Natalia Lazzati

Mathematics for Economics

Note 1: Linear Algebra

Note 1 is based on Searle and Willett (2001) and Simon and Blume (1994, Ch. 6, 7, 8, 9, 11, 26 and 27).

Although most of the models economists are interested in are nonlinear in nature, linear algebra plays a fundamental role in economic theory. The reason is that to study nonlinear systems we often use the meta-principle of calculus: to study the behavior of the solutions to a nonlinear system we examine the behavior of a closely related system of linear equations!

Most of the chapters in Simon and Blume (1994), SB, were assigned to read during summer, so the main purpose of this note is to collect (and connect) the main ideas in a single file. (The notation we use is slightly different from the one in that book.)

Matrix Algebra

Definition of Matrix

A matrix is a rectangular array of elements arranged in rows and columns. All rows are of equal length, as well as all columns. It is a devise frequently used in organizing the presentation of numerical data so that they may be handled with ease mathematically. In this sense we can think of matrix algebra as a vehicle by which mathematical procedures for many problems, both large and small, can be described independently of the size of the problem. The use of matrices is fundamental in many areas of economics, for instance, in econometrics.

In terms of notation a_{ij} indicates the element in the ith row and jth column of a matrix. If a matrix A has m rows and n columns then it can be represented as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}.$$

An alternative representation is

$$A = \{a_{ij}\} \text{ for } i = 1, 2, ..., m \text{ and } j = 1, 2, ..., n.$$

Basic Operations

Pages 153-157 in SB cover five basic matrix operations: addition, substraction, scalar multiplication, matrix multiplication and transposition. These pages also describe the laws of matrix algebra: associative, distributive and commutative. All these operations appear quite often in both statistics and econometrics (so you should try to remember them).

Special Matrices

Several matrices with particular properties receive specific names. Historically, each of them probably originated in a mathematical problem or application.

Some simple special forms are: square, diagonal, identity, lower and upper triangular, symmetric and idempotent matrices (see SB, pp. 160-161). A matrix which is just a single column is called a column vector. Similarly, a row vector is a matrix which is just a single row. In terms of notation we will indicate a column vector by $\mathbf{x}$, and a row vector by $\mathbf{x}^T$. A scalar is just a matrix of dimension 1×1, that is, a single element.

A square matrix is nonsingular if its rank equals the number of its rows (or columns). (Recall that the rank of a matrix is the number of nonzero rows in its raw echelon form.) When such a matrix arises as a coefficient matrix in a system of linear equations, then the system has one and only one solution. So nonsingular matrices are of fundamental importance, and we will offer other characterizations of this property.

Determinants

Definition of Determinant

Determinants are important in the analysis of mathematical models. For example, they help to define if a system of linear equations has a solution, to compute the solution when it exists, and to determine whether a given nonlinear system can be well approximated by a linear one. They are also important to test the sufficient conditions of optimization problems.

A determinant is just a scalar which is the sum of selected products of the elements of the matrix from which it is derived. It is defined only for square matrices—the determinant of a non-square matrix is undefined and does not exist. The customary notation for the determinant of the matrix A is just $|A|$.

Let A be an $n \times n$ matrix, and let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column. The determinant of A_{ij}, $M_{ij} = |A_{ij}|$, receives the name of (i,j) minor of A. In addition $C_{ij} = (-1)^{i+j} M_{ij}$ is the (i,j) cofactor of A.

The determinant of a 1×1 matrix is the value of its sole elements. The determinant of an $n \times n$ matrix A can be defined, inductively, as follows

$$|A| = \sum_{j=1}^{n} a_{1j} (-1)^{1+j} M_{1j} = \sum_{j=1}^{n} a_{1j} C_{1j}. \tag{1}$$

This method of expanding a determinant is known as expansion by the elements of the first row. The selection of the first row is arbitrary, as we can in fact expand the determinant of A by using any row or any column. (Then you should choose the one with more zeros! **Why?**)

Theorem 1. If we define the determinant of A as in (1), then for any i or j,

$$\begin{aligned} |A| &= \sum_{j=1}^{n} a_{ij} (-1)^{i+j} M_{ij} && (2) \\ &= \sum_{i=1}^{n} a_{ij} (-1)^{i+j} M_{ij} && (3) \end{aligned}$$

where (2) expands the determinant through the ith row and (3) uses the jth column.

