Chapter Two

Elements of Linear Algebra

Previously, in chapter one, we have considered single first order differential equations involving a single unknown function. In the next chapter we will begin to consider systems of such equations involving several unknown functions and for these problems, a modicum of linear algebra is required. We recall first definitions for several terms and the notations we will use for them. Then we will proceed to only those aspects of linear algebra that will be needed in order to consider systems of linear differential equations. Several theorems are presented along with their proofs but the proofs can be avoided on first reading.

1. Notation and terminology

Scalars

The term scalar refers to the usual real numbers, and these will be denoted by: $x, \alpha, \ldots$

Vectors

The term "vector" refers to an array of numbers. In particular, an n-vector is an array consisting of n rows and one column. For example,

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

This is called a "column vector" and we denote it by $\vec{x}$.

An array of one row and n columns is called a row vector, for example :

$$\vec{x}^\top = (x_1, \ldots, x_n).$$

When a column vector, $\vec{x}$, is written as a row vector, we use the notation $\vec{x}^\top$ for the row vector.

Matrices

An m by n matrix is an array consisting of m(rows) and n(columns). For example,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

denotes an m by n matrix. Often we will write this more compactly as $A = [a_{ij}]$. In particular, vectors are a special case of matrices, for example,the column vector, $\vec{x}$ is an n by 1 matrix and the corresponding row vector, $\vec{x}^\top$, is a 1 by n matrix.

The matrix obtained by writing the rows of A as columns is called the **transpose** of A and is denoted by $A^\top = [a_{ji}]$. This explains the notation $\vec{x}^\top$ when the column vector $\vec{x}$ is written as a row vector.

If A has the same number of rows as it has columns, it is called a **square matrix**. Square matrices with the special property that $A = A^\top$ are called **symmetric** matrices. We will see later that such matrices occur frequently in applications and have special properties.

Products

There are various products that can be defined between the quantities defined above.

1) ***scalar multiplication***- this refers to the product of a scalar, $\alpha,$ with a vector, $\vec{x},$ or matrix, A:

$$\alpha\vec{x} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}, \qquad \alpha A = \begin{bmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{bmatrix}$$

2) ***inner product***- this refers to the product of two vectors $\vec{x}$ and $\vec{z}$ of the same type (number of components) and the result is a scalar. We use the notation $\vec{x} \cdot \vec{z}$ or $\vec{x}^\top\vec{z}$ for the inner product (also called the "dot product" in some texts). It is defined as follows,

$$\vec{x} \cdot \vec{z} = \vec{x}^\top\vec{z} = x_1 z_1 + \cdots + x_n z_n$$

Note that

$$\vec{x} \cdot \vec{x} = x_1^2 + \cdots + x_n^2 = \|\vec{x}\|^2$$
$$= \text{square of the length of } \vec{x}$$

and $\quad \vec{x} \cdot \vec{z} = 0$ if and only if $\vec{x}$ *and* $\vec{z}$ are **orthogonal** (perpendicular to one another).

3) ***matrix times a vector***- This refers to the product of a matrix A times a vector $\vec{x}.$ This product is defined if and only if the number of columns of the matrix A equals the number of rows in the column vector $\vec{x}.$ The product $A\vec{x}$ is defined as follows

$$A\vec{x} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

It will be useful later to write this product in various ways. For example, the product $A\vec{x}$ can be expressed in terms of the rows of A,

$$A\vec{x} = \begin{pmatrix} \vec{R}_1 \cdot \vec{x} \\ \vdots \\ \vec{R}_m \cdot \vec{x} \end{pmatrix}$$

Here $\vec{R}_j$denotes the $j-th$ row of the $m \times n$ matrix $A,$ and $\vec{R}_j \cdot \vec{x}$ denotes the inner product of this row with the vector $\vec{x}.$ We can also express $A\vec{x}$ in terms of the columns of A

$$A\vec{x} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$
$$= \begin{pmatrix} a_{11}x_1 \\ \vdots \\ a_{m1}x_1 \end{pmatrix} + \cdots + \begin{pmatrix} a_{1n}x_n \\ \vdots \\ a_{mn}x_n \end{pmatrix}$$
$$= x_1\vec{C}_1 + x_2\vec{C}_2 + \cdots + x_n\vec{C}_n$$

where $\vec{C}_k$ denotes the k-th column of the matrix A and the product $A\vec{x}$ can be expressed as the sum $x_1\vec{C}_1 + x_2\vec{C}_2 + \cdots + x_n\vec{C}_n$, where x_k denotes the k-th entry of the vector $\vec{x}$.

4) ***matrix times a matrix***- This refers to the product of the m by n matrix A times the n by p matrix B. The product is defined if and only if the number of columns of A is equal to the number of rows of B. If this is the case then the (j,k) – entry in AB is equal to the inner product of $\vec{R}_j$, the j-th row of the m×n matrix A with $\vec{C}_k$, the k-th column of the n by p matrix B; i.e.,

$$AB = \begin{pmatrix} \vec{R}_1 \bullet \vec{C}_1 & \cdots & \vec{R}_1 \bullet \vec{C}_p \\ \vdots & & \vdots \\ \vec{R}_m \bullet \vec{C}_1 & & \vec{R}_m \bullet \vec{C}_p \end{pmatrix}$$

Clearly, the product AB is not defined unless the number of columns of A equals the number of rows of B.

We recall now the meaning of a few terms regarding matrices:

- **diagonal matrix** all terms d_{jk} *with* $j \neq k$ *are zero*

$$D = \begin{pmatrix} d_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_{nn} \end{pmatrix}$$

- **identity matrix** $I =$ diagonal matrix with $d_{jj} = 1$ *all* j. Note that $IA = AI = A$ for all square matrices A.
- **transpose** if $A = (a_{ij})$ then $A^\top = (a_{ji})$ and $(AB)^\top = B^\top A^\top$
- **symmetric matrix** A is symmetric if $A = A^\top$
- **skew symmetric matrix** A is skew-symmetric if $A = -A^\top$

In order to consider problems involving systems of linear ODE’s it will be necessary to collect a number of concepts from the topic of linear algebra. These concepts can be organized into two groups, those having to do with the so-called "first problem of linear algebra" and those relating to what we will call "the second problem of linear algebra".

2. The First Problem of Linear Algebra: $A\vec{x} = \vec{b}$

Let A denote an n by n matrix, and let $\vec{b}$ denote an arbitrary vector in R^n. The first fundamental problem of linear algebra is the problem of finding a vector $\vec{x}$ in R^n such that $A\vec{x} = \vec{b}$ where A and $\vec{b}$ are given. That is, to find an $\vec{x}$ which solves a system of n linear algebraic equations in n unknowns.

There are two things we want to know about this problem: first, does it have a solution, and second, is the solution unique? Consider for example, the following systems of two equations for two unknowns:

$$1) \quad \begin{aligned} x_1 + x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned}$$

$$2) \quad \begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + 2x_2 &= 2 \end{aligned}$$

$$3) \quad \begin{aligned} x_1 + x_2 &= 1 \\ x_1 - x_2 &= 1 \end{aligned}$$

It should be evident that no solution exists for example 1 since the two equations are contradictory. Likewise, it should be clear that solutions exist for example 2 but the solution is not unique since the two equations are redundant. Finally, in example 3 there is a unique solution. When the number of equations is small, the questions of existence and uniqueness are easily answered and the results understood. We want to be able to deal equally well with systems involving large numbers of equations and unknowns and this will require a certain amount of mathematical machinery which we will now develop. The proofs of the main results here are included only for the sake of completeness and can be omitted on first reading.

