Lecture 6: Linear Algebra

Originally written by Mauricio Caceres Bravo
Revised by Cole Davis
Webpage: cj-davis99.github.io

August 20, 2024

Contents

Notation

- $\forall$ translates to "for all"
- $\exists$ translates to "there exists"
- $\mathbb{N} = \{1, 2, \ldots\}$ is the set of natural numbers
- $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ is the set of integers
- $\mathbb{Q} = \{p/q : p \in \mathbb{Z} \text{ and } q \in \mathbb{Z}\backslash\{0\}\}$ is the set of rational numbers
- $\mathbb{R}$ is the set of real numbers
- $\mathbb{C}$ is the set of complex numbers
- If S is a set and $n \in \mathbb{N}$, then S^n is the n^{th} order Cartesian product of S. E.g., $S^2 = S \times S$
- For any $\varepsilon > 0$, $B_\varepsilon(x)$ is the Euclidean ball around x with radius ε
- Unless otherwise specified, $d(x, y)$ is a metric on the contextual set x, y belong to
- The origin is always denoted as 0 regardless of the dimension of the space considered
- If A is a $N \times M$ matrix, $i \in \{1, \ldots, N\}$, and $j \in \{1, \ldots, M\}$, then $A_{-i,-j}$ is the submatrix of A with the i^{th} row and j^{th} column removed.

1 Motivation

In economics, it is very common to have ***systems of equations***,

$$\begin{aligned} Q_d &= a_0 + a_1 P \\ Q_s &= b_0 + b_1 P \\ Q_d - Q_s &= 0 \end{aligned} \tag{1}$$

We have supply, demand, and equilibrium. This is oozing economics, right? This problem can be expressed in general as

$$Ax = d$$

with A some $N \times M$ matrix times a column vector $x \in \mathbb{R}^M$ equal to some other vector $d \in \mathbb{R}^N$. The above is matrix notation for the system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1M}x_M &= d_1 \\ &\vdots \\ a_{N1}x_1 + a_{N2}x_2 + \ldots + a_{NM}x_M &= d_N \end{aligned}$$

With some wrangling, we can see that (1) can be expressed in this way:

$$\underbrace{\begin{bmatrix} a_0 \\ b_0 \\ 0 \end{bmatrix}}_{d} = \underbrace{\begin{bmatrix} -a_1 & 1 & 0 \\ -b_1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix}}_{x}$$

In this case, A is 3×3 and "invertible" (assuming a_1 and b_1 are not zero) so we can explicitly solve for equilibrium price and quantity.

$$\begin{aligned} \begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix} &= \begin{bmatrix} -a_1 & 1 & 0 \\ -b_1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} a_0 \\ b_0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix} &= -\frac{1}{\det(A)} \begin{bmatrix} 1 & -1 & -1 \\ b_1 & -a_1 & -a_1 \\ b_1 & -a_1 & -b_1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} P \\ Q_d \\ Q_s \end{bmatrix} &= \frac{1}{b_1 - a_1} \begin{bmatrix} a_0 - b_0 \\ a_0 b_1 - a_1 b_0 \\ a_0 b_1 - a_1 b_0 \end{bmatrix} \end{aligned}$$

(the second equality follows from a known matrix inverse formula which we state in Section 4). We can check this is the same answer we'd get if we'd solve this by leveraging the equilibrium condition $Q^s = Q^d$:

$$a_0 + a_1 P = b_0 + b_1 P \implies P = \frac{a_0 - b_0}{b_1 - a_1}$$

Solving for quantity yields,

$$Q^s = Q^d = a_0 + a_1 \frac{a_0 - b_0}{b_1 - a_1} = \frac{a_0 b_1 - a_0 a_1 + a_1 a_0 - a_1 b_0}{b_1 - a_1} = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}$$

In general, we find that

- If A is invertible[1] (more on invertible matrices below) then there exists a ***unique*** solution given by $x^* = A^{-1}d$.
- If $M < N$ then in general there are no solutions (you have more equations than parameters; in general it is not possible to satisfy all the equalities).
- If $N < M$ then in general there are infinitely many solutions (you have more parameters than equations; the additional parameters give you additional degrees of freedom).

Hence it will be useful to study the properties of matrices and vectors, which are the building blocks of these types of systems.

2 Vector Spaces

Definition 1. A ***field***, $\mathbb{F}$, is a set for which addition and multiplication are defined. Explicitly, $\mathbb{F}$ has the following properties:

1. $\forall x, y, z \in \mathbb{F}$
   (a) $x + y \in \mathbb{F}$
   (b) $x + y = y + x$
   (c) $(x + y) + z = x + (y + z)$
2. $\forall x, y, z \in \mathbb{F}$
   (a) $x \cdot y \in \mathbb{F}$
   (b) $x \cdot y = y \cdot x$
   (c) $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
3. $\forall x, y, z \in \mathbb{F}$, $x \cdot (y + z) = x \cdot y + x \cdot z$
4. $\exists 0 \in \mathbb{F}$ such that $\forall x \in \mathbb{F}$, $x + 0 = x$ and $\exists (-x) \in \mathbb{F}$ such that $x + (-x) = 0$
5. $\exists 1 \in \mathbb{F}$ such that $\forall x \in \mathbb{F}$, $x \cdot 1 = x$ and, if $x \neq 0$, $\exists x^{-1} \in \mathbb{F}$ such that $x \cdot x^{-1} = 1$

Most linear algebra texts state all results assuming that the scalar field of interest is either $\mathbb{R}$ or $\mathbb{C}$. Most economics results take this a step further and assume that the scalar field of interest is $\mathbb{R}$. We will operate as the typical economist does and assume that the field of consideration is $\mathbb{R}$, but we state the defintion of the next mathematical object in a general fashion:

[1] Note in this case $N = M$ is necessary, but not sufficient.

