Math Prefresher for Political Scientists

Shiro Kuriwaki and Yon Soo Park

2018-06-19

Contents

Chapter 1

About this Booklet

The documents in this booklet are the product of generations of Math (P)refresher Instructors: Curt Signorino 1996-1997; Ken Scheve 1997-1998; Eric Dickson 1998-2000; Orit Kedar 1999; James Fowler 2000-2001; Kosuke Imai 2001-2002; Jacob Kline 2002; Dan Epstein 2003; Ben Ansell 2003-2004; Ryan Moore 2004-2005; Mike Kellermann 2005-2006; Ellie Powell 2006-2007; Jen Katkin 2007-2008; Patrick Lam 2008-2009; Viridiana Rios 2009-2010; Jennifer Pan 2010-2011; Konstantin Kashin 2011-2012; Sol'{e} Prillaman 2013; Stephen Pettigrew 2013-2014; Anton Strezhnev 2014-2015; Mayya Komisarchik 2015-2016; Connor Jerzak 2016-2017; Shiro Kuriwaki 2017-2018; Yon Soo Park 2018-

Chapter 2

Linear Algebra

Topics: ● Working with Vectors ● Linear Independence ● Basics of Matrix Algebra ● Square Matrices ● Linear Equations ● Systems of Linear Equations ● Systems of Equations as Matrices ● Solving Augmented Matrices and Systems of Equations ● Rank ● The Inverse of a Matrix ● Inverse of Larger Matrices

Much of the material and examples for this lecture are taken from Gill (2006) *Essential Mathematics for Political and Social Scientists*, Simon & Blume (1994) *Mathematics for Economists* and Kolman (1993) *Introductory Linear Algebra with Applications*.

2.1 Working with Vectors

Vector: A vector in n-space is an ordered list of n numbers. These numbers can be represented as either a row vector or a column vector:

$$\mathbf{v} \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

We can also think of a vector as defining a point in n-dimensional space, usually $\mathbf{R}^n$; each element of the vector defines the coordinate of the point in a particular direction.

Vector Addition and Subtraction: If two vectors, $\mathbf{u}$ and $\mathbf{v}$, have the same length (i.e. have the same number of elements), they can be added (subtracted) together:

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 + v_1 & u_2 + v_2 & \cdots & u_k + v_n \end{pmatrix}$$

$$\mathbf{u} - \mathbf{v} = \begin{pmatrix} u_1 - v_1 & u_2 - v_2 & \cdots & u_k - v_n \end{pmatrix}$$

Scalar Multiplication: The product of a scalar c (i.e. a constant) and vector $\mathbf{v}$ is:

$$c\mathbf{v} = \begin{pmatrix} cv_1 & cv_2 & \ldots & cv_n \end{pmatrix}$$

Vector Inner Product: The inner product (also called the dot product or scalar product) of two vectors $\mathbf{u}$ and $\mathbf{v}$ is again defined iff they have the same number of elements

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

If $\mathbf{u} \cdot \mathbf{v} = 0$, the two vectors are orthogonal (or perpendicular).

Vector Norm: The norm of a vector is a measure of its length. There are many different ways to calculate the norm, but the most common of is the Euclidean norm (which corresponds to our usual conception of distance in three-dimensional space):

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1 v_1 + v_2 v_2 + \cdots + v_n v_n}$$

2.2   Linear Independence

Linear combinations: The vector $\mathbf{u}$ is a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ if

$$\mathbf{u} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_k \mathbf{v}_k$$

Linear independence: A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ is linearly independent if the only solution to the equation

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_k \mathbf{v}_k = 0$$

is $c_1 = c_2 = \cdots = c_k = 0$. If another solution exists, the set of vectors is linearly dependent.

A set S of vectors is linearly dependent iff at least one of the vectors in S can be written as a linear combination of the other vectors in S.

Linear independence is only defined for sets of vectors with the same number of elements; any linearly independent set of vectors in n-space contains at most n vectors.

Exercises: Are the following sets of vectors linearly independent?

1.

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

2.

$$\mathbf{v}_1 = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -2 \\ 2 \\ 4 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

2.3   Basics of Matrix Algebra

Matrix: A matrix is an array of real numbers arranged in m rows by n columns. The dimensionality of the matrix is defined as the number of rows by the number of columns, mxn.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Note that you can think of vectors as special cases of matrices; a column vector of length k is a $k \times 1$ matrix, while a row vector of the same length is a $1 \times k$ matrix.

It's also useful to think of matrices as being made up of a collection of row or column vectors. For example,

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_m \end{pmatrix}$$

Matrix Addition: Let $\mathbf{A}$ and $\mathbf{B}$ be two $m \times n$ matrices.

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

Note that matrices $\mathbf{A}$ and $\mathbf{B}$ must have the same dimensionality, in which case they are **conformable for addition**.

Example:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$$

$$\mathbf{A} + \mathbf{B} =$$

Scalar Multiplication: Given the scalar s, the scalar multiplication of $s\mathbf{A}$ is

$$s\mathbf{A} = s \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} sa_{11} & sa_{12} & \cdots & sa_{1n} \\ sa_{21} & sa_{22} & \cdots & sa_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ sa_{m1} & sa_{m2} & \cdots & sa_{mn} \end{pmatrix}$$

Example:

$$s = 2, \qquad \mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

$$s\mathbf{A} =$$

Matrix Multiplication: If $\mathbf{A}$ is an $m \times k$ matrix and $\mathbf{B}$ is a $k \times n$ matrix, then their product $\mathbf{C} = \mathbf{AB}$ is the $m \times n$ matrix where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj}$$

Examples:

1. $\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} =$

2. $\begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} -2 & 5 \\ 4 & -3 \\ 2 & 1 \end{pmatrix} =$

Note that the number of columns of the first matrix must equal the number of rows of the second matrix, in which case they are **conformable for multiplication**. The sizes of the matrices (including the resulting product) must be

$$(m \times k)(k \times n) = (m \times n)$$

Also note that if $\mathbf{AB}$ exists, $\mathbf{BA}$ exists only if $\dim(\mathbf{A}) = m \times n$ and $\dim(\mathbf{B}) = n \times m$.

This does not mean that $\mathbf{AB} = \mathbf{BA}$. $\mathbf{AB} = \mathbf{BA}$ is true only in special circumstances, like when $\mathbf{A}$ or $\mathbf{B}$ is an identity matrix, $\mathbf{A} = \mathbf{B}^{-1}$, or $\mathbf{A} = \mathbf{B}$ and $\mathbf{A}$ is idempotent.

Laws of Matrix Algebra:

1. Associative: $\quad\quad\quad\quad (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
   $\quad\quad\quad\quad\quad\quad\quad\quad (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$

2. Commutative: $\qquad\qquad\qquad$ $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

3. Distributive: $\qquad\qquad\qquad$ $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
$\qquad\qquad\qquad\qquad\qquad$ $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

Commutative law for multiplication does not hold – the order of multiplication matters:

$$\mathbf{AB} \neq \mathbf{BA}$$

Example:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{AB} = \begin{pmatrix} 2 & 3 \\ -2 & 2 \end{pmatrix}, \qquad \mathbf{BA} = \begin{pmatrix} 1 & 7 \\ -1 & 3 \end{pmatrix}$$

Transpose: The transpose of the $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^T$ (also written $\mathbf{A}'$) obtained by interchanging the rows and columns of $\mathbf{A}$.

Examples:

1. $\mathbf{A} = \begin{pmatrix} 4 & -2 & 3 \\ 0 & 5 & -1 \end{pmatrix}, \qquad \mathbf{A}^T = \begin{pmatrix} 4 & 0 \\ -2 & 5 \\ 3 & -1 \end{pmatrix}$

2. $\mathbf{B} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \qquad \mathbf{B}^T = \begin{pmatrix} 2 & -1 & 3 \end{pmatrix}$

The following rules apply for transposed matrices:

1. $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

2. $(\mathbf{A}^T)^T = \mathbf{A}$

3. $(s\mathbf{A})^T = s\mathbf{A}^T$

4. $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$; and by induction $(\mathbf{ABC})^T = \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T$

Example of $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$:

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 \\ 2 & -1 & 3 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 2 & 2 \\ 3 & -1 \end{pmatrix}$$

$$(\mathbf{AB})^T = \left[\begin{pmatrix} 1 & 3 & 2 \\ 2 & -1 & 3 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 2 & 2 \\ 3 & -1 \end{pmatrix} \right]^T = \begin{pmatrix} 12 & 7 \\ 5 & -3 \end{pmatrix}$$

$$\mathbf{B}^T\mathbf{A}^T = \begin{pmatrix} 0 & 2 & 3 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & -1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 12 & 7 \\ 5 & -3 \end{pmatrix}$$

2.4 Square Matrices

Square matrices have the same number of rows and columns; a $k \times k$ square matrix is referred to as a matrix of order k.

The **diagonal** of a square matrix is the vector of matrix elements that have the same subscripts. If $\mathbf{A}$ is a square matrix of order k, then its diagonal is $[a_{11}, a_{22}, \ldots, a_{kk}]'$.

There are several important types of square matrices:

1. **Identity Matrix**: The $n \times n$ identity matrix $\mathbf{I}_n$ is the matrix whose diagonal elements are 1 and all off-diagonal elements are 0. Examples:

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2. **Symmetric Matrix**: A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}'$; this implies that $a_{ij} = a_{ji}$ for all i and j. Examples:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} = \mathbf{A}', \qquad \mathbf{B} = \begin{pmatrix} 4 & 2 & -1 \\ 2 & 1 & 3 \\ -1 & 3 & 1 \end{pmatrix} = \mathbf{B}'$$

3. **Diagonal Matrix**: A matrix $\mathbf{A}$ is diagonal if all of its non-diagonal entries are zero; formally, if $a_{ij} = 0$ for all $i \neq j$ Examples:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

4. **Triangular Matrix**: A matrix is triangular one of two cases. If all entries below the diagonal are zero ($a_{ij} = 0$ for all $i > j$), it is **upper triangular**. Conversely, if all entries above the diagonal are zero ($a_{ij} = 0$ for all $i < j$), it is **lower triangular**.

Examples:

$$\mathbf{A}_{LT} = \begin{pmatrix} 1 & 0 & 0 \\ 4 & 2 & 0 \\ -3 & 2 & 5 \end{pmatrix}, \qquad \mathbf{A}_{UT} = \begin{pmatrix} 1 & 7 & -4 \\ 0 & 3 & 9 \\ 0 & 0 & -3 \end{pmatrix}$$

2.5 Linear Equations

Linear Equation: $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$

a_i are parameters or coefficients. x_i are variables or unknowns.

Linear because only one variable per term and degree is at most 1.

1. $\mathbf{R}^2$: line $\qquad\qquad x_2 = \frac{b}{a_2} - \frac{a_1}{a_2} x_1$

2. $\mathbf{R}^3$: plane $\qquad\qquad x_3 = \frac{b}{a_3} - \frac{a_1}{a_3} x_1 - \frac{a_2}{a_3} x_2$

3. $\mathbf{R}^n$: hyperplane

2.6 Systems of Linear Equations

We are often interested in solving linear systems like\

$$\begin{array}{rrrrr} x & - & 3y & = & -3 \\ 2x & + & y & = & 8 \end{array}$$

More generally, we might have a system of m equations in n unknowns

$$\begin{array}{rrrrrrrr} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & \vdots & & & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

A **solution** to a linear system of m equations in n unknowns is a set of n numbers $x_1, x_2, \cdots, x_n$ that satisfy each of the m equations.

1. $\mathbf{R}^2$: intersection of the lines.

2. $\mathbf{R}^3$: intersection of the planes.

3. $\mathbf{R}^n$: intersection of the hyperplanes.

Example: $x = 3$ and $y = 2$ is the solution to the above 2×2 linear system. Notice from the graph that the two lines intersect at $(3, 2)$.

Does a linear system have one, no, or multiple solutions? For a system of 2 equations in 2 unknowns (i.e., two lines):

1. One solution: The lines intersect at exactly one point.

2. No solution: The lines are parallel.

3. Infinite solutions: The lines coincide.

Methods to solve linear systems:

1. Substitution
2. Elimination of variables
3. Matrix methods

2.7   Systems of Equations as Matrices

Matrices provide an easy and efficient way to represent linear systems such as

$$
\begin{array}{ccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\
\vdots & & & & \vdots & & & & \vdots \\
a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m
\end{array}
$$

as

$$
\mathbf{A}\mathbf{x} = \mathbf{b}
$$

where

1. The $m \times n$ **coefficient matrix A** is an array of mn real numbers arranged in m rows by n columns:

$$
\mathbf{A} = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
$$

2. The unknown quantities are represented by the vector $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

3. The right hand side of the linear system is represented by the vector $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$.

Augmented Matrix: When we append $\mathbf{b}$ to the coefficient matrix $\mathbf{A}$, we get the augmented matrix $\widehat{\mathbf{A}} = [\mathbf{A}|\mathbf{b}]$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & | & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & | & b_2 \\ \vdots & & \ddots & \vdots & | & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & | & b_m \end{pmatrix}$$

2.8 Finding Solutions to Augmented Matrices and Systems of Equations

Row Echelon Form: Our goal is to translate our augmented matrix or system of equations into row echelon form. This will provide us with the values of the vector $\mathbf{x}$ which solve the system. We use the row operations to change coefficients in lower triangle of the augmented matrix to 0. An augmented matrix of the form

$$\begin{pmatrix} \boxed{a'_{11}} & a'_{12} & a'_{13} & \cdots & a'_{1n} & | & b'_1 \\ 0 & \boxed{a'_{22}} & a'_{23} & \cdots & a'_{2n} & | & b'_2 \\ 0 & 0 & \boxed{a'_{33}} & \cdots & a'_{3n} & | & b'_3 \\ 0 & 0 & 0 & \ddots & \vdots & | & \vdots \\ 0 & 0 & 0 & 0 & \boxed{a'_{mn}} & | & b'_m \end{pmatrix}$$

is said to be in row echelon form — each row has more leading zeros than the row preceding it.

