2009 Fall MAS109

Introduction to Linear Algebra

X-Type

January 8, 2010

Contents

1 Systems of Linear Equations **3**
1.1 Introduction to Systems of Linear Equations . . . 3
1.2 Solving Linear Systems by Row Reduction . . . 4

2 Matrices and Matrix Algebra **7**
2.1 Operations on Matrices . . . 7
2.2 Inverses; Algebraic Properties of Matrices . . . 8
2.3 Elementary Matrices; A Method for Finding A^{-1} . . . 10
2.4 Subspaces and Linear Independence . . . 11
2.5 The Geometry of Linear Systems . . . 13
2.6 Matrices with Special Forms . . . 14
2.7 Matrix Factorizations; LU-Decomposition . . . 16
2.8 Partitioned Matrices . . . 17

3 Determinants **19**
3.1 Determinants; Cofactor Expansion . . . 19
3.2 Properties of Determinants . . . 20
3.3 Cramer's Rule; Formula for A^{-1}; Applications of Determinants . 21
3.4 A First Look at Eigenvalues and Eigenvectors . . . 22

4 Linear Transformations **25**
4.1 Matrices as Transformations . . . 25
4.2 Geometry of Linear Operations . . . 26
4.3 Kernel and Range . . . 28
4.4 Composition and Invertibility of Linear Transformations . . . 29

5 Dimension and Structure **30**
5.1 Basis and Dimension . . . 30
5.2 Properties of Bases . . . 31
5.3 The Fundamental Spaces of a Matrix . . . 32
5.4 The Dimension Theorem and Its Implications . . . 33
5.5 The Rank Theorem and Its Implications . . . 35
5.6 The Pivot Theorem and Its Implications . . . 36
5.7 The Projection Theorem and Its Implications . . . 39
5.8 Best Approximation and Least Squares . . . 41
5.9 Orthonormal Bases and the Gram-Schmidt Process . . . 43
5.10 QR-Decomposition; Householder Transformations . . . 44
5.11 Coordinates with Respect to a Basis . . . 46

6 Diagonalization **49**
6.1 Matrix Representations of Linear Transformations 49
6.2 Similarity and Diagonalizability 50
6.3 Orthogonal Diagonalizability; Functions of a Matrix 53
6.4 Quadratic Forms 56
6.5 Singular Value Decomposition 58
6.6 The Pseudoinverse 60

Chapter 1

Systems of Linear Equations

1.1 Introduction to Systems of Linear Equations

1.1.1 Linear Systems and Equations

- A system of linear equations with m equations, n unknowns

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

- Solution set $S = \{(x_1, \cdots, x_n) \in \mathbb{R}^n \mid \text{The system holds}\}$

- Geometrically, S is the intersection of m hyperplanes.

Theorem 1.1.1. There are only three possible cases for the cardinality of the solution set S.

$$\begin{cases} |S| = 0 & \text{inconsistent.} \\ |S| = 1 & \text{consistent.} \\ |S| = \infty & \text{consistent, but has infinitely many solutions.} \end{cases}$$

Sketch of Proof. We can show by examples that there is a system of linear equations which is inconsistent or has unique solution. Also, if we have two (or three, whatever) distinct solutions, we can show that any linear combinations of them are also the solutions. In this case, we have infinitely many solutions. □

1.1.2 Augumented Matrices

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

1.1.3 Elementary Row Operations

i. Multiplying a row by a nonzero constant.

ii. Interchaging two rows.

iii. Adding a multiple of a row to another row.

Note that a row represents the coefficients of a linear equation. Elementary row operations do not change the solution set.

1.1.4 Linear System as Linear Combinations of Vectors

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 3 & 1 \end{bmatrix} \xrightarrow{\text{elementary row operations}} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\therefore (1,2,1) = 2(1,1,1) - 2(1,1,2) + (1,2,3)$$

1.2 Solving Linear Systems by Row Reduction

1.2.1 Reduced Row Echelon Form

$$\begin{bmatrix} 0 & 0 & \mathit{1} & * & 0 & 0 & * & \cdots & * \\ 0 & 0 & 0 & 0 & \mathit{1} & 0 & * & \cdots & * \\ 0 & 0 & 0 & 0 & 0 & \mathit{1} & * & \cdots & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

The matrix shown above is in reduced row echelon form.[1] To be of this form a matrix must have the following properties:

1. *Nonzero rows have leading 1.* If a row does not consist entirely of zeros, then the first nonzero number in the row is a 1. This 1 is called a leading 1.

2. *Zero rows are bottommost.* If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.

3. *Leading 1's are in echelon form.* In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.

4. Each column that contains a leading 1 has zeros everywhere else.

A matrix that has the first three properties is said to be in row echelon form.

- General solutions as linear combinations of column vectors $\rightarrow$ leading variables, free variables.

- Linear System $\rightarrow$ Augumented Matrix $\xrightarrow{\text{elem. row. op.}}$ Reduced Row Echelon Form $\rightarrow$ Solution Set.

[1] Leading 1's are in italic and $*$ denotes any number.

1.2.2 Gauss-Jordan Elimination

1. Forward Phase (Gaussian Elimination)

$$\begin{bmatrix} 1 & \cdots \\ * & \cdots \\ * & \cdots \\ * & \cdots \end{bmatrix} \leftrightarrow \left[\begin{array}{cc|cc} 1 & * & * & \cdots \\ \hline 0 & 0 & 1 & \cdots \\ 0 & 0 & * & \cdots \\ 0 & 0 & * & \cdots \end{array}\right] \longrightarrow \begin{bmatrix} 1 & * & * & * & * & * \\ 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

   1. Locate the leftmost nonzero columns.
   2. Interchange the two row with another row, if necessary, to bring a nonzero entry to the top of the column found in 1.
   3. Multiply the first row in order to introduce a leading 1.
   4. Make all entires below the leading 1 become zeros with elementary row operations.
   5. Repeat 1∼4 without the nonzero column found in 1.

   The resulting matrix is in row echelon form, so you can easily read the solution here by back substitution.

2. Backward Phase

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

   1. Beginning with the last nonzero row and working upward, adding suitable multiples of each row to the rows above to introduce zeros above the leading 1's.

1.2.3 Homogeneous Linear Systems

The augumented matrices of homogeneous linear systems are given by

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & 0 \end{bmatrix}$$

Theorem 1.2.1. There are only two possible cases for the cardinality of the solution set S of a homogeneous linear system.

$$\begin{cases} |S| = 1 & \text{has only trivial solution.} \\ |S| = \infty & \text{has infinitely many solutions.} \end{cases}$$

Sketch of Proof. Refer to Theorem 1.1.1. But we always have the *trivial solution*, $(x_1, x_2, \cdots, x_n) = \mathbf{0}$. □

Theorem 1.2.2 (The Dimension Theorem for H.L.S.)**.** If a homogeneous linear system has n unknowns and the row echelon form of its augumented matrix has r nonzero rows, then the system has $n - r$ free variables.

Sketch of Proof. You can start with the row echelon form and do back substitution. □

Corollary. If a homogeneous linear system has more unknowns than equations, then it has infinitely many solutions.