2.1 Subspaces

In order to answer the questions of existence and uniqueness, it will be helpful to define the notion of a **subspace** of R^n and further define two particular subspaces associated with the matrix A. A collection of vectors M in R^n is a subspace if M is closed under the operation of forming linear combinations. That means simply that if $\vec{v}$ and $\vec{u}$ are any two vectors in the collection M, then for all scalars, C_1 and C_2, the combination, $C_1\,\vec{v} + C_2\,\vec{u}$ must also belong to M. If this is the case, then M is closed under the operation of forming linear combinations and we say M is a subspace of R^n.

Examples of subspaces

1. The set of all vectors $\vec{v}$ in R^n whose length is 1 is an example of a set that is **not** a subspace. For example $\vec{u} = (0,1)$ and $\vec{v} = (1,0)$ are two vectors in R^2 whose length is 1 but $\vec{u} + \vec{v} = (1,1)$ does not have length 1.
2. The set M of all vectors $\vec{v}$ in R^3 whose first entry is 0 is a subspace. For example $\vec{u} = (0,1,1)$, and $\vec{v} = (0,1,-1)$ are in M and for any constants α and β, $\alpha\vec{u} + \beta\vec{v} = (0, \alpha+\beta, \alpha-\beta)$ is also in M.
3. Another example of a set that is a subspace is the set $M = span\{\vec{X}_1, \vec{X}_2, \dots, \vec{X}_p\}$. This is the set consisting of all possible linear combinations of the vectors $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_p$. Clearly M is closed under the operation of forming linear combinations.

Two additional examples of subspaces are the following:

4 N_A = the ***null space*** of A – This is the set of all vectors $\vec{x} \in R^n$ satisfying $A\vec{x} = \vec{0}$.

2. R_A = the ***range*** of A – This is the set of all vectors $\vec{b} \in R^m$ satisfying $A\vec{x} = \vec{b}$. for some $\vec{x} \in R^n$.

Note that the null space of A always contains the zero vector (at least) but there may be nonzero vectors in the null space as well. Similarly the range of A also contains the zero vector (since $A\vec{0} = \vec{0}$) but unless A is composed of all zeroes, it will contain vectors besides the zero vector.

Clearly the system $A\vec{x} = \vec{b}$ has a solution (i.e., a solution exists) if and only if $\vec{b}$ belongs to the range of A. This is just the definition of the range, it is the set of $\vec{b}'s$ for which the system $A\vec{x} = \vec{b}$ has a solution. The best possible situation for existence of solutions is when the range of A is as large as possible, namely if it is the whole space. In that case $A\vec{x} = \vec{b}$ has a solution for all choices of $\vec{b}$.

The system has at most one solution (i.e., any solution is unique) if and only if the null space of A contains only the zero vector. To see this, suppose there are two solutions for the system, e.g., $A\vec{x} = \vec{b}$ and $A\vec{z} = \vec{b}$. Then

$$A\vec{x} - A\vec{z} = \vec{b} - \vec{b} = \vec{0}.$$

That is,

$$A\vec{x} - A\vec{z} = A(\vec{x} - \vec{z}) = \vec{0}, \qquad or \qquad (\vec{x} - \vec{z}) \in N_A$$

But if the null space of A contains only the zero vector, then $\vec{x} - \vec{z} = \vec{0}$, which is to say, $\vec{x} = \vec{z}$, and the two solutions must be equal (so there is really only one solution). Then the best possible situation for uniqueness of solutions is when the null space of A is a small as possible; i.e., when N_A contains only the zero vector.

2.2 Linear independence and dimension

In order to determine whether N_A and R_A contain vectors besides the zero vector, it will be helpful to define the notion of **dimension** for these subspaces. In order to define dimension, we will first have to define the term **linearly independent**.

Definition *a collection of vectors* $\left\{\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_N\right\}$ *is said to be* ***linearly independent*** *if the following two statements are equivalent:*

1) $C_1\vec{X}_1 + C_2\vec{X}_2 + \ldots + C_N\vec{X}_N = \vec{0}$
2) $C_1 = C_2 = \cdots = C_N = 0$

A set of vectors that is not linearly independent is said to be **linearly dependent**.

A set of vectors is linearly independent if none of them can be written as a linear combination of the others. For example, the vectors

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \vec{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

are easily seen to be a linearly independent set in R^4. Note that any vector in R^4 can be written as a linear combination of the $\vec{e}_k'$s. For example, if $\vec{x}^\top = [a,b,c,d]$, then $\vec{x} = a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3 + d\vec{e}_4$. We say that the vectors $\vec{e}_k,\ k = 1,2,3,4$ are a **spanning set** for R^4. Similarly, if M denotes the subspace of R^4 whose first entry is a zero, then $\vec{e}_2,\ \vec{e}_3, \vec{e}_4$ form a spanning set for M. In general, a set of vectors is a spanning set for a subspace M if any vector in M can be expressed as some linear combination of the vectors in the spanning set.

Now we can define the **dimension** of a subspace M to be the maximum number of linearly independent vectors in the subspace. It is clear that the dimension of a subspace of vectors from R^n cannot exceed n. Note that if M has dimension k, then a spanning set for M must contain at least k vectors. A spanning set for M that contains exactly k vectors is called a **basis** for M. A basis is a linearly independent spanning set and, as such, contains the minimal number of elements needed to span M; It is a fact (not proved here) that every basis of a subspace M contains the same number of elements and this number is the dimension of M.

With regard to the question of existence of solutions for $A\vec{x} = \vec{b}$, (where A is an m by n matrix) the system is solvable for any choice of $\vec{b}$ if the range of A equals all of R^n and this happens if the dimension of R_A equals n. With regard to the question of uniqueness, the solution is unique, (if it exists) when N_A contains only the zero vector, i.e., when the dimension of N_A equals zero. That is,

- $A\vec{x} = \vec{b}$ has at least one solution for every $\vec{b} \in R^n$ if $\dim(R_A) = n$
- $A\vec{x} = \vec{b}$ has at most one solution for every $\vec{b} \in R^n$ if $\dim(N_A) = 0$

2.3 Rank of a matrix

In order to determine the dimensions of R_A and N_A, we have to now define two more subspaces associated with A.

the **row space** of A $\quad RS[A] = span\left[\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_m\right]$

and the **column space** of A $\quad CS[A] = span\left[\vec{C}_1, \vec{C}_2, \ldots, \vec{C}_n\right]$

where $\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_m =$ the m rows of A and $\vec{C}_1, \vec{C}_2, \ldots, \vec{C}_n =$ the n columns of A.

Recall that if $M = span\left[\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n\right]$ then M consists of all possible linear combinations of $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$. If the vectors $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ are linearly independent, then they form a basis for M and M has dimension equal to n. In the case of $RS[A]$, the rows of A might not be independent so the dimension of the row space is less than or equal to m. The dimension of the row space of A is called the "row rank of A". Similarly the dimension of the column space is less than or equal to n and we call the dimension of the column space of A the "column rank of A".

