Chapter 2

Matrices and Linear Algebra

2.1 Basics

Definition 2.1.1. A *matrix* is an $m \times n$ array of scalars from a given field F. The individual values in the matrix are called *entries*.

Examples.

$$A = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 2 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

The *size* of the array is—written as $m \times n$, where

$$\underset{\substack{\nearrow \\ \text{number of rows}}}{m} \times \underset{\substack{\nwarrow \\ \text{number of columns}}}{n}$$

Notation

$$A = \begin{array}{cc} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{n1} & a_{n2} & \dots & a_{mn} \end{bmatrix} & \longleftarrow \text{rows} \\ \text{columns} & \end{array}$$

$A :=$ uppercase denotes a matrix

$a :=$ lower case denotes an entry of a matrix $a \in F$.

Special matrices

(1) If $m = n$, the matrix is called *square*. In this case we have

(1a) A matrix A is said to be *diagonal* if

$$a_{ij} = 0 \qquad i \neq j.$$

(1b) A diagonal matrix A may be denoted by $\operatorname{diag}(d_1, d_2, \dots, d_n)$ where

$$a_{ii} = d_i \quad a_{ij} = 0 \qquad j \neq i.$$

The diagonal matrix $\operatorname{diag}(1, 1, \dots, 1)$ is called the *identity* matrix and is usually denoted by

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

or simply I, when n is assumed to be known. $0 = \operatorname{diag}(0, \dots, 0)$ is called the *zero matrix*.

(1c) A square matrix L is said to be *lower triangular* if

$$\ell_{ij} = 0 \qquad i < j.$$

(1d) A square matrix U is said to be upper triangular if

$$u_{ij} = 0 \qquad i > j.$$

(1e) A square matrix A is called *symmetric* if

$$a_{ij} = a_{ji}.$$

(1f) A square matrix A is called *Hermitian* if

$$a_{ij} = \bar{a}_{ji} \quad (\bar{z} := \text{ complex conjugate of } z).$$

(1g) E_{ij} has a 1 in the (i, j) position and zeros in all other positions.

(2) A rectangular matrix A is called *nonnegative* if

$$a_{ij} \geq 0 \quad \text{all} \quad i, j.$$

It is called *positive* if

$$a_{ij} > 0 \quad \text{all} \quad i, j.$$

Each of these matrices has some special properties, which we will study during this course.

Definition 2.1.2. The set of all $m \times n$ matrices is denoted by $M_{m,n}(F)$, where F is the underlying field (usually R or $\mathbb{C}$). In the case where $m = n$ we write $M_n(F)$ to denote the matrices of size $n \times n$.

Theorem 2.1.1. *$M_{m,n}$ is a vector space with basis given by E_{ij}, $1 \le i \le m$, $1 \le j \le n$.*

Equality, Addition, Multiplication

Definition 2.1.3. Two matrices A and B are equal if and only if they have the same size and

$$a_{ij} = b_{ij} \quad \text{all} \quad i, j.$$

Definition 2.1.4. If A is any matrix and $\alpha \in F$ then the scalar multiplication $B = \alpha A$ is defined by

$$b_{ij} = \alpha a_{ij} \quad \text{all} \quad i, j.$$

Definition 2.1.5. If A and B are matrices of the same size then the *sum* A and B is defined by $C = A + B$, where

$$c_{ij} = a_{ij} + b_{ij} \quad \text{all} \quad i, j$$

We can also compute the *difference* $D = A - B$ by summing A and $(-1)B$

$$D = A - B = A + (-1)B.$$

matrix subtraction.

Matrix addition "inherits" many properties from the field F.

Theorem 2.1.2. *If $A, B, C \in M_{m,n}(F)$ and $\alpha, \beta \in F$, then*

(1) $A + B = B + A$ *commutivity*

(2) $A + (B + C) = (A + B) + C$ *associativity*

(3) $\alpha(A + B) = \alpha A + \alpha B$ *distributivity of a scalar*

(4) If $B = 0$ (a matrix of all zeros) then

$$A + B = A + 0 = A$$

(4) $(\alpha + \beta)A = \alpha A + \beta A$

(5) $\alpha(\beta A) = \alpha\beta A$

(6) $0A = 0$

(7) $\alpha\, 0 = 0.$

Definition 2.1.6. If x and $y \in R_n$,

$$x = (x_1 \dots x_n)$$
$$y = (y_1 \dots y_n).$$

Then the <u>scalar</u> or <u>dot</u> product of x and y is given by

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i.$$

Remark 2.1.1. (i) Alternate notation for the scalar product: $\langle x, y\rangle = x \cdot y$.
(ii) The dot product is defined only for vectors of the same length.

Example 2.1.1. Let $x = (1, 0, 3, -1)$ and $y = (0, 2, -1, 2)$ then $\langle x, y\rangle = 1(0) + 0(2) + 3(-1) - 1(2) = -5$.

Definition 2.1.7. If A is $m \times n$ and B is $n \times p$. Let $r_i(A)$ denote the vector with entries given by the i^{th} row of A, and let $c_j(B)$ denote the vector with entries given by the j^{th} row of B. The product $C = AB$ is the $m \times p$ matrix defined by

$$c_{ij} = \langle r_i(A), c_j(B)\rangle$$

where $r_i(A)$ is the vector in R_n consisting of the i^{th} row of A and similarly $c_j(B)$ is the vector formed from the j^{th} column of B. Other notation for $C = AB$

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \quad \begin{array}{l} 1 \le i \le m \\ 1 \le j \le p. \end{array}$$

Example 2.1.2. Let

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

Then

$$AB = \begin{bmatrix} 1 & 2 \\ 11 & 4 \end{bmatrix}.$$

Properties of matrix multiplication

(1) If AB exists, does it happen that BA exists and $AB = BA$? The answer is usually no. First AB and BA exist if and only if $A \in M_{m,n}(F)$ and $B \in M_{n,m}(F)$. Even if this is so the sizes of AB and BA are different (AB is $m \times m$ and BA is $n \times n$) unless $m = n$. However even if $m = n$ we may have $AB \neq BA$. See the examples below. They may be different sizes and if they are the same size (i.e. A and B are square) the entries may be different

$$A = [1, 2] \quad B = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad AB = [1]$$

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

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} B = \begin{bmatrix} -1 & 1 \\ 0 & 1 \end{bmatrix} \qquad AB = \begin{bmatrix} -1 & 3 \\ -3 & 7 \end{bmatrix}$$

$$BA = \begin{bmatrix} 2 & 2 \\ 3 & 4 \end{bmatrix}$$

(2) If A is square we define

$$A^1 = A, \quad A^2 = AA, \quad A^3 = A^2A = AAA$$
$$A^n = A^{n-1}A = A \cdots A \quad (n \text{ factors}).$$

(3) $I = \text{diag}(1, \ldots, 1)$. If $A \in M_{m,n}(F)$ then

$$AI_n = A \quad \text{and}$$
$$I_m A = A.$$

Theorem 2.1.3 (Matrix Multiplication Rules). *Assume $A, B,$ and C are matrices for which all products below make sense. Then*

(1) $A(BC) = (AB)C$

(2) $A(B \pm C) = AB \pm AC$ *and* $(A \pm B)C = AC \pm BC$

(3) $AI = A$ *and* $IA = A$

(4) $c(AB) = (cA)B$

(5) $A0 = 0$ *and* $0B = 0$

(6) For A square

$$A^r A^s = A^s A^r \quad \textit{for all integers} \quad r, s \geq 1.$$

Fact: If AC and BC are equal, it does not follow that $A = B$. See Exercise 60.

Remark 2.1.2. We use an alternate notation for matrix entries. For any matrix B denote the (i, j)-entry by $(B)_{ij}$.

Definition 2.1.8. Let $A \in M_{m,n}(F)$.

(i) Define the *transpose* of A, denoted by A^T, to be the $n \times m$ matrix with entries

$$(A^T)_{ij} = a_{ji}.$$

(ii) Define the *adjoint* of A, denoted by A^*, to be the $n \times m$ matrix with entries

$$(A^*)_{ij} = \bar{a}_{ji} \quad \text{complex conjugate}$$

Example 2.1.3.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 4 & 1 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 5 \\ 2 & 4 \\ 3 & 1 \end{bmatrix}$$

In words... "The rows of A become the columns of A^T, taken in the same order." The following results are easy to prove.

Theorem 2.1.4 (Laws of transposes). *(1)* $(A^T)^T = A$ *and* $(A^*)^* = A$

(2) $(A \pm B)^T = A^T \pm B^T$ *(and for $*$)*

(3) $(cA)^T = cA^T \qquad (cA)^* = \bar{c}A^*$

(4) $(AB)^T = B^T A^T$

(5) If A is symmetric

$$A = A^T$$

(6) If A is Hermitian

$$A = A^*.$$

More facts about symmetry.

Proof. (1) We know $(A^T)_{ij} = a_{ji}$. So $((A^T)^T)_{ij} = a_{ij}$. Thus $(A^T)^T = A$.

(2) $(A \pm B)^T = a_{ji} \pm b_{ji}$. So $(A \pm B)^T = A^T \pm B^T$.

□

Proposition 2.1.1. *(1) A is symmetric if and only if A^T is symmetric.*

(1) A is Hermitian if and only if A^* is Hermitian.*

(2) If A is symmetric, then A^2 is also symmetric.

(3) If A is symmetric, then A^n is also symmetric for all n.

Definition 2.1.9. A matrix is called skew-symmetric if

$$A^T = -A.$$

Example 2.1.4. The matrix

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

is skew-symmetric.

Theorem 2.1.5. *(1) If A is skew symmetric, then A is a square matrix and $a_{ii} = 0$, $i = 1, \ldots, n$.*

(2) For any matrix $A \in M_n(F)$

$$A - A^T$$

is skew-symmetric while $A + A^T$ is symmetric.

(3) Every matrix $A \in M_n(F)$ can be uniquely written as the sum of a skew-symmetric and symmetric matrix.

Proof. (1) If $A \in M_{m,n}(F)$, then $A^T \in M_{n,m}(F)$. So, if $A^T = -A$ we must have $m = n$. Also

$$a_{ii} = -a_{ii}$$

for $i = 1, \ldots, n$. So $a_{ii} = 0$ for all i.

(2) Since $(A - A^T)^T = A^T - A = -(A - A^T)$, it follows that $A - A^T$ is skew-symmetric.

(3) Let $A = B + C$ be a second such decomposition. Subtraction gives

$$\frac{1}{2}(A + A^T) - B = C - \frac{1}{2}(A - A^T).$$

The left matrix is symmetric while the right matrix is skew-symmetric. Hence both are the zero matrix.

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

□

Examples. $A = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right]$ is skew-symmetric. Let

$$\begin{aligned} B &= \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \\ B^T &= \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \\ B - B^T &= \begin{bmatrix} 0 & 3 \\ -3 & 0 \end{bmatrix} \\ B + B^T &= \begin{bmatrix} 2 & 1 \\ 1 & 8 \end{bmatrix}. \end{aligned}$$

Then

$$B = \frac{1}{2}(B - B^T) + \frac{1}{2}(B + B^T).$$

An important observation about matrix multiplication is related to ideas from vector spaces. Indeed, two very important vector spaces are associated with matrices.

Definition 2.1.10. Let $A \in M_{m,n}(\mathbb{C})$.
(i)Denote by

$$c_j(A) := j^{\text{th}} \text{ column of } A$$

$c_j(A) \in \mathbb{C}_m$. We call the subspace of $\mathbb{C}_m$ spanned by the columns of A the **column space** of A. With $c_1(A), \ldots, c_n(A)$ denoting the columns of A

the column space is $\mathfrak{S}\left(c_1\left(A\right),\ldots,c_n\left(A\right)\right).$
(ii) Similarly, we call the subspace of $\mathbb{C}_n$ spanned by the rows of A the **row space** of A. With $r_1\left(A\right),\ldots,r_m\left(A\right)$ denoting the rows of A the row space is therefore $\mathfrak{S}\left(r_1\left(A\right),\ldots,r_m\left(A\right)\right).$

Let $x \in \mathbb{C}_n$, which we view as the $n \times 1$ matrix $x = [x_1 \ldots x_n]^T$. The product Ax is defined and

$$Ax = \sum_{j=1}^{n} x_j c_j(A).$$

That is to say, $Ax \in \mathfrak{S}(c_1(A), \ldots, c_n(A)) =$ *column* space of A.

Definition 2.1.11. Let $A \in M_n(F)$. The matrix A is said to be **invertible** if there is a matrix $B \in M_n(F)$ such that

$$AB = BA = I.$$

In this case B is called the *inverse* of A, and the notation for the inverse is A^{-1}.

Examples.

(i) Let

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

Then

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

(ii) For $n = 3$ we have

$$A = \begin{bmatrix} 1 & 2 & -1 \\ -1 & 3 & -1 \\ -2 & 3 & -1 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 3 & -2 \\ -3 & 7 & -5 \end{bmatrix}$$

A square matrix need not have an inverse, as will be discussed in the next section. As examples, the two matrices below do not have inverses

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

2.2 Linear Systems

The solutions of linear systems is likely the single largest application of matrix theory. Indeed, most reasonable problems of the sciences and economics that have the need to solve problems of several variable almost without exception are reduced to component parts where one of them is the solution of a linear system. Of course the entire solution process may have the linear system solver as a relatively small component, but an essential one. Even the solution of nonlinear problems, especially, employ linear systems to great and crucial advantage.

To be precise, we suppose that the coefficients $a_{ij}, \quad 1 \leq i \leq m$ and $1 \leq j \leq n$ and the data $b_j,\ 1 \leq j \leq m$ are known. We define the *linear system* for the n unknowns $x_1, \ldots, x_n$ to be

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

The *solution set* is defined to be the subset of R_n of vectors $(x_1, \ldots, x_n)$ that satisfy each of the m equations of the system. The question of how to solve a linear system includes a vast literature of theoretical and computation methods. Certain systems form the model of what to do. In the systems below we note that the first one has three highly coupled (interrelated) variables.

$$\begin{aligned} 3x_1 - 2x_2 + 4x_3 &= 7 \\ x_1 - 6x_2 - 2x_3 &= 0 \\ -x_1 + 3x_2 + 6x_3 &= -2 \end{aligned}$$

The second system is more tractable because there appears even to the untrained eye a clear and direct method of solution.

$$\begin{aligned} 3x_1 - 2x_2 - x_3 &= 7 \\ x_2 - 2x_3 &= 1 \\ 2x_3 &= -2 \end{aligned}$$

Indeed, we can see right off that $x_3 = -1$. Substituting this value into the second equation we obtain $x_2 = 1 - 2 = -1$. Substituting both x_2 and x_3 into the first equation, we obtain $2x_1 - 2(-1) - (-1) = 7$, gives $x_1 = 2$. The solution set is the vector $(2,-1,-1)$. The virtue of the second system is that the unknowns can be determined one-by-one, back substituting those already found into the next equation until all unknowns are determined. So if we can convert the given system of the first kind to one of the second kind, we can determine the solution.