Chapter 2

Matrices and Matrix Algebra

2.1 Operations on Matrices

1. Equality: $A = B \Leftrightarrow (A)_{ij} = (B)_{ij}$
2. Sum, Difference: $(A + B)_{ij} = (A)_{ij} + (B)_{ij}$
3. Scalar Product: $(cA)_{ij} = c(A)_{ij}$
4. Matrix Product

   $$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

   From $(AB)\mathbf{x} = A(B\mathbf{x})$, we have

   $$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_n \end{bmatrix}$$

   where the column vectors of B are $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n$.
   If A is a $p \times q$ matrix and B is a $q \times r$ matrix, then AB is $p \times r$ matrix.

   Theorem 2.1.1 (The Row-Column Rule)**.**

   $$(AB)_{ij} = \mathbf{r}_i(A) \cdot \mathbf{c}_j(B)$$

   Corollary (The Row Rule and the Column Rule)**.**

   $$\mathbf{r}_i(AB) = \mathbf{r}_i(A)B \quad \mathbf{c}_j(AB) = A\mathbf{c}_j(B)$$

5. Transpose: $(A^T)_{ij} = (A)_{ji}$
6. Trace: If A is a square matrix, then the trace of A, deonoted by $\mathrm{tr}(A)$, is defined to be the sum of the entires on the main diagonal of A.
7. Inner and Outer Matrix Products: For two column vectors $\mathbf{u}$ and $\mathbf{v}$,

- Matrix Inner Product: $\mathbf{u}^T\mathbf{v}$
- Matrix Outer Product: $\mathbf{u}\mathbf{v}^T$
- $\mathbf{u}^T\mathbf{v} = \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} = \mathbf{v}^T\mathbf{u}$
- $\mathbf{u}^T\mathbf{v} = \mathrm{tr}(\mathbf{u}\mathbf{v}^T) = \mathrm{tr}(\mathbf{v}\mathbf{u}^T)$

2.2 Inverses; Algebraic Properties of Matrices

2.2.1 Properties of Matrix Operations

Theorem 2.2.1.

i. $A + B = B + A$

ii. $A + (B + C) = (A + B) + C$

iii. $(ab)A = a(bA)$

iv. $(a + b)A = aA + bA$

v. $a(A + B) = aA + aB$

Theorem 2.2.2.

i. $A(BC) = (AB)C$

ii. $A(B + C) = AB + AC$

iii. $(B + C)A = BA + CA$

iv. $a(BC) = (aB)C = B(aC)$

2.2.2 Zero Matrices

A matrix whose entires are all zero is called a zero matrix.

2.2.3 Identity Matrices

A square matrix with 1's on the main diagonal and zeros elsewhere is called an identity matrix.

Theorem 2.2.3. If R is the reduced row echelon form of an $n \times n$ matrix A, then either R has a row of zeros or R is the identity matrix I_n.

Sketch of Proof. Let us start with the row echelon form of A. If it does not have any row of zeros, then every column of it has leading 1. This forces R to be I_n. Otherwise, R has a row of zeros. $\square$

2.2.4 Inverse of a Matrix

Definition 2.2.1. A square matrix A of order n is said to be *invertible* (or *nonsignular*) if there exists a matrix A^{-1} (called the *inverse* of A) such that $AA^{-1} = A^{-1}A = I_n$. Otherwise, A is called *singular*.

Theorem 2.2.4. An invertible matrix has a unique inverse.

Sketch of Proof. Let B and C be inverses of A. From the definition of the inverses, $BA = I, AC = I$. Then, $C = BAC = B(AC) = B$. □

Theorem 2.2.5.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \exists\, A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \text{ iff } ad - bc \neq 0$$

Theorem 2.2.6. If $A_1, A_2, \cdots A_n$ are all invertible, then $A_1A_2\cdots A_n$ is also invertible and

$$(A_1A_2\cdots A_n)^{-1} = A_n^{-1}A_{n-1}^{-1}\cdots A_1^{-1}$$

Theorem 2.2.7. A square matrix with a row or column of zeros is singular.

Sketch of Proof. Any matrix products with such square matrix cannot form the identity matrix. □

Theorem 2.2.8. If A is invertible and n is a nonnegative integer, then:

i. A^{-1} is invertible and $\left(A^{-1}\right)^{-1} = A$.

ii. A^n is invertible and $(A^n)^{-1} = \left(A^{-1}\right)^n$.

iii. kA is invertible for any nonzero scalar k, and $(kA)^{-1} = k^{-1}A^{-1}$.

2.2.5 Properties of the Transpose

Theorem 2.2.9.

i. $(A^T)^T = A$

ii. $(kA)^T = kA^T$

iii. $(A+B)^T = A^T + B^T$

iv. $(AB)^T = B^TA^T$

v. $(A^T)^{-1} = (A^{-1})^T$ if A is invertible.

Sketch of Proof for (iv). Use the relationship $\mathbf{r}_i(A)\mathbf{c}_j(B) = \mathbf{r}_j(B^T)\mathbf{c}_i(A^T)$. □

2.2.6 Properties of the Trace

Theorem 2.2.10. If A and B are square matrices with the same size, then:

i. $\text{tr}(A^T) = \text{tr}(A)$

ii. $\text{tr}(cA) = c\,\text{tr}(A)$

iii. $\text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)$

iv. $\text{tr}(AB) = \text{tr}(BA)$

Sketch of Proof for (iv). Use 7 and Theorem 2.1.1 in section 2.1. □

2.3 Elementary Matrices; A Method for Finding A^{-1}

2.3.1 Elementary Matrices

Definition 2.3.1. Elementary matrix is a matrix that results from applying a *single* elementary row operation to an identity matrix.

Theorem 2.3.1. An elementary row operation X corresponds to the left product with the corresponding elementary matrix E. [1]

$$I_n \xrightarrow{X} E \Rightarrow A \xrightarrow{X} EA$$

Therefore, if $A \xrightarrow{E_1} A_1 \xrightarrow{E_2} \cdots \xrightarrow{E_k} A_k$, then $A_k = E_k E_{k-1} \cdots E_2 E_1 A$.

Theorem 2.3.2. An elementary matrix is invertible, and the inverse is also an elementary matrix.

Definition 2.3.2. Two matrices that can be obtained from one another by finite sequences of elementary row operations are said to be *row equivalent*.

Theorem 2.3.3 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. The reduced row echelon form of A is I_n.

ii. A is expressible as a product of elementary matrices.

iii. A is invertible.

Sketch of Proof. Use Theorem 2.2.3. □

We can see by Theorem 2.3.3 that an invertible matrix is row equivalent to the identity matrix.

2.3.2 Inversion Algorithm

For an invertible $n \times n$ matrix A, A is row equivalent to I_n. Therefore we have

$$E_k E_{k-1} \cdots E_2 E_1 A = I_n$$
$$E_k E_{k-1} \cdots E_2 E_1 I_n = A^{-1}$$

So we establish the following theorem.

