Summary of Fraleigh & Beauregard's: LINEAR ALGEBRA

Martin Thøgersen, University of Aarhus
(Based on notes by Bob Gardner*)

26th March 2006

Contents

*http://www.etsu.edu/math/gardner/2010/notes.htm

1 Vectors, Matrices, and Linear Spaces

1.1 Vectors in Euclidean Spaces

Definition. The space $\mathbb{R}^n$, or *Euclidean n-space*, is either **(1)** the collection of all n-tuples of the form $(x_1, x_2, \ldots, x_n)$ where the x_i's are real numbers (the n-tuples are called *points*), or **(2)** the collection of all n-tuples of the form $[x_1, x_2, \ldots, x_n]$ where the x_i's are real numbers (the n-tuples are called *vectors*).

Definition. For $\vec{x} \in \mathbb{R}^n$, say $\vec{x} = [x_1, x_2, \ldots, x_n]$, the *ith component* of $\vec{x}$ is x_i.

Definition. Two vectors in $\mathbb{R}^n$, $\vec{v} = [v_1, v_2, \ldots, v_n]$ and $\vec{w} = [w_1, w_2, \ldots, w_n]$ are *equal* if each of their components are equal. The *zero vector*, $\vec{0}$, in $\mathbb{R}^n$ is the vector of all zero components.

Definition 1.1. Let $\vec{v} = [v_1, v_2, \ldots, v_n]$ and $\vec{w} = [w_1, w_2, \ldots, w_n]$ be vectors in $\mathbb{R}^n$ and let $r \in \mathbb{R}$ be a scalar. Define
1. *Vector addition*: $\vec{v} + \vec{w} = [v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n]$,
2. *Vector subtraction*: $\vec{v} - \vec{w} = [v_1 - w_1, v_2 - w_2, \ldots, v_n - w_n]$, and
3. *Scalar multiplication*: $r\vec{v} = [rv_1, rv_2, \ldots, rv_n]$.

Theorem 1.1. Properties of Vector Algebra in $\mathbb{R}^n$.
Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and let r, s be scalars in $\mathbb{R}$. Then
A1. Associativity of Vector Addition. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
A2. Commutivity of Vector Addition. $\vec{v} + \vec{w} = \vec{w} + \vec{v}$
A3. Additive Identity. $\vec{0} + \vec{v} = \vec{v}$
A4. Additive Inverses. $\vec{v} + (-\vec{v}) = \vec{0}$
S1. Distribution of Scalar Multiplication over Vector Addition.
$$r(\vec{v} + \vec{w}) = r\vec{v} + r\vec{w}$$
S2. Distribution of Scalar Addition over Scalar Multiplication.
$$(r + s)\vec{v} = r\vec{v} + s\vec{v}$$
S3. Associativity. $r(s\vec{v}) = (rs)\vec{v}$
S4. "Preservation of Scale." $1\vec{v} = \vec{v}$

Definition 1.2. Two nonzero vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ are *parallel*, denoted $\vec{v} \parallel \vec{w}$, if one is a scalar multiple of the other. If $\vec{v} = r\vec{w}$ with $r > 0$, then $\vec{v}$ and $\vec{w}$ have the *same direction* and if $\vec{v} = r\vec{w}$ with $r < 0$ then $\vec{v}$ and $\vec{w}$ have *opposite directions.*

Definition 1.3. Given vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{R}^n$ and scalars $r_1, r_2, \ldots, r_k \in \mathbb{R}$, the vector

$$r_1\vec{v_1} + r_2\vec{v_2} + \cdots + r_k\vec{v_k} = \sum_{l=1}^{k} r_l\vec{v_l}$$

is a *linear combination* of the given vectors with the given scalars as *scalar coefficients.*

Definition. The *standard basis vectors in* $\mathbb{R}^2$ are $\hat{\imath} = [1, 0]$ and $\hat{\jmath} = [0, 1]$. The *standard basis vectors in* $\mathbb{R}^3$ are

$$\hat{\imath} = [1, 0, 0], \hat{\jmath} = [0, 1, 0], \text{ and } \hat{k} = [0, 0, 1].$$

Definition. In $\mathbb{R}^n$, the *rth standard basis vector*, denoted $\hat{e_r}$, is

$$\hat{e_r} = [0, 0, \ldots, 0, 1, 0, \ldots, 0],$$

where the rth component is 1 and all other components are 0.

Notice. A vector $\vec{b} \in \mathbb{R}^n$ can be **uniquely** expressed in terms of the standard basis vectors:

$$\vec{b} = [b_1, b_2, \ldots, b_n] = b_1\hat{e_1} + b_2\hat{e_2} + \cdots + b_n\hat{e_n} = \sum_{l=1}^{n} b_l\hat{e_l}.$$

Definition. If $\vec{v} \in \mathbb{R}^n$ is a nonzero vector, then the *line along* $\vec{v}$ is the collection of all vectors of the form $r\vec{v}$ for some scalar $r \in \mathbb{R}$ (notice $\vec{0}$ is on all such lines). For two nonzero nonparallel vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$, the collection of all possible linear combinations of these vectors: $r\vec{v} + s\vec{w}$ where $r, s \in \mathbb{R}$, is the *plane spanned by* $\vec{v}$ *and* $\vec{w}$.

Definition. A *column vector* in $\mathbb{R}^n$ is a representation of a vector as

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

A *row vector* in $\mathbb{R}^n$ is a representation of a vector as

$$\vec{x} = [x_1, x_2, \ldots, x_n].$$

The *transpose* of a row vector, denoted $\vec{x}^T$, is a column vector, and conversely:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T = [x_1, x_2, \ldots, x_n], \text{ and } [x_1, x_2, \ldots, x_n]^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Note. A linear combination of column vectors can easily be translated into a system of linear equations:

$$r\begin{bmatrix} 1 \\ 3 \end{bmatrix} + s\begin{bmatrix} -2 \\ 5 \end{bmatrix} = \begin{bmatrix} -1 \\ 19 \end{bmatrix} \Longleftrightarrow \begin{array}{rcl} r - 2s & = & -1 \\ 3r + 5s & = & 19 \end{array}.$$

Definition 1.4. Let $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{R}^n$. The *span* of these vectors is the **set** of all linear combinations of them, denoted $\operatorname{sp}(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k})$:

$$\begin{aligned} \operatorname{sp}(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}) &= \{r_1\vec{v_1} + r_2\vec{v_2} + \cdots + r_k\vec{v_k} \mid r_1, r_2, \ldots, r_k \in \mathbb{R}\} \\ &= \left\{ \sum_{l=1}^{k} r_l\vec{v_l} \,\middle|\, r_1, r_2, \ldots, r_k \in \mathbb{R} \right\}. \end{aligned}$$

1.2 The Norm and Dot Product

Definition 1.5. Let $\vec{v} = [v_1, v_2, \ldots, v_n] \in \mathbb{R}^n$. The *norm* or *magnitude* of $\vec{v}$ is

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \sqrt{\sum_{l=1}^{n} (v_l)^2}.$$

Theorem 1.2. Properties of the Norm in $\mathbb{R}^n$.
For all $\vec{v}, \vec{w} \in \mathbb{R}^n$ and for all scalars $r \in \mathbb{R}$, we have:
1. $\|\vec{v}\| \geq 0$ and $\|\vec{v}\| = 0$ if and only if $\vec{v} = \vec{0}$.
2. $\|r\vec{v}\| = |r|\|\vec{v}\|$.
3. $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$ (the Triangle Inequality).

Note. **1** and **2** are easy to see and we will prove **3** later in this section.

Definition. A vector with norm 1 is called a *unit vector*. When writing, unit vectors are frequently denoted with a "hat": $\hat{\imath}$.

Definition 1.6. The *dot product* for $\vec{v} = [v_1, v_2, \ldots, v_n]$ and $\vec{w} = [w_1, w_2, \ldots, w_n]$ is

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{l=1}^{n} v_l w_l.$$

Notice. If we let θ be the angle between nonzero vectors $\vec{v}$ and $\vec{w}$, then we get by the Law of Cosines:

$$\|\vec{v}\|^2 + \|\vec{w}\|^2 = \|\vec{v} - \vec{w}\|^2 + 2\|\vec{v}\|\|\vec{w}\| \cos\theta$$

or

$$2\vec{v} \cdot \vec{w} = 2\|\vec{v}\|\|\vec{w}\| \cos\theta$$

or

$$\cos\theta = \frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}. \qquad (*)$$

Definition. The *angle* between nonzero vectors $\vec{v}$ and $\vec{w}$ is

$$\arccos\left(\frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}\right).$$

Theorem 1.4. Schwarz's Inequality.
Let $\vec{v}, \vec{w} \in \mathbb{R}^n$. Then

$$|\vec{v} \cdot \vec{w}| \leq \|\vec{v}\|\|\vec{w}\|.$$

Proof. This follows from $(*)$ and the fact that $-1 \leq \cos\theta \leq 1$. The book gives an algebraic proof. *QED*

Theorem 1.3. Properties of Dot Products.
Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and let $r \in \mathbb{R}$ be a scalar. Then
D1. Commutivity of $\cdot$: $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$.
D2. Distribution of $\cdot$ over vector Addition: $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$.
D3. $r(\vec{v} \cdot \vec{w}) = (r\vec{v}) \cdot \vec{w} = \vec{v} \cdot (r\vec{w})$.
D4. $\vec{v} \cdot \vec{v} \geq 0$ and $\vec{v} \cdot \vec{v} = 0$ if and only if $\vec{v} = \vec{0}$.

Definition. Two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ are *perpendicular* or *orthogonal*, denoted $\vec{v} \perp \vec{w}$, if $\vec{v} \cdot \vec{w} = 0$.

Theorem 1.5. The Triangle Inequality.
Let $\vec{v}, \vec{w} \in \mathbb{R}^n$. Then $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$.

Proof.

$$\begin{aligned}
\|\vec{v} + \vec{w}\|^2 &= (\vec{v} + \vec{w}) \cdot (\vec{v} + \vec{w}) \\
&= \vec{v} \cdot \vec{v} + 2\vec{v} \cdot \vec{w} + \vec{w} \cdot \vec{w} \\
&\leq \|\vec{v}\|^2 + 2\|\vec{v}\|\|\vec{w}\| + \|\vec{w}\|^2 \text{ by Schwarz Inequality} \\
&= (\|\vec{v}\| + \|\vec{w}\|)^2.
\end{aligned}$$

QED

1.3 Matrices and Their Algebra

Definition. A *matrix* is a rectangluar array of numbers. An $m \times n$ matrix is a matrix with m *rows* and n *columns*:

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Definition 1.8. Let $A = [a_{ik}]$ be an $m \times n$ matrix and let $B = [b_{kj}]$ be an $n \times s$ matrix. The *matrix product* AB is the $m \times s$ matrix $C = [c_{ij}]$ where c_{ij} is the dot product of the ith row vector of A and the jth column vector of B: $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$.

Definition. The *main diagonal* of an $n \times n$ matrix is the set $\{a_{11}, a_{22}, \ldots, a_{nn}\}$. A square matrix which has zeros off the main diagonal is a *diagonal matrix*. We denote the $n \times n$ diagonal matrix with all diagonal entires 1 as $\mathcal{I}$ (the $n \times n$ *identity matrix*).

Definition 1.9/1.10. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matices. The *sum* $A + B$ is the $m \times n$ matrix $C = [c_{ij}]$ where $c_{ij} = a_{ij} + b_{ij}$. Let r be a scalar. Then rA is the matrix $D = [d_{ij}]$ where $d_{ij} = ra_{ij}$.

Definition 1.11. Matrix B is the *transpose* of A, denoted $B = A^T$, if $b_{ij} = a_{ji}$. If A is a matrix such that $A = A^T$ then A is *symmetric.*

Example. If A is square, then $A + A^T$ is symmetric.

Note. Properties of Matrix Algebra.
Let A, B be $m \times n$ matrices and r, s scalars. Then
Commutative Law of Addition: $A + B = B + A$
Associative Law of Addition: $(A + B) + C = A + (B + C)$
Additive Identity: $A + 0 = 0 + A = A$ (here "0" represents the $m \times n$ matrix of all zeros)
Left Distribution Law: $r(A + B) = rA + rB$
Right Distribution Law: $(r + s)A = rA + sA$
Associative Law of Scalar Multiplication: $(rs)A = r(sA)$
Scalars "Pull Through": $(rA)B = A(rB) = r(AB)$
Associativity of Matrix Multiplication: $A(BC) = (AB)C$
Matrix Multiplicative Identity: $\mathcal{I}A = A = A\mathcal{I}$

Distributive Laws of Matrix Multiplication: $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$.

Note. Properties of the Transpose Operator.

$$(A^T)^T = A \qquad (A + B)^T = A^T + B^T \qquad (AB)^T = B^T A^T.$$

1.4 Solving Systems of Linear Equations

Definition. A *system of* m *linear equations* in the n unknowns $x_1, x_2, \ldots, x_n$ is a system of the form:

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

Note. The above system can be written as $A\vec{x} = \vec{b}$ where A is the *coefficient matrix* and $\vec{x}$ is the vector of variables. A *solution* to the system is a vector $\vec{s}$ such that $A\vec{s} = \vec{b}$.

Defninition. The *augmented matrix* for the above system is

$$[A \mid \vec{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

Note. We will perform certain operations on the augmented matrix which correspond to the following manipulations of the system of equations:
1. interchange two equations,
2. multiply an equation by a nonzero constant,
3. replace an equation by the sum of itself and a multiple of another equation.

Definition. The following are *elementary row operations*:
1. interchange row i and row j (denoted $R_i \leftrightarrow R_j$),
2. multiplying the ith row by a nonzero scalar s (denoted $R_i \to sR_i$), and
3. adding the ith row to s times the jth row (denoted $R_i \to R_i + sR_j$).
If matrix A can be obtained from matrix B by a series of elementary row operations, then A is *row equivalent* to B, denoted $A \sim B$ or $A \to B$.

Notice. These operations correspond to the above manipulations of the equations and so:

Theorem 1.6. Invariance of Solution Sets Under Row Equivalence. If $[A \mid \vec{b}] \sim [H \mid \vec{c}]$ then the linear systems $A\vec{x} = \vec{b}$ and $H\vec{x} = \vec{c}$ have the same solution sets.

Definition 1.12. A matrix is in *row-echelon form* if
(1) all rows containing only zeros appear below rows with nonzero entries, and
(2) the first nonzero entry in any row appears in a column to the right of the first nonzero entry in any preceeding row.
For such a matrix, the first nonzero entry in a row is the *pivot* for that row.

Note. If an augmented matrix is in row-echelon form, we can use the method of *back substituton* to find solutions.

Definition 1.13. A linear system having no solution is *inconsistent*. If it has one or more solutions, it is *consistent*.