Proof. See SB p. 743. ■

Corollary 2. For any $n \times n$ matrix A, $|A| = \left|A^T\right|$.

Proof. This result follows as a simple corollary of Theorem 1. ■

Properties of the Determinant

Determinants have some useful properties. The next lines assume A is a square matrix.

(a) If we add to any row of A a multiple of another row, then the determinant does not change;

(b) if we form matrix B by multiplying any row of A by a scalar r, then $|B| = r\,|A|$;

(c) if we form matrix B by interchanging two rows of A then $|B| = -\,|A|$;

(d) if two rows of A are the same, then $|A| = 0$;

(e) if A is lower or upper triangular, then its determinant is just the product of the elements of its principal diagonal; and

(f) if R is the row echelon form of A, then $|R| = \pm\,|A|$.

(**Show each of the previous properties. These properties are in terms of rows, explain why the same results hold for the columns.**)

Properties (a)-(f) are quite useful to justify two fundamental theorems that we discuss next. The first one states that the determinant of a matrix defines whether or not the matrix is singular.

Theorem 3. A square matrix A is nonsingular if and only if $|A| \neq 0$.

Proof. By the last property of a determinant, (f), if R is the row echelon form of matrix A then $|R| = \pm\,|A|$. If A is nonsingular then R has all its diagonal elements different from zero. Since R is upper diagonal then the theorem follows by the fifth property of a determinant, (e). ■

The second theorem shows that for two arbitrary $n \times n$ matrices A and B, the determinant of the product is the same as the product of their determinants.

Theorem 4. For arbitrary $n \times n$ matrices A and B we get $|AB| = |A|\,|B|$.

Proof. See SB p. 733. ■

Inverse Matrix

In matrix algebra there is no division; we never divide a matrix. Instead, under certain circumstances we can multiply by what is known as an inverse matrix. This operation between matrices is the counterpart of division. We next define inverse matrices, the conditions for their existence and some important applications.

The inverse of a matrix A is a matrix B that satisfies

$$BA = I = AB. \tag{4}$$

We will indicate the inverse of A, when it exists, as A^{-1}. From the last definition, it follows that a necessary condition for A^{-1} to exist is that A be a square matrix. Moreover, A and A^{-1} will share the same dimension, i.e. if A is an $n \times n$ matrix, so is A^{-1}.

Given that A must be square for A^{-1} to possibly exist, we can use (4) again to derive the conditions under which A^{-1} actually exists. First take the determinant of both sides of (4)

$$\left|A^{-1}A\right| = |I|\,. \tag{5}$$

By the product rule

$$\left|A^{-1}\right||A| = 1. \tag{6}$$

The important consequence is that $|A| \neq 0$ for the inverse of A to exist. Then A^{-1} can exists if A is square, and it does exists if in addition $|A| \neq 0$.

Theorem 5. If A^{-1} exists, then it is unique.

Proof. (**Show this result.**) ■

Let us define adj$A = \{C_{ij}\}$, where C_{ij} is the (i, j) cofactor of A (as described before). The inverse of A, when it exists, can be obtained as follows

$$A^{-1} = \frac{1}{|A|}\left(\mathrm{adj}A\right)^T. \tag{7}$$

Example 1. Consider a matrix $A = \begin{pmatrix} a & x \\ y & b \end{pmatrix}$. Its determinant is $|A| = ab - xy$. Then the inverse exists if and only if $ab - xy \neq 0$. If the latter condition holds, then

$$A^{-1} = \frac{1}{(ab - xy)}\begin{bmatrix} b & -x \\ -y & a \end{bmatrix}. \ \blacktriangle$$

Linearly (In)dependent Vectors

This section and the following one relate to certain mathematical concepts and relationships concerning individual rows (and/or columns) of a matrix. They provide methods for ascertaining if a square matrix A has a zero determinant. These methods are easier than calculating the determinant, and will shed extra light later on the characterization of solutions of systems of equations.

Linear Combination of Vectors

If $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ denote n vectors of the same dimension, then for scalars $a_1, a_2, ..., a_n$,

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_n\mathbf{x}_n \tag{8}$$

is called linear combination of the set of n vectors.

