Simon Fraser University, Department of Economics
Econ 798, Introduction to Mathematical Economics
Prof. B. Antoine
(based on the notes of Prof. A. Karaivanov)

Part 3: Linear Algebra and Matrices

1 Vectors, matrices and operations

1.1 Vectors

An n-dimensional real vector $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ is an ordered n-tuple of real numbers. As a matter of notational convention, whenever we talk of an n-dimensional vector, we mean a column vector; that is, a single column of n rows:

$$x = [x_i]_{i=1,\cdots,n} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

For the set of all real vectors we may define summation and scalar multiplication, as well as inner product, in the ordinary way. Then, the set of all real vectors $x \in \mathbb{R}^n$ is a linear vector space, and endowed with the Euclidean distance/norm forms our well known n-dimensional Euclidean vector space.

1.2 Matrices

A real matrix is defined as a rectangular array of real numbers[1] In particular, the matrix

$$A = [a_{ij}]_{j=1,\cdots,n;i=1,\cdots,m} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{12} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

for $a_{ij} \in \mathbb{R}\ \forall i \in \{1, 2, \cdots, m\}, j \in \{1, 2, \cdots, n\}$ and $m, n \in \mathbb{N} = \{1, 2, \cdots\}$, is a matrix of dimensions (m, n). More compactly, we write this matrix as $A = [a_{ij}]_{m,n}$. Notice that, just as an n-dimensional (real) vector $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$ is an ordered set or ordered n-tuple of

[1] In most of our discussion, we restrict focus to matrices over the set of real numbers, $\mathbb{R}$. All notions, however, can be extended to the complex plane, $\mathbb{C}$, or any arbitrary field. It's a good exercise for the reader to check what (if any) would have to change should, throughout this text, we had set $\mathbb{C}$ wherever $\mathbb{R}$ appears.

real numbers, an (m, n)-matrix is an ordered set or ordered n-tuple of m-dimensional vectors. Thus, we may write a matrix as an array of column vectors

$$A = [a_j] = [a_1 \ a_2 \ \cdots \ a_n]$$

with the understanding that $a_j = (a_{1j}, a_{2j}, \cdots, a_{mj}) \in \mathbb{R}^m$ is an m-dimensional column vector for all $j = 1, \cdots, n$.

1.3 Matrix transposition

The **transpose** A' of an (m, n)-matrix $A = [a_{ij}]$ is an (n, m)-matrix defined simply as

$$A' = [a_{ji}] = \begin{pmatrix} a_1' \\ a_2' \\ \cdots \\ a_n' \end{pmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \cdots & \cdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

1.4 Matrix Addition and Scalar Multiplication

The set of all matrices of the same dimensions forms a vector space. This requires only that we appropriately define matrix addition and scalar multiplication.

Definition 1 *Take any (m, n)-matrices $A = [a_{ij}]$, $B = [b_{ij}]$, $C = [c_{ij}]$, and any real scalars λ, μ. We define **matrix addition** as*

$$C = A + B \Leftrightarrow c_{ij} = a_{ij} + b_{ij} \quad \forall i, j$$

*and **scalar multiplication** as*

$$C = \lambda A \Leftrightarrow c_{ij} = \lambda a_{ij} \quad \forall i, j$$

We can then easily show the following properties:

Lemma 1.1 *Matrix addition and scalar multiplication, defined as above, satisfy:*

(i) Commutative Law: $A + B = B + A$;

(ii) Associative Law: $(A + B) + C = A + (B + C)$;

(iii) Distributive Law: $(\lambda + \mu)A = \lambda A + \mu A$;

(iv) Distributive Law: $\lambda(A + B) = \lambda A + \lambda B$;

(v) Distributive Law: $\lambda(\mu A) = (\lambda\mu)A$.

It then follows that.

Proposition 1.1 *The set of all (m, n) real matrices endowed with matrix addition and scalar multiplication forms a linear vector space.*

Exercise: Persuade yourself that you can prove the above Lemma 1.1 and Proposition 1.1 from first principles.

1.5 Matrix Multiplication

Let A and B be matrices. The matrix product of A and B (denoted AB) is defined when A is (m, n) and B is (n, p). When this holds (i.e., there are as many columns in A as there are rows in B), we say that A and B are conformable. The (i, k)-element of AB is then given by

$$(AB)_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

and AB is an (m, p)-matrix.
Matrix multiplication satisfies the following properties:

Lemma 1.2 *Provided dimension conformity, matrix multiplication satisfies:*

(i) Associative Law: $(AB)C = A(BC)$;

(ii) Associative Law: $(\lambda A)B = \lambda(AB)$;

(iii) Distributive Law: $A(B + C) = AB + AC$;

(iv) Distributive Law: $(A + B)C = AC + BC$.

1.6 Special Matrices

- An (m, n)-matrix A is called **square** if $m = n$.

- A matrix A is **symmetric** if $A' = A$, or equivalently $a_{ij} = a_{ji}, \forall i, j$. This imposes that A is square.

- A matrix A is **upper triangular** if it is square and $a_{ij} = 0$ for $j = i + 1, \cdots, n$ and $i = 1, \cdots, n - 1$. And A is **lower triangular** iff A' is upper triangular.

- A matrix A is **diagonal** if it is both upper and lower triangular; equivalently, iff $a_{ij} = 0$ for $i \neq j$.

- The **null matrix** (denoted 0) is defined by the relation: $A + (-1)A = 0$, or equivalently, $A + 0 = A$. It can thus take any dimension (it must have the same dimension as A), and is the matrix with all elements equal to zero.

- The **identity matrix** (denoted I) is defined by the relation: $IA = AI = A$. It is the diagonal matrix with ones on its diagonal, and takes the dimension necessary for conformability (that is, if A above is (m, n), I is (m, m) in the first part of the above relation and (n, n) in the second part).

Note that if AB exists, BA need not be defined; even if it does exist, it is not generally true that $AB = BA$ (when this is true, we say that A and B are commuting). This is why we need both (iii) and (iv) above; the first distributive law deals with pre-multiplication of a matrix sum by another matrix, while the second deals with post-multiplication of a matrix sum with another matrix.

1.7 Transposition rules

Lemma 1.3 *The following rules hold for matrix transposition:*

- $(A')' = A$;

- $(A + B)' = A' + B'$;

- $(AB)' = B'A'$;

- $A'A = 0 \Leftrightarrow A = A' = 0$;

- $A'A \neq AA \neq A'A'$ *in general, but* $A'A = AA = A'A'$ *for symmetric* A.

Note that AB *defined* $\Rightarrow (AB)' = B'A'$ *defined.*

2 Span, Basis, and Rank

2.1 Linear Combinations

Fix m and n; take a set of n vectors $\{a_j\} = \{a_1, a_2, \cdots, a_n\}$, where a_j is a (column) vector in $\mathbb{R}^m$; take n real numbers $\{x_j\} = \{x_1, \cdots, x_n\}$; and construct a vector y in $\mathbb{R}^m$ as the sum of the a_j weighted by the corresponding $x_j's$:

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

Then the so-constructed vector y is called a **linear combination** of the a_j.

If we let $A = [a_j]$ be the (m,n)-matrix with columns the vectors a_j and x the n-dimensional vector $[x_j]$, then we can write y as

$$y = Ax = \sum_{j=1}^{n} x_j a_j$$

Thus, Ax is a linear combination of the columns of A.

Notice that the dimension of the vector $y = Ax$ is the same as of that of any column a_j. That is, y belongs to the same vector space as the a_j.

2.2 Linear Dependence/Independence

Consider a set of n vectors in $\mathbb{R}^m$, $\{a_j\} = \{a_1, a_2, \cdots, a_n\}$. These vectors are called **linearly dependent** if any one of them can be written as a linear combination of the rest. They are otherwise called **linearly independent**.

Definition 2 *Letting $A = [a_j]$, the vectors a_j are linearly independent if*

$$Ax = 0 \Rightarrow x = 0$$

*They are **linearly dependent** if*

$$\exists x \in \mathbb{R}^m \text{ s.t. } x \neq 0 \text{ and } Ax = 0$$

Recall that two vectors x and z are called orthogonal iff $z'x = 0$. Thus, $Ax = 0$ means that x is orthogonal to all the rows of A. Similarly, $A'x = 0$ (for x now in $\mathbb{R}^m$ since A' is (n,m)) means that x is orthogonal to all the columns a_j of A.