Note. Reducing a Matrix to Row-Echelon Form.
(1) If the first column is all zeros, "mentally cross it off." Repeat this process as necessary.
(2a) Use row interchange if necessary to get a nonzero entry (pivot) p in the top row of the remaining matrix.
(2b) For each row R below the row containing this entry p, add $-r/p$ times the row containing p to R where r is the entry of row R in the column which contains pivot p. (This gives all zero entries below pivot p.)
(3) "Mentally cross off" the first row and first column to create a smaller matrix. Repeat the process **(1)** - **(3)** until either no rows or no columns remain.

Note. The above method is called *Gauss reduction with back substitution.*

Note. The system $A\vec{x} = \vec{b}$ is equivalent to the system

$$x_1\vec{a_1} + x_2\vec{a_2} + \cdots + x_n\vec{a_n} = \vec{b}$$

where $\vec{a_i}$ is the ith column matrix of A. Therefore, $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b}$ is in the span of $\vec{a_1}, \vec{a_2}, \ldots, \vec{a_n}$ (the columns of A).

Definition. A matrix is in *reduced row-echelon form* if all the pivots are 1 and all entries above or below pivots are 0.

Note. The above method is the *Gauss-Jordan method.*

Theorem 1.7. Solutions of $A\vec{x} = \vec{b}$.
Let $A\vec{x} = \vec{b}$ be a linear system and let $[A \mid \vec{b}] \sim [H \mid \vec{c}]$ where H is in row-echelon form.
(1) The system $A\vec{x} = \vec{b}$ is inconsistent if and only if $[H \mid \vec{c}]$ has a row with all entries equal to 0 to the left of the partition and a nonzero entry to the right of the partition.
(2) If $A\vec{x} = \vec{b}$ is consistent and every column of H contains a pivot, the system has a unique solution.
(3) If $A\vec{x} = \vec{b}$ is consistent and some column of H has no pivot, the system has infinitely many solutions, with as many free variables as there are pivot-free columns of H.

Definition 1.14. A matrix that can be obtained from an identity matrix by means of **one** elementary row operation is an *elementary matrix.*

Theorem 1.8. Let A be an $m \times n$ matrix and let E be an $m \times m$ elementary matrix. Multiplication of A **on the left** by E effects the same elementary row operation on A that was performed on the identity matrix to obtain E.

Proof for Row-Interchange. (This is page 71 number 52.) Suppose E results from interchanging rows i and j:

$$\mathcal{I} \overset{R_i \leftrightarrow R_j}{\longrightarrow} E.$$

Then the kth row of E is $[0, 0, \ldots, 0, 1, 0, \ldots, 0]$ where
(1) for $k \notin \{i, j\}$ the nonzero entry if the kth entry,
(2) for $k = i$ the nonzero entry is the jth entry, and
(3) for $k = j$ the nonzero entry is the ith entry.
Let $A = [a_{ij}]$, $E = [e_{ij}]$, and $B = [b_{ij}] = EA$. The kth row of B is $[b_{k1}, b_{k2}, \ldots, b_{kn}]$ and

$$b_{kl} = \sum_{p=1}^{n} e_{kp} a_{pl}.$$

Now if $k \notin \{i, j\}$ then all e_{kp} are 0 except for $p = k$ and

$$b_{kl} = \sum_{p=1}^{n} e_{kp} a_{pl} = e_{kk} a_{kl} = (1) a_{kl} = a_{kl}.$$

Therefore for $k \notin \{i, j\}$, the kth row of B is the same as the kth row of A.

If $k = i$ then all e_{kp} are 0 except for $p = j$ and

$$b_{kl} = b_{il} = \sum_{p=1}^{n} e_{kp}a_{pl} = e_{kj}a_{jl} = (1)a_{jl} = a_{jl}$$

and the ith row of B is the same as the jth row of A. Similarly, if $k = j$ then all e_{kp} are 0 except for $p = i$ and

$$b_{kl} = b_{jl} = \sum_{p=1}^{n} e_{kp}a_{pl} = e_{ki}a_{il} = (1)a_{il} = a_{il}$$

and the jth row of B is the same as the ith row of A. Therefore

$$B = EA \stackrel{R_i \leftrightarrow R_j}{\longrightarrow} A.$$

QED

Note. If A is row equivalent to B, then we can find C such that $CA = B$ and C is a product of elementary matrices.

1.5 Inverses of Square Matrices

Definition 1.15. An $n \times n$ matrix A is *invertible* if there exists an $n \times n$ matrix C such that $AC = CA = \mathcal{I}$. If A is not invertible, it is *singular.*

Theorem 1.9. Uniqueness of an Inverse Matrix.
An invertible matrix has a unique inverse (which we denote A^{-1}).

Proof. Suppose C and D are both inverses of A. Then $(DA)C = \mathcal{I}\mathcal{C} = \mathcal{C}$ and $D(AC) = D\mathcal{I} = \mathcal{D}$. But $(DA)C = D(AC)$ (associativity), so $C = D$. *QED*

Theorem 1.10. Inverses of Products.
Let A and B be invertible $n \times n$ matrices. Then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. By associativity and the assumption that A^{-1} and B^{-1} exist, we have:

$$(AB)(B^{-1}A^{-1}) = [A(BB^{-1})]A^{-1} = (A\mathcal{I})A^{-1} = AA^{-1} = \mathcal{I}.$$

We can similarly show that $(B^{-1}A^{-1})(AB) = \mathcal{I}$. Therefore AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. *QED*

Lemma 1.1. Condition for $A\vec{x} = \vec{b}$ to be Solvable for $\vec{b}$.
Let A be an $n \times n$ matrix. The linear system $A\vec{x} = \vec{b}$ has a solution for every choice of column vector $\vec{b} \in \mathbb{R}^n$ if and only if A is row equivalent to the $n \times n$ identity matrix $\mathcal{I}$.

Theorem 1.11. Commutivity Property.
Let A and C be $n \times n$ matrices. Then $CA = \mathcal{I}$ if and only if $AC = \mathcal{I}$.

Proof. Suppose that $AC = \mathcal{I}$. Then the equation $A\vec{x} = \vec{b}$ has a solution for every column vector $\vec{b} \in \mathbb{R}^n$. Notice that $\vec{x} = C\vec{b}$ is a solution because

$$A(C\vec{b}) = (AC)\vec{b} = \mathcal{I}\vec{b} = \vec{b}.$$

By Lemma 1.1, we know that A is row equivalent to the $n \times n$ identity matrix $\mathcal{I}$, and so there exists a sequence of elementary matrices $E_1, E_2, \ldots, E_t$ such that $(E_t \cdots E_2E_1)A = \mathcal{I}$. By Theorem 1.9, the two equations

$$(E_t \cdots E_2E_1)A = \mathcal{I} \text{ and } AC = \mathcal{I}$$

imply that $E_t \cdots E_2 E_1 = C$, and so we have $CA = \mathcal{I}$. The other half of the proof follows by interchanging the roles of A and C. *QED*

Note. Computation of Inverses.
If $A = [a_{ij}]$, then finding $A^{-1} = [x_{ij}]$ amounts to solving for x_{ij} in:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} = \mathcal{I}.$$

If we treat this as n systems of n equations in n unknowns, then the augmented matrix for these n systems is $[A \mid \mathcal{I}]$. So to compute A^{-1}:
(1) Form $[A \mid \mathcal{I}]$.
(2) Apply Gauss-Jordan method to produce the row equivalent $[\mathcal{I} \mid C]$. If A^{-1} exists, then $A^{-1} = C$.

Note. In the above computations, C is just the product of the elementary matrices that make up A^{-1}.

Theorem 1.12. Conditions for A^{-1} to Exist.
The following conditions for an $n \times n$ matrix A are equivalent:
(*i*) A is invertible.
(*ii*) A is row equivalent to $\mathcal{I}$.
(*iii*) $A\vec{x} = \vec{b}$ has a solution for each $\vec{b}$ (namely, $\vec{x} = A^{-1}\vec{b}$).
(*iv*) A can be expressed as a product of elementary matrices.
(*v*) The span of the column vectors of A is $\mathbb{R}^n$.

Note. In (*iv*) A is the left-to-right product of the inverses of the elementary matrices corresponding to succesive row operations that reduce A to $\mathcal{I}$.

1.6 Homogeneous Systems, Subspaces and Bases

Definition. A linear system $A\vec{x} = \vec{b}$ is *homogeneous* if $\vec{b} = \vec{0}$. The zero vector $\vec{x} = \vec{0}$ is a *trivial solution* to the homogeneous system $A\vec{x} = \vec{0}$. Nonzero solutions to $A\vec{x} = \vec{0}$ are called *nontrivial solutions.*

Theorem 1.13. Structure of the Solution Set of $A\vec{x} = \vec{0}$.
Let $A\vec{x} = \vec{0}$ be a homogeneous linear system. If $\vec{h_1}, \vec{h_2}, \ldots, \vec{h_n}$ are solutions, then any linear combination

$$r_1\vec{h_1} + r_2\vec{h_2} + \cdots + r_n\vec{h_n}$$

is also a solution.

Proof. Since $\vec{h_1}, \vec{h_2}, \ldots, \vec{h_n}$ are solutions,

$$A\vec{h_1} = A\vec{h_2} = \cdots = A\vec{h_n} = \vec{0}$$

and so

$$A(r_1\vec{h_1}+r_2\vec{h_2}+\cdots+r_n\vec{h_n}) = r_1A\vec{h_1}+r_2A\vec{h_2}+\cdots+r_nA\vec{h_n} = \vec{0}+\vec{0}+\cdots+\vec{0} = \vec{0}.$$

Therefore the linear combination is also a solution. *QED*

Definition 1.16. A subset W of $\mathbb{R}^n$ is *closed under vector addition* if for all $\vec{u}, \vec{v} \in W$, we have $\vec{u} + \vec{v} \in W$. If $r\vec{v} \in W$ for all $\vec{v} \in W$ and for all $r \in \mathbb{R}$, then W is *closed under scalar multiplication.* A nonempty subset W of $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$ if it is both closed under vector addition and scalar multiplication.

Theorem 1.14. Subspace Property of a Span.
Let $W = \text{sp}(\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k})$ be the span of $k > 0$ vectors in $\mathbb{R}^n$. Then W is a subspace of $\mathbb{R}^n$. (The vectors $\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}$ are said to *span* or *generate* the subspace.)

Definition. Given an $m \times n$ matrix A, the span of the row vectors of A is the *row space* of A, the span of the column vectors of A is the *column space* of A and the solution set to the system $A\vec{x} = \vec{0}$ is the *nullspace* of A.

Definition 1.17. Let W be a subspace of $\mathbb{R}^n$. A subset $\{\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}\}$ of W is a *basis* for W if every vector in W can be expressed uniquely as a linear combination of $\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}$.

Theorem 1.15. Unique Linear Combinations.
The set $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a basis for $W = \text{sp}(\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k)$ if and only if

$$r_1\vec{w}_1 + r_2\vec{w}_2 + \cdots + r_k\vec{w}_k = \vec{0}$$

implies

$$r_1 = r_2 = \cdots = r_k = 0.$$

Proof. First, if $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a basis for W, then each vector of W can be uniquely written as a linear combination of these $\vec{w}_i$'s. Since $\vec{0} = 0\vec{w}_1 + 0\vec{w}_2 + \cdots + 0\vec{w}_k$ and this is the unique way to write $\vec{0}$ in terms of the $\vec{w}_i$'s, then for any $r_1\vec{w}_1 + r_2\vec{w}_2 + \cdots + r_k\vec{w}_k = \vec{0}$ we must have $r_1 = r_2 = \cdots r_k = 0$.

Second, suppose that the only linear combination of $\vec{w}_i$s that gives $\vec{0}$ is $0\vec{w}_1 + 0\vec{w}_2 + \cdots + 0\vec{v}_k$. We want to show that any vector of W is a unique linear combination of the $\vec{w}_i$'s. Suppose for $\vec{w} \in W$ we have

$$\begin{aligned} \vec{w} &= c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_k\vec{w}_k \text{ and} \\ \vec{w} &= d_1\vec{w}_1 + d_2\vec{w}_2 + \cdots + d_k\vec{w}_k. \end{aligned}$$

Then

$$\begin{aligned} \vec{0} = \vec{w} - \vec{w} &= c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_k\vec{w}_k \\ &\quad - (d_1\vec{w}_1 + d_2\vec{w}_2 + \cdots + d_k\vec{w}_k) \\ &= (c_1 - d_1)\vec{w}_1 + (c_2 - d_2)\vec{w}_2 + \cdots + (c_k - d_k)\vec{w}_k. \end{aligned}$$

So each coefficient must be 0 and we have $c_i = d_i$ for $i = 1, 2, \ldots, k$ and $\vec{w}$ can be written as a linear combination of $\vec{w}_i$'s in only one unique way.

QED

Theorem 1.16. Let A be an $n \times n$ matrix. The following are equivalent:
(1) $A\vec{x} = \vec{b}$ has a unique solution,
(2) A is row equivalent to $\mathcal{I}$,
(3) A is invertible, and
(4) the column vectors of A form a basis for $\mathbb{R}^n$.

Theorem 1.17. Let A be an $m \times n$ matrix. The following are equivalent:
(1) each consistent system $A\vec{x} = \vec{b}$ has a unique solution,
(2) the reduced row-echelon form of A consists of the $n \times n$ identity matrix followed by $m - n$ rows of zeros, and
(3) the column vectors of A form a basis for the column space of A.

Corollary 1. Fewer Equations than Unknowns
If a linear system $A\vec{X} = \vec{b}$ is consistent and has fewer equations than unknowns, then it has an infinite number of solutions.

Corollary 2. The Homogeneous Case
(1) A homogeneous linear system $A\vec{x} = \vec{0}$ having fewer equations then unknowns has a nontrivial solution (i.e. a solution other than $\vec{x} = \vec{0}$),
(2) A square homogeneous system $A\vec{x} = \vec{0}$ has a nontrivial solution if and only if A <u>is not</u> row equivalent to the identity matrix.

Theorem 1.18. Structure of the Solution Set of $A\vec{x} = \vec{b}$.
Let $A\vec{x} = \vec{b}$ be a linear system. If $\vec{p}$ is any particular solution of $A\vec{x} = \vec{b}$ and $\vec{h}$ is a solution to $A\vec{x} = \vec{0}$, then $\vec{p} + \vec{h}$ is a solution of $A\vec{x} = \vec{b}$. In fact, every solution of $A\vec{x} = \vec{b}$ has the form $\vec{p} + \vec{h}$ and the general solution is $\vec{x} = \vec{p} + \vec{h}$ where $A\vec{h} = \vec{0}$ (that is, $\vec{h}$ is an arbitrary element of the nullspace of A).

2 Dimension, Rank, and Linear Transformations

2.1 Independence and Dimension

Definition 2.1. Let $\{\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}\}$ be a set of vectors in $\mathbb{R}^n$. A *dependence relation* in this set is an equation of the form

$$r_1\vec{w_1} + r_2\vec{w_2} + \cdots + r_k\vec{w_k} = \vec{0}$$

with at least one $r_j \neq 0$. If such a dependence relation exists, then $\{\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}\}$ is a *linearly dependent* set. A set of vectors which is not linearly dependent is *linearly independent*.