Example 2. Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{x}_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

A linear combination of the two vectors is given by

$$a_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + a_2 \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} a_1 + 3a_2 \\ 2a_1 + 7a_2 \end{bmatrix}$$

for some scalars a_1 and a_2. ▲

Note that the final line in the last example can be written as

$$\begin{bmatrix} a_1 + 3a_2 \\ 2a_1 + 7a_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = X\mathbf{a}.$$

This is true in general. If

$$X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & ... & \mathbf{x}_n \end{bmatrix} \text{ and } \mathbf{a}^T = (a_1, a_2, ..., a_n) \tag{9}$$

then the linear combination in (8) is a vector that can be written as

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_n\mathbf{x}_n = \textstyle\sum_{i=1}^{n} a_i\mathbf{x}_i = X\mathbf{a}. \tag{10}$$

As a consequence, given a matrix X and a vector $\mathbf{a}$ for which $X\mathbf{a}$ exists, $X\mathbf{a}$ is a column vector, i.e. a linear combination of the columns of X. Similarly, if we let $\mathbf{b}^T = (b, ..., b_m)$ then $\mathbf{b}^T X$ is a row vector, i.e. a linear combination of the rows of X. Moreover, AB is a matrix with rows that are linear combinations of the rows of B, and its columns are linear combinations of the columns of A.

Definition of Linear (In)dependence

As we explained before, the product

$$X\mathbf{a} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_n\mathbf{x}_n \tag{11}$$

is a vector. If there exists a vector $\mathbf{a} \neq \mathbf{0}$, such that $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_n\mathbf{x}_n = \mathbf{0}$, then provided none of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ is null, those vectors are said to be linearly dependent vectors. An alternative statement of the definition is: if $X\mathbf{a} = \mathbf{0}$ for some non-null $\mathbf{a}$, then the columns of X are linearly dependent vectors, provided none is null.

Example 3. Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{x}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Then

$$2\begin{bmatrix} 1 \\ 2 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since neither $\mathbf{x}_1$ nor $\mathbf{x}_2$ are null vectors and $\mathbf{a}^T = (2, -1) \neq 0$, it follows that these two vectors are linearly dependent. ▲

If on the contrary $\mathbf{a} = \mathbf{0}$ is the only vector for which $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_n\mathbf{x}_n = \mathbf{0}$, then provided none of $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ is null, those vectors are said to be linearly independent (LIN) vectors. An alternative statement is: if $X\mathbf{a} = \mathbf{0}$ only for $\mathbf{a} = \mathbf{0}$, then the columns of X are LIN vectors. (**Show that the two vectors in Example 2 are LIN.**)

Assume that $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ are linearly dependent, and suppose that a_1 and a_2 are non-zero (this arbitrary selection is without loss of generality). Then

$$\mathbf{x}_1 + (a_2/a_1)\,\mathbf{x}_2 + ... + (a_n/a_1)\,\mathbf{x}_n = \mathbf{0} \tag{12}$$

so that

$$\mathbf{x}_1 = -\,(a_2/a_1)\,\mathbf{x}_2 - ... - (a_n/a_1)\,\mathbf{x}_n$$

which means that $\mathbf{x}_1$ can be expressed as a linear combination of the other $\mathbf{x}'s$. Therefore, the notion of dependent vectors relates to the possibility of expressing some of them as the linear combinations of some others. It is the existence of such combinations that is important rather than the specific scalars that multiply the vectors in those combinations!

The next theorem captures an important result related to LIN vectors.

Theorem 6. A set of LIN vectors of order n never contains more than n vectors.

Proof. Let $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ be n LIN vectors of order n. Let $\mathbf{u}^*$ be other non-null vector of order n. We show that $\mathbf{u}^*$ and $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$ are linearly dependent.

Since $U = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n]$ has LIN columns, $|U| \neq 0$ and U^{-1} exists. Let $\mathbf{q} = -U^{-1}\mathbf{u}^* \neq \mathbf{0}$, because $\mathbf{u}^* = -U\mathbf{q} \neq \mathbf{0}$. Then $U\mathbf{q} + \mathbf{u}^* = \mathbf{0}$, which can be written as

$$\begin{bmatrix} U & \mathbf{u}^* \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} = \mathbf{0} \text{ with } \begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} \neq \mathbf{0}.$$

Therefore $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n, \mathbf{u}^*$ are linearly dependent. Then there is no other vector $\mathbf{u}^*$ that can be added to the matrix U with the property that the $n+1$ vectors be LIN. ■

Zero Determinants and Inverse

Suppose that n linearly dependent vectors of order n are used as either columns or rows of a matrix. Then the determinant is zero. (**Show this statement using the properties of the determinants.**) Therefore if a square matrix of order n has linearly dependent columns or rows its inverse does not exist.