Reduced Row Echelon Form: We can go one step further and put the matrix into reduced row echelon form. Reduced row echelon form makes the value of $\mathbf{x}$ which solves the system very obvious. For a system of m equations in m unknowns, with no all-zero rows, the reduced row echelon form would be

$$\begin{pmatrix} \boxed{1} & 0 & 0 & 0 & 0 & | & b_1^* \\ 0 & \boxed{1} & 0 & 0 & 0 & | & b_2^* \\ 0 & 0 & \boxed{1} & 0 & 0 & | & b_3^* \\ 0 & 0 & 0 & \ddots & 0 & | & \vdots \\ 0 & 0 & 0 & 0 & \boxed{1} & | & b_m^* \end{pmatrix}$$

Gaussian and Gauss-Jordan elimination: We can conduct elementary row operations to get our augmented matrix into row echelon or reduced row echelon form. The methods of transforming a matrix or system into row echelon and reduced row echelon form are referred to as Gaussian elimination and Gauss-Jordan elimination, respectively.

Elementary Row Operations: To do Gaussian and Gauss-Jordan elimination, we use three basic operations to transform the augmented matrix into another augmented matrix that represents an equivalent linear system – equivalent in the sense that the same values of x_j solve both the original and transformed matrix/system:

1. Interchanging two rows. $\implies$ Interchanging two equations.

2. Multiplying a row by a constant. $\implies$ Multiplying both sides of an equation by a constant.

3. Adding two rows to each other. $\implies$ Adding two equations to each other.

Interchanging Rows: Suppose we have the augmented matrix

$$\widehat{\mathbf{A}} = \begin{pmatrix} a_{11} & a_{12} & | & b_1 \\ a_{21} & a_{22} & | & b_2 \end{pmatrix}$$

If we interchange the two rows, we get the augmented matrix

$$\begin{pmatrix} a_{21} & a_{22} & | & b_2 \\ a_{11} & a_{12} & | & b_1 \end{pmatrix}$$

which represents a linear system equivalent to that represented by matrix $\widehat{\mathbf{A}}$.

Multiplying by a Constant: If we multiply the second row of matrix $\widehat{\mathbf{A}}$ by a constant c, we get the augmented matrix

$$\begin{pmatrix} a_{11} & a_{12} & | & b_1 \\ ca_{21} & ca_{22} & | & cb_2 \end{pmatrix}$$

which represents a linear system equivalent to that represented by matrix $\widehat{\mathbf{A}}$.

Adding (subtracting) Rows: If we add (subtract) the first row of matrix $\widehat{\mathbf{A}}$ to the second, we obtain the augmented matrix

$$\begin{pmatrix} a_{11} & a_{12} & | & b_1 \\ a_{11} + a_{21} & a_{12} + a_{22} & | & b_1 + b_2 \end{pmatrix}$$

which represents a linear system equivalent to that represented by matrix $\widehat{\mathbf{A}}$.

Exercises: Using Gaussian or Gauss-Jordan elimination, solve the following linear systems by putting them into row echelon or reduced row echelon form:

1.

$$\begin{array}{rrrcr} x & - & 3y & = & -3 \\ 2x & + & y & = & 8 \end{array}$$

2.9   Rank — and Whether a System Has One, Infinite, or No Solutions

We previously noted that a 2×2 system had one, infinite, or no solutions if the two lines intersected, were the same, or were parallel, respectively. More generally, to determine how many solutions exist, we can use information about (1) the number of equations m, (2) the number of unknowns n, and (3) the **rank** of the matrix representing the linear system.

Rank: The row rank or column rank of a matrix is the number of nonzero rows or columns in its row echelon form. The rank also corresponds to the maximum number of linearly independent row or column vectors in the matrix. For any matrix $\mathbf{A}$, the row rank always equals column rank, and we refer to this number as the rank of $\mathbf{A}$.

Examples:

1. $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$ Rank=

2. $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 0 \end{pmatrix}$ Rank=

3. $\begin{pmatrix} 1 & 2 & 3 & | & b_1 \\ 0 & 4 & 5 & | & b_2 \\ 0 & 0 & 0 & | & b_3 \end{pmatrix}$, $b_i \neq 0$ Rank=

Let $\mathbf{A}$ be the coefficient matrix and $\widehat{\mathbf{A}} = [\mathbf{A}|\mathbf{b}]$ be the augmented matrix. Then

1. rank $\mathbf{A} \leq$ rank $\widehat{\mathbf{A}}$ — Augmenting $\mathbf{A}$ with $\mathbf{b}$ can never result in more zero rows than originally in $\mathbf{A}$ itself. Suppose row i in $\mathbf{A}$ is all zeros and that b_i is non-zero. Augmenting $\mathbf{A}$ with $\mathbf{b}$ will yield a non-zero row i in $\widehat{\mathbf{A}}$.

2. rank $\mathbf{A} \leq$ rows $\mathbf{A}$ — By definition of "rank."

3. rank $\mathbf{A} \leq$ cols $\mathbf{A}$ — Suppose there are more rows than columns (otherwise the previous rule applies). Each column can contain at most one pivot. By pivoting, all other entries in a column below the pivot are zeroed. Hence, there will only be as many non-zero rows as pivots, which will equal the number of columns.

Existence of Solutions:

1. Exactly one solution: rank $\mathbf{A} =$ rank $\widehat{\mathbf{A}} =$ rows $\mathbf{A} =$ cols $\mathbf{A}$

   Necessary condition for a system to have a unique solution: that there be exactly as many equations as unknowns.

2. Infinite solutions: rank $\mathbf{A} =$ rank $\widehat{\mathbf{A}}$ and cols $\mathbf{A} >$ rank $\mathbf{A}$

   If a system has a solution and has more unknowns than equations, then it has infinitely many solutions.

3. No solution: rank $\mathbf{A} <$ rank $\widehat{\mathbf{A}}$

   Then there is a zero row i in $\mathbf{A}$'s reduced echelon that corresponds to a non-zero row i in $\widehat{\mathbf{A}}$'s reduced echelon. Row i of the $\widehat{\mathbf{A}}$ translates to the equation

$$0x_{i1} + 0x_{i2} + \cdots + 0x_{in} = b_i'$$

   where $b_i' \neq 0$. Hence the system has no solution.

Find the rank and number of solutions for the systems of equations below.:

1.
$$
\begin{array}{rcrcr}
x & - & 3y & = & -3 \\
2x & + & y & = & 8
\end{array}
$$

2.
$$
\begin{array}{rcrcrcr}
x & + & 2y & + & 3z & = & 6 \\
2x & - & 3y & + & 2z & = & 14 \\
3x & + & y & - & z & = & -2
\end{array}
$$

3.
$$
\begin{array}{rcrcrcr}
x & + & 2y & - & 3z & = & -4 \\
2x & + & y & - & 3z & = & 4 \\
3x & + & 6y & - & 9z & = & -12
\end{array}
$$

4.

$$\begin{array}{ccccccccc}
x & + & 2y & + & 3z & + & 4w & = & 5 \\
x & + & 3y & + & 5z & + & 7w & = & 11 \\
x & & & - & z & - & 2w & = & -6
\end{array}$$

2.10 The Inverse of a Matrix

Inverse Matrix: An $n \times n$ matrix $\mathbf{A}$ is **nonsingular** or **invertible** if there exists an $n \times n$ matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

where $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$. If there is no such $\mathbf{A}^{-1}$, then $\mathbf{A}$ is singular or noninvertible.

Example: Let

$$\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 2 & 2 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} -1 & \frac{3}{2} \\ 1 & -1 \end{pmatrix}$$

Since

$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}_n$$

we conclude that $\mathbf{B}$ is the inverse, $\mathbf{A}^{-1}$, of $\mathbf{A}$ and that $\mathbf{A}$ is nonsingular.

Properties of the Inverse:

1. If the inverse exists, it is unique.

2. $\mathbf{A}$ nonsingular $\implies$ $\mathbf{A}^{-1}$ nonsingular $\qquad\qquad (\mathbf{A}^{-1})^{-1} = \mathbf{A}$

3. $\mathbf{A}$ nonsingular $\implies$ $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$

Procedure to Find $\mathbf{A}^{-1}$: We know that if $\mathbf{B}$ is the inverse of $\mathbf{A}$, then

$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}_n$$

Looking only at the first and last parts of this

$$\mathbf{A}\mathbf{B} = \mathbf{I}_n$$

Solving for $\mathbf{B}$ is equivalent to solving for n linear systems, where each column of $\mathbf{B}$ is solved for the corresponding column in $\mathbf{I}_n$. We can solve the systems simultaneously by augmenting $\mathbf{A}$ with $\mathbf{I}_n$ and performing Gauss-Jordan elimination on $\mathbf{A}$. If Gauss-Jordan elimination on $[\mathbf{A}|\mathbf{I}_n]$ results in $[\mathbf{I}_n|\mathbf{B}]$, then $\mathbf{B}$ is the inverse of $\mathbf{A}$. Otherwise, $\mathbf{A}$ is singular.\ [12pt] To summarize: To calculate the inverse of $\mathbf{A}$

1. Form the augmented matrix $[\mathbf{A}|\mathbf{I}_n]$

2. Using elementary row operations, transform the augmented matrix to reduced row echelon form.

3. The result of step 2 is an augmented matrix $[\mathbf{C}|\mathbf{B}]$.

   (a) If $\mathbf{C} = \mathbf{I}_n$, then $\mathbf{B} = \mathbf{A}^{-1}$.

   (b) If $\mathbf{C} \neq \mathbf{I}_n$, then $\mathbf{C}$ has a row of zeros. $\mathbf{A}$ is singular and $\mathbf{A}^{-1}$ does not exist.

Exercise:

Find the inverse of $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{pmatrix}$

2.11 Linear Systems and Inverses

Let's return to the matrix representation of a linear system

$$\mathbf{Ax} = \mathbf{b}$$

If $\mathbf{A}$ is an $n \times n$ matrix, then $\mathbf{Ax} = \mathbf{b}$ is a system of n equations in n unknowns. Suppose $\mathbf{A}$ is nonsingular. Then $\mathbf{A^{-1}}$ exists. To solve this system, we can premultiply each side by $\mathbf{A^{-1}}$ and reduce it as follows:

$$
\begin{aligned}
\mathbf{A^{-1}(Ax)} &= \mathbf{A^{-1}b} \\
\mathbf{(A^{-1}A)x} &= \mathbf{A^{-1}b} \\
\mathbf{I_n x} &= \mathbf{A^{-1}b} \\
\mathbf{x} &= \mathbf{A^{-1}b}
\end{aligned}
$$

Hence, given $\mathbf{A}$ and $\mathbf{b}$ and given that $\mathbf{A}$ is nonsingular, then $\mathbf{x} = \mathbf{A^{-1}b}$ is a unique solution to this system.

Notice also that the requirements for $\mathbf{A}$ to be nonsingular correspond to the requirements for a linear system to have a unique solution:

$$\text{rank}\mathbf{A} = \text{rows}\mathbf{A} = \text{cols}\mathbf{A}$$

2.12 Determinants

Singularity: Determinants can be used to *determine* whether a square matrix is nonsingular.

A square matrix is nonsingular iff its determinant is not zero.

Determinant of a 1×1 matrix, $\mathbf{A}$, equals a_{11}

Determinant of a 2×2 matrix, $\mathbf{A}$, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$:

$$
\begin{aligned}
\det(\mathbf{A}) &= |\mathbf{A}| \\
&= a_{11}|a_{22}| - a_{12}|a_{21}| \\
&= a_{11}a_{22} - a_{12}a_{21}
\end{aligned}
$$

We can extend the second to last equation above to get the definition of the determinant of a 3×3 matrix:

$$
\begin{aligned}
\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})
\end{aligned}
$$

Let's extend this now to any $n \times n$ matrix. Let's define $\mathbf{A}_{ij}$ as the $(n-1) \times (n-1)$ submatrix of $\mathbf{A}$ obtained by deleting row i and column j. Let the (i,j)th **minor** of $\mathbf{A}$ be the determinant of $\mathbf{A}_{ij}$:

$$
M_{ij} = |\mathbf{A}_{ij}|
$$

Then for any $n \times n$ matrix $\mathbf{A}$

$$
|\mathbf{A}| = a_{11}M_{11} - a_{12}M_{12} + \cdots + (-1)^{n+1}a_{1n}M_{1n}
$$

Example: Does the following matrix have an inverse?

$$
\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{pmatrix}
$$

1. Calculate its determinant.

$$
\begin{aligned}
&= 1(2 - 15) - 1(0 - 15) + 1(0 - 10) \\
&= -13 + 15 - 10 \\
&= -8
\end{aligned}
$$

2. Since $|\mathbf{A}| \neq 0$, we conclude that $\mathbf{A}$ has an inverse.

Determinant of Triangular or Diagonal Matrices: For any upper-triangular, lower-triangular, or diagonal matrix, the determinant is just the product of the diagonal terms.

Example: Suppose we have the following square matrix in row echelon form (i.e., upper triangular)

$$
\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{pmatrix}
$$

Then

$$
|\mathbf{R}| = r_{11}\begin{vmatrix} r_{22} & r_{23} \\ 0 & r_{33} \end{vmatrix} = r_{11}r_{22}r_{33}
$$

Properties of Determinants:

1. $|\mathbf{A}| = |\mathbf{A}^T|$

2. If $\mathbf{B}$ results from $\mathbf{A}$ by interchanging two rows, then $|\mathbf{B}| = -|\mathbf{A}|$.

3. If two rows of $\mathbf{A}$ are equal, then $|\mathbf{A}| = 0$.   (Notice that in this case rank $\mathbf{A} \neq$ rows $\mathbf{A}$, which was one of the conditions for the existence of a unique solution.)