Theorem 2.1. Alternative Characterization of Basis
Let W be a subspace of $\mathbb{R}^n$. A subset $\{\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}\}$ of W is a basis for W if and only if
(1) $W = \text{sp}(\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k})$ and
(2) the vector $\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}$ are linearly independent.

Note. The proof of Theorem 2.1 follows directly from the definitions of *basis* and *linear independence*.

Theorem. Finding a Basis for $W = \text{sp}(\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k})$.
Form the matrix A whose jth column vector is $\vec{w_j}$. If we row-reduce A to row-echelon form H, then the set of all $\vec{w_j}$ such that the jth column of H contains a pivot, is a basis for W.

Theorem 2.2. Relative Sizes of Spanning and Independent Sets.
Let W be a subspace of $\mathbb{R}^n$. Let $\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}$ be vectors in W that span W and let $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m}$ be vectors in W that are independent. Then $k \geq m$.

Corollary. Invariance of Dimension.
Any two bases of a subspace of $\mathbb{R}^n$ contains the same number of vectors.

Definition 2.2. Let W be a subspace of $\mathbb{R}^n$. The number of elements in a basis for W is the *dimension* of W, denoted $\dim(W)$.

Theorem 2.3. Existence and Determination of Bases.
(1) Every subspace W of $\mathbb{R}^n$ has a basis and $\dim(W) \leq n$.
(2) Every independent set of vectors in $\mathbb{R}^n$ can be enlarged to become a basis of $\mathbb{R}^n$.

(3) If W is a subspace of $\mathbb{R}^n$ and $\dim(W) = k$ then

(a) every independent set of k vectors in W is a basis for W, and

(b) every set of k vectors in W that spans W is a basis of W.

2.2 The Rank of a Matrix

Note. In this section, we consider the relationship between the dimensions of the column space, row space and nullspace of a matrix A.

Theorem 2.4. Row Rank Equals Column Rank.
Let A be an $m \times n$ matrix. The dimension of the row space of A equals the dimension of the column space of A. The common dimension is the *rank* of A.

Note. The dimension of the column space is the number of pivots of A when in row-echelon form, so by page 129, the rank of A is the number of pivots of A when in row-echelon form.

Note. Finding Bases for Spaces Associated with a Matrix.
Let A be an $m \times n$ matrix with row-echelon form H.
(1) for a basis of the row space of A, use the nonzero rows of H,
(2) for a basis of the column space of A, use the columns of A corresponding to the columns of H which contain pivots, and
(3) for a basis of the nullspace of A use H to solve $H\vec{x} = \vec{0}$ as before.

Theorem 2.5. Rank Equation.
Let A be $m \times n$ with row-echelon form H.
(1) The dimension of the nullspace of A is

$$\begin{aligned} \text{nullity}(A) &= (\#\text{ free variables in solution of } A\vec{x} = \vec{0}) \\ &= (\#\text{ pivot-free columns of } H). \end{aligned}$$

(2) $\text{rank}(A) = (\#\text{ of pivots in } H)$.
(3) Rank Equation:

$$\text{rank}(A) + \text{nullity}(A) = \#\text{ of columns of } A.$$

Theorem 2.6. An Invertibility Criterion.
An $n \times n$ matrix A is invertible if and only if $\text{rank}(A) = n$.

Example. If A is square, then $\text{nullity}(A) = \text{nullity}(A^T)$.

Proof. The column space of A is the same as the row space of A^T, so $\text{rank}(A) = \text{rank}(A^T)$ and since the number of columns of A equals the number of columns of A^T, then by the Rank Equation:

$$\text{rank}(A) + \text{nullity}(A) = \text{rank}(A^T) + \text{nullity}(A^T)$$

and the result follows. *QED*

2.3 Linear Transformations of Euclidean Spaces

Definition. A *linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function whose *domain* is $\mathbb{R}^n$ and whose *codomain* is $\mathbb{R}^m$, where
(1) $T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v})$ for all $\vec{u}, \vec{v} \in \mathbb{R}^n$, and
(2) $T(r\vec{u}) = rT(\vec{u})$ for all $\vec{u} \in \mathbb{R}^n$ and for all $r \in \mathbb{R}$.

Note. Combining **(1)** and **(2)** gives

$$T(r\vec{u} + s\vec{v}) = rT(\vec{u}) + sT(\vec{v})$$

for all $\vec{u}, \vec{v} \in \mathbb{R}^n$ and $r, s \in \mathbb{R}$. As the book says, "linear transformations preserve linear combinations."

Note. $T(\vec{0}) = T(0\vec{0}) = 0T(\vec{0}) = \vec{0}$.

Theorem 2.7. Bases and Linear Transformations.
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $B = \{\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n}\}$ be a basis for $\mathbb{R}^n$. For any vector $\vec{v} \in \mathbb{R}^n$, the vector $T(\vec{v})$ is uniquely determined by $T(\vec{b_1}), T(\vec{b_2}), \ldots, T(\vec{b_n})$..

Proof. Let $\vec{v} \in \mathbb{R}^n$. Then since B is a basis, there exist unique scalars $r_1, r_2, \ldots, r_n$ such that

$$\vec{v} = r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_n\vec{b_n}.$$

Since T is linear, we have

$$T(\vec{v}) = r_1T(\vec{b_1}) + r_2T(\vec{b_2}) + \cdots + r_nT(\vec{b_n}).$$

Since the coefficients r_i are uniquely determined by $\vec{v}$, it follows that the value of $T(\vec{v})$ is completely determined by the vectors $T(\vec{b_i})$. *QED*

Corollary. Standard Matrix Representation of Linear Transformations.
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear, and let A be the $m \times n$ matrix whose jth column is $T(\hat{e_j})$. Then $T(\vec{x}) = A\vec{x}$ for each $\vec{x} \in \mathbb{R}^n$. A is the *standard matrix representation* of T.

Proof. For any matrix A, $A\hat{e_j}$ is the jth column of A. So if A is the matrix described, then $A\hat{e_j} = T(\hat{e_j})$, and so T and the linear transformation T_A given by $T_A(\vec{x}) = A\vec{x}$ agree on the standard basis $\{\hat{e_1}, \hat{e_2}, \ldots, \hat{e_n}\}$ of $\mathbb{R}^n$. Therefore by Theorem 2.7, $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. *QED*

Theorem/Definition. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix representation A.
(1) The *range* $T[\mathbb{R}^n]$ of T is the column space of A.
(2) The *kernel* of T is the nullspace of A, denoted $\ker(T)$.
(3) If W is a subspace of $\mathbb{R}^n$, then $T[W]$ is a subspace of $\mathbb{R}^m$ (i.e. T preserves subspaces).

Notice. If A is the standard matrix representation for T, then from the rank equation we get:

$$\dim(\text{range } T) + \dim(\ker T) = \dim(\text{domain } T).$$

Definition. For a linear transformation T, we define *rank* and *nullity* in terms of the standard matrix representation A of T:

$$\text{rank}(T) = \dim(\text{range } T), \qquad \text{nullity}(T) = \dim(\ker T).$$

Definition. If $T : \mathbb{R}^n \to \mathbb{R}^m$ and $T' : \mathbb{R}^m \to \mathbb{R}^k$, then the *composition* of T and T' is $(T' \circ T) : \mathbb{R}^n \to \mathbb{R}^k$ where $(T' \circ T)\vec{x} = T'(T(\vec{x}))$.

Theorem. Matrix Multiplication and Composite Transformations.
A composition of two linear transformations T and T' with standard matrix representation A and A' yields a linear transformation $T' \circ T$ with standard matrix representation $A'A$.

Definition. If $T : \mathbb{R}^n \to \mathbb{R}^n$ and there exists $T' : \mathbb{R}^n \to \mathbb{R}^n$ such that $T \circ T'(\vec{x}) = \vec{x}$ for all $\vec{x} \in \mathbb{R}^n$, then T' is the *inverse* of T denoted $T' = T^{-1}$. (Notice that if $T : \mathbb{R}^m \to \mathbb{R}^n$ where $m \neq n$, then T^{-1} is not defined — there are domain/range size problems.)

Theorem. Invertible Matrices and Inverse Transformations.
Let $T : \mathbb{R}^n \to \mathbb{R}^n$ have standard matrix representation A: $T(\vec{x}) = A\vec{x}$. Then T is invertible if and only if A is invertible and $T^{-1}(\vec{x}) = A^{-1}\vec{x}$.

2.4 Linear Transformations of the Plane (in brief)

Note. If A is a 2×2 matrix with rank 0 then it is the matrix

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

and all vectors in $\mathbb{R}^2$ are mapped to $\vec{0}$ under the transformation with asociated matrix A (We can view $\vec{0}$ as a 0 dimensional space). If the rank$(A) = 1$, then the column space of A, which is the range of T_A, is a one dimensional subspace of $\mathbb{R}^2$. In this case, T_A *projects* a vector onto the column space. See page 155 for details.

Note. We can *rotate* a vector in $\mathbb{R}^2$ about the origin through an angle θ by applying T_A where

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

This is an example of a *rigid* transformation of the plane since lengths are not changed under this transformation.

Note. We can *reflect* a vector in $\mathbb{R}^2$ about the x-axis by applying T_X where

$$X = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We can *reflect* a vector in $\mathbb{R}^2$ about the y-axis by applying T_Y where

$$Y = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We can *reflect* a vector in $\mathbb{R}^2$ about the line $y = x$ by applying T_Z where

$$Z = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Notice that X, Y, and Z are elementary matrices since they differ from $\mathcal{I}$ by an operation of row scaling (for X and Y), or by an operation of row interchange (for Z).

Note. Transformation T_A where

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

is a *horizontal expansion* if $r > 1$, and is a *horizontal contraction* if $0 < r < 1$. Transformation T_B where

$$B = \begin{bmatrix} 1 & 0 \\ 0 & r \end{bmatrix}$$

is a *vertical expansion* if $r > 1$, and is a *vertical contraction* if $0 < r < 1$. Notice that A and B are elementary matrices since they differ from $\mathcal{I}$ by an operation of row scaling.

Note. Transformation T_A where

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

is a *vertical shear* (see Figure 2.2.16 on page 163). Transformation T_B where

$$B = \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}$$

is a *horizontal shear*. Notice that A and B are elementary matrices since they differ from $\mathcal{I}$ by an operation of row addition.

Theorem. Geometric Description of Invertible Transformations of $\mathbb{R}^2$.
A linear transformation T of the plane $\mathbb{R}^2$ into itself is invertible if and only if T consists of a finite sequence of:
- Reflections in the x-axis, the y-axis, or the line $y = x$;
- Vertical or horizontal expansions or contractions; and
- Vertical or horizontal shears.

Proof. Each elementary operation corresponds to one of these types of transformations (and conversely). Each of these transformations correspond to elementary matrices as listed above (and conversely). Also, we know that a matrix is invertible if and only if it is a product of elementary matrices by Theorem 1.12(iv). Therefore T is invertible if and only if its associated matrix is a product of elementary matrices, and so the result follows. *QED*

2.5 Lines, Planes, and Other Flats

Definitions 2.4, 2.5. Let S be a subset of $\mathbb{R}^n$ and let $\vec{a} \in \mathbb{R}^n$. The set $\{\vec{x}+\vec{a} \mid \vec{x} \in S\}$ is the *translate* of S by $\vec{a}$, and is denoted by $S+\vec{a}$. The vector $\vec{a}$ is the *translation vector*. A *line* in $\mathbb{R}^n$ is a translate of a one-dimensional subspace of $\mathbb{R}^n$.

Definition. If a line L in $\mathbb{R}^n$ contains point $(a_1, a_2, \ldots, a_n)$ and if vector $\vec{d}$ is parallel to L, then $\vec{d}$ is a *direction vector* for L and $\vec{a} = [a_1, a_2, \ldots, a_n]$ is a *translation vector* of L.

Note. With $\vec{d}$ as a direction vector and $\vec{a}$ as a translation vector of a line, we have $L = \{t\vec{d} + \vec{a} \mid t \in \mathbb{R}\}$. In this case, t is called a *parameter* and we can express the line *parametrically* as a vector equation:

$$\vec{x} = t\vec{d} + \vec{a}$$

or as a collection of component equations:

$$\begin{aligned} x_1 &= td_1 + a_1 \\ x_2 &= td_2 + a_2 \\ &\vdots \\ x_n &= td_n + a_n. \end{aligned}$$

Definition 2.6. A *k-flat* in $\mathbb{R}^n$ is a translate of a k-dimensional subspace of $\mathbb{R}^n$. In particular, a 1-flat is a *line*, a 2-flat is a *plane*, and an $(n-1)$-flat is a *hyperplane*. We consider each point of $\mathbb{R}^n$ to be a *zero-flat*.

Note. We can also talk about a translate of a k-dimensional subspace W of $\mathbb{R}^n$. If a basis for W is $\{\vec{d_1}, \vec{d_2}, \ldots, \vec{d_k}\}$, then the k-flat through the point $(a_1, a_2, \ldots, a_n)$ and parallel to W is

$$\vec{x} = t_1\vec{d_1} + t_2\vec{d_2} + \cdots + t_k\vec{d_k} + \vec{a}$$

where $\vec{a} = [a_1, a_2, \ldots, a_n]$ and $t_1, t_2, \ldots, t_k \in \mathbb{R}$ are *parameters*. We can also express this k-flat parametrically in terms of components.

Note. We can now clearly explain the geometric interpretation of solutions of linear systems in terms of k-flats. Consider $A\vec{x} = \vec{b}$, a system of m equations in n unknowns that has at least one solution $\vec{x} = \vec{p}$. By Theorem 1.18 on page 97, the solution set of the system consists of all vectors of the form $\vec{x} = \vec{p} + \vec{h}$ where $\vec{h}$ is a solution of the homogeneous system $A\vec{x} = \vec{0}$. Now the solution set of $A\vec{x} = \vec{0}$ is a subspace of $\mathbb{R}^n$, and so the solution of $A\vec{x} = \vec{b}$ is a k-flat (where k is the nullity of A) passing through point $(p_1, p_2, \ldots, p_n)$ where $\vec{p} = [p_1, p_2, \ldots, p_n]$.

3 Vector Spaces

3.1 Vector Spaces

Definition 3.1. A *vector space* is a set V of *vectors* along with an operation of addition $+$ of vectors and multiplication of a vector by a scalar (real number), which satisfies the following. For all $\vec{u}, \vec{v}, \vec{w} \in V$ and for all $r, s \in \mathbb{R}$:
(A1) $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
(A2) $\vec{v} + \vec{w} = \vec{w} + \vec{v}$
(A3) There exists $\vec{0} \in V$ such that $\vec{0} + \vec{v} = \vec{v}$
(A4) $\vec{v} + (-\vec{v}) = \vec{0}$
(S1) $r(\vec{v} + \vec{w}) = r\vec{v} + r\vec{w}$
(S2) $(r + s)\vec{v} = r\vec{v} + s\vec{v}$
(S3) $r(s\vec{v}) = (rs)\vec{v}$
(S4) $1\vec{v} = \vec{v}$

Definition. $\vec{0}$ is the *additive identity.* $-\vec{v}$ is the *additive inverse* of $\vec{v}$.