Note that the maximum number of linearly independent vectors that we can have in any given collection of n-dimensional vectors is n. To see this, note that the equation $Ax = 0$ can be thought of as a system of n equations (where n is the dimension of A), with the number of unknowns being equal to the dimension of x, which is also the number of columns (vectors) in A. But, it is a fact that a homogenous system of equations (i.e., one with zeros on the right-hand side of every equation) with more unknowns than equations must have infinitely many solutions, all but one of which are nonzero. On the other hand, we know that we can write n linearly independent vectors of dimension n - the n-dimensional identity matrix consists of just such a collection.

2.3 The Span and the Nullspace of a Matrix, and Linear Projections

Consider an (m,n)-matrix $A = [a_j]$, with a_j denoting its typical column. Consider then the set of all possible linear combinations of the a_j. This set is called the span of the a_j, or the column span of A.

Definition 3 *(i) The **(column) span** of an* (m,n)*-matrix* A *is*

$$\begin{aligned} S(A) &\equiv S[a_1, \cdots, a_n] \\ &\equiv \{y \in \mathbb{R}^m \; s.t. \; y = Ax = \sum_{j=1}^{n} x_j a_j \; for \; some \; x = [x_j] \in \mathbb{R}^n \end{aligned}$$

We define the nullspace of A as the set of all vectors that are orthogonal to the rows of A.

Definition 4 *The **nullspace** or **Kernel** of an* (m,n)*-matrix* A *is*

$$\begin{aligned} N(A) &\equiv N[a_1, ..., a_n] \\ &\equiv \{x \in \mathbb{R}^n \; s.t. \; Ax = 0\} \end{aligned}$$

- **Exercise:** Given A, show that $S(B) \subseteq S(A)$ and $N(A') \subseteq N(B')$ whenever $B = AX$ for some matrix X. What is the geometric interpretation?

Notice that $N(A')$ is the set of all vectors that are orthogonal to the a_j. Thus,

$$z \in N(A') \Leftrightarrow z \perp S(A)$$

which means that $N(A')$ is the orthogonal complement subspace of $S(A)$. That is[2].

$$S(A) + N(A') = \mathbb{R}^m$$

- **Exercise** Given an (m,n)-matrix A, show that $S(A)$, $N(A)$ and $N(A')$ are all linear subspaces. Show further that $S(A)$ and $N(A')$ are orthogonal subspaces, in the sense that $z \in S(A)$, $u \in N(A') \;\; \Rightarrow \;\; z'u = 0$. Show further that $S(A) + N(A') = \mathbb{R}^m$, in the sense that for every $y \in \mathbb{R}^m$ there are vectors $z \in S(A)$ and $u \in N(A')$ such that $y = z + u$.

Remark: z is then called the (linear) projection of y on $S(A)$, or the regression of y on A, and u is called the residual, or the projection off $S(A)$. Does this reminds you something relevant to econometrics?

The last results are thus summarized in the following:

Lemma 2.1 *$S(A)$ and $N(A')$ form an orthogonal partition for $\mathbb{R}^m$; that is,*

$$S(A) \perp N(A') \; and \; S(A) + N(A') = \mathbb{R}^m$$

[2]Recall that, for any sets X, Y, their sum is defined as $X + Y \equiv \{z \; s.t. \; z = x + y, \text{ some } x \in X, \, y \in Y.\}$

2.4 Basis of a Vector Space

Let X be a vector space. We say that the set of vectors $\{a_1, \cdots, a_n\} \subset X$, or the matrix $A = [a_j]$, **spans** X iff $S(a_1, \cdots, a_n) = S(A) = X$.

If A spans X, it must be the case that any $x \in X$ can be written as a linear combination of the a_j. That is, for any $x \in \mathbb{R}^n$, there are real numbers $\{c_1, \cdots, c_n\} \subset \mathbb{R}$, or $c \in \mathbb{R}^n$, such that

$$x = c_1 a_1 + \cdots + c_n a_n \quad \text{or} \quad x = Ac$$

There may be only one or many c such that $x = Ac$. But if for each $x \in X$ there is only a unique c such that $x = Ac$, then c_j is called the j-th coordinate of x with respect to A. And then A is indeed a basis for X.

Definition 5 *A basis for a vector space X is any set of linearly independent vectors, or a matrix, that spans the whole X.*

• **Example** In $\mathbb{R}^n$, the usual basis is given by $\{e_1, \cdots, e_n\}$ where e_i is a vector with a unit in the i-th position and zeros elsewhere; alternatively, the (n, n)-identity matrix

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

is a basis for $\mathbb{R}^n$. If $x = (x_1, \cdots, x_n) \in \mathbb{R}^n$, then x_j are simply the coordinates of x with respect to I; that is,

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n = Ix$$

This means that $\{e_1, \cdots, e_n\}$, or I, spans $\mathbb{R}^n$. And trivially, the e_j are linearly independent, because

$$xI = x, \quad \text{or} \quad x_1 e_1 + \cdots + x_n e_n = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and thus $xI = 0 \Rightarrow x = 0$.

Observe that any space may admit many-many different bases! For instance:

• **Exercise** Show that if $\{e_j\}$ is a basis for X, then so is $\{f_j\} = \{\mu e_j\}$ for any scalar $\mu \neq 0$. And a bit less trivial:

• **Exercise** Suppose $\{e_j\}$ is a basis for X; let $P = [p_{ij}]$ be any nonsingular (n, n)-matrix, and let $f_j = \sum_i p_{ij} e_i$. Show then that $\{f_j\}$ is a basis for X as well.

In other words,

Lemma 2.2 *If E is a basis for X, then so is $F = EP$ for any nonsingular P.*

Notice that, with F so-constructed, P is then simply the coordinates of the basis F with respect to the basis E.

• **Example** For instance, all of the following matrices are bases for $\mathbb{R}^2$:

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad , \quad \begin{bmatrix} 3 & -1 \\ 2 & 5 \end{bmatrix} \quad , \quad \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix} \quad , \quad \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \quad ,$$

provided $\alpha\delta - \beta\gamma \neq 0$.

• **Example** Can you figure out what are the coordinates of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ with respect to the above alternative bases? What about any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$?

In the above example for $\mathbb{R}^2$, we found many different bases, but they all had something in common: They were all made of just 2 vectors, and we know well that 2 is the dimension of $\mathbb{R}^2$. But, what is the dimension of a vector space, and is this unique despite the multiplicity of bases? In answering this the following helps:

Lemma 2.3 *Let $\{e_j\} = \{e_1, \cdots, e_n\}$ be a basis for X, and let $\{b_j\} = \{b_1, \cdots, b_m\}$ be any set of vectors with $m > n$. Then $\{b_j\}$ can not be linearly independent.*

• **Exercise** Can you prove this? Here is a hint. Write $b_j = \sum_i c_{ij} e_i$ for some $\{c_{ij}\}$. Let $C = [c_{ij}]$ and let $x \neq 0$ be some solution to $Cx = 0$. [Which lemma/proposition ensures that such a solution exists?] Use that to show

$$\sum_i \lambda_i e_i = 0 \quad \text{for} \quad \lambda_i = \left(\sum_j x_j c_{ij} \right) \neq 0 \forall i$$

But isn't that a contradiction, which indeed completes the proof?

From the last lemma it follows that all bases of a given space will have the same number of elements. Therefore, we can unambiguously define:

Definition 6 *The dimension of a vector space X, $\dim(X)$, is the number of elements in any of its bases. On the other hand, if such a basis with finite elements does not exist, then the space is infinite dimensional.*

2.5 The Rank and the Nullity of a Matrix

The rank of matrix $A = [a_j]$ is defined as the maximum number of independent columns a_j of this matrix.

Definition 7 *The rank of a matrix A is the dimension of its span. The nullity of A is the dimension of its nullspace. That is,*

$$rank(A) \equiv \dim(S(A)) \quad and \quad null(A) \equiv \dim(N(A))$$

A useful result to keep in mind is the following.