This procedure for solving linear systems is therefore the applications of operations to effect the gradual elimination of unknowns from the equations until a new system results that can be solved by direct means. The operations allowed in this process must have precisely one important property: *They must not change the solution set by either adding to it or subtracting from it.* There are exactly three such operations needed to reduce any set of linear equations so that it can be solved directly.

(E1) Interchange two equations.

(E2) Multiply any equation by a nonzero constant.

(E3) Add a multiple of one equation to another.

This can be summarized in the following theorem

Theorem 2.2.1. *Given the linear system (*). The set of equation operations E1, E2, and E3 on the equations of (*) does not alter the solution set of the system (*).*

We leave this result to the exercises. Our main intent is to convert these operations into corresponding operations for matrices. Before we do this we clarify which linear systems can have a soltution. First, the system can be converted to matrix form by setting A equal to the $m \times n$ matrix of coefficients, b equal to the $m \times 1$ vector of data, and x equal to the $n \times 1$ vector of unknowns. Then the system (*) can be written as

$$Ax = b$$

In this way we see that with $c_i(A)$ denoting the i^{th} column of A, the system is expressible as

$$x_1 c_1(A) + \cdots + x_n c_n(A) = b$$

From this equation it is clear that the system has a solution if and only if the vector b is in $\mathfrak{S}(c_1(A), \cdots, c_n(A))$. This is summarized in the following theorem.

Theorem 2.2.2. *A necessary and sufficient condition that $Ax = b$ has a solution is that $b \in \mathfrak{S}(c_1(A) \dots c_n(A))$.*

In the general matrix product $C = AB$, we note that the column space of $C \subset$ column space of A. In the following definition we regard the matrix A as a function acting upon vectors in one vector space with range in another vector space. This is entirely similar to the domain-range idea of function theory.

Definition 2.2.1. The **range** of $A = \{Ax \mid x \in R_n \text{ (or} \mathbb{C}_n \text{)}\}$.

It follows directly from our discussion above that the range of A equals $\mathfrak{S}(c_1(A), \dots, c_n(A))$.
Row operations: To solve $Ax = b$ we use a process called *Gaussian elimination*, which is based on row operations.

Type 1: Interchange two rows. (Notation: $R_i \longleftrightarrow R_j$)
Type 2: Multiply a row by a nonzero constant. (Notation: $cR_i \to R_i$)
Type 3: Add a multiple of one row to another row. (Notation: $cR_i + R_j \to R_j$)

Gaussian elimination is the process of reducing a matrix to its RREF using these row operations. Each of these operations is the respective analogue of the equation operations described above, and each can be realized by left matrix multiplication. We have the following.
Type 1

$$
\begin{array}{cc}
E_1 = \left[\begin{array}{cccccccccc}
1 & & \vdots & & & & \vdots & & & \\
 & 1 & \vdots & & & & \vdots & & & \\
\cdots & \cdots & 0 & \cdots & \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\
 & & \vdots & 1 & & & \vdots & & & \\
 & & \vdots & & \ddots & & \vdots & & & \\
 & & \vdots & & & 1 & \vdots & & & \\
\cdots & \cdots & 1 & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & \cdots \\
 & & \vdots & & & & \vdots & 1 & & \\
 & & \vdots & & & & \vdots & & \ddots & \\
 & & \vdots & & & & \vdots & & & 1
\end{array}\right]
&
\begin{array}{c}
\\ \\ \text{row } i \\ \\ \\ \\ \text{row j} \\ \\ \\ \\
\end{array}
\\
\begin{array}{cc}
\qquad\qquad \begin{array}{c}\text{column}\\ i\end{array} & \qquad \begin{array}{c}\text{column}\\ j\end{array}
\end{array}
&
\end{array}
$$

Notation: $R_i \leftrightarrow R_j$

Type 2

$$E_2 = \begin{array}{cl} \begin{bmatrix} 1 & & & \vdots & & & \\ & \ddots & & \vdots & & & \\ & & 1 & \vdots & & & \\ \cdots & \cdots & \cdots & c & \cdots & \cdots & \cdots \\ & & & \vdots & 1 & & \\ & & & \vdots & & \ddots & \\ & & & \vdots & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \\ \quad \text{row } i \\ \\ \\ \\ \end{array} \\ \text{column } i & \end{array}$$

Notation: cR_i

Type 3

$$E_3 = \begin{array}{cl} \begin{bmatrix} 1 & & \vdots & & & & \\ & 1 & \vdots & & & & \\ & & \vdots & & & & \\ & & \vdots & \ddots & & & \\ \cdots & \cdots & c & \cdots & \cdots & \cdots & \cdots \\ & & \vdots & & & 1 & \\ & & \vdots & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \\ \quad \text{row } j \\ \\ \\ \\ \end{array} \\ \begin{array}{c} \text{column} \\ i \end{array} & \end{array}$$

Notation: $cR_i + R_j$, the abbreviated form of $cR_i + R_j \to R_j$

Example 2.2.1. The operations

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} \begin{array}{c} R_1 \longleftrightarrow R_2 \\ \to \end{array} \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{array}{c} 4R_3 \\ \to \end{array} \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -4 & 0 & 8 \end{bmatrix}$$

can also be realized as

$$R_1 \longleftrightarrow R_2: \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

$$4R_3 \quad : \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 0 \\ -4 & 0 & 8 \end{bmatrix}$$

The operations

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ -1 & 0 & 2 \end{bmatrix} \underset{2R_1 + R_3}{\overset{-3R_1 + R_2}{\rightarrow}} \begin{bmatrix} 2 & 1 & 0 \\ -6 & -1 & 1 \\ 3 & 2 & 2 \end{bmatrix}$$

can be realized by the left matrix multiplications

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

Note there are two matrix multiplications them, one for each Type 3 elementary operation.

Row-reduced echelon form. To each $A \in M_{m,n}(E)$ there is a canonical form also in $M_{m,n}(E)$ which may be obtained by row operations. Called the RREF, it has the following properties.

(a) Each nonzero row has a 1 as the first nonzero entry (:= *leading one*).

(b) All column entries above and below a leading one are zero.

(c) All zero rows are at the bottom.

(d) The leading one of one row is to the left of leading ones of all lower rows.

Example 2.2.2.

$$B = \begin{bmatrix} 1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{is in RREF.}$$

Theorem 2.2.3. *Let $A \in M_{m,n}(F)$. Then the RREF is necessarily unique.*

We defer the proof of this result. Let $A \in M_{m,n}(F)$. Recall that the *row space* of A is the subspace of R_n (or $\mathbb{C}_n$) spanned by the rows of A. In symbols the row space is

$$\mathfrak{S}(r_1(A), \ldots, r_m(A)).$$

Proposition 2.2.1. *For $A \in M_{m,n}(F)$ the rows of its RREF span the rows space of A.*

Proof. First, we know the nonzero rows of the RREF are linearly independent. And all row operations are linear combinations of the rows. Therefore the row space generated from the RREF is contained in the row space of A. If the containment is proper. That is there is a row of A that is linearly independent from the row space of the RREF, this is a contradiction because every row of A can be obtained by the inverse row operations from the RREF. □

Proposition 2.2.2. *If $A \in M_{m,n}(F)$ and a row operation is applied to A, then linearly dependent columns of A remain linearly dependent and linearly independent columns of A remain linearly independent.*

Proposition 2.2.3. *The number of linearly independent columns of $A \in M_{m,n}(F)$ is the same as the number of leading ones in the RREF of A.*

Proof. Let $S = \{i_1 \ldots i_k\}$ be the columns of the RREF of A having a leading one. These columns of the RREF are linearly independent Thus these columns were originally linearly independent. If another column is linearly independent, this column of the RREF is linearly dependent on the columns with a leading one. This is a contradiction to the above proposition. □

Proof of Theorem 2.2.3. By the way the RREF is constructed, left-to-right, and top-to-bottom, it should be apparent that if the right most row of the RREF is removed, there results the RREF of the $m \times (n-1)$ matrix formed from A by deleting the n^{th} column. Similarly, if the bottom row of the RREF is removed there results a new matrix in RREF form, though not simply related to the original matrix A.

To prove that the RREF is unique, we proceed by a double induction, first on the number of columns. We take it as given that for an $m \times 1$ matrix the RREF is unique. It is either the zero $m \times 1$ matrix, which would be the case if A was zero or the matrix with a 1 in the first row and zeros in the other rows. Assume therefore that the RREF is unique if the number of columns is less than n. Assume there are two RREF forms, B_1 and B_2 for A. Now the RREF of A is therefore unique through the $(n-1)^{\text{st}}$ columns. The only difference between the RREF's B_1 and B_2 must occur in the n^{th} column. Now proceed by induction on the number of nonzero rows. Assume that $A \neq 0$. If A has just one row, the RREF of A is simply **the** scalar multiple of A that makes the first nonzero column entry a one. Thus it is unique. If $A = 0$, the RREF is also zero. Assume now that the RREF is unique for matrices with less than m rows. By the comments above that the only difference between the RREF's B_1 and B_2 can occur at the (m, n)-entry. That is $(B_1)_{m,n} \neq (B_2)_{m,n}$. They are therefore not leading ones. (Why?) There is a leading one in the m^{th} row, however, because it is a non zero row. Because the row spaces of B_1 and B_2 are identical, this results in a contradiction, and therefore the (m, n)-entries must be equal. Finally, $B_1 = B_2$. This completes the induction. (Alternatively, the two systems pertaining to the RREF's must have the same solution set to the system $Ax = 0$. With $(B_1)_{m,n} \neq (B_2)_{m,n}$, it is easy to see that the solution sets to $B_1x = 0$ and $B_2x = 0$ must differ.) □

Definition 2.2.2. Let $A \in M_{m,n}$ and $b \in R_m$ (or $\mathbb{C}_n$). Define

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

$[A|b]$ is called the *augmented matrix* of A by b. $[A|b] \in M_{m,n+1}(F)$. The augmented matrix is a useful notation for finding the solution of systems using row operations.

Identical to other definitions for solutions of equations, the equivalence of two systems is defined via the idea of equality of the solution set.

Definition 2.2.3. Two linear systems $Ax = b$ and $Bx = c$ are called **equivalent** if one can be converted to the other by elementary equation operations.

It is easy to see that this implies the following

Theorem 2.2.4. *Two linear systems $Ax = b$ and $Bx = c$ are equivalent if and only if both $[A|b]$ and $[B|c]$ have the same row reduced echelon form.*

We leave the prove to the reader. (See Exercise 23.) Note that the solution set need not be a single vector; it can be null or infinite.

2.3 Rank

Definition 2.3.1. The **rank** of any matrix A, denote by $r(A)$, is the dimension of its column space.

Proposition 2.3.1. *(i) The rank of A equals the number of nonzero rows of the RREF of A, i.e. the number of leading ones.*
(ii) $r(A) = r(A^T)$.

Proof. (i) Follows from previous results.
(ii) The number of linearly independent rows equals the number of linearly independent columns. The number of linearly independent rows is the number of linearly independent columns of A^T—by definition. Hence $r(A) = r(A^T)$. □

Proposition 2.3.2. *Let $A \in M_{m,n}(\mathbb{C})$ and $b \in \mathbb{C}_m$. Then $Ax = b$ has a solution if and only if $r(A) = r([A|b])$, where $[A|b]$ is the augmented matrix.*

Remark 2.3.1. Solutions may exist and may not. However, even if a solution exists, it may not be unique. Indeed if it is not unique, there is an infinity of solutions.

Definition 2.3.2. When $Ax = b$ has a solution we say the system is *consistent.*

Naturally, in practical applications we want our systems to be consistent. When they are not, this can be an indicator that something is wrong with the underlying physical model. In mathematics, we also want consistent systems; they are usually far more interesting and offer richer environments for study.

In addition to the column and row spaces, another space of great importance is the so-called null space, the set of vectors $x \in R_n$ for which $Ax = 0$. In contrast, when solving the simple single variable linear equation $ax = b$ with $a \neq 0$ we know there is always a unique solution $x = b/a$. In solving even the simplest higher dimensional systems, the picture is not as clear.

Definition 2.3.3. Let $A \in M_{m,n}(F)$. The **null space** of A is defined to be

$$\text{Null}(A) = \{x \in R_n \mid Ax = 0\}.$$

It is a simple consequence of the linearity of matrix multiplication that $\text{Null}(A)$ is a linear subspace of R_n. That is to say, $\text{Null}(A)$ is closed under vector addition and scalar multiplication. In fact, $A(x + y) = Ax + Ay = 0 + 0 = 0$, if $x, y \in \text{Null}(A)$. Also, $A(\alpha x) = \alpha Ax = 0$, if $x \in \text{Null}(A)$. We state this formally as

Theorem 2.3.1. *Let $A \in M_{m,n}(F)$. Then Null(A) is a subspace of R_n while the range of A is in R_m.*

Having such solutions gives valuable information about the solution set of the linear system $Ax = b$. For, if we have found a solution, x, and have any vector $z \in \text{Null}(A)$, then $x + z$ is a solution of the same linear system. Indeed, what is easy to see is that if u and v are both solutions to $Ax = b$, then $A(u - v) = Au - Av = 0$, or what is the same $x - y \in \text{Null}(A)$. This means that to find *all* solutions to $Ax = b$, we need only find a single solution and the null space. We summarize this as the following theorem.

Theorem 2.3.2. *Let $A \in M_{m,n}(F)$ with null space Null(A). Let x be any nonzero solution to $Ax = b$. Then the set $x +$ Null(A) is the entire solution set to $Ax = b$.*

Example 2.3.1. Find the null space of $A = \begin{bmatrix} 1 & 3 \\ -3 & -9 \end{bmatrix}$.

Solution. Solve $Ax = 0$. The RREF for A is $\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$. Solving $x_1 + 3x_2 = 0$, take $x_2 = t$, a "free" parameter and solve for x_1 to get $x_1 = -3t$. Thus every solution to $Ax = 0$ can be written in the form

$$x = \begin{bmatrix} -3t \\ t \end{bmatrix} = t \begin{bmatrix} -3 \\ 1 \end{bmatrix} \qquad t \in R$$

Expressed this way we see that $\text{Null}(A) = \left\{ t \begin{bmatrix} -3 \\ 1 \end{bmatrix} \mid t \in R \right\}$, a subspace of R_2 of dimension 1.

