Math Camp 2025 – Linear Algebra[*]

Camilo Abbate and Sofia Olguin [†]

Department of Economics, UC Santa Barbara

August 19, 2025

1 Vectors

A vector can be defined in many different ways, but let us broadly define a vector as a list of objects that is used to represent and conceptualize ideas.

Definition 1.1. A **column vector** is a $n \times 1$ matrix:[1]

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

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

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

Example 1.2. Consider 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.[2]
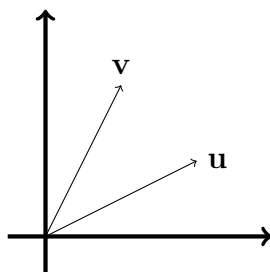
Definition 1.3. Two vectors in $\mathbb{R}^n$ are **equal** if and only if their corresponding elements are equal.

Example 1.4. Our vectors from above, $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, are ⎯⎯⎯⎯⎯⎯.

---

[*]This lecture notes are for personal use only and is not intended for reproduction, distribution, or citation.

[†]This lecture notes were prepared for the 2025 UCSB Math Camp for Ph.D. students in economics. It incorporates materials from previous instructors, including Seonmin Will Heo, Eunseo Kang, and James Banovetz.

[1]A matrix is a rectangular array of elements, arranged in rows and columns.

[2]Alternatively, vectors can be thought of as arrows that point from the origin to a specific point in the plane.

Definition 1.5. Given two vectors $\mathbf{u}$ and $\mathbf{v}$, their **sum** is the vector $\mathbf{u} + \mathbf{v}$, obtained by adding the corresponding elements (assuming the vectors "live" in the same space).

Example 1.6. Consider the vectors $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

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

Definition 1.7. 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.

Example 1.8. Consider the vector $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and the scalar $c = 3$. Then

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

Aside. Try to imagine each element as a scalar. For example, a vector $[2, 3]$ can be represented as $2\mathbf{i} + 3\mathbf{j}$, where $\mathbf{i}$ and $\mathbf{j}$ are the "basis vectors" of the xy coordinate system. A basis is a minimal set of vectors that can be used to uniquely express any other vector within a given vector space.

Definition 1.9. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **parallel** if and only if there exists a real number $c \in \mathbb{R} \backslash \{0\}$ such that

$$\mathbf{u} = c\mathbf{v}$$

(*Basically if one is a scalar multiple of the other.*)

Definition 1.10. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ and scalars $c_1, c_2, \cdots, c_k$, 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 $\mathbf{v}_1, \cdots, \mathbf{v}_k$ with weights $c_1, \cdots, c_k$.

Definition 1.11. If $\mathbf{v}_1, \cdots, \mathbf{v}_k \in \mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \cdots, \mathbf{v}_k$ is called **span** of $\mathbf{v}_1, \cdots, \mathbf{v}_k$. That is, the span is the collection of all vectors that can be written in the form

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

Definition 1.12. Let $\mathbb{W} \subseteq \mathbb{R}^n$. Then $\mathbf{v}_1, \cdots, \mathbf{v}_k \in \mathbb{R}^n$ spans $\mathbb{W}$ if given any $\mathbf{w} \in \mathbb{W}$,

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

Definition 1.13. A set of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k \in \mathbb{R}^n$ 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 vectors are **linearly independent**.

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

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

has only the trivial solution.

Definition 1.14. trivial solution

From the vector equation

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

it is clear that $c_1 = 0, c_2 = 0, \ldots, c_k = 0$ is a solution to such a system; it is called the trivial solution. Any solution in which at least one coefficient has nonzero value is called a non-trivial solution.

Aside. Let us assume that $\mathbf{v}_1, \cdots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly dependent. WLOG, there exists an $i \in \{1, \cdots, k\}$ such that $c_i \neq 0$. Then

$$c_i\mathbf{v}_i + \sum_{j=1,\ j\neq i}^{k} c_j\mathbf{v}_j = \mathbf{0}_n \quad \text{and} \quad \mathbf{v}_i = -\sum_{j=1,\ j\neq i}^{k} \frac{c_j}{c_i}\mathbf{v}_j$$

Example 1.15. 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.

Definition 1.16. $\mathbf{v}_1, \cdots, \mathbf{v}_k \in \mathbb{W} \subset \mathbb{R}^n$ forms a **basis** of $\mathbb{W}$ if

1. $\mathbf{v}_1, \cdots, \mathbf{v}_k$ spans $\mathbb{W}$, and

2. $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly independent.