2.4 Existence and Uniqueness for $A\vec{x} = \vec{b}$

Note the following facts about the row space and column space of A.

Since

$$A\vec{x} = \begin{bmatrix} \vec{R}_1 \bullet \vec{x} \\ \vec{R}_2 \bullet \vec{x} \\ \vdots \\ \vec{R}_n \bullet \vec{x} \end{bmatrix}$$

it follows that:

$\vec{x} \in N_A$ if and only if $\vec{R}_1 \bullet \vec{x} = \vec{R}_2 \bullet \vec{x} = \cdots = \vec{R}_n \bullet \vec{x} = 0,$

i.e., if and only if $\vec{x}$ is orthogonal to all the rows of A

Likewise, since $A\vec{x} = x_1\vec{C}_1 + x_2\vec{C}_2 + \cdots + x_n\vec{C}_n$

it follows that

$A\vec{x} = \vec{b}$ if and only if $\vec{b} = x_1\vec{C}_1 + x_2\vec{C}_2 + \cdots + x_n\vec{C}_n$

i.e., if and only if $\vec{b} \in CS[A] = span\left[\vec{C}_1, \vec{C}_2, \ldots, \vec{C}_n\right]$

These two observations can be stated more concisely as:

1. $N_A =$ the orthogonal complement of $RS[A] =: (RS[A])^{\perp}$
2. $R_A = CS[A]$

Here, $M^{\perp}$ denotes the the **orthogonal complement** of subspace M and consists of all vectors which are orthogonal to every vector in M. It is always the case that $M^{\perp}$ is a subspace and the only vector that belongs to both M and $M^{\perp}$ is the zero vector.

The importance of these two observations increases when they are combined with the following two facts. The proofs of these results can be found in the appendix to this chapter.

Theorem ***1***- *Let A be any m by n matrix. Then the column rank of A (number of linearly independent columns) equals the row rank of A (number of linearly independent rows); i.e.,* $\dim CS(A) = \dim RS(A)$.

Since the row rank and column rank of A are equal, we will refer simply to the **rank** of A.

Theorem ***2***- *If M is any subspace in* R^n *then* $\dim M + \dim M^{\perp} = n$.

Corollary *Let A be any* $m \times n$ *matrix. Then* $\dim R_A + \dim N_A = n$

The corollary to theorem 2 is an existence and uniqueness theorem for the system $A\vec{x} = \vec{b}$. It asserts that $\dim N_A = n - r$, where $r = rank\, A$. Then if we can determine the rank of A, the questions of existence and uniqueness can be quickly answered. In particular, in the case of a square matrix A, we have $m = n$ so if $r = n$ the system has a unique solution for every choice of $\vec{b}$. On the other hand, if $r < n$, then the solution is not unique and no solution exists for some choices of $\vec{b}$. It remains to see how to determine the rank of A.

2.5 Echelon form and the rank of A

We will now show how to compute the rank of a matrix A. Beginning with the matrix A, we perform a series of **row operations** on A to reduce it to **echelon form**. The **row operations** consist of:

1. multiply row i by a nonzero scalar α: αR_i
2. multiply row i by a nonzero scalar, α, and add it to row j : $R_j + \alpha R_i$
3. interchange rows i and j : R_{ij}

The notations, αR_i, $R_j + \alpha R_i$ and R_{ij} are the symbols we will use for the row operations used to reduce a matrix to echelon form. A matrix is said to be in echelon form if

- all trivial rows (rows composed of all zeroes) lie below all nontrivial rows (rows that have at least one nonzero entry)
- the first nonzero entry in any row lies to the right of the first nonzero entry in the row above

These conditions imply that all entries below the matrix diagonal are zeroes. Then the row operations are generally selected to produce zeroes below the diagonal as efficiently as possible. e.g., here we begin with a matrix A and employ 5 row operations that reduce A to an echelon form matrix B:

$$A = \begin{bmatrix} 2 & -2 & -4 \\ -3 & -2 & 6 \\ 2 & -1 & 2 \end{bmatrix} \Rightarrow \tfrac{1}{2}R_1 \Rightarrow \begin{bmatrix} 1 & -1 & -2 \\ -3 & -2 & 6 \\ 2 & -1 & 2 \end{bmatrix}$$

$$R_2 + 3R_1 \Rightarrow \begin{bmatrix} 1 & -1 & -2 \\ 0 & -5 & 0 \\ 2 & -1 & 2 \end{bmatrix} \Rightarrow R_3 - 2R_1 \Rightarrow \begin{bmatrix} 1 & -1 & -2 \\ 0 & -5 & 0 \\ 0 & 1 & 6 \end{bmatrix}$$

$$\tfrac{-1}{5}R_2 \Rightarrow \begin{bmatrix} 1 & -1 & -2 \\ 0 & 1 & 0 \\ 0 & 1 & 6 \end{bmatrix} \Rightarrow R_3 - R_2 \Rightarrow \begin{bmatrix} 1 & -1 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix} = B \text{ (echelon form)}$$

Two matrices are said to be ***row equivalent*** if one is obtained from the other by a sequence of row operations. For example, the matrices A and B above are row equivalent. The matrix B above is in ***echelon form***; i.e., in each row, the first nonzero entry lies to the right of the first nonzero entry in the row above and in each column, all entries below the diagonal entry are zeroes Now we can define the ***rank*** of A to be the number of nontrivial rows in a row equivalent matrix that is in echelon form. Thus the rank of the matrix A in the example above is 3 since the row equivalent matrix B has 3 nontrivial rows. In the following example,

$$A = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 3 & -1 & 2 \end{bmatrix}, , \quad \Rightarrow \text{row operations} \;\Rightarrow: B = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \text{echelon form}$$

we see that A has rank equal to 2 since A is row equivalent to B whose rank is clearly equal to two. Evidently, the rank of an n by n matrix can equal any number between zero (all entries of A would have to be zeroes) and n (the rows of A are a linearly independent set of vectors).

Theorems 1 and 2 of the previous section assert that, .

- $\dim[RS[A]] + \dim[N_A] = n$
- $\dim[RS[A]] = \dim[CS[A]] = \dim[R_A]$

These now can be expressed in the form:

- $\dim[N_A] = n - r$
- $r = rank\, A = \dim[R_A]$

It is evident that if the n by n matrix A has rank equal to $n,$ then $\dim[R_A] = n$ and $R_A = R^n.$ In this case $A\vec{x} = \vec{b}$ is solvable for any $\vec{b} \in R^n.$ It is also evident that if $rank[A] = n,$ then $\dim[N_A] = 0,$ which is to say N_A contains only the zero vector and the solution to $A\vec{x} = \vec{b}$ is unique.