Definition 2. A ***vector space*** V over a field $\mathbb{F}$ is a collection (or a set) of objects called vectors endowed with addition and scalar multiplication. Explicitly, V has the following properties:

1. ***Additive closure***: $v + w \in V$ for any $v, w \in V$
2. ***Commutativity***: $v + w = w + v$ for any $v, w \in V$
3. ***Associativity***: $(v + w) + u = v + (w + u)$ for any $v, w \in V$.
4. ***Additive identity***: $\exists 0 \in V$ s.t. $v + 0 = v$ for any $v \in V$ (the 0 vector or the origin).
5. ***Additive inverse***: $\forall v \in V \;\; \exists -v \in V$ s.t. $v + (-v) = 0$.
6. ***Multiplicative closure***: $\forall \alpha \in \mathbb{F}$ and $\forall v \in V$, $\alpha v \in V$
7. ***Unit rule*** or ***multiplicative identity***: $1v = v$ for any $v \in V$.
8. ***Multiplicative associativity***: $(\alpha\beta)v = \alpha(\beta v)$ for any $v \in V$ and $\alpha, \beta \in \mathbb{F}$.
9. ***Distributivity***: For any $\alpha, \beta \in \mathbb{F}$, $v, w \in V$, we have

$$\alpha(u + v) = \alpha u + \alpha v \quad \text{and} \quad (\alpha + \beta)v = \alpha v + \beta v$$

Remark 1. If you are like me, you might think it's a bit odd that we are making a big deal of associativity, commutativity, and distributivity (like we did back in primary school). The reason is that a more formal treatment of linear spaces would not take any of these properties for granted, and would be very careful in discussing everything in this section (and henceforth) using just the definitions.

It can be shown that all finite dimensional vector spaces vector spaces over $\mathbb{R}$ are equivalent to the vector space $\mathbb{R}^N$. This can be of course be proven formally, but for this class I will simply discuss linear algebra results in $\mathbb{R}^N$.

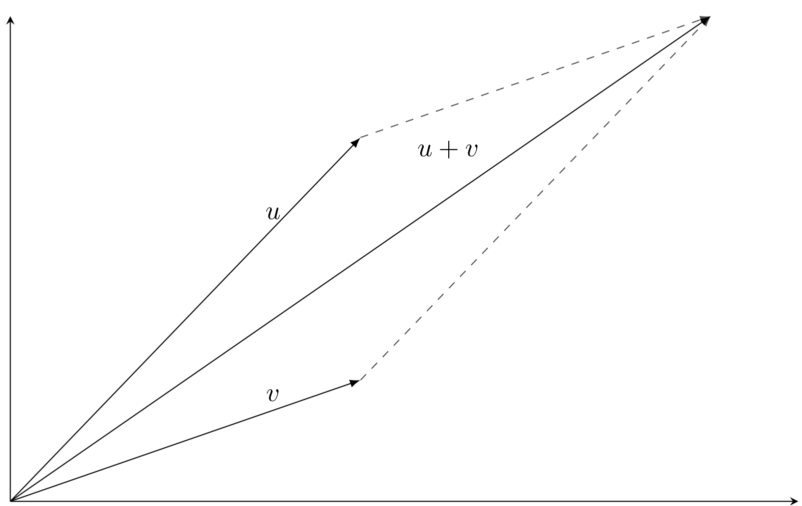


Figure 1: Example of Vector Addition

Definition 3. Let V be a vector space and let $\{v_1, \dots v_N\} \subseteq V$. A ***linear combination*** of $\{v_1, \dots v_N\}$ is given by

$$\sum_{i=1}^{N} \alpha_i v_i = \alpha_1 v_1 + \dots + \alpha_N v_N$$

for some arbitrary set of coefficients $\{\alpha_1, \dots, \alpha_N\} \subseteq \mathbb{R}$.

Remark 2. A vector space is a space that contains all its linear combinations. That is, for any $\{v_i\}_{i=1}^N \subseteq V$ and $\{\alpha_i\}_{i=1}^N \subseteq \mathbb{R}$ we have

$$\sum_{i=1}^{N} \alpha_i v_i \in V$$

if and only if V is a vector space. This is not obvious (and in a more formal treatment we would prove it from the definitions), but it turns out one characterization of a vector space is a space that has all its linear combinations. Hence the reason why vector spaces are also known as ***linear spaces***.