Lemma 2.4 *Let any matrix A, and A' its transpose. Then, the rank of A and A' coincide:*

$$rank(A) = rank(A')$$

This simply means that a matrix always has as many linearly independent columns as linearly independent rows. Equivalently, a matrix and its transpose span subspaces of the same dimension. But, is there any relation between the rank and the nullity of a matrix? There is indeed, and this constitutes the "**fundamental theorem of linear algebra**":

Theorem 2.1 *Let any (m, n)-matrix $A = [a_j]$, with n columns $a_j \in \mathbb{R}^m$. Then, its rank and its nullity sum up to n:*

$$rank(A) + null(A) = n = \sharp aj$$

• **Exercise** Here is a sketch of the proof [not an easy one]; you have to fill the details. Let $k = null(A) \equiv dim(N(A))$. [Check that $k \leq m$.] Take a basis $\{e_1, \cdots, e_k, e_{k+1}, \cdots, e_n\}$ of $\mathbb{R}^n$ such that $\{e_1, \cdots, e_k\}$ is a basis for $N(A)$. For any $x \in \mathbb{R}^n$, there are $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$ such that

$$x = \lambda_1 e_1 + \cdots + \lambda_n e_n$$

and

$$Ax = \lambda_{k+1} Ae_{k+1} + \cdots + \lambda_n Ae_n$$

because

$$\{e_1, \cdots, e_k\} \subset N(A) \Rightarrow Ae_1 = \cdots = Ae_k = 0$$

Show further that the set $\{Ae_{k+1}, \cdots, Ae_n\}$ is linearly independent as well. Assume not and get a contradiction that $\{e_1, \cdots, e_n\}$ would then be linearly dependent. Conclude that $\{Ae_{k+1}, \cdots, Ae_n\}$ forms a basis for $S(A)$. Notice that $\{Ae_{k+1}, \cdots, Ae_n\}$ has (nk) elements, and thus $rank(A) = dim(S(A)) = n - k = n - null(A)$. QED

A related result is the following:

• **Exercise** Using the last theorem and the previous lemma, show that

$$\begin{aligned} rank(A) + null(A') &= m \\ null(A) - null(A') &= n - m \end{aligned}$$

Remark: Recall that $S(A)$ and $N(A')$ form a partition (an orthogonal one, indeed) of $\mathbb{R}^m$. It is not thus surprising that $dim(S(A)) + dim(N(A')) = dim(\mathbb{R}^m)$, or $rank(A) + null(A') = m$. From a 'transpose' view, $S(A')$ and $N(A)$ form a partition of $\mathbb{R}^n$, and thus $rank(A') + null(A) = n$, or $rank(A) + null(A) = n$, using the fact that $rank(A') = rank(A)$. Does this provide you with a clear geometric intuition for the above theorem?

2.6 Nonsingularity and Matrix Inversion

Definition 8 *A square matrix A of dimension (n, n) is nonsingular if $rank(A) = n$. Equivalently, A is nonsingular if $null(A) = 0$.*

Lemma 2.5 *If A is nonsingular then there exists a nonsingular matrix A^{-1} such that*

$$AA^{-1} = A^{-1}A = I$$

where I is the identity matrix.

Proof. Recall that if A is nonsingular, then its columns are linearly independent. Therefore, the equation $Ax = c$ has a unique solution x. In particular, $Ax = e_i$ has a unique solution, where e_i is the $(n, 1)$-column vector with a 1 as its ith element and zeroes elsewhere. We can stack n such equations $(Ax_1 = e_1, Ax_2 = e_2, \cdots, Ax_n = e_n)$ to show that $AX = I$ is satisfied by a unique matrix X; this matrix is a right inverse of A. To show that A has a left inverse, note that A nonsingular implies that the rows of A are linearly independent as well, and so the equation $yA = e_i'$, where e_i' is the transpose of e_i, has a unique solution x, where y is a row vector. Continuing as before we see that $YA = I$ is satisfied by a unique matrix Y, and this is a left inverse of A.

Finally, to show that $X = Y$ (and that both can therefore truly be called A^{-1}), we will show that if $AX = YA = I$, then $X = Y$. To see this, suppose it's not true; $AX = YA = I$ but $X \neq Y$. Then premultiply the left-hand side of the inequality by YA and post-multiply the right-hand side by AX (which changes nothing, since both are equal to the identity matrix) and we have $YAX \neq YAX$, which is false and establishes the assertion. ∎

The following facts about matrix inverses are useful (assuming invertibility):

Lemma 2.6 $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. Observe that if we postmultiply AB by its inverse $(AB)^{-1}$ we get I by the definition of matrix inverse, and similarly if we premultiply AB by $B^{-1}A^{-1}$ we get I, so we can apply the previous lemma to conclude that $(AB)^{-1} = B^{-1}A^{-1}$. ∎

Lemma 2.7 $(A')^{-1} = (A^{-1})'$.

Proof. We want to show that $A'(A^{-1})' = (A^{-1})'A' = I$. But this follows directly from our transpose rules, taking the transpose on both sides of the equalities $AA^{-1} = I$ (which gives $(A^{-1})'A' = I$) and $A^{-1}A = I$ (which gives $A'(A^{-1})' = I$). ∎

Note that the fact shown in the first of the two proofs above implies that the left-hand inverse and the right-hand inverse of a matrix are the same, so that we can talk meaningfully of the inverse of a nonsingular square matrix

2.7 Powers of a Matrix

For any square matrix A we may define its k-th power, denoted by A^k, for any $k \in \mathbb{N}$. A^k is defined inductively by $A^0 \equiv I$ and $A^k \equiv AA^{k-1}$ for $k \in N_* = \{1, 2, , \cdots\}$.

If A is invertible, then we also define $A^{-k} \equiv (A^{-1})^k$, for any $k \in \mathbb{N}$. Actually it can be shown that $A^{-k} = (A^k)^{-1}$.(See the practice exercise)

Remark: So far we have not defined A^k for a non-integer but real number k. This will become possible for nonsingular and symmetric, or any diagonalizable matrices, but we have to defer till the point we discuss matrix eigensystem and diagonalization.

2.8 The Determinant of a Matrix

The determinant, $\det(A)$ or $|A|$, of a matrix A is defined iff A is square.

Definition 9 *The determinant of a square matrix A with dimension (n, n) is a mapping $A \to \det(A)$ such that*

i) $\det(.)$ *is linear in each row of A.*

ii) if rank$(A) < n$ then $\det(A) = 0$ *and vice versa.*

iii) $\det(I) = 1$

We can view the matrix A as a collection of n row vectors $\{a_1, a_2, \cdots, a_n\}$ where each $a_i \in \mathbb{R}^n$. The determinant is then a function mapping the set of vectors $\{a_1, a_2, \cdots, a_n\}$ into $\mathbb{R}$. (We will not prove this here, but this mapping exists and it is unique, for any square matrix A). We can write $D(a_1, a_2, \cdots, a_n) \to \mathbb{R}$. Property i) of the previous definition means that

$$D(a_1, a_2, \cdots, \lambda a_i + \mu\tilde{a}_i, \cdots, a_n) = \lambda D(a_1, a_2, \cdots, a_i, \cdots, a_n) + \mu D(a_1, a_2, \cdots, \tilde{a}_i, \cdots, a_n)$$

for any scalars $\lambda, \mu \in \mathbb{R}$ and any vectors $a_i, \tilde{a}_i \in \mathbb{R}$. In particular note that

$$D(a_1, a_2, \cdots, \lambda a_i, \cdots, a_n) = \lambda D(a_1, a_2, \cdots, a_i, \cdots, a_n)$$

Alternatively, we can give an inductive definition, providing a computational algorithm for the determinant.

Definition 10 *Let an (n, n)-matrix $A = [a_{ij}]$. If $n = 1$, and hence $A = [a_{11}]$, then* $\det(A) \equiv a_{11}$. *For any $n > 1$, we let*

$$|A| \equiv \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j}) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots \pm a_{1n}\det(A_{1n})$$

where $[a_{1j}]$ is the first row of A and A_{1j} is the $((n-1), (n-1))$-submatrix constructed by erasing the first row and the j-th column of A.