Example 1.17. 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.

Aside. At this point, I highly encourage everyone to watch "Lecture 3: Essence of linear algebra" by 3b1b in case you haven't watched it. If you have ever been confused about some notions in linear algebra, this video may provide you a different way to interpret and understand them.

2   Systems of Linear Equations and Matrix Equations

Definition 2.1. A **linear equation** in the variables $x_1, x_2, \cdots, x_n$ 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.

Example 2.2. Consider the following system of linear equations:

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

Definition 2.3. The compact rectangular array containing the coefficients of each variable aligned in columns is called the **coefficient matrix**.

Example 2.4. From the system of linear equations above, the coefficient matrix is

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

Definition 2.5. The coefficient matrix, concatenated with a column consisting of the right-hand-side constants is called the **augmented matrix**.

Example 2.6. From the system of linear equations above, the augmented matrix is

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

Definition 2.7. The **size** of a matrix tells how many rows and columns it has. An $m \times n$ matrix has m rows and n columns:

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

Definition 2.8. The **element** a_{ij} is the value in row i, column j.

Definition 2.9. The three basic operations that are used to solve the system of equations 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

Example 2.10. Consider the augmented matrix from our system above:

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{bmatrix} \quad \rightarrow \text{Replacing } R_3 \text{ with } R_3 + 4R_1: \rightarrow \quad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{bmatrix}$$

Multiplying R_2 by $1/2$

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & -3 & 13 & -9 \end{bmatrix} \quad \rightarrow \text{Replacing } R_3 \text{ with } R_3 + 3R_2: \rightarrow \quad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Replacing R_1 with $2R_2 + R_1$:

$$\begin{bmatrix} 1 & 0 & -7 & 8 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix} \quad \rightarrow \text{Replacing } R_1 \text{ with } 7R_3 + R_1: \rightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Replacing R_2 with $4R_3 + R_2$:

$$\begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix} \quad \rightarrow \text{The simplified system:} \rightarrow \quad \begin{matrix} x_1 & & & = & 29 \\ & x_2 & & = & 16 \\ & & x_3 & = & 3 \end{matrix}$$

Aside. 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.

Definition 2.11. If $\mathcal{A}$ is an $m \times n$ matrix with columns $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathbb{R}^m$, and if $\mathbf{x} \in \mathbb{R}^n$, then the **product** of $\mathcal{A}$ and $\mathbf{x}$, denoted $\mathcal{A}\mathbf{x}$, is the linear combination of the columns of $\mathcal{A}$ using the corresponding entries in $\mathbf{x}$ as weights

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

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

Theorem 2.12. *If $\mathcal{A}$ 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 $\mathcal{A}\mathbf{x} = \mathbf{b}$ has the same solution set as the linear equations whose augmented matrix is*

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

Example 2.13. 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 $\mathcal{A}$ be the coefficient matrix:

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

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

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

Aside. We will work with matrix equations quite a bit during the first year, especially when we learn a regression framework. However, we will almost never apply elementary row operations to convert matrices into the reduced row echelon form by hand. We are revisiting concepts typically taught in a linear algebra course, and we would be remiss not to understand them thoroughly. The key takeaway is that these elementary row operations do not alter the solution set. It is a good time to start thinking about how the matrix equations are written and what they mean.

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

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

---

3 Matrix Operations

Definition 3.1. Let $\mathcal{A}$ and $\mathcal{B}$ be $m \times n$ matrices. Let a_{ij} and b_{ij} denote the elements in the ith row and jth column of the matrices $\mathcal{A}$ and $\mathcal{B}$, respectively. Then $\mathcal{A}$ and $\mathcal{B}$ are **equal**, denoted $\mathcal{A} = \mathcal{B}$, if and only if $a_{ij} = b_{ij}$ for all i and all j.

Aside. 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.

Definition 3.2. If $\mathcal{A}$ and $\mathcal{B}$ are both $m \times n$ matrices, then the **sum** $\mathcal{A} + \mathcal{B}$ is the $m \times n$ matrix whose elements are the sums of the corresponding elements from $\mathcal{A}$ and $\mathcal{B}$. The **difference** $\mathcal{A} - \mathcal{B}$ is the $m \times n$ matrix whose elements are the differences of the corresponding elements of $\mathcal{A}$ and $\mathcal{B}$.

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

Definition 3.3. If k is a scalar and A is a matrix, then the **scalar multiple** $k\mathcal{A}$ is the matrix whose elements are k times the corresponding elements in $\mathcal{A}$.

