Math Camp 2019 - Linear Algebra*

Created by James Banovetz

Modified by Woongchan Jeon

Department of Economics, UC Santa Barbara

August 22, 2019

1. **Systems of Linear Equations and Matrices**

    (a) <u>Definition</u>. A **linear equation** in the variables $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form

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

    where b and the coefficients $a_1, \ldots, a_n$ are real or complex numbers. A **system of linear equations** is a collection of one or more linear equations involving the same variables.

    (b) <u>Example</u>. Consider the following system of linear equations:

    $$
    \begin{aligned}
    x_1 - 2x_2 + x_3 &= 0 \\
    2x_2 - 8x_3 &= 8 \\
    -4x_1 + 5x_2 + 9x_3 &= -9
    \end{aligned}
    $$

    (c) <u>Defintion</u>. The compact rectangular array containing the coefficients of each variable aligned in columns is called the **coefficient matrix**.

    (d) <u>Example</u>. From the system of linear equations above, the coefficient matrix is

    $$
    \begin{bmatrix}
    1 & -2 & 1 \\
    0 & 2 & -8 \\
    -4 & 5 & 9
    \end{bmatrix}
    $$

    (e) <u>Definition</u>. The coefficient matrix, concatenated with a column consisting of the right-hand-side constants is called the **augmented matrix**.

    (f) <u>Example</u>. From the system of linear equations above, the augmented matrix is

    $$
    \left[
    \begin{array}{ccc|c}
    1 & -2 & 1 & 0 \\
    0 & 2 & -8 & 8 \\
    -4 & 5 & 9 & -9
    \end{array}
    \right]
    $$

---

*These lecture notes are drawn principally from *Linear Algebra and its Applications*, 4th ed., by David C. Lay and *Fundamental Methods of Mathematical Economics*, 4th ed., by Alpha C. Chiang and Kevin Wainwright. The material posted on this website is for personal use only and is not intended for reproduction, distribution, or citation.

(g) <u>Definition</u>. The **size** of a matrix tells how many rows and columns it has. An $m \times n$ matrix has m rows and n columns:

$$
\begin{bmatrix}
a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\
a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\
a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn}
\end{bmatrix}
$$

(h) <u>Definition</u>. The **element** a_{ij} is the value in row i, column j

(i) <u>Definition</u>. To solve the system of equations, the three basic operations are known as **elementary row operations**. These operations consist of:

- Scaling - Multiply all entries in a row by a non-zero constant

- Replacement - Replace one row by the sum of itself and the multiple of another row

- Interchange - Interchange any two rows

(j) <u>Example</u>. Consider the augmented matrix from our system above:

$$
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 2 & -8 & 8 \\
-4 & 5 & 9 & -9
\end{array}\right]
\rightarrow \text{Multiplying } R_2 \text{ by } 1/2: \rightarrow
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 1 & -4 & 4 \\
-4 & 5 & 9 & -9
\end{array}\right]
$$

Replacing R_3 with $R_3 + 4R_1$:

$$
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 1 & -4 & 4 \\
0 & -3 & 13 & -9
\end{array}\right]
\rightarrow \text{Replacing } R_3 \text{ with } R_3 + 3R_2: \rightarrow
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 1 & -4 & 4 \\
0 & 0 & 1 & 3
\end{array}\right]
$$

Replacing R_2 with $R_2 + 4R_3$:

$$
\left[\begin{array}{ccc|c}
1 & -2 & 1 & 0 \\
0 & 1 & 0 & 16 \\
0 & 0 & 1 & 3
\end{array}\right]
\rightarrow \text{Replacing } R_1 \text{ with } R_1 - R_3: \rightarrow
\left[\begin{array}{ccc|c}
1 & -2 & 0 & -3 \\
0 & 1 & 0 & 16 \\
0 & 0 & 1 & 3
\end{array}\right]
$$

Replacing R_1 with $R_1 + 2R_2$:

$$
\left[\begin{array}{ccc|c}
1 & 0 & 0 & 29 \\
0 & 1 & 0 & 16 \\
0 & 0 & 1 & 3
\end{array}\right]
\rightarrow \text{The simplified system:} \rightarrow
\begin{array}{rcl}
x_1 & = & 29 \\
x_2 & = & 16 \\
x_3 & = & 3
\end{array}
$$

(k) <u>Aside</u>. This matrix is now in what we call **reduced row echelon form**. Don't worry about the terminology, however, what's important is the row operations, which may come in handy a few times throughout the year. Also note that not all linear systems have unique solutions; it is possible to have a unique solution, infinite solutions, or no solutions.

2. Vectors and Vector Equations

(a) <u>Definition</u>: A **column vector** is a $n \times 1$ matrix:

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

While a **row vector** is a $1 \times m$ matrix:

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}$$

(b) <u>Example</u>. Consider the 2×1 vectors $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. These vectors "live" in $\mathbb{R}^2$ (the set of ordered pairs of real numbers). Geometrically, we can interpret $\mathbb{R}^2$ as a 2-dimensional plane and vectors as points on the plane:
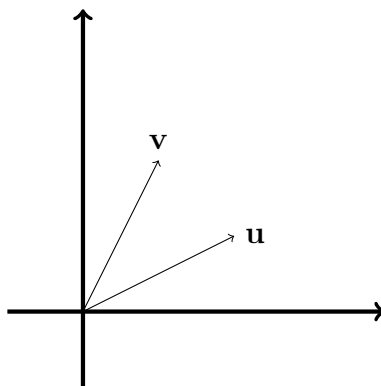
(c) <u>Aside</u>. A vector is a quantity that has magnitude and direction. A directed line segment is a line segment that has both a starting and an end point ,so it has a direction.

(d) <u>Aside</u>. Note that analagous interpretations apply to higher-dimensional spaces (e.g., $\mathbb{R}^3$, $\mathbb{R}^4$, etc.), but it's visually much easier to picture things in 2-dimensions.

(e) <u>Definition</u>. Two vectors in $\mathbb{R}^n$ are **equal** if and only if their corresponding elements are equal.

(f) <u>Example</u>. Our vectors from above, $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, are *not* equal.

(g) <u>Definition</u>. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, their **sum** is the vector $\mathbf{u} + \mathbf{v}$, obtained by adding the corresponding elements.

(h) <u>Example</u>. Consider the vectors $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

(i) <u>Definition</u>. Given a vector $\mathbf{u}$ and a real number c, the **scalar multiple** of $\mathbf{u}$ by c is the vector obtained by multiplying each element in $\mathbf{u}$ by c.

(j) <u>Example</u>. Consider the vector $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and the scalar $c = 3$. Then:

$$c\mathbf{u} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$$

(k) <u>Definition</u>. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and scalars $c_1, c_2, \ldots, c_k$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

is called a **linear combination** of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with weights $c_1, \ldots, c_n$.

(l) <u>Definition</u>. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is **linearly dependent** if and only if at least one of the vectors can be written as a linear combination of the others. If no vector is a linear combination of the others, the set of fectors are **linearly independent**.

Remark A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is **linearly independent** if and only if

$$\left\{ (c_1, \cdots, c_k) \in \mathbb{R}^k : \sum_{i=1}^{k} c_i\mathbf{v}_i = \mathbf{0} \right\} \cap \left(\mathbb{R}^k \setminus \{\mathbf{0}\} \right) = \emptyset$$

(m) <u>Example</u>. Consider the vectors:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

By inspection, $\mathbf{v}_3 = -2\mathbf{v}_1 + \mathbf{v}_2$, so the vectors are linearly dependent.

(n) <u>Definition</u>. If $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is called **span** of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. That is, the span is collection of all vectors that can be written in the form

$$\left\{ \mathbf{w} \in \mathbb{R}^n : \mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k \text{ for any } (c_1, \cdots, c_k) \in \mathbb{R}^k \right\}$$

(o) <u>Definition</u>. $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ spans $\mathbb{W}$ if given any $\mathbf{w} \in \mathbb{W}$,

$$\exists\, (c_1, \cdots, c_k) \in \mathbb{R}^k \ni \mathbf{w} = \sum_{i=1}^{k} c_i\mathbf{v}_i$$

(p) <u>Definition</u>. $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{W} \subset \mathbb{R}^n$ forms a **basis** of $\mathbb{W}$ if

    i. $\mathbf{v}_1, \ldots, \mathbf{v}_k$ spans $\mathbb{W}$, and

    ii. $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent.

(q) <u>Example</u>. Consider the vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$. These vectors span $\mathbb{R}^2$; any vector in $\mathbb{R}^2$ can be written as a linear combination of these two vectors.

3. **The Matrix Equation**

  (a) <u>Definition</u>. If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$, and if $\mathbf{x} \in \mathbb{R}^n$, then the **product** of A and $\mathbf{x}$, denoted $A\mathbf{x}$, is the linear combination of the columns of A using the corresponding entries in x as weights

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \qquad \textbf{(Column Expansion)}$$

and the equation $A\mathbf{x} = \mathbf{b}$ is called a **matrix equation**.