Theorem 2.3.3 (Fundamental theorem on rank). *$A \in M_{m,n}(F)$. The following are equivalent*

(a) $r(A) = k$.

(b) There exist exactly k linearly independent columns of A.

(c) There exist exactly k linearly independent rows of A.

(d) The dimension of the column space of A is k (i.e. $\dim(\text{Range } A) = k$).

(e) There exists a set S of exactly k vectors in R_m for which $Ax = b$ has a solution for each $b \in \mathfrak{S}(S)$.

(f) The null space of A has dimension $n - k$.

Proof. The equivalence of (a), (b), (c) and (d) follow from previous considerations. To establish (e), let $S = \{c_{\ell_1}, c_{\ell_2}, \dots, c_{\ell_k}\}$ denote the linearly independent column vectors of A. Let $T = \{e_{\ell_1}, e_{\ell_2}, \dots, e_{\ell_k}\} \subset R_n$ be the standard vectors. Then $Ae_{\ell_j} = c_{\ell_j}$. If $b \in \mathfrak{S}(S)$, then $b = a_1 c_{\ell_1} + a_2 c_{\ell_2} + \cdots + a_k c_{\ell_k}$. A solution to $Ax = b$ is given by $x = a_1 e_{\ell_1} + a_2 e_{\ell_2} + \cdots + a_k e_{\ell_k}$. Conversely, if (e) holds, then the set S must be linearly independent for otherwise S could be reduced to $k-1$ or fewer vectors. Similarly if A has $k+1$ linearly independent columns then set S can be expanded. Therefore, the column space of A must have exactly k vectors.

To prove (f) we assume that $S = \{v_1, \dots, v_k\}$ is a basis for the column space of A. Let $T = \{w_1, \dots, w_k\} \subset R_n$ for which $Aw_i = v_i$, $i = 1, \dots, k$. By our extension theorem, we select $n-k$ vectors $w_{k+1}, \dots, w_n$ such that $U = \{w_1, \dots, w_k, w_{k+1}, \dots, w_n\}$ is a basis of R_n. We must have that $Aw_{k+1} \in \mathfrak{S}(S)$. Hence there are scalars $b_1, \dots, b_k$ such that

$$Aw_{k+1} = A(b_1 w_1 + \cdots + b_k w_k)$$

and thus $w'_{k+1} = w_{k+1} - (b_1 w_1 + \cdots + b_k w_k)$ is in the null space of A. Repeat this process for each w_{k+j}, $j = 1, \dots, n-k$. We generate a total of $n-k$ vectors $\{w'_{k+1}, \dots, w'_n\}$ in this manner. This set must be linearly independent. (Why?) Therefore, the dimension of the null space must be at least $n-k$. Now we consider a new basis which consists of the original vectors and the $n-k$ vectors $\{w'_{k+1}, w'_{k+2}, \dots, w'_n\}$ for which $Aw = 0$. We assert that the dimension of the null space is exactly $n-k$. For if $z \in R_n$ is a vector for which $Az = 0$, then z can be uniquely written as a component z_1 from $\mathfrak{S}(T)$ and a component z_2 from $\mathfrak{S}(\{w'_{k+1}, \dots, w'_n\})$. But $Az_1 \neq 0$ and $Az_2 = 0$. Therefore $Az = 0$ is impossible unless the component $z_1 = 0$.

Conversely, if (f) holds we take a basis for the null space $T = \{u_1, u_2, \dots, u_{n-k}\}$ and extend the basis

$$T' = T \cup \{u_{n-k+1}, \ldots, u_n\}$$

to R_n. Next argue similarly to above that

$$Au_{n-k+1},\ Au_{n-k+2}, \dots, Au_n$$

must be linearly independent, for otherwise there is yet another linearly independent vector that can be added to its basis, a contradiction. Therefore the column space must have dimension at least, and hence equal to k. $\square$

The following corollary assembles many consequences of this theorem.

Corollary 2.3.1. *(1)* $r(A) \leq \min(m, n)$.

(2) $r(AB) \leq \min(r(A), r(B))$.

(3) $r(A + B) \leq r(A) + r(B)$.

(4) $r(A) = r(A^T) = r(A^*) = r(\bar{A})$.

(5) If $A \in M_m(F)$ *and* $B \in M_{m,n}(F)$, *and if* A *is invertible, then*

$$r(AB) = r(B).$$

Similarly, if $C \in M_n(F)$ *is invertible and* $B \in M_{m,n}(F)$

$$r(BC) = r(B).$$

(6) $r(A) = r(A^T A) = r(A^* A)$.

(7) Let $A \in M_{m,n}(F)$, *with* $r(A) = k$. *Then* $A = XBY$ *where* $X \in M_{m,k}$, $Y \in M_{k,n}$ *and* $B \in M_k$ *is invertible.*

(8) In particular, every rank 1 matrix has the form $A = xy^T$, *where* $x \in R_m$ *and* $y \in R_n$. *Here*

$$xy^T = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_n \\ \vdots & \vdots & & \vdots \\ x_m y_1 & x_m y_2 & \dots & x_m y_n \end{bmatrix}.$$

Proof. (1) The rank of any matrix is the number of linearly independent rows, which is the same as the number of linearly independent columns. The maximum this value can be is therefore the maximum of the minimum of the dimensions of the matrix, or $r(A) \leq \min(m, n)$.

(2) The product AB can be viewed in two ways. The first is as a set of linear combinations of the rows of B, and the other is as a set of linear combinations of the columns of A. In either case the number of linear independent rows (or columns as the case may be) In other words, the rank of the product AB cannot be greater than the number of linearly independent columns of A nor greater than the number of linearly independent rows of B. Another way to express this is as $r(AB) \leq \min(r(A), r(B))$

(3) Now let $S = \{v_1, \dots v_{r(A)}\}$ and $T = \{w_1, \dots, w_{r(B)}\}$ be basis of the column spaces of A and B respectively. Then, the dimension of the union $S \cup T = \{v_1, \dots v_{r(A)}, w_1, \dots, w_{r(B)}\}$ cannot exceed $r(A) + r(B)$. Also, every vector in the column space of $A + B$ is clearly in the span of $S \cup T$. The result follows.

(4) The rank of A is the number of linearly independent rows (and columns) of A, which in turn is the number of linearly independent columns of A^T, which in turn is the rank of A^T. That is, $r(A) = r\left(A^T\right)$. Similar proofs hold for A^* and $\bar{A}$.

(5) Now suppose that $A \in M_m(F)$ is invertible and $B \in M_{m,n}$. As we have emphasized many times the rows of the product AB can be viewed as a set of linear combinations of the rows of B. Since A has rank m any set of linearly independent rows of B remains linearly independent. To see why, let $r_i(AB)$ denote the i^{th} row of the product AB. Then it is easy to see that

$$r_i(AB) = \sum_{j=1}^{m} a_{ij} r_j(B)$$

Suppose we can determine constants $c_{1,} \dots, c_m$ not all zero so that

$$\begin{aligned} 0 &= \sum_{j=1}^{m} c_i r_i(AB) = \sum_{i=1}^{m} c_i \sum_{j=1}^{m} a_{ij} r_j(B) \\ &= \sum_{j=1}^{m} r_j(B) \sum_{i=1}^{m} c_i a_{ij} \end{aligned}$$

This linear combination of the rows of B has coefficient given by $A^T c$, where $c = [c_{1,} \dots, c_k]^T$. Because the rank of A (and A^T) is m, we can solve this system for any vector $d \in R_m$. Suppose that the row vectors $r_{j_l}(B)$, $\ell = 1, \dots, r(B)$, are linearly independent. Arrange that the components of d to be zero for indices not included in the set j_l, $\ell = 1, \dots, r(B)$ and not all zero otherwise. Then the conclusion $0 = \sum_{j=1}^{m} r_j(B) \sum_{i=1}^{m} c_i a_{ij} = \sum_{l=1}^{r(B)} r_{j_l}(B) d_{j_l}$ is impossible. Indeed, the same basis of the row space of B will be a basis of the row space of AB. This proves the result.

(6) We postpone the proof of this result until we discuss orthogonality.

(7) Place A in RREF, say A_{RREF}. Since $r(A) = k$ we know the top k rows of A_{RREF} are linearly independent and the remaining rows are zero. Define Y to be the $k \times n$ matrix consisting of these top k rows. Define $B = I_k$. Now the rows of A are linear combinations of these rows. So, define the $m \times k$ matrix X to have rows as follows: The first row of consists of the coefficients so that $\sum x_{1j} r_j(Y) = r_1(A)$. In general, the i^{th} row of X is selected so that

$$\sum x_{ij} r_j(Y) = r_i(A)$$

(8) This is an application of (7) noting in this special case that X is an $m \times 1$ matrix that can be interpretted as a vector $x \in R_m$. Similarly, Y is an $1 \times n$ matrix that can be interpretted as a vector $y \in R_n$. Thus, with $I = [1]$, we have

$$A = xy^T$$

$\square$

Example 2.3.2. Here is the decomposition of the form given in Lemma 2.3.1 (7). The 3×4 matrix A has rank 2.

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

The matrix Y is the RREF of A.

Example 2.3.3. Let $x = [x_1, x_2, \ldots, x_m]^T \in R_m$ and $y = [y_1, y_2, \ldots y_n]^T \in R_n$. Then the rank one $m \times n$ matrix xy^T has the form

$$xy^T = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & & x_2y_n \\ \vdots & & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

In particular, with $x = [1, 3, 5]^T$, and $y = [-2, 7]^T$, the rank one 3×2 matrix xy^T is given by

$$xy^T = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} [-2,\ 7] = \begin{bmatrix} -2 & 7 \\ -6 & 21 \\ -10 & 35 \end{bmatrix}$$

Invertible Matrices

A subclass matrices $A \in M_n(F)$ that have only the zero kernel is very important in applications and theoretical developments.

Definition 2.3.4. $A \in M_n$ is called **nonsingular** if $Ax = 0$ implies that $x = 0$.

In many texts such matrices are introduced though an equivalent alternate definition involving rank.

Definition 2.3.5. $A \in M_n$ is **nonsingular** if $r(A) = n$.

We also say that nonsingular matrices have *full* rank. That nonsingular matrices are invertible and conversely together with many other equivalences is the content of the next theorem.

Theorem 2.3.4. *[Fundamental theorem on inverses] Let $A \in M_n(F)$. Then the following statements are equivalent.*

(a) *A is nonsingular.*

(b) *A is invertible.*

(c) $r(A) = n$.

(d) *The rows and columns of A are linearly independent.*

(e) $\dim(\text{Range}(A)) = n$.

(f) $\dim(\text{Null}(A)) = 0$.

(g) *$Ax = b$ is consistent for all $b \in R_n$ (or $\mathbb{C}_n$).*

(h) *$Ax = b$ has a unique solution for every $x \in R_n$ (or $\mathbb{C}_n$).*

(i) *$Ax = 0$ has only the zero solution.*

*(j)** *0 is not an eigenvalue of A.*

*(k)** $\det A \neq 0$.

* The statements about eigenvalues and the determinant ($\det A$) of a matrix will be clarified later after they have been properly defined. They are included now for completeness.

Definition 2.3.6. Two linear systems $Ax = b$ and $Bx = c$ are called **equivalent** if one can be converted to the other by elementary equation operations. Equivalently, the systems are equivalent if $[A \mid b]$ can be converted to $[B \mid c]$ by elementary row operations.

Alternatively, the systems are equivalent if they have the same solution set which means of course that both can be reduced to the same RREF.

Theorem 2.3.5. *If $A \in M_n(F)$ and $B \in M_n(F)$ with $AB = I$, then B is unique.*

Proof. If $AB = I$ then for every $e_1 \dots e_n$ there is a solution to the system $Ab_i = e_i$ for all $1 = 1, 2, \dots, n$. Thus the set $\{b_i\}_{i=1}^n$ is linearly independent (because the set $\{e_i\}$ is) and moreover a basis. Similarly if $AC = I$ then $A(C - B) = 0$, and there are $c_i \in R_n$ (or $\mathbb{C}_n$), $i = 1, \dots, n$ such that $Ac_i = e_i$. Suppose for example that $c_1 - b_1 \neq 0$. Since the $\{b_i\}_{i=1}^n$ is a basis it follows that $c_1 - b_1 = \Sigma \alpha_j b_j$, where not all α_j are zero. Therefore,

$$A(c_1 - b_1) = \Sigma \alpha_j A b_j = \Sigma \alpha_j e_j \neq 0.$$

and this is a contradiction. □

Theorem 2.3.6. *Let $A \in M_n(F)$. If B is a right inverse, $AB = I$, then B is a left inverse.*

Proof. Define $C = BA - I + B$, and assume $C \neq B$ or what is the same thing that B is not a left inverse. Then

$$\begin{aligned} AC &= ABA - A + AB = (AB)(A) - A + AB \\ &= A - A + AB = I \end{aligned}$$

This implies that C is another right inverse of A, contradicting Theorem 2.3.5. □

2.4 Orthogonality

Let V be a vector space over $\mathbb{C}$. We define an inner product $\langle \cdot, \cdot \rangle$ on $V \times V$ to be a function from V to $\mathbb{C}$ that satisfies the following properties:

1. $\langle av, w \rangle = a \langle v, w \rangle$ and $\langle v, aw \rangle = \overline{a} \langle v, w \rangle$ ($\overline{a}$ is the complex conjugate of a)
2. $\langle v, w \rangle = \overline{\langle w, v \rangle}$

3. $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ (linearity)

4. $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$

5. $\langle v, v\rangle \geq 0$ with $\langle v, v\rangle = 0$ if and only if $v = 0$.

For inner products over real vector spaces, we neglect the complex conjugate operation. In addition, we want our inner products to define a *norm* as follows:

6. For any $v \in V$, $\|v\|^2 = \langle v, v\rangle$

We assume thoughout the text that all vector spaces with inner products have norms defined exactly in this way. With the norm and vector v can be *normalized* by dilating it to have length 1, say $v_n = v\frac{1}{\|v\|}$. The simplest type of inner product on $\mathbb{C}_n$ is given by

$$\langle v, w\rangle = \sum_{i=1}^{n} x_i \bar{y}_i$$

We call this the **standard** inner product.