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

Definition 3.4. If $\mathcal{A}$ is a $m \times n$ matrix and $\mathcal{B}$ is a $n \times k$ matrix with columns $\mathbf{b}_1, \cdots, \mathbf{b}_k$, then the **product** $\mathcal{A}\mathcal{B}$ is the $m \times k$ matrix whose columns are $\mathcal{A}\mathbf{b}_1, \cdots, \mathcal{A}\mathbf{b}_k$. Note that the number of columns of $\mathcal{A}$ must match the number of rows in $\mathcal{B}$ for the matrices to be conformable.

Example 3.5. Suppose we have two matrices: $\mathcal{A}_{1\times2}$ and $\mathcal{B}_{2\times3}$. Then $\mathcal{A}\mathcal{B}$ is permissible, but $\mathcal{B}\mathcal{A}$ is not. If we have $\mathcal{A}_{1\times2}\mathcal{B}_{2\times3} = \mathcal{C}_{1\times3}$, then

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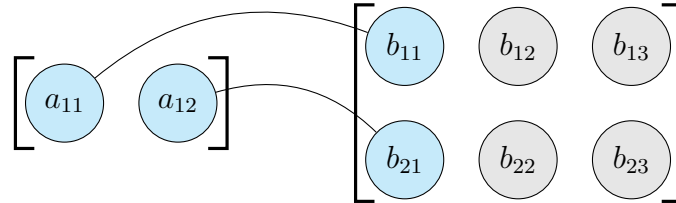
- For c_{12}: $c_{12} = a_{11}b_{12} + a_{12}b_{22}$
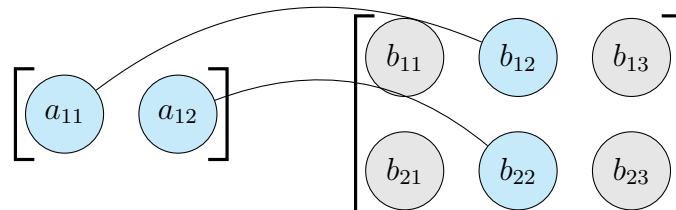
Aside. There are a few special matrices with which you should be familiar. For example:

1. **Identity Matrix** (a diagonal matrix of ones)

$$\mathcal{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

   Note that for all n by n matrices $\mathcal{A}$:

$$\mathcal{A}\mathcal{I}_n = \mathcal{I}_n\mathcal{A} = \mathcal{A}$$

2. **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 $\mathcal{A}$ and "conformable" matrices $\mathbf{0}$:

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

3. **Idempotent Matrix** (a matrix that is the product of itself)

$$\mathcal{A} = \mathcal{A}\mathcal{A}$$

Theorem 3.6. *Let $\mathcal{A}$ be a $m \times n$ matrix, and let $\mathcal{B}$ and $\mathcal{C}$ have dimensions for which the indicated sums and products are defined. Then*

$$\mathcal{A}(\mathcal{B}\mathcal{C}) = (\mathcal{A}\mathcal{B})\mathcal{C} = \mathcal{A}\mathcal{B}\mathcal{C} \qquad\qquad \textit{(Associative Law)}$$
$$\mathcal{A}(\mathcal{B} + \mathcal{C}) = \mathcal{A}\mathcal{B} + \mathcal{A}\mathcal{C} \qquad\qquad \textit{(Left Distributive Law)}$$
$$(\mathcal{B} + \mathcal{C})\mathcal{A} = \mathcal{B}\mathcal{A} + \mathcal{C}\mathcal{A} \qquad\qquad \textit{(Right Distributive Law)}$$

Aside. Note that in matrix algebra, we *do not* have commutativity; that is, $\mathcal{A}\mathcal{B} \neq \mathcal{B}\mathcal{A}$ in general. It may be true of specific matrices, but in most cases (as in an example above), the matrices won't even be conformable in both orders.

Definition 3.7. Given a $m \times n$ matrix $\mathcal{A}$, the **transpose** of $\mathcal{A}$ is the $n \times m$ matrix, denoted $\mathcal{A}^T$ or $\mathcal{A}'$, whose columns are formed from the corresponding rows of $\mathcal{A}$.

Example 3.8. Consider the following matrices:

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

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

Remark. A property of matrices that occasionally comes in handy is **symmetry**, which means $\mathcal{A} = \mathcal{A}^T$ (or $a_{ij} = a_{ji}, \ \forall i, j$). Later on, we'll talk about Hessian matrices, which have this property.