4

(b) <u>Theorem</u> (Lay THM 1.3). If A is a $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ and if $\mathbf{b} \in \mathbb{R}^m$, then the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution set as the linear equations whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n & | & \mathbf{b} \end{bmatrix}$$

(c) <u>Example</u>. Consider our first example, the linear system:

$$x_1 - 2x_2 + x_3 = 0$$
$$2x_2 - 8x_3 = 8$$
$$-4x_1 + 5x_2 + 9x_3 = -9$$

Let A be the coefficient matrix:

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix}$$

Then we can rewrite the linear system as a matrix equation $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ -9 \end{bmatrix}$$

(d) <u>Aside</u>. We won't worry too much about the matrix equation right now, but note that this is reminiscent of our matrix representation of OLS. We'll work with this type of equation quite a bit during the first year.

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik-1}\beta_{k-1} + \beta_k + u_i \qquad \forall\, i = 1, \cdots, n$$
$$= \mathbf{x}_i^T \boldsymbol{\beta} + u_i \qquad \forall\, i = 1, \cdots, n$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

4. **Matrix Operations**

(a) <u>Definition</u>. Let A and B be $m \times n$ matrices. Let a_{ij} and b_{ij} denote the elements in the ith row and jth column of the matrices A and B, respectively. Then A and B are **equal**, denoted $A = B$, if and only if $a_{ij} = b_{ij}$ for all i and all j.

(b) <u>Aside</u>. Note that this is simply an extension of our definition for vector equality; indeed, we could define matrix equality in terms of vector equality, i.e., two matrices are equal if and only if their corresponding vectors are equal.

(c) <u>Definition</u>. If A and B are both $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix whose elements are the sums of the corresponding elements from A and B. The **difference** $A - B$ is the $m \times n$ matrix whose elements are the differences of the corresponding elements of A and B.

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

(d) <u>Definition</u>. If k is a scalar and A is a matrix, then the **scalar multiple** kA is the matrix whose elements are k times the corresponding elements in A.

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}$$

5

(e) <u>Definition</u>. If A is a $m \times n$ matrix and B is a $n \times k$ matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_k$, then the **product** AB is the $m \times k$ matrix whose columns are $A\mathbf{b}_1, \ldots, A\mathbf{b}_k$. note that the number of columns of A must match the number of rows in B for the matrices to be conformable.

(f) <u>Aside</u>. Depending on the source, we will frequently say that A is post-multiplied by B and B is pre-multiplied by A. This helps keep straight where the matrices are in an equation, as they need to be conformable for calculations.

(g) <u>Example</u>. Suppose we have two matrices: $A_{1 \times 2}$ and $B_{2 \times 3}$. Then AB is permissible, but BA is not. If we have $A_{1 \times 2} B_{2 \times 3} = C_{1 \times 3}$, then

- For c_{11}: $c_{11} = a_{11}b_{11} + a_{12}b_{21}$
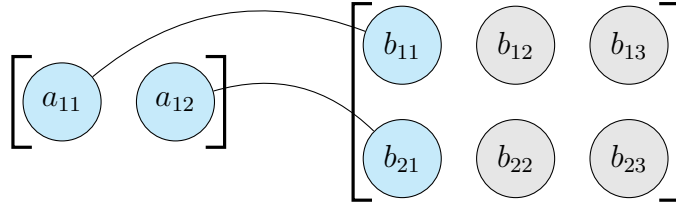
- For c_{12}: $c_{12} = a_{11}b_{12} + a_{12}b_{22}$
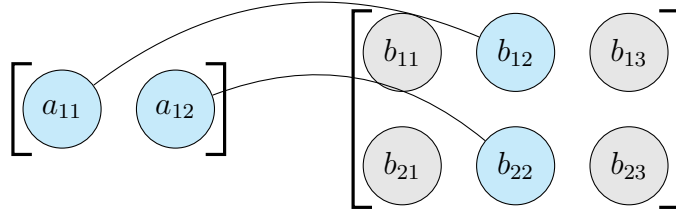
(h) <u>Aside.</u> There are a few special matrices with which you should be familiar. For example:

i. The Identity Matrix (a diagonal matrix of ones)

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Note that for all n by n matrices A:

$$AI = IA = A$$

ii. The Null Matrix (a matrix of zeros)

$$\mathbf{0}_{m \times n} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Note that for all matrices A and "conformable" matrices $\mathbf{0}$:

$$A + \mathbf{0} = A \qquad\qquad \mathbf{0} + A = A$$
$$\mathbf{0}A = \mathbf{0} \qquad\qquad A\mathbf{0} = \mathbf{0}$$

iii. Idempotent Matrices (a matrix that is the product of itself)

$$A = AA$$

(i) <u>Theorem</u> (Lay THM 2.2) Let A be a $m \times n$ matrix, and let B and C have dimensions for which the indicated sums and products are defined. Then

$$A(BC) = (AB)C = ABC \qquad \text{(Associative Law)}$$
$$A(B + C) = AB + AC \qquad \text{(Left Distributive Law)}$$
$$(B + C)A = BA + CA \qquad \text{(Right Distributive Law)}$$

(j) <u>Aside</u>. Note that in matrix algebra, we *do not* have commutativity; that is, $AB \neq BA$ in general. It may be true of specific matrices, but in many cases (as in an example above), the matrices won't even be conformable in both orders.

(k) <u>Definition</u>. Given a $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted A^T or A', whose columns are formed from the corresponding rows of A.

(l) <u>Example</u>. Consider the following matrices:

$$A = \begin{bmatrix} 3 & 8 & -9 \\ 1 & 0 & 4 \end{bmatrix} \qquad\qquad B = \begin{bmatrix} 3 & 4 \\ 1 & 7 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 3 & 1 \\ 8 & 0 \\ -9 & 4 \end{bmatrix} \qquad\qquad B^T = \begin{bmatrix} 3 & 1 \\ 4 & 7 \end{bmatrix}$$

Remark A property of matrices that occasionally comes in handy is symmetry, when $A = A^T$. Later on, we'll talk about Hessian matrices, which have this property.

(m) <u>Theorem</u> (Lay THM 2.3). Let A and B denote matrices whose sizes are appropriate for the following sums and products.

$$(A^T)^T = A$$
$$(A + B)^T = A^T + B^T$$
$$(AB)^T = B^T A^T$$

(n) <u>Definition</u>. The **determinant** of a square matrix A, denoted by $|A|$ or $\det(A)$, is a uniquely defined scalar associated with the matrix. For matrices of various sizes:

  i. 1×1 Matrix: $A = [a_{11}]$

$$|A| = a_{11}$$

  ii. 2×2 Matrix: $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

  iii. Larger Matrix: Check the appendix

(o) <u>Theorem</u> (CW SEC 5.3).

   i. Taking the transpose does not affect the determinant:
$$|A| = |A^T|$$

   ii. Scaling a row by k will change the value of the determinant k-fold, e.g.:
$$\begin{vmatrix} ka & kb \\ c & d \end{vmatrix} = k \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

   iii. Replacement–adding a multiple of a row (or column) to another row (or column)–leaves the determinant unchanged, e.g.:
$$\begin{vmatrix} a & b \\ c + ka & d + kb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

   iv. Interchanging any two rows (or columns) reverses the sign of the determinant (but does not change the absolute value), e.g.:
$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = - \begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

(p) <u>Example</u>. $f_i - \lambda g_i = 0$ $\forall i = 1, \cdots, n$ and $\lambda > 0$

$$\begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ g_2 & f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_n & f_{n1} & \cdots & \cdots & f_{nn} \end{vmatrix} = \begin{vmatrix} 0 & \frac{1}{\lambda}f_1 & \frac{1}{\lambda}f_2 & \cdots & \frac{1}{\lambda}f_n \\ \frac{1}{\lambda}f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ \frac{1}{\lambda}f_2 & f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\lambda}f_n & f_{n1} & \cdots & \cdots & f_{nn} \end{vmatrix} = \left(\frac{1}{\lambda}\right)^2 \cdot \begin{vmatrix} 0 & f_1 & f_2 & \cdots & f_n \\ f_1 & f_{11} & f_{12} & \cdots & f_{1n} \\ f_2 & f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_n & f_{n1} & \cdots & \cdots & f_{nn} \end{vmatrix}$$

If we assume that the preference is convex and that the budget constraint is linear, then FOC guarantees the utility maximization.

(q) <u>Definition</u>. A matrix A is **singular** if and only if
$$|A| = 0$$
which occurs only when rows (or columns) are linearly dependent.

Remark I'm bringing this to your attention primarily for econometric reasons; perfect-collinearity means that two of your explanatory variables are linear combinations of one another. In this case, STATA, R, or whatever program you ultimately use to run regressions will fail (as they use matrix algebra to do OLS).

(r) <u>Definition</u>. A $n \times n$ matrix is said to be **invertible** if there is an $n \times n$ matrix C such that
$$CA = I_n \quad \text{and} \quad AC = I_n$$

In this case, C is the **inverse** of A. In fact, the inverse is unique, so it is typically denoted A^{-1}.

(s) <u>Theorem</u> (Lay THM 2.4). If A is a 2×2 matrix
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and $ad - bc \neq 0$ (i.e., the determinant is non-zero/the matrix is non-singular), then the inverse is given by the formula:
$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Check the appendix for a larger matrix.