Definition 4. Let V be a vector space. We have $\{v_i\}_{i=1}^N \subseteq V$ is a ***basis*** of V if for every $u \in V$ there exists a *unique* linear combination of $\{v_i\}_{i=1}^N$ s.t.

$$u = \sum_{i=1}^{N} \alpha_i v_i$$

The coefficients $\{\alpha_i\}_{i=1}^N$ are called the ***coordinates*** of u in V with respect to the basis $\{v_i\}_{i=1}^N$.

The easiest example of a basis is the standard basis in $\mathbb{R}^N$:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \dots \quad e_N = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Naturally any $u \in \mathbb{R}^N$ can be uniquely represented as a linear combination of the standard basis if we take $\alpha_i = u_i \ \forall i$; that is, if we set the coefficients equal to each of the entries in u.

Definition 5. Let V be a vector space; we say that $\{v_i\}_{i=1}^N \subseteq V$ is a ***spanning set*** of V if for every $u \in V$ there exists a linear combination of $\{v_i\}_{i=1}^N$ s.t.

$$u = \sum_{i=1}^{N} \alpha_i v_i$$

Note that in the definition above, the linear combination does not have to be unique. Hence every basis of a vector space V is a spanning set of V, but not every spanning set of V can be a basis. Hence the question: When is a spanning set a basis?

Definition 6. A linear combination $\{\alpha_i\}_{i=1}^N$ is called ***trivial*** if $\alpha_i = 0$ for every i.

Definition 7. Let V be a vector space; we say that $\{v_i\}_{i=1}^N \subseteq V$ are ***linearly independent*** if the only linear combination of v_i that is equal to 0 is trivial. That is

$$\sum_{i} \alpha_i v_i = 0 \implies \alpha_i = 0$$

Definition 8. Let V be a vector space; if $\{v_i\}_{i=1}^N \subseteq V$ are not linearly independent, we say they are ***linearly***

dependent.

Theorem 1. *Let V be a vector space; $\{v_i\}_{i=1}^N \subseteq V$ is a basis for $V \iff \{v_i\}_{i=1}^N$ are a linearly independent spanning set of V.*

Claim 1. *Let $\{v_i\}_{i=1}^N, \{u_i\}_{i=1}^M$ be any basis for a vector space V. Then $N = M$.*

That is, bases of a vector space have the same number of elements. This leads to the following:

Definition 9. A vector space V has ***dimension***, $\dim(V)$, equal to the number of elements in any of its bases.

3 Linear Transformations

Definition 10. Let V, W be vector spaces. $T : V \to W$ is a ***linear transformation*** if

$$T(\alpha v + w) = \alpha T(v) + T(w)$$

for any $v \in V, w \in W, \alpha \in \mathbb{R}$.

Theorem 2. *Let V, W be vector spaces, $T : V \to W$ a linear transformation, and $\{v_i\}_{i=1}^N$ a basis for V. Then*

$$T(v) = \sum_i \alpha_i T(v_i)$$

for any $v \in V$.

Proof. Since $\{v_i\}_{i=1}^N$ is a basis for V, we know that $\{v_i\}_{i=1}^N$ spans V. That is, $\forall v \in V$, there exists an $\alpha_1, \ldots, \alpha_N$ such that

$$v = \sum_{i=1}^N \alpha_i v_i$$

Fix an arbitrary $v \in V$. Then

$$T(v) = T\left(\sum_{i=1}^N \alpha_i v_i\right)$$

We can then proceed by induction, using the properties of a linear transformation we saw above. Since V is a linear space, it has all its linear combinations. Thus $\alpha_1 v_1 \in V$ and $\sum_{i>1} \alpha_i v_i \in V$. Applying the definition,

$$T(v) = T\left(\alpha_1 v_1 + \sum_{i>1} \alpha_i v_i\right) = \alpha_1 T(v_1) + T\left(\sum_{i>1} \alpha_i v_i\right)$$

Iterating the above:

$$T(v) = \sum_{i<k} \alpha_i T(v_i) + T\left(\alpha_k v_k + \sum_{i>k} \alpha_k v_k\right) = \sum_{i<k} \alpha_i T(v_i) + \alpha_k T(v_k) + T\left(\sum_{i>k} \alpha_k v_k\right)$$

Therefore we can simply write

$$T(v) = \sum_i \alpha_i T(v_i)$$

□

Theorem 3. *Any linear transformation $T : V \to W$ can be represented by a matrix.*

Proof. Suppose $\dim(V) = N$ and $\dim(W) = M$. Further, let $\{v_i\}_{i=1}^N$ be a basis for V. An $M \times N$ matrix A is a collection of N M-dimensional vectors $\{a_1, \ldots, a_N\} \subseteq W$. We know that for any $x \in V$, there exists $x_1, \ldots, x_N \in \mathbb{R}$ such that $x = \sum_{i=1}^N x_i v_i \implies T(x) = \sum_{i=1}^N x_i T(v_i)$. Define $a_i = T(v_i) \in W$ for all i. Then

$$T(x) = \sum_{i=1}^{N} a_i x_i = Ax$$

where $A = (a_1, a_2, \ldots a_N)$ and $x = (x_1, x_2, \ldots, x_N)^\top$. This completes the proof. □