Theorem 3.9. *Let $\mathcal{A}$ and $\mathcal{B}$ denote matrices whose sizes are appropriate for the following sums and products.*

$$(\mathcal{A}^T)^T = \mathcal{A}$$
$$(\mathcal{A} + \mathcal{B})^T = \mathcal{A}^T + \mathcal{B}^T$$
$$(\mathcal{A}\mathcal{B})^T = \mathcal{B}^T \mathcal{A}^T$$

Definition 3.10. The **determinant** of a square matrix $\mathcal{A}$, denoted by $|\mathcal{A}|$ or $\det(\mathcal{A})$, is a uniquely defined scalar associated with the matrix. For matrices of various sizes:

1. 1×1 Matrix: $\mathcal{A} = [a]$

$$|\mathcal{A}| = a$$

2. 2×2 Matrix: $\mathcal{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$|\mathcal{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

3. Larger Matrix: Check the appendix 1.

Theorem 3.11.   *1. Taking the transpose does not affect the determinant:*

$$|\mathcal{A}| = |\mathcal{A}^T|$$

2. Scaling a row by k will change the value of the determinant k-fold, e.g.:

$$\begin{vmatrix} ka & kb \\ c & d \end{vmatrix} = k \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

3. Replacement – adding a multiple of a row (or column) to another row (or column) – leaves the determinant unchanged, e.g.:

$$\begin{vmatrix} a & b \\ c + ka & d + kb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

4. Interchanging any two rows (or columns) reverses the sign of the determinant (but does not change the absolute value), e.g.:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = - \begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

Definition 3.12. A matrix $\mathcal{A}$ is **singular** if and only if

$$|\mathcal{A}| = 0$$

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).

Definition 3.13. A $n \times n$ matrix is said to be **invertible** if there is an $n \times n$ matrix $\mathcal{C}$ such that

$$\mathcal{C}\mathcal{A} = \mathcal{I}_n \quad \text{and} \quad \mathcal{A}\mathcal{C} = \mathcal{I}_n$$

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

Theorem 3.14. *If $\mathcal{A}$ is a 2×2 matrix*

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

and $ad - bc \neq 0$ (i.e., the determinant is non-zero/the matrix is _____), then the inverse is given by the formula:

$$\mathcal{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Check the appendix 2 for a larger matrix.

Example 3.15. Consider the matrix

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

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

$$
\begin{aligned}
|\mathcal{A}| &= -2 && \text{(the determinant)} \\
\mathcal{A}^{-1} &= \frac{1}{-2} \begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} && \text{(by THM 2.4)} \\
\mathcal{A}^{-1} &= \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix} && \text{(simplifying)}
\end{aligned}
$$

Checking to make sure we did everything properly:

$$
\begin{aligned}
\mathcal{A}\mathcal{A}^{-1} &= \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix} && \text{(multiplying matrices)} \\
\mathcal{A}\mathcal{A}^{-1} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} && \text{(simplifying)}
\end{aligned}
$$

Theorem 3.16. *If $\mathcal{A}$ and $\mathcal{B}$ are $n \times n$ invertible matrices, then*

1. *$\mathcal{A}^{-1}$ is invertible and*
$$(\mathcal{A}^{-1})^{-1} = \mathcal{A}$$

2. *$\mathcal{A}\mathcal{B}$ is invertible and*
$$(\mathcal{A}\mathcal{B})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}$$

3. *$\mathcal{A}^T$ is invertible and*
$$(\mathcal{A}^T)^{-1} = (\mathcal{A}^{-1})^T$$

Example 3.17. Given a $n \times k$ matrix $\mathbf{X}$ where $n > k$, assume that $\mathbf{X}^T\mathbf{X}$ is invertible.

$$
\begin{aligned}
\left(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)^T &= (\mathbf{X}^T)^T\left((\mathbf{X}^T\mathbf{X})^{-1}\right)^T\mathbf{X}^T \\
&= \mathbf{X}\left((\mathbf{X}^T\mathbf{X})^T\right)^{-1}\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T && \text{(Symmetric)} \\
\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T &= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathcal{I}_k\mathbf{X}^T \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T && \text{(Idempotent)}
\end{aligned}
$$

4 Connecting Concepts

Theorem 4.1. *Let $\mathcal{A}$ be a $n \times n$ matrix. Then the following statements are equivalent:*

1. *$|\mathcal{A}| \neq 0$*

2. *$\mathcal{A}$ is invertible*

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

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

5. *The columns of $\mathcal{A}$ form a linearly independent set*

6. *The columns of $\mathcal{A}$ span $\mathbb{R}^n$*

7. *$\mathcal{A}$ has full rank*

Aside. 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:

1. Standard substitution (potentially very time consuming)

2. Row operations on an augmented matrix

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

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.