An alternative means for determining whether the matrix A has rank equal to n is to compute the determinant of $A.$ Computing the determinant of an n by n matrix A when n is large is impractical so this approach is more theoretical than practical. At any rate, it can be shown that the determinant of A is different from zero if and only if the rank of A equals $n.$ Then, at least in cases of small $n,$ the determinant computation can be used to decide the rank of $A.$

These observations lead to the fundamental result regarding the first problem of linear algebra:

Theorem *3- The following are equivalent statements about the n by n matrix, A :*

1. $A\vec{x} = \vec{b}$ is uniquely solvable for every $\vec{b} \in R^n$
2. $rank[A] = n$
3. $\det(A) \neq 0$
4. there exists a unique n by n matrix, A^{-1} such that $AA^{-1} = A^{-1}A = I$

Exercises

For each of the following matrices, reduce to row echelon form, determine the rank and tell whether the problem $A\vec{x} = \vec{b}$ is uniquely solvable for every $\vec{b} \in R^n$:

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

3. $A = \begin{bmatrix} 3 & -1 & -1 \\ -3 & 2 & 3 \\ -1 & -2 & 1 \end{bmatrix}$ 4. $A = \begin{bmatrix} -2 & 1 & 2 \\ -2 & 2 & 2 \\ 3 & 0 & -3 \end{bmatrix}$

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

Theorem 3 asserts that when the rank of the n by n matrix A is equal to $n,$ the system $A\vec{x} = \vec{b}$ has a unique solution for every $\vec{b} \in R^n.$ The following theorem covers the case when the rank of A is less than $n.$ Here we make use of the n by $n+1$ augmented matrix $\left[A|\vec{b}\right]$, obtained by joining the column vector $\vec{b}$ to the matrix $A.$

Theorem 4- Suppose the n by n matrix, A has rank $r < n.$

1. If $rank[A] < rank\left[A|\vec{b}\right]$, then $A\vec{x} = \vec{b}$ has no solution.

2. If $rank[A] = rank\left[A|\vec{b}\right] = r < n,$ then $A\vec{x} = \vec{b}$ has an infinite family of solutions. The general solution in this case can be written as

$$\vec{x} = \vec{x}_p + C_1\vec{x}_1 + \cdots + C_q\vec{x}_q$$

where $q = n - r,$ $\vec{x}_p$ is any solution of $A\vec{x} = \vec{b},$ $C_1,\ldots,C_q$ are arbitrary constants and $\vec{x}_1,\ldots,\vec{x}_q$ are a basis for $N_A.$ The general solution is a q-parameter family of solutions where $C_1,\ldots,C_q$ are the parameters and $\vec{x}$ is said to be in **parametric form**.

We will illustrate these two theorems with some examples:

Examples

Solve the following systems to find the unique solution if one exists. If the solution is not unique, find the general solution in parametric form. If no solution exists indicate how you know this.

1.

$$\begin{aligned} x_2 - 4x_3 &= 6 \\ x_1 + 2x_2 - x_3 &= 2 \\ 2x_1 + x_2 - 2x_3 &= 4 \end{aligned}$$

Form the augmented matrix, $\left[A|\vec{b}\right]$:

$$\begin{bmatrix} 0 & 1 & -4 & 6 \\ 1 & 2 & -1 & 2 \\ 2 & 1 & -2 & 4 \end{bmatrix}$$

Use row operations to reduce the augmented matrix to the following "reduced" echelon form:

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

Clearly $rank[A] = rank\left[A|\vec{b}\right] = 3,$ so a unique solution exists and with the augmented matrix in this reduced echelon form, we see that the solution is $x_1 = 1/2,$ $x_2 = 0,$ and $x_3 = -3/2.$

2.

$$\begin{aligned} x_1 + x_2 - 2x_3 &= 2 \\ 3x_1 - x_2 - x_3 &= 3 \\ 5x_1 + x_2 - 5x_3 &= 7 \end{aligned}$$

Form the augmented matrix, $\left[A|\vec{b}\right]$:

$$\begin{bmatrix} 1 & 1 & -2 & 2 \\ 3 & -1 & -1 & 3 \\ 5 & 1 & -5 & 7 \end{bmatrix}$$

Use row operations to reduce the augmented matrix to the following "reduced" echelon form:

$$\begin{bmatrix} 1 & 0 & -\frac{3}{4} & \frac{5}{4} \\ 0 & 1 & -\frac{5}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Here $rank[A] = rank\left[A|\vec{b}\right] = 2 < 3,$ so solutions exist but are not unique. To express the solution in parametric form, we write

$$\vec{x} = \begin{bmatrix} \frac{5}{4} + \frac{3}{4}x_3 \\ \frac{3}{4} + \frac{5}{4}x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} \\ \frac{3}{4} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{3}{4}x_3 \\ \frac{5}{4}x_3 \\ x_3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{5}{4} \\ \frac{3}{4} \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{3}{4} \\ \frac{5}{4} \\ 1 \end{bmatrix} = \vec{x}_p + C\,\vec{x}_H$$

Here $\vec{x}_p = \left[\frac{5}{4}, \frac{3}{4}, 0\right]^\top$ is a particular solution of $A\vec{x} = \vec{b}$, and $\vec{x}_H = \left[\frac{3}{4}, \frac{5}{4}, 1\right]^\top$ is a basis for the (one dimensional) null space of A. Note that since any multiple of a vector in N_A is also in N_A, we could also use $[3,5,4]$ as a basis for N_A.

3.

$$\begin{aligned} 2x_1 - 2x_2 + x_3 &= -1 \\ x_1 - x_2 - x_3 &= 1 \\ 3x_3 &= -1 \end{aligned}$$

Form the augmented matrix, $\left[A|\vec{b}\right]$:

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

and use row operations to reduce it to row echelon form:

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

Here the rank of A equals two, while the rank of $\left[A|\vec{b}\right]$ is three. Then it follows from theorem 4 that there is no solution to this system since the equations are inconsistent.

4. Find a basis for N_A

$$i)\ A = \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \qquad ii)\ A = \begin{bmatrix} 1 & 1 & 0 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Since the rank of this 3 by 3 matrix is clearly equal to 2, it follows from theorem 2 that $\dim N_A = 3 - 2 = 1$. Then we are looking for the general solution of $A\vec{x} = \vec{0}$ and this means $\vec{x} = [x_1, x_2, x_3]^\top$ where

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

Then $\vec{x} = [5x_3, -2x_3, x_3]^\top = x_3[5, -2, 1]^\top$ and it is evident that all $\vec{x} \in N_A$ are multiples of $[5, -2, 1]^\top$, and this vector forms a basis for N_A.

In the case of the 2 by 4 matrix, the rank is apparently 2 so theorem 2 implies $\dim N_A = 4 - 2 = 2.$ If $A\vec{x} = \vec{0},$ then

$$\begin{aligned} x_1 + x_2 + 4x_4 &= 0 \\ x_3 + 2x_4 &= 0. \end{aligned}$$

Now we can solve these equations for two of the $x_i's$ in terms of the other two and it doesn't matter which two we solve for. If we solve for x_1 and x_3 in terms of x_2 and $x_4,$ then

$$\begin{aligned} \vec{x} &= [-x_2 - 4x_4, x_2, -2x_4, x_4]^\top \\ &= x_2[-1, 1, 0, 0]^\top + x_4[-4, 0, -2, 1]^\top. \end{aligned}$$

We conclude that $[-1, 1, 0, 0]^\top$ and $[-4, 0, -2, 1]^\top$ are a basis for $N_A.$ Alternatively, if we decided to solve for x_1 and x_4 in terms of x_2 and $x_3,$ then we would find,

$$x_4 = -\frac{1}{2}x_3 \text{ and } x_1 = -x_2 + 2x_3,$$

$$\text{so} \quad \vec{x} = \left[-x_2 + 2x_3, x_2, x_3, -\frac{1}{2}x_3\right]^\top.$$

Then the general solution of $A\vec{x} = \vec{0}$ can be written $\vec{x} = x_2[-1, 1, 0, 0]^\top + x_3\left[2, 0, 1, -\frac{1}{2}\right]^\top$. In this case the basis for N_A consists of $[-1, 1, 0, 0]^\top$ and $\left[2, 0, 1, -\frac{1}{2}\right]^\top$, which is seen to be essentially the same as the previous basis.