Theorem 2.3.4. For an invertible $n \times n$ matrix A,

$$[A|I] \xrightarrow{\text{elem. row. op.}} [I|A^{-1}]$$

1. Solving Linear Systems: $A\mathbf{x} = \mathbf{b} \Rightarrow \mathbf{x} = A^{-1}\mathbf{b}$

2. Solving Homogeneous Linear Systems: If A is invertible, then the system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$

[1] Also we can think of elementary column operations. An elementary column operation corresponds to the *right* product with the corresponding elementary *column* matrix.

3. Solving Multiple Linear Systems with a Common Coefficient Matrix

$$[A|\mathbf{b}_1|\mathbf{b}_2|\cdots|\mathbf{b}_k] \to [I_n|\mathbf{x}_1|\mathbf{x}_2|\cdots|\mathbf{x}_n]$$

4. The Consistency Problem: For a given matrix A, find all matrices $\mathbf{b}$ for which the linear system $A\mathbf{x} = \mathbf{b}$ is consistent.

Theorem 2.3.5 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. Theorem 2.3.3.

ii. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

iii. $A\mathbf{x} = \mathbf{b}$ is consistent for every vector $\mathbf{b}$ in $\mathbb{R}^n$.

iv. $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every vector $\mathbf{b}$ in $\mathbb{R}^n$.

Sketch of Proof. Let us start with (iv). Let $A\mathbf{x}_1 = \mathbf{e}_1, \cdots, A\mathbf{x}_n = \mathbf{e}_n$. Then,

$$\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

is the inverse of A. Therefore, A is invertible. $\square$

Theorem 2.3.6. If A and B are square matrices such that $AB = I$ *or* $BA = I$, then A and B are both invertible, and each is the inverse of the other.

Sketch of Proof. Without loss of generality, assume $BA = I$. It suffices to prove that A is invertible. Consider the homogeneous system $A\mathbf{x} = \mathbf{0}$. If $\mathbf{x}$ is any solution of this system, then:

$$\mathbf{x} = (BA)\mathbf{x} = B(A\mathbf{x}) = B\mathbf{0} = \mathbf{0}$$

Thus, $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, which proves that A is invertible. $\square$

Corollary. If A and B are square matrices whose product AB is invertible, then A and B are invertible.

2.4 Subspaces and Linear Independence

2.4.1 Subspaces of $\mathbb{R}^n$

Definition 2.4.1. A nonempty set of vectors in $\mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ if it is closed under scalar multiplication and addition, or equivalently, closed under linear combination.

Example.

1. Trivial Subspaces of $\mathbb{R}^n$: zero subspace $\{\mathbf{0}\}$ and entire of $\mathbb{R}^n$
2. All subspaces of $\mathbb{R}^2$ fall into one of three categories: The zero subspace, lines through the origin, and entire $\mathbb{R}^2$
3. All subspaces of $\mathbb{R}^3$ fall into one of four categories: The zero subspace, line through the origin, plane through the origin, and entire $\mathbb{R}^3$

If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$ are vectors in $\mathbb{R}^n$, then the set

$$W = \{\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_m\mathbf{v}_m \,|\, (t_1, t_2, \cdots, t_m) \in \mathbb{R}^m\}$$

is a subspace of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \cdots, \mathbf{v}_m$. Alternatively, W is denoted by

$$W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$$

Theorem 2.4.1. If $A\mathbf{x} = \mathbf{0}$ is a homogeneous linear system with n unknowns, then its solution set is a subspace of $\mathbb{R}^n$.

Theorem 2.4.2. If A is a matrix with n columns, then the solution space of the homogeneous system $A\mathbf{x} = \mathbf{0}$ is entire $\mathbb{R}^n$ iff $A = 0$.

Sketch of Proof. Assume the solution space of the homogeneous system $A\mathbf{x} = \mathbf{0}$ is entire $\mathbb{R}^n$. Then we have

$$A\mathbf{e}_1 = \mathbf{0}, A\mathbf{e}_2 = \mathbf{0}, \cdots, A\mathbf{e}_n = \mathbf{0}$$

Therefore,

$$0 = \begin{bmatrix} A\mathbf{e}_1 & A\mathbf{e}_2 & \cdots & A\mathbf{e}_n \end{bmatrix} = AI_n = A$$

The converse is trivial. $\square$

Corollary. If A and B are matrices with n columns, then $A = B$ iff $A\mathbf{x} = B\mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^n$.

2.4.2 Linear Independence

Definition 2.4.2. A nonempty set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ in $\mathbb{R}^m$ is said to be *linearly independent* if the solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = 0$$

is only $c_1 = c_2 = \cdots = c_n = 0$.

From the definition above, if there exists $\mathbf{v}_k$ which is a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\} \setminus \{\mathbf{v}_k\}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is linearly dependent.

Theorem 2.4.3. S is linearly independent iff $A\mathbf{x} = \mathbf{0}$ has only trivial solution where

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

Theorem 2.4.4. A set with more than n vectors in $\mathbb{R}^n$ is linearly dependent.

Theorem 2.4.5 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. Theorem 2.3.5.

ii. The column vectors of A are linearly independent.

iii. The row vectors of A are linearly independent.

2.4.3 Translated Subspaces

Definition 2.4.3. Suppose that W is a subspace of $\mathbb{R}^n$ and $\mathbf{x}_0$ is a vector in $\mathbb{R}^n$. Then,

$$\mathbf{x}_0 + W = \mathbf{x}_0 + t_1\mathbf{v}_1 + \cdots + t_m\mathbf{v}_m$$

is called a *translated subspace*. (Also called linear manifold or affine flat.)

Note that a translated subspace is not a subspace, since $\mathbf{x}_0 + W$ does not include the origin.

2.5 The Geometry of Linear Systems

2.5.1 The relationship between $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$

Theorem 2.5.1. Suppose that A is an $m \times n$ matrix, and $\mathbf{b}$ is a vector in $\mathbb{R}^m$, and let W be the solution space of $A\mathbf{x} = 0$; $\mathbf{x}_0$ be a solution of $A\mathbf{x} = \mathbf{b}$. [2] Then the solution space of $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}_0 + W$.

Theorem 2.5.2. A nonhomogeneous linear system with more unknowns than equations is either inconsistent or has infinitely many solutions.

2.5.2 Consistency of a linear system

Theorem 2.5.3. A linear system $A\mathbf{x} = \mathbf{b}$ is consistent iff $\mathbf{b}$ is in the column space of A.

Sketch of Proof. Let A be $\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. Then, $A\mathbf{x} = \mathbf{b}$ is equivalent to $\mathbf{b} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$. A subspace spanned by the column vectors of A is called the column space of A. □

2.5.3 Hyperplanes

Definition 2.5.1. The solution of linear equation of the form:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $a_1, a_2, \cdots, a_n$ are all nonzero, is called a hyperplane in $\mathbb{R}^n$.

The hyperplane in definition 2.5.1 is also denoted by $\mathbf{a}^\perp$, since $\mathbf{a} = (a_1, \cdots, a_n)$ is orthogonal to every vector on the hyperplane. Note that the dimension of the hyperplane in $\mathbb{R}^n$ is $n - 1$.