(t) <u>Example</u>. Consider the matrix

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$$

Since the determinant is $|A| = (3 * 6) - (4 * 5) = -2 \neq 0$, A is intvertible:

$$|A| = -2 \qquad \text{(the determinant)}$$

$$A^{-1} = \frac{1}{-2} \begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} \qquad \text{(by THM 2.4)}$$

$$A^{-1} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix} \qquad \text{(simplifying)}$$

Checking to make sure we did everything properly:

$$AA^{-1} = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix} \qquad \text{(multiplying matrices)}$$

$$= \begin{bmatrix} (-9+10) & (6-6) \\ (-15+15) & (10-9) \end{bmatrix} \qquad \text{(the product)}$$

$$AA^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{(simplifying)}$$

(u) <u>Theorem</u> (Lay THM 2.6). If A and B are $n \times n$ invertible matrices, then

    i. A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

    ii. AB is invertible and

$$(AB)^{-1} = B^{-1}A^{-1}$$

    iii. A^T is invertible and

$$(A^T)^{-1} = (A^{-1})^T$$

(v) <u>Theorem</u>. If A is a $n \times n$ invertible matrix, the inverse may be found by row reducing the augmented matrix $\begin{bmatrix} A \mid I_n \end{bmatrix}$, which is row equivalent to $\begin{bmatrix} I_n \mid A^{-1} \end{bmatrix}$.

5. **Connecting Concepts**

(a) <u>Theorem</u> (Lay THM 2.8). Let A be a $n \times n$ matrix. Then the following statements are equivalent:

    i. $|\mathbf{A}| \neq 0$

    ii. A is invertible

    iii. The equation $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ for each $\mathbf{b} \in \mathbb{R}^n$

    iv. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial (i.e., $\mathbf{x} = \mathbf{0}$) solution

    v. The columns of A form a linearly independent set

    vi. The columns of A span $\mathbb{R}^n$

(b) <u>Aside</u>. With the determinant and inverse, we now have a few different ways to solve linear systems that we may come across during the first year:

    i. Standard substitution (potentially very time consuming)

    ii. Row operations on an augmented matrix

    iii. Inverting $\mathbf{A}$ to find $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

In reality, however, you aren't asked to solve entire linear systems all that frequently; the first two tools frequently suffice. Understanding how inverses work, however, is important for OLS. There is one additional solution concept that you may also find helpful.

(c) Theorem (Lay THM 3.7). Given a system of equations $A\mathbf{x} = \mathbf{b}$, where A is a $n \times n$ invertible matrix and $\mathbf{b} \in \mathbb{R}$, let $A_i(\mathbf{b})$ be the matrix obtained from A by replacing column i with $\mathbf{b}$:

$$A_i(\mathbf{b}) = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \ldots & \mathbf{a}_n \end{bmatrix}$$

Then for any $\mathbf{b} \in \mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by:

$$x_i = \frac{|A_i(\mathbf{b})|}{|A|}, \quad i = 1, 2, \ldots, n$$

Check the appendix for the example in economics.

(d) Definition. The **rank** of a matrix is the number of nonzero rows in its row echelon form.

(e) Theorem. Let A be a $m \times n$ matrix and B be a $n \times k$ matrix. Then

   i. $\text{rank}(A) = \text{rank}(A^T)$

   ii. $\text{rank}(A) \leq \min\{m, n\}$

   iii. $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$

   iv. $\text{rank}(A) = \text{rank}(A^T A) = \text{rank}(AA^T)$

6. **Quadratic Forms**

(a) Definition. A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x} \in \mathbb{R}^n$ can be written in the form
$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$$
where A is a $n \times n$ symmetric matrix.

(b) Example. Consider a two-variable quadratic form

$$Q(\mathbf{x}) = ax_1^2 + 2bx_1x_2 + cx_2^2 \qquad \text{(a two-variable quadratic form)}$$

$$Q(\mathbf{x}) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{(rewriting in matrix form)}$$

Note that the matrix of constants is not unique!

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

(c) Aside. Frequently, you will also see quadratic forms written in double-sum notation:

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Where a_{ij} are elements of A, which is symmetric, i.e., $a_{ij} = a_{ji} \quad \forall i, j$.
If off-diagonal elements are all zero, then

$$\mathbf{x}^T A \mathbf{x} = \sum_{i=1}^{n} a_{ii} x_i^2$$

10

(d) <u>Example</u>.

$$-\sum_{i=1}^{n}\frac{u_i^2}{2\sigma^2} = -\frac{1}{2}\mathbf{u}^T(\sigma^2 I_n)^{-1}\mathbf{u}$$

(e) <u>Definition</u>. Let $Q(\mathbf{x})$ be a quadratic form. Then $Q(\mathbf{x})$ is:

   i. **Positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$

   ii. **Negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$

   iii. **Indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values

Quadratic forms can also be **postive semi-definite** or **negative semi-definite** if we replace the strict inequalities with weak inequalities.

(f) <u>Example</u>. Consider the quadratic form:

$$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

This is a positive definite quadratic form, e.g., if $\mathbf{x}^T = \begin{bmatrix} 1 & 1 \end{bmatrix}$:

$$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{(plugging in values)}$$

$$\mathbf{x}^T A\mathbf{x} = 2$$

Similarly, if $\mathbf{x}^T = \begin{bmatrix} -3 & 2 \end{bmatrix}$:

$$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix} -3 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -3 \\ 2 \end{bmatrix} \qquad \text{(plugging in values)}$$

$$\mathbf{x}^T A\mathbf{x} = 38$$

Indeed, any vector in $\mathbb{R}^2$ (besides the zero-vector) will produce a strictly positive value.

(g) <u>Example</u>. We can prove that a positive definite matrix $A_{n\times n}$ is nonsingular. In other words,

$$\text{Positive Definite} \implies \text{Nonsingular}$$

An easy way to do this is to prove by contradiction.

- $P \implies Q$ is equivalent to $\neg P \vee Q$

- $\neg(P \implies Q)$ is equivalent to $\neg(\neg P \vee Q)$, or $P \wedge \neg Q$

- Def. of Positive Definite: $\forall \mathbf{x} \neq \mathbf{0}, \mathbf{x}^T A\mathbf{x} > 0$

- Theorem (T1).: if A is singular, $\exists\, \mathbf{x} \neq \mathbf{0} \ni A\mathbf{x} = \mathbf{0}$

Proof by contradiction <u>to show:</u> $\forall\, \mathbf{x}, \mathbf{x}^T A\mathbf{x} > 0$ and $\exists\, y \ni \mathbf{x}^T A\mathbf{x} = 0$.

Proof.

Let A be positive definite	(by hypothesis)
$\implies \forall\, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^T A\mathbf{x} > 0$	(by def. of positive definite)
Suppose that A is singular	(towards a contradiction)
$\implies \exists\, \mathbf{y} \neq \mathbf{0} \ni A\mathbf{y} = \mathbf{0}$	(by T1)
$\implies \mathbf{y}^T A\mathbf{y} = 0$	(premultiplying by $\mathbf{y}^T$)
Thus, a contradiction	
$\implies \mathbf{A}$ is nonsingular	

$\square$

(h) <u>Definition</u>. Let A be a $n \times n$ matrix with elements a_{ij}. Then a **principle minor** is a minor produced by the deletion of the row and column associated with an element such that $i = j$. A **nth-order principle minor** is produced by the deletion of rows and columns associed with elements such that $i = j$, leaving a nth order determinant.

(i) <u>Example</u>. Consider the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

The 1st-order principle minors are:

$$|a_{11}| = a_{11} \qquad \text{(deleting 2nd and 3rd rows/columns)}$$

$$|a_{22}| = a_{22} \qquad \text{(deleting 1st and 3rd rows/columns)}$$

$$|a_{33}| = a_{33} \qquad \text{(deleting 1st and 2nd rows/columns)}$$

The 2nd-order principle minors are:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \qquad \text{(deleting the 3rd row/column)}$$

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{23}a_{32} \qquad \text{(deleting the 2nd row/column)}$$

$$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{11}a_{33} - a_{31}a_{13} \qquad \text{(deleting the 1st row/column)}$$

The 3rd-order principle minor is:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = |A| \qquad \text{(deleting no rows/columns)}$$

(j) <u>Definition</u>. The **leading principle minors** are the principal minors that are associated with contiguous, upper-left elements of the matrix.

(k) <u>Example</u>. For the 3×3 matrix A in the preceding example, the leading principle minors are

$$|A_1| = |a_{11}| \qquad \text{(the 1st-order LPM)}$$

$$|A_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad \text{(the 2nd-order LPM)}$$

$$|A_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \qquad \text{(the 3rd-order LPM)}$$

(l) <u>Aside</u>. Note that in the example, we are using Simon and Blume's notation. This is by no means consistent throughout the various mathematical economics textbooks available; indeed, notation and terminology vary quite a bit. We always start enumerating leading principle minors, however, with element a_{11}, which is the upper-left corner of *every* leading principle minor.

(m) <u>Theorem</u> (CW SEC 11.3 and SB THM 16.2). Let A be a $n \times n$ symmetric matrix. then the quadratic form $\mathbf{x}^T A \mathbf{x}$ is:

- positive definite if and only if all of its *leading principle mionors* are strictly positive

- positive semi-definite if and only if all of its *principle minors* are non-negative

- negative definite if and only if all of its *leading principle minors* follow the signing convention

$$|A_1| < 0 \qquad |A_2| > 0 \qquad |A_3| < 0 \ldots$$

- negative semi-definite if and only if all of its *principle minors* of an odd-order are ≤ 0 and all of it's *principle minors* of an even-order are ≥ 0.