It is worth observing that the general solution for the linear system, $A\vec{x} = \vec{b},$ has the same structure as the general solution to a linear ODE $L[y(t)] = f(t).$ Recall that the general solution for the linear ODE is equal to $y_G(t) = y_p(t) + y_H(t),$ where y_p is a particular solution of the inhomgeneous ODE and y_H denotes the general homogeneous solution. If the m by n matrix A has a null space of dimension $q,$ then the general homogeneous solution for A is of the form $\vec{x}_H = C_1\vec{x}_1 + \cdots + C_q\vec{x}_q$ and the general solution of $A\vec{x} = \vec{b}$ is $\vec{x} = \vec{x}_p + \vec{x}_H,$ for any particular solution, $\vec{x}_p,$ of $A\vec{x} = \vec{b}.$

Exercises

In each of the following, solve $Ax = b.$ Express the solution in the form $x = x_p + x_H$ where x_p is a particular solution and $Ax_H = 0.$ Express x_H in parametric form.

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

2. $A = \begin{bmatrix} 2 & 0 & 0 & -4 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

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

4. $A = \begin{bmatrix} 5 & -2 & -5 \\ -3 & 0 & 3 \\ 0 & -3 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}$

5. $A = \begin{bmatrix} -3 & 3 & 3 & 0 \\ -4 & 2 & 4 & -2 \\ 1 & 0 & -1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix}$

6. $A = \begin{bmatrix} 2 & 0 & 0 & -4 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

7. $A = \begin{bmatrix} 1 & 0 & 2 & -1 \\ 2 & 2 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

8. $A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 2 & -2 & 1 & 0 & 2 \\ 3 & 0 & 1 & 1 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix}$

Find a basis for the null space of A :

9. $A = \begin{bmatrix} 1 & 1 & 1 \\ -5 & -2 & -5 \\ 1 & 0 & 1 \end{bmatrix}$ 10. $A = \begin{bmatrix} 2 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \\ 1 & 1 & 3 & 2 \\ -3 & 3 & 3 & 0 \end{bmatrix}$

11. Does span $\left\{ \begin{bmatrix} 1 \\ -4 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \right\}$ contain $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ or $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$?

12. Does span $\left\{ \begin{bmatrix} 1 \\ -4 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right\}$ contain $\begin{bmatrix} -3 \\ 2 \\ 7 \end{bmatrix}$?

3. The Second Problem of Linear Algebra: $A\vec{x} = \lambda\vec{x}$

Let A denote an n by n matrix. It is clear that for any choice of the scalar, λ, the equation $A\vec{x} = \lambda\vec{x}$, has the solution $\vec{x} = \vec{0}$. However, if λ is such that $A\vec{x} = \lambda\vec{x}$, has nontrivial solutions, $\vec{x} \neq \vec{0}$, then we say that λ is an **eigenvalue for A**, and the corresponding nontrivial solutions are said to be **eigenvectors for A** corresponding to the eigenvalue λ. Note that if $\vec{x} \neq \vec{0}$, *and* $A\vec{x} = \lambda\vec{x}$, then for every nonzero scalar α, $A(\alpha\vec{x}) = \lambda(\alpha\vec{x})$, so any nonzero scalar multiple of an eigenvector for A corresponding to the eigenvalue λ is again an eigenvector for A corresponding to the eigenvalue λ. The problem of finding the eigenvalues and eigenvectors for an n by n matrix A is the second problem of linear algebra.

3.1 Finding Eigenvalues and Eigenvectors of a Matrix

If $\vec{x}$ is an eigenvector for A corresponding to the eigenvalue λ, then $\vec{x} \neq \vec{0}$ and $A\vec{x} - \lambda\vec{x} = (A - \lambda I)\vec{x} = \vec{0}$; i.e, $\vec{x} \in N_{A-\lambda I} =: N_\lambda$ *and* $\vec{x} \neq \vec{0}$. Then $\dim N_\lambda > 0$ and the results of the previous sections imply that the rank of $A - \lambda I$ is less than n. In this case, it is known from theorem 2.2 that the determinant of $A - \lambda I$ must be zero. It is also well known that since A is n by n, $\det(A - \lambda I)$ is a polynomial in λ of degree n. Then we have the following equivalent ways of characterizing eigenvalues and eigenvectors,

Theorem 1. λ is an eigenvalue for A if:

1. $A\vec{x} = \lambda\vec{x}$, for some $\vec{x} \neq \vec{0}$,
2. $\det(A - \lambda I) = P_n(\lambda) = 0$
3. $rank(A - \lambda I) < n$
4. $\dim N_\lambda > 0$

Then $\vec{x}$ is an eigenvector for A corresponding to the eigenvalue λ if:

1. $\vec{x} \neq \vec{0}$, and $A\vec{x} = \lambda\vec{x}$,
2. $\vec{x} \in N_{A-\lambda I} = N_\lambda$

Examples

1. Consider the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, Then

$$det(A - \lambda I) = \det\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}$$
$$= 3 - 4\lambda + \lambda^2 = P_2(\lambda)$$

and the eigenvalues of A are the roots of the equation, $P_2(\lambda) = 0$; *i.e.*, $\lambda_1 = 1$, $\lambda_2 = 3$.

Next, we find the eigenvectors of A. The eigenvector corresponding to $\lambda = 1$ lies in the null space of

$$(A - \lambda I)|_{\lambda=1} = (A - I) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Since

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

if and only if $x_1 + x_2 = 0$; i.e., $\vec{x} \in N_{A-I}$ if and only if $x_2 = -x_1$. Then $\vec{x} = [x_1, -x_1]^\top = x_1[1,-1]$ which is to say, $\vec{x} \in N_{A-I}$ if and only if $\vec{x}$ is a multiple of $\vec{E}_1 = [1,-1]^\top$. Then all eigenvectors of A associated with λ_1 are multiples of $\vec{E}_1$.

The eigenvector corresponding to $\lambda = 3$ lies in the null space of

$$(A - \lambda I)|_{\lambda=3} = (A - 3I) = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

For the null space of $(A - 3I)$ we find in the same way that, $\vec{x} \in N_{A-3I}$, if and only if $\vec{x}$ is a multiple of $\vec{E}_2 = [1,1]^\top$.

Note that both $\lambda_1 = 1$, $\lambda_2 = 3$ are real numbers and $\vec{E}_1 \cdot \vec{E}_2 = 0$. This is not accidental. We will show that for any real symmetric matrix, the eigenvalues are real and the eigenvectors associated with distinct eigenvalues are orthogonal.