This definition implies for a $(2,2)$-matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

that $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

For a $(3,3)$-matrix, $A = [a_{ij}]$, the determinant can be computed by the Sarrus Rule, which works as follows:

- First border the matrix at its right with its first two columns;

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{matrix}$$

- then take the sum of the products of the elements along the parallels of the principal diagonal minus the sum of the products of the elements along the parallels to the other diagonal; this gives the determinant of A:

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

Sometime useful in speeding up computation, the following lemma allows us to work out the determinant along any row or column (also known as expansion by cofactors).

Lemma 2.8 *For any fixed $i = 1, \cdots, n$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) = (-1)^{i+1}(a_{i1} \det(A_{i1}) a_{i2} \det(A_{i2}) + \cdots \pm a_{in} \det(A_{in}))$$

and for any fixed $j = 1, \cdots, n$

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) = (-1)^{j+1}(a_{1j} \det(A_{1j}) a_{2j} \det(A_{2j}) + \cdots \pm a_{nj} \det(A_{nj}))$$

where $[a_{ij}]_{j=1,\cdots,n}$ is the i-th row of A, $[a_{ij}]_{i=1,\cdots,n}$ is the j-th column of A, and A_{ij} is the $((n-1),(n-1))$ submatrix constructed by erasing the i-th row and the j-th column of A.

Finally, keep in mind the following:

Lemma 2.9 *i)* $\det(AB) = \det(A)\det(B)$

ii) $\det(A) = \det(A')$

iii) $\det(A^{-1}) = 1/\det(A)$

The determinant of a matrix allows an interesting interpretation in terms of the surface (more generally volume) of the vectors comprising a matrix.

2.9 Matrix Inversion, part II

The ordinary test for invertibility of a matrix is whether its determinant equals zero or not.

Theorem 2.2 *A matrix A is invertible iff* $\det(A) \neq 0$.

Provided $\det(A) \neq 0$, A^{-1} exists for sure. However, the computation of A^{-1} is generally a pain, especially when $n > 3$. An algorithm to compute the inverse of a given (n,n)-matrix $A = [a_{ij}]$ works as follows:

- For $n = 1$, and hence $A = [a_{11}]$, then $A^{-1} = \left[\frac{1}{a_{11}}\right] = \left[\frac{1}{\det(A)}\right]$.
- For $n > 1$: Let A_{ij} denote the $((n-1),(n-1))$-submatrix constructed by erasing the i-th row and the j-th column of A. Let $\det(A_{ij})$ be called the (i,j)-first-order minor of A and define the (i,j) cofactor of A as

$$c_{ij} = (-1)^{i+j} \det(A_{ij})$$

Notice by the way that

$$\begin{aligned}
\det(A) &= \sum_i a_{ij} c_{ij} \\
&= \sum_j a_{ij} c_{ij} \\
&= \sum_j (-1)^{i+j} a_{ij} \det(A_{ij})
\end{aligned}$$

For example,

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(A_{1n})$$

Construct the (n,n)-matrix $C = [c_{ij}]$ of all first-order cofactors of A and define the adjoint of A as the transpose of C,

$$adj A \equiv C' = [c_{ji}]$$

Then, the inverse of A is

$$\begin{aligned}
A^{-1} &= \frac{1}{\det(A)} adj A \\
&= \frac{1}{\det(A)} \begin{bmatrix} +\det(A_{11}) & -\det(A_{21}) & \cdots & (-1)^{n+1}\det(A_{n1}) \\ -\det(A_{12}) & +\det(A_{22}) & \cdots & (-1)^{n+2}\det(A_{n2}) \\ \vdots & \vdots & \vdots & \vdots \\ (-1)^{n+1}\det(A_{1n}) & (-1)^{n+2}\det(A_{1n}) & \cdots & +\det(A_{nn}) \end{bmatrix}
\end{aligned}$$

This algorithm works pretty well for manual computation if $n = 2$ or 3.

• **Exercise** Show that for a (2,2) matrix

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

provided $\det(A) = \alpha\delta - \beta\gamma \neq 0$, the inverse is

$$A^{-1} = \frac{1}{\alpha\delta - \beta\gamma} \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix}$$

But, otherwise, thanks to modern technology, computers can be used in computing the inverse of a huge matrix.

2.10 Matrix Inversion and Linear Independence

This is probably the most important characterization result to remember.

Theorem 2.3 *Let A be a (n, n) square matrix. Then, the following conditions are equivalent*

(i) A is nonsingular; i.e., A^{-1} exists;

(ii) A has a non-zero determinant, $\det(A) \neq 0$;

(iii) the columns $\{a_j\}$ of A are linearly independent; i.e., $Ax = 0 \Rightarrow x = 0$;

(iv) A forms a basis for $\mathbb{R}^n$;

(v) A spans the whole n-dimensional space, $S(A) = \mathbb{R}^n$;

(vi) the kernel of A is null, $N(A) = \{0\}$.

Further, the following lemma will prove useful when we analyze linear equations systems.

Lemma 2.10 *Let $1 \leq m < n$ and an (m, n) matrix $A = [a_{ij}]$. Then there is $x \in \mathbb{R}^n$, $x \neq 0$ such that $Ax = 0$. This in turn implies $null(A) \equiv \dim(N(A)) \geq 1$.*

Remark: If we interpret $N(A)$ as the set of solutions to the system $Ax = 0$ (see next section for details), the above lemma says that, whenever there are more unknowns than equations $(n > m)$, then the system $Ax = 0$ is underdetermined and admits a continuum of solutions.

• **Exercise** Construct an (n, n) matrix $B = [b_{ij}]$ by $b_{ij} = a_{ij}$ for $i = 1, \cdots, m$ and $b_{ij} = 0$ for $i = m + 1, \cdots, n$, $\forall j$; more compactly, $B = \begin{bmatrix} A \\ 0 \end{bmatrix}$. What is $\det(B)$? What does this imply for Bx? Notice that $Bx = \begin{bmatrix} Ax \\ 0 \end{bmatrix}$. Does this help you prove the above lemma?

3 Systems of Equations

3.1 Linear Systems

A linear system of m equations in n unknowns (or, for brevity, an (m,n) linear system) is generally written as:

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

where the a_{ij}'s and b_i's ($i = 1, \cdots, m$ and $j = 1, \cdots, n$) are coefficients and constants taken from $\mathbb{R}$ and the x_j's are the unknowns. A solution to (1) is any n-tuple of real numbers x_j's that satisfy simultaneously all m equations in (1). We may write the above system in a more compact format if we let A be an (m,n) matrix formed by the coefficients a_{ij}'s, x the $(n,1)$ column vector of the unknowns x_j's, and b the $(m,1)$ column vector of constants b_i's:

$$
A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}
x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}
$$

Then we rewrite (1) simply as

$$
Ax = b \tag{2}
$$

We call this an (m,n) system, and we call it square iff $m = n$.

The set of solutions for (2) is simply the set

$$
X^* \equiv X^*(A,b) \equiv \{x \in \mathbb{R}^n \ s.t. \ Ax = b\}
$$

In general, for given A and b, X^* may be empty (no solution), or be a singleton (unique solution), or have more than one elements (multiple solutions). Our task is now to identify conditions on the given A and b that give rise to each case (none, unique, or many solutions).

Remark: We emphasize that, in our context, whenever we talk of a 'solution' we mean a solution in the field of real numbers, not in the field of complex numbers. But this is not at all a restriction. In fact, as long as A and b are real, then any complex solution to $Ax = b$ has to be real. Why so? Simply because if x was not real, while A real, then Ax would not be real, contradicting $Ax = b$ and b real.

Example: Consider the following three (2,2) linear systems:

$$
\begin{aligned}
x_1 + x_2 &= 0 \\
2x_1 + 2x_2 &= 1
\end{aligned}
\tag{3}
$$

or

$$x_1 + x_2 = 0 \qquad (4)$$
$$x_1 + 2x_2 = 1$$

or

$$x_1 + x_2 = 0 \qquad (5)$$
$$2x_1 + 2x_2 = 0$$

The question we ask is: Why does (3) admit no solution at all, (4) only one solution, and (5) a continuum of solutions? [Can you verify that claim? Can you find the set of solutions yourself?]