Example 1. Consider the following operation

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 6 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 8 \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Which we express as a linear transform

$$T(e_1) = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \qquad T(e_2) = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \qquad T(e_3) = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

That is,

$$T \begin{pmatrix} 4 \\ 6 \\ 8 \end{pmatrix} = 4T(e_1) + 6T(e_2) + 8T(e_3)$$

In general we can define matrix multiplication as linear operations on vector spaces:

$$AB = A[b_1 \ldots b_N] = [Ab_1 \ldots Ab_N]$$

Some properties of ***matrix multiplication***:

1. ***Associativity***: $(AB)C = A(BC)$
2. ***Distributivity***: $A(B + C) = AB + AC$

However, in general ***matrix multiplication is not commutative***. That is, $AB \neq BA$ in general (in fact, BA may not even be well-defined). Although matrix addition *is* commutative: $A + B = B + A$.

4 Matrix Inverse, Rank, and Determinant

Definition 11. The vector space spanned by the columns of a matrix A is the ***column space*** of A. The ***rank*** of a matrix, $\text{rank}(A)$, is the dimension of the column space (the maximum number of linearly independent columns).[2] A square matrix $N \times N$ is called ***full-rank*** if $\text{rank}(A) = N$ and ***rank-deficient*** if $\text{rank}(A) < N$.

Definition 12. A $N \times N$ matrix A is called ***diagonal*** if all its non-diagonal entries are 0.

Definition 13. The ***identity matrix*** is an $N \times N$ diagonal matrix with all diagonal entries equal to 1 (and non-diagonal entries equal to 0). For any $N \times N$ matrix A, $AI = IA = A$.

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

Definition 14. The $N \times N$ matrix A is said to be ***left invertible*** if $\exists C_L$ s.t. $C_L A = I$ the identity. We say it is ***right invertible*** if $\exists C_R$ s.t. $AC_R = I$. If A is left and right invertible with $C_L = C_R$, we say that A is ***invertible*** (or ***non-singular***), and we denote $C_R = C_L = A^{-1}$ the ***inverse*** of A.[3]

Definition 15. If A is not invertible, we say that A is ***singular***.

Theorem 4. *Let A be a $N \times N$ matrix. A is invertible* $\iff \text{rank}(A) = N$.

Some other properties of the rank:

- $\text{rank}(A) = \text{rank}(A^T) = \text{rank}(AA^T) = \text{rank}(A^T A)$.
- $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$.
- $\text{rank}(CAB) = \text{rank}(A)$ if C, B are non-singular.

Example 2. Can we find the inverse of the matrix

$$B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

[2] It turns out that the row-rank of a matrix, the maximum number of linearly independent rows, is the same as the column-rank of a matrix. Hence we can just talk about the rank without clarifying row or column.

[3] The left and right inverses are the same for a square matrix if either exist, but this is a result, not an assumption. Further, note you can define the left and right inverses for non-square matrices, but such matrices needn't be invertible as only one of C_L or C_R might exist.

using elementary row operations?

$$\begin{bmatrix} 2 & 1 & \to & 1 & 0 \\ 1 & 1 & & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1/2 & \to & 1/2 & 0 \\ 1 & 1 & & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1/2 & \to & 1/2 & 0 \\ 0 & -1/2 & & 1/2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \to & 1 & -1 \\ 0 & -1/2 & & 1/2 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & \to & 1 & -1 \\ 0 & 1 & & -1 & 2 \end{bmatrix}$$

The steps here were:

1. Multiply the first row by $1/2$.
2. Multiply the second row by -1 and add the first row to the second row.
3. Add the second row to the first row.
4. Multiply the second row by -2.

In general, however, we have a shortcut for 2×2 matrices:

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

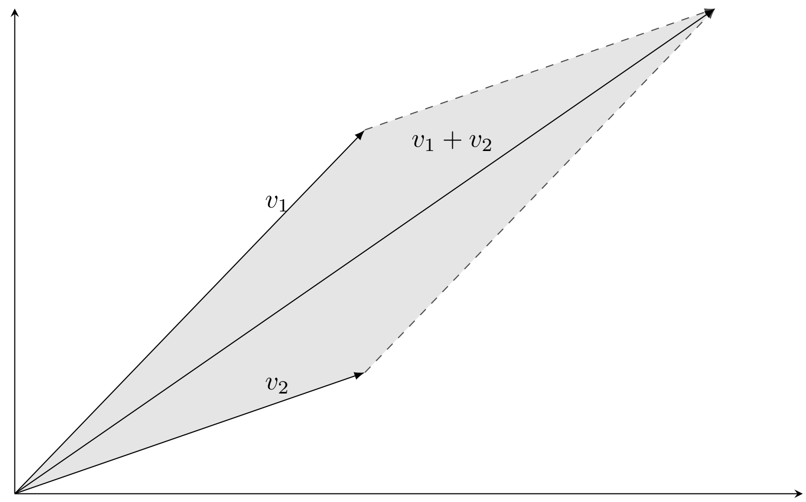
You can check this formula gives exactly what we found using row operations.