4. If a row of $\mathbf{A}$ consists of all zeros, then $|\mathbf{A}| = 0$.   (Same as 3.)

5. If $\mathbf{B}$ is obtained by multiplying a row of $\mathbf{A}$ by a scalar s, then $|\mathbf{B}| = s|\mathbf{A}|$.

6. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by adding to the ith row of $\mathbf{A}$ the jth row $(i \neq j)$ multiplied by a scalar s, then $|\mathbf{B}| = |\mathbf{A}|$.   (i.e., If the row isn't simply multiplied by a scalar and left, then the determinant remains the same.)

7. If no row interchanges and no scalar multiplications of a single row are used to compute the row echelon form $\mathbf{R}$ from the $n \times n$ coefficient matrix $\mathbf{A}$, then $|\mathbf{A}| = |\mathbf{R}|$.   (Implied by the previous properties.)

8. A square matrix is nonsingular iff its determinant $\neq 0$.   (Implied by the previous properties.)

9. $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$

10. If $\mathbf{A}$ is nonsingular, then $|\mathbf{A}| \neq 0$ and $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$.

2.13 Getting Inverse of a Matrix using its Determinant and Matrix of Cofactors

Thus far, we have a number of algorithms to

1. Find the solution of a linear system,
2. Find the inverse of a matrix

but these remain just that — algorithms. At this point, we have no way of telling how the solutions x_j change as the parameters a_{ij} and b_i change, except by changing the values and "rerunning" the algorithms.

With determinants, we can 1. Provide an explicit formula for the inverse, and 2. Provide an explicit formula for the solution of an $n \times n$ linear system.

Hence, we can examine how changes in the parameters and b_i affect the solutions x_j.

Determinant Formula for the Inverse of a 2×2: The determinant of a 2×2 matrix $\mathbf{A}$ $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is defined as:

$$\frac{1}{\det(\mathbf{A})} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

2.14 Inverse of Larger Matrices

Cofactors and Adjoint Matrices:

First define the (i,j)th **cofactor** C_{ij} of $\mathbf{A}$ as $(-1)^{i+j} M_{ij}$. Recall that M_{ij} is the minor of $\mathbf{A}$, defined as the determinant of the matrix that results from removing row i and column j from $\mathbf{A}$.

Then define the **adjoint** of $\mathbf{A}$ as the $n \times n$ matrix whose (i,j)th entry is C_{ji} (notice the switch in indices!). In other words, $\mathrm{adj}(\mathbf{A})$ is the transpose of the cofactor matrix of $\mathbf{A}$.

Then the inverse of $\mathbf{A}$ is defined as the reciprocal of the determinant of $\mathbf{A}$ times its adjoint, given by the formula

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathrm{adj}\,\mathbf{A} = \begin{pmatrix} \frac{C_{11}}{|\mathbf{A}|} & \frac{C_{21}}{|\mathbf{A}|} & \cdots & \frac{C_{n1}}{|\mathbf{A}|} \\ \frac{C_{12}}{|\mathbf{A}|} & \frac{C_{22}}{|\mathbf{A}|} & \cdots & \frac{C_{n2}}{|\mathbf{A}|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{C_{1n}}{|\mathbf{A}|} & \frac{C_{2n}}{|\mathbf{A}|} & \cdots & \frac{C_{nn}}{|\mathbf{A}|} \end{pmatrix}$$

Exercises:

Let $\mathbf{A} = \begin{pmatrix} 5 & 1 & 3 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{pmatrix}$

1. Find the determinant of $\mathbf{A}$.

$$\begin{aligned} \det \mathbf{A} \quad &= \\ &= \\ &= \\ &= \end{aligned}$$

2. Find the adjoint matrix of $\mathbf{A}$.

$\mathrm{adj}(\{\mathbf{A}\}) =$

3. **Find the inverse of $\{\mathbf{A}\}$.**

$\mathbf{A}^{-1} =$

Chapter 3

Functions and Notation

Topics Dimensionality; Interval Notation for $\mathbf{R}^1$; Neighborhoods: Intervals, Disks, and Balls; Introduction to Functions; Domain and Range; Some General Types of Functions; $\log$, $\ln$, and $\exp$; Other Useful Functions; Graphing Functions; Solving for Variables; Finding Roots; Limit of a Function; Continuity; Sets, Sets, and More Sets.

Much of the material and examples for this lecture are taken from Simon & Blume (1994) *Mathematics for Economists*, **Boyce & Diprima (1988)** *Calculus*, **and Protter & Morrey (1991)** *A First Course in Real Analysis*

3.1   Dimensionality

$\mathbf{R}^1$ **is the set of all real numbers extending from $-\infty$ to $+\infty$ — i.e., the real number line. $\mathbf{R}^n$ is an n-dimensional space (often referred to as Euclidean space), where each of the n axes extends from $-\infty$ to $+\infty$.**

- $\mathbf{R}^1$ **is a one dimensional line.**
- $\mathbf{R}^2$ **is a two dimensional plane.**
- $\mathbf{R}^3$ **is a three dimensional space.**
- $\mathbf{R}^4$ **could be 3-D plus time (or temperature, etc).**

Points in $\mathbf{R}^n$ are ordered n-tuples, where each element of the n-tuple represents the coordinate along that dimension.

- $\mathbf{R}^1$**: (3)**
- $\mathbf{R}^2$**: (-15, 5)**
- $\mathbf{R}^3$**: (86, 4, 0)**

3.2   Interval Notation for $\mathbf{R}^1$

Open interval:

$$(a, b) \equiv \{x \in \mathbf{R}^1 : a < x < b\}$$

x **is a one-dimensional element in which x is greater than a and less than b**

Closed interval:

$$[a, b] \equiv \{x \in \mathbf{R}^1 : a \leq x \leq b\}$$

x **is a one-dimensional element in which x is greater or equal to than a and less than or equal to b**

Half open, half closed:
$$(a, b] \equiv \{x \in \mathbf{R}^1 : a < x \leq b\}$$

x **is a one-dimensional element in which x is greater than a and less than or equal to b**

3.3 Neighborhoods: Intervals, Disks, and Balls

In many areas of math, we need a formal construct for what it means to be "near" a point c in $\mathbf{R}^n$. This is generally called the neighborhood of c. It's represented by an open interval, disk, or ball, depending on whether $\mathbf{R}^n$ is of one, two, or more dimensions, respectively.

Given the point c, these are defined as

- ϵ**-interval in $\mathbf{R}^1$:** $\{x : |x - c| < \epsilon\}$**. x is in the neighborhood of {c} if it is in the open interval** $(c - \epsilon, c + \epsilon)$**.**
- ϵ**-disk in $\mathbf{R}^2$:** $\{x : ||x - c|| < \epsilon\}$**. x is in the neighborhood of {c} if it is inside the circle or disc with center c and radius ϵ.**
- ϵ**-ball in $\mathbf{R}^n$:** $\{x : ||x - c|| < \epsilon\}$**. x is in the neighborhood of {c} if it is inside the sphere or ball with center c and radius ϵ.**

3.4 Introduction to Functions

A function (in $\mathbf{R}^1$) is a mapping, or transformation, that relates members of one set to members of another set. For instance, if you have two sets: set A and set B, a function from A to B maps every value a in set A such that $f(a) \in B$. Functions can be "many-to-one", where many values or combinations of values from set A produce a single output in set B, or they can be "one-to-one", where each value in set A corresponds to a single value in set B.

Examples: Mapping notation

- **Function of one variable:** $f : \mathbf{R}^1 \to \mathbf{R}^1 \backslash$ $f(x) = x + 1$**. For each x in $\mathbf{R}^1$, $f(x)$ assigns the number $x + 1$.**
- **Function of two variables:** $f : \mathbf{R}^2 \to \mathbf{R}^1$**.** $f(x, y) = x^2 + y^2$**. For each ordered pair (x, y) in $\mathbf{R}^2$, $f(x, y)$ assigns the number $x^2 + y^2$.**

We often use variable x as input and another y as output, e.g. $y = x + 1$

3.5 Domain and Range/Image

Some functions are defined only on proper subsets of $\mathbf{R}^n$.

- **Domain: the set of numbers in X at which $f(x)$ is defined.**
- **Range: elements of Y assigned by $f(x)$ to elements of X, or**

$$f(X) = \{y : y = f(x), x \in X\}$$

  Most often used when talking about a function $f : \mathbf{R}^1 \to \mathbf{R}^1$.
- **Image: same as range, but more often used when talking about a function $f : \mathbf{R}^n \to \mathbf{R}^1$.**

3.6 Some General Types of Functions

Monomials: $f(x) = ax^k \backslash$ a **is the coefficient.** k **is the degree.$\backslash$ Examples:** $y = x^2$**,** $y = -\frac{1}{2}x^3$

Polynomials: sum of monomials. Examples: $y = -\frac{1}{2}x^3 + x^2$, $y = 3x + 5$

The degree of a polynomial is the highest degree of its monomial terms. Also, it's often a good idea to write polynomials with terms in decreasing degree.

Rational Functions: ratio of two polynomials.

Examples: $y = \frac{x}{2}$, $y = \frac{x^2+1}{x^2-2x+1}$

Exponential Functions: Example: $y = 2^x$

Trigonometric Functions: Examples: $y = \cos(x)$, $y = 3\sin(4x)$

3.7 log, ln, and exp

Relationship of logarithmic and exponential functions:

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

The log function can be thought of as an inverse for exponential functions. a **is referred to as the "base" of the logarithm.**

Common Bases: The two most common logarithms are base 10 and base e**.**

1. **Base 10:** $y = \log_{10}(x) \iff 10^y = x$. **The base 10 logarithm is often simply written as "log(x)" with no base denoted.**
2. **Base** e**:** $y = \log_e(x) \iff e^y = x$. **The base** e **logarithm is referred to as the "natural" logarithm and is written as "ln(x)".**

Properties of exponential functions:

- $a^x a^y = a^{x+y}$
- $a^{-x} = 1/a^x$
- $a^x/a^y = a^{x-y}$
- $(a^x)^y = a^{xy}$
- $a^0 = 1$

Properties of logarithmic functions (any base):

Generally, when statisticians or social scientists write $\log(x)$ **they mean** $\log_e(x)$**. In other words:** $\log_e(x) \equiv \ln(x) \equiv \log(x)$

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

and

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

- $\log(xy) = \log(x) + \log(y)$
- $\log(x^y) = y\log(x)$
- $\log(1/x) = \log(x^{-1}) = -\log(x)$
- $\log(x/y) = \log(x \cdot y^{-1}) = \log(x) + \log(y^{-1}) = \log(x) - \log(y)$
- $\log(1) = \log(e^0) = 0$

Change of Base Formula: Use the change of base formula to switch bases as necessary:

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}$$

Example:

$$\log_{10}(x) = \frac{\ln(x)}{\ln(10)}$$

- $\log_{10}(\sqrt{10}) = \log_{10}(10^{1/2}) =$

- $\log_{10}(1) = \log_{10}(10^0) =$

- $\log_{10}(10) = \log_{10}(10^1) =$

- $\log_{10}(100) = \log_{10}(10^2) =$

- $\ln(1) = \ln(e^0) =$

- $\ln(e) = \ln(e^1) =$

3.8 Other Useful Functions

Factorials!:

$$x! = x \cdot (x-1) \cdot (x-2) \cdots (1)$$

Modulo: Tells you the remainder when you divide one number by another. Can be extremely useful for programming: x mod y or x % y

- $17 \mod 3 = 2$
- $100 \% 30 = 10$

Summation:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \cdots + x_n$$

- $\sum_{i=1}^{n} cx_i = c \sum_{i=1}^{n} x_i$

- $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

- $\sum_{i=1}^{n} c = nc$

Product:

$$\prod_{i=1}^{n} x_i = x_1 x_2 x_3 \cdots x_n$$

Properties:

- $\prod_{i=1}^{n} cx_i = c^n \prod_{i=1}^{n} x_i$

- $\prod_{i=1}^{n} (x_i + y_i) = $ **a total mess**

- $\prod_{i=1}^{n} c = c^n$

You can use logs to go between sum and product notation. This will be particularly important when you're learning maximum likelihood estimation.

$$
\begin{aligned}
\log\left(\prod_{i=1}^{n} x_i\right) &= \log(x_1 \cdot x_2 \cdot x_3 \cdots x_n) \\
&= \log(x_1) + \log(x_2) + \log(x_3) + \cdots + \log(x_n) \\
&= \sum_{i=1}^{n} \log(x_i)
\end{aligned}
$$

Therefore, you can see that the log of a product is equal to the sum of the logs. We can write this more generally by adding in a constant, c:

$$
\begin{aligned}
\log\left(\prod_{i=1}^{n} cx_i\right) &= \log(cx_1 \cdot cx_2 \cdots cx_n) \\
&= \log(c^n \cdot x_1 \cdot x_2 \cdots x_n) \\
&= \log(c^n) + \log(x_1) + \log(x_2) + \cdots + \log(x_n) \\
&= n\log(c) + \sum_{i=1}^{n} \log(x_i)
\end{aligned}
$$

3.9 Graphing Functions

What can a graph tell you about a function?

- Is the function increasing or decreasing? Over what part of the domain?
- How "fast" does it increase or decrease?
- Are there global or local maxima and minima? Where?
- Are there inflection points?
- Is the function continuous?
- Is the function differentiable?
- Does the function tend to some limit?
- Other questions related to the substance of the problem at hand.

3.10 Solving for Variables and Finding Inverses

Sometimes we're given a function $y = f(x)$ and we want to find how x varies as a function of y. If f is a one-to-one mapping, then it has an inverse.

