y \geq K^2) \leq \frac{E[Y]}{K^2} = \frac{Var[X]}{K^2}.$$

where the inequality follows from Markov's inequality. □

7.6 The Moment Generating Function & Select Distributions

The moment generating function, or **mgf**, for a random variable X with cdf $F(x)$ is

$$M(t) = E[e^{tX}]$$

Since $M(t) = \int_{-\infty}^{\infty} e^{tx} dF(x)$, we have

$$M^{(k)}(t) = \int_{-\infty}^{\infty} x^k e^{tx} dF(x), \qquad \text{so } M^{(k)}(0) = E[X^k]$$

where $M^{(k)}$ denotes the kth derivative of $M(t)$. The moment generating function may not exist for all random variables, but if it does exist on some interval $(-\xi, \xi)$ centered at 0, then it uniquely characterizes the cdf of X. More formally,

Theorem 37. *Suppose* X *and* Y *are random variables with cdfs* F_X *and* F_Y *and mgfs* $M_X(t)$ *and* $M_Y(t)$ *that exist for* t *sufficiently close to* 0. *For all* $\epsilon > 0$ *suppose there exists* $\xi > 0$ *such that*

$$|M_X(t) - M_Y(t)| < \epsilon \qquad \forall t \in (-\xi, \xi).$$

Then there is a function $\delta(\epsilon)$ *such that at all continuity points* $a \in \mathbb{R}$ *of* F_X *and* F_Y,

$$|F_X(a) - F_Y(a)| < \delta(\epsilon) \qquad \text{and} \qquad \lim_{\epsilon \to 0} \delta(\epsilon) = 0.$$

Thus, when the mgfs of two random variables are close, their cdfs are close as well. The proof of this theorem is outside the scope of this class. We will use this result, though, to prove the Central Limit Theorem below.

Exercise 76. Suppose that the random variables $X_1, ..., X_K$ are all (mutually) independent with mgfs $M_1(t), ..., M_K(t)$ that exist for t sufficiently close to 0. Show that the mgf of the random variable created by taking the sum of these variables $Y = \sum_k X_k$ is the product of the mgfs, denoted $\prod_k M_k(t)$.

Exercise 77. Let X be a discrete random variable with pmf $f(x) = (1/2)^x$, $x = 1, 2, 3, ...$ Find $E[X]$ and the mgf $M(t)$ of X for $|t| < \log 2$.

Exercise 78. Suppose the mgf of X and Y are $M_X(t) = M_Y(t) = e^{t^2+3t}$ and X and Y are independent. What is the mgf of $Z = 2X - 3Y + 4$?

I now derive the mgf (and some properties) of several well-known discrete and continuous random variables/distributions.

Bernoulli distribution The pmf of X is $f(1) = p$, $f(0) = 1 - p$ and $f(x) = 0$ for all $x \notin \{0, 1\}$. p is said to be the **parameter** of the Bernoulli distribution. The mgf is

$$M(t) = E[e^{tX}] = pe^{1t} + (1-p)e^{0t} = 1 - p + pe^t.$$

so the mean of the distribution is $E[X] = p$ and the variance is $Var[X] = p(1-p)$.

Binomial distribution Suppose that $X_1, ..., X_n$ are independent and identically distributed (**iid**) Bernoulli random variables each with parameter p. Then $Y = \sum_i X_i$ has the Binomial distribution with support $\mathsf{supp}\, Y = \{0, 1, ..., n\}$ and for $y \in \mathsf{supp}\, Y$, the pmf of Y is

$$f_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

By the result in Exercise 76, the mgf of Y is

$$M(t) = (1 - p + pe^t)^n.$$

The mean of the distribution is therefore np and the variance is $np(1-p)$.

Poisson distribution X takes values on $\mathsf{supp}\, X = \{0, 1, 2, ...,\}$ and for $x \in \mathsf{supp}\, X$, the pmf is

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The mgf is

$$M(t) = \sum_{n=0}^{\infty} e^{tn} \frac{\lambda^n e^{-\lambda}}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(e^t\lambda)^n}{n!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}.$$

The mean and the variance are therefore both equal λ, which is the parameter of this distribution. Note also that the Poisson distribution is the limit as $n \to \infty$ of a sequence of

Binomial distributions with parameters $p_n = \lambda/n$; to see this, note that the limit as $n \to \infty$ of the sequence of corresponding mgfs is

$$\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty} \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^t\right)^n = e^{\lambda(e^t-1)}$$

where the second inequality follows from Exercise 38. You can therefore think of the Poisson distribution as modeling the probability λ of one success over a period of unit time. The number of successes in non-overlapping periods are independent.