Definition 16. The ***determinant*** of a $N \times N$ square matrix A is

$$\det(A) = \sum_{j=1}^{N} (-1)^{i+j} a_{ij} \det(A_{-i,-j})$$

with i any fixed row of A and $A_{-i,-j}$ the sub-matrix of A that results from moving row i and column j.

Consider this picture:



The set

$$\{v : v = t_1 v_1 + t_2 v_2, \quad t_1, t_2 \in [0, 1]\}$$

will give the parallelogram, and the determinant of the 2×2 matrix $V = [v_1 \; v_2]$ will be be the area of the parallelogram. (Rather, the absolute value of the determinant; if the determinant is negative that just says something about the direction of the vectors, but the intuition remains). Some properties of the determinant:

- If A and B are both square matrices, then $\det(AB) = \det(A)\det(B)$.
- $\det(A) = \prod_i a_{ii}$ if A is diagonal, upper triangular (i.e., only zeros entries below diagonal), or lower triangular (i.e., only zero entries above diagonal).
- $\det(I) = 1$ (immediate consequence of the point above).
- $\det(\alpha A) = \alpha^N \det(A)$ for any $\alpha \in \mathbb{R}$.
- If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.

Another point to make about the determinant is that it is related to the rank, that is, whether the columns (or rows) of a square matrix A are linearly independent.

1. If A has a zero column or zero row, then $\det(A) = 0$.

   To see this, note that from the formula of the determinant, if we have a row or column of all 0s, then each element of the sum will eventually be multiplied by 0, and the entire term will be 0. In the 2×2 case, if one of the vectors comprising a "side" of the parallelogram is the origin, the parallelogram will just be line which we know has no area.

2. If A has two columns or rows that are equal, then $\det(A) = 0$.

   If two columns of A are equal, WLOG let them be the first two columns, consider the matrix P with all diagonal entries equal to 1 and all but one non-diagonal entries equal to 0. Let $P_{21} = -1$. That is,

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

What is AP? In this case, with $A = [a_1 \quad a_2 \quad \cdots \quad a_N]$, we have

$$AP = \begin{bmatrix} a_1 - a_2 & a_2 & \cdots & a_N \end{bmatrix} = \begin{bmatrix} 0 & a_2 & \cdots & a_N \end{bmatrix}$$

with the first column equal to 0 since we assumed the first two columns were equal. Finally, we found in the previous bullet that if a column was all 0s then the determinant was 0. Hence

$$0 = \det(AP) = \det(A)\det(P) \implies \det(A) = 0 \quad \text{or} \quad \det(P) = 0$$

Note $\det(P) = 1$, so it must be that $\det(A) = 0$.

3. If A has a columns that is a multiple of another column, then $\det(A) = 0$.

   Again WLOG suppose these are the first two columns and let A be as above. If $a_1 = ca_2$ for some $c \in \mathbb{R}$, then we can define P almost identically, except $P_{21} = -\alpha$. That is,

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -c & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}$$

   And again we have $AP = [a_1 - ca_2 \;\; a_2 \;\; \cdots \;\; a_N] = [0 \;\; a_2 \;\; \cdots \;\; a_N]$. Hence $\det(AP) = \det(A)\det(P) = 0$, but again we have $\det(P) = 1$, so $\det(A) = 0$.

4. If A has a columns or rows that are linearly dependent, then $\det(A) = 0$.

   This is really the point we wanted to make: A determinant of 0 is tied to the linear dependence of the columns of A. This follows from the work we did above: In this case, $\exists\{c_i\}$ s.t. $c_i \neq 0$ for some i s.t.

$$a_j = \sum_{i \neq j} c_i a_i$$

for some j. WLOG let $j = 1$, and the matrix P becomes

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -c_2 & 1 & 0 & \cdots & 0 \\ -c_3 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \\ -c_N & & & & 1 \end{bmatrix}$$

So that

$$AP = \begin{bmatrix} a_1 - c_2 a_2 - c_3 a_3 \ldots & a_2 & \cdots & a_N \end{bmatrix} = \begin{bmatrix} a_1 - \sum_{i \neq 1} c_i a_i & a_2 & \cdots & a_N \end{bmatrix} = \begin{bmatrix} 0 & a_2 & \cdots & a_N \end{bmatrix}$$

So, one last time, $\det(AP) = \det(A)\det(P) = 0$ and $\det(P) = 1$ implies $\det(A) = 0$.

Theorem 5. *A is rank-deficient* $\iff \det(A) = 0$.. *Equivalently, A is non-invertible* $\iff \det(A) = 0$.

Definition 17. For a $N \times N$ matrix A, the i, j ***minor***, denoted as M_{ij}, is the *determinant* of the $(N-1) \times (N-1)$ sub-matrix obtained by deleting the ith row and jth column of A.

Definition 18. The i, j ***cofactor*** of the square $N \times N$ matrix A is given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

Thus we can rewrite Definition 16 as $\det(A) = \sum_{j=1}^{N} a_{ij} C_{ij}$ for any given i.