First, however, we consider another example. For the matrix $A = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$, the eigenvalues are found to be : $\lambda_1 = 2i$, $\lambda_2 = -2i$ and the eigenvectors are :

$$\vec{E}_1 = \begin{bmatrix} -i \\ 1 \end{bmatrix} \text{ associated with } \lambda = 2i,$$

$$\vec{E}_2 = \begin{bmatrix} i \\ 1 \end{bmatrix} \text{ associated with } \lambda = -2i$$

In this case note that $\vec{E}_1 \cdot \vec{E}_1 = \vec{E}_2 \cdot \vec{E}_2 = 0$ *but* $\vec{E}_1 \cdot \vec{E}_2 \neq 0$. This is bad. Neither $\vec{E}_1$ nor $\vec{E}_2$ is the zero vector but each appears to have zero length. The problem here lies in the definition of the inner product. We have to modify the previous definition to account for vectors having complex entries. The new definition for the inner product is,

$$\vec{x} \cdot \vec{z} = x_1\bar{z}_1 + \cdots + x_n\bar{z}_n$$

where $\bar{z}_k$ = the complex conjugate of z_k.

Recall that the complex conjugate of the complex number $z_k = \alpha_k + i\beta_k$ is defined as $\bar{z}_k = \alpha_k - i\beta_k$. With this definition, $\vec{E}_1 \cdot \vec{E}_1 \neq 0$, $\vec{E}_2 \cdot \vec{E}_2 \neq 0$ *and* $\vec{E}_1 \cdot \vec{E}_2 = 0$. In addition, when the vectors involved both have only real components, then the new definition reduces to the old one. For the new definition, we have the following properties,

$$\vec{x} \cdot \vec{z} = \overline{\vec{z} \cdot \vec{x}}$$

and $$(\alpha\vec{x}) \cdot \vec{z} = \alpha(\vec{x} \cdot \vec{z})$$

hence, for the complex inner product, scalar multiplication has the following property,

$$\vec{x} \cdot (\alpha\vec{z}) = \bar{\alpha}\,(\vec{x} \cdot \vec{z}).$$

Now we can show:

Theorem ***2- (Spectral Theorem for Symmetric Matrices)*** *Suppose A is an n by n symmetric* *real matrix. Then :*

1. the eigenvalues of A are all real.
2. eigenvectors associated with distinct eigenvalues of A are orthogonal
3. A has n mutually orthogonal eigenvectors $\vec{E}_1, \ldots \vec{E}_n$ which form an orthogonal basis for R^n.

Proof of 1) and 2)- For any matrix A (symmetric or not) and all vectors $\vec{x}, \vec{z}$

$$(A\vec{x}) \cdot \vec{z} = \vec{x} \cdot (A^\top \vec{z}).$$

If A is symmetric (i.e., $A = A^\top$) this becomes,

$$(A\vec{x}) \cdot \vec{z} = \vec{x} \cdot (A\vec{z}).$$

If $A\vec{x} = \lambda\vec{x}$ then $(A\vec{x}) \cdot \vec{x} = \vec{x} \cdot (A\,\vec{x})$ becomes $(\lambda\vec{x}) \cdot \vec{x} = \vec{x} \cdot (\lambda\,\vec{x})$. Then by the scalar multiplication property of the complex inner product,

$$\lambda(\vec{x} \cdot \vec{x}) = \bar{\lambda}(\vec{x} \cdot \vec{x}).$$

Finally, since $\vec{x}$ is an eigenvector, $\vec{x} \cdot \vec{x} \neq 0$ so $\lambda = \bar{\lambda}$, which is to say, any eigenvalue of A is real.

To show the second assertion of this theorem, suppose

$$A\vec{x} = \lambda\vec{x} \qquad \text{and} \qquad A\vec{z} = \mu\vec{z},\ \lambda \neq \mu.$$

Then $(A\vec{x}) \cdot \vec{z} = \vec{x} \cdot (A\,\vec{z})$ or $(\lambda\vec{x}) \cdot \vec{z} = \vec{x} \cdot (\mu\,\vec{z})$. Since λ, μ are eigenvalues of A, they are both real so that

$$\lambda(\vec{x} \cdot \vec{z}) = \mu(\vec{x} \cdot \vec{z})$$

$$\text{or } (\lambda - \mu)(\vec{x} \cdot \vec{z}) = 0$$

Finally, since $\lambda \neq \mu$, it follows that $(\vec{x} \cdot \vec{z}) = 0.$ ■

Corollary *If A is skew symmetric ($A = -A^\top$), then the eigenvalues occur in complex conjugate pairs and assertions 2 and 3 of the theorem hold.*

The meaning of 3 is that if $\vec{x}$ is an arbitrary vector in R^n then we can find unique constants $\alpha_1, \ldots, \alpha_n$ such that

$$\vec{x} = \alpha_1\vec{E}_1 + \ldots + \alpha_n\vec{E}_n.$$

In fact, the $\alpha' s$ can easily be found by using the orthogonality of the eigenvectors:

for each k, $1 \leq k \leq n$,

$$\begin{aligned} \vec{x} \cdot \vec{E}_k &= \left(\alpha_1\vec{E}_1 + \ldots + \alpha_n\vec{E}_n\right) \cdot \vec{E}_k \\ &= \alpha_1\vec{E}_1 \cdot \vec{E}_k + \ldots + \alpha_n\vec{E}_n \cdot \vec{E}_k \\ &= \alpha_k\vec{E}_k \cdot \vec{E}_k \end{aligned}$$

Here, we used the fact that $\vec{E}_j \cdot \vec{E}_k = 0$ if $j \neq k$.

Since the $\vec{E}_k' s$ are eigenvectors, $\vec{E}_k \cdot \vec{E}_k \neq 0$ so that

$$\alpha_k = \frac{\vec{x} \cdot \vec{E}_k}{\vec{E}_k \cdot \vec{E}_k} \quad \text{for each } k,\ 1 \leq k \leq n.$$

The fact that the eigenvectors of A form a basis for R^n is very useful in solving systems of linear constant coefficient differential equations. The following result is also useful,

> **Theorem** ***3*** *Suppose A is an n by n real matrix having n distinct, real eigenvalues. Then the corresponding n eigenvectors are linearly independent (in general not orthogonal).*

Since the eigenvectors in this case are linearly independent, they form a basis for $R^n,$ but since they are not, in general, orthogonal, finding scalars $\alpha_1, \ldots, \alpha_n$ such that

$$\vec{x} = \alpha_1 \vec{E}_1 + \ldots + \alpha_n \vec{E}_n.$$

requires solving the system of equations, $[E]\vec{\alpha} = \vec{x},$ where $[E]$ is the matrix whose columns are the eigenvectors of A, and $\vec{\alpha}$ denotes the vector whose entries are the coefficients $\alpha_1, \ldots, \alpha_n.$

Examples

Find the eigenvalues and eigenvectors for the following matrices:

1. $A = \begin{bmatrix} 1 & 3 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix},$

Here

$$\begin{aligned} \det(A - \lambda I) &= (1-\lambda)^2(4-\lambda) - 9(4-\lambda) \\ &= (4-\lambda)[1 - 2\lambda + \lambda^2 - 9] \\ &= (4-\lambda)(\lambda^2 - 2\lambda - 8) \\ &= (4-\lambda)^2(\lambda+2) \end{aligned}$$

and it follows that the eigenvalues are $\lambda = 4, 4, -2.$ We say in this case that $\lambda = 4$ has **algebraic multiplicity** 2 and $\lambda = -2$ has algebraic multiplicity 1.