Using any inner product, we can define an angle between vectors.

Definition 2.4.1. The **angle** θ_{xy} between vectors x and y in R_n is defined by

$$\begin{aligned}\cos\theta_{xy} &= \frac{\langle x, y\rangle}{\|x\|\|y\|} \\ &= \frac{\langle x, y\rangle}{(\langle x, x\rangle)^{1/2}(\langle y, y\rangle)^{1/2}}.\end{aligned}$$

This comes from the well known result in R_2

$$x \cdot y = \|x\|\|y\| \cos\theta$$

which can be proved using the law of cosines. With angle comes the notion of orthogonality.

Definition 2.4.2. Two vectors u and v are said to be **orthogonal** if the angle between them if $\frac{\pi}{2}$ or what is the same thing $\langle u, v\rangle = 0$. In this case we commonly write $x \perp y$. We extend this notation to sets U writing $x \perp U$ to mean that $x \perp u$ for every $u \in U$. Similarly two sets U and V are called orthogonal if $u \perp v$ for every $u \in U$ and $v \in V$.

Remark 2.4.1. It is important to note that the notion of orthogonality depends completely on the inner product. For example, the **weighted inner product** defined by $\langle v, w\rangle = \sum_{i=1}^{n} w_i\, x_i \bar{y}_i$ where the $w_i > 0$ gives very different orthogonal vectors from the standard inner product.

Example 2.4.1. In R_n or $\mathbb{C}_n$ the standard unit vectors are orthogonal with respect to the standard inner product.

Example 2.4.2. In the R_3 the vectors $u = (1, 2, -1)$ and $v = (1, 1, 3)$ are orthogonal because

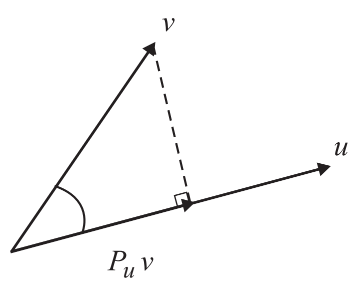
$$\langle x, y\rangle = 1\,(1) + 2\,(2) - 1\,(3) = 0$$

Note that in R_3 the complex conjugate is not written. The set of vectors $(x_1, x_2, x_3) \in R_3$ orthogonal to $u = (1, 2, -1)$ satisfies the equation $x_1 + 2x_2 - x_3 = 0$ is recognizable as the plane with normal vector u.

Definition 2.4.3. We define the *projection* $P_u v$ of one vector v in the direction of an other vector u to be

$$P_u v = \frac{\langle u, v\rangle}{\|u\|^2} u$$

As you can see, we have merely written an expression for the more intuitive version of the projection in question given by $\|v\| \cos\theta_{uv} \frac{u}{\|u\|}$. In the figure below, we show the fundamental diagram for the projection of one vector in the direction of another.



If the vectors u and v are orthogonal, it is easy to see that $P_u v = 0$. (Why?)

Example 2.4.3. Find the projection of the vector $v = (1, 2, 1)$ on the vector $u = (-2, 1, 3)$

Solution. We have

$$\begin{aligned} P_u v &= \frac{\langle u, v\rangle}{\|u\|^2} u = \frac{\langle (1,2,1), (-2,1,3)\rangle}{\|(-2,1,3)\|^2}(-2,1,3) \\ &= \frac{1(-2)+2(1)+1(3)}{14}(-2,1,3) \\ &= \frac{5}{14}(-2,1,3) \end{aligned}$$

We are now ready to find *orthogonal sets* of vectors and orthogonal bases. First we make an important definition.

Definition 2.4.4. Let V be a vector space with an inner product. A set of vectors $S = \{x_1, \ldots, x_n\}$ in V is said to be *orthogonal* if $\langle x_i, x_j\rangle = 0$ for $i \neq j$. It is called **orthonormal** if also $\langle x_i, x_i\rangle = 1$. If, in addition, S is a basis it is called an orthogonal basis or orthonomal basis.

Note: Sometimes the conditions for orthonormality are written as

$$\langle x_i, x_j\rangle = \delta_{ij}$$

where δ_{ij} is the "Dirac" delta: $\delta_{ij} = 0$, $i \neq j$, $\delta_{ii} = 1$.

Theorem 2.4.1. *Suppose U is a subspace of the (inner product) vector space V and that U has the basis $S = \{x_1 \ldots x_k\}$, then U has an orthogonal basis.*

Proof. Define $y_1 = \frac{x_1}{\||x|\|}$. Thus y_1 is the "normalized" x_1. Now define the new orthonormal basis recursively by

$$\begin{aligned} y'_{j+1} &= x_{j+1} - \sum_{i=1}^{j} \langle y_i, x_{j+1}\rangle y_i \\ y_{j+1} &= \frac{y'_{j+1}}{\|y'_{j+1}\|} \end{aligned}$$

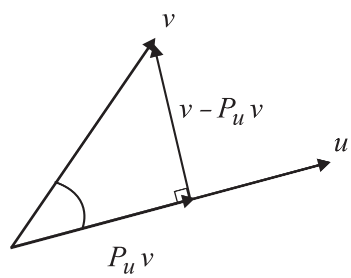
for $j = 1, 2, \ldots, k-1$. Then

(1) y_{j+1} is orthogonal to $y_1, \ldots, y_j$

(2) $y_{j+1} \neq 0$.

In the language above we have $\langle y_i, y_j\rangle = \delta_{ij}$. ☐

Basically, what the proof accomplishes is to take the differences of the vector from the projections to the others. Referring to the figure above we compute $v - P_u v$ as noted in the figure below. The process of orthogonalization described above is called the **Gram–Schmidt** process.



Representation of vectors

One of the great advantages of orthonormal bases is that they make the representation of vectors particularly easy. It is as simple as computing an inner product. Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$ and with subspace U having basis $S = \{u_1, u_2, \ldots, u_k\}$. Then for every $u \in U$ we know there are constants $a_1, a_2, \ldots, a_k$ such that

$$x = a_1 u_1 + a_2 u_2 + \cdots + a_k u_{k.}$$

Taking the inner product of both sides with u_j and applying the orthogonality relations

$$\begin{aligned} \langle x, u_j \rangle &= \langle a_1 u_1 + a_2 u_2 + \cdots + a_k u_{k.}, u_j \rangle \\ &= \sum_{j=1}^{k} a_i \langle u_{i.}, u_j \rangle = a_j \end{aligned}$$

Thus $a_j = \langle x, u_j \rangle$, $j = 1, 2, \ \ldots, \ k$, and

$$x = \sum_{j=1}^{k} \langle u_{.}, u_j \rangle \, u_j$$

Example 2.4.4. One basis of R_2 is given by the orthonormal vectors $S = \{u_1, u_2\}$, where $u_1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T$ and $u_2 = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]^T$. The representation of $x = [3, 2]^T$ is given by

$$x = \sum_{j=1}^{2} \langle u_{.}, u_j \rangle \, u_j = \frac{5}{2}\sqrt{2}\left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^T + \frac{1}{2}\sqrt{2}\left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]^T$$

Orthogonal subspaces

Definition 2.4.5. For any set of vectors S we define

$$S^{\perp} = \{v \in V \mid v \perp S\}$$

That is, $S^{\perp}$ is the set of vectors orthogonal to S. Often, $S^{\perp}$ is called the **orthogonal complement** or **orthocomplement** of S.

For example the orthocomplement of any vector $v = [v_1, v_2, v_3]^T \in R_3$ is the (unique) plane passing through the origin that is orthogonal to v. It is easy to see that the equation of the plane is $x_1 v_1 + x_2 v_2 + x_3 v_3 = 0$.

For any set of vectors S the orthocomplement $S^{\perp}$ has the remarkable property of being a subspace of V, and therefore it is must have an orthogonal basis.

Proposition 2.4.1. *Suppose that V is a vector space with an inner product, and $S \subset V$. Then $S^{\perp}$ is a subspace of V.*

Proof. If $y_1, \ldots, y_m \in S^{\perp}$ then $\Sigma a_i y_i \in S^{\perp}$ for every set of coefficients $a_1, \ldots, a_m$ in R (or $\mathbb{C}$). □

Corollary 2.4.1. *Suppose that V is a vector space with an inner product, and $S \subset V$.*

(i) If S is a basis of V, $S^{\perp} = \{0\}$.

(ii) If $U = \mathfrak{S}(S)$, then $U^{\perp} = S^{\perp}$.

The proofs of these facts are elementary consequences of the proposition. An important decomposition result is based on orthogonality of subspaces. For example, suppose that V is a finite dimensional inner product space and that U is a subspace of V. Let $U^{\perp}$ be the orthocomplement of U, and let $S = \{u_1, u_2, \ldots, u_k\}$ be an orthonormal basis of U. Let $x \in V$. Define $x_1 = \sum_{j=1}^{k} \langle x, u_j \rangle u_j$, and $x_2 = x - x_1$. Then it follows that $x_1 \in U$ and $x_2 \in U^{\perp}$. Moreover, $x = x_1 + x_2$. We summarize this in the following.

Proposition 2.4.2. *Let V is a vector space with inner product $\langle \cdot, \cdot \rangle$ and with subspace U. Then every vector $x \in V$ can be written as a sum of two orthogonal vectors $x = x_1 + x_2$, where $x_1 \in U$ and $x_2 \in U^{\perp}$.*

Geometrically what this results asserts is that for a given subspace of an inner product space, every vector has an orthogonal decomposition as two unique sum of a vector from the subspace and its orthocomplement. We write the vector components as the respective projections of the given vector to the orthogonal subspaces

$$\begin{aligned} x_1 &= P_U\, x \\ x_2 &= P_{U^\perp}\, x \end{aligned}$$

Such decompositions are important in the analysis of vector spaces and matrices. In the case of vector spaces, of course, the representation of vectors is of great value. In the case of matrices, this type of decomposition serves to allow reductions of the matrices while preserving the information they carry.

2.4.1 An important equality for matrix multiplication and the inner product

Let $A \in M_{mn}(\mathbb{C})$. Then we know that both A^*A and AA^* (Alternatively, A^TA and AA^T exist) exist, and we can surely inquire about the rank of these matrices. The main result of this section is on the rank of A^TA, namely that $r\,(A) = r(A^*A) = r(AA^*)$. The proof is quite simple but requires an important equality. Let $A \in M_{mn}(\mathbb{C})$ and $v \in \mathbb{C}_n$ and $w \in \mathbb{C}_m$. Then

$$\begin{aligned} \langle Av, w\rangle &= \left\langle \sum_{i=1}^{m} (Av)_i\,, w_i \right\rangle \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}v_j\ \bar{w}_i \\ &= \sum_{j=1}^{n} v_j\ \sum_{i=1}^{m} a_{ij}\bar{w}_i \\ &= \sum_{j=1}^{n} v_j \overline{\sum_{i=1}^{m} \bar{a}_{ij}w_i} \\ &= \sum_{j=1}^{n} v_j\ \overline{(A^*w)_j} \\ &= \langle v, A^*w\rangle \end{aligned}$$

As a consequence we have $\langle A^*Av, w\rangle = \langle Av, Aw\rangle$ if both v, $w \in \mathbb{C}_n$. This important equality allows the adjoint or transpose matrices to be used on either side of the inner product, as needed. Indeed we shall use this below.

Proposition 2.4.3. *Let $A \in M_{mn}(\mathbb{C})$ have rank $r(A)$. Then*

$$r(A) = r(A^*A) = r(AA^*)$$

Proof. Assume that $r(A) = k$. Then there are k standard vectors $e_{j_1}, \ldots, e_{j_k}$ such that for each $l = 1, 2, \ldots k$, the vectors Ae_{j_l} is one of the linearly independent columns of A. Moreover, it also follows that for every set of constants $a_1, \ldots, a_k$ the vector $A\left(\sum a_l e_{l_j}\right) \neq 0$. Now $A^*A\left(\sum a_l e_{l_j}\right) \neq 0$ follows because

$$\begin{aligned}\left\langle A^*A\left(\sum a_l e_{l_j}\right), \left(\sum a_l e_{l_j}\right)\right\rangle &= \left\langle A\left(\sum a_l e_{l_j}\right), A\left(\sum a_l e_{l_j}\right)\right\rangle \\ &= \left\|A\left(\sum a_l e_{l_j}\right)\right\|^2 \neq 0\end{aligned}$$

This in turn establishes that A^*A cannot be zero on a linear space of dimension k except for the zero element of course, and since the rank of A^*A cannot be larger than k the result is proved. □

Remark 2.4.2. This establishes (6) of the Corollary 2.3.1 above. Also, it is easy to see that the result is also true for real matrices.

2.4.2 The Legendre Polynomials

When a vector space has an inner product, it is possible to construct an orthogonal basis from any given basis. We do this now for the polynomial space $P_n(-1, 1)$ and a particular basis.

Consider the space the polynomials of degree n defined on the interval $[-1, 1]$ over the reals. Recall that this is a vector space and has as a basis the *monomials* $\left\{1, x, x^2, \ldots, x^n\right\}$. We can define an assortment of inner products on this space, but the most common inner product is given by

$$\langle p, q\rangle = \int_{-1}^{1} p(x)\, q(x)\ dx$$

Verifying the inner product properties is fairly straight forward and we leave it as an exercise. This inner product also defines a norm

$$\|p\|^2 = \int_{-1}^{1} |p(x)|^2\ dx$$

This norm satisfies the triangle inequality requires an integral version of the Cauchy-Schwartz inequality.

Now that we have an inner product and norm, we could proceed to find an othogonal basis of $P_n(-1,1)$ by applying the Gram-Schmidt procedure to the basis $\left\{1, x, x^2, \ldots, x^n\right\}$. This procedure can be clumsy and tedious. It is easier to build an orthogonal basis from scratch. Following tradition we will use capital letters $P_0, P_1, \ldots$ to denote our orthogonal polynomials. Toward this end take $P_0 = 1$. Note we are numbering from 0 onwards so that the polynomial degree will agree with the index. Now let $P_1 = ax + b$. For orthogonality, we need

$$\int_{-1}^{1} P_0(x) P_1(x) \ dx = \int_{-1}^{1} 1 \cdot (ax + b) \ dx = 2b = 0$$

Thus $b = 0$ and a can be arbitrary. We take $a = 1$. This gives $y_1 = x$. Now we assume the model for the next orthogonal function to be $y_2 = ax^2 + bx + c$. This time there are two orthogonality conditions to satisfy.

$$\begin{aligned}
\int_{-1}^{1} P_0(x) P_2(x) \ dx &= \int_{-1}^{1} 1 \cdot \left(ax^2 + bx + c\right) \ dx = \frac{2}{3}a + 2c = 0 \\
\int_{-1}^{1} P_1(x) P_2(x) \ dx &= \int_{-1}^{1} x \cdot \left(ax^2 + bx + c\right) \ dx = \frac{2}{3}b = 0
\end{aligned}$$

We conclude that $b = 0$. From the equation $\frac{2}{3}a + 2c = 0$, we can assign one of the variables and solve for the other one. Following tradition we take $c = -\frac{1}{2}$ and solve for a to get $a = \frac{3}{2}$.