Exercise 79. Let X and Y be independent random variables both distributed Poisson with $E[X] = \lambda_X$ and $E[Y] = \lambda_Y$. Find the distribution of $X + Y$.

Uniform distribution X takes values on an interval $\mathsf{supp}\, X = [a, b] \subset \mathbb{R}$ with every value in this interval equally "likely." That is, the pmf is $f(x) = 1/(b-a)$ on $[a, b]$ and 0 everywhere else. The mgf is therefore

$$M(t) = \int_a^b e^{tx}\frac{1}{b-a}dx = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

if $t \neq 0$ and 1 if $t = 0$. The mean of the distribution is the midpoint of the interval $E[X] = \frac{1}{2}(a+b)$ and the variance is $Var[X] = \frac{1}{12}(b-a)^2$.

Exercise 80. Consider random variable X with cdf $F(x)$. Let $h(x) = F(x)$ and let the inverse $h^{-1}(y)$ be defined as the smallest x such that $F(x) = y$ (which exists because F is right continuous). Show that $Y = h(X)$ has uniform distribution with support $[0, 1]$.

Exercise 81. If X has a uniform distribution on support $[0, 1]$, find the distribution of $1/X$. Does $E[1/X]$ exist? If so, find it.

Normal distribution X takes on values on $\mathsf{supp}\, X = \mathbb{R}$ and for parameters (μ, σ) where $\sigma > 0$ the pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Note that the moment generating function is

$$\begin{aligned} M(t) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2 + xt} dx \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-(\mu+\sigma^2 t))^2} dx \\ &= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \end{aligned}$$

since the second integral is the integral of the pdf of the normal distribution with parameters $((\mu + t\sigma^2), \sigma^2)$ and so must integrate to 1. The distribution is often denoted $\mathcal{N}(\mu, \sigma^2)$ since $E[X] = \mu$ is the mean of the distribution and $Var[X] = \sigma^2$ the variance. The **standard normal** is the distribution $\mathcal{N}(0, 1)$ with zero mean and unit variance.

Exercise 82. Suppose X is a random variable with standard normal distribution. Show that the random variable $Y = cX$ where c is a constant has distribution $\mathcal{N}(0, c^2)$.

Exercise 83. Suppose that $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables each having Poisson distribution. Suppose the parameter of the distribution of X_n is n for all n. Consider the corresponding sequence of **standardized** Poisson random variables $\{(X_n - n)/\sqrt{n}\}_{n=1}^{\infty}$. Show that the corresponding sequence of distributions of the standardized sequence converges to the standard normal distribution.

Exercise 84. Suppose that X is distributed $\mathcal{N}(\mu, \sigma^2)$. Show that $E[|X - \mu|] = \sigma\sqrt{2/\pi}$.

Chi-squared distribution Let $X_1, ..., X_n$ be a sequence of iid random variables each with a standard normal distribution. Then $Y = \sum_i (X_i)^2$ is said to have a chi-squared distribution with n degrees of freedom. The distribution is denoted χ_n^2.

Exercise 85. Show that the mean of the chi-squared distribution with k degrees of freedom is $E[Y] = k$ and the variance is $Var[Y] = 2k$.

$F-$distribution Let Y_1 and Y_2 be two independent random variables each where Y_1 has distribution χ_k^2 and Y_2 has distribution χ_l^2. Then $Q = (Y_1/k)/(Y_2/l)$ is said to be have an F_{kl} distribution with k degrees in the numerator and l in the denominator.

$t-$distribution Let Z be a random variable with standard normal distribution and Y be a random variable with a χ_k^2 distribution. If Y and Z are independent, then $T = Z/\sqrt{Y/k}$ is said to have the **student's** t_k-distribution, with k degrees of freedom.

The Multivariate Normal Distribution Consider the random variables X_1 and X_2, construct the pair (X_1, X_2) and treat it as a vector. $X = (X_1, X_2)$ has the bivariate normal distribution if and only if for all vectors $\alpha \in \mathbb{R}^2$, the random variable defined $\alpha' X$ is normally distributed. If X_1 and X_1 are jointly bivariate normal, then they are each normally distributed. The mean and covariance matrix of (X_1, X_2) both exist and we denote them by μ (a vector of size 2) and Σ (a square matrix of order 4). Let α be a vector of size two. Then $Y = \alpha' X$ is normal with mean and variance $E[\alpha' X] = \mu'\alpha$ and $Var[\alpha' X] = \alpha'\Sigma\alpha$. Therefore,

$$M_X(\alpha) = E[e^{\alpha' X}] = E[e^Y] = M_Y(1) = e^{\alpha'\mu + \frac{1}{2}\alpha'\Sigma\alpha}$$

where M_X and M_Y are the mgfs of X and Y, respectively. By the uniqueness of the mgf, this implies that the distribution of X is completely characterized by μ and Σ, and we write $X = (X_1, X_2) \sim \mathcal{N}_2(\mu, \Sigma)$. These things generalize to the case where $X = (X_1, ..., X_n)$, in which case X has the multivariate normal distribution.