Theorem 4.2. *Given a system of equations $\mathcal{A}\mathbf{x} = \mathbf{b}$, where $\mathcal{A}$ is a $n \times n$ invertible matrix and $\mathbf{b} \in \mathbb{R}$, let $\mathcal{A}_i(\mathbf{b})$ be the matrix obtained from A by replacing column i with $\mathbf{b}$:*

$$\mathcal{A}_i(\mathbf{b}) = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \dots & \mathbf{a}_n \end{bmatrix}$$

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

$$x_i = \frac{|\mathcal{A}_i(\mathbf{b})|}{|A|}, \quad i = 1, 2, \dots, n$$

Check the Appendix 3(Cramer's Rule) for the example in economics.

Definition 4.3. The **rank** of a matrix is the number of nonzero rows in its row echelon form.

Theorem 4.4. *Let $\mathcal{A}$ be a $m \times n$ matrix and $\mathcal{B}$ be a $n \times k$ matrix. Then*

1. *$rank(\mathcal{A}) = rank(\mathcal{A}^T)$*

2. *$rank(\mathcal{A}) \leq \min\{m, n\}$*

3. *$rank(\mathcal{A}\mathcal{B}) \leq \min\{rank(\mathcal{A}), rank(\mathcal{B})\}$*

4. *$rank(\mathcal{A}) = rank(\mathcal{A}^T\mathcal{A}) = rank(\mathcal{A}\mathcal{A}^T)$*

Aside. Given a $n \times k$ matrix $\mathbf{X}$ where $n > k$, suppose that $\text{rank}(\mathbf{X}) = k$ (full rank). Then

$$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}^T\mathbf{X}) = k$$

Therefore, $\mathbf{X}^T\mathbf{X}$ is invertible.

5   Quadratic Forms

Definition 5.1. 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

$$Q(\mathbf{x}) = \mathbf{x}^T \mathcal{A} \mathbf{x}$$

where $\mathcal{A}$ is a $n \times n$ symmetric matrix.

Example 5.2. Consider a two-variable quadratic form

$$Q(\mathbf{x}) = ax^2 + 2bxy + cy^2 \qquad \text{(a two-variable quadratic form)}$$

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

Note that the matrix of constants($\mathcal{A}$) is not unique!

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

Aside. The matrix associated with a quadratic form need not be symmetric. However, no loss of generality is obtained by assuming $\mathcal{A}$ is symmetric. We can always take definite and semidefinite matrices to be symmetric since they are defined by a quadratic form. Specifically consider a nonsymmetric matrix $\mathcal{A}$ and replace it with $\frac{1}{2}(\mathcal{A} + \mathcal{A}^T)$. Then $\mathcal{A}$ is now symmetric and still it has the same $Q(\mathbf{x})$.

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

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

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

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

Definition 5.3. Let $Q(\mathbf{x})$ be a quadratic form. Then $Q(\mathbf{x})$ is:

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

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

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

Quadratic forms can also be **positive semi-definite** or **negative semi-definite** if we replace the strict inequalities with weak inequalities. The matrix $\mathcal{A}$ that satisfies certain definite or semi-definite condition in Definition 6.3 is called a **positive (semi)definite/negative (semi)definite matrix**.

Example 5.4. Consider the quadratic form:

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

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

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

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

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

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

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

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

Example 5.5. We can prove that a positive definite matrix $\mathcal{A}_{n \times n}$ is nonsingular. In other words,

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

An easy way to do this is to prove by contradiction(from the logic class).

- $P \implies Q$ 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: $\forall \mathbf{x} \neq \mathbf{0}, \mathbf{x}^T \mathcal{A} \mathbf{x} > 0$

- Theorem 4.1: if $\mathcal{A}$ is singular, $\exists \, \mathbf{x} \neq \mathbf{0} \ni \mathcal{A} \mathbf{x} = \mathbf{0}$

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

Proof.