- indefinite if none of the above hold

(n) <u>Example</u>. Consider the matrix

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{bmatrix}$$

We can determine the definiteness of the matrix by first evaluating the leading principle minors:

$$|1| = 1$$

$$\begin{vmatrix} 1 & -1 \\ -1 & 6 \end{vmatrix} = 5$$

$$\begin{vmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{vmatrix} = 11$$

All the LPMs are positive, so A is positive definite. Note that if one of the leading principle minors was equal to zero, we'd need to check all of the other PMs for positive semi-definiteness.

(o) <u>Aside</u>. This has a close relationship to optimization. Consider a function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ which is of class $\mathcal{C}^2$. Then the Hessian matrix is given by

$$H(\mathbf{x}) = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}$$

Where f_{ij} denotes the second derivative $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$. Consider a two-variable function $z = f(x, y)$. Then the second-order total differential is given by:

$$dz = f_x dx + f_y dy \qquad\qquad\qquad \text{(1st-order total differential)}$$

$$d^2 z = f_{xx} dx^2 + f_{xy} dx dy + f_{yx} dy dx + f_{yy} dy^2 \qquad\qquad \text{(2nd-order total differential)}$$

$$d^2 z = f_{xx} dx^2 + 2 f_{xy} dx dy + f_{yy} dy^2 \qquad\qquad \text{(by Young's Theorem)}$$

We can re-write this in matrix notation:

$$d^2 z = \begin{bmatrix} dx & dy \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} \qquad\qquad \text{(the quadratic form)}$$

We can rewrite the 2nd-order total differential in quadratic form, using the Hessian.

i. <u>Theorem</u>. Let f be a twice-continuously differentiable function then if the associated Hessian is

    A. positive definite, then $f(\cdot)$ is strictly convex

    B. positive semi-definite, then $f(\cdot)$ is convex

    C. negative definite, then $f(\cdot)$ is strictly concave

    D. negative semi-definite, then $f(\cdot)$ is concave

Basically, the intutition is that if we're at a point where the derivatives equal zero, then for any small movements dx and dy, we will get a positive movement d^2z if the Hessian is positive definite, which means that the function is convex. We will talk about these topics more when we talk about contstrained optimization.

7. **Eigenvalues and Eigenvectors**

(a) <u>Definition</u>. For an $n \times n$ matrix A, the **trace** is the sum of the elements along the main diagonal:

$$\text{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

(b) <u>Example</u>. Consider the identity matrix:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The trace is the sum along the main diagonal, so $\text{tr}(I_3) = 3$ (indeed, the trace of any identity matrix is the dimensionality of that matrix).

(c) <u>Theorem</u>. Let A and B be $n \times n$ matrices. Traces have the following properties:

i. Linearity

$$\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$$
$$c\,\text{tr}(A) = \text{tr}(cA)$$

ii. Product/Transpose Rules

$$\text{tr}(AB) = \text{tr}(BA)$$
$$\text{tr}(A^T) = \text{tr}(A)$$

(d) <u>Definition</u>. A **eigenvector** of a $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$. Note that we can rewrite this equation to be:

$$(A - \lambda I_n)\mathbf{x} = \mathbf{0}$$

(e) <u>Example</u>. Consider the matrix and vector:

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$$

14

We can show that $\mathbf{u}$ is an eigenvector of A:

$$Au = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} \qquad \text{(the product)}$$

$$= \begin{bmatrix} (6 - 30) \\ (30 - 10) \end{bmatrix} \qquad \text{(multiplying)}$$

$$= \begin{bmatrix} -24 \\ 20 \end{bmatrix} \qquad \text{(simplifying)}$$

$$Au = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix} \qquad \text{(factoring)}$$

Thus $\mathbf{u}$ and -4 are an eigenvector and an eigvenvalue, respectively.

(f) <u>Aside</u>. Recall that one of our theorems states that $B\mathbf{x} = \mathbf{0}$ has a non-trivial solution $\mathbf{x}$ if and only if B is singular. Thus, for there to be a non-trivial eigenvector, it must be that $(A - \lambda I_n)$ is singular (even if A itself is *not* singular).

(g) <u>Definition</u>. For a $n \times n$ matrix A, the scalar equation

$$|(A - \lambda I_n)| = 0$$

is called the **characteristic equation** of A.

(h) <u>Theorem</u>. A scalar λ is an eigenvalue of a $n \times n$ matrix A if and only if λ satisfies the charactersitic equation of A.

(i) <u>Example</u>. Suppose we have the following 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

To find the eigenvalues, we want to pick the λs such that:

$$\begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = (a - \lambda)(d - \lambda) - bc = 0$$

This is the characteristic equation for the matrix.

(j) <u>Example</u>. Consider the matrix

$$A = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$$

We can find the associated eigenvalues by employing the characteristic equation:

$$|(A - \lambda I_2)| = \begin{vmatrix} 2 - \lambda & 2 \\ 2 & -1 - \lambda \end{vmatrix} \qquad \text{(the characteristic equation)}$$

$$= (2 - \lambda)(-1 - \lambda) - 4 \qquad \text{(the determinant)}$$

$$= \lambda^2 - \lambda - 6 \qquad \text{(simplifying)}$$

$$\lambda^2 - \lambda - 6 = 0 \qquad \text{(setting equal to zero)}$$

$$(\lambda - 3)(\lambda + 2) = 0 \qquad \text{(factoring)}$$

Thus, $\lambda_1 = 3$ and $\lambda_2 = -2$ are the eigenvalues. To find associated eigenvectors, recall that we need $(A - \lambda I_2)\mathbf{x} = \mathbf{0}$. Thus, for our first eigen value:

$$(A - \lambda_1 I_n)\mathbf{x} = 0 \qquad \text{(our condition)}$$

$$\begin{bmatrix} 2 - 3 & 2 \\ 2 & -1 - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \text{(plugging in values)}$$

$$\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \text{(simplifying)}$$

Note that the rows are linear combinations of one another; this reduces to the equation:

$$x_1 = 2x_2 \qquad \text{(reducing the system)}$$

Note that eigenvectors are not unique; to force a standardized solution, we frequently normalize the vectors (i.e., make them length one):

$$\sqrt{x_1^2 + x_2^2} = 1 \qquad \text{(normalizing)}$$

$$x_1^2 + x_2^2 = 1 \qquad \text{(squaring both sides)}$$

$$(2x_2)^2 + x_2 = 1 \qquad \text{(plugging in for } x_1\text{)}$$

$$5x_2^2 = 1 \qquad \text{(simplifying)}$$

$$x_2 = \frac{1}{\sqrt{5}} \qquad \text{(solving for } x_2 \text{ using the positive root)}$$

Thus, our eigenvector associated with λ_1 is

$$\mathbf{x}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

If we performed the same steps for our second eigenvalue, λ_2:

$$\mathbf{x}_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Note that in general, we may have n distinct eigenvalues and n associated eigenvectors, all of which we normalized to length one.

(k) <u>Theorem</u>. Let A be a $n \times n$ matrix. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then:

i. The trace equals the sum of the eigenvalues

$$\text{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

ii. The determinant of A equals the product of the eigenvalues

$$|A| = \prod_{i=1}^{n} \lambda_i$$

16

iii. The trace of A^k ($\mathbf{A}$ multiplied by itself k times) is the sum of the eigenvalues raised to the kth power

$$\text{tr}(A^k) = \sum_{i=1}^{n} \lambda_i^k$$

iv. Let $A^T = A$.

   A. Eigenvalues of A are real.

   B. Eigenvectors corresponding to distinct eigenvalues are pairwise orthogonal.

   Proof. Pick $i, j \in \{1, \cdots, n\}$ such that $\lambda_i \neq \lambda_j$.

   $$\lambda_i \mathbf{x}_j^T \mathbf{x}_i = \mathbf{x}_j^T (\lambda_i \mathbf{x}_i) = \mathbf{x}_j^T (A\mathbf{x}_i) = \mathbf{x}_j^T (A^T \mathbf{x}_i) = (\mathbf{x}_j^T A^T)\mathbf{x}_i = (A\mathbf{x}_j)^T \mathbf{x}_i = (\lambda_j \mathbf{x}_j)^T \mathbf{x}_i = \lambda_j \mathbf{x}_j^T \mathbf{x}_i$$

   $$(\lambda_i - \lambda_j)\mathbf{x}_j^T \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{x}_j^T \mathbf{x}_i = 0$$

   $\square$

   C. A can be diagonalized.
   $$AX = X\Lambda \quad \text{where} \quad X^T X = I_n$$
   $$X^T A X = \Lambda$$

v. Let $A^T = A$.

   A. Positive (semi)definite if and only if the eigenvalues of A are all (weakly) positive.

   Proof. Given any $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $\mathbf{v}^T (X^T A X)\mathbf{v} = (X\mathbf{v})^T A(X\mathbf{v})$ and $\mathbf{v}^T X^T A X \mathbf{v} = \mathbf{v}^T \Lambda \mathbf{v}$. Therefore, $(X\mathbf{v})^T A(X\mathbf{v}) = \mathbf{v}^T \Lambda \mathbf{v}$. $\square$

   B. Negative (semi)definite if and only if the eigenvalues of A are all (weakly) negative.

   C. Indefinite if and only if A has both positive and negative eigenvalues.

vi. Let $A^T = A$ and A is idempotent.