The next polynomial will be modeled as $P_3(x) = ax^3 + bx^2 + cx + d$. Three orthogonality relations need to be satisfied.

$$\begin{aligned}
\int_{-1}^{1} P_0(x) P_3(x) \ dx &= \int_{-1}^{1} 1 \cdot \left(ax^3 + bx^2 + cx + d\right) \ dx = \frac{2}{3}b + 2d = 0 \\
\int_{-1}^{1} P_1(x) P_3(x) \ dx &= \int_{-1}^{1} x \cdot \left(ax^3 + bx^2 + cx + d\right) \ dx = \frac{2}{5}a + \frac{2}{3}c = 0 \\
\int_{-1}^{1} P_2(x) P_3(x) \ dx &= \int_{-1}^{1} \frac{1}{2}(3x - 1)\left(ax^3 + bx^2 + cx + d\right) \ dx \\
&= \frac{3}{5}a - \frac{1}{3}b + c - d = 0
\end{aligned}$$

It is easy to see that $b = d = 0$ (why?) and from $\frac{2}{5}a + \frac{2}{3}c = 0$, we select

$c = -\frac{3}{2}$ and $a = \frac{5}{2}$. Our table of orthogonal polynomials so far is

k	$P_k(x)$
0	1
1	x
2	$\frac{1}{2}(3x-1)$
3	$\frac{1}{2}\left(5x^3-3x\right)$

Continue in this fashion, generating polynomials of increasing order each orthogonal to all of the lower order ones.

$$\begin{aligned}
P_0(x) &= 1 \\
P_1(x) &= x \\
P_2(x) &= 3/2\,x^2 - 1/2 \\
P_3,x) &= 5/2\,x^3 - 3/2\,x \\
P_4(x) &= \frac{35}{8}\,x^4 - \frac{15}{4}\,x^2 + 3/8 \\
P_5(x) &= \frac{63}{8}\,x^5 - \frac{35}{4}\,x^3 + \frac{15}{8}\,x \\
P_6(x) &= \frac{231}{16}\,x^6 - \frac{315}{16}\,x^4 + \frac{105}{16}\,x^2 - \frac{5}{16} \\
P_7(x) &= \frac{429}{16}\,x^7 - \frac{693}{16}\,x^5 + \frac{315}{16}\,x^3 - \frac{35}{16}\,x \\
P_8(x) &= \frac{6435}{128}\,x^8 - \frac{3003}{32}\,x^6 + \frac{3465}{64}\,x^4 - \frac{315}{32}\,x^2 + \frac{35}{128} \\
P_9(x) &= \frac{12155}{128}\,x^9 - \frac{6435}{32}\,x^7 + \frac{9009}{64}\,x^5 - \frac{1155}{32}\,x^3 + \frac{315}{128}\,x \\
P_{10}(x) &= \frac{46189}{256}\,x^{10} - \frac{109395}{256}\,x^8 + \frac{45045}{128}\,x^6 - \frac{15015}{128}\,x^4 + \frac{3465}{256}\,x^2 - \frac{63}{256}
\end{aligned}$$

2.4.3 Orthogonal matrices

Besides sets of vectors being orthogonal, there is also a definition of orthogonal matrices. The two notions are closely linked.

Definition 2.4.6. We say a matrix $A \in M_n(\mathbb{C})$ is *orthogonal* if $A^*A = I$. The same definition applies to matrices $A \in M_n(R)$ with A^* replaced by A^T.

For example, the rotation matrices (Exercise **??**) $B_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ are all orthogonal.

A simple consequence of this definition is that the rows and the columns of A are **orthonormal**. We see for example that when A is orthogonal then $(A^*)^2 = (A^{-1})^2 = A^{-1}A^{-1} = (A^2)^{-1}$. Such a definition applies, as well to higher powers. For instance, if A is orthogonal then A^m is orthogonal for every positive integer m.

One way to generate orthogonal matrices in $\mathbb{C}_n$ (or R_n) is to begin with an orthonormal basis and arrange it into an $n\times n$ matrix either as its columns or rows.

Theorem 2.4.2. *(i) Let $\{x_i\}$, $i = 1, \ldots, n$ be an orthonormal basis of $\mathbb{C}_n$ or (R_n). Then the matrices*

$$U = \begin{bmatrix} x_1 & \cdots & x_n \\ \downarrow & \cdots & \downarrow \\ \cdot & & \cdot \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} x_1 & \longrightarrow & \cdot \\ \vdots & \vdots & \\ x_n & \longrightarrow & \cdot \end{bmatrix}$$

formed by arranging the vectors x_i as its respective columns or rows are orthogonal.

(ii) Conversely, U is an orthogonal matrix, the sets of its rows and columns are each orthonormal, and moreover each forms a basis of $\mathbb{C}_n$ or (R_n).

The proofs are entirely trivial. We shall consider these types of results in more detail later in Chapter 4. In the meantime there are a few more interesting results that are direct consequences of the definition and facts about the transpose (adjoint).

Theorem 2.4.3. *Let $A, B \in M_n(C)$ (or $M_n(R)$) be orthogonal matrices. Then*
(a) A is invertible and $A^{-1} = A^*$.
(b) For each integer $k = 0, \pm 1, \pm 2, \ldots$, both A^k and $-A^k$ are orthogonal.
(c) AB is orthogonal.

2.5 Determinants

This section is about determinants that can be regarded as a measure of singularity of a matrix. More generally, in many applied situations that deal with complex objects, a single number is sought that will in some way classify an aspect of those objects. The determinant is such a measure for singularity of the matrix. The determinant is difficult to calculate and of not much practical use. However, it has considerable theoretical value and certainly has a place of historical interest.

Definition 2.5.1. Let $A \in M_n(F)$. Define the **determinant** of A to be the value in F

$$\det A = \sum_{\sigma} \left(\prod_{i=1}^{n} a_{i\sigma(i)} \right) \cdot \operatorname{sgn} \sigma$$

where σ is a permutation of the integers $\{1, 2, \ldots, n\}$ and

(1) $\sum\limits_{\sigma}$ denotes the sum over all permutations

(2) $\operatorname{sgn} \sigma = \text{sign of } \sigma = \pm 1$

A **transposition** is the exchange of two elements of an ordered list with all others staying the same. With respect to permutations, a transposition of one permutation is another permutation formed by the exchange of two values. For example a transposition of $\{1, 4, 3, 2\}$ is $\{1, 3, 4, 2\}$. The **sign** of a given permutation σ is

(a) $+1$, if the number of transpositions required to bring σ to $\{1, 2, \ldots, n\}$ is even.

(b) -1, if the number of transpositions required to bring σ to $\{1, 2, \ldots, n\}$ is odd.

Alternatively, and what is the same thing, we may count the number m of transpositions required to bring σ to $\{1, 2, \ldots, n\}$ and to compute the sign is $(-1)^m$.

Example 2.5.1.

$$\begin{array}{lll}
\sigma_1 = \{2, 1, 3\} \xrightarrow{1 \leftrightarrow 2} \{1, 2, 3\} & & \text{odd} \\
\sigma_2 = \{2, 3, 1\} \xrightarrow{3 \leftrightarrow 1} \{2, 1, 3\} \xrightarrow{1 \leftrightarrow 2} \{1, 2, 3\} & & \text{even} \\
\operatorname{sgn} \sigma_1 = -1 \qquad \operatorname{sgn} \sigma_2 = +1 & &
\end{array}$$

Proposition 2.5.1. *Let $A \in M_n$.*

(i) If two rows of A are interchanged to obtain B, then

$$\det B = -\det A.$$

(ii) Given $A \in M_n(F)$. If any row is multiplied by a scalar c, the resulting matrix B has determinant

$$\det B = c \det A.$$

(iii) If any two rows of $A \in M_n(F)$ are equal,

$$\det A = 0.$$

Proof. (i) Suppose rows i_1 and i_2 are interchanged. Now for the given permutations σ apply the transposition $i_1 \leftrightarrow i_2$ to get σ_1. Then

$$\prod_{i=1}^{n} a_{i_1\sigma(i)} = \prod_{i=1}^{n} b_{i_2\sigma_1(i)}$$

because

$$a_{i_1\sigma(i_1)} = b_{i_2\sigma_1(i_2)}$$

as

$$b_{i_2 j} = a_{i_1 j} \quad \text{and} \quad \sigma_1(i_2) = \sigma_2(i_1)$$

and similarly $a_{i_2\sigma(i_2)} = b_{i_1\sigma_1(i_1)}$. All other terms are equal. In the computation of the full determinant with signs of the permutations, we see that the change is caused only by the fact $\text{sgn}(\sigma_1) = -\text{sgn}(\sigma)$. Thus,

$$\det B = -\det A.$$

(ii) Is trivial.

(iii) If two rows are equal then by part (i)

$$\det A = -\det A$$

and this implies $\det A = 0$.

∎

Corollary 2.5.1. *Let $A \in M_n$. If A has two rows equal up to a multiplicative constant, it has has determinant zero.*

What happens to the determinant when two matrices are added. The result is too complicated to write down is not very important. However, when a single vector is added to a row or a column of a matrix, then the result can be simply stated.

Proposition 2.5.2. *Suppose $A \in M_n(F)$. Suppose B is obtained from A by adding a vector v to a given row (resp. column) and C is obtained from A by replacing the given row (resp. column) by the vector v. Then*

$$\det B = \det A + \det C.$$

Proof. Assume the j^{th} row is altered. Using the definition of the determinant,

$$\begin{aligned}
\det A &= \sum_\sigma sgn\,(\sigma) \prod_i b_{i\sigma(i)} = \sum_\sigma sgn\,(\sigma) \left(\prod_{i \neq j} b_{i\sigma(i)} \right) b_{j\sigma(j)} \\
&= \sum_\sigma sgn\,(\sigma) \left(\prod_{i \neq j} a_{i\sigma(i)} \right) (a+v)_{j\sigma(j)} \\
&= \sum_\sigma sgn\,(\sigma) \left(\prod_{i \neq j} a_{i\sigma(i)} \right) a_{j\sigma(j)} + \sum_\sigma sgn\,(\sigma) \left(\prod_{i \neq j} a_{i\sigma(i)} \right) v_{j\sigma(j)} \\
&\quad \det A + \det C
\end{aligned}$$

For column replacement the proof is similar, particularly using the alternate representation of the determinant given in Exercise 15. □

Corollary 2.5.2. *Suppose $A \in M_n(F)$ and B is obtained by multiplying a given row (resp. column) of A by a scalar and adding it to another row (resp. column), then*

$$\det B = \det A.$$

Proof. First note that in applying Proposition 2.5.2 C has two rows equal up to a multiplicative constant. Thus $\det C = 0$. □

Computing determinants is usually difficult and many techniques have been devised to compute them out. As is evident from counting, computing the determinant of an $n \times n$ matrix using the definition above would require the expression of all $n!$ permutations of the integers $\{1, 2, \ldots, n\}$ and the determination of their signs together with all the concommitant products and summation. This method is prohibitively costly. Using elementary row operations and Gaussian elimination, the evaluation of the determinant becomes more manageable. First we need the result below.

Theorem 2.5.1. *For the elementary matrices the following results hold.*

(a) for Type 1 (row interchange) E_1

$$\det E_1 = -1$$

(b) for Type 2 (multiply a row by a constant c) E_2

$$\det E_2 = c$$

(c) for Type 3 (add a multiple of one row to another row) E_3

$$\det E_3 = 1.$$

Note that (c) is a consequence of Corollary 2.5.2. Proof of parts (a) and (b) are left as exercises. Thus for any matrix $A \in M_n(F)$ we have

$$\begin{aligned}
\det(E_1 A) &= -\det A &&= \det E_1 \det A \\
\det(E_2 A) &= c \det A &&= \det E_2 \det A \\
\det(E_3 A) &= \det A &&= \det E_3 \det A.
\end{aligned}$$

Suppose $F_1 \ldots F_k$ is a sequence of row operations to reduce A to its RREF. Then

$$F_k F_{k-1} \ldots F_1 A = B.$$

Now we see that

$$\begin{aligned}
\det B &= \det(F_k F_{k-1} \ldots F_1 A) \\
&= \det(F_k) \det(F_{k-1} \ldots F_1 A) \\
&= \quad \vdots \\
&= \det(F_k) \det(F_{k-1}) \ldots \det(F_1) \det A.
\end{aligned}$$

For B in RREF *and* $B \in M_n(F)$, we have that B is upper triangular. The next result establishes Theorem 2.3.4(k) about the determinant of singular and non singular matrices. Moreover, the determinant of triangular matrices is computed simply as the product of its diagonal elements.

Proposition 2.5.3. *Let $A \in M_n$. Then*

(i) If $r(A) < n$, then $\det A = 0$.

(ii) If A is triangular then

$$\det A = \prod a_{ii}$$

(iii) If $r(A) = n$, then $\det A \neq 0$.

Proof. (i) If $r(A) < n$, then its RREF has a row of zeros, and $\det A = 0$ by Theorem 2.5.1. (ii) If A is triangular the only product without possible zero entries is $\prod a_{ii}$. Hence $\det A = \prod a_{ii}$. (iii) If If $r(A) = n$, then its RREF has no nonzero rows. Since it is square and has a leading one in each column, it follows that the RREF is the identity matrix. Therefore $\det A \neq 0$. ☐

Now let $A, B \in M_n$. If A is singular the RREF must have a zero row. It follows that $\det A = 0$. If A is singular it follows that AB is singular. Therefore

$$0 = \det AB = \det A \det B.$$

The same reasoning applies if B is singular. If A and B are not singular both A and B can be row reduced to the identity. Let $F_1 \dots F_{k_1}$ be the row operations that reduce A to I, and $G_1 \dots G_{k_B}$ be the row operations that reduce B to I. Then

$$\begin{aligned} \det A &= [\det(F_1) \dots \det(F_{k_A})]^{-1} \\ \det B &= [\det(G_1) \dots \det(G_{k_B})]^{-1}. \end{aligned}$$

Also

$$I = (G_{k_B} \dots G_1)(F_{k_B} \dots F_1)AB$$

and we have

$$\det I = (\det A)^{-1}(\det B)^{-1} \det AB.$$

This proves the

Theorem 2.5.2. *If $A, B \in M_n(F)$,* $\det AB = \det A \det B$.