$$\begin{aligned}
&\text{Let } \mathcal{A} \text{ be positive definite} && \text{(by hypothesis)} \\
&\implies \forall \, \mathbf{x} \neq \mathbf{0}, \mathbf{x}^T \mathcal{A} \mathbf{x} > 0 && \text{(by def. of positive definite)} \\
&\text{Suppose that } \mathcal{A} \text{ is singular} && \text{(towards a contradiction)} \\
&\implies \exists \, \mathbf{y} \neq \mathbf{0} \ni \mathcal{A} \mathbf{y} = \mathbf{0} && \text{(by Theorem 5.1)} \\
&\implies \mathbf{y}^T \mathcal{A} \mathbf{y} = 0 && \text{(premultiplying by } \mathbf{y}^T) \\
&\text{Thus, a contradiction} && \\
&\implies \mathcal{A} \text{ is nonsingular} &&
\end{aligned}$$

$\square$

Definition 5.6. Let $\mathcal{A}$ be a $n \times n$ matrix with elements a_{ij}. Then a **principal minor** is a minor[3] produced by the deletion of the row and column associated with an element such that $i = j$. A n**th-order principal minor** is produced by the deletion of rows and columns associated with elements such that $i = j$, leaving a nth order determinant.

Example 5.7. Consider the matrix

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

---

[3]A minor of a matrix is the determinant of a submatrix

The 1st-order principal 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 principal 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 principal minor is:

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

Definition 5.8. The **leading principal minors** are the principal minors that are associated with contiguous, upper-left elements of the matrix.

Example 5.9. For the 3×3 matrix $\mathcal{A}$ in the preceding example, the leading principal minors are

$$|\mathcal{A}_1| = |a_{11}| \qquad \text{(the 1st-order LPM)}$$

$$|\mathcal{A}_2| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad \text{(the 2nd-order LPM)}$$

$$|\mathcal{A}_3| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \qquad \text{(the 3rd-order LPM)}$$

Aside. 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 principal minors, however, with element a_{11}, which is the upper-left corner of *every* leading principal minor.

Theorem 5.10. *Let A be a* $n \times n$ *symmetric matrix. then the quadratic form* $\mathbf{x}^T \mathcal{A} \mathbf{x}$ *is:*

- *positive definite if and only if all of its* leading principal minors *are strictly positive*

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

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

$$|\mathcal{A}_1| < 0 \qquad |\mathcal{A}_2| > 0 \qquad |\mathcal{A}_3| < 0 \dots$$

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

- *indefinite if none of the above hold*

Example 5.11. Consider the matrix
$$\mathcal{A} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 6 & -2 \\ 0 & -2 & 3 \end{bmatrix}$$

We can determine the definiteness of the matrix by first evaluating the leading principal 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 principal minors was equal to zero, we'd need to check all of the other PMs for positive semi-definiteness.

Below (connecting definiteness to Hessian) is going to be covered in the Analysis session.

Aside. 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:

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

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

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

We can re-write this in matrix notation:
$$d^2 z = \begin{bmatrix} dx & dy \end{bmatrix} \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} \qquad \text{(the quadratic form)}$$

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

Theorem 5.12. *Let f be a twice-continuously differentiable function then if the associated Hessian is*

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

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

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

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

Basically, the intuition 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 $d^2 z$ 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.

6 Eigenvalues and Eigenvectors

Definition 6.1. For an $n \times n$ matrix $\mathcal{A}$, the **trace** is the sum of the elements along the main diagonal:

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

Example 6.2. Consider the identity matrix:

$$\mathcal{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The trace is the sum along the main diagonal, so $\text{tr}(\mathcal{I}_3) = 3$ (indeed, the trace of any identity matrix is the dimensionality of that matrix).

Theorem 6.3. *Traces have the following properties:*

1. *(Linearity) Let $\mathcal{A}$ and $\mathcal{B}$ be $n \times n$ matrices and $c \in \mathbb{R}$.*

$$tr(\mathcal{A} + \mathcal{B}) = tr(\mathcal{A}) + tr(\mathcal{B})$$
$$c \cdot tr(\mathcal{A}) = tr(c\mathcal{A})$$

2. *(Transpose) Let $\mathcal{A}$ be a $n \times n$ matrix.*

$$tr(\mathcal{A}^T) = tr(\mathcal{A})$$

3. *(Product) Let $\mathcal{A}$ be a $m \times n$ matrix and let $\mathcal{B}$ be a $n \times m$ matrix.*

$$tr(\mathcal{A}\mathcal{B}) = tr(\mathcal{B}\mathcal{A})$$

Example 6.4. Given a $n \times k$ matrix $\mathbf{X}$ where $n > k$, suppose that $\text{rank}(\mathbf{X}) = k$ (full rank). Then

$$\text{tr}\left(\mathcal{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right) = \text{tr}\left(\mathcal{I}_n\right) - \text{tr}\left(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right) \quad \underline{\hspace{3cm}}$$
$$= n - \text{tr}\left(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right) \quad \underline{\hspace{3cm}}$$
$$= n - \text{tr}(\mathcal{I}_k) \quad \underline{\hspace{3cm}}$$
$$= n - k \quad \underline{\hspace{3cm}}$$