Example 4. Let

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 2 \\ -6 & 5 & 1 \\ 9 & -5 & 1 \end{bmatrix}.$$

Note that $\mathbf{x}_1 = -\,(3/2)\,\mathbf{x}_2 + (3/2)\,\mathbf{x}_3$. So substracting $-\,(3/2)\,\mathbf{x}_2 + (3/2)\,\mathbf{x}_3$ from $\mathbf{x}_1$ we get a new matrix with all 0's in its first column. Since this operation does not change the determinant

(**Why?**), then

$$\begin{vmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{vmatrix} = \left| \begin{bmatrix} 0 & 0 & 2 \\ 0 & 5 & 1 \\ 0 & -5 & 1 \end{bmatrix} \right| = 0$$

by expanding the determinant through the first column. ▲

Rank

Definition of Rank and Inverse

The rank is an extremely important and useful characteristic of every matrix. It concerns its number of LIN rows (and columns).

In the previous section we stated that a determinant of a matrix is zero when either any of its rows is a linear combination of other rows or any of its columns is a linear combination of other columns. Then a square matrix cannot have simultaneously LIN rows and linearly dependent columns. A natural question arises: What is the relationship between the number of LIN rows and LIN columns in a matrix? The next theorem provides an answer.

Theorem 7. The number of LIN rows in a matrix is the same as the number of LIN columns.

Proof. See Searle and Willett (2001), pp. 162-163. ■

The last theorem states that every matrix has the same number of LIN rows and LIN columns. This result guarantees the consistency of the next definition.

Definition 8. **(Rank)** The rank of a matrix is the number of LIN rows (and columns) in the matrix.

In terms of notation we will write r_A for the rank of matrix A. An alternative way to define r_A is by the number of non-zero rows in the row echelon form of matrix A. This definition offers a procedure to find the rank of any given matrix, but is less transparent with respect to the meaning of the concept.

Table 1 summarizes the relationships between matrix inverse, rank and linearly independence.

Table 1: Equivalent Statements for a Square Matrix A of order n

Inverse Existing	Inverse Non-Existing
A^{-1} exists	A^{-1} does not exist
$\|A\| \neq 0$	$\|A\| = 0$
$r_A = n$	$r_A < n$
A has n linearly independent rows	A has fewer than n linearly independent rows
A has n linearly independent columns	A has fewer than n linearly independent columns

Spanning sets, Bases and Subspaces

The space $\mathbb{R}^n$ is a special case of a wider concept called vector space. We say a set $S \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ if the next two conditions hold

(a) for every pair $\mathbf{x}_i$ and $\mathbf{x}_j$ such that $\mathbf{x}_i \in S$ and $\mathbf{x}_j \in S$, we also find that $\mathbf{x}_i + \mathbf{x}_j \in S$; and

(b) for every real scalar a and $\mathbf{x}_i \in S$, we also find that $a\mathbf{x}_i \in S$.

Suppose that every vector in S can be expressed as a linear combination of the t vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t$. Then this set of t vectors is said to span or generate S, and it is called a spanning set for S. When the vectors of a spanning set are also LIN then the set is said to be a basis for S. It is a basis and not the basis because for any S there may be many bases. All bases of a subspace S have the same number of vectors and that number is called the dimension of S, or $\dim(S)$.

Example 5. Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ and } \mathbf{x}_3 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

be three vectors in $\mathbb{R}^3$. Any two of them form a basis for the vector space in $\mathbb{R}^3$ that has 0 as the last element. The dimension of the space is 2. This space is in fact a subspace of $\mathbb{R}^3$. ▲

(**How many vectors form a basis for $S = \mathbb{R}^n$? Provide one simple basis.**)

For an $m \times n$ matrix A, the rows of A have n components. We write Row(A), row space of A, for the subspace of $\mathbb{R}^n$ spanned by the m rows of A. If $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_m \in \mathbb{R}^n$ are the m vectors that form the rows of A, then

$$\text{Row}(A) \equiv \{a_1\mathbf{r}_1 + a_2\mathbf{r}_2 + ... + a_m\mathbf{r}_m : a_1, a_2, ..., a_m \in \mathbb{R}\}.$$

By definition, $\dim[\text{Row}(A)] = r_A$.