Theorem 2.5.4. The solution space of $A\mathbf{x} = \mathbf{0}$ consists of all vectors in $\mathbb{R}^n$ that are orthogonal to every row vector of A.

[2] Assume $A\mathbf{x} = \mathbf{b}$ is consistent. $\mathbf{x}_0$ is called a particular solution of $A\mathbf{x} = \mathbf{b}$.

2.6 Matrices with Special Forms

2.6.1 Diagonal Matrices

Definition 2.6.1. A diagnonal matrix is a square matrix in which all entries off the main diagonal. Thus, a general $n \times n$ diagonal matrix has the form

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

where $d_1, d_2, \cdots, d_n$ are any real numbers.

Powers of diagonal matrices are easily computed by the following formula:

$$D^k = \begin{bmatrix} d_1^k & 0 & \cdots & 0 \\ 0 & d_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^k \end{bmatrix}$$

where k is an nonnegative integer. This formula is valid for every integer k if $d_1, \cdots, d_n$ are all nonzero.
Also, matrix products by diagonal matrices are given by:

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} d_1a_{11} & d_1a_{12} & \cdots & d_1a_{1m} \\ d_2a_{21} & d_2a_{22} & \cdots & d_2a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_na_{n1} & d_na_{n2} & \cdots & d_na_{mn} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_m \end{bmatrix} = \begin{bmatrix} d_1a_{11} & d_2a_{12} & \cdots & d_ma_{1m} \\ d_1a_{21} & d_2a_{22} & \cdots & d_ma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_1a_{n1} & d_2a_{n2} & \cdots & d_ma_{mn} \end{bmatrix}$$

Thus, to multiply a matrix A on the left (resp. right) by a diagonal matrix D, multiply successive row (resp. column) vectors of A by the successive diagonal entires of D.

2.6.2 Triangular Matrices

Definition 2.6.2. An *upper triangular matrix* is a square matrix $U = [u_{ij}]$ with $u_{ij} = 0$ if $i > j$ and a *lower triangular matrix* is a square matrix $L = [l_{ij}]$ with $l_{ij} = 0$ if $i < j$.

That is, an upper triangular matrix is a square matrix that has zero entries below the main diagonal and a lower triangular matrix is a square matrix that has zero entires above the main diagonal.

Definition 2.6.3. A strictly upper triangular matrix is a square matrix $U = [u_{ij}]$ with $u_{ij} = 0$ if $i \geq j$ and a lower triangular matrix is a square matrix $L = [l_{ij}]$ with $l_{ij} = 0$ if $i \leq j$.

Theorem 2.6.1.

i. The transpose of a lower triangular matrix is upper triangular, and the transpose of an upper triangular matrix is lower triangular.

ii. A product of lower triangular matrices is lower triangular, and a product of upper triangular matrices is upper triangular.

iii. A triangular matrix is invertible iff its diagonal entries are all nonzero.

iv. The inverse of an invertible lower triangular matrix is lower triangular, and the inverse of an invertible upper triangular matrix is upper triangular.

2.6.3 Symmetric and Skew Symmetric Matrices

Definition 2.6.4. A square matrix A is called *symmetric* if $A^T = A$.[3] and *skew-symmetric* if $A^T = -A$.

Theorem 2.6.2. If A and B are symmetric matrices with the same size, then:

i. A^T is symmetric.

ii. $A + B$ and $A - B$ are symmetric.

iii. kA is symmetric for any scalar k.

iv. If A is invertible, then A^{-1} is also symmetric.

Theorem 2.6.3. The product of two symmetric matrices is symmetric iff the matrices commute.

AA^T, A^TA are always symmetric for any $m \times n$ matrix A.

Theorem 2.6.4. If A is a square matrix, then the matrices A, AA^T, A^TA are either all invertible or all singular.

2.6.4 Nilpotent of a Matrix

Definition 2.6.5. A square matrix A is said to be *nilpotent* if there is a positive integer such that $A^k = 0$. The smallest positive power for which $A^k = 0$ is called the *index of nilpotency*.

Theorem 2.6.5. If A is a square matrix, and if A is nilpotent and has index of nilpotency k, then:

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^{k-1}$$

Theorem 2.6.6. Every strictly triangular matrix is nilpotent.

Theorem 2.6.7 (Power Series Representation)**.** If A is an $n \times n$ matrix for which the sum of the absolute values of the entries in each column (or each row) is less than 1, then $I - A$ is invertible and can be expressed as

$$(I - A)^{-1} = I + A + A^2 + \cdots$$

[3]That is, $(A)_{ij} = (A)_{ji}$.

2.7 Matrix Factorizations; LU-Decomposition

2.7.1 LU-Decomposition

Let A be LU. Then $A\mathbf{x} = \mathbf{b}$ becomes $LU\mathbf{x} = \mathbf{b}$. Substituting with $U\mathbf{x} = \mathbf{y}$ results the following,

$$U\mathbf{x} = \mathbf{y}, L\mathbf{y} = \mathbf{b}$$

which is called the method of *LU-decomposition*. Thus, obtaining the solution of $A\mathbf{x} = \mathbf{b}$ is decomposed to:

1. Solving $L\mathbf{y} = \mathbf{b}$ with respect to $\mathbf{y}$ by forward substitution.
2. Solving $U\mathbf{x} = \mathbf{y}$ with respect to $\mathbf{x}$ by backward substitution.

Theorem 2.7.1. If a square matrix A can be reduced to row echelon form by Gaussian elimination with no row changes, then A has an LU-decomposition.

An algorithm for LU-decomposition is described as follows:

1. Reduce A to row echelon form U without using row interchanges, keeping track of the multipliers used to introduce the leading 1's and the multipliers used to introduce zero's below the leading 1's.
2. In each position along (resp. below) the main diagonal of L, place the reciprocal (resp. negative) of the multiplier that introduced the leading 1 in that position in U.

Example.

$$\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix} \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 2 & 0 & 0 \\ 3 & * & 0 \\ * & * & * \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 1 \\ 3 & 7 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 3 & * & 0 \\ 1 & * & * \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 8 & 5 \end{bmatrix} \rightarrow$$

$$\therefore L = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 4 & 1 \end{bmatrix}, \; U = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that we used *pivots* instead of leading 1's in U.

2.7.2 LDU-Decomposition

In addition, L can be decomposed to $L'D$ where D is a diagonal matrix. The columns of L' are obtained by dividing each entry in the column by the diagonal entry in the column.

2.7.3 PLU-Decomposition

Definition 2.7.1. A matrix obtained by reordering the rows of I_n in some way is called a *permutation* matrix.

In the case where A cannot reduced to a row echelon form without row interchanges, we can reduce $P^{-1}A$ to a row echelon form U without row interchanges if we choose a suitable permutation matrix P. This is called the *PLU-decomposition*.