Example 6.5. Given a vector $\widehat{\mathbf{u}} \in \mathbb{R}_n$,
$$\widehat{\mathbf{u}}^T\widehat{\mathbf{u}} = \text{tr}(\widehat{\mathbf{u}}^T\widehat{\mathbf{u}})$$

Definition 6.6. An **eigenvector** of a $n \times n$ matrix $\mathcal{A}$ 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:
$$(A - \lambda I_n)\mathbf{x} = \mathbf{0}$$

Example 6.7. Consider the matrix and vector:

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

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

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

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

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

$$\mathcal{A}\mathbf{u} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix} \qquad \text{(factoring)}$$

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

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

Definition 6.8. For a $n \times n$ matrix $\mathcal{A}$, the scalar equation

$$|(\mathcal{A} - \lambda I_n)| = 0$$

is called the **characteristic equation** of $\mathcal{A}$.

Theorem 6.9. *A scalar λ is an eigenvalue of a $n \times n$ matrix $\mathcal{A}$ if and only if λ satisfies the charactersitic equation of $\mathcal{A}$.*

Example 6.10. 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.

Example 6.11. Consider the matrix

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

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

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

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

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

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

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

Thus, $\lambda_1 = 3$ and $\lambda_2 = -2$ are the eigenvalues. To find associated eigenvectors, recall that we need $(\mathcal{A} - \lambda I_2)\mathbf{x} = \mathbf{0}_2$. Thus, for our first eigen value:

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

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

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

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):

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

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

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

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

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.

Definition 6.12. A square $n \times n$ matrix $\mathcal{A}$ is called **diagonalizable** if there exists an $n \times n$ invertible matrix $\mathcal{P}$ such that $\mathcal{P}^{-1}\mathcal{A}\mathcal{P}$ is a diagonal matrix.

$$A : \text{diagonalizable} \iff \exists P : P^{-1}AP \text{ is diagonal}$$

Theorem 6.13. *Let $\mathcal{A}$ be a $n \times n$ matrix. If $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of $\mathcal{A}$, then:*

1. *The trace equals the sum of the eigenvalues*

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

2. *The determinant of A equals the product of the eigenvalues*

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

18

3. The trace of $\mathcal{A}^k$ is the sum of the eigenvalues raised to the kth power

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

4. Let $\mathcal{A}^T = \mathcal{A}$.

   (a) Eigenvalues of $\mathcal{A}$ are real.

   (b) Eigenvectors corresponding to distinct eigenvalues are pairwise orthogonal.

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

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

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

   $\square$

   (c) $\mathcal{A}$ can be diagonalized.
   $$\mathcal{A}X = X\Lambda \quad where \quad X^T X = I_n$$
   $$\mathcal{A} = X\Lambda X^T$$

5. Let $\mathcal{A}^T = \mathcal{A}$.

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

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

   (b) Negative (semi)definite if and only if the eigenvalues of $\mathcal{A}$ are all (weakly) negative.

   (c) Indefinite if and only if $\mathcal{A}$ has both positive and negative eigenvalues.

6. Let $\mathcal{A}^T = \mathcal{A}$ and $\mathcal{A}$ is idempotent.

   (a) Eigenvalues of $\mathcal{A}$ are either 0 or 1.

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

   (b) $\mathcal{A}$ is positive semidefinite.

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

   (c) $rank(\mathcal{A}) = tr(A)$.

7. Let $\mathcal{A}^T = \mathcal{A}$ and $\mathcal{A}$ is positive definite. Then there exists C such that $\mathcal{A} = C^T C$.