Just as we formed the row space of a given matrix A as the set spanned by the rows of A, we can do the same for the columns of A. If A is an $m \times n$ matrix, then its columns have m components. We write Col(A), column space of A, for the subspace of $\mathbb{R}^m$ spanned by the n columns of A. If $\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n \in \mathbb{R}^m$ are the n columns of A, then

$$\text{Col}(A) \equiv \{a_1\mathbf{c}_1 + a_2\mathbf{c}_2 + ... + a_n\mathbf{c}_n : a_1, a_2, ..., a_n \in \mathbb{R}\}.$$

By Theorem 7 the number of LIN rows of a matrix equals the number of LIN columns, then $\dim[\text{Col}(A)] = \dim[\text{Row}(A)] = r_A$. This is not the same as to say Col(A) is equal to Row(A). (**Provide an example illustrating this claim.)**

System of Linear Equations

All previous results are extremely important in the analysis of systems of linear equations.

Definition of System of Linear Equations and Characterization of the Solution(s)

Linear equations in several unknowns can be represented as

$$A\mathbf{x} = \boldsymbol{\alpha} \tag{13}$$

where $\mathbf{x}$ is the vector of unknowns, $\boldsymbol{\alpha}$ is a vector of known values and A is the matrix of coefficients. In economics we often refer to $\mathbf{x}$ as the endogenous variables and to $\boldsymbol{\alpha}$ as the exogenous ones (or parameters). While the former are determined within the model, the latter are exogenously given.

Example 6. Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}.$$

Then (13) takes the form of either

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \quad \text{or} \quad \begin{array}{rcl} a_{11}x_1 + a_{12}x_2 & = & \alpha_1 \\ a_{21}x_1 + a_{22}x_2 & = & \alpha_2 \end{array}. \blacktriangle$$

System (13) can also be written as

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + ... + x_n\mathbf{c}_n = \boldsymbol{\alpha} \tag{14}$$

where $\mathbf{c}_j$ is the jth column of A, for $j = 1, ..., n$. When the system of linear equations is expressed in this way, it is apparent that $A\mathbf{x} = \boldsymbol{\alpha}$ has a solution if and only if $\boldsymbol{\alpha} \in \text{Col}(A)$, i.e. if $\boldsymbol{\alpha}$ belongs to the space generated by the columns of the coefficient matrix. Therefore, the statement [$A\mathbf{x} = \boldsymbol{\alpha}$ has a solution for all $\boldsymbol{\alpha} \in \mathbb{R}^m$] is equivalent to the statement [$\text{Col}(A)$ is all of $\mathbb{R}^m$]. The next theorem captures these observations.

Theorem 9. Let A be an $m \times n$ matrix, then

(a) $A\mathbf{x} = \boldsymbol{\alpha}$ has a solution for a particular $\boldsymbol{\alpha} \in \mathbb{R}^m$ if and only if $\boldsymbol{\alpha} \in \text{Col}(A)$; and

(b) $A\mathbf{x} = \boldsymbol{\alpha}$ has a solution for every $\boldsymbol{\alpha} \in \mathbb{R}^m$ if and only if $r_A = m$.

When the system of equations $A\mathbf{x} = \boldsymbol{\alpha}$ has a solution we say the system is consistent. Consistent equations may have more that one solution. To describe the set of solutions of a consistent system $A\mathbf{x} = \boldsymbol{\alpha}$, it is often easier to start studying the corresponding homogenous system $A\mathbf{x} = \mathbf{0}$.

The next theorem states that the solution set of a homogenous linear system of n variables is a subspace of $\mathbb{R}^n$.

Theorem 10. Let A be an $m \times n$ matrix. The set V of solutions to the system of equations $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$.

Proof. To this end we need to show that V is closed under vector addition and multiplication. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V, and let $r\mathbf{u} + s\mathbf{v}$ be a linear combination of the two vectors. Then

$$A(r\mathbf{u} + s\mathbf{v}) = Ar\mathbf{u} + As\mathbf{v} = rA\mathbf{u} + sA\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

since $\mathbf{u}$ and $\mathbf{v}$ are in V. Thus V is closed under linear combinations and it is a subspace of $\mathbb{R}^n$. ■

We next define null-space of A.