7.7 Convergence Concepts & Results

Consider a probability space $(\Omega, \mathcal{F}, P)$, a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ each defined on the space, and another random variable X also defined on the same space. Then we have the following notions of convergence.

1. Let
$$A = \left\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}$$
be the outcomes on which $\{X_n\}$ converges to X. (Perhaps A is empty.) Then the sequence $\{X_n\}$ is said to **converge almost surely** to X (denoted $X_n \xrightarrow{a.s.} X$) if
$$P(A) = 1.$$

2. Alternatively, for any $\varepsilon > 0$ let
$$p_n(\varepsilon) = P(|X_n - X| > \varepsilon)$$
be the probability that X_n differs by more than ϵ from X. Then if for all ε, the sequence $\{p_n(\varepsilon)\}$ converges to 0, i.e. if
$$\lim_{n\to\infty} p_n(\varepsilon) = 0, \qquad \forall \varepsilon > 0,$$
then $\{X_n\}$ is said to **converge in probability** to X (denoted $X_n \xrightarrow{p} X$ or sometimes as plim $X_n = X$).

3. Suppose that
$$\lim_{n\to\infty} E\left[|X_n - X|^r\right] = 0$$
for $r = 1$. Then $\{X_n\}$ is said to **converge in mean** to X (denoted $X_n \xrightarrow{m} X$). If the same limit holds for $r = 2$ then $\{X_n\}$ is said to **converge in mean square** (denoted $X_n \xrightarrow{m.s.} X$).

4. Let $\{F_n\}$ denote the corresponding sequence of cdfs for the sequence $\{X_n\}$ of random variables and let F be the cdf of X. If for all points at which F is continuous,
$$\lim_{n\to\infty} F_n(x) = F$$
then $\{X_n\}$ is said to **converge in distribution** to X (denoted $X_n \xrightarrow{d} X$). Sometimes we also say that $\{X_n\}$ **converges weakly** to X.

Exercise 86. (i) Show that convergence in mean square implies convergence in mean. *Hint:* Use Jensen's inequality. (ii) Show that convergence in mean implies convergence in probability *Hint:* Use Markov's inequality. *Extra credit:* (iii) Show that almost sure convergence implies convergence in probability. (iv) Finally, show that convergence in probability implies convergence in distribution.

Theorem 38. (**weak law of large numbers**) *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of random variables with means denoted μ_i and variances denoted σ_i^2. Suppose that they are uncorrelated meaning the covariances are $\sigma_{ij} = 0$ for all $i \neq j$. Let*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i, \qquad \bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2$$

and suppose that

$$\lim_{n\to\infty} \bar{\sigma}_n^2/n = 0.$$

Then we have

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

Proof. Note that by Chebychev's inequality we have for all $\varepsilon > 0$,

$$P(|\bar{X}_n - \bar{\mu}_n| > \varepsilon) \leq \frac{E\left[(\bar{X}_n - \bar{\mu}_n)^2\right]}{\varepsilon^2} = \frac{\frac{1}{n^2} E\left[(\sum_{i=1}^{n}(X_i - \mu_i))^2\right]}{\varepsilon^2} = \frac{\frac{1}{n}\bar{\sigma}_n^2}{\varepsilon^2}$$

Therefore $P(|\bar{X}_n - \bar{\mu}_n| > \varepsilon)$ converges to 0 for all ε since the right side of this expression converges to 0 in the limit as $n \to \infty$. □

Theorem 39. (**central limit theorem**) *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid random variables each with mean μ and variance σ^2 and mgf $M(t)$ that exists for all t in some interval $(-\xi, \xi)$. Suppose that $M''(t)$ is continuous at 0 and let*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then,