Definition 19. The ***adjoint*** of an $N \times N$ matrix A is the transpose of the cofactor matrix given by

$$\text{adj}(A) = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1N} \\ C_{21} & C_{22} & & \vdots \\ \vdots & & \ddots & \\ C_{N1} & C_{N2} & \cdots & C_{NN} \end{bmatrix}^\top = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{N1} \\ C_{12} & C_{22} & & \vdots \\ \vdots & & \ddots & \\ C_{1N} & C_{2N} & \cdots & C_{NN} \end{bmatrix}$$

Note the above is a matrix of scalars, since C_{ij} is the product of a determinant, which is a scalar, and $(-1)^{i+j}$, which is a scalar as well.

Theorem 6. *If A is a non-singular $N \times N$ matrix then*

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$$

For instance, for a 2×2 matrix we have

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

Applying the theorem,

$$B^{-1} = \frac{1}{\det(B)} \begin{bmatrix} C_{11} & C_{21} \\ C_{12} & C_{22} \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Theorem 7 (Cramer's Rule). *Let A be a $N \times N$ non-singular matrix s.t. $Ax = b$. Then*

$$x_i = \frac{\det(\widetilde{A}_i)}{\det(A)}$$

for

$$\widetilde{A}_i = \begin{bmatrix} a_1 & \cdots & a_{i-1} & b & a_{i+1} & \cdots & a_N \end{bmatrix}$$

That is, $\widetilde{A}$ is the matrix obtained by replacing the ith column of A with b.

Definition 20. The ***trace*** of an $N \times N$ matrix A is the sum of its diagonal elements,

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

Some properties of the trace:

- $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$
- $\text{trace}(AB) = \text{trace}(BA)$ if both products exist.
- $\text{trace}(\alpha A) = \alpha \, \text{trace}(A)$ for any $\alpha \in \mathbb{R}$.

An example which comes up in econometrics is that for any $N \times K$ matrix X s.t. $\left(X^T X\right)^{-1}$ exists, we have

$$\text{trace}\left(\underbrace{X}_{A}\overbrace{(X^T X)^{-1} X^T}^{B}\right) = \text{trace}((X^T X)^{-1} X^T X) = \text{trace}(I_{K \times K}) = K$$

Further,

$$\text{trace}\left(I_{N \times N} - X(X^T X)^{-1} X^T\right) = \text{trace}\left(I_{N \times N}\right) - \text{trace}(X(X^T X)^{-1} X^T\big) = N - K$$

5 Eigenvalues and Eigenvectors

Take

$$T(v) = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

Note that for the standard basis,

$$\begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} e_1 = 3e_1 \qquad \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} e_2 = -e_2$$

For this particular example, we can see the standard basis is exactly scaled by 3 and -1. Hence, T acting on v scales v_1 by 3 and scales v_2 by -1. Unfortunately, not all linear transformations have diagonal matrix representations that allow us to easily determine how they geometrically transform their arguments. This leads us to ask the following question: can we build a theory that helps us better understand the geometry of any given linear transformation?

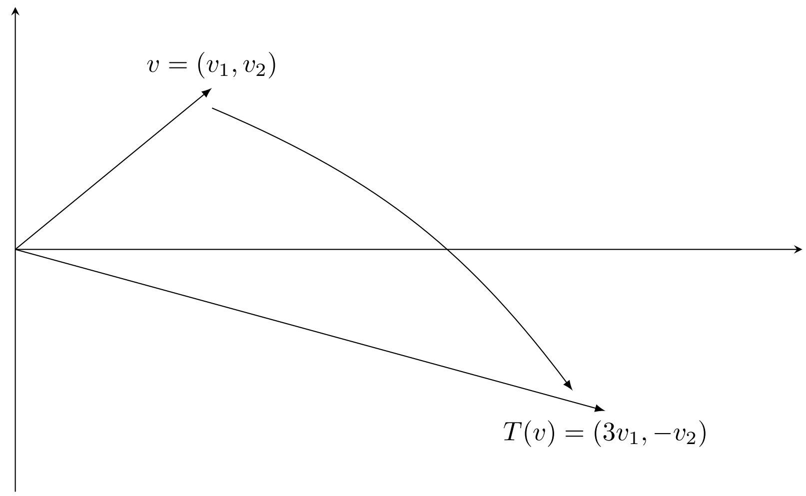


Figure 2: Visualization of a Transformation

Definition 21. Let A be a $N \times N$ square matrix. The $N \times 1$ vector $v \neq 0$ is an ***eigenvector*** (or characteristic vector) of A with corresponding ***eigenvalue*** or (characteristic root) λ if

$$Av = \lambda v$$

To understand the geometry of a matrix A, we want to find a basis $\{v_i\}$ of the column space of A s.t.

$$Av_i = \lambda_i v_i$$

that is, a basis of eigenvectors. The intuition here is that if v_i lies on an axis in the column space of A, then applying A to v_i is only going to, at most, scale v_i. Identifying these axes and their associated scaling factor fully describes the geometry of the transformation that results from applying A. Note that

$$Av = \lambda v \iff (A - \lambda I)v = 0$$

This means that for any $v \neq 0$, we have that

$$p(\lambda) \equiv \det(A - \lambda I) = 0$$

$p(\lambda)$ is called the ***characteristic polynomial*** of A. (Recall that a matrix has a zero determinant if it is singular; in this case if $(A - \lambda I)v = 0$ for any non-zero vector then the columns of $A - \lambda I$ are not linearly independent, and thus the matrix is singular and has a zero determinant.)