Definition 11. (Null-space) The subspace of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$ is called the null-space of A, and is written as $\text{Null}(A)$.

The solution sets to non-homogeneous equations are not subspaces, i.e. they do not have a linear structure. They are affine subspaces, which is to say they are translates of a subspace.

Definition 12. Let V be a subspace of $\mathbb{R}^n$ and let $\mathbf{d} \in \mathbb{R}^n$ be a fixed vector. The set

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{d} + \mathbf{v} \text{ for some } \mathbf{v} \in V\}$$

is called the set of translates of V by $\mathbf{d}$ and is written $\mathbf{d} + V$. Subsets of $\mathbb{R}^n$ of the form $\mathbf{d} + V$, where V is a subspace of $\mathbb{R}^n$, are called affine subspaces of $\mathbb{R}^n$.

Theorem 13. Let $A\mathbf{x} = \boldsymbol{\alpha}$ be an $m \times n$ system of linear equations. Let $\mathbf{d}$ in $\mathbb{R}^n$ be a solution of this system. Then every other solution $\mathbf{x}$ of $A\mathbf{x} = \boldsymbol{\alpha}$ can be written as $\mathbf{x} = \mathbf{d} + \mathbf{v}$ where $\mathbf{v}$ is a vector in the null-space of A. Then the solution set of $A\mathbf{x} = \boldsymbol{\alpha}$ is the affine subspace $\mathbf{d}+ \text{Null}(A)$.

Proof. Let $\mathbf{x}'$ solve $A\mathbf{x} = \boldsymbol{\alpha}$. Then

$$A\left(\mathbf{x}' - \mathbf{d}\right) = A\mathbf{x}' - A\mathbf{d} = \boldsymbol{\alpha} - \boldsymbol{\alpha} = \mathbf{0}$$

so $\mathbf{v} = \mathbf{x}' - \mathbf{d} \in \text{Null}(A)$.

Conversely, if $\mathbf{v} \in \text{Null}(A)$, then $A(\mathbf{d} + \mathbf{v}) = A\mathbf{d} + A\mathbf{v} = \boldsymbol{\alpha} + \mathbf{0} = \boldsymbol{\alpha}$. ■

The next theorem is often called Fundamental Theorem of Linear Algebra. It describes the dimension of $\text{Null}(A)$ in terms of r_A.

Theorem 14. Let A be an $m \times n$ matrix. Then

$$\dim \text{Null}(A) = n - r_A.$$

Proof. See Searle and Willett (2001), pp. 178-179. ■

Combining this result with Theorem 9 we get that when the system $A\mathbf{x} = \boldsymbol{\alpha}$ is consistent, then the solution set is an affine subspace whose dimension equals the number of variables minus the rank of A. (**Use this comment to provide necessary and sufficient conditions in terms of n, m and r_A for $A\mathbf{x} = \boldsymbol{\alpha}$ to have one, and only one, solution for all $\boldsymbol{\alpha} \in \mathbb{R}^m$.**)

Example 7. Let the system of linear equations be

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \end{pmatrix}. \tag{15}$$

Here $m = 2$, $n = 3$, and the rank of the coefficient matrix is 2. (**Why?**) By Theorem 9(b) the system (15) has a solution. To characterize the solution set let us study the corresponding homogenous system. Notice that $\dim\mathrm{Null}(A) = n - r_A = 1$. Then the $\mathrm{Null}(A)$ is a line through the origin in a three-dimensional space. This line can described by $(a, -a, 0)$ with $a \in \mathbb{R}$. In addition, notice that $\mathbf{d} = (1, 1, 1)$ is a particular solution to system (15). By Theorem 13, the solution set of (15) is given by all vector $\mathbf{x}$ that satisfies $\mathbf{x}^T = (1, 1, 1) + (a, -a, 0)$ with $a \in \mathbb{R}$. ▲

Crammer's Rule for Linear Equations

Crammer's rule is an easy method for solving linear equations. Assume A is a square matrix, and let us write the system of equations $A\mathbf{x} = \boldsymbol{\alpha}$ as in (14)

$$x_1\mathbf{c}_1 + ... + x_j\mathbf{c}_j + ... + x_n\mathbf{c}_n = \boldsymbol{\alpha}.$$

Crammer's rule states that x_j can be obtained as

$$x_j = \frac{\left| \begin{bmatrix} \mathbf{c}_1 & ... & \mathbf{c}_{j-1} & \boldsymbol{\alpha} & \mathbf{c}_{j+1} & ... & \mathbf{c}_n \end{bmatrix} \right|}{|A|}$$

where the numerator for x_j is just the determinant of a matrix derived from A by substituting its jth column by $\boldsymbol{\alpha}$.