   A. Eigenvalues of A are either 0 or 1.

   Proof. $\Lambda^2 = X^T A^2 X = X^T A X = \Lambda$ Therefore, $\lambda \in \{0, 1\}$. $\square$

   B. A is positive semidefinite.

   Proof. Given any $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $\mathbf{v}^T A \mathbf{v} = \mathbf{v}^T (X^T \Lambda X)\mathbf{v} = (X\mathbf{v})^T \Lambda (X\mathbf{v}) \geq 0$. $\square$

   C. $\text{rank}(A) = \text{tr}(A)$.

vii. Let $A^T = A$ and A is positive definite. Then there exists C such that $A = C^T C$.
   Proof. $A = X\Lambda X^T = X\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}X^T = (\Lambda^{\frac{1}{2}}X^T)^T (\Lambda^{\frac{1}{2}}X^T)$ $\square$

(1) <u>Example</u>. Let X be a $n \times k$ matrix with $\text{rank}(X) = k$ ($n \geq k$).

$$M_X := I_n - X(X^T X)^{-1}X^T$$

   i. M_X is symmetric.

   ii. M_X is idempotent.

8. **Vector Spaces and Norms**

   (a) <u>Aside</u>. Building on what we know about already about vectors, we can start thinking about vector spaces. Recall that earlier, we mentioned that two linearly independent vectors $\mathbf{v}, \mathbf{u} \in \mathbb{R}^2$ span $\mathbb{R}^2$. Indeed, $\mathbb{R}^2$ is the vector space defined by the various linear combinations of $\mathbf{v}$ and $\mathbf{u}$. Indeed, $\mathbb{R}^n$ is the vector space that we usually are dealing with; however, it probably is a good idea for us to know a more formal definition.

   (b) <u>Definition</u>. A **vector space** is a set of vectors V equipped with addition and scalar multiplication that satisfies the following properties for vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ and scalars a and b:

   - Associativity of addition: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

   - Commutativity of addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

   - The identity element of addition: $\exists\, \mathbf{0} \in V \ni \mathbf{u} + \mathbf{0} = \mathbf{u}\ \forall\, \mathbf{u} \in V$

   - The inverse element of addition: $\forall\, \mathbf{u} \in V\ \exists - \mathbf{u} \in V \ni \mathbf{u} + (-\mathbf{u}) = 0$

   - Compatibility of sclar and field multiplication: $a(b\mathbf{u}) = (ab)\mathbf{u}$

   - The identity element of scalar multiplication: $1\mathbf{u} = \mathbf{u}$

   - Distributivity of scalar multiplication with respect to vector addition: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$

   - Distributivity of scalar multiplication with respect to scalar addition: $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$

   (c) <u>Definition</u>. Given a vector space V, a subset $W \subset V$ is a **subspace** if and only if given any $\lambda, \mu \in \mathbb{R}$ and $v, w \in W$,
   $$\lambda v + \mu w \in W$$

   Remark
   - The zero vector $\mathbf{0}$ is in W

   - If $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} \in W$

   - If $\mathbf{u}$ is in W and k is a scalar, then $k\mathbf{u}$ is in W

   (d) <u>Aside</u>. Again, the concepts of vector spaces and subspaces will probably not be particularly relevant to you in terms of your first-year coursework, but they are concepts with which you should be familiar. In particular, there are a few subspaces which may come up, particularly in econometrics.

   (e) <u>Definition</u>. The **column space** of a matrix A, denoted $\mathrm{Col}(A)$, is the set of all linear combinations of the columns of A.

   (f) <u>Definition</u>. The **null space** of a matrix A, denoted $\mathrm{Nul}(A)$, is the set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

   (g) <u>Aside</u>. Let A be a $m \times n$ matrix.

   $$\mathrm{Col}(A) = \{\mathbf{b} \in \mathbb{R}^m\ :\ A\mathbf{x} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}$$
   $$\mathrm{Nul}(A) = \{\mathbf{x} \in \mathbb{R}^n\ :\ A\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}$$
   $$\mathrm{Col}(A^T) = \{\mathbf{b} \in \mathbb{R}^n\ :\ A^T\mathbf{x} = \mathbf{b} \text{ for some } \mathbf{x} \in \mathbb{R}^m\}$$
   $$\mathrm{Nul}(A^T) = \{\mathbf{x} \in \mathbb{R}^m\ :\ A^T\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \in \mathbb{R}^m\}$$

(h) Example. Consider the matrix

$$B = \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A basis for this matrix is the set of vectors

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{b}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

These are linearly independent, and all other columns of the matrix are linear combinations of these.

(i) Definition. A **metric space** (X, d) is a set X and a function, $d : X \times X \to \mathbb{R}$, such that for any $x, y, z \in X$, it satisfies the following properties:

- $d(x, y) \geq 0$
- $d(x, y) = 0 \iff x = y$
- $d(x, y) = d(y, x)$
- $d(x, z) \leq d(x, y) + d(y, z)$

(j) Definition. A **normed vector space** $(V, ||\cdot||)$ is a vector space V and a function, $||\cdot|| : V \to \mathbb{R}$, such that for any $\mathbf{x}, \mathbf{y} \in V$ and for any $\alpha \in \mathbb{R}$, it satisfies the following properties:

- $||\mathbf{x}|| \geq 0$
- $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
- $||\alpha\mathbf{x}|| = |\alpha| \cdot ||\mathbf{x}||$
- $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$

(k) Aside. Norms are the first step in getting us a notion of distance in a space. While we won't end up using norms explictly all that frequently during the first year, they come about a great deal implicitely, any time we talk about things like minimum distance, or minimum distance estimators. We typically deal with one norm, but there are a few others.

(l) Examples. The following are all norms, which may be familiar:

i. Euclidean Norm (occasionally denoted $||\mathbf{x}||_2$):

$$|\mathbf{x}| = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}$$

This is the "as the crow flies" distance from the origin in $\mathbb{R}^2$, and is (by far) the most frequent norm we will use.

ii. Manhattan Norm (occastionally denoted $||\mathbf{x}||_1$):

$$||\mathbf{x}|| = \sum_{i=1}^{n} |x_i|$$

This is akin to measuring distance from the origin on a map laid out as a grid of streets.

(m) <u>Example</u>. Consider $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. The most common metric we use is the notion of Euclidean distance, i.e.,

$$d(x, y) = \left((x_1 - y_1)^2 + (x_2 - y_2)^2\right)^{\frac{1}{2}}$$

This is equivalent to stating $||\mathbf{x} - \mathbf{y}||_2$.

(n) <u>Example</u>. Let $X \subset \mathbb{R}^n$ and $S = C(X)$ be the set of all continuous and bounded functions $f : X \to \mathbb{R}$. Define the metric $d : S \times X \to \mathbb{R}$ such that given any $g, h \in S$

$$d(g, h) = \sup_{\mathbf{x} \in X} |g(\mathbf{x}) - h(\mathbf{x})|$$

Then (S, d) is a metric space.

(o) <u>Aside</u>. As we move forward in econometrics, micro theory, or anything else during the first year sequences, we will implicitly assume that any "minimum distance" is minimizing Euclidean distance unless specifically told otherwise. A notion of distance, however, will be used implicitly (or explicitly) in classes, talks, papers, etc., so it is a concept with which you should be familiar.

9. **Orthogonal Projections**

(a) <u>Definition</u>. Given a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ and a vector $\mathbf{y} \in \mathbb{R}^n$, then $\mathbf{y}$ may be written as a sum of two vectors

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{x}$ for some scalar α and $\mathbf{u}$ is orthogonal to $\mathbf{x}$. Then $\hat{\mathbf{y}}$ is the **orthogonal projection** of $\mathbf{y}$ onto $\mathbf{x}$, and $\hat{\mathbf{u}}$ is the component of $\mathbf{y}$ that is orthgonal to $\mathbf{x}$.

(b) <u>Aside</u>. Alternative terminology would state that $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto the span of $\mathbf{x}$. For us, it does not matter which terminology we use. The first is probably more common, but the second is perhaps a bit easier to understand graphically.

(c) <u>Theorem</u>. The α that satisfies the definition of the orthongal projection is given by

$$\alpha = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}$$

Proof.

$$f(a) = ||a\mathbf{x} - \mathbf{y}|| = (a\mathbf{x} - \mathbf{y})^T(a\mathbf{x} - \mathbf{y}) = a^2\mathbf{x}^T\mathbf{x} - 2a\mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{y}$$
$$\frac{\partial f(\alpha)}{\partial a} = 0$$
$$2\alpha\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{y} = 0$$
$$\alpha = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}$$

$\square$

The orthgonal projection $\hat{\mathbf{y}}$ is given by the equation:

$$\hat{\mathbf{y}} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}\mathbf{x}$$

(d) <u>Theorem</u>. Let $\theta \in (0, \pi)$ be angle between two vectors $\mathbf{x}, \mathbf{y}$. Then

$$\mathbf{x}^T \mathbf{y} = ||\mathbf{x}|| \cdot ||\mathbf{y}|| \cdot \cos(\theta)$$

Proof.

$$\alpha \mathbf{x} = \frac{||\mathbf{y}|| \cdot \cos(\theta)}{||\mathbf{x}||} \mathbf{x}$$

$$\alpha \mathbf{x} = \frac{\mathbf{x}'\mathbf{y}}{||\mathbf{x}||^2} \mathbf{x}$$

□

(e) <u>Aside</u>. Two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are orthogonal when an angle between them is right ($\theta = \frac{1}{2}\pi$).