Example 3. Consider

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix} = 0 \iff (1-\lambda)^2 - 4 = 0 \iff (\lambda+1)(\lambda-3) = 0$$

Hence $\lambda = \{-1, 3\}$. These are the ***eigenvalues***. Now we find the ***eigenvectors***. For $\lambda = -1$

$$(A - \lambda I)v = 0 \iff \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} v = 0$$

Which gives that v must be s.t. $v_1 = -v_2$. For $\lambda = 3$

$$(A - \lambda I)v = 0 \iff \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} v = 0$$

Which gives that v must be s.t. $v_1 = v_2$. It is often useful for the norm of the basis to be 1 where the norm $\|v\|$ for any $v \in \mathbb{R}^N$ is

$$\|v\| = \left(\sum_{i=1}^{N} v_i^2 \right)^{1/2}$$

Hence the normal eigenvectors are given by

$$v = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad u = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Note that we can express the above as a system:

$$AP = P\Lambda$$

where

$$P = \begin{bmatrix} v & u \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \Lambda = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}$$

In this case, note $v^T u = 0$ and $v^T v = u^T u = 1$. Hence

$$P^T P = \begin{bmatrix} v^T \\ u^T \end{bmatrix} \begin{bmatrix} v & u \end{bmatrix} = \begin{bmatrix} v^T v & v^T u \\ u^T v & u^T u \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and if we pre-multiply the eigen-system by P^T we get

$$P^T AP = \Lambda$$

Definition 22. A an $N \times N$ matrix is ***diagonalizable*** if $\exists P$ and diagonal matrix Λ s.t.

$$P^{-1}AP = \Lambda$$

Note we can equivalently write $A = P\Lambda P^{-1}$.

It will be useful to note if A is diagonalizable

- A^m for any $m \in \mathbb{N}$ can computed as $P\Lambda^m P^{-1}$.

$$A^m = A \times \ldots \times A = P\Lambda P^{-1} \times \ldots \times P\Lambda P^{-1} \times \ldots \times P\Lambda P^{-1} = P\Lambda^m P^{-1}$$

  since each of the $P^{-1}P$ cancel and $\Lambda \times \Lambda = \Lambda^2$ and so on.

- $A^{1/m}$ for any $m \in \mathbb{N}$ can be computed as $P\Lambda^{1/m}P^{-1}$.

$$\left(P\Lambda^{1/m}P^{-1}\right)^m = P\Lambda^{1/m}P^{-1} \times \ldots \times P\Lambda^{1/m}P^{-1} = P\Lambda^{1/m} \times \ldots \times \Lambda^{1/m}P^{-1} = P\Lambda P^{-1} = A$$

- If Λ is invertible, then A^{-m} for any $m \in \mathbb{N}$ can computed as $P\Lambda^{-m}P^{-1}$.

$$A^{-m} = (A^m)^{-1} = \left(P\Lambda^m P^{-1}\right)^{-1} = P\Lambda^{-m}P^{-1}$$

  and we can check

$$A^m A^{-m} = P\Lambda^m P^{-1} P\Lambda^{-m} P^{-1} = P\Lambda^m \Lambda^{-m} P^{-1} = PP^{-1} = I$$

- Similarly, if Λ is invertible $A^{-1/m}$ for any $m \in \mathbb{N}$ can computed as $P\Lambda^{-1/m}P^{-1}$.
- Let $q \in \mathbb{Q}$ (so $q = m/n$ for $m, n \in \mathbb{Z}$). Then $A^q = (A^m)^{1/n}$ is $P(\Lambda^m)^{1/n}P^{-1} = P\Lambda^q P^{-1}$.
- Finally, $A^r = P\Lambda^r P^{-1}$ for any $r \geq 0$ (if in addition Λ is invertible, this holds for any $r \in \mathbb{R}$).

Claim 2. *If A is diagonalizable then P is a matrix of eigenvectors and Λ is a matrix of eigenvalues.*

Proof. If A is diagonalizable then

$$AP = P\Lambda$$

Let v_k be the kth column of P. Then $Av_k = \lambda_k v_k$ we can see v_k is an eigenvector and λ_k an eigenvalue. □

Theorem 8. *Let A be a $N \times N$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_N$.*

- $\text{trace}(A) = \sum_{i=1}^N \lambda_i$
- $\det(A) = \prod_{i=1}^N \lambda_i$.

Proof. Let us show this assuming A is diagonalizable.[4]

$$\text{trace}(A) = \text{trace}(P^{-1}\Lambda P) = \text{trace}(\Lambda P P^{-1}) = \text{trace}(\Lambda) = \sum_i \lambda_i$$

$$\det(A) = \det(P^{-1}\Lambda P) = \det(P^{-1})\det(\Lambda)\det(P) = \det(\Lambda) = \prod_i \lambda_i$$

where we utilized $\text{trace}(AB) = \text{trace}(BA)$ and $\det(A^{-1}) = \det(A)^{-1}$ □

Theorem 9. *If A is symmetric then A is diagonalizable and P can be chosen to be orthonormal ($P^{-1} = P^T$ with each column of P equal to a unit vector). This is known as the* ***spectral decomposition****.*

Recall in Example 3, the matrix A we considered was symmetric, and we found a diagonalization where $P^{-1} = P^T$. This theorem states that we can do this for any symmetric matrix. Below I present an example for a matrix that is not symmetric.