2.8 Partitioned Matrices

2.8.1 Block Multiplication

A matrix can be *partitioned* (subdivided) into *submatrices* (also called blocks) in various ways by inserting lines between selected rows and columns. Operations can be performed on partitioned matrices by treating the submatrices as if they were numerical entries. For example, if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}, \ B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

and if the sizes of the submatrices conform for the required operations, then

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \\ A_{31}B_{11} + A_{32}B_{21} & A_{31}B_{12} + A_{32}B_{22} \end{bmatrix}$$

This procedure is called *block multiplication.*

2.8.2 Block Diagonal Matrices

A partitioned square matrix A is said to be *block diagonal* if the matrices on the main diagonal are square and all matrices off the main diagonal are zero; that is, the matrix is partitioned as

$$A = \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{bmatrix}$$

where $D_1, D_2, \cdots, D_n$ are square matrices.
Powers of block diagonal matrices are easily computed by the following formula:

$$A^k = \begin{bmatrix} D_1^k & 0 & \cdots & 0 \\ 0 & D_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_n^k \end{bmatrix}$$

where k is an nonnegative integer. This formula is valid for every integer k iff $D_1, \cdots, D_n$ are all invertible.

2.8.3 Block Triangular Matrices

A partitioned square matrix A is said to be *block upper triangular* if the matrices on the main diagonal are square and all matrices below the main diagonal are zero; that is, the matrix is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ 0 & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn} \end{bmatrix}$$

where $A_{11}, A_{12}, \cdots, A_{nn}$ are square matrices. The definition of a *block lower triangular matrix* is similar.

Chapter 3

Determinants

3.1 Determinants; Cofactor Expansion

3.1.1 Determinants

The determinant of a 2×2 matrix is given by

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

and the determinant of a 3×3 matrix is given by

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

Generally, the determinant of an $n \times n$ matrix is defined as following:

Definition 3.1.1. Let A be an $n \times n$ matrix. Then,

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} (A)_{i,\sigma(i)}$$

in which:

- S_n is a set of permutation $\{1, 2, \cdots, n\}$, so σ denotes one way to permutate $\{1, 2, \cdots, n\}$.
- $\operatorname{sgn}(\sigma)$ is determined by the minimum number of interchanges of adjacent elements required to put the permutation sgn into the right order $\{1, 2, \cdots, n\}$. If it is even, $\operatorname{sgn}(\sigma) = 1$, otherwise, $\operatorname{sgn}(\sigma) = -1$.

In this definition, $\prod_{i=1}^{n} (A)_{i,\sigma(i)}$ is called an *elementary product*. Note that an elementary product has an entry from each row and column. If the elementary product is multiplied by $\operatorname{sgn}(\sigma)$, it is called the *signed* elementary product.

Theorem 3.1.1. If A is a square matrix with a row or a column of zeros, then $\det(A) = 0$.

Theorem 3.1.2. If A is a triangular matrix, then $\det(A)$ is the product of the entries on the main diagonal.

3.1.2 Cofactor Expansion

Definition 3.1.2. The *minor* M_{ij} of an $n \times n$ matrix A is the determinant of the submatrix that remains when ith row and jth column of A are deleted, and the *cofactor* C_{ij} is defined by $C_{ij} = (-1)^{i+j} M_{ij}$.

Theorem 3.1.3. The determinant of an $n \times n$ matrix A is given by the *cofactor expansions*

i. $\det(A) = (A)_{1i}C_{1i} + (A)_{2i}C_{2i} + \cdots + (A)_{ni}C_{ni}$ (along ith column)

ii. $\det(A) = (A)_{j1}C_{j1} + (A)_{j2}C_{j2} + \cdots + (A)_{jn}C_{jn}$ (along jth row)

Sketch of Proof. This theorem can be proved by mathematical induction. □

3.2 Properties of Determinants

Theorem 3.2.1. If A is a square matrix, then $\det(A) = \det(A^T)$

Sketch of Proof. The *signed* elementary products are the same. □

Theorem 3.2.2. Let A be an $n \times n$ matrix. If B is the matrix that results when

i. a single row or column of A is multiplied by a scalar k, then $\det(B) = k\det(A)$.

ii. two rows or columns of A are interchanged, then $\det(B) = -\det(A)$.

iii. a multiple of one row (resp. one column) of A is added to another row (resp. another column), then $\det(B) = \det(A)$.

Corollary. Let A be an $n \times n$ matrix. $\det(kA) = k^n \det(A)$ for a scalar k.

Corollary. If A has two proportional rows or columns, then $\det(A) = 0$.

Theorem 3.2.3 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. Theorem 2.4.5.

ii. $\det(A) \neq 0$.

Sketch of Proof. Let R be a reduced row echelon form of A. Since R is derived from A by a sequence of elementary row operations, it follows that $\det(A) and \det(R)$ are both zero or both nonzero. Let us assume that $\det(A) \neq 0$. Thus $\det(R) \neq 0$, which implies R does not have any zero rows and hence $R = I_n$ by Theorem 2.2.3. Therefore A is invertible by Theorem 2.3.3. The converse is also true again by Theorem 2.3.3. □

Theorem 3.2.4. If A and B are square matrices of the same size, then $\det(AB) = \det(A) \cdot \det(B)$.

Sketch of Proof.

(i) If A is not invertible, then the Corollary of Theorem 2.3.6 implies that AB is not invertible.

(ii) If A is invertible, then A can be expressed by a product of elementary matrices by Theorem 2.3.3, namely, $A = E_1E_2 \cdots E_k$. Therefore, $\det(AB) = \det(E_1E_2 \cdots E_kB) = \det(E_1)\det(E_2E_3 \cdots E_kB) = \cdots = \det(E_1)\det(E_2)\cdots \det(E_k)\det(B) = \det(E_1E_2 \cdots E_k)\det(B)$.[1]

Combining the cases above, $\det(AB) = \det(A) \cdot \det(B)$. □

Corollary. For every nonnegative integer n, $\det(A^n) = \{\det(A)\}^n$. If A is invertible, this relationship is also valid for every negative integer n.

3.3 Cramer's Rule; Formula for A^{-1}; Applications of Determinants

3.3.1 Adjoint

Theorem 3.3.1. If the entries in any row (or column) of a square matrix are multiplied by the cofactors of the corresponding entries in a different row (or column), then the sum of the products are zero.

Sketch of Proof. The products are not changed if the different row (resp. column) is replaced by the row (resp. column), since the cofactor of every entry in the different row (resp. column) excludes the entries of itself. Then, the sum of the products are the determinant of the resulting matrix, which is zero, since it has two same rows (resp. columns). □

Theorem 3.3.2. If A is an invertible matrix, then

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

where the *adjoint* of A, denoted by $\operatorname{adj}(A)$, is the transpose of the *matrix of cofactors* $[C_{ij}]$ from A.

Sketch of Proof. By the direct computation of $A\operatorname{adj}(A) = \det(A)\,I$. □

Theorem 3.3.3 (Cramer's Rule)**.** If $A\mathbf{x} = \mathbf{b}$ is a linear system of n equations in n unknowns, then the system has a unique solution iff $\det(A) \neq 0$, in which case the solution is

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \cdots, \quad x_n = \frac{\det(A_n)}{\det(A)}$$

where A_i is obtained by replacing $\mathbf{c}_i(A)$ with vector b from A.