(f) <u>Aside</u>. Note that these formulas look virtually identical to our typical matrix formula for the coefficient estimates in OLS. This is not a coicidence; indeed, if we have only one explantory variable and no constant, these is the *exact* formula for $\hat{\beta}$ and the fitted values.

(g) <u>Example</u>. Consider the two vectors

$$\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

We can find the orthogonal project of $\mathbf{y}$ onto the span of $\mathbf{x}$, then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in span($\mathbf{x}$) and one orthogonal to $\mathbf{x}$:

$$\mathbf{x}'\mathbf{y} = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 6 \end{bmatrix} = 40 \qquad \text{(the numerator)}$$

$$\mathbf{x}'\mathbf{x} = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20 \qquad \text{(the denominator)}$$

$$\hat{\mathbf{y}} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}} \mathbf{x} = \frac{40}{20} \begin{bmatrix} 4 \\ 2 \end{bmatrix} \qquad \text{(the projection formula)}$$

$$\hat{\mathbf{y}} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \qquad \text{(simplifying)}$$

So the projection of $\mathbf{y}$ onto $\mathbf{u}$ is $[8 \quad 4]'$. The orthogonal component is $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$:

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} \qquad \text{(by def. of } \hat{\mathbf{u}}\text{)}$$

$$\hat{\mathbf{u}} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \text{(simplifying)}$$

Thus, the vector $\mathbf{y}$ can be written as

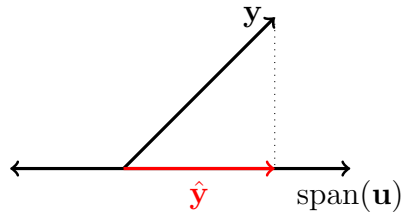$$\mathbf{y} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

(h) <u>Example</u>. Consider the proejction of $\mathbf{y}$ onto $\mathbf{x}$ where

$$\mathbf{y} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

By our formula, we know that

$$\hat{\mathbf{y}} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Graphically:

(i) <u>Aside</u>. This idea extends to vectors that don't span the x-axis (it's just a bit easier to draw). Similarly, we can think about projections in an analagous way for multiple dimensions (although it is much harder to draw and much harder to picture mentally). Indeed, higher dimensionality is important–the fitted values from OLS, for example, are the projection of $\mathbf{y}$ onto the column space of the data matrix X. Luckily, we have theorems to help us handle more dimensions.

10. **Orthogonality**

(a) <u>Definition</u>. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Then the **inner product** of the vectors is a scalar given by $\mathbf{u}^T\mathbf{v}$ or $\mathbf{u}\cdot\mathbf{v}$.

$$\mathbf{u}^T\mathbf{v} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n$$

This is also known as the **dot product**.

(b) <u>Aside</u>. Note that we can write our usual notion of length, the Euclidian norm, in terms of inner product:

$$||\mathbf{x}||_2 = \sqrt{\mathbf{x}^T\mathbf{x}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}$$

(c) <u>Definition</u>. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if and only if $\mathbf{u}^T\mathbf{v} = 0$.

(d) <u>Theorem</u> (Lay THM 6.2). Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if and only if

$$||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$$

This is the linear algebra version of the Pythagorean theorem.

(e) <u>Aside</u>. We can work through some of the algebra to show that this must be true if we're using the Euclidean norm. Note that $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = 0$ if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal:

$$\begin{aligned} ||\mathbf{x} + \mathbf{y}||^2 &= (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{x} + \mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{y} \\ &= ||\mathbf{x}||^2 + 0 + 0 + || + \mathbf{y}||^2 \\ ||\mathbf{x} + \mathbf{y}||^2 &= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 \end{aligned}$$

Now, the same steps can be used to show that $||\mathbf{x} - \mathbf{y}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2$. This implies that $||\mathbf{x} + \mathbf{y}||^2 = ||\mathbf{x} - \mathbf{y}||^2$, or that $\mathbf{x}$ is equidistant to $\mathbf{y}$ and $-\mathbf{y}$. Graphically:
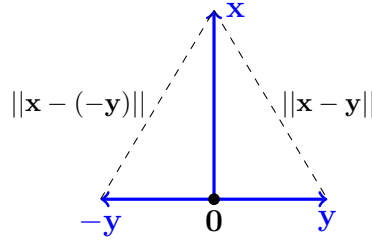
In other words, orthogonal vectors are perpendicular to each other.

(f) <u>Definition</u>. Nonempty "sets" $V, W \subset \mathbb{R}^n$ are orthogonal if and only if given any $\mathbf{v} \in V$, $\mathbf{w} \in W$

$$\mathbf{v}^T \mathbf{w} = 0$$

Remark The collection of vectors is not necessarily a vector space.

(g) <u>Theorem</u>. If $V, W \subset \mathbb{R}^n$ are orthogonal, then

$$V \cap W = \emptyset \ \lor \ V \cap W = \{\mathbf{0}\}$$

(h) <u>Definition</u>. Let V be a subset of $\mathbb{R}^n$. The orthogonal complement of V, denoted by $V^\perp$, is the set of all vectors $w \in \mathbb{R}^n$ that are orthogonal to the set V.

(i) <u>Aside</u>.

    i. $V^\perp$ is a subspace of $\mathbb{R}^n$.

    ii. $\text{span}(V) = (V^\perp)^\perp$

(j) <u>Theorem</u> (Lay THM 6.3). Let A be a $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row}(A))^\perp = \text{Nul}(A) \quad \text{and} \quad (\text{Col}(A))^\perp = \text{Nul}(A^T)$$

Proof.

$$
\begin{aligned}
\text{Row}(A) &= \text{Col}(A^T) \\
\text{Nul}(A) &= \{\mathbf{x} \ : \ A\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \in \mathbb{R}^n\} \\
&= S^\perp \qquad\qquad\qquad\qquad\qquad \text{where } S \text{ is the collection of rows of } A. \\
&= \text{span}(S)^\perp \\
&= \text{Col}(A^T)^\perp
\end{aligned}
$$

$\square$

(k) <u>Aside</u>. This concept is important to us in terms of the linear algebra characterization of OLS. The column space of our data matrix X and it's orthogonal complement are key components in understanding regressions.

(l) <u>Definition</u>. A set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ in $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from the set are orthogonal. An orthogonal basis for a subspace $W \subset \mathbb{R}^n$ is a basis for W that is also an orthogonal set.

(m) <u>Theorem</u> (Lay THM 6.8). Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$$

where $\hat{\mathbf{y}}$ is in W and $\hat{\mathbf{u}} \in W^{\perp}$. If $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is a basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}'\mathbf{x}_1}{\mathbf{x}_1'\mathbf{x}_1}\mathbf{x}_1 + \cdots + \frac{\mathbf{y}'\mathbf{x}_k}{\mathbf{x}_k'\mathbf{x}_k}\mathbf{x}_k$$

and $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}$. Here, $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto W. This is known as the orthogonal decomposition theorem.

(n) <u>Aside</u>. To quick points are in order. First, this gives us a tool to understand how we can project $\mathbf{y}$ onto a higher-dimensional space than line spanned by a single vector. Second, note that the theorem states that any orthogonal basis can be used; it is the projection itself (not the basis) that is unique. Scale any vector in the basis (or analogously, any column in the data matrix X) and it does not change the fitted values.

(o) <u>Theorem</u> (Lay THM 6.9). Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y} \in \mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the point in W closest to $\mathbf{y}$:

$$||\mathbf{y} - \hat{\mathbf{y}}|| < ||\mathbf{y} - \mathbf{v}|| \text{ for all } \mathbf{v} \in W \ni \mathbf{v} \neq \hat{\mathbf{y}}$$

Proof.

$$||\mathbf{y} - \mathbf{v}|| = ||(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})|| \hspace{3cm} (\text{adding } \mathbf{0})$$

Note that $\hat{\mathbf{y}} - \mathbf{v} \in W$ (both vectors "live" in W, so their difference does as well. Further, note $\mathbf{y} - \hat{\mathbf{y}} \in W^{\perp}$ (by definition of orthogonal projection). Thus:

$$||\mathbf{y} - \mathbf{v}||^2 = ||\mathbf{y} - \hat{\mathbf{y}}||^2 + ||\hat{\mathbf{y}} - \mathbf{v}||^2 \hspace{2cm} (\text{by the Pythaogrean Theorem})$$

$$||\hat{\mathbf{y}} - \mathbf{v}|| > 0 \hspace{2cm} (\mathbf{v} \neq \hat{\mathbf{y}} \text{ and by def. of a norm})$$

$$\implies ||\mathbf{y} - \mathbf{v}||^2 > ||\mathbf{y} - \hat{\mathbf{y}}||^2 \hspace{2cm} (\text{by } ||\hat{\mathbf{y}} - \mathbf{v}|| > 0)$$

$$\implies ||\mathbf{y} - \mathbf{v}|| > ||\mathbf{y} - \hat{\mathbf{y}}|| \hspace{2cm} (\text{by non-negativity of norms})$$

$\square$

11. **OLS as a Projection**

(a) <u>Aside</u>. Suppose We have large-dimensional problem $\mathbf{X}\mathbf{b} = \mathbf{y}$ (where $\mathbf{X}$ is $n \times k$, $\mathbf{b}$ is $k \times 1$, and $\mathbf{y}$ is $n \times 1$) that has no solution. Indeed, we frequently (i.e., ALWAYS as economists) have an inconsistent system. Thanks to the idea of orthogonal projections, however, we still have a solution concept.