Use algebra to move x to the left hand side (LHS) of the equation and so that the right hand side (RHS) is only a function of y.\

Examples: (we want to solve for x)

$$
y = 3x + 2 \quad \implies \quad -3x = 2 - y \quad \implies \quad 3x = y - 2 \quad \implies \quad x = \frac{1}{3}(y - 2)
$$

$$
y = 3x - 4z + 2 \quad \implies \quad y + 4z - 2 = 3x \quad \implies \quad x = \frac{1}{3}(y + 4z - 2)
$$

$$y = e^x + 4 \implies \quad y - 4 = e^x$$
$$\implies \ln(y - 4) = \ln(e^x)$$
$$\implies \quad x = \ln(y - 4)$$

Sometimes (often?) the inverse does not exist. For example, we're given the function $y = x^2$ (a parabola). Solving for x, we get $x = \pm\sqrt{y}$. For each value of y, there are two values of x.

3.11   Finding the Roots or Zeroes of a Function

Solving for variables is especially important when we want to find the roots of an equation: those values of variables that cause an equation to equal zero. Especially important in finding equilibria and in doing maximum likelihood estimation.

Procedure: Given $y = f(x)$, set $f(x) = 0$. Solve for x.

Multiple Roots:

$$f(x) = x^2 - 9 \quad \implies \quad 0 = x^2 - 9 \quad \implies \quad 9 = x^2 \quad \implies \quad \pm\sqrt{9} = \sqrt{x^2} \quad \implies \quad \pm 3 = x$$

Quadratic Formula: For quadratic equations $ax^2 + bx + c = 0$, use the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Examples:

1. $f(x) = 3x + 2$

2. $f(x) = e^{-x} - 10$

3. $f(x) = x^2 + 3x - 4 = 0$

3.12   The Limit of a Function

We're often interested in determining if a function f approaches some number L as its independent variable x moves to some number c (usually 0 or $\pm\infty$). If it does, we say that the limit of $f(x)$, as x approaches c, is L: $\lim_{x \to c} f(x) = L$.

For a limit L to exist, the function $f(x)$ must approach L from both the left and right.

Limit of a function. Let $f(x)$ be defined at each point in some open interval containing the point c. Then L equals $\lim_{x \to c} f(x)$ if for any (small positive) number ϵ, there exists a corresponding number $\delta > 0$ such that if $0 < |x - c| < \delta$, then $|f(x) - L| < \epsilon$.

Note: f(x) does not necessarily have to be defined at c for $\lim_{x \to c}$ to exist.

Uniqueness: $\lim_{x \to c} f(x) = L$ and $\lim_{x \to c} f(x) = M \implies L = M$

Properties: Let f and g be functions with $\lim_{x \to c} f(x) = A$ and $\lim_{x \to c} g(x) = B$.

- $\lim_{x \to c}[f(x) + g(x)] = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$

- $\lim_{x \to c} \alpha f(x) = \alpha \lim_{x \to c} f(x)$
- $\lim_{x \to c} f(x)g(x) = [\lim_{x \to c} f(x)][\lim_{x \to c} g(x)]$
- $\lim_{x \to c} \frac{f(x)}{g(x)} = \frac{\lim_{x \to c} f(x)}{\lim_{x \to c} g(x)}$, **provided** $\lim_{x \to c} g(x) \neq 0$.

Note: In a couple days we'll talk about L'Hˆopital's Rule, which uses simple calculus to help find the limits of functions like this.

Examples:

$$\lim_{x \to c} k =$$

$$\lim_{x \to c} x =$$

$$\lim_{x \to 2} (2x - 3) =$$

$$\lim_{x \to c} x^n =$$

Types of limits:

1. **Right-hand limit:** The value approached by $f(x)$ when you move from right to left.
2. **Left-hand limit:** The value approached by $f(x)$ when you move from left to right.
3. **Infinity:** The value approached by $f(x)$ as x grows infinitely large. Sometimes this may be a number; sometimes it might be ∞ or $-\infty$.
4. **Negative infinity:** The value approached by $f(x)$ as x grows infinitely negative. Sometimes this may be a number; sometimes it might be ∞ or $-\infty$.

3.13 Continuity

Continuity: Suppose that the domain of the function f includes an open interval containing the point c. Then f is continuous at c if $\lim_{x \to c} f(x)$ exists and if $\lim_{x \to c} f(x) = f(c)$. Further, f is continuous on an open interval (a, b) if it is continuous at each point in the interval.

Examples: Continuous functions.

Properties:

1. If f and g are continuous at point c, then $f + g$, $f - g$, $f \cdot g$, $|f|$, and αf are continuous at point c also. f/g is continuous, provided $g(c) \neq 0$.
2. **Boundedness:** If f is continuous on the closed bounded interval $[a, b]$, then there is a number K such that $|f(x)| \leq K$ for each x in $[a, b]$.
3. **Max/Min:** If f is continuous on the closed bounded interval $[a, b]$, then f has a maximum and a minimum on $[a, b]$. They may be located at the end points.

3.14 Sets

Interior Point: The point x is an interior point of the set S if x is in S and if there is some ϵ-ball around x that contains only points in S. The interior of S is the collection of all interior points in S. The interior can also be defined as the union of all open sets in S.

- If the set S is circular, the interior points are everything inside of the circle, but not on the circle's rim.
- **Example:** The interior of the set $\{(x, y) : x^2 + y^2 \leq 4\}$ is $\{(x, y) : x^2 + y^2 < 4\}$.

Boundary Point: The point x is a boundary point of the set S if every ϵ-ball around x contains both points that are in S and points that are outside S. The boundary is the collection of all boundary points.

- If the set S is circular, the boundary points are everything on the circle's rim.
- Example: The boundary of $\{(x,y) : x^2 + y^2 \leq 4\}$ is $\{(x,y) : x^2 + y^2 = 4\}$.

Open: A set S is open if for each point x in S, there exists an open ϵ-ball around x completely contained in S.

- If the set S is circular and open, the points contained within the set get infinitely close to the circle's rim, but do not touch it.
- Example: $\{(x,y) : x^2 + y^2 < 4\}$

Closed: A set S is closed if it contains all of its boundary points.

- If the set S is circular and closed, the set contains all points within the rim as well as the rim itself.
- Example: $\{(x,y) : x^2 + y^2 \leq 4\}$
- Note: a set may be neither open nor closed. Example: $\{(x,y) : 2 < x^2 + y^2 \leq 4\}$

Complement: The complement of set S is everything outside of S.

- If the set S is circular, the complement of S is everything outside of the circle.
- Example: The complement of $\{(x,y) : x^2 + y^2 \leq 4\}$ is $\{(x,y) : x^2 + y^2 > 4\}$.

Closure: The closure of set S is the smallest closed set that contains S.

- Example: The closure of $\{(x,y) : x^2 + y^2 < 4\}$ is $\{(x,y) : x^2 + y^2 \leq 4\}$

Bounded: A set S is bounded if it can be contained within an ϵ-ball.

- Examples: Bounded: any interval that doesn't have ∞ or $-\infty$ as endpoints; any disk in a plane with finite radius.
- Unbounded: the set of integers in $\mathbf{R}^1$; any ray.

Compact: A set is compact if and only if it is both closed and bounded.

Empty: The empty (or null) set is a unique set that has no elements, denoted by $\{\}$ or .

- Examples: The set of squares with 5 sides; the set of countries south of the South Pole.
- The set, S, denoted by $\{\}$ is technically *not* empty. That is because this set contains the empty set within it, so S is not empty.

Chapter 4

Calculus I

Topics: Sequences; Limit of a Sequence; Series; Derivatives; Higher-Order Derivatives; Composite Functions and The Chain Rule; Derivatives of Exp and Ln; Maxima and Minima; Partial Derivatives; L'Hˆopital's Rule; Taylor Approximation; Derivative Calculus in 6 Steps

Much of the material and examples for this lecture are taken from Simon & Blume (1994) *Mathematics for Economists* and from Boyce & Diprima (1988) *Calculus*

4.1   Sequences

A sequence $\{y_n\} = \{y_1, y_2, y_3, \ldots, y_n\}$ is an ordered set of real numbers, where y_1 is the first term in the sequence and y_n is the nth term. Generally, a sequence is infinite, that is it extends to $n = \infty$. We can also write the sequence as $\{y_n\}_{n=1}^{\infty}$.

Examples:

Think of sequences like functions. Before, we had $y = f(x)$ with x specified over some domain. Now we have $\{y_n\} = \{f(n)\}$ with $n = 1, 2, 3, \ldots$.

Three kinds of sequences:

1. Sequences like **1** above that converge to a limit.
2. Sequences like **2** above that increase without bound.
3. Sequences like **3** above that neither converge nor increase without bound — alternating over the number line.

Boundedness and monotonicity:

1. **Bounded:** if $|y_n| \leq K$ for all n
2. **Monotonically Increasing:** $y_{n+1} > y_n$ for all n
3. **Monotonically Decreasing:** $y_{n+1} < y_n$ for all n

Subsequence: Choose an infinite collection of entries from $\{y_n\}$, retaining their order.

4.2   The Limit of a Sequence

We're often interested in whether a sequence converges to a limit. Limits of sequences are conceptually similar to the limits of functions addressed in the previous lecture.

Limit of a sequence. The sequence $\{y_n\}$ has the limit L, that is $\lim_{n \to \infty} y_n = L$, if for any $\epsilon > 0$ there is an integer N (which depends on ϵ) with the property that $|y_n - L| < \epsilon$ for each $n > N$. $\{y_n\}$ is said to converge to L. If the above does not hold, then $\{y_n\}$ diverges.

Uniqueness: If $\{y_n\}$ **converges, then the limit** L **is unique.**

Properties: Let $\lim\limits_{n\to\infty} y_n = A$ **and** $\lim\limits_{n\to\infty} z_n = B$. **Then**

1. $\lim\limits_{n\to\infty} [\alpha y_n + \beta z_n] = \alpha A + \beta B$

2. $\lim\limits_{n\to\infty} y_n z_n = AB$

3. $\lim\limits_{n\to\infty} \frac{y_n}{z_n} = \frac{A}{B}$, **provided** $B \neq 0$

Examples

1. $\lim\limits_{n\to\infty} \left\{ 2 - \frac{1}{n^2} \right\} =$

2. $\lim\limits_{n\to\infty} \left\{ \frac{4^n}{n!} \right\} =$

Finding the limit of a sequence in $\mathbf{R}^n$ **is similar to that in** $\mathbf{R}^1$.

Limit of a sequence of vectors. The sequence of vectors $\{\mathbf{y_n}\}$ **has the limit L, that is** $\lim\limits_{n\to\infty} \mathbf{y_n} = \mathbf{L}$, **if for any** ϵ **there is an integer** N **where** $\|\mathbf{y_n} - \mathbf{L}\| < \epsilon$ **for each** $n > N$. **The sequence of vectors** $\{\mathbf{y_n}\}$ **is said to converge to the vector L — and the distances between** $\mathbf{y_n}$ **and L converge to zero.**

Think of each coordinate of the vector $\mathbf{y_n}$ **as being part of its own sequence over** n. **Then a sequence of vectors in** $\mathbf{R}^n$ **converges if and only if all** n **sequences of its components converge.**

Examples:

1. **The sequence** $\{y_n\}$ **where** $y_n = \left(\frac{1}{n}, 2 - \frac{1}{n^2} \right)$ **converges to** $(0, 2)$.

2. **The sequence** $\{y_n\}$ **where** $y_n = \left(\frac{1}{n}, (-1)^n \right)$ **does not converge, since** $\{(-1)^n\}$ **does not converge.**

Bolzano-Weierstrass Theorem: Any sequence contained in a compact (i.e., closed and bounded) subset of $\mathbf{R}^n$ **contains a convergent subsequence.\ Example: Take the sequence** $\{y_n\} = \{(-1)^n\}$, **w** **two accumulating points, but no limit. However, i** **closed and bounded.**

4.3   Series

The sum of the terms of a sequence is a series. As there are both finite and infinite sequences, there are finite and infinite series.

The series associated with the sequence $\{y_n\} = \{y_1, y_2, y_3, \dots, y_n\} = \{y_n\}_{n=1}^{\infty}$ **is** $\sum_{n=1}^{\infty} y_n$. **The** n**th partial sum** S_n **is defined as** $S_n = \sum_{k=1}^{n} y_k$, **the sum of the first** n **terms of the sequence.**

Convergence: A series $\sum y_n$ **converges if the sequence of partial sums** $\{S_1, S_2, S_3, \dots\}$ **converges, i.e has a finite limit.**

A geometric series is a series that can be written as $\sum_{n=0}^{\infty} r^n$, **where** r **is called the ratio. A geometric series converges to** $\frac{1}{1-r}$ **if** $|r| < 1$ **and diverges otherwise. For example,** $\sum_{n=0}^{\infty} \frac{1}{2^n} = 2$.

Examples of other series:

1. $\sum_{n=0}^{\infty} \frac{1}{n!} = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots = e$.
   This one is especially useful in statistics and probability.

2. $\sum_{n=1}^{\infty} \frac{1}{n} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots = \infty$ **(harmonic series)**

4.4 Derivatives

The derivative of f at x is its rate of change at x — i.e., how much $f(x)$ changes with a change in x.

- For a line, the derivative is the slope.\
- For a curve, the derivative is the slope of the line tangent to the curve at x.

Derivative: Let f be a function whose domain includes an open interval containing the point x. The derivative of f at x is given by

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

If $f'(x)$ exists at a point x, then f is said to be **differentiable** at x. Similarly, if $f'(x)$ exists for every point a long an interval, then f is differentiable along that interval. For f to be differentiable at x, f must be both continuous and "smooth" at x. The process of calculating $f'(x)$ is called **differentiation**.