Example. Some examples of vector spaces are:
(1) The set of all polynomials of degree n or less, denoted $\mathcal{P}_n$.
(2) All $m \times n$ matrices.
(3) The set of all functions integrable f with domain $[0, 1]$ such that $\displaystyle\int_0^1 |f(x)|^2\,dx < \infty$. This vector space is denoted $L^2[0, 1]$:

$$L^2[0,1] = \left\{ f \,\middle|\, \int_0^1 |f(x)|^2\,dx < \infty \right\}.$$

Theorem 3.1. Elementary Properties of Vector Spaces.
Every vector space V satisfies:
(1) the vector $\vec{0}$ is the unique additive identity in a vector space,
(2) for each $\vec{v} \in V$, $-\vec{v}$ is the unique additive inverse of $\vec{v}$,
(3) if $\vec{u} + \vec{v} = \vec{u} + \vec{w}$ then $\vec{v} = \vec{w}$,
(4) $0\vec{v} = \vec{0}$ for all $\vec{v} \in V$,
(5) $r\vec{0} = \vec{0}$ for all scalars $r \in \mathbb{R}$,
(6) $(-r)\vec{v} = r(-\vec{v}) = -(r\vec{v})$ for all $r \in \mathbb{R}$ and for all $\vec{v} \in V$.

Proof of (1) and (3). Suppose that there are two additive identities, $\vec{0}$ and

$\vec{0}'$. Then consider:

$$\begin{aligned}\vec{0} &= \vec{0} + \vec{0}' \qquad \text{(since } \vec{0}' \text{ is an additive identity)} \\ &= \vec{0}' \text{(since } \vec{0} \text{ is an additive identity).}\end{aligned}$$

Therefore, $\vec{0} = \vec{0}'$ and the additive identity is unique.

Suppose $\vec{u} + \vec{v} = \vec{u} + \vec{w}$. Then we add $-\vec{u}$ to both sides of the equation and we get:

$$\vec{u} + \vec{v} + (-\vec{u}) = \vec{u} + \vec{w} + (-\vec{u})$$

$$\vec{v} + (\vec{u} - \vec{u}) = \vec{w} + (\vec{u} - \vec{u})$$

$$\vec{v} + \vec{0} = \vec{w} + \vec{0}$$

$$\vec{v} = \vec{w}$$

The conclusion holds. *QED*

3.2 Basic Concepts of Vector Spaces

Definition 3.2. Given vectors $\vec{v_1}, \vec{v_2}, \dots, \vec{v_k} \in V$ and scalars $r_1, r_2, \dots, r_k \in \mathbb{R}$,

$$\sum_{l=1}^{k} r_l \vec{v_l} = r_1 \vec{v_1} + r_2 \vec{v_2} + \cdots + r_k \vec{v_k}$$

is a *linear combination* of $\vec{v_1}, \vec{v_2}, \dots, \vec{v_k}$ with *scalar coefficients* $r_1, r_2, \dots, r_k$.

Definition 3.3. Let X be a subset of vector space V. The *span* of X is the set of all linear combinations of elements in X and is denoted $\mathrm{sp}(X)$. If $V = \mathrm{sp}(X)$ for some finite set X, then V is *finitely generated.*

Definition 3.4. A subset W of a vector space V is a *subspace* of V if W is itself a vector space.

Theorem 3.2. Test for Subspace.
A subset W of vector space V is a subspace if and only if
(1) $\vec{v}, \vec{w} \in W \Rightarrow \vec{v} + \vec{w} \in W$,
(2) for all $r \in \mathbb{R}$ and for all $\vec{v} \in W$ we have $r\vec{v} \in W$.

Defninition 3.5. Let X be a set of vectors from a vector space V. A *dependence relation* in X is an equation of the form

$$\sum_{l=1}^{k} r_l \vec{v_l} = r_1 \vec{v_1} + r_2 \vec{v_2} + \cdots + r_k \vec{v_k} = \vec{0}$$

with some $r_j \neq 0$ and $\vec{v_i} \in X$. If such a relation exists, then X is a *linearly dependent* set. Otherwise X is a *linearly independent* set.

Definition 3.6. Let V be a vector space. A set of vectors in V is a *basis* for V if
(1) the set of vectors span V, and
(2) the set of vectors is linearly independent.

Theorem 3.3. Unique Combination Criterion for a Basis.
Let B be a set of nonzero vectors in vector space V. Then B is a basis for V if and only if each vector in V can by uniquely expressed as a linear combination of the vectors in set B.

Proof. Suppose that B is a basis for vector space V. Then by the first part of Definition 3.6 we see that any vector $\vec{v} \in V$ can be written as a linear combination of the elements of B, say

$$\vec{v} = r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_k\vec{b_k}.$$

Now suppose that there is some other linear combination of the vectors in B which represents $\vec{v}$ (we look for a contradiction):

$$\vec{v} = s_1\vec{b_1} + s_2\vec{b_2} + \cdots + s_k\vec{b_k}.$$

If we subtract these two representations of $\vec{v}$ then we get that

$$\vec{0} = (r_1 - s_1)\vec{b_1} + (r_2 - s_2)\vec{b_2} + \cdots + (r_k - s_k)\vec{b_k}.$$

By the second part of Definition 3.6, we know that $r_1 - s_1 = r_2 - s_2 = \cdots = r_k - s_k = 0$. Therefore there is only one linear combination of elements of B which represent $\vec{v}$.

Now suppose that each vector in V can be uniquley represented as a linear combination of the elements of B. We wish to show that B is a basis. Clearly B is a spanning set of V. Now we can write $\vec{0}$ as a linear combination of elements of B by taking all coefficients as 0. Since we hypothesize that each vector can be *uniquely* represented, then

$$\vec{0} = r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_k\vec{b_k}$$

only for $r_1 = r_2 = \cdots = r_k = 0$. Hence the elements of B are linearly independent and so B is a basis. *QED*

Definition. A vector space is *finitely generated* if it is the span of some finite set.

Theorem 3.4. Relative Size of Spanning and Independent Sets. Let V be a vector space. Let $\vec{w_1}, \vec{w_2}, \ldots, \vec{w_k}$ be vectors in V that span V and let $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m}$ be vectors in V that are independent. Then $k \geq m$.

Corollary. Invariance of Dimension for Finitely Generated Spaces. Let V be a finitely generated vector space. Then any two bases of V have the same number of elements.

Definition 3.7. Let V be a finitely generated vector space. The number of elements in a basis for V is the *dimension* of V, denoted $\dim(V)$.

3.3 Coordinatization of Vectors

Definition. An *ordered basis* $(\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n})$ is an "ordered set" of vectors which is a basis for some vector space.

Definition 3.8. If $B = (\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n})$ is an ordered basis for V and $\vec{v} = r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_n\vec{b_n}$, then the vector $[r_1, r_2, \ldots, r_n] \in \mathbb{R}^n$ is the *coordinate vector of $\vec{v}$ relative to B*, denoted $\vec{v_B}$.

Note. To find $\vec{v_B}$:
(1) write the basis vectors as column vectors to form $[\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n} \mid \vec{v}]$,
(2) use Gauss-Jordan elimination to get $[\mathcal{I} \mid \vec{v_B}]$.

Definition. An *isomorphism* between two vector spaces V and W is a one-to-one and onto function α from V to W such that:
(1) if $\vec{v_1}, \vec{v_2} \in V$ then

$$\alpha(\vec{v_1} + \vec{v_2}) = \alpha(\vec{v_1}) + \alpha(\vec{v_2}), \text{ and}$$

(2) if $\vec{v} \in V$ and $r \in \mathbb{R}$ then $\alpha(r\vec{v}) = r\alpha(\vec{v})$.
If there is such an α, then V and W are *isomorphic*, denoted $V \cong W$.

Note. An isomorphism is a one-to-one and onto linear transformation.

Theorem. The Fundamental Theorem of Finite Dimensional Vectors Spaces.
If V is a finite dimensional vector space (say $\dim(V) = n$) then V is isomorphic to $\mathbb{R}^n$.

Proof. Let $B = (\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n})$ be an ordered basis for V and for $\vec{v} \in V$ with $\vec{v_B} = [r_1, r_2, \ldots, r_n]$ define $\alpha : V \to \mathbb{R}^n$ as

$$\alpha(\vec{v}) = [r_1, r_2, \ldots, r_n].$$

Then "clearly" α is one-to-one and onto. Also for $\vec{v}, \vec{w} \in V$ suppose

$$\vec{v_B} = [r_1, r_2, \ldots, r_n] \text{ and } \vec{w_B} = [s_1, s_2, \ldots, s_n]$$

and so

$$\begin{aligned} \alpha(\vec{v} + \vec{w}) &= [r_1 + s_1, r_2 + s_2, \ldots, r_n + s_n] \\ &= [r_1, r_2, \ldots, r_n] + [s_1, s_2, \ldots, s_n] \\ &= \alpha(\vec{v}) + \alpha(\vec{w}). \end{aligned}$$

For a scalar $t \in \mathbb{R}$,

$$\alpha(t\vec{v}) = [tr_1, tr_2, \ldots, tr_n] = t[r_1, r_2, \ldots, r_n] = t\alpha(\vec{v}).$$

So α is an isomorphism and $V \cong \mathbb{R}^n$. *QED*

Example. Prove the set $\{(x-a)^n, (x-a)^{n-1}, \ldots, (x-a), 1\}$ is a basis for $\mathcal{P}_n$.

Proof. Let $\vec{v_0}, \vec{v_1}, \ldots, \vec{v_n}$ be the coordinate vectors of $1, (x-a), \ldots, (x-a)^n$ in terms of the ordered basis $\{1, x, x^2, \ldots, x^n\}$. Form a matrix A with the $\vec{v_i}$s as the columns:

$$A = [\vec{v_0}\vec{v_1} \cdots \vec{v_n}].$$

Notice that A is "upper triangular:'

$$A = \begin{bmatrix} 1 & -a & a^2 & \cdots & (-a)^n \\ 0 & 1 & -2a & \cdots & \vdots \\ 0 & 0 & 1 & \cdots & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

and so the $\vec{v_i}$ are linearly independent. Since $\dim(\mathcal{P}_n) = n+1$ and the set

$$\{(x-a)^n, (x-a)^{n-1}, \ldots, (x-a), 1\}$$

is a set of $n+1$ linearly independent vectors, then this set is a basis for $\mathcal{P}_n$. *QED*

3.4 Linear Transformations

Note. We have already studied linear transformations from $\mathbb{R}^n$ into $\mathbb{R}^m$. Now we look at linear transformations from one general vector space to another.

Definition 3.9. A function T that maps a vector space V into a vector space V' is a *linear transformation* if it satisfies:
(1) $T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v})$, and **(2)** $T(r\vec{u}) = rT(\vec{u})$,
for all vectors $\vec{u}, \vec{v} \in V$ and for all scalars $r \in \mathbb{R}$.

Definition. For a linear transformation $T : V \to V'$, the set V is the *domain* of T and the set V' is the *codomain* of T. If W is a subset of V, then $T[W] = \{T(\vec{w}) \mid \vec{w} \in W\}$ is the *image* of W under T. $T[V]$ is the *range* of T. For $W' \subset V'$, $T^{-1}[W'] = \{\vec{v} \in V \mid T(\vec{v}) \in W'\}$ is the *inverse image* of W' under T. $T^{-1}[\{\vec{0'}\}]$ if the *kernal* of T, denoted $\ker(T)$.

Definition. Let V, V' and V'' be vector spaces and let $T : V \to V'$ and $T' : V' \to V''$ be linear transformations. The *composite transformation* $T' \circ T : V \to V''$ is defined by $(T' \circ T)(\vec{v}) = T'(T(\vec{v}))$ for $\vec{v} \in V$.

Example. Let F be the vector space of all functions $f : \mathbb{R} \to \mathbb{R}$, and let D be its subspace of all differentiable functions. Then differentiation is a linear transformation of D into F.

Theorem 3.5. Preservation of Zero and Subtraction
Let V and V' be vectors spaces, and let $T : V \to V'$ be a linear transformation. Then
(1) $T(\vec{0}) = \vec{0}'$, and
(2) $T(\vec{v_1} - \vec{v_2}) = T(\vec{v_1}) - T(\vec{v_2})$,
for any vectors $\vec{v_1}$ and $\vec{v_2}$ in V.

Proof of (1). Consider

$$T(\vec{0}) = T(0\vec{0}) = 0T(\vec{0}) = \vec{0}'.$$

QED

Theorem 3.6. Bases and Linear Transformations.
Let $T : V \to V'$ be a linear transformation, and let B be a basis for V. For any vector $\vec{v}$ in V, the vector $T(\vec{v})$ is uniquely determined by the vectors $T(\vec{b})$ for all $\vec{b} \in B$. In other words, if two linear transformations have the

same value at each basis vector $\vec{b} \in B$, then the two transformations have the same value at each vector in V.

Proof. Let T and $\overline{T}$ be two linear transformations such that $T(\vec{b_i}) = \overline{T}(\vec{b_i})$ for each vector $\vec{b_i} \in B$. Let $\vec{v} \in V$. Then for some scalars $r_1, r_2, \ldots, r_k$ we have

$$\vec{v} = r_1\vec{b_1} + r_2\vec{v_2} + \cdots + r_k\vec{b_k}.$$

Then

$$\begin{aligned} T(\vec{v}) &= T(r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_k\vec{v_k}) \\ &= r_1T(\vec{b_1}) + r_2T(\vec{b_2}) + \cdots + r_kT(\vec{v_k}) \\ &= r_1\overline{T}(\vec{b_1}) + r_2\overline{T}(\vec{b_2}) + \cdots + r_k\overline{T}(\vec{v_k}) \\ &= \overline{T}(r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_k\vec{v_k}) \\ &= \overline{T}(\vec{v}). \end{aligned}$$

Therefore T and $\overline{T}$ are the same tranformations. *QED*

Theorem 3.7. Preservation of Subspaces.
Let V and V' be vector spaces, and let $T : V \to V'$ be a linear transformation.
(1) If W is a subspace of V, then $T[W]$ is a subspace of V'.
(2) If W' is a subspace of V', then $T^{-1}[W']$ is a subspace of V.

Theorem. Let $T : V \to V'$ be a linear transformation and let $T(\vec{p}) = \vec{b}$ for a particular vector $\vec{p}$ in V. The solution set of $T(\vec{x}) = \vec{b}$ is the set $\{\vec{p} + \vec{h} \mid \vec{h} \in \ker(T)\}$.