3.2 Solution of linear system of equations

Consider the (m, n) system,

$$Ax = b$$

where $A = [a_j] = [a_{ij}]$ is the (m, n) matrix of coefficients, $x = [x_j]$ is the $(n, 1)$ vector of unknowns, and $b = [b_i]$ the $(m, 1)$ vector of constants; let also $a_j \in \mathbb{R}^m$ be the j-th column of A, so that $A = [a_1 \ \cdots \ a_n]$.

What does $b = Ax$ mean? Notice that

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

This is just a linear combination of the columns a_j's of A, with the x_j's being the corresponding weights. Therefore, $b = Ax$ simply means that the given vector $b \in \mathbb{R}^m$ can be written as a linear combination of the columns a_j's of A. Equivalently, $b = Ax$ means that b falls into the subspace spanned by A.

But recall that b falls into the span of A, and can be written as a linear combination of the a_j's, if and only if the matrix formed by stacking b together with all a_j's is singular, which also means that its span coincides with that of A alone. Thus, letting

$$\begin{aligned} S(A) &\equiv S[a_1, \cdots, a_n] \\ &\equiv \{y \in \mathbb{R}^m \ s.t. \ y = Ax = \sum_{j=1}^{n} x_j a_j \quad \text{for some } x = [x_j] \in \mathbb{R}^n\} \end{aligned}$$

be the span of $A = [a_j]$ and

$$[A, b] = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ a_{21} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix}$$

be the bordered matrix of coefficients, and by appealing to Theorem 2.3, we have:

Lemma 3.1 *The set of solutions to $Ax = b$ is nonempty if and only if b falls into the span of A; and this holds if and only if the bordered matrix $[A, b]$ is of the same rank with A, or equivalently spans the same space with A:*

$$\begin{aligned} X^*(A,b) \neq 0 &\Leftrightarrow b \in S(A) \\ &\Leftrightarrow S[A,b] = S(A) \\ &\Leftrightarrow rank[A,b] = rank(A) \end{aligned}$$

Also, recall that, given an (m, n) matrix A, the span $S(A)$ and the nullspace $N(A')$ of A form an orthogonal partition for the whole $\mathbb{R}^m$. Thus the set of solutions is empty if and only if the residual of the projection of b on $S(A)$ is nonzero, or equivalently b is not orthogonal to $N(A')$.

We now consider two complementary subclasses of linear systems: those that have $b = 0$, and those that have $b \neq 0$. We call the former homogeneous and the latter non-homogeneous.

3.3 Homogeneous Linear Systems

We here consider linear systems $Ax = b$ with $b = 0$. Such systems are called homogeneous.

Recall that $Ax = b$ has a solution (at least one) if and only if $b \in S(A)$; here, if and only if $0 \in S(A)$. Observe then that, since the span $S(A)$ of any matrix A is a subspace, the zero vector is always an element of the span of $S(A)$. Thus, if $b = 0$, the system $Ax = 0$ always has a solution, whatever the matrix A is. Indeed, $x = 0$ is always a solution to $Ax = 0$:

$$b = 0 \Rightarrow 0 \in X^* \qquad (*)$$

Moreover, by definition indeed, for a homogeneous system we have that

$$b = 0 \Rightarrow X^* = \{x | Ax = 0\} \equiv N(A)$$

so that the set of solutions to $Ax = 0$ is simply the nullspace of the coefficient matrix A. And since $N(A)$ is a vector space as well, $0 \in N(A)$, which is an alternative way to say (*).

But in general $x = 0$ may not be the unique solution. When will $x = 0$ be the unique solution to $Ax = 0$? From Theorem 2.3 we know that $Ax = 0$ implies that $x = 0$ necessarily, if and only if all the columns a_j of A are linearly independent. That is, $x = 0$ is the unique solution to $Ax = 0$ if and only if $rank(A) = n$. In this case, the span of A is the whole space, $S(A) = \mathbb{R}^n$ and $rank(A) = \dim(\mathbb{R}^n)$, and its nullspace is zero-dimensional, $N(A) = \{0\}$ and $null(A) = 0$.

But we know that $rank(A) \leq \min\{n, m\}$, and thus $rank(A) = n$ implies $m \geq n$ necessarily. That is, for the solution to be unique we need at least as many equations as unknowns. It follows that if there are less equations than unknowns, $m < n$, then $Ax = 0$ must have more than one solution. With $m < n$, indeed, the columns of A are necessarily linearly dependent;

in particular, the nullspace of A has dimension at least 1, and exactly as many as $null(A) = n - rank(A) \geq 1$ columns of A can be written as linear combinations of the rest.

In this case, $Ax = 0$ has a nonzero solution as well. Since $Ax = 0$ implies $Az = 0$ for any $z \in S(x)$, and $x \neq 0 \Rightarrow S(x) = \mathbb{R}$, it follows that $Ax = 0$ has indeed a continuum of solutions. More precisely, if we can find as many as k linearly independent solutions $x_1, \cdots, x_k$ to $Ax = 0$, then any $z \in S[x_1, \cdots, x_k]$ is a solution as well.

- **Exercise:** Prove the above claim. That is, prove that $Ax_1 = Ax_2 = 0 \Rightarrow Az = 0 \quad \forall z \in S[x_1, x_2]$.

We observe that the maximum number k of independent solutions $x_1, \cdots, x_k$ to $Ax = 0$ is simply (by definition indeed) the dimension of the nullspace $N(A) \equiv \{x | Ax = 0\}$ of A; that is $k = null(A) = \dim(N(A))$ of A. Recall then that, for an (m, n) matrix A, it is true that $null(A) = n - rank(A)$. Therefore, there are as many independent solutions to $Ax = 0$ as $k = n - rank(A)$.

Lemma 3.2 *For any homogeneous system $Ax = 0$,*

$$X^* = N(A) \quad \textit{and} \quad \dim(X^*) = n - rank(A)$$

Further, if $m > rank(A)$, then as many as $m - rank(A)$ equations are redundant, in the sense that they are implied by the rest $rank(A)$ equations. Thus, any (m, n) homogeneous system with $m > \rho$ can be reduced to an (ρ, n) system, where $\rho = rank(A)$. Therefore, without any loss of generality, from now on consider only (m, n) systems with $m = \rho = rank(A) \leq n$. We can then distinguish two cases:
either $m = n = rank(A)$,
or $m = rank(A) < n$.
Then $m = rank(A)$ is the 'effective' number of equations (that is, the number of linearly independent equations), while n is the number of unknowns.

- Case I: $n = m = rank(A)$
  In this case we have as many equations as unknowns and A is a nonsingular (n, n) matrix. Then, $Ax = 0$ if and only if $x = 0$. Hence, $x = 0$ is the unique solution to $Ax = 0$, $X^* = \{0\}$, and $\dim(X^*) = 0 = n - m$.

- Case II: $n > m = rank(A)$
  In this case we have more unknowns than equations and A is a singular (m, n) matrix with $rank(A) = m < n$. Then, $Ax = 0$ has as many independent solutions as $k = n - rank(A) = n - m$. This means that we may freely choose $(n - m)$ values for, say, the first $(n - m)$ unknowns $(x_1, \cdots, x_{n-m})$ and then the system $Ax = 0$ pins down the values for the rest m unknowns $(x_{n-m+1}, \cdots, x_n)$. And then $\dim(X^*) = n - m \geq 1$.