Notation for derivatives:

1. y', $f'(x)$ **(Prime or Lagrange Notation)**

2. Dy, $Df(x)$ **(Operator Notation)**

3. $\frac{dy}{dx}$, $\frac{df}{dx}(x)$ **(Leibniz's Notation)**

Properties of derivatives: Suppose that f and g are differentiable at x and that α is a constant. Then the functions $f \pm g$, αf, fg, and f/g (provided $g(x) \neq 0$) are also differentiable at x. Additionally,

1. **Power rule:** $[x^k]' = kx^{k-1}$

2. **Sum rule:** $[f(x) \pm g(x)]' = f'(x) \pm g'(x)$

3. **Constant rule:** $[\alpha f(x)]' = \alpha f'(x)$

4. **Product rule:** $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$

5. **Quotient rule:** $[f(x)/g(x)]' = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$, $g(x) \neq 0$

4.5 Higher-Order Derivatives or, Derivatives of Derivatives of Derivatives

We can keep applying the differentiation process to functions that are themselves derivatives. The derivative of $f'(x)$ with respect to x, would then be

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

and so on. Similarly, the derivative of $f''(x)$ would be denoted $f'''(x)$.

First derivative: $f'(x)$, y', $\frac{df(x)}{dx}$, $\frac{dy}{dx}$

Second derivative: $f''(x)$, y'', $\frac{d^2 f(x)}{dx^2}$, $\frac{d^2 y}{dx^2}$

nth derivative: $\frac{d^n f(x)}{dx^n}, \frac{d^n y}{dx^n}$

Example:

$$f(x) = x^3 f'(x) \qquad\qquad = 3x^2 f''(x) = 6x f'''(x) \qquad\qquad = 6f''''(x) = 0$$

4.6   Composite Functions and the Chain Rule

Composite functions are formed by substituting one function into another and are denoted by

$$(f \circ g)(x) = f[g(x)]$$

To form $f[g(x)]$, the range of g must be contained (at least in part) within the domain of f. The domain of $f \circ g$ consists of all the points in the domain of g for which $g(x)$ is in the domain of f.

Examples:

1. $f(x) = \ln x$,
   $g(x) = x^2$
   $(f \circ g)(x) = \ln x^2$,
   $(g \circ f)(x) = [\ln x]^2$,
   Notice that $f \circ g$ and $g \circ f$ are not the same functions.

2. $f(x) = 4 + \sin x$,
   $g(x) = \sqrt{1 - x^2}$,
   $(f \circ g)(x) = 4 + \sin \sqrt{1 - x^2}$,
   $(g \circ f)(x)$ **does not exist.** $\sqrt{1 - (4 + \sin(x))^2}$ **is not a real number because** $1 - (4 + \sin(x))^2$ **is always negative.**

Chain Rule: Let $y = (f \circ g)(x) = f[g(x)]$. The derivative of y with respect to x is

$$\frac{d}{dx}\{f[g(x)]\} = f'[g(x)]g'(x)$$

which can also be written as

$$\frac{dy}{dx} = \frac{dy}{dg(x)}\frac{dg(x)}{dx}$$

(Note: the above does not imply that the $dg(x)$'s cancel out, as in fractions. They are part of the derivative notation and you can't separate them out or cancel them.)

The chain rule can be thought of as the derivative of the "outside" times the derivative of the "inside", remembering that the derivative of the outside function is evaluated at the value of the inside function.

Generalized Power Rule: If $y = [g(x)]^k$, then $dy/dx = k[g(x)]^{k-1}g'(x)$.

Examples:

1. **Find dy/dx for $y = (3x^2 + 5x - 7)^6$.** Let $f(z) = z^6$ and $z = g(x) = 3x^2 + 5x - 7$. **Then,** $y = f[g(x)]$ **and**

$$\frac{dy}{dx} \quad =$$

2. **Find dy/dx for $y = \sin(x^3 + 4x)$.** (Note: the derivative of $\sin x$ is $\cos x$.) **Let $f(z) = \sin z$ and $z = g(x) = x^3 + 4x$. Then, $y = f[g(x)]$ and**

$$\frac{dy}{dx} \quad =$$

4.7 Derivatives of Euler's number and natural logs

Derivatives of Exp or e:

1. $\frac{d}{dx} e^x = e^x$

2. $\frac{d}{dx} \alpha e^x = \alpha e^x$

3. $\frac{d^n}{dx^n} \alpha e^x = \alpha e^x$

4. $\frac{d}{dx} e^{u(x)} = e^{u(x)} u'(x)$

Examples: Find dy/dx for

1. $y = e^{-3x}$

2. $y = e^{x^2}$

3. $y = e^{\sin 2x}$

Derivatives of $\ln$:

1. $\frac{d}{dx} \ln x = \frac{1}{x}$

2. $\frac{d}{dx} \ln x^k = \frac{d}{dx} k \ln x = \frac{k}{x}$

3. $\frac{d}{dx} \ln u(x) = \frac{u'(x)}{u(x)}$ **(by the chain rule)**

4. **For any positive base b, $\frac{d}{dx} b^x = (\ln b)(b^x)$.**

Examples: Find dy/dx for

1. $y = \ln(x^2 + 9)$

2. $y = \ln(\ln x)$

3. $y = (\ln x)^2$

4. $y = \ln e^x$

4.8 Applications of the Derivative: Maxima and Minima

The first derivative, $f'(x)$, identifies whether the function $f(x)$ at the point x is increasing or decreasing at x.

1. Increasing: $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $f'(x) > 0$

2. Decreasing: $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $f'(x) < 0$

3. Neither increasing nor decreasing: $\qquad\qquad\qquad\qquad$ $f'(x) = 0$
   i.e. a maximum, minimum, or saddle point

Examples:

The second derivative $f''(x)$ identifies whether the function $f(x)$ at the point x is

1. Concave down: $\qquad\qquad$ $f''(x) < 0$

2. Concave up: $\qquad\qquad$ $f''(x) > 0$

Maximum (Minimum): x_0 is a local maximum (minimum) if $f(x_0) > f(x)$ $(f(x_0) < f(x))$ for all x within some open interval containing x_0. x_0 is a ____ global maximum (minimum)____ if $f(x_0) > f(x)$ $(f(x_0) < f(x))$ for all x in the domain of f.

Critical points: Given the function f **defined over domain** D, **all of the following are defined as critical points:**

1. Any interior point of D where $f'(x) = 0$.
2. Any interior point of D where $f'(x)$ does not exist.
3. Any endpoint that is in D.

The maxima and minima will be a subset of the critical points.

Second Derivative Test of Maxima/Minima: We can use the second derivative to tell us whether a point is a maximum or minimum of $f(x)$.

Local Maximum: $\qquad\qquad\qquad$ $f'(x) = 0$ and $f''(x) < 0$

Local Minimum: $\qquad\qquad\qquad$ $f'(x) = 0$ and $f''(x) > 0$

Need more info: $\qquad\qquad\qquad$ $f'(x) = 0$ and $f''(x) = 0$

Global Maxima and Minima. Sometimes no global max or min exists — e.g., $f(x)$ not bounded above or below. However, there are three situations where we can fairly easily identify global max or min.

1. **Functions with only one critical point.** If x_0 is a local max or min of f and it is the only critical point, then it is the global max or min.

2. **Globally concave up or concave down functions.** If $f''(x)$ is never zero, then there is at most one critical point. That critical point is a global maximum if $f'' < 0$ and a global minimum if $f'' > 0$.

3. **Functions over closed and bounded intervals** must have both a global maximum and a global minimum.

Examples: Find any critical points and identify whether they're a max, min, or saddle point:

1. $f(x) = x^2 + 2$

2. $f(x) = x^3 + 2$

3. $f(x) = |x^2 - 1|$, $x \in [-2, 2]$

4.9   Partial Derivatives

Suppose we have a function f now of two (or more) variables and we want to determine the rate of change relative to one of the variables. To do so, we would find it's partial derivative, which is defined similar to the derivative of a function of one variable.

Partial Derivative: Let f be a function of the variables $(x_1, \ldots, x_n)$. The partial derivative of f with respect to x_i is

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}$$

Only the ith variable changes — the others are treated as constants.

We can take higher-order partial derivatives, like we did with functions of a single variable, except now we the higher-order partials can be with respect to multiple variables.

Examples:

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

2. $f(x, y) = x^3 y^4 + e^x - \ln y$
   $\frac{\partial f}{\partial x}(x, y) =$
   $\frac{\partial f}{\partial y}(x, y) =$
   $\frac{\partial^2 f}{\partial x^2}(x, y) =$
   $\frac{\partial^2 f}{\partial x \partial y}(x, y) =$

4.10   L'H^opital's Rule

In studying limits, we saw that $\lim_{x \to c} f(x)/g(x) = \left(\lim_{x \to c} f(x) \right) / \left(\lim_{x \to c} g(x) \right)$, provided that $\lim_{x \to c} g(x) \neq 0$.

If both $\lim_{x \to c} f(x) = 0$ and $\lim_{x \to c} g(x) = 0$, then we get an indeterminate form of the type $0/0$ as $x \to c$. However, a limit may still exist. We can use L'H^opital's rule to find the limit.

L'H^opital's Rule: Suppose f and g are differentiable on some interval $a < x < b$ and that either

1. $\lim_{x \to a^+} f(x) = 0$ and $\lim_{x \to a^+} g(x) = 0$, or

2. $\lim_{x \to a^+} f(x) = \pm\infty$ and $\lim_{x \to a^+} g(x) = \pm\infty$

Suppose further that $g'(x)$ is never zero on $a < x < b$ and that

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

then

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

%**Put more simply, if** $\lim\limits_{x \to a f(x)}$ **is of the form** $0/0$ **or** $\pm\infty/\pm\infty$**, then**

$$\lim_{x \to a} f(x) = \lim_{x \to a} f'(x)$$

\

And if $\lim\limits_{x \to a} \frac{f'(x)}{g'(x)} = 0/0$ **or** $\pm\infty/\pm\infty$ **then you can apply L'H^opital's rule a second time, and continue applying it until you have a solution.**

Examples: Use L'Hôpital's rule to find the following limits:

1. $\lim\limits_{x \to 0^+} \frac{\ln(1+x^2)}{x^3}$

2. $\lim\limits_{x \to 0^+} \frac{e^{1/x}}{1/x}$

3. $\lim\limits_{x \to 2} \frac{x-2}{(x+6)^{1/3}-2}$

4.11   Taylor Series Approximation

Taylor series (also known as the delta method) are used commonly to represent functions as infinite series of the function's derivatives at some point a**. For example, Taylor series are very helpful in representing** *nonlinear* **functions as linear functions. One can thus** *approximate* **functions by using lower-order, finite series known as Taylor polynomials. If** $a = 0$**, the series is called a Maclaurin series.**

Specifically, a Taylor series of a real or complex function $f(x)$ **that is infinitely differentiable in the neighborhood of point** a **is:**

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$
$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

Taylor Approximation: We can often approximate the curvature of a function $f(x)$ **at point** a **using a 2nd order Taylor polynomial around point** a**:**

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + R_2$$

R_2 **is the Lagrange remainder and often treated as negligible, giving us:**

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2$$

Taylor series expansion is easily generalized to multiple dimensions.\

Curvature and The Taylor Polynomial as a Quadratic Form: The Hessian is used in a Taylor polynomial approximation to $f(\mathbf{x})$ and provides information about the curvature of $f(\mathbf{x})$ at x — e.g., which tells us whether a critical point x^* is a min, max, or saddle point.

1. The second order Taylor polynomial about the critical point $\mathbf{x}^*$ is

$$f(\mathbf{x}^* + \mathbf{h}) = \mathbf{f}(\mathbf{x}^*) + \nabla \mathbf{f}(\mathbf{x}^*)\mathbf{h} + \frac{1}{2}\mathbf{h}^\mathbf{T}\mathbf{H}(\mathbf{x}^*)\mathbf{h} + \mathbf{R}(\mathbf{h})$$

2. Since we're looking at a critical point, $\nabla f(\mathbf{x}^*) = 0$; and for small h, $R(\mathbf{h})$ is negligible. Rearranging, we get

$$f(\mathbf{x}^* + \mathbf{h}) - \mathbf{f}(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{h}^\mathbf{T}\mathbf{H}(\mathbf{x}^*)\mathbf{h}$$

3. The RHS is a quadratic form and we can determine the definiteness of $\mathbf{H}(\mathbf{x}^*)$.

4.12 Summary: Derivative calculus in 6 steps

With these six rules (and decent algebra and trigonometry skills) you can figure out the derivative of anything.

1. **Sum rule:**
$$[f(x) \pm g(x)]' = f'(x) \pm g'(x)$$

2. **Product rule:**
$$[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$$

3. **Power rule:**
$$[x^k]' = kx^{k-1}$$

4. **Chain rule:**
$$\frac{d}{dx}\{f[g(x)]\} = f'[g(x)]g'(x)$$

5. e^x:
$$\frac{d}{dx}e^x = e^x$$

6. **Trig identity:**
$$\frac{d}{dx}\sin(x) = \cos(x)$$

Chapter 5

Calculus II

Today's Topics: **The Indefinite Integral: The Antiderivative; Common Rules of Integration; The Definite Integral: The Area under the Curve; Integration by Substitution; Integration by Parts**

Much of the material and examples for this lecture are taken from Simon & Blume (1994) *Mathematics for Economists* **and from Boyce & Diprima (1988)** *Calculus*

5.1   The Indefinite Integral: The Antiderivative

So far, we've been interested in finding the derivative $g = f'$ of a function f. However, sometimes we're interested in exactly the reverse: finding the function f for which g is its derivative. We refer to f as the {antiderivative} of g.

Let DF be the derivative of F. And let $DF(x)$ be the derivative of F evaluated at x. Then the antiderivative is denoted by D^{-1} (i.e., the inverse derivative). If $DF = f$, then $F = D^{-1}f$.

Indefinite Integral: Equivalently, if F is the antiderivative of f, then F is also called the indefinite integral of f and written $F(x) = \int f(x)dx$.