Proof. (Page 229 number 46) Let $\vec{p}$ be a solution of $T(\vec{x}) = \vec{b}$. Then $T(\vec{p}) = \vec{b}$. Let $\vec{h}$ be a solution of $T(\vec{x}) = \vec{0'}$. Then $T(\vec{h}) = \vec{0'}$. Therefore

$$T(\vec{p} + \vec{h}) = T(\vec{p}) + T(\vec{h}) = \vec{b} + \vec{0'} = \vec{b},$$

and so $\vec{p} + \vec{h}$ is indeed a solution. Also, if $\vec{q}$ is any solution of $T(\vec{x}) = \vec{b}$ then

$$T(\vec{q} - \vec{p}) = T(\vec{q}) - T(\vec{p}) = \vec{b} - \vec{b} = \vec{0'},$$

and so $\vec{q} - \vec{p}$ is in the kernal of T. Therefore for some $\vec{h} \in \ker(T)$, we have $\vec{q} - \vec{p} = \vec{h}$, and $\vec{q} = \vec{p} + \vec{h}$. *QED*

Definition. A transformation $T : V \to V'$ is *one-to-one* if $T(\vec{v_1}) = T(\vec{v_2})$ implies that $\vec{v_1} = \vec{v_2}$ (or by the contrapositive, $\vec{v_1} \neq \vec{v_2}$ implies $T(\vec{v_1}) \neq T(\vec{v_2})$). Transformation T is *onto* if for all $\vec{v'} \in V'$ there is a $\vec{v} \in V$ such that $T(\vec{v}) = \vec{v'}$.

Corollary. A linear transformation T is one-to-one if and only if $\ker(T) = \{\vec{0}\}$.

Proof. By the previous theorem, if $\ker(T) = \{\vec{0}\}$, then for all relevant $\vec{b}$, the equation $T(\vec{x}) = \vec{b}$ has a unique solution. Therefore T is one-to-one.

Next, if T is one-to-one then for any nonzero vector $\vec{x}$, $T(\vec{x})$ is nonzero. Therefore by Theorem 3.5 Part (1), $\ker(T) = \{\vec{0}\}$. *QED*

Definition 3.10. Let V and V' be vector spaces. A linear transformation $T : V \to V'$ is *invertible* if there exists a linear transformation $T^{-1} : V' \to V$ such that $T^{-1} \circ T$ is the identity transformation on V and $T \circ T^{-1}$ is the identity transformation on V'. Such T^{-1} is called an *inverse transformation* of T.

Theorem 3.8. A linear transformation $T : V \to V'$ is invertible if and only if it is one-to-one and onto V'.

Proof. Suppose T is invertible and is not one-to-one. Then for some $\vec{v_1} \neq \vec{v_2}$ both in V, we have $T(\vec{v_1}) = T(\vec{v_2}) = \vec{v'}$. But then $T^{-1} \circ T(\vec{v'}) = \vec{v_1}$ and $T^{-1} \circ T(\vec{v'}) = \vec{v_2}$, a contradiction. Therefore if T is invertible then T is one-to-one.

From definition 3.10, if T is invertible then for any $\vec{v'} \in V'$ we must have $T^{-1}(\vec{v'}) = \vec{v}$ for some $\vec{v} \in V$. Therefore the image of $\vec{v}$ is $\vec{v'} \in V'$ and T is onto.

Finally, we need to show that if T is one-to-one and onto then it is invertible. Suppose that T is one-to-one and onto V'. Since T is onto V', then for each $\vec{v'} \in V'$ we can find $\vec{v} \in V$ such that $T(\vec{v}) = \vec{v'}$. Because T is one-to-one, this vector $\vec{v} \in V$ is unique. Let $T^{-1} : V' \to V$ be defined by $T^{-1}(\vec{v'}) = \vec{v}$. Then

$$(T \circ T^{-1})(\vec{v'}) = T(T^{-1}(\vec{v'})) = T(\vec{v}) = \vec{v'}$$

and

$$(T^{-1} \circ T)(\vec{v}) = T^{-1}(T(\vec{v})) = T^{-1}(\vec{v'}) = \vec{v},$$

and so $T \circ T^{-1}$ is the identity map on V' and $T^{-1} \circ T$ is the identity map on V.

Now we need only show that T^{-1} is linear. Suppose $T(\vec{v_1}) = \vec{v_1'}$ and $T(\vec{v_2}) = \vec{v_2'}$. Then

$$T^{-1}(\vec{v_1'} + \vec{v_2'}) = T^{-1}(T(\vec{v_1}) + T(\vec{v_2})) = T^{-1}(T(\vec{v_1} + \vec{v_2}))$$

$$= (T^{-1} \circ T)(\vec{v_1} + \vec{v_2}) = \vec{v_1} + \vec{v_2} = T^{-1}(\vec{v_1'}) + T^{-1}(\vec{v_2'}).$$

Also

$$T^{-1}(r\vec{v_1'}) = T^{-1}(rT(\vec{v_1})) = T^{-1}(T(r\vec{v_1})) = r\vec{v_1} = rT^{-1}(\vec{v_1'}).$$

Therefore T^{-1} is linear. *QED*

Theorem 3.9. Coordinatization of Finite-Dimensional Spaces. Let V be a finite-dimensional vector space with ordered basis $B = (\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n})$. The map $T : V \to \mathbb{R}^n$ defined by $T(\vec{v}) = \vec{v_B}$, the coordinate vector of $\vec{v}$ relative to B, is an isomorphism.

Theorem 3.10. Matrix Representations of Linear Transformations. Let V and V' be finite-dimensional vector spaces and let $B = (\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n})$ and $B' = (\vec{b_1'}, \vec{b_2'}, \ldots, \vec{b_m'})$ be ordered bases for V and V', respectively. Let $T : V \to V'$ be a linear transformation, and let $\overline{T} : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation such that for each $\vec{v} \in V$, we have $\overline{T}(\vec{v_B}) = T(\vec{v})_{B'}$. Then the standard matrix representation of $\overline{T}$ is the matrix A whose jth column vector is $T(\vec{b_j})_{B'}$, and $T(\vec{v})_{B'} = A\vec{v_B}$ for all vectors $\vec{v} \in V$.

Definition 3.11. The matrix A of Theorem 3.10 is the *matrix representation of T relative to B, B'.*

Theorem. The matrix representation of T^{-1} relative to B', B is the inverse of the matrix representation of T relative to B, B'.

3.5 Inner-Product Spaces

Note. In this section, we generalize the idea of dot product. We use this more general idea to define length and angle.

Note. Motivated by the properties of dot product on $\mathbb{R}^n$, we define the following:

Definition 3.12. An *inner product* on a vector space V is a function that associates with each ordered pair of vectors $\vec{v}, \vec{w} \in V$ a real number, written $\langle \vec{v}, \vec{w} \rangle$, satisfying the following properties for all $\vec{u}, \vec{v}, \vec{w} \in V$ and for all scalars r:
P1. Symmetry: $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$
P2. Additivity: $\langle \vec{u}, \vec{v} + \vec{w} \rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u}, \vec{w} \rangle$,
P3. Homogeneity: $r\langle \vec{v}, \vec{w} \rangle = \langle r\vec{v}, \vec{w} \rangle = \langle \vec{v}, r\vec{w} \rangle$,
P4. Positivity: $\langle \vec{v}, \vec{v} \rangle \geq 0$, and $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = \vec{0}$.
An *inner-product space* is a vector space V together with an inner product on V.

Example. Dot product on $\mathbb{R}^n$ is an example of an inner product: $\langle \vec{v}, \vec{w} \rangle = \vec{v} \cdot \vec{w}$ for $\vec{v}, \vec{w} \in \mathbb{R}^n$.

Example. Show that the space $P_{0,1}$ of all polynomial functions with real coefficients and domain $0 \leq x \leq 1$ is an inner-product space if for p and q in $P_{0,1}$ we define

$$\langle p, q \rangle = \int_0^1 p(x)q(x)\,dx.$$

Definition 3.13. Let V be an inner-product space. The *magnitude* or *norm* of a vector $\vec{v} \in V$ is $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$. The *distance* between $\vec{v}$ and $\vec{w}$ in an inner-product space V is $d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\|$.

Theorem 3.11. Schwarz Inequality.
Let V be an inner-product space, and let $\vec{v}, \vec{w} \in V$. Then

$$\langle \vec{v}, \vec{w} \rangle \leq \|\vec{v}\|\|\vec{w}\|.$$

Proof. Let $r, s \in \mathbb{R}$. Then by Definition 3.12

$$\begin{aligned} \|r\vec{v} + s\vec{w}\|^2 &= \langle r\vec{v} + s\vec{w}, r\vec{v} + s\vec{w} \rangle \\ &= r^2\langle \vec{v}, \vec{v} \rangle + 2rs\langle \vec{v}, \vec{w} \rangle + s^2\langle \vec{w}, \vec{w} \rangle \\ &\geq 0. \end{aligned}$$

Since this equation holds for all $r, s \in \mathbb{R}$, we are free to choose particular values of r and s. We choose $r = \langle \vec{w}, \vec{w} \rangle$ and $s = -\langle \vec{v}, \vec{w} \rangle$. Then we have

$$\begin{aligned} &\langle \vec{w}, \vec{w} \rangle^2 \langle \vec{v}, \vec{v} \rangle - 2\langle \vec{w}, \vec{w} \rangle \langle \vec{v}, \vec{w} \rangle^2 + \langle \vec{v}, \vec{w} \rangle^2 \langle \vec{w}, \vec{w} \rangle \\ &= \langle \vec{w}, \vec{w} \rangle^2 \langle \vec{v}, \vec{v} \rangle - \langle \vec{w}, \vec{w} \rangle \langle \vec{v}, \vec{w} \rangle^2 \\ &= \langle \vec{w}, \vec{w} \rangle [\langle \vec{w}, \vec{w} \rangle \langle \vec{v}, \vec{v} \rangle - \langle \vec{v}, \vec{w} \rangle^2] \geq 0. \end{aligned} \tag{13}$$

If $\langle \vec{w}, \vec{w} \rangle = 0$ then $\vec{w} = 0$ by Theorem 3.12 Part (P4), and the Schwarz Inequality is proven (since it reduces to $0 \geq 0$). If $\|\vec{w}\|^2 = \langle \vec{w}, \vec{w} \rangle \neq 0$, then by the above inequality the other factor of inequality (13) must also be nonnegative:

$$\langle \vec{w}, \vec{w} \rangle \langle \vec{v}, \vec{v} \rangle - \langle \vec{v}, \vec{w} \rangle^2 \geq 0.$$

Therefore

$$\langle \vec{v}, \vec{w} \rangle^2 \leq \langle \vec{v}, \vec{v} \rangle \langle \vec{w}, \vec{w} \rangle = \|\vec{v}\|^2 \|\vec{w}\|^2.$$

Taking square roots, we get the Schwarz Inequality. *QED*

Theorem. The Triangle Inequality.
Let $\vec{v}, \vec{w} \in V$ (where V is an inner-product space). Then

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|.$$

Proof. We have

$$\begin{aligned} \|\vec{v} + \vec{w}\|^2 &= \langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle \\ &= \langle \vec{v}, \vec{v} \rangle + 2\langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{w} \rangle \text{ (by Definition 3.12)} \\ &= \|\vec{v}\|^2 + 2\langle \vec{v}, \vec{w} \rangle + \|\vec{w}\|^2 \text{ (by Definition 3.13)} \\ &\leq \|\vec{v}\|^2 + 2\|\vec{v}\|\|\vec{w}\| + \|\vec{w}\|^2 \text{ (by Schwarz Inequality)} \\ &= (\|\vec{v}\| + \|\vec{w}\|)^2 \end{aligned}$$

Taking square roots, we have the Triangle Inequality. *QED*

Definition. Let $\vec{v}, \vec{w} \in V$ where V is an inner-product space. Define the *angle between vectors $\vec{v}$ and $\vec{w}$* as

$$\theta = \arccos \frac{\langle \vec{v}, \vec{w} \rangle}{\|\vec{v}\|\|\vec{w}\|}.$$

In particular, $\vec{v}$ and $\vec{w}$ are *orthogonal* (or *perpendicular*) if $\langle \vec{v}, \vec{w} \rangle = 0$.

4 Determinants

4.1 Areas, Volumes, and Cross Products

Note. Area of a Parallelogram.
Consider the parallelogram determined by two vectors $\vec{a}$ and $\vec{b}$. Its area is

$$A = \text{Area} = (\text{base}) \times (\text{height}) = \|\vec{a}\|\|\vec{b}\| \sin\theta$$

$$= \|\vec{a}\|\|\vec{b}\|\sqrt{1-\cos^2\theta}.$$

Squaring both sides:

$$\begin{aligned} A^2 &= \|\vec{a}\|^2\|\vec{b}\|^2(1-\cos^2\theta) \\ &= \|\vec{a}\|^2\|\vec{b}\|^2 - \|\vec{a}\|^2\|\vec{b}\|^2\cos^2\theta \\ &= \|\vec{a}\|^2\|\vec{b}\|^2 - (\vec{a}\cdot\vec{b})^2. \end{aligned}$$

Converting to components $\vec{a} = [a_1, a_2]$ and $\vec{b} = [b_1, b_2]$ gives

$$A^2 = (a_1b_2 - a_2b_1)^2$$

or $A = |a_1b_2 - a_2b_1|$.

Definition. For a 2×2 matrix $A = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}$, define the *determinant* of A as

$$\det(A) = a_1b_2 - a_2b_1 = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}.$$

Definition. For two vectors $\vec{b} = [b_1, b_2, b_3]$ and $\vec{c} = [c_1, c_2, c_3]$ define the *cross product* of $\vec{b}$ and $\vec{c}$ as

$$\vec{b} \times \vec{c} = \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} \hat{\imath} - \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} \hat{\jmath} + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \hat{k}.$$

Note. We can take dot products and find that $\vec{b} \times \vec{c}$ is perpendicular to both $\vec{b}$ and $\vec{c}$.

Definition. For a 3×3 matrix $A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$ define the *determinant* as

$$\det(A) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}.$$

Note. We can now see that cross products can be computed using determinants:

$$\vec{b} \times \vec{c} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

Theorem. The area of the parallelogram determined by $\vec{b}$ and $\vec{c}$ is $||\vec{b} \times \vec{c}||$.