Theorem 10. *Let A be a symmetric $N \times N$ matrix. Then*

1. *A is PD (ND) $\iff$ $\lambda_i > 0$ (< 0) for all eigenvalues λ_i of A.*
2. *A is PSD (NSD) $\iff$ $\lambda_i \geq 0$ (≤ 0) for all eigenvalues λ_i of A.*

Proof. If A is symmetric, then

$$v'Av = v'P\Lambda P^T v = u^T\Lambda u \tag{2}$$

for $u = P^T v$ (since P is non-singular, $u \neq 0$ if $v \neq 0$). Hence the definiteness of A is the same as the definiteness of Λ. From here we can see that

$$u^T\Lambda u = \sum_{i=1}^N \lambda_i u_i^2 \tag{3}$$

where $u_i^2 \geq 0$ (strict for at least one i). If $\lambda_i > 0$ for all i, $v'Av = \sum_{i=1}^N \lambda_i u_i^2 > 0$, so it is PD. Conversely, if A is PD, we know $v'Av > 0$ for any $v \neq 0$. In particular, it must be true for v equal to each of the columns of P. Since the columns of P are orthonormal, for $v = p_i$ the ith column of P, $P^T p_i = e_i$ the ith standard vector (i.e. 1 in the ith entry and 0 elsewhere). Hence $v'Av = \lambda_i$ when $v = p_i$; since all such quadratic forms are positive, $\lambda_i > 0$. The logic for < 0 and ND, ≥ 0 and PSD, and ≤ 0 and NSD is completely analogous. □

[4] For the general proof, recall $\lambda_1, \ldots, \lambda_N$ are the roots of the polynomial given by

$$\det(A - \lambda I) = (\lambda_1 - \lambda) \times (\lambda_2 - \lambda) \times \ldots \times (\lambda_N - \lambda)$$

For the determinant, set $\lambda = 0$ and find $\det(A) = \prod_{i=1}^N \lambda_i$. For the trace, we need to leverage something known as "Vieta's formulas." The relevant result is that for a polynomial of order N, the $N-1$ coefficient is the sum of the roots of the polynomial (this sounds esoteric, but think about how you learned to expand formulas like $(\lambda_1 - \lambda)(\lambda_2 - \lambda)$; the coefficient of the "middle" term is $-(\lambda_1 + \lambda_2)$, and this is just the generalization). If we can show trace(A) is the $N-1$ coefficient of the characteristic polynomial we'd be done. To see it, note that λ^{N-1} will only appear if all the diagonal elements are multiplied (any other permutation will give at most a polynomial of order $N-2$). Hence λ^{N-1} only appears as part of the term

$$\prod_i (a_{ii} - \lambda)$$

Here the roots are a_{ii}, so the coefficient on λ^{N-1} is $\sum_i a_{ii} = \text{trace}(A)$.

5.1 Eigen Example: Eigen Decomposition of an Asymmetric Matrix

Let A be a 2×2 matrix defined as follows

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

We can find that $\det(A - \lambda I) = 0$ gives

$$0 = (1-\lambda)^2 + 4 = \lambda^2 - 2\lambda + 5 \qquad \lambda = \frac{2 \pm \sqrt{4 - 4 \cdot 5}}{2}$$

Hence $\lambda = 1 \pm 2i$. We first consider the eigenvectors corresponding to the eigenvalue $1 + 2i$:

$$\begin{aligned} Av &= (1+2i)v \\ \iff \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} &= \begin{bmatrix} (1+2i)v_1 \\ (1+2i)v_2 \end{bmatrix} \\ \implies v_1 + 2v_2 &= (1+2i)v_1 \\ \implies v_2 &= iv_1 \end{aligned}$$

We can follow the same logic for $\lambda = 1 - 2i$ to arrive at the condition that defines the second set of eigenvectors:

$$\begin{aligned} Au &= (1-2i)u \\ \implies u_2 &= -iu_1 \end{aligned}$$

For simplicity, let $v_1 = u_1 = 1$. Then our eigenvectors are

$$v = \begin{bmatrix} 1 \\ i \end{bmatrix} \quad u = \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

Note that we could scale v and u to make them normal vectors; however, there is nothing we can do to make v and u orthogonal (i.e., $u'v = v'u = 0$). Thus, the matrix

$$P = \begin{bmatrix} v & u \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

is not an orthogonal matrix and instead of $P^{-1} = P^T$, we have that

$$P^{-1} = \begin{bmatrix} 1/2 & -1/2i \\ 1/2 & 1/2i \end{bmatrix}$$

Hence

$$A = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 1+2i & 0 \\ 0 & 1-2i \end{bmatrix} \begin{bmatrix} 1/2 & -1/2i \\ 1/2 & 1/2i \end{bmatrix}$$

Index