To find the eigenvectors associated with $\lambda = 4,$ we must find a basis for the null space of $A - 4I.$ Now,

$$A - 4I = \begin{bmatrix} -3 & 3 & 0 \\ 3 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and we see that $\vec{x} = [x_1, x_2, x_3]^\top \in N_4$ if $-3x_1 + 3x_2 = 0;$ i.e., if $\vec{x} = [x_1, x_1, x_3]^\top = x_1[1,1,0]^\top + x_3[0,0,1]^\top.$ Then

$$\vec{E}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \vec{E}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

are a basis for $N_4.$ Since the dimension of N_4 is $2,$ we say $\lambda = 4$ has **geometric multiplicity** of $2.$

The eigenvector associated with $\lambda = -2$ (recall that the eigenvectors of A form an orthogonal basis for R^3 and there are already two vectors associated with $\lambda = 4$ so there can only be one more eigenvector) is in the null space of

$$A + 2I = \begin{bmatrix} 3 & 3 & 0 \\ 3 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Then $\vec{x} = [x_1, x_2, x_3]^\top \in N_{-2}$ if $3x_1 + 3x_2 = 0$ and $x_3 = 0$; i.e., if $\vec{x} = [x_1, -x_1, 0]^\top = x_1[1, -1, 0]^\top$. Evidently, $\lambda = -2$ has geometric multiplicity equal to one and the associated eigenvector is

$$\vec{E}_3 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Clearly the eigenvalues of the symmetric matrix A are real and the eigenvectors $\{\vec{E}_1, \vec{E}_2, \vec{E}_3\}$ form an orthogonal basis for R^3 as predicted by theorem 2.

2. $A = \begin{bmatrix} 1 & 3 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$,

In this example,

$$\begin{aligned} \det(A - \lambda I) &= (1-\lambda)^2(4-\lambda) + 9(4-\lambda) \\ &= (4-\lambda)[1 - 2\lambda + \lambda^2 + 9] \\ &= (4-\lambda)(\lambda^2 - 2\lambda + 10) \\ &= (4-\lambda)(\lambda - 1 + 3i)(\lambda - 1 - 3i) \end{aligned}$$

and we see that this skew symmetric matrix has one real eigenvalue and a conjugate pair of complex eigenvalues as asserted in the corollary to theorem 2. A matrix with real entries can only have complex eigenvalues that occur in conjugate pairs. Thus if $\alpha + i\beta$ is an eigenvalue of the matrix A, then $\alpha - i\beta$ must also be an eigenvalue for A.

To find the eigenvectors associated with $\lambda = 4$, we proceed as in the previous example to find a basis for the null space of $A - 4I$. Now,

$$A - 4I = \begin{bmatrix} -3 & 3 & 0 \\ -3 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and we see that $\vec{x} = [x_1, x_2, x_3]^\top \in N_4$ if $x_1 = 0$, and $x_2 = 0$, with no restriction placed on x_3; i.e.,$\vec{x} \in N_4$ if $\vec{x} = [0, 0, x_3]^\top = x_3[0, 0, 1]^\top$. Then

$$\vec{E}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

is a basis for N_4.

To find the other eigenvectors, consider first $\lambda = 1 - 3i$,

$$A-(1-3i)I=\begin{bmatrix} 3i & 3 & 0 \\ -3 & 3i & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Then $\vec{x}=[x_1,x_2,x_3]^\top \in N_{1-3i}$ if $ix_1+x_2=0$ and $x_3=0$; i.e., $\vec{x}=[x_1,-ix_1,0]^\top = x_1[1,-i,0]^\top$. Notice that we used only the first equation in the system $A-(1-3i)I=0$ in order to solve for x_1 and x_2. This is because the first and second equations in the system must be dependent so the second equation contains the same information as the first. Thus only the first equation need be considered. Furthermore, since the eigenvector $\vec{E}_2$ associated with $\lambda=1-3i$ is $[1,-i,0]^\top$, the eigenvector associated with the conjugate eigenvalue, $\lambda=1+3i$, must necessarily be the conjugate of $\vec{E}_2$. That is, $\vec{E}_3=[1,i,0]^\top$ is the eigenvector associated with $\lambda=1+3i$. Note that the eigenvectors $\left\{\vec{E}_1,\vec{E}_2,\vec{E}_3\right\}$ of the skew symmetric matrix A form an orthogonal basis for R^3 as predicted by the corollary to theorem 2. Each of the eigenvalues in this example has algebraic and geometric multiplicity equal to one.

3. $A=\begin{bmatrix} 1 & 3 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

Since A has only zeroes below the diagonal, it is easy to see that $\det(A-\lambda I)=(\lambda-1)(\lambda+1)(\lambda-4)$ so the eigenvalues are just the diagonal entries in A, $\lambda=1,-1,4$. Then according to theorem 3, the eigenvectors will be linearly independent but not necessarily orthogonal.

We have, in this example

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

from which we see that in the eigenvector associated with $\lambda=1$, $\vec{E}_1=[x_1,x_2,x_3]^\top \in N_1$ if $x_3=x_2=0$. Then $\vec{E}_1=[x_1,0,0]^\top = x_1[1,,0,0]^\top$. Similarly,

$$A+I=\begin{bmatrix} 2 & 3 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

which implies that $\vec{E}_2=[x_1,x_2,x_3]^\top \in N_{-1}$ if $x_3=0$ and $2x_1+3x_2=0$. Then $\vec{E}_2=x_2\left[-\frac{3}{2},1,0\right]^\top$. Finally,

$$A-4I=\begin{bmatrix} -3 & 3 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which tells us that $\vec{E}_3=[x_1,x_2,x_3]^\top \in N_4$ if $x_1=x_2=0$ and $\vec{E}_3=x_3[0,0,1]^\top$. Since the eigenvectors we have found are determined only up to the multiplicative constants $x_1,x_2,$ and x_3 we are free to choose any value we like for the constants. In particular, if we choose $x_1 = x_3 = 1$ and $x_2 = 2$ then we have $\vec{E}_1 = [1,0,0]^\top$, $\vec{E}_2 = [-3,2,0]^\top$ and $\vec{E}_3 = [0,0,1]^\top$. These eigenvectors are independent but not mutually orthogonal.

4. $A = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

Just as in the previous example, the eigenvalues of this matrix are easily found to be $\lambda = 1,1,4$ where now the eigenvalue $\lambda = 1$ has algebraic multiplicity 2. The eigenvector for $\lambda = 4$ is the same as before but for $\lambda = 1$ we have

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

so the (single) eigenvector in this case is $\vec{E}_1 = [1,0,0]^\top$. Since there are not three distinct eigenvalues, theorem 3 does not guarantee three eigenvectors. Here the geometric multiplicity of $\lambda = 1$ is just 1 which is less than the algebraic multiplicity. In general the geometric multiplicity of an eigenvalue is less than or equal to its algebraic multiplicity. When the geometric multiplicity of an eigenvalue is strictly less than its algebraic multiplicity, the matrix will not have a full set of eigenvectors.

The examples here all involve matrices that are 3 by 3. In applications it will frequently be necessary to find the eigenvalues and eigenvectors of matrices that are much larger. Since this involves solving a polynomial equation of high degree, the analysis is usually done by computer. For purposes of illustration here we have used the largest matrices where we can easily find the eigenvalues and eigenvectors without resorting to computers.

Exercises

For each of the following matrices, find the eigenvalues and the eigenvectors.