2.5.1 Minors and Determinants

The method of row reduction is one of the simplest methods to compute the determinant of a matrix. Indeed, it is not necessary to use Type 2 elementary transformation. This results in the computing the determinant as the product of the diagonal elements of the resulting triangular matrix possibly multiplied by a minus sign. An alternate approach to computing determinants using *minors* is both interesting and useful. However, unless the matrix has some special form, it does not provide a computational alternative to row reduction.

Definition 2.5.2. Let $A \in M_n(C)$. For any row i and column j define the (ij)-*minor* of A by

$$M_{ij} = \det A\Big|_{\substack{i^{\text{th}} \text{ row removed} \\ j^{\text{th}} \text{ column removed}}}$$

The notation

$$A\Big|_{\substack{i^{\text{th}} \text{ row removed} \\ j^{\text{th}} \text{ column removed}}}$$

denotes the $(n-1)\times(n-1)$ matrix formed from A by removing the i^{th} row and j^{th} column. With minors an alternative formulation of the determinant can be given. This method, while not of great value computationally, has some theoretical importance. For example, the inverse of a matrix can be expressed using minors. We begin by consideration the determinant.

Theorem 2.5.3. *Let $A \in M_n(C)$. (i) Fix any row, say row k. The determinant of A is given by*

$$\det A = \sum_{j=1}^{n} a_{kj} \;\; (-1)^{k+j} M_{kj}$$

(ii) Fix any column, say column m. The determinant of A is given by

$$\det A = \sum_{j=1}^{n} a_{jm} \;\; (-1)^{m+j} M_{jm}$$

Proof. (i) Suppose that $k = 1$. Consider the quantity

$$a_{11}M_{11}$$

We observe that this is equivalent to all the products of the form

$$sgn\,(\sigma)\,a_{11}\cdot a_{2\sigma(2)}\cdot\cdots\cdot a_{n\sigma(n)}$$

where only permutations that fix the integer (i.e. position) 1 are taken. Thus $\sigma(1)=1$. Since this position is fixed the signs taken in the determinant M_{11} for permutations of $n-1$ integers are respectively the same as the signs for the new permutation of n integers.

Now consider all permutations that fix the integer 2 in the sense that $\sigma(1)=2$. The quantity $a_{12}\,(-1)^{1+2}\,M_{12}$ consists of all the products of the form

$$sgn\,(\sigma)\,a_{12}\cdot a_{2\sigma(1)}a_{3\sigma(3)}\cdot\cdots\cdot a_{n\sigma(n)}$$

We need here the extra sign change because if the part of the permutation σ of the integers $\{1,3,4,\ldots,n\}$ is of one sign, which is the sign used in the computation of $\det M_{ij}$, then the permutation of σ of the integers $\{1,2,3,4,\ldots,n\}$ is of the other sign, and that sign is $sgn\,(\sigma)$.

When we proceed to the k^{th} component, we consider permutations that fix the integer k. That is, $\sigma(1)=k$. In this case the quantity $a_{1k}\,(-1)^{1+k}\,M_{1k}$ consists of all products of the form

$$sgn\,(\sigma)\,a_{1k}a_{2\sigma(1)}\cdot\cdots\cdot a_{k-1\sigma(k-1)}a_{k+1\sigma(k+1)}\cdot\cdots\cdot a_{n\sigma(n)}$$

Continuing in this way we exhaust all possible products $a_{1\sigma(1)}\cdot a_{2\sigma(2)}\cdot\cdots\cdot a_{n\sigma(n)}$ over all possible permutations of the integers $\{1,2,\ldots,n\}$. This proves the assertion. The proof for expanding from any row is similar, with only a possible change of sign needed, which is a prescribed.

(ii) The proof is similar. □

Example 2.5.2. Find the determinant of

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

expanding across the first row and then expanding down the second column.

Solution. Expanding across the first row gives

$$\begin{aligned}\det A &= a_{11}M_{11}-a_{12}M_{12}+a_{13}M_{12}\\ &= 3\det\begin{bmatrix}1&3\\2&-1\end{bmatrix}-2\det\begin{bmatrix}0&3\\1&-1\end{bmatrix}-1\det\begin{bmatrix}0&1\\1&2\end{bmatrix}\\ &= 3\,(-7)-2\,(-3)-(-1)\\ &= -14\end{aligned}$$

Expanding across the second column gives

$$\begin{aligned}
\det A &= -2\det\begin{bmatrix} 0 & 3 \\ 1 & -1 \end{bmatrix} + 1\det\begin{bmatrix} 3 & -1 \\ 1 & -1 \end{bmatrix} - 2\det\begin{bmatrix} 3 & -1 \\ 0 & 3 \end{bmatrix} \\
&= -2\,(-3) + (-2) - 2\,(9) = -14
\end{aligned}$$

The inverse of the matrix can be formulated in terms of minors, which is formulated below.

Definition 2.5.3. Let $A \in M_n(\mathbb{C})$ (or $M_n(R)$). Define the **adjugate** (or **adjoint**) matrix $\hat{A}$ by

$$\hat{A}_{ij} = (-1)^{i+j} M_{ji}$$

where M_{ji} is the ji minor.

The adjugate has traditionally been called the "adjoint", but that terminology is somewhat ambiguous in light of the previous definition as complex conjugate transpose. Note that it is defined for all square matrices; when restricted to invertible matrices the inverse appears.

Theorem 2.5.4. *Let $A \in M_n(\mathbb{C})$ (or $M_n(R)$) be invertible. Then $A^{-1} = \frac{1}{\det A}\hat{A}$*

Proof. A quick examination of the ij-entry of the product $A\hat{A}$ yields the following sum

$$\sum_{j=1}^{n} a_{ij}\hat{A}_{jk} = \frac{1}{\det(A)}\sum_{j=1}^{n} a_{ij}\,(-1)^{k+j} M_{kj}$$

There are two possibilities. (1) If $i = k$, then the summation above is the summation to form the determinant as described in Theorem 2.5.3. (2) If $i \neq k$, the summation is the computation of the determinant of the matrix A with the k^{th} row replaced by the i^{th} row. Thus the determinant of a matrix with two identical rows is represented above and this must be zero. We conclude that $\left(A\hat{A}\right)_{ij} = \delta_{ij}$, the usual Kronecker 'delta,' and the result is proved. □

Example 2.5.1. Find the adjugate and inverse of

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

It is easy to see that

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

Also $\det A = -6$. Therefore, the inverse

$$A^{-1} = -\frac{1}{6}\begin{bmatrix} 1 & -4 \\ -2 & 2 \end{bmatrix}$$

Remark 2.5.1. The notation for cofactors of a square matrix A is often used

$$\hat{a}_{ij} = (-1)^{i+j} M_{ji}$$

Note the reversed order of the subscripts ij and then ji above.

Cramer's Rule

We know now that the solution to the system $Ax = b$ is given by $x = A^{-1}b$. Moreover, the inverse A^{-1} is given by $A^{-1} = \dfrac{\hat{A}}{\det A}$, where $\hat{A}$ is the adjugate matrix. The the i^{th} component of the solution vector is therefore

$$\begin{aligned} x_i &= \frac{1}{\det A}\sum_{j=1}^{n} \hat{a}_{ij} b_j \\ &= \frac{1}{\det A}\sum_{j=1}^{n} (-1)^{i+j} M_{ji} b_j \\ &= \frac{\det A_i}{\det A} \end{aligned}$$

where we define the matrix A_i to be the modification to A by replacing its i^{th} column by the vector b. In this way we obtain a very compact formula for the solution of a linear system. Called Cramer's rule we state this conclusion as

Theorem 2.5.1. *(Cramer's Rule.) Let $A \in M_n(\mathbb{C})$ be invertible and $b \in C_n$. For each $i = 1, , n$, define the matrix A_i to be the modification of A by replacing its i^{th} column by the vector b. Then the solution to the linear system $Ax = b$ is given by components $x_i = \dfrac{\det A_i}{\det A}$, $i = 1, , n$.*

Example 2.5.2. Given the matrix $A = \begin{bmatrix} 2 & 4 \\ 2 & 1 \end{bmatrix}$, and the vector $b = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Solve the system $Ax = b$ by Cramer's rule.

We have

$$A_1 = \begin{bmatrix} 2 & 4 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix}$$

and $\det A_1 = 6$, $\det A_2 = -6$, $\det A = -6$. Therefore

$$x_1 = -1 \text{ and } x_2 = 1$$

A curious formula

The useful formula using cofactors given below will have some consequence when we study positive definite operators in Chapter **??**.

Proposition 2.5.1. *Consider the matrix*

$$B = \begin{bmatrix} 0 & x_1 & x_2 & \cdots & x_n \\ x_1 & a_{11} & a_{12} & \cdots & a_{1n} \\ x_2 & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & a_{n1} & a_{n2} & & a_{nn} \end{bmatrix}$$

Then

$$\det B = -\sum \hat{a}_{ij} x_i x_j$$

where $\hat{a}_{ij}$ is the ij-cofactor of A.

Proof. Expand by minors along the top row to get

$$\det B = \sum (-1)^j x_j M_{1j}(B)$$

Now expand the matrix of $M_{1j}(B)$ down the first column. This gives

$$M_{1j}(B) = \sum (-1)^{i-1} x_i M_{ij}(A)$$

Combining we obtain

$$\begin{aligned}
\det B &= \sum (-1)^j \, x_j M_{1j}\,(B) \\
&= \sum (-1)^j \, x_j \left[\sum (-1)^{i-1} \, x_i M_{ij}\,(A)\right] \\
&= \sum \sum (-1)^{i+j-1} \, x_j x_i M_{ij}\,(A) \\
&= -\sum \sum \hat{a}_{ij} x_j x_i
\end{aligned}$$

□

The reader may note that in the last line of the equation above, we should have used $\hat{a}_{ij}$. However, the formulation given is correct, as well. (Why?)

2.6 Partitioned Matrices

It is convenient to study partitioned or "blocked" matrices, or more graphically said, matrices whose entries are themselves matrices. For example, with I_2 denoting the 2×2 identity matrix we can create the 4×4 matrix written in partitioned form and expanded form.

$$A = \begin{bmatrix} aI_2 & cI_2 \\ cI_2 & dI_2 \end{bmatrix} = \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{bmatrix}$$

Partitioning matrices allows our attention to focus on certain structural properties. In many applications partititioned matrices appear in a natural way, with the particular blocks having some system context. Many similar subclasses and processes apply to partitioned matrices. In specific situations they can be added, multiplied, and inverted, just like regular matrices. It is even possible to perform "blocked" version of Gaussian elimination. In the few results here, we touch on some of these possibilities.

Definition 2.6.1. For each $1 \leq i \leq m$ and $1 \leq j \leq n$, let A_{ij} be an $m_i \times n_j$ matrices where . Then the matrix

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

is a partitioned matrix of order $(\sum m)_i \times (\sum n_j)$.

The usual operations of addition and multiplication of partitioned matrices can be performed provided each of the operations makes sense. For addition of two partitioned matrices A and B it is necessary to have the same numbers of blocks of the respective same sizes. Then

$$\begin{aligned} A+B &= \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{2n} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & B_{2n} & \cdots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \cdots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{2n}+B_{2n} & \cdots & A_{2n}+B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \cdots & A_{mn}+B_{mn} \end{bmatrix} \end{aligned}$$

For multiplication, the situation is a bit more complicated. For definiteness, suppose that B is a partitioned matrix with block sizes $s_i \times t_j$, where $1 \le i \le p$ and $1 \le j \le q$ The usual operations to construct $C = AB$,

$$\sum_{j=1}^{n} A_{ij} B_{jk}$$

then make sense provided $p = n$ and $n_j = s_j$, $1 \le j \le n$.

A special category of partitioned matrices are the so-called **quasi-triangular** matrices, wherein $A_{ij} = 0$ if $i > j$ for the "lower" triangular version. The special subclass of quasi-triangular matrices wherein $A_{ij} = 0$ if $i \ne j$ are called **quasi-diagonal.** In the case of the multiplication of partitioned matrices ($C = AB$) with the left multiplicand A a quasi-diagonal matrix, we have $C_{ik} = A_{ii} B_{ik}$. Thus the multiplication is similar in form to the usual multiplication of matrices where the left multiplicand is a diagonal matrix. In the case of the multiplication of partitioned matrices ($C = AB$) with the right multiplicand B a quasi-diagonal matrix, we have $C_{ik} = A_{ik} B_{kk}$. For quasi-triangular matrices with square diagonal blocks, there is an interesting result about the determinant.

Theorem 2.6.1. *Let A be a quasi-triangular matrix, where the diagonal blocks A_{ii} are square. Then*

$$\det A = \prod_i \det A_{ii}$$

Proof. Apply row operations on each vertical block without row interchanges between blocks, without any Type 2 operations. The resulting matrix in each diagonal block position (i, i) is triangular. Be sure to multiply one of the diagonal entries by ± 1, reflecting the number of row interchanges within a block. The resulting matrix can still be regarded as partitioned, though the diagonal blocks are now actually upper triangular. Now apply Proposition 2.5.3, noting that the product of each of the diagonal entries pertaining to the i^{th} block is in fact $\det A_{ii}$. □

A simple consequence of this result, proved alá Gaussian elimination, is contained in the following corollary.

Corollary 2.6.1. *Consider the partitioned matrix*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with square diagonal blocks and with A_{11} invertible. Then the rank of A is the same as the rank of A_{11} if and only if $A_{22} = A_{21}A_{11}^{-1}A_{12}$.

Proof. Multiplication of A by the elementary partitioned matrix

$$E = \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}$$

yields

$$\begin{aligned} EA &= \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \\ &= \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \end{aligned}$$

Since E has full rank, it follows that $\operatorname{rank}(EA) = \operatorname{rank} A$. Since EA is quasi-triangular, it follows that the rank of A is the same as the rank of A_{11} if and only if $A_{22} - A_{21}A_{11}^{-1}A_{12} = 0$. □

2.7 Linear Transformations

Definition 2.7.1. A mapping T from R_n to R_m is called a **linear transformation** if

$$\begin{aligned} T(x+y) &= Tx + Ty \qquad && \forall x, y \in R_n \\ T(ax) &= aTx && \forall a \in F. \end{aligned}$$

Note: We normally write Tx instead of $T(x)$.