So, now let us generalize to arbitrary number of equations and unknowns. Let A be an (m,n) matrix for arbitrary m, n and let $\rho \leq \min\{m,n\}$ be its rank. Then, partition A and x as follows:

$$A = \begin{bmatrix} D & B \\ C & \tilde{A} \end{bmatrix} \qquad x = \begin{bmatrix} z \\ \tilde{x} \end{bmatrix} \tag{6}$$

where $\tilde{A}$ is a full-rank (ρ,ρ) matrix, for $\rho = rank(A) = rank(\tilde{A})$, $z = (x_1, \cdots, x_{n\rho})$ is $((n-\rho),1)$ and $\tilde{x} = (x_{n-\rho+1}, \cdots, x_n)$ is $(\rho, 1)$. Check the dimensions of B, C, D and notice that

$$Ax = \begin{bmatrix} Dz + B\tilde{x} \\ Cz + \tilde{A}\tilde{x} \end{bmatrix}$$

so that

$$Ax = 0 \Leftrightarrow \begin{cases} Dz + B\tilde{x} = 0 \\ Cz + \tilde{A}\tilde{x} = 0 \end{cases}$$

As we explained before, the first $(m\rho)$ equations are redundant. That is, $Cz + \tilde{A}\tilde{x} = 0$ implies $Dz + B\tilde{x} = 0$ as well. Thus,

$$Ax = 0 \Leftrightarrow Cz + \tilde{A}\tilde{x} = 0$$

Since $\tilde{A}$ has full rank, it is invertible, and therefore we get

$$Ax = 0 \Leftrightarrow \tilde{x} = -\tilde{A}^{-1}Cz$$

This means that any $x = (z, \tilde{x})$ such that $\tilde{x} = -\tilde{A}^{-1}Cz$, for any $z \in \mathbb{R}^{n-\rho}$, is a solution to $Ax = 0$. Therefrom it also follows that $\dim(X^*) = n - \rho$. And conversely, x is a solution to $Ax = 0$ only if a partition like the above is possible. Therefore, we can summarize our results so far in the following theorem.

Theorem 3.1 *Consider the (m,n) homogeneous system $Ax = 0$. The set of solutions always includes 0 and thus is nonempty; and is given by the nullspace of A:*

$$X^* = N(A) \equiv \{x | Ax = 0\}$$

The dimension of X^* is simply the nullity of A:

$$\dim(X^*) = null(A) = n - rank(A) \geq 0$$

Whenever $m > rank(A)$, as many equations as $m - rank(A)$ are redundant. Further, the solution is unique at $x = 0$ if and only if A is of full rank,

$$X^* = \{0\} \Leftrightarrow rank(A) = n$$

Otherwise, there is a continuum of solutions of the form

$$X^* = \{x \in \mathbb{R}^n | x = (z, -\tilde{A}^{-1}Cz) \text{ for some } z \in \mathbb{R}^{n-rank(A)}\}$$

with $\tilde{A}$ being any square submatrix of A with $rank(\tilde{A}) = rank(A)$ and C then being as in (6).

- **Exercise:** Consider the (3,3) system $Ax = 0$ for

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

Show that $rank(A) = 2$, and partition A appropriately so as to apply what we did before. What is the set of solutions?

3.4 Non-homogeneous Linear Systems

In the previous subsection we consider linear equation systems with $b = 0$. Now consider systems

$$Ax = b \quad \text{for} \quad b \neq 0$$

We repeat that existence of a solution means that $b \neq 0$ can be written as a linear combination of the columns in A, or that b falls into the span of A.

Consider first the case that $m = n = rank(A)$. Then A is square and has full rank, so that it is nonsingular and is a basis for the whole $\mathbb{R}^n$. It follows that $b \in S(A)$ necessarily. Moreover, since A is invertible,

$$Ax = b \Leftrightarrow x = A^{-1}b$$

Thus in this case the set of solutions is singleton, $X^* = \{A^{-1}b\}$. The result works conversely as well, and even if $b = 0$. Thus we have:

Lemma 3.3 *A square (n, n) system $Ax = b$ has a unique solution if and only if A has full rank, which means that A is nonsingular, or equivalently* $\det(A) \neq 0$. *Then, $x = A^{-1}b$ is the unique solution.*

Remark: Notice that, in the above case, A is a basis for $\mathbb{R}^n$, where b belongs. Hence the geometric interpretation of the solution is that $x = A^{-1}b$ gives the (unique) coordinates of b with respect to the basis A. Now suppose that A is not of full rank, but still $rank[A, b] = rank(A)$. This still implies $b \in S(A)$, and at least one solution exists. But now (with $b \neq 0$), the solution is not unique! Instead, we have a whole continuum of solutions!

On the other hand, letting $[A, b]$ be the bordered matrix formed as

$$[A, b] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{bmatrix}$$

we observe that if $S[A, b]$ is strictly bigger than $S(A)$, which is equivalent to $rank[A, b] > rank(A)$, then it must be the case that b cannot be written as a linear combination of the columns of A; that is, $b \notin S(A)$ and rather the projection of b on $N(A)$ is nonzero. Therefore:

Lemma 3.4 *The (nonhomogeneous) system $Ax = b$ has no solution if and only if the rank of $[A, b]$ exceeds that of A,*

$$X^* = \emptyset \Leftrightarrow rank[A, b] > rank(A)$$

The situation is indeed similar to the homogeneous case. In particular, we may rewrite $Ax = b$ equivalently as

$$[A, b]y = 0$$

where

$$y = \begin{bmatrix} x \\ -1 \end{bmatrix} = (x_1, \cdots, x_n, -1)$$

Notice that $[A, b]$ is $(m, (n + 1))$ and y is $(n + 1), 1)$, with $y \neq 0$ by construction.. This way, we have in fact transformed the non-homogeneous system $Ax = b$ into a homogeneous one, $[A, b]y = 0$. The important constraint is only that we require, by construction indeed, that $y \neq 0$. Thus, for $Ax = b$ to have any solution we need that $[A, b]y$ has a nonzero solution. But the latter, as we showed before, is possible if and only if the bordered matrix $[A, b]$ is singular. If instead $[A, b]$ is nonsingular, and $null[A, b] = 0$, then $Ax = b$ has no solution.

Moreover, if $null[A, b] = 1$, then the set of solutions y of $[A, b]y = 0$ is a single-dimensional line, and thus $Ax = b$ has a unique solution. In particular, the point y of this line that has -1 as its last coordinate gives us the unique solution to $Ax = b$. In fact:

- **Exercise** Show that $null[A, b] = 1$ if and only if $rank[A, b] = rank(A)$.

If $null[A, b]2$, then the set of solutions y of $[A, b]y = 0$ is a hyperplane of dimension equal to $null[A, b] \geq 2$, and thus the set of solutions x of $Ax = b$ is a hyperplane with dimension equal to $null[A, b] - 1 \geq 1$.

- **Exercise** Persuade yourself that, if $rank([A, b]) = rank(A)$, then and only then $X^* \neq \emptyset$, and further $\dim(X^*) = null[A, b] - 1$.

We can thus summarize our results in the following theorem.

Theorem 3.2 *Consider the (m, n) system $Ax = b$, with either $b \neq 0$ or $b = 0$. We distinguish the following cases:*

- *(Unique Solution) If $rank[A,b] = rank(A) = n \leq m$, then and only then the system has a unique solution. In this case, indeed, as many as $m-n$ equations are redundant, and, provided an appropriate partition, $X^* = \{\tilde{A}^{-1}\tilde{b}\}$.*

- *(No Solution) If $rank[A,b] > rank(A)$, which necessarily implies $b \neq 0$ and $m > rank(A)$, then and only then the system has no solution, $X^* = \emptyset$.*

- *(Multiple Solutions) If $rank[A,b] = rank(A)$, but $rank(A) < n$, then and only then the system has multiple solutions, and then $\dim(X^*) = null[A,b] - 1 = n - rank(A) \geq 1$.*

When there is a unique solution, we say that the system is **exactly determined**. When there is no, the system in **overdetermined**. When there are many solutions, the system is **under-determined** (or indeterminate).

• **Exercise** Let $m = n$, and consider $Ax = b$. Suppose that $\det(A) = 0$, so that the system is either underdetermined or overdetermined. What of the two cases arises if $b = 0$? And what happens if $b \neq 0$? Next consider $m > n = rank(A)$ and characterize the appropriate partition that gives $X^* = \{\tilde{A}^{-1}\tilde{b}\}$.