3.3.2 Geometric Interpretation of Determinants

Theorem 3.3.4.

i. If A is a 2×2 matrix, then $|\det(A)|$ represents the area of the parallelogram spanned by the two column vectors of A.

[1] If E is an $n \times n$ elementary matrix, then $\det(EB) = \det(E) \cdot \det(B)$.

ii. If A is a 3×3 matrix, then $|\det(A)|$ represents the volume of the parallelepiped spanned by the three column vectors of A.

Theorem 3.3.5. Suppose that a triangle in the xy-plane has vertices $P_1(x_1, y_1)$, $P_2(x_2, y_2), P_3(x_3, y_3)$ and that the labeling is such that the triangle is traversed counterclockwise from P_1 to P_2 to P_3. Then the area of the triangle is given by

$$\text{area}\, \triangle P_1P_2P_3 = \frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

Note that if the labeling is such that the triangle is traversed clockwise, then we have the negative determinant.

3.3.3 Polynomial Interpolation; The Vandermonde Determinant

Theorem 3.3.6 (Polynomial Interpolation)**.** Given any n points in the xy-plane that have distinct x-coordinates, then there is a unique polynomial of degree $n-1$ or less whose graph passes through those points.

Sketch of Proof. Suppose that n points are $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ and the interpolating polynomial is

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}$$

Then the coefficients in this polynomial satisfy $M\mathbf{v} = \mathbf{y}$ where

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The determinant of the matrix M is called the *Vandermonde Determinant*, which is given by

$$\det(M) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

Since $x_1, x_2, \cdots, x_n$ are all distinct, $\det(M) \neq 0$ hence M is invertible. □

3.4 A First Look at Eigenvalues and Eigenvectors

3.4.1 Fixed Points

Definition 3.4.1. A *fixed point* of an $n \times n$ matrix A is a solution of the homogeneous linear system $A\mathbf{x} = \mathbf{x} \Leftrightarrow (I - A)\mathbf{x} = \mathbf{0}$.

Every square matrix A has at least one fixed point, namely $\mathbf{x} = \mathbf{0}$, which is called the *trivial* fixed point.

Theorem 3.4.1. If A is an $n \times n$matrix, then the following statements are equivalent.

i. A has nontrivial fixed points.

ii. $I - A$ is singular.

iii. $\det(I - A) = 0$.

3.4.2 Eigenvalues and Eigenvectors

Definition 3.4.2. If A is an $n \times n$ matrix, then a scalar λ is called an *eigenvalue* of A if there is a *nonzero* vector such that $A\mathbf{x} = \lambda\mathbf{x} \Leftrightarrow (\lambda I - A)\mathbf{x} = \mathbf{0}$. The solution space of this equation is called the *eigenspace* of A corresponding to the eigenvalue λ. A nonzero vector of the eigenspace is called an *eigenvector* of A corresponding to λ.

Theorem 3.4.2. If A is an $n \times n$ matrix, then the following statements are equivalent.

i. λ is an eigenvalue of A.

ii. λ is a solution of the equation $\det(\lambda I - A) = 0$.

iii. The linear system $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

$\det(\lambda I - A) = 0$ is called the *characteristic equation* of A.

Theorem 3.4.3. If A is a triangular matrix, then the eigenvalues of A are the entries on the main diagonal of A.

Theorem 3.4.4. If λ is an eigenvalue of a matrix A and $\mathbf{x}$ is a corresponding eigenvector, and if k is any positive integer, then λ^k is an eigenvalue of A^k and $\mathbf{x}$ is a corresponding eigenvector.

Theorem 3.4.5 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. Theorem 3.2.3.

ii. $\lambda = 0$ is not an eigenvalue of A.

Theorem 3.4.6. If A is an $n \times n$ matrix, then the characteristic polynomial of A can be expressed as

$$\det(\lambda I - A) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_1)^{m_2} \cdots (\lambda - \lambda_k)^{m_k}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct (possibly complex) eigenvalues of A and $m_1 + m_2 + \cdots + m_k = n$.

Here, the exponent m_i, called the *algebraic multiplicity* of the eigenvalue λ_i, tells how many times that eigenvalue is repeated in the complete factorization of the characteristic polynomial.

Theorem 3.4.7. If A is an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ (repeated according to multiplicity), then:

i. $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$

ii. $\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$

3.4.3 Eigenvalue Analysis of 2×2 Matrices

Theorem 3.4.8. If A is a 2×2 matrix with real entries, then:

i. A has two distinct real eigenvalues if $\mathrm{tr}(A)^2 - 4\det(A) > 0$.

ii. A has one repeated real eigenvalue if $\mathrm{tr}(A)^2 - 4\det(A) = 0$.

iii. A has two conjugate imaginary eigenvalues $\mathrm{tr}(A)^2 - 4\det(A) < 0$.

Theorem 3.4.9. Suppose that A is a symmetric 2×2 matrix with real entries. Then A has real eigenvalues and has one repeated eigenvalue, namely $\lambda = a$, iff A is of the form

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

Sketch of Proof. If the 2×2 symmetric matrix is

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

then the characteristic equation of A is $(a - d)^2 + 4b^2$. □

Theorem 3.4.10.

i. If a 2×2 symmetric matrix with real entries has one repeated eigenvalue, then the eigenspace corresponding to that eigenvalue is $\mathbb{R}^2$.

ii. If a 2×2 symmetric matrix with real entries has tow distinct eigenvalues, then the eigenspaces corresponding to those eigenvalues are perpendicular lines through the origin of $\mathbb{R}^2$.

Chapter 4

Linear Transformations

4.1 Matrices as Transformations

4.1.1 Matrix Transformations

For an $m \times n$ matrix A, the transformation A, denoted by $T_A : \mathbb{R}^n \to \mathbb{R}^m$, is defined as $T_A(\mathbf{x}) = A\mathbf{x}$, or, $\mathbf{x} \xrightarrow{T_A} A\mathbf{x}$. If $m = n$, we have $T_A : \mathbb{R}^n \to \mathbb{R}^n$, which is called a matrix operator on $\mathbb{R}^n$.

4.1.2 Linear Transformations

Definition 4.1.1. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ if

i. $\forall c \in \mathbb{R} \forall \mathbf{u} \in \mathbb{R}^n \quad T(c\mathbf{u}) = cT(\mathbf{u})$ (Homogeneity)

ii. $\forall \mathbf{u} \in \mathbb{R}^n \, \forall \mathbf{v} \in \mathbb{R}^n \quad T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (Additivity)

In special case where $n = m$, T is called a linear operator on $\mathbb{R}^n$.

Note that $T(\mathbf{0}) = \mathbf{0}$ for any linear transformation T.

Theorem 4.1.1 (Superposition Principle)**.** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ are vectors in $\mathbb{R}^n$ and $c_1, c_2, \cdots, c_k$ are any scalars, then

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_kT(\mathbf{v}_k)$$

Theorem 4.1.2. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and vectors be in column form. Then $T(\mathbf{x}) = [T]\mathbf{x}$ where [T]=$\begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$ which is called the *standard matrix* for T.