(b) <u>Definition</u>. If $\mathbf{X}$ is a $n \times k$ matrix and $\mathbf{y} \in \mathbb{R}^n$, a **least-squares solution** of $\mathbf{X}\mathbf{b} = \mathbf{y}$ is a $\hat{\boldsymbol{\beta}} \in \mathbb{R}^k$ such that

$$||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}|| \leq ||\mathbf{y} - \mathbf{X}\mathbf{b}|| \text{ for all } \mathbf{b} \in \mathbb{R}^k$$

(c) <u>Aside</u>. Note that no matter what, the answer we pick is going to "live" in the column space of X, $\text{Col}(\mathbf{X})$. Thus, we're trying to find the point in $\text{Col}(\mathbf{X})$ closest to $\mathbf{y}$. To do this, we'll use our best approximation theorem, which tells us that the closest point is the orthogonal projection. $\hat{\mathbf{y}} \in \text{Col}(\mathbf{X})$, so a solution to the system $\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{y}}$ *will* exist.

(d) <u>Theorem</u> (Lay THM 6.14). Let $\mathbf{X}$ be a $n \times k$ matrix. The following statements are logically equivalent:

   i. The equation $\mathbf{Xb} = \mathbf{y}$ has a unique least-squares solution for each $\mathbf{y} \in \mathbb{R}^n$

   ii. the columns of A are linearly independent

   iii. the matrix $A^T A$ is invertible

When these statements are true, the least-squares solution $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Proof.

$$0 = \mathbf{X}^T(\mathbf{y} - \hat{\mathbf{y}}) \qquad \text{(by def. of orthogonal)}$$

$$0 = \mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \qquad \text{(plugging in for } \hat{\mathbf{y}})$$

$$0 = \mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} \qquad \text{(distributing)}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \qquad \text{(solving for } \hat{\boldsymbol{\beta}})$$

$\square$

The least-squares solution will satisfy this equation. If the matrix $\mathbf{X}$ is of full rank, then our estimate is the familiar

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \left(\frac{1}{n}\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}^T\mathbf{y}\right) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{y}_i\right)$$

(e) <u>Definition</u>. Let $\mathbf{X}$ be a $n \times k$ matrix of full rank and let $\mathbf{y} \in \mathbb{R}^n$. Then the matrix

$$P_x = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

is the **projection matrix** that projects $\mathbf{y}$ onto $\text{Col}(\mathbf{X})$.

(f) <u>Theorem</u> (Rudd SEC 2.6.2) Let P_x be the projection matrix where $P_x = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Then P_x is

   i. Symmetric

   ii. Idempotent

   iii. Positive Semidefinite

      Proof. $\mathbf{v}^T P_x \mathbf{v} = \mathbf{v}^T P_x^2 \mathbf{v} = (P_x\mathbf{v})^T(P_x\mathbf{v}) = ||P_x\mathbf{v}|| \geq 0$ $\square$

(g) <u>Definition</u>. Let P_x be the projection matrix associated with the columnn space of a $n \times k$ matrix X. Then the **annihilator matrix** is defined as

$$\mathbf{M}_x = I_n - \mathbf{P}_x$$

M_x projects values of $\mathbf{y} \in \mathbb{R}^n$ onto $\text{Col}(\mathbf{X})^\perp$.

Remark Given $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$,

   i. M_x is symmetric, idempotent, and positive semidefinite.

   ii. $M_x\mathbf{X} = \mathbf{0}$

   iii. $\hat{\mathbf{u}} = M_x\mathbf{y} = M_x\mathbf{u}$ and $\hat{\mathbf{u}}^T\hat{\mathbf{u}} = \mathbf{u}^T M_x\mathbf{u}$

   iv. $\hat{\mathbf{u}} \in \text{Nul}(\mathbf{X}^T)$ and $\mathbf{X}^T\hat{\mathbf{u}} = \mathbf{0}$ $(n - k$ degrees of freedom$)$

   v. $\text{rank}(M_x) = \text{tr}(M_x) = n - k$

(h) <u>Aside</u>. Depending on the presentation of OLS, the projection and annihilator matrices are important tools in developing the theory behind econometrics. Even if we focus on a more statistics-motivated approach to OLS, it is valuable to understand these matrices and what they do.

12. **Matrix Differentiation**

(a) <u>Aside</u>. We won't get into the details, but it is occasionally helpful to employ differentiation at a matrix level; for example, we can derive an estimator $\hat{\beta}$ for the regression equation $y = X\beta + \varepsilon$ by minimizing $\varepsilon'\varepsilon$ via matrix differentiation. Here are a few of the formulas you'll likely run across:

   i. Vector-by-Vector $y \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$

      A. In general:

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} \quad \text{and} \quad \frac{\partial y}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$$

      B. For a vector of constants $a \in \mathbb{R}$:

$$\frac{\partial a}{\partial \mathbf{x}} = \mathbf{0}_n$$

      C. For a derivative of a vector with respect to itself:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} = I_n$$

   ii. Matrix-by-Vector

      A. Linear form:

$$\frac{\partial A\mathbf{x}}{\partial \mathbf{x}^T} = A \qquad\qquad\qquad \text{(post-multiplied by } \mathbf{x})$$

      B. Quadratic form:

$$\frac{\partial \mathbf{x}^T A\mathbf{x}}{\partial \mathbf{x}^T} = \mathbf{x}^T(A + A^T) \qquad\qquad \text{(if } A \text{ is not symmetric)}$$

$$\frac{\partial \mathbf{x}^T A\mathbf{x}}{\partial \mathbf{x}^T} = 2\mathbf{x}^T A \qquad\qquad\qquad \text{(if } A \text{ is symmetric)}$$

(b) <u>Example</u>. Ordinary Least Square

$$\hat{\boldsymbol{\beta}} \in arg \min_{b \in \mathbb{R}^k} S(\mathbf{b}) \quad \text{where} \quad S(\mathbf{b}) = (\mathbf{y} - \mathbf{Xb})^T (\mathbf{y} - \mathbf{Xb})$$

$$\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}^T}\bigg|_{\mathbf{b} = \hat{\boldsymbol{\beta}}} = \mathbf{0}$$

$$2\hat{\boldsymbol{\beta}}^T (\mathbf{X}^T \mathbf{X}) - 2\mathbf{y}^T \mathbf{X} = 0$$
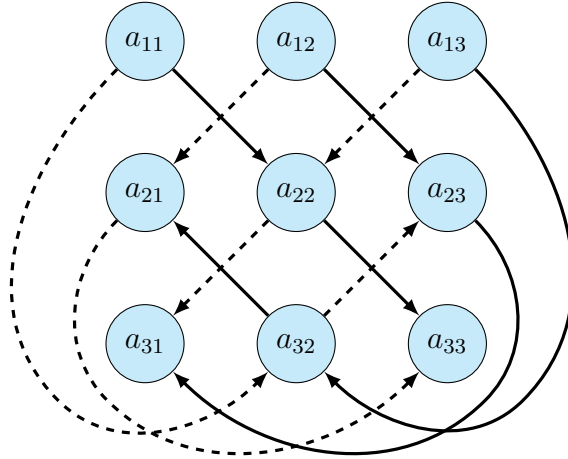
$$(\mathbf{X}^T \mathbf{X})\hat{\boldsymbol{\beta}} - \mathbf{X}^T \mathbf{y} = 0$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

13. **Appendix 1: Determinant**

(a) 3×3 Matrix: $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

First method, multiply three elements diagonally to the right and sum; multiply three elements diagonally to the left and subtract. Visually, (solid = positive, dashed = negative):


$$|\mathbf{A}| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{32}a_{23}$$

(b) For matrices 3×3 and larger, use the Laplace Expansion.

   i. <u>Definition</u>. The determinant of the matrix that results from deleting rows and columns associated with element a_{ij} is the **minor** $|M_{ij}|$.