3.5 Finding the Solution: Cramer's rule

We have identified the conditions under which a square system $Ax = b$ has a unique solution: this is so if and only if A is invertible. Then and only then the unique solution is given by

$$x = A^{-1}b$$

Calculating this requires that we first calculate the inverse A^{-1}. (Recall the algorithm that was presented in section 2.9)

The inverse of A is then given as

$$\begin{aligned} A^{-1} &= \frac{1}{\det(A)} adj(A) \\ &= \frac{1}{\det(A)} \begin{bmatrix} +\det(A_{11}) & -\det(A_{21}) & \cdots & (-1)^{n+1}\det(A_{n1}) \\ -\det(A_{12}) & +\det(A_{22}) & \cdots & (-1)^{n+2}\det(A_{n2}) \\ \vdots & \vdots & \cdots & \vdots \\ (-1)^{n+1}\det(A_{1n}) & (-1)^{n+2}\det(A_{1n}) & \cdots & +\det(A_{nn}) \end{bmatrix} \end{aligned}$$

where A_{ij} is the $((n-1),(n-1))$ matrix formed by erasing the i-th row and the j-th column of A, and $\det(A_{ij})$ is the (i,j) minor of A.

An alternative way to calculate the solution is to use the Cramer's Rule. Let B_j be the (n,n) matrix formed by taking A and substituting its j-th column, a_j, with the constants vector, b.

For instance, for $j = 2$,

$$B_2 = [a_1 \; b \; a_3 \; \cdots \; a_n] = \begin{bmatrix} a_{11} & b_1 & a_{13} & \cdots & a_{1n} \\ a_{21} & b_2 & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & b_n & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

and so on. Let $\det(A) \neq 0$ and $\det(B_j)$ be the determinants of A and B_j, respectively. Cramer's rule then says that the j-th element of the solution $x = A^{-1}b$ is given by

$$x_j = \frac{\det(B_j)}{\det(A)} \quad \forall j = 1, \cdots, n$$

NB: Cramer's rule is good to know, but if you ever have to invert a numerical matrix, you'd better turn to a software like Matlab or Mathematica.

4 Eigensystems

In this section we study eigenvalues and eigenvectors of a given matrix A. These can be used to transform the matrix A into a simpler form which is useful for solving systems of linear equations and analyzing the properties of the mapping described by A.

Definition 11 *We say that λ is an eigenvalue of an (n, n) matrix A with corresponding eigenvector v if*

$$Av = \lambda v$$

for some $v \neq 0$. Conversely, we say that $v \neq 0$ satisfying the equation is an eigenvector corresponding to the eigenvalue λ.

Note that we can rewrite the above equation as $(A - \lambda I)v = 0$. This homogenous system of equations has a nontrivial solution if and only if the matrix $A - \lambda I$ is singular, which in turn holds iff $\det(A - \lambda I) = 0$. This leads to a characterization of the eigenvalues as solutions to the equation $\det(A - \lambda I) = 0$. Note that $\det(A - \lambda I)$ is a polynomial of degree n in λ (why?). It thus out that it has at most n, possibly complex, roots.

Definition 12 *Let A be any (n, n) matrix and I the (n, n) identity matrix. The characteristic polynomial of A is $\xi(\lambda) \equiv \det(A - \lambda I)$. Its characteristic equation is $\xi(\lambda) = 0$, and the solutions to it are called characteristic roots, or eigenvalues. Any vector $v \neq 0$ that satisfies $Av = \lambda v$ for some λ is a characteristic vector, or eigenvector.*

• **Exercise** Let a $(2,2)$ matrix

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

Show that $\xi(\lambda) \equiv \det(A - \lambda I) = \lambda^2 - (\alpha + \delta)\lambda + (\alpha\delta - \beta\gamma)$, and find the eigenvalues.

• **Exercise** By using the inductive definition of the determinant show, or at least persuade yourself, that if A is (n, n) then $\xi(\lambda) \equiv \det(A - \lambda I)$ is an n-th order polynomial.

By the fundamental theorem of algebra, any n-th order polynomial has exactly n roots. Of course, some of these roots may be imaginary rather than real, or they might be repeated. In any case, there are (complex or real) numbers $\{\lambda_1, \cdots, \lambda_n\}$ such that

$$\xi(\lambda) \equiv \det(A - \lambda I) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

These $\{\lambda_1, \cdots, \lambda_n\}$ are the eigenvalues of A.

Remark: We emphasize that, in what follows, we do not assume the eigenvalues or the eigenvectors to be real. A real matrix A may well have nonreal eigenvalues and nonreal eigenvectors. Nor do we assume that all roots are distinct. If a root appears once, then it is called a distinct root, while if it is repeated $r > 1$ times, then it is called a r-fold root. For example, the only eigenvalue of the identity matrix is 1, appearing with multiplicity n; i.e., it is an n-fold root. Moreover, in this example the eigenvectors are not unique, either. Indeed, all nonzero vectors $v \in \mathbb{R}^n$ are eigenvectors of the identity matrix associated to eigenvalue 1.

Returning to the eigenvectors of A, we observe that the singularity (since $\xi(\lambda_j) \equiv \det(A - \lambda_j I) = 0$) of matrix $(A - \lambda_j I)$, implies that there exists a vector $v_j \neq 0$ such that $(A - \lambda_j I)v_j = 0$. Rearranging we get $Av_j = \lambda_j v_j$.

Lemma 4.1 *To any eigenvalue of A, λ_j such that* $\det(A - \lambda_j I) = 0$, *there is associated at least one eigenvector $v_j \neq 0$ such that $Av_j = \lambda_j v_j$.*

We showed that each eigenvalue has at least one eigenvector associated with it. In fact, it has a whole continuum of associated eigenvectors. Indeed, take $v_j \neq 0$ such that $Av_j = \lambda_j v_j$ and let $w_j = \mu v_j$ for any scalar $\mu \neq 0$. Then $w_j \neq 0$ and

$$Av_j = \lambda_j v_j \Rightarrow A(\mu v_j) = \lambda_j(\mu v_j) \Rightarrow Aw_j = \lambda_j w_j$$

meaning that w_j is as well an eigenvector associated with eigenvalue λ_j. Given this intrinsic multiplicity, we define the following.

Theorem 4.1 *Suppose k $(k \leq n)$ eigenvalues $\{\lambda_1, \cdots, \lambda_k\}$ of A are distinct, and take any corresponding eigenvectors $\{v_1, \cdots, v_k\}$, defined by $v_j \neq 0$, $Av_j = \lambda_j v_j$ for $j = 1, \cdots, k$ Then, $\{v_1, \cdots, v_k\}$ are linearly independent.*

Proof. First consider two such eigenvectors. Suppose we have eigenvalue λ with eigenvector v, and eigenvalue λ with eigenvector w, $\lambda \neq \mu$. We will show that $\alpha v + \beta w = 0 \Rightarrow \alpha = \beta = 0$, implying that v and w are linearly independent.

So suppose we have

$$\begin{aligned} & \alpha v + \beta w = 0 && (7) \\ \Rightarrow \quad & \alpha A v + \beta A w = 0 && (8) \\ \Rightarrow \quad & \alpha \lambda v + \beta \mu w = 0 && (9) \end{aligned}$$

Now multiply (7) by $-\lambda$ and add to (9) to get

$$\beta(\mu - \lambda) w = 0 \tag{10}$$

which implies that $\beta = 0$, and plugging this back into (7) implies that $\alpha = 0$ as well.