Example 2.7.1. $T\colon\ R_n \to R_m$. Let $a \in R_m$ and $y \in R_n$. Then for each $x \in R_n$, $Tx = \langle x, y\rangle a$ is a linear transformation. Let $S = \{v_1 \dots v_n\}$ be a basis of R_n, and define the $m \times n$ matrix with columns given by the coordinates of $Tv_1, Tv_2, \dots, Tv_n$. Then this matrix

$$A = \begin{bmatrix} Tv_1 & Tv_2 & & Tv_n \\ \downarrow & \downarrow & \iddots & \downarrow \end{bmatrix}$$

is the matrix representation of T with respect to the basis S. Thus, if $x = \Sigma a_i v_i$, whence $[x]_S = (a_1 \dots a_n)$, we have

$$[Tx]_S = A[x]_S$$

- ⋆ There is a duality between all linear transformations from R_n to R_m and the set $M_{m,n}(F)$.

Note that $M_{m,n}(F)$ is itself a vector space over F. Hence $\mathcal{L}(F_n, F_m)$, the set of linear transformations from F_n to F_m is likewise. As such it has subspaces.

Example 2.7.2. (1) Let $\bar{x} \in F_n$. Definite $\mathcal{J} = \{T \in \mathcal{L} \mid T\bar{x} = 0\}$. Then $\mathcal{J}$ is a subspace of $\mathcal{L}(F_n, F_m)$.

(2) Let $\mathcal{U} = \{T \in \mathcal{L}(R_n, R_n) \mid Tx \geq 0 \text{ if } x \geq 0\}$, where $\{x \geq 0\}$ means the positive orthant of R_n. $\mathcal{U}$ is *not* a linear subspace of $\mathcal{L}(R_n, R_n)$, though it is a convex set.

(3) Define $T\colon\ P_n \to P_n$ by $Tp = \dfrac{d}{dx}p$. T is a linear transformation.

Example 2.7.3. Express the linear transformation $D\colon\ P_3 \to P_3$ given by $Dp = \frac{d}{dx}p(x)$ as a matrix with respect to the basis. $S = \{1, x, x^2, x^3\}$. We have $D1 = 0 = 0 + 0x + 0x^2 + 0x^3$. Also

$$[D1]_S = [0, 0, 0, 0]^T$$

similarly

$$\begin{aligned} [Dx]_S &= [1, 0, 0, 0]^T \\ [Dx^2]_S &= [0, 2, 0, 0]^T \\ [Dx^3]_S &= [0, 0, 3, 0]^T. \end{aligned}$$

Hence

$$[D]_S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In this context the differentiation operator is rather simple.

Example 2.7.4. Consider the linear transformation T defined by $Tq = 3x\frac{d}{dx}q + x^2q$ for $q \in P_2$. Find the matrix representation of T.

Solution. First off we notice that this transformation has range in P_4. Let's use the standard bases for this problem. We then determine the coordinates of T for vectors in the P_2 basis $\{1, x, x^2\}$in the P_4 basis $\{1, x, x^2, x^3, x^4\}$. Compute

$$\begin{aligned} T(1) &= x^2 \\ T(x) &= 3x + x^3 \\ T\left(x^2\right) &= 6x^2 + x^4 \end{aligned}$$

The coordinates of the input vectors we know are $[1, 0, 0]^T$, $[0, 1, 0]^T$, and $[0, 0, 1]^T$. For the output vectors the coordinates are $[0, 0, 1, 0, 0]^T$, $[0, 3, 0, 1, 0]^T$, and $[0, 0, 6, 0, 1]^T$. So, with respect to these two bases, the matrix of the transformation is

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & 0 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Observe that the dimensionality corresponds with the dimentionality of the respective spaces.

Example 2.7.5. Let $V = R_2$, with $S_0 = \{v_1, v_2\} = \{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]\}$, $S_1 = \{w_1, w_2\} = \left\{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right], \left[\begin{smallmatrix}-2\\1\end{smallmatrix}\right]\right\}$, and $T = I$ the identity. The vectors above are expressed in the standard $E = \{e_1, e_2\}$, $Tv_j = Iv_j = v_j$. To find $[v_j]_{S_1}$ we solve

$$v_j = a_j w_1 + \beta_j w_2$$

$$v_1\colon \quad \begin{bmatrix}1 & -2\\ 2 & 1\end{bmatrix}\begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix} = \begin{bmatrix}1\\ 0\end{bmatrix} \longrightarrow \begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix} = \begin{bmatrix}\frac{1}{5}\\ -\frac{2}{5}\end{bmatrix}$$

$\nwarrow$ solve linear
$\swarrow$ systems

$$v_2\colon \quad \begin{bmatrix}1 & -2\\ 2 & 1\end{bmatrix}\begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix} = \begin{bmatrix}1\\ 1\end{bmatrix} \longrightarrow \begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix} = \begin{bmatrix}\frac{3}{5}\\ -\frac{1}{5}\end{bmatrix}$$

Therefore

$$_{S_1}[I]_{S_0} = \begin{bmatrix}\frac{1}{5} & \frac{3}{5}\\ -\frac{2}{5} & -\frac{1}{5}\end{bmatrix} \quad \leftarrow$$

change of
basis
matrix

If

$$[x]_{S_0} = \begin{bmatrix}-1\\ 2\end{bmatrix} \qquad [x]_{S_1} = {}_{S_1}[I]_{S_0}\begin{bmatrix}-1\\ 2\end{bmatrix}$$
$$= \frac{1}{5}\begin{bmatrix}1 & 3\\ -2 & -1\end{bmatrix}\begin{bmatrix}-1\\ 2\end{bmatrix} = \frac{1}{5}\begin{bmatrix}5\\ 0\end{bmatrix} = \begin{bmatrix}1\\ 0\end{bmatrix}.$$

Note the necessity of using the standard basis to express the vectors in both bases S_0 and S_1.

2.8 Change of Basis

Let V be a vector space with bases $S_0 = \{v_1 \dots v_n\}$ and $S_1 = \{w_1 \dots w_n\}$, and suppose $T\colon\ V \to\ V$ is a linear transformation. We want to find the representation of T as a matrix that takes a vector x given in terms of its S_0 coordinates and produces the vector Tx given in terms of its S_1 coordinates.

We know that $x \to [x]_{S_0}$ is well defined. The action of T is known if the n vector $[x]_{S_0} = \begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$ and the vectors $Tv_1, Tv_2, \dots, Tv_n$ are known, for if $x = \Sigma c_j v_j$, then $Tx = \Sigma c_j Tv_j$, by linearity.

To determine $[Tx]_{S_1}$ we need to convert the $Tv_j,\ j = 1, \dots, n$ to coordinates in the other S_1 basis, This is done as follows. Find

$$[Tv_j]_{S_1} = \begin{bmatrix}t_{1j}\\ t_{2j}\\ \vdots\\ t_{nj}\end{bmatrix} \qquad j = 1, 2, \dots, n.$$

Then if $x \in V$

$$
\begin{aligned}
[Tx]_{S_1} &= [\Sigma c_j T v_j]_{S_1} = \Sigma c_j [T v_j]_{S_1} \\
&= \left[\sum_j t_{ij} c_j \right] \\
&= \begin{bmatrix} t_{11} & \dots & t_{1n} \\ & & \\ t_{n1} & & t_{nn} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.
\end{aligned}
$$

This $n \times n$ array $[t_{ij}]$ depends on T, S_0 and S_1 but not on x. We define the $S_0 \to S_1$ basis representation of T to be $[t_{ij}]$, and we write this as

$$
{}_{S_1}[T]_{S_0} = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nn} \end{bmatrix}.
$$

In the special case that T is the identity operator the matrix ${}_{S_1}[I]_{S_0}$ converts the coordinates of a vector in the basis S_0 to coordinates in the basis S_1. It is easy to see that ${}_{S_0}[I]_{S_1}$ must be the inverse of ${}_{S_1}[I]_{S_0}$ and thus

$$
{}_{S_0}[I]_{S_1} \cdot {}_{S_1}[I]_{S_0} = I.
$$

We can also establish the equality

$$
{}_{S_1}[T]_{S_1} = {}_{S_1}[I]_{S_0}\,{}_{S_0}[T]_{S_0}\,{}_{S_0}[I]_{S_1}.
$$

In this way we see that the matrix representation of T depends on the bases involved. If X is any invertible matrix in $M_n(F)$ we can write

$$
B = X^{-1} A X.
$$

The interpretation in this context is clear

$$
\begin{aligned}
X&\colon \text{ change of coordinate from one basis to another } S_0 \to S_1 \\
X^{-1}&\colon \text{ change of coordinate } S_1 \to S_0 \\
A&\colon \text{ matrix of the linear transformation in the basis } S_0 \\
B&\colon \text{ matrix of the same linear transformation in the basis } S_1.
\end{aligned}
$$

With this in mind it seems prudent to study linear transformations in the basis that makes their matrix representation as simple as possible.

Example 2.8.1. Let $A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}$ be the matrix representation of a linear transformation given with respect to the standard basis $S_0 = \{e_1,\ e_2\} = \{(1,0)\,,(0,1)\}$ Find the matrix representation of this transformation with resepect to the basis $S_1 = \{v_1, v_2\} = \{(2,1)\,,(1,1)\}$.
Solution. According to the analysis above we need to determine ${}_{S_0}[I]_{S_1}$ and ${}_{S_1}[I]_{S_0}$. Of course ${}_{S_1}[I]_{S_0} = {}_{S_0}[I]_{S_1}^{-1}$. Since the coordinates of the vectors in S_1 are expressed in terms of the basis vectors S_0 we obtain directly

$$
{}_{S_0}[I]_{S_1} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}
$$

Its inverse is given by

$$
{}_{S_0}[I]_{S_1}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}
$$

Assembling these matrices we have the final matrix converted to the new basis.

$$
\begin{aligned}
{}_{S_1}[A]_{S_1} &= {}_{S_1}[I]_{S_0}\ A\ {}_{S_0}[I]_{S_1} \\
&= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 6 & 4 \\ -7 & -4 \end{bmatrix}
\end{aligned}
$$

Example 2.8.2. Consider the same problem as above except that the matrix A is given in the basis S_1. Find matrix representation of this transformation with resepect to the basis S_0.
Solution. To solve this problem we need to determine ${}_{S_0}[A]_{S_0} = {}_{S_0}[I]_{S_1}\ A\ {}_{S_1}[I]_{S_0}$. As we already have these matrices, we determine that

$$
\begin{aligned}
{}_{S_0}[A]_{S_0} &= {}_{S_0}[I]_{S_1}\ A\ {}_{S_1}[I]_{S_0}\,. \\
&= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\
&= \begin{bmatrix} -6 & 13 \\ -4 & 8 \end{bmatrix}
\end{aligned}
$$

2.9 Appendix A — Solving linear systems

The key to solving linear systems is to reduce the augmented system to RREF and solve the resulting equations. While this may be so, there is an intermediate step that occurs about half way through the computation of the RREF where the reduced matrix achieves an upper triangular form. At this point the solution can be determined directly by back substitution. To clarify the rules on back substitution, suppose that we have the triangular form

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{nn} & b_n \end{array}\right]$$

Assuming that the diagonal part consists of all nonzero terms, we can solve this system by back substitution. First solve for $x_n = \frac{b_n}{a_{nn}}$. Now inductively solve for the remaining solution coordinates using the formula

$$x_{n-j} = \frac{1}{b_{n-j,n-j}} \left[\sum_{k=0}^{j-1} a_{n-j,n-k} x_{n-k} \right], \qquad j = 1, 2, \ldots, \ n-1$$

This inconvenient looking formula can be replaced by

$$x_j = \frac{1}{b_{jj}} \left[\sum_{k\ =\ j+1}^{n} a_{jk} x_k \right], \qquad j = n-1, \ n-2, \ldots, \ 1$$

where the index runs from $j = n-1$ up to $j = 1$. The upshot is that the row reduction process can be halted when a triangular-like form has been attained. The applies as well to nonsingular and non square systems, where the the process is stopped when all the leading ones have been identified, entries *below* them have been zeroed out, and all the zero rows are present. The principle reason for using back substitution is to reduce the number of computations required, an important consideration in numerical linear algebra. In the example below we solve a 3×3 nonsingular system.

Example 2.9.1. Solve $Ax = b$ where

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

Solution. Find the RREF of $[A \mid b]$. Then solve $Ax = b$.

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 2 & 2 & -1 & 6 \\ -1 & 3 & 2 & -2 \end{array}\right] \begin{array}{c} -2R_1 + R_2 \\ \rightarrow \\ R_1 + R_3 \end{array} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & -2 & -1 & 0 \\ 0 & 5 & 2 & 1 \end{array}\right]$$

$$\begin{array}{c} -\frac{1}{2}R_2 \\ \rightarrow \end{array} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 5 & 2 & 1 \end{array}\right]$$

$$\begin{array}{c} -5R_2 + R_3 \\ \rightarrow \end{array} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} & 1 \end{array}\right] \qquad (*)$$

$$\begin{array}{c} -\frac{1}{2}R_3 + R_2 \\ \rightarrow \end{array} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

$$\begin{array}{c} -2R_3 \\ \rightarrow \end{array} \left[\begin{array}{ccc|c} 1 & 2 & 0 & 3 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

$$\begin{array}{c} -2R_2 + R_1 \\ \rightarrow \end{array} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

Hence solving we obtain $x_3 = -2,\quad x_2 = 1,$ and $x_1 = 1$. This is fine, but there is a faster way to solve this system. Stop the reduction when the system attains a triangular form at $(*)$. From this point solve to obtain $x_3 = -2$. Now back substitute $x_3 = 2$ into the second row (equation) to solve for x_2. Thus $x_2 = -\frac{1}{2}(-2) = 1$. Finally, back substitute $x_3 = 2$ and $x_2 = 1$ into the first row (equation) to solve for x_1. Thus $x_1 = 3 - 2(1) = 1$. Sometimes the form $(*)$ is called the row reduced form.

Example 2.9.2. Given the augmented system for $Ax = b$ is in RREF.

$$\left[\begin{array}{cccccc|c} 1 & 2 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Find the solution.