   ii. <u>Example</u>. Consider again the 3×3 matrix A (defined above). The minor associated with the element a_{11} is $|M_{11}| = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$. Visually, this can be represented:
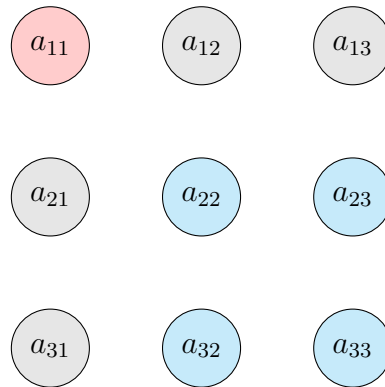
The minor associated with a_{22} is $|M_{22}| = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$
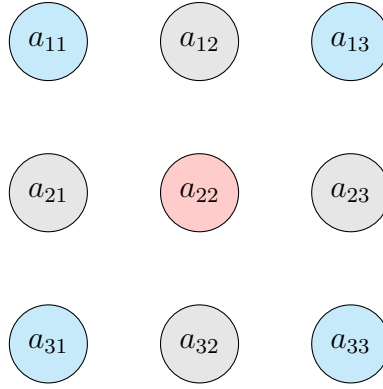
iii. <u>Definition</u>. The **cofactor** $|C_{ij}|$ is a minor modified by a prescribed algebraic sign that follows the convention:

$$|C_{ij}| \equiv (-1)^{i+j}|M_{ij}|$$

iv. <u>Definition</u>. The determinant found by **Laplace Expansion** is the value found by "expanding" along any row or column:

$$|A| = \sum_{j=1}^{n} a_{ij}|C_{ij}| \qquad \text{(expansion by the } i\text{th row)}$$

$$|A| = \sum_{i=1}^{n} a_{ij}|C_{ij}| \qquad \text{(expansion by the } j\text{th column)}$$

v. <u>Example</u>. Considering one again the 3×3 matrix C (defined above), performing the Laplace Expansion using the first row:

$$|\mathbf{A}| = a_{11}|C_{11}| + a_{12}|C_{12}| + a_{13}|C_{13}|$$

$$|\mathbf{A}| = (-1)^{1+1} \cdot a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{1+2} \cdot a_{12} \begin{vmatrix} a_{21} & a_{31} \\ a_{23} & a_{33} \end{vmatrix} + (-1)^{1+3} \cdot a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

vi. <u>Example</u>. Consider the 3×3 matrix $A = \begin{bmatrix} -7 & 0 & 3 \\ 9 & 1 & 4 \\ 0 & 6 & 5 \end{bmatrix}$. Expanding along the top row:

$$|A| = (-7) \begin{vmatrix} 1 & 4 \\ 6 & 5 \end{vmatrix} - (0) \begin{vmatrix} 9 & 4 \\ 0 & 5 \end{vmatrix} + (3) \begin{vmatrix} 9 & 1 \\ 0 & 6 \end{vmatrix}$$

$$= -7(5 - 24) + 0 + 3(54 - 0)$$

$$= 133 + 0 + 162 = 295$$

14. **Appendix 2: Inverse**

(a) <u>Example</u>. Suppose we have an invertible matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Then we could find the inverse of A by performing row operations on the augmented matrix

$$\begin{bmatrix} A \mid I_3 \end{bmatrix} = \left[\begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array} \right]$$

until we arrived at a matrix of the form

$$\begin{bmatrix} I_3 \mid A^{-1} \end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 & b_{31} & b_{32} & b_{33} \end{array} \right]$$

(b) <u>Theorem</u> (CW SEC 5.4) If A is a $n \times n$ invertible matrix, the inverse may be found by using the determinant and cofactors of the matrix A.

   i. <u>Definition</u>. The **cofactor matrix** of A is a matrix defined by replacing the elements of A with their associated cofactors:

$$C = \begin{bmatrix} |C_{11}| & |C_{12}| & \cdots & |C_{1n}| \\ |C_{21}| & |C_{22}| & \cdots & |C_{2n}| \\ \vdots & \vdots & \ddots & \vdots \\ |C_{n1}| & |C_{n2}| & \cdots & |C_{nn}| \end{bmatrix}$$

   ii. <u>Definition</u>. The **adjoint** of A is the transpose of the cofactor matrix:

$$\mathrm{adj}(A) = C^T = \begin{bmatrix} |C_{11}| & |C_{21}| & \cdots & |C_{n1}| \\ |C_{12}| & |C_{22}| & \cdots & |C_{n2}| \\ \vdots & \vdots & \ddots & \vdots \\ |C_{1n}| & |C_{2n}| & \cdots & |C_{nn}| \end{bmatrix}$$

The inverse of a $n \times n$ matrix A can be found via the formula:

$$A^{-1} = \frac{1}{|A|}\mathrm{adj}(A)$$

(c) <u>Example</u>. Consider the 3×3 matrix A:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 2 & 1 & 2 \\ 4 & 0 & 0 \end{bmatrix}$$

First, we can take the determinant to establish invertibility:

$$|A| = 0 \begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} - 1 \begin{vmatrix} 2 & 2 \\ 4 & 0 \end{vmatrix} + 0 \begin{vmatrix} 2 & 1 \\ 4 & 0 \end{vmatrix} \qquad \text{(expanding on 1st row)}$$

$$= 0 - 1(0 - 8) + 0 = 8 \qquad \text{(simplifying)}$$

Since $|A| \neq 0$, we can inverst this matrix. Next, find the cofactor matrix C_A:

$$C_A = \begin{bmatrix} \begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} & -\begin{vmatrix} 2 & 2 \\ 4 & 0 \end{vmatrix} & \begin{vmatrix} 2 & 1 \\ 4 & 0 \end{vmatrix} \\[12pt] -\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} & \begin{vmatrix} 0 & 0 \\ 4 & 0 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 4 & 0 \end{vmatrix} \\[12pt] \begin{vmatrix} 1 & 0 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 0 & 0 \\ 2 & 2 \end{vmatrix} & \begin{vmatrix} 0 & 1 \\ 2 & 1 \end{vmatrix} \end{bmatrix} \qquad \text{(the cofactor matrix)}$$

$$C_A = \begin{bmatrix} 0 & 8 & -4 \\ 0 & 0 & 4 \\ 2 & 0 & -2 \end{bmatrix} \qquad \text{(simplifying)}$$

Recall that the adjoint of A is the transpose of C_A:

$$\text{adj}(A) = C_A^T \qquad \text{(by def. of the adjoint)}$$

$$\text{adj}(A) = \begin{bmatrix} 0 & 0 & 2 \\ 8 & 0 & 0 \\ -4 & 4 & -2 \end{bmatrix} \qquad \text{(the adjoint matrix)}$$

Finally, we can calculate the inverse of A:

$$A^{-1} = \frac{1}{|A|}\text{adj}(A) \qquad \text{(the inverse formula)}$$

$$= \frac{1}{8}\begin{bmatrix} 0 & 0 & 2 \\ 8 & 0 & 0 \\ -4 & 4 & -2 \end{bmatrix} \qquad \text{(plugging in values)}$$

$$A^{-1} = \begin{bmatrix} 0 & 0 & 1/4 \\ 1 & 0 & 0 \\ -1/2 & 1/2 & -1/4 \end{bmatrix} \qquad \text{(the inverse)}$$

15. **Appendix 3: Cramer's Rule**

(a) <u>Example</u> (CW SEC 5.6). Consider a two-commodity, linear market model. For the first market:

$$Q_{d1} = a_0 + a_1 P_1 + a_2 P_2 \qquad \text{(demand 1)}$$

$$Q_{s1} = b_0 + b_1 P_1 + b_2 P_2 \qquad \text{(supply 1)}$$

$$0 = Q_{d1} - Q_{s1} \qquad \text{(S = D)}$$

For the second:

$$Q_{d2} = \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 \qquad \text{(demand 2)}$$

$$Q_{s2} = \beta_0 + \beta_1 P_1 + \beta_2 P_2 \qquad \text{(supply 2)}$$

$$0 = Q_{d2} - Q_{s2} \qquad \text{(S = D)}$$

This reduces into a two-equation, two-unknown system:

$$0 = (a_0 - b_0) + (a_1 - b_1) P_1 + (a_2 - b_2) P_2 \qquad \text{(for market 1)}$$

$$0 = (\alpha_0 - \beta_0) + (\alpha_1 - \beta_1) P_1 + (\alpha_2 - \beta_2) P_2 \qquad \text{(for market 2)}$$

Rewriting for expediency, defining $c_i = a_i - b_i$ (and analogously for greek letters):

$$c_1 P_1 + c_2 P_2 = -c_0 \qquad \text{(market 1)}$$

$$\gamma_1 P_1 + \gamma_2 P_2 = -\gamma_0 \qquad \text{(market 2)}$$

Rewriting in matrix notation:

$$\begin{bmatrix} c_1 & c_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} -c_0 \\ -\gamma_0 \end{bmatrix} \qquad \text{(the linear system)}$$

Solving for P_1^* (the market clearing price for good 1):

$$P_1^* = \frac{\begin{vmatrix} -c_0 & c_2 \\ -\gamma_0 & \gamma_2 \end{vmatrix}}{\begin{vmatrix} c_1 & c_2 \\ \gamma_1 & \gamma_2 \end{vmatrix}} = \frac{c_2 \gamma_0 - c_0 \gamma_2}{c_1 \gamma_2 - c_2 \gamma_1} \qquad \text{(by Cramer's Rule)}$$

Solving for P_2^* (the market clearing price for good 2):

$$P_2^* = \frac{\begin{vmatrix} c_1 & -c_0 \\ \gamma_1 & -\gamma_0 \end{vmatrix}}{\begin{vmatrix} c_1 & c_2 \\ \gamma_1 & \gamma_2 \end{vmatrix}} = \frac{c_0 \gamma_1 - c_1 \gamma_0}{c_1 \gamma_2 - c_2 \gamma_1} \qquad \text{(by Cramer's Rule)}$$