Now consider any 3 eigenvectors (v, w, u) with distinct eigenvalues (λ, μ, ν); we proceed in much the same manner:

$$\begin{aligned} & \alpha v + \beta w + \gamma u = 0 && (11) \\ \Rightarrow \quad & \alpha A v + \beta A w + \gamma A u = 0 && (12) \\ \Rightarrow \quad & \alpha \lambda v + \beta \mu w + \gamma \nu u = 0 && (13) \\ \Rightarrow \quad & \beta(\mu - \lambda) w + \gamma(\nu - \lambda) u = 0 \quad \text{with } (-\lambda(11) + (13)) && (14) \end{aligned}$$

But this is a linear combination of two eigenvectors, and we have just shown that they must be linearly independent. So we have $\beta = \gamma = 0$, which implies that $\alpha = 0$ as well. We can continue in this manner to show that any k eigenvectors with distinct eigenvalues are linearly independent. ■

- **Example:** Consider the matrix:

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

Its characteristic polynomial is

$$\det(A - \lambda I) = \lambda^2 - 2\lambda - 3 = (\lambda + 1)(\lambda 3)$$

Hence, its eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 3$. To find the eigenvectors we have to solve $Av_1 = \lambda_1 v_1$ for v_1 and $Av_2 = \lambda_2 v_2$ for v_2. Let's find first an eigenvector $v_1 = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}$ corresponding to $\lambda_1 = -1$:

$$\begin{aligned} (A - \lambda_1 I) v_1 = 0 \quad & \Leftrightarrow \quad \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ & \Leftrightarrow \quad 3 v_{11} v_{21} = 0 \end{aligned}$$

Thus, any $v_1 = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}$ such that $3v_{11} = v_{21} \neq 0$ is an eigenvector for $\lambda_1 = -1$, and vice versa, any eigenvector corresponding to λ_1 is of the form $v_1 = \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}$ such that $3v_{11} = v_{21} \neq 0$. Now consider the eigenvector $v_2 = \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}$ corresponding to $\lambda_2 = 3$:

$$(A - \lambda_2 I)v_2 = 0 \quad \Leftrightarrow \quad v_{12} + v_{22} = 0$$

Hence, the eigenvectors corresponding to $\lambda_2 = 3$ are of the form $v_2 = \begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix}$ such that $v_{12} = -v_{22} \neq 0$.

For symmetric matrices we can say something stronger. Here we discuss the eigenvectors of distinct eigenvalues; a more general theorem follows below.

Theorem 4.2 *Suppose k ($k \leq n$) eigenvalues $\{\lambda_1, \cdots, \lambda_k\}$ of A are distinct with A symmetric, and take any corresponding eigenvectors $\{v_1, \cdots, v_k\}$, defined by $v_j \neq 0$, $Av_j = \lambda_j v_j$ for $j = 1, \cdots, k$. Then, $\{v_1, \cdots, v_k\}$ are orthogonal.*

Proof. Suppose v is an eigenvector for λ and w is an eigenvector for μ. Then

$$w'Av = \lambda w'v$$

and

$$w'Av = \mu w'v$$

(where the second equation is derived by taking the transpose of $Aw = \mu w$ and postmultiplying by v). Thus, $0 = (\lambda - \mu)w'v$ implying that $w'v = 0$, or that w and v are orthogonal. ■

Remark: Finally, we notice that the computation of eigenvalues for diagonal or triangular matrices is trivial: These are given simply by the diagonal elements.

Lemma 4.2 *Let A be an (n, n) diagonal or triangular matrix with diagonal elements $\{a_{jj}\}$. Then its eigenvalues are $\lambda_j = a_{jj}$, all $j = 1, \cdots, n$.*

• **Exercise** Here is how to work out the proof: Take an upper triangular matrix A, and form the matrix $C = \lambda I - A$. Notice that $C = \lambda I - A$ is upper triangular as well. Compute the determinant $\xi(\lambda) \equiv \det(\lambda I - A) = \det(C)$ inductively starting from the first row and going down. Show thereby that

$$\begin{aligned} \det(C) &= c_{11}\det(C_{11}) - c_{12}\det(C_{12}) + \cdots \pm c_{1n}\det(C_{1n}) \\ &= (\lambda - a_{11})\det(C_{11}) + 0 \\ &= \cdots \\ &= (\lambda - a_{11})\cdots(\lambda - a_{nn}) \end{aligned}$$

to conclude the proof.

Lemma 4.3 *If A is idempotent (defined by $AA = A$) then the eigenvalues of A are 0 or 1.*

Proof. $Ax = \lambda x \Rightarrow Ax = AAx = \lambda Ax = \lambda^2 x$, so $\lambda^2 = \lambda$ which implies $\lambda = 0$ or $\lambda = 1$.

5 Quadratic form

We define a quadratic form as a function $Q : \mathbb{R}^n \to \mathbb{R}$ such that

$$Q(x) = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

Without loss of generality, we may assume $a_{ij} = a_{ji}$ and thus A to be symmetric. Note that $Q(x)$ is a **scalar**.

- **Exercise** Is the assumption that A is symmetric really 'without loss of generality?' Take $Q(x) = x'Ax$ for arbitrary A and show that there exists symmetric B such that $Q(x) = x'Ax = x'Bx$ (but see next Theorem for a hint).

Definition 13 *A quadratic form Q is positive (negative) semidefinite iff $Q(x) \geq 0 (\leq 0)$ for all x. It is positive (negative) definite iff $Q(x) > 0 (< 0)$ for all $x \neq 0$.*

Note that positive (negative) definiteness implies positive (negative) semidefiniteness, but not the converse. If a quadratic form satisfies none of these conditions, we say it is **indefinite**.

In many economic applications (e.g., static or dynamic optimization, econometrics, etc.), it is important to determine whether a symmetric matrix is positive/negative definite/semidefinite. The diagonalization of the symmetric matrix A can indeed help us easily characterize the quadratic form $Q(x) = x'Ax$.

Theorem 5.1 *Let B be an (n, n)-matrix, let $Q(x) = x'Ax$ be the corresponding quadratic form with $A = (B + B')/2$ symmetric***, and let $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of A (possibly not all distinct). Then:*

*(i) A is **positive definite** if and only if all eigenvalues are positive,*

$$Q(x) > 0 \quad \forall x \neq 0 \Leftrightarrow \lambda_i > 0 \quad \forall i$$

*(ii) A is **negative definite** if and only if all eigenvalues are negative,*

$$Q(x) < 0 \quad \forall x \neq 0 \Leftrightarrow \lambda_i < 0 \quad \forall i$$

(iii) A is ***positive semidefinite*** *if and only if all eigenvalues are nonnegative,*

$$Q(x) \geq 0 \quad \forall x \Leftrightarrow \lambda_i \geq 0 \quad \forall i$$

(iv) A is ***negative semidefinite*** *if and only if all eigenvalues are nonnegative,*

$$Q(x) \leq 0 \quad \forall x \Leftrightarrow \lambda_i \leq 0 \quad \forall i$$

(v) A is indefinite if and only if there are eigenvalues with opposite signs.

***Note that this theorem must make use of the eigenvalues of A, not B! For one thing, we cannot guarantee that B is diagonalizable, but we know this is true of A. Moreover, consider the following:

Question: for a (n, n) matrix A (not necessarily symmetric) to be positive definite (in the sense that $x'Ax > 0$ for any nonzero $x \in \mathbb{R}^n$), is it necessary and/or sufficient that its real eigenvalues are all positive?

Answer: It is necessary. Indeed, assume that $\lambda < 0$ is an eigenvalue of A and v is an eigenvector for this eigenvalue. Then $v'Av = \lambda v'v = \lambda |v|^2 < 0$, so A is not positive definite.

On the other hand, it is not sufficient. Consider $A = \begin{pmatrix} 1 & -5 \\ 0 & 1 \end{pmatrix}$. Its only eigenvalue is 1, but for $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ we have $x'Ax = -3$.

However, positive definiteness is inherently a property of a quadratic form, not of a matrix (although it can be defined, as above, in terms of a matrix). Remember that there exists infinitely many matrices representing a particular quadratic form (that is, such matrices A that $Q(x) = x'Ax$), all with generally different eigenvalues, and exactly one of them is symmetric. What you want to do to establish positive definiteness (or lack thereof) of a quadratic form is to find this symmetric matrix representing it (if you have any matrix B then $(B + B')/2$ is what you are looking for) and test whether its eigenvalues are all positive either by finding them all or by applying the principal minor method or otherwise).

For example, the symmetric matrix representing the same quadratic form as $\begin{pmatrix} 1 & -5 \\ 0 & 1 \end{pmatrix}$ is $\begin{pmatrix} 1 & -2.5 \\ -2.5 & 1 \end{pmatrix}$; its determinant is negative, so clearly it does not have both eigenvalues positive and hence the quadratic form is not positive definite, as demonstrated explicitly above.

- **Exercise** Let X be an arbitrary real matrix. Show that $X'X$ is positive semidefinite.

Theorem 5.2 *Let X be an (m, n) matrix with $m \geq n$ and $rank(X) = n$. Then $X'X$ is positive definite.*

Proof. See practice exercise 7.