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. Let $Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$ and $Y_i = (X_i - \mu)/\sigma$ so that $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i$. Let

$$\psi(t) := E\left[e^{Y_i t}\right] = E\left[e^{(X_i/\sigma - \mu/\sigma)t}\right] = e^{-\mu t/\sigma} M(t/\sigma)$$

be the mgf of Y_i which exists on the interval $(-\xi\sigma, \xi\sigma)$. Note that $\psi(0) = 1$, $\psi'(0) = 0$ and $\psi''(0) = 1$. By Taylor's theorem, taking $a = 0$, we have

$$\begin{aligned}\psi(t) &= \psi(0) + \psi'(0)t + \frac{1}{2}\psi''(0)t^2 + \frac{1}{2}(\psi''(\tau(t)) - \psi''(0))t^2 \\ &= 1 + \frac{1}{2}t^2 + \frac{1}{2}\big(\psi''(\tau(t)) - 1\big)t^2\end{aligned}$$

where $\tau(t)$ is a number between 0 and t such that $\lim_{t\to 0}\tau(t) = 0$. Since ϕ'' is continuous at zero (because it is differentiable), we also have $\lim_{t\to 0}\psi''(\tau(t)) = \psi''(0) = 1$. Then, the

mgf of Z_n on the interval $(-\xi\sigma, \xi, \sigma)$ is

$$\begin{aligned} M_{Z_n}(t) &= (\phi(t/\sqrt{n}))^n \\ &= \left(1 + \frac{1}{2}(t/\sqrt{n})^2 + \frac{1}{2}\big(\psi''(\tau(t/\sqrt{n})) - 1\big)(t/\sqrt{n})^2\right)^n \\ &= \left(1 + \frac{\frac{1}{2}t^2 + \frac{1}{2}(\psi''(\tau(t/\sqrt{n})) - 1)t^2}{n}\right)^n \end{aligned} \tag{79}$$

Now as n goes to $+\infty$ the right hand side of this converges to $e^{t^2/2}$, which is the mgf of the standard normal distribution. Then by invoking Theorem 37, the proof is complete. □

Exercise 87. The assertion that (79) converges to $e^{t^2/2}$ in n follows from the fact that if $\{z_n\}$ is a real sequence converging to z then $\lim_{n\to\infty}(1 + z_n/n)^n = e^z$. Look back at how you proved exercise 38 and convince yourself that this assertion also holds.

Theorem 40. (**the Delta Method**) *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables such that $\sqrt{n}(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ and let $g : \mathbb{R} \to \mathbb{R}$ be continuously differentiable at μ. Then,*

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, (g'(\mu))^2\sigma^2).$$

Proof. By the mean value theorem

$$g(X_n) - g(\mu) = (X_n - \mu)g'(\tilde{X}_n)$$

for some $\tilde{X}_n(\omega)$ between μ and $X_n(\omega)$. Since X_n converges in probability to μ, $\tilde{X}_n$ converges in probability to μ as well. Then, there is a theorem called the "continuous mapping theorem" (look it up) that says that $g'(\tilde{X}_n)$ converges in probability to $g'(\mu)$ since g' is continuous. Thus, $\sqrt{n}(g(X_n) - g(\mu))$ converges in distribution to $\mathcal{N}(0, (g'(\mu))^2\sigma^2)$. □

The Delta method generalizes to vectors of random variables. Let $\{X_n\}$ be a sequence of vectors of random variables each of size k such that $\sqrt{n}(X_n - \mu)$ converges in distribution to $\mathcal{N}_k(0, \Sigma)$, the k-variate normal distribution, and let $g : \mathbb{R}^k \to \mathbb{R}^l$ be continuously differentiable at the vector μ of size k. Let A denote the $l \times k$ Jacobian matrix of first derivatives of g at μ. Then $\sqrt{n}(g(X_n) - g(\mu))$ converges in distribution to $\mathcal{N}_l(0, A\Sigma A')$.

A Appendix

A.1 Proof of the Heine-Borel Theorem

We begin the proof by first establishing two lemmata.

Lemma A.1. *Let $\{x_i\}$ be an increasing sequence contained in a bounded set $X \subset \mathbb{R}$. Then the sequence converges to a limit that equals* $\sup\{x_i\}$.

Proof. For $\epsilon > 0$ we know that $\sup\{x_i\} - \epsilon$ is not an upper bound for the sequence, so there exists $x_N \in \{x_i\}$ such that for all $n > N$ we have $\sup\{x_i\} - \epsilon < x_n \leq \sup\{x_i\}$ (which follows because the sequence is increasing). This rearranges to $0 < \sup\{x_i\} - x_n < \epsilon$. Since ϵ was arbitrary, this shows that $\sup\{x_i\}$ is a limit of the sequence $\{x_i\}$. □

Lemma A.2. *Let $Z := [-z, z]$ for some $z > 0$. Then for all n, the set Z^n is compact.*

Proof. We show that Z is compact; the fact that Z^n is compact follows immediately from this and the definition of convergence of vectors.