Solution. The leading ones occur in columns 1, 3, and 5. The values in columns 2, 4, and 6 can be taken as free parameters. So, take $x_2 = r,\ x_4 = s$, and $x_6 = t$. Now solving for the other varables we have

$$\begin{aligned} x_1 &= 4 - 2r \\ x_3 &= 1 - 2s + t \\ x_5 &= 1 - 3t \end{aligned}$$

The solution set is comprised of the vector

$$\begin{aligned} x &= [4 - 2r, r, 1 - 2s + t,\ s, 1 - t, t]^T \\ &= [4, 0, 1, 0, 1, 0]^T + r\,[-2, 1, 0, 0, 0, 0]^T + s\,[0, 0, -2, 1, 0, 0]^T + t\,[0, 0, 1, 0 - 3, 1]^T \end{aligned}$$

for all $r,\ s,$ and t. We can rewrite this as the set

$$S = \left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + r \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -3 \\ 1 \end{bmatrix} \;\middle|\; r, s, t \in R \text{ or } \mathbb{C} \right\}$$

This representation shows better the connection between the free constants and the component vectors that make up the solution. Note this expression also reveals the solution of the homogeneous solution $Ax = 0$ as the set

$$\left\{ r\,[-2, 1, 0, 0, 0, 0]^T + s\,[0, 0, -2, 1, 0, 0]^T + t\,[0, 0, 1, 0 - 3, 1]^T \;\middle|\; r, s, t \in R \text{ or } \mathbb{C} \right\}$$

Indeed, this is a full subspace.

Example 2.9.3. The RREF can be used to determine the inverse, as well. Given the matrix $A \in M_n$, the inverse is given by the matrix X for which $AX = I$. In turn with $x_1, \ \ldots, \ x_n$ representing the columns of X and $e_1, \ \ldots, \ e_n$ representing the standard vectors we see that $Ax_j = e_j,\ j = 1, 2, \ldots, n$. To solve for these vectors, form the augmented matrix $[A \mid e_j],\ j = 1, 2, \ldots,\ n$ and row reduce as above. A massive short cut to this process is to augment all the standard vectors at one and row reduce the resulting $n \times 2n$ matrix $[A \mid I]$. If A is invertible, its RREF is the identity. Therefore,

$$[A \mid I] \quad \underset{\text{operations}}{\overset{\text{row}}{\rightarrow}} \quad [I \mid X]$$

and, of course, $A^{-1} = X$. Thus, for

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

we row reduce $[A \mid I]$ as follows

$$\begin{aligned} [A \mid I] \quad &= \left[\begin{array}{ccc|ccc} -2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 3 & -2 & -1 & 0 & 0 & 1 \end{array}\right] \\ &\underset{\text{operations}}{\overset{\text{row}}{\rightarrow}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & 1 & 2 \\ 0 & 1 & 0 & 7 & 2 & 4 \\ 0 & 0 & 1 & -5 & -1 & -3 \end{array}\right] \end{aligned}$$

2.10 Exercises

1. Consider the differential operator $T = 2x\frac{d}{dx}(\cdot) - 4$ acting on the vector space of cubic polynomials, P_3. Show that T is a linear transformation and find a matrix representation of it. Assume the basis is given by $\{1, x, x^2, x^3\}$.

2. (i) Find matrices A and B, each with positive rank, for which $r(A + B) = r(A) + r(B)$. (ii) Find matrices A and B, each with positive rank, for which $r(A+B) = 0$. (iii) Give a method to find two nonzero matrices A and B for which the sum has any preassigned rank. Of course, the matrix sizes may depend on this value.

3. Find square matrices A and B for which $r(A) = r(B) = 2$ and for which $r(AB) = 0$.

4. Suppose that A is an $m \times n$ matrix and that x is a solution of $Ax = b$ over the prescribed field. Show that every solution of $Ax = b$ have the form $x + x_0$, where x_0 is a solution of $Ax_0 = 0$.

5. Find a matrix $A \in M_n$ of rank $n-1$ for which $r(A^k) = n-k$ for $k \leq n$. Is it possible to begin this process with a matrix $A \in M_n$ of rank n and for which $r(A^k) = n-k+1$ for $k \leq n$?

6. Show that $Ax = b$ has a solution if and only if $y^T b = 0$ if and only if $y^T A = 0$ for some column vector.

7. In R_2 the linear transformation that rotates any vector by θ radians counter clockwise T has matrix representation with respect to the standard basis given by

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

   What is the matrix representation with respect to the standard basis of the transformation that rotates any vector by θ radians clockwise? What is the relation between the matrices?

8. Show that if $B, C \in M_n(F)$, where B is symmetric and C is skew-symmetric, then $B = C$ implies that $B = C = 0$.

9. Prove the general formula for the inverse of the 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $A^{-1} = \dfrac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

10. Prove Theorem 2.5.1(a).

11. Prove Theorem 2.5.1(b).

12. Prove that every permutation σ must have an inverse σ^{-1} (i.e. $\sigma^{-1}(\sigma(j)) = j$), and the signs of σ^{-1} and σ are the same.

13. Show that the sign of every transposition is -1.

14. Prove that $\det A = \sum_\sigma (-1)^{sgn(\sigma)} \left(\prod_i a_{\sigma(i)\ i} \right)$

15. Prove Proposition 2.5.2 using minors.

16. Suppose that the $n \times n$ matrix A is singular. Show that each column of the adjugate matrix $\hat{A}$ is a solution of $Ax = 0$. (McDuffee, Chapter 3, Theorem 29.)

17. Suppose that A is an $(n-1)\times n$ matrix, and consider the homogeneous system $Ax = 0$ for $x \in R_n$. Define h_i to be the determinant of the $(n-1)\times(n-1)$ matrix formed by removing the i^{th} column of A. Show that the vector $h = (h_1, \ldots, h_n)^T$ is a solution to $Ax = 0$. (McDuffee, Chapter 3, Corollary 29.)

18. Show that $A = \left[\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right]$ has no inverse by trying to solve $AB = I$ That is, assume the form
$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
multiply the matrices (A and B) together, and then solve for the unknowns a, b, c, and d. (This is not a very efficient way to determine inverses of matrices. Try the same thing for any 3×3 matrix.)

19. Prove that the elementary equation operations do not change the solution set of a linear system.

20. Find the inverses of E_1, E_2, and E_3.

21. Find the matrix representation of linear transformation T that rotates any vector by θ radians counter clockwise (ccw) with respect to the basis $S = \{(2,1), (1,1)\}$.

22. Consider R_3. Suppose that we have angles $\{\theta_i\}_{i=1}^3$ and pairs of coordinate vectors $\{(e_1, e_2), (e_1, e_3), (e_2, e_3)\}$. Let T be the linear transformation that successively rotates a vector in the respective planes $\{(e_{i_1}, e_{i_2})\}_{i=1}^k$ through the respective angles $\{\theta_i\}_{i=1}^k$. Find the matrix representation of T with respect to the standard basis. Prove that it is invertible.

23. Prove Theorem 2.2.4.

24. Prove or disprove the equivalence of the linear systems.
$$\begin{aligned} 2x - 3y &= -1 \\ x + 4y &= 5 \end{aligned} \qquad\qquad \begin{aligned} -x + 4y &= 3 \\ x + 2y &= 3 \end{aligned}$$

25. Find basis for the orthocomplement of the subspace of R_3 spanned by the vectors $\{[2,1,1]^T,\ [1,1,2]^T\}$.

26. Find basis for the orthocomplement of the subspace of R_3 spanned by the vector $[1,1,1]^T$.

27. Consider planar rotations in R_n with respect to the standard bases elements. Prove that there must be $\frac{n(n-1)}{2}$ of them — discounting the particular angle. Display the general representation of any of them. Prove or disprove that any two of them are commutative. That is for two angles $\{\theta_i\}_{i=1}^{2}$ and pairs of coordinate vectors $\{(e_{i_1}, e_{i_2})\}_{i=1}^{k}$ the respective counter clockwise rotations are commutative.

28. Suppose that $A \in M_{mk}$, $B \in M_{kn}$ and both have rank k. Show that the rank of AB is k.

29. Suppose that $A \in M_{mk}$ has rank k. Prove that $A_{\text{RREF}} = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$ where I_k is the identity matrix of size k and 0 is the $m-k \times k$ zero matrix.

30. Suppose that $B \in M_{kn}$ has rank k. Prove that $B_{\text{RREF}} = \begin{bmatrix} I_k & 0 \end{bmatrix}$ where I_k is the identity matrix of size k and 0 is the$k$ $\times n-k$ zero matrix.

31. Determine and prove a version of Corollary 2.6.1 for 3×3 blocked matrices, where we assume the diagonal blocks A_{11} is invertible and wish to conclude the result that the rank of A is the equal to the rank of A_{11}.

32. Suppose that we have angles $\{\theta_i\}_{i=1}^{k}$ and pairs of coordinate vectors $\{(e_{i_1}, e_{i_2})\}_{i=1}^{k}$. Let T be the linear transformation that successively rotates a vector in the respective planes $\{(e_{i_1}, e_{i_2})\}_{i=1}^{k}$ through the respective angles $\{\theta_i\}_{i=1}^{k}$. Prove that the matrix representation of the linear transformation with respect to any basis must be invertible.

33. The super-diagonal of a matrix is the set of elements $a_{i,i+1}$. The subdiagonal of a matrix is the set of elements $a_{i-1,i}$. A tri-banded matrix is one for which the entries are zero above the super-diagonal and below the subdiagonal. Suppose that for an $n \times n$ tri-banded matrix T, we have $a_{i-1,i} = a$, $a_{ii} = 0$, and $a_{i,i+1} = c$. Prove the following facts:

    (a) If n is odd $\det A = 0$.

    (b) If $n = 2m$ is even $\det T = (-1)^m a^m c^m$.

34. For the banded matrix of the previous example, prove the following for the powers T^p of T.

(a) If p is odd, prove that $(T^p)_{ij} = 0$ if $i + j$ is even.

(b) If p is even, prove that $(T^p)_{ij} = 0$ if $i + j$ is odd.

35. Consider the vector space $P_2(1,2)$ with inner product defined by $\langle p, q\rangle = \int_1^2 p(x)\, q(x)\, dx$. Find an orthogonal basis of $P_2(1,2)$. (Hint. Begin with the standard basis $\{1, x, x^2\}$. Apply the Gram-Schmidt procedure.)

36. For what values of a and b is the matrix below singular

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

37. The Vandermonde matrix, defined for a sequence of numbers $\{x_1, \ldots x_n\}$, is given by the $n \times n$ matrix

$$V_n = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

Prove that the determinant is given by

$$\det V_n = \prod_{i>j=1}^{n} (x_i - x_j)$$

38. In the case the x-values are the integers $\{1, \ldots, n\}$, prove that $\det V_n$ is divisible by $\prod\limits_{i=1}^{n} (i-1)!$. (These numbers are called superfactorials.)

39. Prove that for the weighted functional defined in Remark 2.4.1, it is necessary and sufficient that the weights be strictly positive for it to be an inner product.

40. For what values of a and b is the matrix below singular

$$A = \begin{bmatrix} b & 0 & a & 0 \\ 0 & 0 & b & a \\ 0 & a & 0 & b \\ b & a & 0 & 0 \end{bmatrix}$$

41. Find an orthogonal basis for R_2 from the vectors $\{(1,2),(2,1)\}$.

42. Find an orthogonal basis of the subspace of R_3 spanned by $\{(1,0,1),(0,1,-1)\}$

43. Suppose that V is a vector space with an inner product, and $S \subset V$. Show that if S is a basis of V, $S^{\perp} = \{0\}$.

44. Suppose that V is a vector space with an inner product, and $S \subset V$. Show that if $U = \mathfrak{S}(S)$, then $U^{\perp} = S^{\perp}$.

45. Let $A \in M_{mn}(F)$. Show it may not be true that $r(A) = r(A^T A) = r(AA^T)$ unless $F = R$, in which case it is true.

46. If $A \in M_n(C)$ is orthogonal, show that the rows and columns of A are orthogonal.

47. If A is orthogonal then A^m is orthogonal for every positive integer m. (This is a part of Theorem 2.4.3(b).)

48. Consider the polynomial space $P_n[-1,1]$ with the inner product $\langle p,q\rangle = \int_{-1}^{1} p(t)\, q(t)\ dt$. Show that every polynomial $p \in P_n$ for which $p(1) = p(-1) = 0$ is orthogonal to its derivative.

49. Consider the polynomial space $P_n[-1,1]$ with the inner product $\langle p,q\rangle = \int_{-1}^{1} p(t)\, q(t)\ dt$. Show that the subspace of polynomials in even powers (e.g. $p(t) = t^2 - 5t^6$) is orthogonal to the subspace of polynomials in odd powers.

50. Let $A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}$ be the matrix representation of a linear transformation given with respect to the standard basis $S_0 = \{e_1,\ e_2\} = \{(1,0),(0,1)\}$ Find the matrix representation of this transformation with resepect to the basis $S_1 = \{v_1, v_2\} = \{(2,-3),(1,-2)\}$.

51. Show that the sign of every transposition from the set $\{1, 2, \ldots, n\}$ is -1.

52. What are the signs of the permutations $\{7,\ 6,\ 5,\ 4,\ 3,\ 2,\ 1\}$ and $\{7,1,\ 6,\ 4,\ 3,\ 5,\ 2\}$ of the integers $\{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7\}$?

53. Prove that the sign of the permutation $\{m,\ m-1, \ldots,\ 2,\ 1\}$ is $(-1)^m$.

54. Suppose $A, B \in M_n(F)$. If A is singular, use a row space argument to show that $\det AB = 0$.

55. Show that if $A \in M_{m,n}$ and $B \in M_{m,n}$ and both $r(A) = r(B) = m$. If $r(AB) = m - k$, what can be said about n?

56. Prove that

$$\det \begin{vmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & -x_2 & x_1 \end{vmatrix} = \left(x_1^2 + x_2^2 + x_3^2 + x_4^2\right)^2$$

57. Prove that there is no invertible 3×3 matrix that has all the same cofactors. What similar statement can be made for $n \times n$ matrices?

58. Show by example that there are matrices A and B for which $\lim_{n\to\infty} A^n$ and $\lim_{n\to\infty} B^n$ both exist, but for which $\lim_{n\to\infty} (AB)^n$ does not exist.

59. Let $A \in M_2(\mathbb{C})$. Show that there is no matrix solution $B \in M_2(\mathbb{C})$ to $AB - BA = I$. What can you say about the same problem with $A, B \in M_n(\mathbb{C})$?

60. Show by example that if $AC = BC$ then it does not follow that $A = B$. However, show that if C is inveritble the conclusion $A = B$ is valid.