Examples:	1. $\int \frac{1}{x^2}dx =$ 2. $\int 3e^{3x}dx =$ 3. $\int (x^2 - 4)dx =$

Notice from these examples that while there is only a single derivative for any function, there are multiple antiderivatives: one for any arbitrary constant c. c just shifts the curve up or down on the y-axis. If more info is present about the antiderivative — e.g., that it passes through a particular point — then we can solve for a specific value of c. \end{itemize}

5.2   Common Rules of Integration

- $\int af(x)dx = a\int f(x)dx$

- $\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$

- $\int x^n dx = \frac{1}{n+1}x^{n+1} + c$

- $\int e^x dx = e^x + c$

- $\int \frac{1}{x}dx = \ln x + c$

- $\int e^{f(x)} f'(x) dx = e^{f(x)} + c$

- $\int [f(x)]^n f'(x) dx = \frac{1}{n+1} [f(x)]^{n+1} + c$

- $\int \frac{f'(x)}{f(x)} dx = \ln f(x) + c$

Examples:

1. $\int 3x^2 dx =$

2. $\int (2x + 1) dx =$

3. $\int e^x e^{e^x} dx =$

5.3   The Definite Integral: The Area under the Curve

Riemann Sum: Suppose we want to determine the area $A(R)$ of a region R defined by a curve $f(x)$ and some interval $a \leq x \leq b$. One way to calculate the area would be to divide the interval $a \leq x \leq b$ into n subintervals of length Δx and then approximate the region with a series of rectangles, where the base of each rectangle is Δx and the height is $f(x)$ at the midpoint of that interval. $A(R)$ would then be approximated by the area of the union of the rectangles, which is given by

$$S(f, \Delta x) = \sum_{i=1}^{n} f(x_i) \Delta x$$

and is called a Riemann sum.

As we decrease the size of the subintervals Δx, making the rectangles "thinner," we would expect our approximation of the area of the region to become closer to the true area. This gives the limiting process

$$A(R) = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i) \Delta x$$

Riemann Integral: If for a given function f the Riemann sum approaches a limit as $\Delta x \to 0$, then that limit is called the Riemann integral of f from a to b. Formally,

$$\int_a^b f(x) dx = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i) \Delta x$$

Definite Integral: We use the notation $\int_a^b f(x) dx$ to denote the definite integral of f from a to b.

In words, the definite integral $\int_a^b f(x) dx$ is the area under the "curve" f(x) from $x = a$ to $x = b$.

First Fundamental Theorem of Calculus: Let the function f be bounded on $[a, b]$ and continuous on (a, b). Then the function

$$F(x) = \int_a^x f(t) dt, \quad a \leq x \leq b$$

has a derivative at each point in (a, b) and

$$F'(x) = f(x), \quad a < x < b$$

This last point shows that differentiation is the inverse of integration.

Second Fundamental Theorem of Calculus: Let the function f be bounded on $[a, b]$ and continuous on (a, b). Let F be any function that is continuous on $[a, b]$ such that $F'(x) = f(x)$ on (a, b). Then

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

Procedure to calculate a "simple" definite integral $\int_a^b f(x)dx$:

1. Find the indefinite integral $F(x)$.
2. Evaluate $F(b) - F(a)$.

Examples:

1. $\int_1^3 3x^2 dx =$

2. $\int_{-2}^2 e^x e^{e^x} dx =$

Properties of Definite Integrals:

1. $\int_a^a f(x)dx = 0$ **There is no area below a point.**

2. $\int_a^b f(x)dx = -\int_b^a f(x)dx$ **Reversing the limits changes the sign of the integral.**

3. $\int_a^b [\alpha f(x) + \beta g(x)]dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx$

4. $\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$

Examples:

1. $\int_1^1 3x^2 dx =$

2. $\int_0^4 (2x + 1)dx =$

3. $\int_{-2}^0 e^x e^{e^x} dx + \int_0^2 e^x e^{e^x} dx =$

5.4   Integration by Substitution

Sometimes the integrand doesn't appear integrable using common rules and antiderivatives. A method one might try is integration by substitution, which is related to the Chain Rule.

Suppose we want to find the indefinite integral $\int g(x)dx$ and assume we can identify a function $u(x)$ such that $g(x) = f[u(x)]u'(x)$. Let's refer to the antiderivative of f as F. Then the chain rule tells us that $\frac{d}{dx}F[u(x)] = f[u(x)]u'(x)$. So, $F[u(x)]$ is the antiderivative of g. We can then write

$$\int g(x)dx = \int f[u(x)]u'(x)dx = \int \frac{d}{dx}F[u(x)]dx = F[u(x)] + c$$

Procedure to determine the indefinite integral $\int g(x)dx$ by the method of substitution:

1. **Identify some part of $g(x)$ that might be simplified by substituting in a single variable u (which will then be a function of x).**
2. **Determine if $g(x)dx$ can be reformulated in terms of u and du.**
3. **Solve the indefinite integral.**
4. **Substitute back in for x**

Substitution can also be used to calculate a definite integral. Using the same procedure as above,

$$\int\limits_a^b g(x)dx = \int\limits_c^d f(u)du = F(d) - F(c)$$

where $c = u(a)$ and $d = u(b)$.

Example: $\int x^2\sqrt{x+1}dx$

The problem here is the $\sqrt{x+1}$ term. However, if the integrand had $\sqrt{x}$ times some polynomial, then we'd be in business. Let's try $u = x + 1$. Then $x = u - 1$ and $dx = du$. Substituting these into the above equation, we get

$$
\begin{aligned}
\int x^2\sqrt{x+1}dx &= \int (u-1)^2\sqrt{u}du \\
&= \int (u^2 - 2u + 1)u^{1/2}du \\
&= \int (u^{5/2} - 2u^{3/2} + u^{1/2})du
\end{aligned}
$$

We can easily integrate this, since it's just a polynomial. Doing so and substituting $u = x + 1$ back in, we get

$$\int x^2\sqrt{x+1}dx = 2(x+1)^{3/2}\left[\frac{1}{7}(x+1)^2 - \frac{2}{5}(x+1) + \frac{1}{3}\right] + c$$

For the above problem, we could have also used the substitution $u = \sqrt{x+1}$. Then $x = u^2 - 1$ and $dx = 2udu$. Substituting these in, we get

$$\int x^2\sqrt{x+1}dx = \int (u^2 - 1)^2 u2udu$$

which when expanded is again a polynomial and gives the same result as above.

Another Example $\int\limits_0^1 \frac{5e^{2x}}{(1+e^{2x})^{1/3}}dx$

When an expression is raised to a power, it's often helpful to use this expression as the basis for a substitution. So, let $u = 1 + e^{2x}$. Then $du = 2e^{2x}dx$ and we can set $5e^{2x}dx = 5du/2$. Additionally, $u = 2$ when $x = 0$ and $u = 1 + e^2$ when $x = 1$. Substituting all of this in, we get

$$\int_0^1 \frac{5e^{2x}}{(1+e^{2x})^{1/3}}dx \;=\; \frac{5}{2}\int_2^{1+e^2}\frac{du}{u^{1/3}}$$

$$=\; \frac{5}{2}\int_2^{1+e^2} u^{-1/3}du$$

$$=\; \frac{15}{4}u^{2/3}\Big|_2^{1+e^2}$$

$$=\; 9.53$$

5.5 Integration by Parts, or Ultraviolet Voodoo

Another useful integration technique is {integration by parts}, which is related to the Product Rule of differentiation. The product rule states that

$$\frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

Integrating this and rearranging, we get

$$\int u\frac{dv}{dx}dx = uv - \int v\frac{du}{dx}dx$$

or

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx$$

More frequently remembered as

$$\int u\,dv = uv - \int v\,du$$

where $du = u'(x)dx$ and $dv = v'(x)dx$.

For definite integrals: $\int_a^b u\frac{dv}{dx}dx = uv\big|_a^b - \int_a^b v\frac{du}{dx}dx$

Our goal here is to find expressions for u and dv that, when substituted into the above equation, yield an expression that's more easily evaluated.

Examples:

1. $\int xe^{ax}dx$

2. $\int x^n e^{ax}dx$

3. $\int x^3 e^{-x^2}dx$

Chapter 6

Optimization

Topics:

• **Quadratic Forms** • **Definiteness of Quadratic Forms** • **Maxima and Minima in $\mathbf{R}^n$** • **First Order Conditions** • **Second Order Conditions** • **Global Maxima and Minima** • **Constrained Optimization** • **Equality Constraints** • **Inequality Constraints** • **Kuhn-Tucker Conditions**

Much of the material and examples for this lecture are taken from Simon & Blume (1994) *Mathematics for Economists* and Ecker & Kupferschmid (1988) *Introduction to Operations Research*

6.1   Quadratic Forms

Quadratic forms important because 1. Approximates local curvature around a point — e.g., used to identify max vs min vs saddle point. 2. Simple, so easy to deal with. 3. Have a matrix representation.

Quadratic Form: A polynomial where each term is a monomial of degree 2 in any number of variables:

$$\textbf{One variable: } Q(x_1) = a_{11}x_1^2$$
$$\textbf{Two variables: } Q(x_1, x_2) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2$$
$$\textbf{N variables: } Q(x_1, \cdots, x_n) = \sum_{i \leq j} a_{ij}x_ix_j$$

which can be written in matrix terms:

$$\textbf{One variable: } Q(\mathbf{x}) = x_1^T a_{11} x_1$$

$$\textbf{N variables: } Q(\mathbf{x}) = \begin{pmatrix} x1 & x2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} & \cdots & \frac{1}{2}a_{1n} \\ \frac{1}{2}a_{12} & a_{22} & \cdots & \frac{1}{2}a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}a_{1n} & \frac{1}{2}a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \mathbf{x}^T \mathbf{A} \mathbf{x}$$

Examples:

1. **Quadratic on $\mathbf{R}^2$:**

$$\begin{aligned} Q(x_1, x_2) &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & \frac{1}{2}a_{12} \\ \frac{1}{2}a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{22}x_2^2 \end{aligned}$$

2. **Quadratic on $\mathbf{R}^3$:** $Q(x_1, x_2, x_3) =$

6.2   Concavity of Quadratic Forms

Concavity helps identify the curvature of a function, $f(x)$, in **2** dimensional space.

• **Concave: A function f is strictly concave over the set S if** $\forall x_1, x_2 \in S$ **and** $\forall a \in (0, 1)$,

$$f(ax_1 + (1 - a)x_2) > af(x_1) + (1 - a)f(x_2)$$

Any line connecting two points on a concave function will lie *below* the function.

• **Convex: A function f is strictly convex over the set S if** $\forall x_1, x_2 \in S$ **and** $\forall a \in (0, 1)$,

$$f(ax_1 + (1 - a)x_2) < af(x_1) + (1 - a)f(x_2)$$

Any line connecting two points on a convex function will lie *above* the function.

• **Quasiconcave: A function f is quasiconcave over the set S if** $\forall x_1, x_2 \in S$ **and** $\forall a \in (0, 1)$,

$$f(ax_1 + (1 - a)x_2) \geq \min(f(x_1), f(x_2))$$

No matter what two points you select, the *lowest* valued point will always be an end point.

• **Quasiconvex: A function f is quasiconcave over the set S if** $\forall x_1, x_2 \in S$ **and** $\forall a \in (0, 1)$,

$$f(ax_1 + (1 - a)x_2) \leq \max(f(x_1), f(x_2))$$

No matter what two points you select, the *highest* valued point will always be an end point.

• **Quasilinear: A function that is both quasiconvex and quasiconcave is quasilinear.**

Second Derivative Test of Concavity: The second derivative can be used to understand concavity.

If

$$\begin{aligned} f''(x) < 0 &\Rightarrow \textbf{Concave} \\ f''(x) > 0 &\Rightarrow \textbf{Convex} \end{aligned}$$

6.3 Definiteness of Quadratic Forms

Definiteness helps identify the curvature of a function, $Q(\mathbf{x})$, in n dimensional space.

Definiteness: By definition, a quadratic form always takes on the value of zero when $x = 0$, $Q(\mathbf{x}) = 0$ at $\mathbf{x} = 0$. The definiteness of the matrix $\mathbf{A}$ is determined by whether the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$ is greater than zero, less than zero, or sometimes both over all $\mathbf{x} \neq 0$.

1. **Positive Definite** $\mathbf{x}^T \mathbf{A}\mathbf{x} > 0, \quad \forall \mathbf{x} \neq 0$ Min

    ex: $Q(x_1, x_2) = x_1^2 + x_2^2$

2. **Positive Semidefinite** $\mathbf{x}^T \mathbf{A}\mathbf{x} \geq 0, \quad \forall \mathbf{x} \neq 0$

    ex: $Q(x_1, x_2) = x_1^2 + 2x_1 x_2 + x_2^2$

3. **Negative Definite** $\mathbf{x}^T \mathbf{A}\mathbf{x} < 0, \quad \forall \mathbf{x} \neq 0$ Max

    ex: $Q(x_1, x_2) = -x_1^2 - x_2^2$

4. **Negative Semidefinite** $\mathbf{x}^T \mathbf{A}\mathbf{x} \leq 0, \quad \forall \mathbf{x} \neq 0$

    ex: $Q(x_1, x_2) = -x_1^2 - 2x_1 x_2 - x_2^2$

5. **Indefinite** $\mathbf{x}^T \mathbf{A}\mathbf{x} > 0$ **for some x $\neq 0$ and**   Neither
    $\mathbf{x}^T \mathbf{A}\mathbf{x} < 0$ **for other x $\neq 0$**

    ex: $Q(x_1, x_2) = x_1^2 - x_2^2$

Examples:

1. **Positive Definite:**
$$Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{x}$$
$$= x_1^2 + x_2^2$$

2. **Positive Semidefinite:**

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \mathbf{x}$$
$$= (x_1 - x_2)^2$$

3. **Indefinite:**

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \mathbf{x}$$
$$= x_1^2 - x_2^2$$

6.4 First Order Conditions

When we examined functions of one variable x, we found critical points by taking the first derivative, setting it to zero, and solving for x. For functions of n variables, the critical points are found in much the same way, except now we set the partial derivatives equal to zero.[1]

Gradient ($\nabla f(\mathbf{x})$): Given a function $f(\mathbf{x})$ in n variables, the gradient $\nabla f(\mathbf{x})$ is a column vector, where the ith element is the partial derivative of $f(\mathbf{x})$ with respect to x_i:

---

[1] We will only consider critical points on the interior of a function's domain.