Consider an infinite sequence $\{x_i\} \subset Z$. For each j, let $z_j = \inf\{x_j, x_{j+1}, ...\}$. Then $\{z_j\}$ is an increasing sequence that is bounded by $Z \subset \mathbb{R}$. (This follows from Exercise 27.) According to the previous lemma, it therefore converges to a limit, $z := \sup\{z_j\}$. We now show that z is the limit of a subsequence of $\{x_i\}$. Given $\epsilon > 0$, and any number N, there is $n > N$ such that $|z_n - z| < \epsilon/2$ since z is the limit of sequence $\{z_n\}$. Since $z_n = \inf\{x_n, x_{n+1}, ...\}$ there is $m \geq n$ such that $|x_m - z_n| < \epsilon/2$. Combining these using the triangle inequality,

$$|x_m - z| < |x_m - z_n| + |z_n - z| < \epsilon.$$

Thus we can construct a subsequence of $\{x_i\}$ that converges to z. Note that $z \in Z$ since ϵ was arbitrary; otherwise, some elements of the subsequence would lie outside Z. □

Now we can prove the Heine-Borel theorem.

First let us show that a compact set, X, is closed and bounded. To show that it is closed, take any convergent sequence $\{x_k\} \subset X$. Since X is compact, this sequence has a convergent subsequence $\left\{x_{m(k)}\right\}$ whose limit is in X. By the uniqueness of the limit, this is also the limit of $\{x_k\}$. Hence X is closed.

If X is not bounded, then for each n, there is $x_n \in X$ such that $||x_n|| > n$. Then the sequence $\{x_n\} \subset X$ does not have a convergent subsequence. To see this, suppose that it did and let x be the limit of such a subsequence. For all $m > 2||x||$ we have

$$||x_m - x|| \geq ||x_m|| - ||x|| \geq m - ||x|| > m$$

where the first inequality follows from the triangle inequality for vectors (see Exercise 7). This however, violates the condition for convergence (that for all ϵ there is a number N

such that $||x_n - x|| < \epsilon$ for all $n \geq N$; simply take ϵ to be smaller than m). This proves that X must be bounded.

We must now show the reverse: that if X is closed and bounded set then it must be compact. By boundedness, there is a number $z > 0$ such that $|x_i| \leq z$ for all $x \in X$ and all i, where x_i is the ith component of the vector x. Lemma A.2 above shows that $Z \equiv [-z, z] \times \cdots \times [-z, z]$ is compact. Obviously, $X \subset Z$. If we can show that a closed subset of a compact set is also compact, then we are done. To do this last step, take any sequence in X. Since $X \subset Z$, this is also a sequence in Z, which is a compact set. So it must have a convergent subsequence with limit in Z. But since X is closed, and this subsequence lies in X, the limit must also lie in X. Therefore, X is compact.

A.2 Finishing the Proof of the Implicit Function Theorem

We prove the claim that $\sqrt{s^2 + t^2}/s$ in the proof of the theorem is bounded for small values of s.

Note that for s small enough we have

$$|\varepsilon| < \min \left\{ \frac{1}{2} \left| \frac{\partial F(x,y)}{\partial y} \right|, \left| \frac{\partial F(x,y)}{\partial x_j} \right| \right\} \tag{80}$$

since ε goes to 0 and s goes to 0. Then, we have

$$\begin{aligned} \left| \frac{\partial F(x,y)}{\partial y} \right| |t| &\leq |s| \left| \frac{\partial F(x,y)}{\partial x_j} \right| + |\varepsilon| \sqrt{s^2 + t^2} \\ &\leq |s| \left| \frac{\partial F(x,y)}{\partial x_j} \right| + |s||\varepsilon| + |t||\varepsilon| \\ &\leq |s| \left| \frac{\partial F(x,y)}{\partial x_j} \right| + |s| \left| \frac{\partial F(x,y)}{\partial x_j} \right| + |t| \frac{1}{2} \left| \frac{\partial F(x,y)}{\partial y} \right| \end{aligned}$$

where the first inequality follows from (37), the second from the triangle inequality, and the third from (80). Solving the final inequality for $|t|$ gives

$$|t| \leq 4|s| \left| \frac{\partial F(x,y)/\partial y}{\partial F(x,y)/\partial x_j} \right|$$

Therefore, we have

$$\left| \frac{\sqrt{s^2 + t^2}}{s} \right| \leq \frac{|s| + |t|}{|s|} = 1 + 4 \left| \frac{\partial F(x,y)/\partial y}{\partial F(x,y)/\partial x_j} \right|$$

which shows that $\sqrt{s^2 + t^2}/s$ is indeed bounded for small values of s.