Proof. We know from the first note of this section that the area squared is $A^2 = ||\vec{c}|| ||\vec{b}|| - (\vec{c} \cdot \vec{b})^2$. In terms of components we have

$$A^2 = (c_1^2 + c_2^2 + c_3^2)(b_1^2 + b_2^2 + b_3^2) - (c_1 b_1 + c_2 b_2 + c_3 b_3)^2.$$

Multiplying out and regrouping we find that

$$A^2 = \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix}^2 + \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix}^2 + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}^2.$$

Taking square roots we see that the claim is verified. *QED*

Theorem. The volume of a box determined by vectors $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$ is

$$V = |a_1(b_2 c_3 - b_3 c_2) - a_2(b_1 c_3 - b_3 c_1) + a_3(b_1 c_2 - b_2 c_1)| = |\vec{a} \cdot \vec{b} \times \vec{c}|.$$

Proof. Consider the box determined by $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$. The volume of the box is the height times the area of the base. The area of the base is $||\vec{b} \times \vec{c}||$ by the previous theorem. Now the height is

$$h = ||\vec{a}|| |\cos\theta| = \frac{||\vec{b} \times \vec{c}|| ||\vec{a}|| |\cos\theta|}{||\vec{b} \times \vec{c}||} = \frac{|(\vec{b} \times \vec{c}) \cdot \vec{a}|}{||\vec{b} \times \vec{c}||}.$$

(Notice that if $\vec{b} \times \vec{c}$ is in the opposite direction as given in the illustration above, then θ would be greater than $\pi/2$ and $\cos\theta$ would be negative. Therefore the absolute value is necessary.) Therefore

$$V = (\text{Area of base})(\text{height}) = \|\vec{b} \times \vec{c}\| \frac{|(\vec{b} \times \vec{c}) \cdot \vec{a}|}{\|\vec{b} \times \vec{c}\|} = |(\vec{b} \times \vec{c}) \cdot \vec{a}|.$$

QED

Note. The volume of a box determined by $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$ can be computed in a similar manner to cross products:

$$V = |\det(A)| = \left| \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \right|.$$

Theorem 4.1. Properties of Cross Product.
Let $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$.
(1) Anticommutivity: $\vec{b} \times \vec{c} = -\vec{c} \times \vec{b}$
(2) Nonassociativity of $\times$: $\vec{a} \times (\vec{b} \times \vec{c}) \neq (\vec{a} \times \vec{b}) \times \vec{c}$ (That is, equality does not in general hold.)
(3) Distributive Properties: $\vec{a} \times (\vec{b} + \vec{c}) = (\vec{a} \times \vec{b}) + (\vec{a} \times \vec{c})$
$(\vec{a} + \vec{b}) \times \vec{c} = (\vec{a} \times \vec{c}) + (\vec{b} \times \vec{c})$
(4) Perpendicular Property: $\vec{b} \cdot (\vec{b} \times \vec{c}) = (\vec{b} \times \vec{c}) \cdot \vec{c} = 0$
(5) Area Property: $\|\vec{b} \times \vec{c}\| =$ Area of the parallelogram determined by $\vec{b}$ and $\vec{c}$
(6) Volume Property: $\vec{a} \cdot (\vec{b} \times \vec{c}) = (\vec{a} \times \vec{b}) \cdot \vec{c} = \pm$Volume of the box determined by $\vec{a}$, $\vec{b}$, and $\vec{c}$.
(7) $\vec{a} \times (\vec{b} \times \vec{c}) = (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c}$

Proof of (1). We have

$$\begin{aligned}
\vec{b} \times \vec{c} &= \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} \hat{\imath} - \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} \hat{\jmath} + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \hat{k} \\
&= (b_2c_3 - b_3c_2)\hat{\imath} - (b_1c_3 - b_3c_1)\hat{\jmath} + (b_1c_2 - b_2c_1)\hat{k} \\
&= -\Big((b_3c_2 - b_2c_3)\hat{\imath} - (b_3c_1 - b_1c_3)\hat{\jmath} + (b_2c_1 - b_1c_2)\hat{k} \Big) \\
&= -\left(\begin{vmatrix} c_2 & c_3 \\ b_2 & b_3 \end{vmatrix} \hat{\imath} - \begin{vmatrix} c_1 & c_3 \\ b_1 & b_3 \end{vmatrix} \hat{\jmath} + \begin{vmatrix} c_1 & c_2 \\ b_1 & b_2 \end{vmatrix} \hat{k} \right) \\
&= -\vec{c} \times \vec{b}
\end{aligned}$$

QED

4.2 The Determinant of a Square Matrix

Definition. The *minor matrix* A_{ij} of an $n \times n$ matrix A is the $(n-1) \times (n-1)$ matrix obtained from it by eliminating the ith row and the jth column.

Definition. The determinant of A_{ij} times $(-1)^{i+j}$ is the *cofactor* of entry a_{ij} in A, denoted a'_{ij}.

Definition 4.1. The *determinant* of a 1×1 matrix is its single entry. Let $n > 1$ and assume the determinants of order less than n have been defined. Let $A = [a_{ij}]$ be an $n \times n$ matrix. The *cofactor* of a_{ij} in A is $a'_{ij} = (-1)^{i+j} \det(A_{ij})$. The *determinant* of A is

$$\det(A) = a_{11}a'_{11} + a_{12}a'_{12} + \cdots + a_{1n}a'_{1n} = \sum_{i=1}^{n} a_{1i}a'_{1i}.$$

Theorem 4.2. General Expansion by Minors.
The determinant of A can be calculated by expanding about any row or column:

$$\begin{aligned} \det(A) &= a_{r1}a'_{r1} + a_{r2}a'_{r2} + \cdots a_{rn}a'_{rn} \\ &= a_{1s}a'_{1s} + a_{2s}a'_{2s} + \cdots a_{ns}a'_{ns} \end{aligned}$$

for any $1 \leq r \leq n$ or $1 \leq s \leq n$.

Proof. Use mathematical induction.

Theorem. Properties of the Determinant.
Let A be a square matrix:

1. $\det(A) = \det(A^T)$.

2. If H is obtained from A by interchanging two rows, then $\det(H) = -\det(A)$.

3. If two rows of A are equal, then $\det(A) = 0$.

4. If H is obtained from A by multiplying a row of A by a scalar r, then $\det(H) = r\det(A)$.

5. If H is obtained from A by adding a scalar times one row to another row, then $\det(H) = \det(A)$.

Proof of 2. We will prove this by induction. The proof is trivial for $n = 2$. Assume that $n > 2$ and that this row interchange property holds for square matrices of size smaller that $n \times n$. Let A be an $n \times n$ matrix and let B be the matrix obtained from A by interchanging the ith row and the rth row. Since $n > 2$, we can choose a kth row for expansion by minors, where $k \notin \{r, i\}$. Consider the cofactors

$$(-1)^{k+j}|A_{kj}| \text{ and } (-1)^{k+j}|B_{kj}|.$$

These numbers must have opposite signs, by our induction hypothesis, since the minor matrices A_{kj} and B_{kj} have size $(n-1) \times (n-1)$, and B_{kj} can be obtained from A_{kj} by interchanging two rows. That is, $|B_{kj}| = -|A_{kj}|$. Expanding by minors on the kth row to find $\det(A)$ and $\det(B)$, we see that $\det(A) = -\det(B)$. *QED*

Note. Property 1 above implies that each property of determinants stated for "rows" also holds for "columns."

Theorem 4.3. Determinant Criterion for Invertibility.
A square matrix A is invertible if and only if $\det(A) \neq 0$.

Theorem 4.5 The Multiplicative Property.
If A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.

4.3 Computation of Determinants and Cramer's Rule

Note. Computation of A Determinant.
The determinant of an $n \times n$ matrix A can be computed as follows:

1. Reduce A to an echelon form using only row (column) addition and row (column) interchanges.

2. If any matrices appearing in the reduction contain a row (column) of zeros, then $\det(A) = 0$.

3. Otherwise,

$$\det(A) = (-1)^r \cdot (\text{product of pivots})$$

where r is the number of row (column) interchanges.

Theorem 4.5. Cramer's Rule.
Consider the linear system $A\vec{x} = \vec{b}$, where $A = [a_{ij}]$ is an $n \times n$ invertible matrix,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

The system has a unique solution given by

$$x_k = \frac{\det(B_k)}{\det(A)} \text{ for } k = 1, 2, \ldots, n,$$

where B_k is the matrix obtained from A by replacing the kth column vector of A by the column vector $\vec{b}$.

Proof. Since A is invertible, we know that the linear system $A\vec{x} = \vec{b}$ has a unique solution by Theorem 1.16. Let $\vec{x}$ be this unique solution. Let X_k be the matrix obtained from the $n \times n$ identity matrix by replacing its kth column vector by the column vector $\vec{x}$, so

$$X_k = \begin{bmatrix} 1 & 0 & 0 & \cdots & x_1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & x_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & x_3 & 0 & 0 & \cdots & 0 \\ & & & & \vdots & & & & \\ 0 & 0 & 0 & \cdots & x_k & 0 & 0 & \cdots & 0 \\ & & & & \vdots & & & & \\ 0 & 0 & 0 & \cdots & x_n & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

We now compute the product AX_k If $j \neq k$, then the jth column of AX_k is the product of A and the jth column of the identity matrix, which is just the jth column of A. If $j = k$, then the jth column of AX_k is $A\vec{x} = \vec{b}$. Thus AX_k is the matrix obtained from A by replacing the kth column of A by the column vector $\vec{b}$. That is, AX_k is the matrix B_k described in the statement of the theorem. From the equation $AX_k = B_k$ and the multiplicative property of determinants, we have

$$\det(A) \cdot \det(X_k) = \det(B_k).$$

Computing $\det(X_k)$ by expanding by minors across the kth row, we see that $\det(X_k) = x_k$ and thus $\det(A) \cdot x_k = \det(B_k)$. Because A is invertible, we know that $\det(A) \neq 0$ by theorem 4.3, and so $x_k = \det(B_k)/\det(A)$ as claimed. *QED*

Note. Recall that a'_{ij} is the determinant of the minor matrix associated with element a_{ij} (i.e. the *cofactor* of a_{ij}).

Definition. For an $n \times n$ matrix $A = [a_{ij}]$, define the *adjoint* of A as

$$\text{adj}(A) = (A')^T$$

where $A' = [a'_{ij}]$.

Theorem 4.6. Property of the Adjoint.
Let A be $n \times n$. Then

$$(\text{adj}(A))A = A\,\text{adj}(A) = (\det(A))\mathcal{I}.$$

Corollary. A Formula for A^{-1}.
Let A be $n \times n$ and suppose $\det(A) \neq 0$. Then

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

Note. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\mathrm{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ and $\det(a) = ad - bc$, so

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

5 Eigenvalues and Eigenvectors

5.1 Eigenvalues and Eigenvectors

Definition 5.1. Let A be an $n \times n$ matrix. A scalar λ is an *eigenvalue* of A if there is a nonzero column vector $\vec{v} \in \mathbb{R}^n$ such that $A\vec{v} = \lambda\vec{v}$. The vector $\vec{v}$ is then an *eigenvector* of A corresponding to λ.

Note. If $A\vec{v} = \lambda\vec{v}$ then $A\vec{v} - \lambda\vec{v} = \vec{0}$ and so $(A - \lambda\mathcal{I})\vec{v} = \vec{0}$. This equation has a nontrivial solution only when $\det(A - \lambda\mathcal{I}) = 0$.

Definition. $\det(A-\lambda\mathcal{I})$ is a polynomial of degree n (where A is $n\times n$) called the *characteristic polynomial* of A, denoted $p(\lambda)$, and the equation $p(\lambda) = 0$ is called the *characteristic equation*.

Theorem 5.1. Properties of Eigenvalues and Eigenvectors.
Let A be an $n \times n$ matrix.

1. If λ is an eigenvalue of A with $\vec{v}$ as a corresponding eigenvector, then λ^k is an eigenvalue of A^k, again with $\vec{v}$ as a corresponding eigenvector, for any positive integer k.

2. If λ is an eigenvalue of an invertible matrix A with $\vec{v}$ as a corresponding eigenvector, then $\lambda \neq 0$ and $1/\lambda$ is an eigenvalue of A^{-1}, again with $\vec{v}$ as a corresponding eigenvector.

3. If λ is an eigenvalue of A, then the set E_λ consisting of the zero vector together with all eigenvectors of A for this eigenvalue λ is a subspace of n-space, the *eigenspace* of λ.

Proof of (2). By definition, $\lambda \neq 0$. If λ is an eigenvalue of A with eigenvector $\vec{v}$, then $A\vec{v} = \lambda\vec{v}$. Therefore $A^{-1}A\vec{v} = A^{-1}\lambda\vec{v}$ or $\vec{v} = \lambda A^{-1}\vec{v}$. So $A^{-1}\vec{v} = (1/\lambda)\vec{v}$ and $1/\lambda$ is an eigenvalue of A^{-1}. *QED*

Definition 5.2. Eigenvalues and Eigenvectors.
Let T be a linear transformation of a vector space V into itself. A scalar λ is an *eigenvalue* of T if there is a nonzero vector $\vec{v} \in V$ such that $T(\vec{v}) = \lambda\vec{v}$. The vector $\vec{v}$ is then an *eigenvector* of T corresponding to λ.

5.2 Diagonalization

Note. In this section, the theorems stated are valid for matrices and vectors with complex entries and complex scalars, unless stated otherwise.

Theorem 5.2. Matrix Summary of Eigenvalues of A.
Let A be an $n \times n$ matrix and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be (possibly complex) scalars and $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ be nonzero vectors in n-space. Let C be the $n \times n$ matrix having $\vec{v_j}$ as jth column vector and let

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Then $AC = CD$ if and only if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A and $\vec{v_j}$ is an eigenvector of A corresponding to λ_j for $j = 1, 2, \ldots, n$.

Proof. We have

$$\begin{aligned} CD &= \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{v_1} & \vec{v_2} & \cdots & \vec{v_n} \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\ &= \begin{bmatrix} \vdots & \vdots & & \vdots \\ \lambda_1\vec{v_1} & \lambda_2\vec{v_2} & \cdots & \lambda_n\vec{v_n} \\ \vdots & \vdots & & \vdots \end{bmatrix}. \end{aligned}$$

Also,

$$AC = A \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{v_1} & \vec{v_2} & \cdots & \vec{v_n} \\ \vdots & \vdots & & \vdots \end{bmatrix}.$$

Therefore, $AC = CD$ if and only if $A\vec{v_j} = \lambda_j\vec{v_j}$. *QED*

Note. The $n \times n$ matrix C is invertible if and only if $\text{rank}(C) = n$ — that is, if and only if the column vectors of C form a basis of n-space. In this case, the criterion $AC = CD$ in Theorem 5.2 can be written as $D = C^{-1}AC$. The equation $D = C^{-1}AC$ transforms a matrix A into a diagonal matrix D that is much easier to work with.

Definition 5.3. Diagonalizable Matrix.
An $n \times n$ matrix A is *diagonalizable* if there exists an invertible matrix C such that $C^{-1}AC = D$ is a diagonal matrix. The matrix C is said to *diagonalize* the matrix A.

Corollary 1. A Criterion for Diagonalization.
An $n \times n$ matrix A is diagonalizable if and only if n-space has a basis consisting of eigenvectors of A.

Corollary 2. Computation of A^k.
Let an $n \times n$ matrix A have n eigenvectors and eigenvalues, giving rise to the matrices C and D so that $AC = CD$, as described in Theorem 5.2. If the eigenvectors are independent, then C is an invertible matrix and $C^{-1}AC = D$. Under these conditions, we have $A^k = CD^kC^{-1}$.