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

Critical Point: $\mathbf{x}^*$ is a critical point iff $\nabla f(\mathbf{x}^*) = 0$. If the partial derivative of f(x) with respect to x^* is 0, then $\mathbf{x}^*$ is a critical point. To solve for $\mathbf{x}^*$, find the gradient, set each element equal to 0, and solve the system of equations.

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix}$$

Example: Given a function $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$, find the:

1. Gradient $\nabla f(\mathbf{x}) =$

2. Critical Point $x^* =$

\end{itemize}

6.5   Second Order Conditions

When we found a critical point for a function of one variable, we used the second derivative as an indicator of the curvature at the point in order to determine whether the point was a min, max, or saddle (second derivative test of concavity). For functions of n variables, we use *second order partial derivatives* as an indicator of curvature.

Hessian (H(x)): Given a function $f(\mathbf{x})$ in n variables, the hessian $\mathbf{H}(\mathbf{x})$ is an $n \times n$ matrix, where the (i, j)th element is the second order partial derivative of $f(\mathbf{x})$ with respect to x_i and x_j:

$$\mathbf{H}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}$$

Note that the hessian will be a symmetric matrix because $\frac{\partial f(\mathbf{x})}{\partial x_1 \partial x_2} = \frac{\partial f(\mathbf{x})}{\partial x_2 \partial x_1}$.

Also note that given that $f(\mathbf{x})$ is of quadratic form, each element of the hessian will be a constant.

Second Order Conditions:\[6pt] Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = 0$,

1. **H(x*) Positive Definite** $\implies$ **Strict Local Min**

2. **H(x) Positive Semidefinite** $\implies$ **Local Min**
   $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$

3. **H(x*) Negative Definite** $\implies$ **Strict Local Max**

4. **H(x) Negative Semidefinite** $\implies$ **Local Max**
   $\forall \mathbf{x} \in B(\mathbf{x}^*, \epsilon)$

5. **H(x*) Indefinite** $\implies$
   Saddle Point

Example: We found that the only critical point of $f(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 + 1$ is at $\mathbf{x}^* = (1, 0)$. Is it a min, max, or saddle point?

1. **Hessian H(x) =**

2. **Leading principal minors of H(x) =**

3. **Definiteness?**

4. **Maxima, Minima, or Saddle Point?**

5. **Alternate check of definiteness. Is $\mathbf{H(x^*)} \gtrless 0 \quad \forall \quad \mathbf{x} \neq 0$?**

6.6   Definiteness and Concavity

Although definiteness helps us to understand the curvature of an n-dimensional function, it does not necessarily tell us whether the function is globally concave or convex.

We need to know whether a function is globally concave or convex to determine whether a critical point is a global min or max.

Testing for Global Concavity: We can use the definiteness of the Hessian to determine whether a function is globally concave or convex:

1. **H(x) Positive Semidefinite $\forall \mathbf{x}$** $\implies$ **Globally Convex**

2. **H(x) Negative Semidefinite $\forall \mathbf{x}$** $\implies$ **Globally Concave**

Notice that the definiteness conditions must be satisfied over the entire domain.

6.7 Global Maxima and Minima

Global Max/Min Conditions: Given a function $f(\mathbf{x})$ and a point $\mathbf{x}^*$ such that $\nabla f(\mathbf{x}^*) = 0$,

    1. $f(\mathbf{x})$ **Globally Convex** $\implies$ **Global Min**

    2. $f(\mathbf{x})$ **Globally Concave** $\implies$ **Global Max**

Note that showing that $\mathbf{H}(\mathbf{x}^*)$ is negative semidefinite is not enough to guarantee $\mathbf{x}^*$ is a local max. However, showing that $\mathbf{H}(\mathbf{x})$ is negative semidefinite for all $\mathbf{x}$ guarantees that x^* is a global max. (The same goes for positive semidefinite and minima.)

Example: Take $f_1(x) = x^4$ and $f_2(x) = -x^4$. Both have $x = 0$ as a critical point. Unfortunately, $f_1''(0) = 0$ and $f_2''(0) = 0$, so we can't tell whether $x = 0$ is a min or max for either. However, $f_1'''(x) = 12x^2$ and $f_2'''(x) = -12x^2$. For all x, $f_1''(x) \geq 0$ and $f_2''(x) \leq 0$ — i.e., $f_1(x)$ is globally convex and $f_2(x)$ is globally concave. So $x = 0$ is a global min of $f_1(x)$ and a global max of $f_2(x)$.

Example

Given $f(\mathbf{x}) = x_1^3 - x_2^3 + 9x_1 x_2$, find any maxima or minima.

    1. **First order conditions.**

        (a) **Gradient $\nabla f(\mathbf{x}) =$**

        (b) **Critical Points $\mathbf{x}^* =$**

    2. **Second order conditions.**

        (a) **Hessian $\mathbf{H}(\mathbf{x}) =$**

        (b) **Hessian $\mathbf{H}(\mathbf{x}_1^*) =$**

        (c) **Leading principal minors of $\mathbf{H}(\mathbf{x}_1^*) =$**

        (d) **Definiteness of $\mathbf{H}(\mathbf{x}_1^*)$?**

    (e) **Maxima, Minima, or Saddle Point for x_1^*?**

    (f) **Hessian $H(x_2^*) =$**

    (g) **Leading principal minors of $H(x_2^*) =$**

    (h) **Definiteness of $H(x_2^*)$?**

    (i) **Maxima, Minima, or Saddle Point for x_2^*?**

3. **Global concavity/convexity.**

    (a) **Is f(x) globally concave/convex?**

    (b) **Are any x^* global minima or maxima?**

6.8 Constrained Optimization

We have already looked at optimizing a function in one or more dimensions over the whole domain of the function. Often, however, we want to find the maximum or minimum of a function over some restricted part of its domain.\ ex: Maximizing utility subject to a budget constraint

Types of Constraints: For a function $f(x_1, \ldots, x_n)$, there are two types of constraints that can be imposed:

1. Equality constraints: constraints of the form $c(x_1, \ldots, x_n) = r$. Budget constraints are the classic example of equality constraints in social science.

2. Inequality constraints: constraints of the form $c(x_1, \ldots, x_n) \leq r$. These might arise from non-negativity constraints or other threshold effects.

In any constrained optimization problem, the constrained maximum will always be less than or equal to the unconstrained maximum. If the constrained maximum is less than the unconstrained maximum, then the constraint is binding. Essentially, this means that you can treat your constraint as an equality constraint rather than an inequality constraint.

For example, the budget constraint binds when you spend your entire budget. This generally happens because we believe that utility is strictly increasing in consumption, i.e. you always want more so you spend everything you have.

Any number of constraints can be placed on an optimization problem. When working with multiple constraints, always make sure that the set of constraints are not pathological; it must be possible for all of the constraints to be satisfied simultaneously.

Set-up for Constrained Optimization:

$$\max_{x_1, x_2} f(x_1, x_2) \ \textbf{s.t.} \ c(x_1, x_2)$$

$$\min_{x_1, x_2} f(x_1, x_2) \ \textbf{s.t.} \ c(x_1, x_2)$$

This tells us to maximize/minimize our function, $f(x_1, x_2)$, with respect to the choice variables, x_1, x_2, subject to the constraint.

Example:

$$\max_{x_1, x_2} f(x_1, x_2) = -(x_1^2 + 2x_2^2) \ \textbf{s.t.} \ x_1 + x_2 = 4$$

It is easy to see that the *unconstrained* maximum occurs at $(x_1, x_2) = (0, 0)$, but that does not satisfy the constraint. How should we proceed?

6.9   Equality Constraints

Equality constraints are the easiest to deal with because we know that the maximum or minimum has to lie on the (intersection of the) constraint(s).

The trick is to change the problem from a constrained optimization problem in n variables to an unconstrained optimization problem in $n+k$ variables, adding *one* variable for *each* equality constraint. We do this using a lagrangian multiplier.

Lagrangian function: The Lagrangian function allows us to combine the function we want to optimize and the constraint function into a single function. Once we have this single function, we can proceed as if this were an *unconstrained* optimization problem.\

For each constraint, we must include a Lagrange multiplier (λ_i) as an additional variable in the analysis. These terms are the link between the constraint and the Lagrangian function.\

Given a *two dimensional* set-up:

$$\max_{x_1, x_2} / \min_{x_1, x_2} f(x_1, x_2) \ \textbf{s.t.} \ c(x_1, x_2) = a$$

We define the Lagrangian function $L(x_1, x_2, \lambda_1)$ as follows:

$$L(x_1, x_2, \lambda_1) = f(x_1, x_2) - \lambda_1(c(x_1, x_2) - a)$$

More generally, in *n dimensions*:

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k) = f(x_1, \ldots, x_n) - \sum_{i=1}^{k} \lambda_i(c_i(x_1, \ldots, x_n) - r_i)$$

Getting the sign right:\ Note that above we subtract the lagrangian term *and* we subtract the constraint constant from the constraint function. Occasionally, you may see the following alternative form of the Lagrangian, which is *equivalent*:

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k) = f(x_1, \ldots, x_n) + \sum_{i=1}^{k} \lambda_i(r_i - c_i(x_1, \ldots, x_n))$$

Here we add the lagrangian term *and* we subtract the constraing function from the constraint constant.

Using the Lagrangian to Find the Critical Points: To find the critical points, we take the partial derivatives of lagrangian function, $L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k)$, with respect to each of its

variables (all choice variables x *and* all lagrangian multipliers λ). At a critical point, *each* of these partial derivatives must be equal to zero, so we obtain a system of $n + k$ *equations in* $n + k$ *unknowns*:

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \sum_{i=1}^{k} \lambda_i \frac{\partial c_i}{\partial x_1} \quad = \quad 0$$

$$\vdots \quad = \quad \vdots$$

$$\frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} - \sum_{i=1}^{k} \lambda_i \frac{\partial c_i}{\partial x_n} \quad = \quad 0$$

$$\frac{\partial L}{\partial \lambda_1} = c_1(x_i, \ldots, x_n) - r_1 \quad = \quad 0$$

$$\vdots \quad = \quad \vdots$$

$$\frac{\partial L}{\partial \lambda_k} = c_k(x_i, \ldots, x_n) - r_k \quad = \quad 0$$

We can then solve this system of equations, because there are $n + k$ equations and $n + k$ unknowns, to calculate the critical point $(x_1^*, \ldots, x_n^*, \lambda_1^*, \ldots, \lambda_k^*)$.

Second-order Conditions and Unconstrained Optimization: There may be more than one critical point, i.e. we need to verify that the critical point we find is a maximum/minimum. Similar to unconstrained optimization, we can do this by checking the second-order conditions.

Example:

$$\max_{x_1, x_2} f(x) = -(x_1^2 + 2x_2^2) \text{ s.t. } x_1 + x_2 = 4$$

1. **Begin by writing the Lagrangian:**

$$L(x_1, x_2, \lambda) =$$

2. **Take the partial derivatives and set equal to zero:**

$$\frac{\partial L}{\partial x_1} = \quad\quad\quad = \quad 0$$

$$\frac{\partial L}{\partial x_2} = \quad\quad\quad = \quad 0$$

$$\frac{\partial L}{\partial \lambda} = \quad\quad\quad = \quad 0$$

3. **Solve the system of equations:**

Notice that when we take the partial derivative of L with respect to the lagranigian multiplier and set it equal to 0, we return exactly our constraint! This is why signs matter.

6.10   Inequality Constraints

Inequality constraints define the boundary of a region over which we seek to optimize the function. This makes inequality constraints more challenging because we do not know if the

maximum/minimum lies along one of the constraints (the constraint binds) or in the interior of the region.

We must introduce more variables in order to turn the problem into an unconstrained optimization.

Slack: For *each* inequality constraint $c_i(x_1, \ldots, x_n) \leq a_i$, we define a slack variable s_i^2 for which the expression $c_i(x_1, \ldots, x_n) \leq a_i - s_i^2$ would hold with equality. These slack variables capture how close the constraint comes to binding. We use s^2 rather than s to ensure that the slack is positive.

Slack is just a way to transform our constraints.

Given a *two-dimensional* set-up and these edited constraints:

$$\max_{x_1,x_2} / \min_{x_1,x_2} f(x_1, x_2) \text{ s.t. } c(x_1, x_2) \leq a_1$$

Adding in Slack:

$$\max_{x_1,x_2} / \min_{x_1,x_2} f(x_1, x_2) \text{ s.t. } c(x_1, x_2) \leq a_1 - s_1^2$$

We define the Lagrangian function $L(x_1, x_2, \lambda_1, s_1)$ as follows:

$$L(x_1, x_2, \lambda_1, s_1) = f(x_1, x_2) - \lambda_1(c(x_1, x_2) + s_1^2 - a_1)$$

More generally, in *n dimensions*:

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k, s_1, \ldots, s_k) = f(x_1, \ldots, x_n) - \sum_{i=1}^{k} \lambda_i(c_i(x_1, \ldots, x_n) + s_i^2 - a_i)$$

Finding the Critical Points: To find the critical points, we take the partial derivatives of the lagrangian function, $L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k, s_1, \ldots, s_k)$, with respect to each of its variables (all choice variables x, all lagrangian multipliers λ, and all slack variables s). At a critical point, *each* of these partial derivatives must be equal to zero, so we obtain a system of $n+2k$ *equations in* $n + 2k$ *unknowns*:

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} - \sum_{i=1}^{k} \lambda_i \frac{\partial c_i}{\partial x_1} = 0$$

$$\vdots \quad = \quad \vdots$$

$$\frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} - \sum_{i=1}^{k} \lambda_i \frac{\partial c_i}{\partial x_n} = 0$$

$$\frac{\partial L}{\partial \lambda_1} = c_1(x_i, \ldots, x_n) + s_1^2 - b_1 = 0$$

$$\vdots \quad = \quad \vdots$$

$$\frac{\partial L}{\partial \lambda_k} = c_k(x_i, \ldots, x_n) + s_k^2 - b_k = 0$$

$$\frac{\partial L}{\partial s_1} = 2s_1\lambda_1 = 0$$

$$\vdots \quad = \quad \vdots$$

$$\frac{\partial L}{\partial s_k} = 2s_k\lambda_k = 0$$

Complementary slackness conditions: The last set of first order conditions of the form $2s_i\lambda_i = 0$ (the partials taken with respect to the slack variables) are known as complementary slackness conditions. These conditions can be satisfied one of three ways:

1. $\lambda_i = 0$ and $s_i \neq 0$: This implies that the slack is positive and thus *the constraint does not bind.*

2. $\lambda_i \neq 0$ and $s_i = 0$: This implies that there is no slack in the constraint and *the constraint does bind.*

3. $\lambda_i = 0$ and $s_i = 0$: In this case, there is no slack but the *constraint binds trivially,* without changing the optimum.