Example 8. Consider a firm that produces two outputs, y_1 and y_2, by using only capital, K, and labor, L. The production function is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{bmatrix} K \\ L \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

where a, b, c and d are all strictly positive coefficients. (**Provide precise conditions under which this system has a unique solution for all y_1 and y_2.**)

By Crammer's rule

$$K = \frac{dy_1 - by_2}{ad - cb} \text{ and } L = \frac{ay_2 - cy_1}{ad - cb}. \blacktriangle$$

Linear Forms

At the beginning of this note, we defined a matrix as a rectangular array of elements arranged in rows and columns. This section states that there is a one-to-one correspondence between linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $m \times n$ matrices, i.e. each linear function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is an $\mathbf{f}_A$ for a unique $m \times n$ matrix A. So matrices are more than just rectangular arrays of elements, they are also representations of linear functions.

There is a convenient way to think about functions of the form $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ that we will invoke often. We can define mapping $\mathbf{f}$ as a vector of component functions f_i, each of which is a real-valued function of n variables

$$\mathbf{f} = (f_1, f_2, ..., f_m)^T \text{ with } f_i : \mathbb{R}^n \to \mathbb{R} \text{ for } i = 1, ..., m.$$

We next formalize the notion of linear function.

Definition 15. (Linear function) A linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function $\mathbf{f}$ that preserves the vector space structure

$$\mathbf{f}(\mathbf{x} + \mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{y}) \text{ and } \mathbf{f}(r\mathbf{x}) = r\mathbf{f}(\mathbf{x})$$

for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and all r in $\mathbb{R}$. Linear functions are sometimes called linear transformations.

An example of a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$ with $m = 1$ is the function

$$f(\mathbf{x}) = \mathbf{a^T x} = a_1 x_1 + a_2 x_2 + ... + a_n x_n$$

for some vector $\mathbf{a}^T = (a_1, a_2, ..., a_n)$ in $\mathbb{R}^n$. (**Check this function satisfies Definition 15.**)

The next theorem states that every linear real-valued function is of the form $f(\mathbf{x}) = \mathbf{a^T x}$.

Theorem 16. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function. Then, there exists a vector $\mathbf{a} \in \mathbb{R}^n$ such that $f(\mathbf{x}) = \mathbf{a^T x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. We develop the proof for $n = 3$. Let

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

be the canonical basis for $\mathbb{R}^3$. Let $a_i = f(\mathbf{e}_i)$ for $i = 1, 2, 3$; let $\mathbf{a}^T = (a_1, a_2, a_3)$. Then, for any vector $\mathbf{x} \in \mathbb{R}^3$,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$$

and

$$\begin{aligned} f(\mathbf{x}) &= f(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3) \\ &= f(x_1\mathbf{e}_1) + f(x_2\mathbf{e}_2) + f(x_3\mathbf{e}_3) \\ &= x_1 f(\mathbf{e}_1) + x_2 f(\mathbf{e}_2) + x_3 f(\mathbf{e}_3) \\ &= \mathbf{a}^{\mathbf{T}}\mathbf{x} \end{aligned}$$

which completes the proof. (**Generalize the proof for an arbitrary** n.) ■

Theorem 16 implies that every real-valued linear function on $\mathbb{R}^n$ can be associated with a unique vector $\mathbf{a} \in \mathbb{R}^n$ so that

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}.$$

Conversely, every such $\mathbf{a}$ induces a linear map.

The same correspondence between linear functions and matrices carries over to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. If A is an $m \times n$ matrix, then $\mathbf{f}_A(\mathbf{x}) = A\mathbf{x}$ is a linear function. The following theorem states the converse.

Theorem 17. Let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Then, there exists an $m \times n$ matrix A such that $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. See SB, p. 288. ■