   Proof. $\mathcal{A} = X\Lambda X^T = X\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}X^T = (\Lambda^{\frac{1}{2}}X^T)^T(\Lambda^{\frac{1}{2}}X^T)$ $\square$

Example 6.14. Let X be a $n \times k$ matrix with rank($\mathbf{X}$)$= k$ ($n \geq k$).

$$\mathcal{M}_{\mathbf{X}} := \mathcal{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$
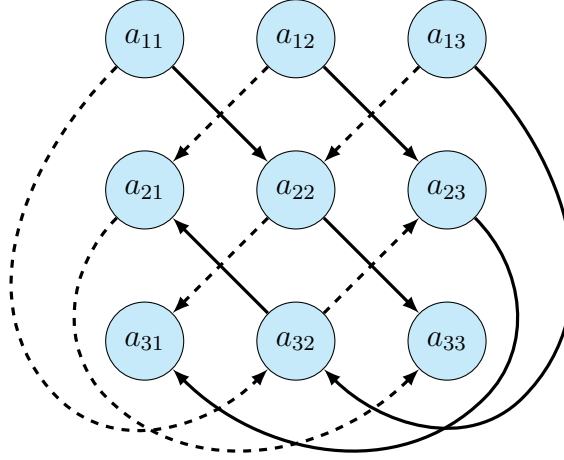
1. $\mathcal{M}_{\mathbf{X}}$ is symmetric.

2. $\mathcal{M}_{\mathbf{X}}$ is idempotent.

7 Appendices

7.1 Appendix 1: determinant

1. 3×3 Matrix: $\mathcal{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

   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):


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

2. For matrices 3×3 and larger, use the Laplace Expansion.

   Definition 7.1. The determinant of the matrix that results from deleting rows and columns associated with element a_{ij} is the **minor** $|M_{ij}|$.

   Example 7.2. 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:
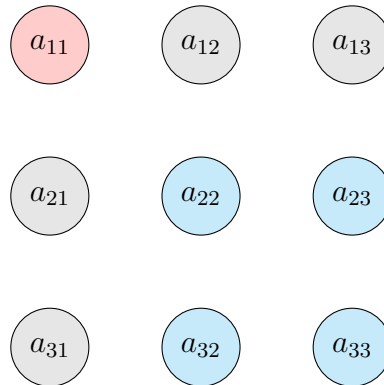
   The minor associated with a_{22} is $|M_{22}| = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$
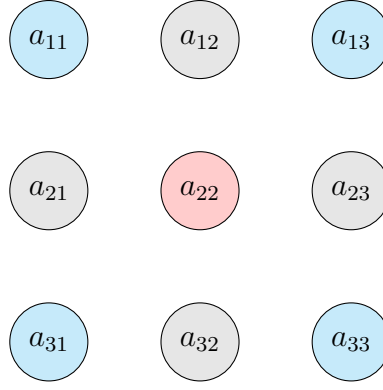
Definition 7.3. The **cofactor** $|C_{ij}|$ is a minor modified by a prescribed algebraic sign that follows the convention:

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

Definition 7.4. The determinant found by **Laplace Expansion** is the value found by "expanding" along any row or column:

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

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

Example 7.5. Considering one again the 3×3 matrix C (defined above), performing the Laplace Expansion using the first row:

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

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

7.2   Appendix 2: inverse

Example 7.6. Suppose we have an invertible matrix

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

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

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

until we arrived at a matrix of the form

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

Theorem 7.7. *If $\mathcal{A}$ is a $n \times n$ invertible matrix, the inverse may be found by using the determinant and cofactors of the matrix $\mathcal{A}$.*

Definition 7.8. The **cofactor matrix** of $\mathcal{A}$ is a matrix defined by replacing the elements of $\mathcal{A}$ with their associated cofactors:

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

Definition 7.9. The **adjoint** of $\mathcal{A}$ is the transpose of the cofactor matrix:

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

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

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

Example 7.10. Consider the 3×3 matrix $\mathcal{A}$:

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

First, we can take the determinant to establish invertibility:

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

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

Since $|\mathcal{A}| \neq 0$, we can invert this matrix. Next, find the cofactor matrix $C_\mathcal{A}$:

$$C_\mathcal{A} = \begin{bmatrix} \begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} & -\begin{vmatrix} 2 & 2 \\ 4 & 0 \end{vmatrix} & \begin{vmatrix} 2 & 1 \\ 4 & 0 \end{vmatrix} \\[1.5em] -\begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} & \begin{vmatrix} 0 & 0 \\ 4 & 0 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 4 & 0 \end{vmatrix} \\[1.5em] \begin{vmatrix} 1 & 0 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 0 & 0 \\ 2 & 2 \end{vmatrix} & \begin{vmatrix} 0 & 1 \\ 2 & 1 \end{vmatrix} \end{bmatrix} \qquad \text{(the cofactor matrix)}$$

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

Recall that the adjoint of A is the transpose of $C_\mathcal{A}$:

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

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

Finally, we can calculate the inverse of $\mathcal{A}$:

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

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

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

7.3 Appendix 3: Cramer's Rule

Example 7.11. 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):

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

Glossary