4.1.3 Linear Operators $\mathbb{R}^2 \to \mathbb{R}^2$

1. Rotation operator R_θ: Rotation around the origin by θ.

$$R_\theta = \begin{bmatrix} R_\theta(\mathbf{e}_1) \mid R_\theta(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

2. Reflection operator H_θ: Reflection by the line $y = \tan\theta \cdot x$.

$$H_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}$$

3. Projection operator P_θ: Orthogonal projection onto $y = \tan\theta \cdot x$.

$$P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}$$

4. Scaling operator with factor k

$$[T] = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$$

5. Compression/Expansion with factor k

$$[T] = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \quad [T] = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$$

in the x direction and in the y direction, respecitvely.

6. Shears with factor k

$$[T] = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \quad [T] = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$$

in the x direction and in the y direction, respecitvely.

4.2 Geometry of Linear Operations

4.2.1 Norm-Preserving Linear Operators

Theorem 4.2.1. If T is a linear operator on $\mathbb{R}^n$, then the following statements are equivalent.

i. $\forall \mathbf{x} \in \mathbb{R}^n \quad \|T(\mathbf{x})\| = \|\mathbf{x}\|$ (T is orthogonal (i.e., norm preserving))

ii. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ (T is dot product preserving)

Sketch of Proof. Use the relationship $\mathbf{x} \cdot \mathbf{y} = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right)$. □

A linear operator T with the norm-preserving (or length-preserving) property is called an *orthogonal operator* or a *linear isometry*.

4.2.2 Orthogonal Matrices

Definition 4.2.1. A square matrix A is said to be orthogonal if $A^{-1} = A^T$.

Theorem 4.2.2. If A, B are orthogonal matrices, then:

i. A^T is orthogonal.

ii. A^{-1} is orthogonal.

iii. AB is orthogonal.

iv. $\det(A) = \pm 1$.

Theorem 4.2.3. If A is an $m \times n$ matrix, then the following statements are equivalent.

i. $A^T A = I$.

ii. $\forall \mathbf{x} \in \mathbb{R}^n \quad \|A\mathbf{x}\| = \|\mathbf{x}\|$

iii. $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

iv. The column vectors of A are orthonormal.

v. The row vectors of A are orthonormal. (If $m = n$)

Sketch of Proof. Let us prove (iii)$\Rightarrow$(iv). Since T_A preserves length and orthogonoality, hence orthonormality, the column vectors of A, namely $A\mathbf{e}_1$, $A\mathbf{e}_2$, $\cdots$, $A\mathbf{e}_n$, preserve orthonormality of $\mathbf{e}_1, \cdots, \mathbf{e}_n$. $\square$

4.2.3 Orthogonal Linear Operators on $\mathbb{R}^2$ and $\mathbb{R}^3$

Theorem 4.2.4. If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is an orthogonal linear operator, then

i. T is a rotation about the origin iff $|[T]| = 1$.

ii. T is a reflection about a line through the origin iff $|[T]| = -1$.

Theorem 4.2.5. Every orthogonal linear operator on $\mathbb{R}^3$ corresponds to linear operator on $\mathbb{R}^3$ of the following types:

i. Rotation about a line through the origin.

ii. Reflection about a plane through the origin.

iii. Rotation about a line through the origin followed by a reflection about the plane through the origin that is perpendicular to the line.

Note that every point on the axis of rotation is a fixed point of the standard matrix of the rotation, A. Also, if we have a vector $\mathbf{x}$ on the plane perpendicular to the axis, then we can calculate the angle of rotation by calculating the angle between $\mathbf{x}$ and $A\mathbf{x}$.

Example. The standard matrices of rotations about the positive x, y, z axes through an angle θ are respectively

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

4.3 Kernel and Range

4.3.1 Kernel

Definition 4.3.1. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then the set of vectors in $\mathbb{R}^n$ that T maps into $\mathbf{0}$ is called the kernel of T and is denoted by $\ker(T)$.

Theorem 4.3.1. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, $\ker(T)$ is a subspace of $\mathbb{R}^n$.

Definition 4.3.2. For an $m \times n$ matrix A, the solution space of $A\mathbf{x} = \mathbf{0}$, or equivalently, the kernel of T_A, is called the *null sapce* of the matrix A and is denoted by $\mathrm{null}(A)$.

Theorem 4.3.2. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then T maps subspaces of $\mathbb{R}^n$ into subspaces of $\mathbb{R}^m$.

4.3.2 Range

Theorem 4.3.3. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $\mathrm{ran}(T)$ is a subspace of $\mathbb{R}^m$.

Theorem 4.3.4. If A is an $m \times n$ matrix, then $\mathrm{ran}(T_A)$ is the column space of A.

4.3.3 Existence and Uniqueness

Definition 4.3.3. $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *onto* (or surjective) if $\mathrm{ran}(T) = \mathbb{R}^m$. That is, for every vector $\mathbf{b} \in \mathbb{R}^n$, the equation $T(\mathbf{x}) = \mathbf{b}$ is always consistent.

Definition 4.3.4. $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *one-to-one* (or injective) if $T(\mathbf{x}_1) = T(\mathbf{x}_2)$ implies $\mathbf{x}_1 = \mathbf{x}_2$. T is said to be bijective if T is surjective and injective.

Theorem 4.3.5. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then the following statements are equivalent.

i. T is one-to-one.

ii. $\ker(T) = \{\mathbf{0}\}$.

Theorem 4.3.6. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator on $\mathbb{R}^n$, then T is one-to-one iff T is onto.

Theorem 4.3.7 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. Theorem 3.4.5.

ii. T_A is one-to-one.

iii. T_A is onto.

4.4 Composition and Invertibility of Linear Transformations

4.4.1 Compositions of Linear Transformations

Theorem 4.4.1. For two linear transformations $T_1 : \mathbb{R}^n \to \mathbb{R}^k, T_2 : \mathbb{R}^k \to \mathbb{R}^m$, $(T_2 \circ T_1) : \mathbb{R}^n \to \mathbb{R}^m$ is also a linear transformation.

Theorem 4.4.2. If A is an $m \times k$ matrix and B is an $k \times n$ matrix, then $T_{AB} = T_A \circ T_B$.

Theorem 4.4.3. If A is an invertible 2×2 matrix, then T_A is a composition of shears, compressions, expansions in the directions of the coordinate axes, and reflections about the coordinate axes and $y = x$.

Sketch of Proof. The standard matrices of these transformations are 2×2 elementary matrices. □

4.4.2 Inverse of Linear Transformations

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one linear transformation, then $T^{-1} : \mathrm{ran}(T) \to \mathbb{R}^n$ can be defined as $T^{-1}(\mathbf{y}) = \mathbf{x}$ if $T(\mathbf{x}) = \mathbf{y}$. T^{-1} is also one-to-one.

Theorem 4.4.4. If T is a one-to-one linear operator on $\mathbb{R}^n$, then the standard matrix A of T is invertible and $T^{-1} = T_{A^{-1}}$.