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

4. $A = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$ 5. $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$.

6. $A = \begin{bmatrix} -2 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$ 7. $A = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$.

8. $A = \begin{bmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{bmatrix}$ 9. $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & -1 & 4 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}$ 11. $A = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 12. $A = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 2 & 4 \\ 0 & 0 & 1 \end{bmatrix}$.

3.2 Discussion of Eigenvalues and Eigenvectors of a Matrix

Students often ask whether there is any physical significance for eigenvalues and eigenvectors for a matrix A. The answer is, "it depends". For example, if A represents the inertia tensor of a rigid body (the inertia tensor for a three dimensional solid is a 3 by 3 matrix) then the eigenvectors of A are the principal axes of rotation and the eigenvalues are the associated moments of inertia of the rigid body. If you are not a student of rigid body dynamics then this example is not very meaningful. There are examples from quantum mechanics which provide interpretations of eigenvalues and eigenvectors but, again, unless you are familiar with quantum mechanics, these examples are not enlightening.

As we will see in the next chapter, our interest in eigenvalues and eigenvectors arises in connection with the solving of systems of linear ordinary differential equations of the form, $\frac{d}{dt}\vec{x}(t) = A\vec{x}(t)$. If such a system is the result of modelling some sort of vibrational system (these are nearly always represented schematically as a system of masses connected to one another by springs), then the eigenvalues can usually be interpreted as the "natural frequencies" of the system and the eigenvectors are the so called "mode shapes" that correspond to these frequencies. The mode shapes describe the patterns of deflections assumed by the masses in the system when it is oscillating at one of the natural frequencies.

More generally, the eigenvalues and eigenvectors for a matrix A do not have a specific physical meaning but they do have a mathematical interpretation. If A has n linearly independent eigenvectors then these vectors form a natural basis in R^n for representing the solution of the system $\frac{d}{dt}\vec{x}(t) = A\vec{x}(t)$. In fact, the solution of this system can be written in the form

$$\vec{x}(t) = e^{tA}\vec{x}(0)$$

but the meaning of this expression can only be interpreted through the use of the eigenvalues and eigenvectors.

Appendix to Chapter Two

The following two theorems and the corollary are the central results bearing on the first problem of linear algebra.

Theorem ***1-*** *Let A be any m by n matrix. Then the column rank of A (number of linearly independent columns) equals the row rank of A (number of linearly independent rows); i.e.,* $\dim CS(A) = \dim RS(A)$.

Since the row rank and column rank of A are equal, we will refer simply to the ***rank*** of A.

Proof- Suppose row rank of $A = r$ and let $\vec{x}_1, \ldots \vec{x}_r$ denote a basis for $RS(A)$. Then the vectors $A\vec{x}_1, \ldots, A\vec{x}_r$ all belong to the range of A, that is, to $CS(A)$. Now suppose

$$C_1A\vec{x}_1 + \ldots + C_rA\vec{x}_r = \vec{0}$$

Then

$$A[C_1\vec{x}_1 + \ldots + C_r\vec{x}_r] = \vec{0}$$

which implies that $C_1\vec{x}_1 + \ldots + C_r\vec{x}_r$ belongs to N_A. But $C_1\vec{x}_1 + \ldots + C_r\vec{x}_r$ also belongs to $RS(A)$ which means $C_1\vec{x}_1 + \ldots + C_r\vec{x}_r$ belongs $N_A \cap N_A^{\perp}$. This means $C_1\vec{x}_1 + \ldots + C_r\vec{x}_r = \vec{0}$. Since $\vec{x}_1, \ldots \vec{x}_r$ are a basis for $RS(A)$, they are linearly independent and this implies the constants are all zero and it follows finally that the vectors $A\vec{x}_1, \ldots, A\vec{x}_r$ are linearly independent. This proves that $\dim CS(A) \geq \dim RS(A)$. Since this holds for any matrix, it follows that it holds for A^T, so $\dim CS(A^T) \geq \dim RS(A^T)$. But $CS(A^T) = RS(A)$ and $RS(A^T) = CS(A)$ so this last result means that $\dim RS(A) \geq \dim CS(A)$. Together these imply $\dim CS(A) = \dim RS(A)$. ■

Theorem *2- Let A be any $m \times n$ matrix. Then* $\dim R_A + \dim N_A = n$

Proof- Let $\dim N_A = p$ and $\dim R_A = q$. Then we want to show that $p + q = n$. Let $\vec{v}_1, \ldots, \vec{v}_p$ denote a basis for N_A. Then we can q additional vectors to extend this basis to $\{\vec{v}_1, \ldots, \vec{v}_p, \vec{w}_1, \ldots, \vec{w}_q\}$ so as to get a basis for R^n. Then for an arbitrary $\vec{x}$ in R^n we have

$$\vec{x} = C_1\vec{v}_1 + \ldots + C_p\vec{v}_p + D_1\vec{w}_1 + \ldots + D_q\vec{w}_q$$

Then

$$\begin{aligned} A\vec{x} &= A[C_1\vec{v}_1 + \ldots + C_p\vec{v}_p + D_1\vec{w}_1 + \ldots + D_q\vec{w}_q] \\ &= \vec{0} + A[D_1\vec{w}_1 + \ldots + D_q\vec{w}_q] \\ &= D_1A\vec{w}_1 + \ldots + D_qA\vec{w}_q \end{aligned}$$

Clearly, the vectors $\{A\vec{w}_1, \ldots, A\vec{w}_q\}$ span R_A and it remains to show that they form a basis for R_A. (for this shows $\dim R_A = q$). Suppose then that

$$a_1A\vec{w}_1 + \ldots + a_qA\vec{w}_q = \vec{0}$$

Then

$$A[a_1\vec{w}_1 + \ldots + a_q\vec{w}_q] = \vec{0}$$

which implies $[a_1\vec{w}_1 + \ldots + a_q\vec{w}_q] \in N_A$. But $\vec{v}_1, \ldots, \vec{v}_p$ is a basis for N_A so there exist scalars $b_1, \ldots, b_p$ such that

$$a_1\vec{w}_1 + \ldots + a_q\vec{w}_q = b_1\vec{v}_1 + \ldots + b_p\vec{v}_p$$

But in this case $a_1\vec{w}_1 + \ldots + a_q\vec{w}_q - b_1\vec{v}_1 - \ldots - b_p\vec{v}_p = \vec{0}$, which means $a_1 = a_2 = \ldots = a_q = b_1 = \ldots = b_p = 0$ since $\{\vec{v}_1, \ldots, \vec{v}_p, \vec{w}_1, \ldots, \vec{w}_q\}$ is a basis for R^n. Then it follows that the vectors $\{A\vec{w}_1, \ldots, A\vec{w}_q\}$ are linearly independent and therefore form a basis for R_A. Then $\dim R_A = q$ and $p + q = n$. ■

Corollary- If M is any subspace in R^n then $\dim M + \dim M^{\perp} = n$.

Proof- If M is a subspace of R^n of dimension $p,$ let $\{\vec{v}_1,\ldots,\vec{v}_p\}$ denote a basis for M and let A be a matrix whose rows are these vectors. Then $N_A = M^{\perp}.$ Now theorem 1 implies $\dim M = \dim R_A = p$ and theorem 2 implies $\dim M + \dim M^{\perp} = n.$