Proof. By Corollary 1, if the eigenvectors of A are independent, then A is diagonalizable and so C is invertible. Now consider

$$
\begin{aligned}
A^k &= \underbrace{(CDC^{-1})(CDC^{-1})\cdots(CDC^{-1})}_{k \text{ factors}} \\
&= CD(C^{-1}C)D(C^{-1}C)D(C^{-1}C)\cdots(C^{-1}C)DC^{-1} \\
&= CD\mathcal{I}D\mathcal{I}D\cdots\mathcal{I}DC^{-1} \\
&= C\underbrace{DDD\cdots D}_{k \text{ factors}}C^{-1} = CD^kC^{-1}
\end{aligned}
$$

QED

Theorem 5.3. Independence of Eigenvectors.
Let A be an $n \times n$ matrix. If $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively, the set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ is linearly independent and A is diagonalizable.

Proof. We prove this by contradiction. Suppose that the conclusion is false and the hypotheses are true. That is, suppose the eigenvectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ are linearly dependent. then one of them is a linear combination of its predecessors (see page 203 number 37). Let $\vec{v_k}$ be the first such vector, so that

$$\vec{v_k} = d_1\vec{v_1} + d_2\vec{v_2} + \cdots + d_{k-1}\vec{v}_{k-1} \qquad (2)$$

and $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v}_{k-1}\}$ is independent. Multiplying (2) by λ_k, we obtain

$$\lambda_k\vec{v_k} = d_1\lambda_k\vec{v_1} + d_2\lambda_k\vec{v_2} + \cdots + d_{k-1}\lambda_k\vec{v}_{k-1}. \qquad (3)$$

Also, multiplying (2) on the left by the matrix A yields

$$\lambda_k\vec{v_k} = d_1\lambda_1\vec{v_1} + d_2\lambda_2\vec{v_2} + \cdots + d_{k-1}\lambda_{k-1}\vec{v}_{k-1} \qquad (4),$$

since $A\vec{v_i} = \lambda_i\vec{v_i}$. Subtracting (4) from (3), we see that

$$\vec{0} = d_1(\lambda_k - \lambda_1)\vec{v_1} + d_2(\lambda_k - \lambda_2)\vec{v_2} + \cdots + d_{k-1}(\lambda_k - \lambda_{k-1})\vec{v}_{k-1}.$$

But this equation is a dependence relation since not all d_i's are 0 and the λ's are hypothesized to be different. This contradicts the linear independence of the set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v}_{k-1}\}$. This contradiction shows that $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ is independent. From Corollary 1 of Theorem 5.2 we see that A is diagonalizable. *QED*

Definition 5.4. An $n \times n$ matrix P is *similar* to an $n \times n$ matrix Q if there exists an invertible $n \times n$ matrix C such that $C^{-1}PC = Q$.

Definition. The *algebraic multiplicity* of an eigenvalue λ_i of A is its multiplicity as a root of the characteristic equation of A. Its *geometric multiplicity* is the dimension of the eigenspace E_{λ_i}.

Theorem. The geometric multiplicity of an eigenvalue of a matrix A is less than or equal to its algebraic multiplicity.

Theorem 5.4. A Criterion for Diagonalization.
An $n \times n$ matrix A is diagonalizable if and only if the algebraic multiplicity of each (possibly complex) eigenvalue is equal to its geometric multiplicity.

Theorem 5.5. Diagonalization of Real Symmetric Matrices.
Every real symmetric matrix is real diagonalizable. That is, if A is an $n \times n$ symmetric real matrix with real-number entries, then each eigenvalue of A is a real number, and its algebraic multiplicity equals its geometric multiplicity.

Note. The proof of Theorem 5.5 is in Chapter 9 and uses the *Jordan canonical form* of matrix A.

6 Orthogonality

6.1 Projections

Note. We want to find the projection $\vec{p}$ of vector $\vec{F}$ on $\text{sp}(\vec{a})$. We see that $\vec{p}$ is a multiple of $\vec{a}$. Now $(1/\|\vec{a}\|)\vec{a}$ is a unit vector having the same direction as $\vec{a}$, so $\vec{p}$ is a scalar multiple of this unit vector. We need only find the appropriate scalar, which is $\|\vec{F}\|\cos\theta$. If $\vec{p}$ is in the opposite direction of $\vec{a}$ and $\theta \in [\pi/2, 3\pi/2]$, then the appropriate scalar is again given by $\|\vec{F}\|\cos\theta$. Thus

$$\vec{p} = \frac{\|\vec{F}\|\cos\theta}{\|\vec{a}\|}\vec{a} = \frac{\|\vec{F}\|\|\vec{a}\|\cos\theta}{\|\vec{a}\|\|\vec{a}\|}\vec{a} = \frac{\vec{F}\cdot\vec{a}}{\vec{a}\cdot\vec{a}}\vec{a}.$$

We use this to motivate the following definition.

Definition. Let $\vec{a}, \vec{b} \in \mathbb{R}^n$ The *projection $\vec{p}$ of $\vec{b}$ on sp($\vec{a}$)* is

$$\vec{p} = \frac{\vec{b}\cdot\vec{a}}{\vec{a}\cdot\vec{a}}\vec{a}.$$

Definition 6.1. Let W be a subspace of $\mathbb{R}^n$. The set of all vectors in $\mathbb{R}^n$ that are orthogonal to every vector in W is the *orthogonal complement* of W and is denoted by $W^\perp$.

Note. To find the orthogonal complement of a subspace of $\mathbb{R}^n$:

1. Find a matrix A having as *row* vectors a generating set for W.

2. Find the nullspace of A — that is, the solution space of $A\vec{x} = \vec{0}$. This nullspace is $W^\perp$.

Theorem 6.1. Properties of $W^\perp$.
The orthogonal complement $W^\perp$ of a subspace W of $\mathbb{R}^n$ has the following properties:

1. $W^\perp$ is a subspace of $\mathbb{R}^n$.

2. $\dim(W^\perp) = n - \dim(W)$.

3. $(W^\perp)^\perp = W$.

4. Each vector $\vec{b} \in \mathbb{R}^n$ can be expressed uniquely in the form $\vec{b} = \vec{b}_W + \vec{b}_{W^\perp}$ for $\vec{b}_W \in W$ and $\vec{b}_{W^\perp} \in W^\perp$.

Proof of (1) and (2). Let $\dim(W) = k$, and let $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ be a basis for W. Let A be the $k \times n$ matrix having $\vec{v_i}$ as its ith row vector for $i = 1, 2, \ldots, k$.

Property (1) follows from the fact that $W^\perp$ is the nullspace of matrix A and therefore is a subspace of $\mathbb{R}^n$.

For Property (2), consider the rank equation of A:

$$\text{rank}(A) + \text{ nullity}(A) = n.$$

Since $\dim(W) = \text{ rank}(A)$ and since $W^\perp$ is the nullspace of A, then $\dim(W^\perp) = n - \dim(W)$. *QED*

Definition 6.2. Let $\vec{b} \in \mathbb{R}^n$, and let W be a subspace of $\mathbb{R}^n$. Let $\vec{b} = \vec{b}_W + \vec{b}_{W^\perp}$ be as described in Theorem 6.1. Then $\vec{b}_W$ is the *projection of* $\vec{b}$ *on* W.

Note. To find the projection of $\vec{b}$ on W, follow these steps:

1. Select a basis $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ for the subspace W.

2. Find a basis $\{\vec{v}_{k+1}, \vec{v}_{k+2}, \ldots, \vec{v}_n\}$ for $W^\perp$.

3. Find the coordinate vector $\vec{r} = [r_1, r_2, \ldots, r_n]$ of $\vec{b}$ relative to the basis $(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n})$ so that

$$\vec{b} = r_1\vec{v_1} + r_2\vec{v_2} + \cdots + r_n\vec{v_n}.$$

4. Then $\vec{b}_W = r_1\vec{v_1} + r_2\vec{v_2} + \cdots + r_k\vec{v_k}$.

Note. We can perform projections in inner product spaces by replacing the dot products in the formulas above with inner products.

Example. Consider the inner product space $\mathcal{P}_{[0,1]}$ of all polynomial functions defined on the interval $[0, 1]$ with inner product

$$\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)\,dx.$$

Find the projection of $f(x) = x$ on $\text{sp}(1)$ and then find the projection of x on $\text{sp}(1)^\perp$.

6.2 The Gram Schmidt Process

Definition. A set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ of nonzero vectors in $\mathbb{R}^n$ is *orthogonal* if the vectors $\vec{v_j}$ are mutually perpendicular — that is, if $\vec{v_i} \cdot \vec{v_j} = 0$ for $i \neq j$.

Theorem 6.2. Orthogonal Bases.
Let $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$. Then this set is independent and consequently is a basis for the subspace $\text{sp}(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k})$.

Proof. Let j be an integer between 2 and k. Consider

$$\vec{v_j} = s_1\vec{v_1} + s_2\vec{v_2} + \cdots + s_{j-1}\vec{v}_{j-1}.$$

If we take the dot product of each side of this equation with $\vec{v_j}$ then, since the set of vectors is orthogonal, we get $\vec{v_j} \cdot \vec{v_j} = 0$, which contradicts the hypothesis that $\vec{v_j} \neq \vec{0}$. Therefore no $\vec{v_j}$ is a linear combination of its predecessors and by Exercise 37 page 203, the set is independent. Therefore the set is a basis for its span. *QED*

Theorem 6.3. Projection Using an Orthogonal Basis.
Let $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$, and let $\vec{b} \in \mathbb{R}^n$. The projection of $\vec{b}$ on W is

$$\vec{b}_W = \frac{\vec{b} \cdot \vec{v_1}}{\vec{v_1} \cdot \vec{v_1}}\vec{v_1} + \frac{\vec{b} \cdot \vec{v_2}}{\vec{v_2} \cdot \vec{v_2}}\vec{v_2} + \cdots + \frac{\vec{b} \cdot \vec{v_k}}{\vec{v_k} \cdot \vec{v_k}}\vec{v_k}.$$

Proof. We know from Theorem 6.1 that $\vec{b} = \vec{b}_W + \vec{b}_{W^\perp}$ where $\vec{b}_W$ is the projection of $\vec{b}$ on W and $\vec{b}_{W^\perp}$ is the projection of $\vec{b}$ on $W^\perp$. Since $\vec{b}_W \in W$ and $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is a basis of W, then

$$\vec{b}_W = r_1\vec{v_1} + r_2\vec{v_2} + \cdots + r_k\vec{v_k}$$

for some scalars $r_1, r_2, \ldots, r_k$. We now find these r_i's. Taking the dot product of $\vec{b}$ with $\vec{v_i}$ we have

$$\begin{aligned}
\vec{b} \cdot \vec{v_i} &= (\vec{b}_W \cdot \vec{v_i}) + (\vec{b}_{W^\perp} \cdot \vec{v_i}) \\
&= (r_1\vec{v_1} \cdot \vec{v_i} + r_2\vec{v_2} \cdot \vec{v_i} + \cdots + r_k\vec{v_k} \cdot \vec{v_i}) + 0 \\
&= r_i\vec{v_i} \cdot \vec{v_i}
\end{aligned}$$

Therefore $r_i = (\vec{b} \cdot \vec{v_i})/(\vec{v_i} \cdot \vec{v_i})$ and so

$$r_i \vec{v_i} = \frac{\vec{b} \cdot \vec{v_i}}{\vec{v_i} \cdot \vec{v_i}} \vec{v_i}.$$

Substituting these values of the r_i's into the expression for $\vec{b}_W$ yields the theorem. *QED*

Definition 6.3. Let W be a subspace of $\mathbb{R}^n$. A basis $\{\vec{q_1}, \vec{q_2}, \ldots, \vec{q_k}\}$ for W is *orthonormal* if

1. $\vec{q_i} \cdot \vec{q_j} = 0$ for $i \neq j$, and

2. $\vec{q_i} \cdot \vec{q_i} = 1$.

That is, each vector of the basis is a unit vector and the vectors are pairwise orthogonal.

Note. If $\{\vec{q_1}, \vec{q_2}, \ldots, \vec{q_k}\}$ is an orthonormal basis for W, then

$$\vec{b}_W = (\vec{b} \cdot \vec{q_1})\vec{q_1} + (\vec{b} \cdot \vec{q_2})\vec{q_2} + \cdots + (\vec{b} \cdot \vec{q_k})\vec{q_k}.$$

Theorem 6.4. Orthonormal Basis (Gram-Schmidt) Theorem.
Let W be a subspace of $\mathbb{R}^n$, let $\{\vec{a_1}, \vec{a_2}, \ldots, \vec{a_k}\}$ be any basis for W, and let

$$W_j = \text{sp}(\vec{a_1}, \vec{a_2}, \ldots, \vec{a_j}) \text{ for } j = 1, 2, \ldots, k.$$

Then there is an orthonormal basis $\{\vec{q_1}, \vec{q_2}, \ldots, \vec{q_k}\}$ for W such that $W_j = \text{sp}(\vec{q_1}, \vec{q_2}, \ldots, \vec{q_j})$.

Note. The proof of Theorem 6.4 is computational. We summarize the proof in the following procedure:

Gram-Schmidt Process.
To find an orthonormal basis for a subspace W of $\mathbb{R}^n$:

1. Find a basis $\{\vec{a_1}, \vec{a_2}, \ldots, \vec{a_k}\}$ for W.

2. Let $\vec{v_1} = \vec{a_1}$. For $j = 1, 2, \ldots, k$, compute in succession the vector $\vec{v_j}$ given by subtracting from $\vec{a_j}$ its projection on the subspace generated by its predecessors.

3. The $\vec{v_j}$ so obtained form an orthogonal basis for W, and they may be normalized to yield an orthonormal basis.

Note. We can recursively describe the way to find $\vec{v_j}$ as:

$$\vec{v_j} = \vec{a_j} - \left(\frac{\vec{a_j} \cdot \vec{v_1}}{\vec{v_1} \cdot \vec{v_1}} \vec{v_1} + \frac{\vec{a_j} \cdot \vec{v_2}}{\vec{v_2} \cdot \vec{v_2}} \vec{v_2} + \cdots + \frac{\vec{a_j} \cdot \vec{v}_{j-1}}{\vec{v}_{j-1} \cdot \vec{v}_{j-1}} \vec{v}_{j-1} \right).$$

If we normalize the $\vec{v_j}$ as we go by letting $\vec{q_j} = (1/\|\vec{v_j}\|)\vec{v_j}$, then we have

$$\vec{v_j} = \vec{a_j} - ((\vec{a_j} \cdot \vec{q_1})\vec{q_1} + (\vec{a_j} \cdot \vec{q_2})\vec{q_2} + \cdots + (\vec{a_j} \cdot \vec{q}_{j-1})\vec{q}_{j-1}).$$

Corollary 2. Expansion of an Orthogonal Set to an Orthogonal Basis.
Every orthogonal set of vectors in a subspace W of $\mathbb{R}^n$ can be expanded if necessary to an orthogonal basis of W.