Example: Find the critical points for the following constrained optimization:

$$\max_{x_1, x_2} f(x) = -(x_1^2 + 2x_2^2) \text{ s.t. } x_1 + x_2 \leq 4$$

1. Rewrite with the slack variables:

$$\max_{x_1, x_2} f(x) = -(x_1^2 + 2x_2^2) \text{ s.t. } x_1 + x_2 \leq 4 - s_1^2$$

2. Write the Lagrangian:

$$L(x_1, x_2, \lambda_1, s_1) = -(x_1^2 + 2x_2^2) - \lambda_1(x_1 + x_2 + s_1^2 - 4)$$

3. Take the partial derivatives and set equal to 0:

$$\frac{\partial L}{\partial x_1} = -2x_1 - \lambda_1 = 0$$

$$\frac{\partial L}{\partial x_2} = -4x_2 - \lambda_1 = 0$$

$$\frac{\partial L}{\partial \lambda_1} = -(x_1 + x_2 + s_1^2 - 4) = 0$$

$$\frac{\partial L}{\partial s_1} = -2s_1\lambda_1 = 0$$

4. Consider all ways that the complementary slackness conditions are solved:

Hypothesis	s_1	λ_1	x_1	x_2	$f(x_1, x_2)$
$s_1 = 0\ \lambda_1 = 0$	No solution				
$s_1 \neq 0\ \lambda_1 = 0$	2	0	0	0	0
$s_1 = 0\ \lambda_1 \neq 0$	0	$\frac{-16}{3}$	$\frac{8}{3}$	$\frac{4}{3}$	$-\frac{32}{3}$
$s_1 \neq 0\ \lambda_1 \neq 0$	No solution				

This shows that there are two critical points: $(0,0)$ and $\left(\frac{8}{3}, \frac{4}{3}\right)$.

5. **Find maximum:** Looking at the values of $f(x_1, x_2)$ at the critical points, we see that $f(x_1, x_2)$ is maximized at $x_1^* = 0$ and $x_2^* = 0$.

Example: Find the critical points for the following constrained optimization:

$$\max_{x_1, x_2} f(x) = -(x_1^2 + 2x_2^2) \text{ s.t. } \begin{array}{l} x_1 + x_2 \leq 4 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{array}$$

1. **Rewrite with the slack variables:**

2. **Write the Lagrangian:**

3. **Take the partial derivatives and set equal to zero:**

4. **Consider all ways that the complementary slackness conditions are solved:**

Hypothesis	s_1	s_2	s_3	λ_1	λ_2	λ_3	x_1	x_2	$f(x_1, x_2)$
$s_1 = s_2 = s_3 = 0$									
$s_1 \neq 0, s_2 = s_3 = 0$									
$s_2 \neq 0, s_1 = s_3 = 0$									
$s_3 \neq 0, s_1 = s_2 = 0$									
$s_1 \neq 0, s_2 \neq 0, s_3 = 0$									
$s_1 \neq 0, s_3 \neq 0, s_2 = 0$									
$s_2 \neq 0, s_3 \neq 0, s_1 = 0$									
$s_1 \neq 0, s_2 \neq 0, s_3 \neq 0$									

5. **Find maximum:**

6.11   Kuhn-Tucker Conditions

As you can see, this can be a pain. When dealing explicitly with *non-negativity constraints*, this process is simplified by using the Kuhn-Tucker method.

Kuhn-Tucker conditions: Because the problem of maximizing a function subject to inequality and non-negativity constraints arises frequently in economics, the Kuhn-Tucker approach provides a method that often makes it easier to both calculate the critical points and identify points that are (local) maxima.\

Given a *two-dimensional set-up*:

$$\max_{x_1,x_2} / \min_{x_1,x_2} f(x_1, x_2) \text{ s.t.} \quad \begin{array}{c} c(x_1, x_2) \le a_1 \\ x_1 \ge 0 \\ gx_2 \ge 0 \end{array}$$

We define the Lagrangian function $L(x_1, x_2, \lambda_1)$ the same as if we did not have the non-negativity constraints:

$$L(x_1, x_2, \lambda_2) = f(x_1, x_2) - \lambda_1(c(x_1, x_2) - a_1)$$

More generally, in *n dimensions*:

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k) = f(x_1, \dots, x_n) - \sum_{i=1}^{k} \lambda_i(c_i(x_1, \dots, x_n) - a_i)$$

Kuhn-Tucker and Complementary Slackness Conditions: To find the critical points, we first calculate the Kuhn-Tucker conditions by taking the partial derivatives of the lagrangian function, $L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_k)$, with respect to each of its variables (all choice variable sx and all lagrangian multipliers λ) <u>and</u> we calculate the complementary slackness conditions by multiplying each partial derivative by its respective variable <u>and</u> include non-negativity conditions for all variables (choice variables x and lagrangian multipliers λ).

Kuhn-Tucker Conditions

$$\frac{\partial L}{\partial x_1} \le 0, \quad \dots, \quad \frac{\partial L}{\partial x_n} \le 0$$
$$\frac{\partial L}{\partial \lambda_1} \ge 0, \quad \dots, \quad \frac{\partial L}{\partial \lambda_m} \ge 0$$

Complementary Slackness Conditions

$$x_1 \frac{\partial L}{\partial x_1} = 0, \quad \dots, \quad x_n \frac{\partial L}{\partial x_n} = 0$$
$$\lambda_1 \frac{\partial L}{\partial \lambda_1} = 0, \quad \dots, \quad \lambda_m \frac{\partial L}{\partial \lambda_m} = 0$$

Non-negativity Conditions

$$x_1 \ge 0 \quad \dots \quad x_n \ge 0$$
$$\lambda_1 \ge 0 \quad \dots \quad \lambda_m \ge 0$$

Note that some of these conditions are set equal to 0, while others are set as inequalities!

Note also that to minimize the function $f(x_1, \dots, x_n)$, the simplest thing to do is maximize the function $-f(x_1, \dots, x_n)$; all of the conditions remain the same after reformulating as a maximization problem.

There are additional assumptions (notably, f(x) is quasi-concave and the constraints are convex) that are sufficient to ensure that a point satisfying the Kuhn-Tucker conditions is a global max; if these assumptions do not hold, you may have to check more than one point.

Finding the Critical Points with Kuhn-Tucker Conditions: Given the above conditions, to find the critical points we solve the above system of equations. To do so, we must check *all* border and interior solutions to see if they satisfy the above conditions.

In a two-dimensional set-up, this means we must check the following cases:

(1) $x_1 = 0, x_2 = 0$ **Border Solution**

(2) $x_1 = 0, x_2 \neq 0$ **Border Solution**

(3) $x_1 \neq 0, x_2 = 0$ **Border Solution**

(4) $x_1 \neq 0, x_2 \neq 0$ **Interior Solution**

Example:

$$\max_{x_1, x_2} f(x) = -(x_1^2 + 2x_2^2) \textbf{ s.t. } \begin{array}{l} x_1 + x_2 \leq 4 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{array}$$

1. **Write the Lagrangian:**

$$L(x_1, x_2, \lambda) = -(x_1^2 + 2x_2^2) - \lambda(x_1 + x_2 - 4)$$

2. **Find the First Order Conditions:**
   Kuhn-Tucker Conditions

$$\frac{\partial L}{\partial x_1} = -2x_1 - \lambda \ \leq \ 0$$

$$\frac{\partial L}{\partial x_2} = -4x_2 - \lambda \ \leq \ 0$$

$$\frac{\partial L}{\partial \lambda} = -(x_1 + x_2 - 4) \ \geq \ 0$$

   Complementary Slackness Conditions

$$x_1 \frac{\partial L}{\partial x_2} = x_1(-2x_1 - \lambda) = \ \ 0$$

$$x_2 \frac{\partial L}{\partial x_2} = x_2(-4x_2 - \lambda) \ = \ \ 0$$

$$\lambda \frac{\partial L}{\partial \lambda} = -\lambda(x_1 + x_2 - 4) \ = \ \ 0$$

   Non-negativity Conditions

$$x_1 \ \geq \ \ 0$$

$$x_2 \ \geq \ \ 0$$

$$\lambda \ \geq \ \ 0$$

3. **Consider all border and interior cases:**

Hypothesis	λ	x_1	x_2	$f(x_1, x_2)$
$x_1 = 0, x_2 = 0$	**0**	**0**	**0**	**0**
$x_1 = 0, x_2 \neq 0$	**-16**	**0**	**4**	**-32**
$x_1 \neq 0, x_2 = 0$	**-8**	**4**	**0**	**-16**
$x_1 \neq 0, x_2 \neq 0$	$-\frac{16}{3}$	$\frac{8}{3}$	$\frac{4}{3}$	$-\frac{32}{3}$

4. **Find Maximum:**
   Three of the critical points violate the requirement that $\lambda \geq 0$, so the point $(0, 0, 0)$ is the maximum.

Example:

$$\max_{x_1, x_2} f(x) = \frac{1}{3}\log(x_1 + 1) + \frac{2}{3}\log(x_2 + 1) \textbf{ s.t. } \begin{array}{l} x_1 + 2x_2 \leq 4 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{array}$$

1. **Write the Lagrangian:**

2. **Find the First Order Conditions:**
   Kuhn-Tucker Conditions

   Complementary Slackness Conditions

   Non-negativity Conditions

3. **Consider all border and interior cases:**

Hypothesis	λ	x_1	x_2	$f(x_1, x_2)$
$x_1 = 0, x_2 = 0$				
$x_1 = 0, x_2 \neq 0$				
$x_1 \neq 0, x_2 = 0$				
$x_1 \neq 0, x_2 \neq 0$				

4. **Find Maximum:**

Chapter 7

Introduction

You can label chapter and section titles using {#label} after them, e.g., we can reference Chapter 7. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter ??.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```r
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 7.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 7.1.

```r
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the bookdown package (Xie, 2018) in this sample book, which was built on top of R Markdown and knitr (Xie, 2015).
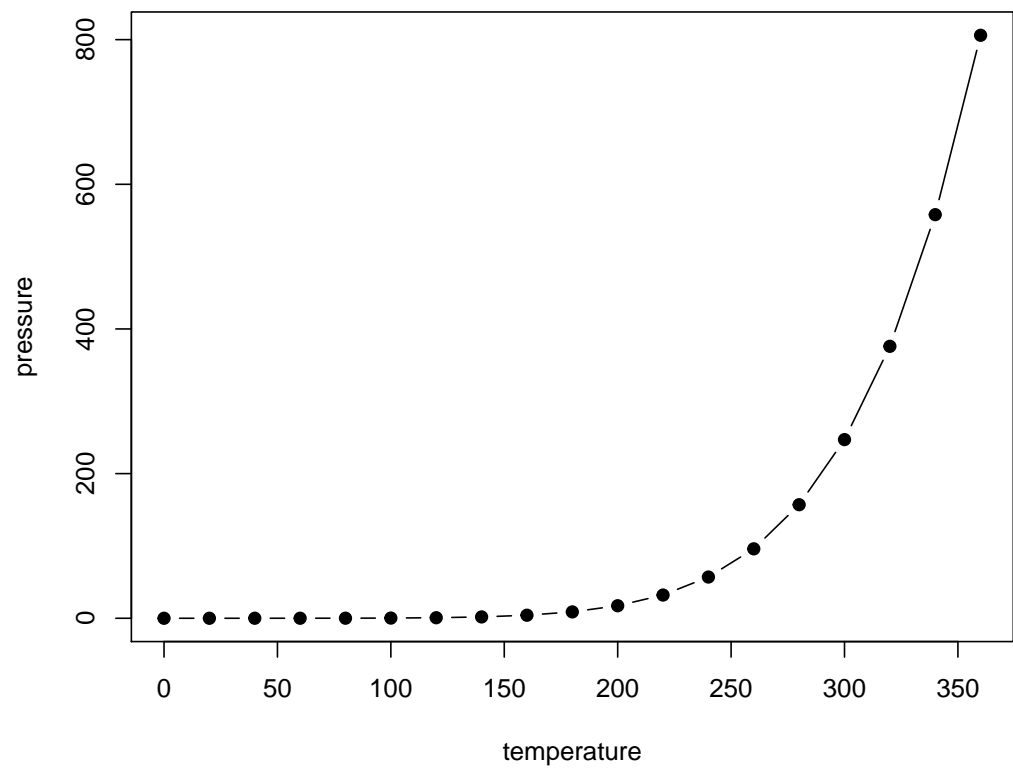
Figure 7.1: Here is a nice figure!

Table 7.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2018). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.7.