6.3 Orthogonal Matrices

Definition 6.4. An $n \times n$ matrix A is *orthogonal* if $A^T A = \mathcal{I}$.

Note. We will see that the columns of an orthogonal matrix must be unit vectors and that the columns of an orthogonal matrix are mutually orthogonal (inspiring a desire to call them *orthonormal matrices*, but this is not standard terminology).

Theorem 6.5. Characterizing Properties of an Orthogonal Matrix. Let A be an $n \times n$ matrix. The following conditions are equivalent:

1. The rows of A form an orthonormal basis for $\mathbb{R}^n$.

2. The columns of A form an orthonormal basis for $\mathbb{R}^n$.

3. The matrix A is orthogonal — that is, A is invertible and $A^{-1} = A^T$.

Proof. Suppose the columns of A are vectors $\vec{a_1}, \vec{a_2}, \ldots, \vec{a_n}$. Then A is orthogonal if and only if

$$\mathcal{I} = A^T A = \begin{bmatrix} \cdots & \vec{a_1} & \cdots \\ \cdots & \vec{a_2} & \cdots \\ & \vdots & \\ \cdots & \vec{a_n} & \cdots \end{bmatrix} \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{a_1} & \vec{a_2} & \cdots & \vec{a_n} \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

and we see that the diagonal entries of the product are $\vec{a_j} \cdot \vec{a_j} = 1$ therefore each vector is a unit vector. All off-diagonal entries of $\mathcal{I}$ are 0 and so for $i \neq j$ we have $\vec{a_i} \cdot \vec{a_j} = 0$. Therefore the columns of A are orthonormal (and conversely if the columns of A are orthonormal then $A^T A = \mathcal{I}$). Now $A^T = A^{-1}$ if and only if A is orthogonal, so A is orthogonal if and only if $AA^T = \mathcal{I}$ or $(A^T)^T A^T = \mathcal{I}$. So A is orthogonal if and only if A^T is orthogonal, and hence the rows of A are orthonormal if and only if A is orthogonal. *QED*

Theorem 6.6. Properties of $A\vec{x}$ for an Orthogonal Matrix A.
Let A be an orthogonal $n \times n$ matrix and let $\vec{x}$ and $\vec{y}$ be any column vectors in $\mathbb{R}^n$. Then

1. $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$,

2. $\|A\vec{x}\| = \|\vec{x}\|$, and

3. The angle between nonzero vectors $\vec{x}$ and $\vec{y}$ equals the angle between $A\vec{x}$ and $A\vec{y}$.

Proof. Recall that $\vec{x} \cdot \vec{y} = (\vec{x}^T)\vec{y}$. Then since A is orthogonal,

$$[(A\vec{x}) \cdot (A\vec{y})] = (A\vec{x})^T A\vec{y} = \vec{x}^T A^T A\vec{y} = \vec{x}^T \mathcal{I}\vec{y} = \vec{x}^T\vec{y} = [\vec{x} \cdot \vec{y}]$$

and the first property is established.

For the second property,

$$\|A\vec{x}\| = \sqrt{A\vec{x} \cdot A\vec{x}} = \sqrt{\vec{x} \cdot \vec{x}} = \|\vec{x}\|.$$

Since dot products and norms are preserved under multiplication by A, then the angle

$$\cos^{-1}\left(\frac{\vec{x} \cdot \vec{y}}{\sqrt{\vec{x} \cdot \vec{x}}\sqrt{\vec{y} \cdot \vec{y}}}\right) = \cos^{-1}\left(\frac{(A\vec{x}) \cdot (A\vec{y})}{\sqrt{(A\vec{x}) \cdot (A\vec{x})}\sqrt{(A\vec{y}) \cdot (A\vec{y})}}\right).$$

QED

Theorem 6.7. Orthogonality of Eigenspaces of a Real Symmetric Matrix.
Eigenvectors of a real symmetric matrix that correspond to different eigenvalues are orthogonal. That is, the eigenspaces of a real symmetric matrix are orthogonal.

Proof. Let A be an $n \times n$ symmetric matrix, and let $\vec{v_1}$ and $\vec{v_2}$ be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2, respectively. Then

$$A\vec{v_1} = \lambda_1\vec{v_1} \text{ and } A\vec{v_2} = \lambda_2\vec{v_2}.$$

We need to show that $\vec{v_1}$ and $\vec{v_2}$ are orthogonal. Notice that

$$\lambda_1(\vec{v_1} \cdot \vec{v_2}) = (\lambda_1\vec{v_1}) \cdot \vec{v_2} = (A\vec{v_1}) \cdot \vec{v_2} = (A\vec{v_1})^T\vec{v_2} = (\vec{v_1}^T A^T)\vec{v_2}.$$

Similarly

$$[\lambda_2(\vec{v_1} \cdot \vec{v_2})] = \vec{v_1}^T A\vec{v_2}.$$

Since A is symmetric, then $A = A^T$ and

$$\lambda_1(\vec{v_1} \cdot \vec{v_2}) = \lambda_2(\vec{v_1} \cdot \vec{v_2}) \text{ or } (\lambda_1 - \lambda_2)(\vec{v_1} \cdot \vec{v_2}) = 0.$$

Since $\lambda_1 - \lambda_2 \neq 0$, then it must be the case that $\vec{v_1} \cdot \vec{v_2} = 0$ and hence $\vec{v_1}$ and $\vec{v_2}$ are orthogonal. *QED*

Theorem 6.8. Fundamental Theorem of Real Symmetric Matrices. Every real symmetric matrix A is diagonalizable. The diagonalization $C^{-1}AC = D$ can be achieved by using a real orthogonal matrix C.

Proof. By Theorem 5.5, matrix A has only real roots of its characteristic polynomial and the algebraic multiplicity of each eigenvalue is equal to its geometric multiplicity. Therefore we can find a basis for $\mathbb{R}^n$ which consists of eigenvectors of A. Next, we can use the Gram-Schmidt process to create an orthonormal basis for each eigenspace. We know by Theorem 6.7 that the basis vectors from different eigenspaces are perpendicular, and so we have a basis of mutually orthogonal eigenvectors of unit length. As in Section 5.2, we make matrix C by using these unit eigenvectors as columns and we have that $C^{-1}AC = D$ where D consists of the eigenvalues of A. Since the columns of C form an orthonormal set, matrix C is a real orthogonal matrix, as claimed. *QED*

Note. The converse of Theorem 6.8 is also true. If $D = C^{-1}AC$ is a diagonal matrix and C is an orthogonal matrix, then A is symmetric (see Exercise 24). The equation $D = C^{-1}AC$ is called the *orthogonal diagonalization* of A.

Definition 6.5. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is *orthogonal* if it satisfies $T(\vec{v}) \cdot T(\vec{w}) = \vec{v} \cdot \vec{w}$ for all $\vec{v}, \vec{w} \in \mathbb{R}^n$.

Theorem 6.9. Orthogonal Transformations vis-à-vis Matrices. A linear transformation T of $\mathbb{R}^n$ into itself is orthogonal if and only if its standard matrix representation A is an orthogonal matrix.

Proof. By definition, T preserves dot products if and only if it is orthogonal, and so its standard matrix A must preserve dot products and so by Theorem 6.5 A is orthogonal. Conversely, we know that the columns of A are $T(\vec{e_1}), T(\vec{e_2}), \ldots, T(\vec{e_n})$ where $\vec{e_j}$ is the jth unit coordinate vector of $\mathbb{R}^n$, by Theorem 3.10. We have

$$T(\vec{e_i}) \cdot T(\vec{e_j}) = \vec{e_i} \cdot \vec{e_j} = \begin{cases} 0 & \text{if} \quad i \neq j, \\ 1 & \text{if} \quad i = j \end{cases}$$

and so the columns of A form an orthonormal basis of $\mathbb{R}^n$. So A is an orthogonal matrix. *QED*

7 Change of Basis

7.1 Coordinatization and Change of Basis

Recall. Let $B = \{\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n}\}$ be an ordered basis for a vector space V. Recall that if $\vec{v} \in V$ and $\vec{v} = r_1\vec{b_1} + r_2\vec{b_2} + \cdots + r_n\vec{b_n}$, then the coordinate vector of $\vec{v}$ relative to B is $\vec{v}_B = [r_1, r_2, \ldots, r_n]$.

Definition. Let M_B be the matrix having the vectors in the ordered basis B as column vectors. This is the *basis matrix* for B:

$$M_B = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{b_1} & \vec{b_2} & \cdots & \vec{b_n} \\ \vdots & \vdots & & \vdots \end{bmatrix}.$$

Note. We immediately have that $M_B\vec{v}_B = \vec{v}$. If B' is another ordered basis of $\mathbb{R}^n$, then similarly $M_{B'}\vec{v}_{B'} = \vec{v}$ and so $\vec{v} = M_{B'}\vec{v}_{B'} = M_B\vec{v}_B$. Since the columns of $M_{B'}$ are basis vectors for $\mathbb{R}^n$ (and so independent), then

$$\vec{v}_{B'} = M_{B'}^{-1}M_B\vec{v}_B.$$

Notice that this equation gives a relationship between the expression of $\vec{v}$ relative to basis B and the expression of $\vec{v}$ relative to basis B'. We can define $C_{B,B'} = M_{B'}^{-1}M_B$ and then C can be used to convert $\vec{v}_B$ into $\vec{v}_{B'}$ by multiplication: $\vec{v}_{B'} = C_{B,B'}\vec{v}_B$. We can show that $C = M_{B'}^{-1}M_B$ is the unique matrix which can accomplish this conversion (see Page 48 number 41(b)).

Definition 7.1. Let B and B' be ordered bases for a finite dimensional vector space V. The *change-of-coordinates matrix* from B to B' is the unique matrix $C_{B,B'}$ such that $C_{B,B'}\vec{v}_B = \vec{v}_{B'}$.

Note. Of course we can convert $\vec{v}_{B'}$ to $\vec{v}_B$ using C^{-1}: $\vec{v}_B = C^{-1}\vec{v}_{B'}$. In terms of the change-of-coordinates matrix, we have $C_{B',B} = C_{B,B'}^{-1}$.

Note. Finding the Change-of-Coordinates Matrix from B to B' in $\mathbb{R}^n$.
Let $B = (\vec{b_1}, \vec{b_2}, \ldots, \vec{b_n})$ and $B' = (\vec{b_1}', \vec{b_2}', \ldots, \vec{b_n}')$ be ordered bases of $\mathbb{R}^n$. The change-of-coordinates matrix from B to B' is the matrix $C_{B,B'}$ obtained

by the row reduction

$$\left[\begin{array}{cccc|cccc} \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vec{b_1}' & \vec{b_2}' & \cdots & \vec{b_n}' & \vec{b_1} & \vec{b_2} & \cdots & \vec{b_n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \end{array}\right] \sim [\mathcal{I} \mid C_{B,B'}].$$

Note. Recall that the coordinate vector $(\vec{b}_j)_{B'}$ of $\vec{b}_j$ relative to B' is found by reducing the augmented matrix $[M_{B'} \mid \vec{b}_j]$. So all n coordinate vectors $(\vec{b}_j)_{B'}$ can be found at once by reducing the augmented matrix $[M_{B'} \mid M_B]$. Therefore

$$C_{B,B'} = \left[\begin{array}{cccc} \vdots & \vdots & & \vdots \\ (\vec{b_1})_{B'} & (\vec{b_2})_{B'} & \cdots & (\vec{b_n})_{B'} \\ \vdots & \vdots & & \vdots \end{array}\right].$$

7.2 Matrix Representations and Similarity

Theorem 7.1. Similarity of Matrix Representations of T.
Let T be a linear transformation of a finite-dimensional vector space V into itself, and let B and B' be ordered bases of V. Let R_B and $R_{B'}$ be the matrix representations of T relative to B and B', respectively. Then

$$R_{B'} = C^{-1} R_B C$$

where $C = C_{B',B}$ is the change-of-coordinates matrix from B' to B. Hence, $R_{B'}$ and R_B are similar matrices.

Theorem. Significance of the Similarity Relationship for Matrices.
Two $n \times n$ matrices are similar if and only if they are matrix representations of the same linear transformation T relative to suitable ordered bases.

Proof. Theorem 7.1 shows that matrix representations of the same transformation relative to different bases are similar. Now for the converse. Let A be an $n \times n$ matrix representing transformation T, and let F be similar to A, say $F = C^{-1}AC$. Since C is invertible, its columns are independent and form a basis for $\mathbb{R}^n$. Let B be the ordered basis having as jth vector the jth column vector of C. Then C is the change-of-coordinates matrix from B to the standard ordered basis E. That is, $C = C_{B,E}$. Therefore $F = C^{-1}AC = C_{E,B}AC_{B,E}$ is the matrix representation of T relative to basis B. *QED*

Note. Certain properties of matrices are independent of the coordinate system in which they are expressed. These properties are called *coordinate-independent*. For example, we will see that the eigenvalues of a matrix are coordinate-independent quantities.

Theorem 7.2. Eigenvalues and Eigenvectors of Similar Matrices.
Let A and R be similar $n \times n$ matrices, so that $R = C^{-1}AC$ for some invertible $n \times n$ matrix C. Let the eigenvalues of A be the (not necessarily distinct) numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$.

1. The eigenvalues of R are also $\lambda_1, \lambda_2, \ldots, \lambda_n$.

2. The algebraic and geometric multiplicity of each λ_i as an eigenvalue of A remains the same as when it is viewed as an eigenvalue of R.

3. If $\vec{v_i} \in \mathbb{R}^n$ is an eigenvector of the matrix A corresponding to λ_i, then $C^{-1}\vec{v_i}$ is an eigenvector of the matrix R corresponding to λ_i.

Proof of (1). The characteristic equation for matrix R is $\det(R - \lambda\mathcal{I})$ and so

$$\begin{aligned}
\det(R - \lambda\mathcal{I}) &= \det(C^{-1}AC - \lambda\mathcal{I}) \\
&= \det(C^{-1}AC - \lambda C^{-1}C) \\
&= \det(C^{-1}(A - \lambda\mathcal{I})C) \\
&= \det(C^{-1})\det(A - \lambda\mathcal{I})\det(C) \text{ by Theorem 4.4} \\
&= \frac{1}{\det(C)}\det(A - \lambda\mathcal{I})\det(C) \text{ by Page 262 number 31} \\
&= \det A - \lambda\mathcal{I}.
\end{aligned}$$

Therefore the characteristic equation of R and A are the same, and so R and A have the same eigenvalues. *QED*

Definition. The *geometric multiplicity* of an eigenvalue λ of a transformation T is the dimension of the eigenspace $E_\lambda = \{\vec{v} \in V \mid T(\vec{v}) = \lambda\vec{v}\}$. The *algebraic multiplicity* λ is the algebraic multiplicity of the λ as a root of the characteristic polynomial of T (technically, the characteristic polynomial of the matrix which represents T).

Definition 7.2. A linear transformation T of a finite-dimensional vector space V into itself is *diagonalizable* if V has an ordered basis consisting of eigenvectors of T.