Theorem 4.4.5. If T is an invertible linear operator on $\mathbb{R}^2$, then:

i. The image of a line is a line.

ii. The image of a line passes through the origin iff the original line passes through the origin.

iii. The images of two lines are parallel iff the original lines are parallel.

iv. The images of three points are colinear iff the original points are colinear.

v. The images of the line segment joining two points is the line segment joining the images of those points.

Chapter 5

Dimension and Structure

5.1 Basis and Dimension

5.1.1 Basis for Subspaces

Definition 5.1.1. A set of vectors in a subspace W of $\mathbb{R}^n$ is said to be a *basis* for W if it is linearly independent and spans W. Zero subspaces have no basis.

Example. $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$.

Theorem 5.1.1. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is a set of two or more nonzero vectors in $\mathbb{R}^n$, then S is linearly dependent iff some vector in S is a linear combination of its predecessors.

Sketch of Proof. Assume S is linearly dependent. Then there exist scalars $t_1, t_2, \cdots, t_k$, not all zero, such that

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k = \mathbf{0}$$

Let t_i be the nonzero scalar that has the largest index. Then $\mathbf{v}_i$ is a linear combination of its predecessors.[1] $\square$

Theorem 5.1.2. If W is a nonzero subspace of $\mathbb{R}^n$, then there exists a basis for W at most n vectors.

Sketch of Proof. We can choose linear independent vectors in W until they span W. But the number of them cannot exceed n, as guaranteed by Theorem 2.4.4. $\square$

Theorem 5.1.3. All bases for a nonzero subspace of $\mathbb{R}^n$ have the same number of vectors.

Sketch of Proof. Assume that there are two bases B_1, B_2 for a subspace W of $\mathbb{R}^n$ that vary in the number of vectors. Suppose

$$B_1 = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$$
$$B_2 = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$$

[1] By the definition of t_i, all terms beyond the index i must be zero.

with $m > k$. Then we shall prove that B_2 is linear dependent, which leads to a contradiction. So consider the following equation:

$$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_m\mathbf{w}_m = 0$$

Since $\mathbf{w}_1, \cdots, \mathbf{w}_m \in W$, we can write

$$c_i\mathbf{w}_i = c_i\left(a_{1i}\mathbf{v}_1 + a_{2i}\mathbf{v}_2 + \cdots + a_{ki}\mathbf{v}_k\right)$$

Taking summation of both sides from $i = 1$ to $i = m$ yields

$$\sum_{i=1}^{m} c_i\mathbf{w}_i = \mathbf{v}_1\sum_{i=1}^{m} c_ia_{1i} + \mathbf{v}_2\sum_{i=1}^{m} c_ia_{2i} + \cdots + \mathbf{v}_k\sum_{i=1}^{m} c_ia_{ki}$$

Let us consider a linear system of equations of k equations in m unknowns:

$$\sum_{i=1}^{m} c_ia_{1i} = 0,\ \sum_{i=1}^{m} c_ia_{2i} = 0, \cdots,\ \sum_{i=1}^{m} c_ia_{ki} = 0$$

This system has nontrivial solutions, since it is homogeneous and has more unknowns than equations (Corollary of Theorem 1.2.2). So the proof is now complete. □

5.1.2 Dimension

Definition 5.1.2. The *dimension* $\dim(W)$ of a nonzero subspace W of $\mathbb{R}^n$ is defined to be the number of vectors in a basis for W. Zero subspaces have 0 dimension.

If A is an $m \times n$ matrix, then the solution space of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ is given by Gauss-Jordan elimination and is

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k$$

where $t_1, \cdots, t_k \in \mathbb{R}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a set of vectors in $\mathbb{R}^n$ that is linearly independent. Here $\mathbf{x}$ is called the *canonical solutions* of $A\mathbf{x} = \mathbf{0}$, and $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is called the *canonical basis* for the solution space.

Theorem 5.1.4. If $\mathbf{a}$ is a nonzero vector in $\mathbb{R}^n$, then $\dim(\mathbf{a}^\perp) = n - 1$.

5.2 Properties of Bases

5.2.1 Properties of Bases

Theorem 5.2.1. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is a basis for subspace W of $\mathbb{R}^n$, then every vector $\mathbf{v}$ in W can be expressed in exactly one way as a linear combination of the vectors in S.

Theorem 5.2.2. Let S be a finite set of vectors in a nonzero subspace W of $\mathbb{R}^n$.

i. If S spans W, but is not a basis for W, then a basis for W can be obtained by removing appropriate vectors from S.

ii. If S is a linearly independent set, but not a basis for W, then a basis for W can be obtained by adding appropriate vectors from W to S.

Theorem 5.2.3. If W is a nonzero subspace of $\mathbb{R}^n$, then $\dim(W)$ is the maximum number of linearly independent vectors in W.

5.2.2 Subspaces of Subspaces

Theorem 5.2.4. If V and W are subspaces of $\mathbb{R}^n$, and if V is a subspace of W, then:

i. $0 \leq \dim(V) \leq \dim(W) \leq n$

ii. $V = W$ iff $\dim(V) = \dim(W)$

Theorem 5.2.5. Let S be a nonempty set of vectors in $\mathbb{R}^n$, and let S' be a set that results by adding additional vectors in $\mathbb{R}^n$ to S.

i. $\operatorname{span}(S') = \operatorname{span}(S) \Leftrightarrow$ the additional vectors are in $\operatorname{span}(S)$.

ii. $\dim(\operatorname{span}(S')) = \dim(\operatorname{span}(S)) \Leftrightarrow \operatorname{span}(S') = \operatorname{span}(S)$.

Theorem 5.2.6 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. Theorem 4.3.7.

ii. The column vectors of A span $\mathbb{R}^n$.

iii. The row vectors of A span $\mathbb{R}^n$.

iv. The column vectors of A form a basis for $\mathbb{R}^n$.

v. The row vectors of A form a basis for $\mathbb{R}^n$.

5.3 The Fundamental Spaces of a Matrix

5.3.1 The Fundamental Spaces of a Matrix

Definition 5.3.1. The fundamental spaces of a matrix A are as follows:

i. $\operatorname{row}(A)$, the *row space* of A: The subspace spanned by row vectors of A.

ii. $\operatorname{col}(A)$, the *column space* of A: The subspace spanned by column vectors of A.

iii. $\operatorname{null}(A)$, the *null space* of A: The solution space of the system $A\mathbf{x} = \mathbf{0}$.

iv. $\operatorname{null}(A^T)$, the null space of A^T: The solution space of the system $A^T\mathbf{x} = \mathbf{0}$.

Definition 5.3.2. The dimension of the row space of a matrix A is called the *rank* of A and is denoted by $\operatorname{rank}(A)$; and the dimension of the null space of A is called the *nullity* of A and is denoted by $\operatorname{nullity}(A)$.

5.3.2 Orthogonal Complements

Definition 5.3.3. For a nonempty set S in $\mathbb{R}^n$, the *orthogonal complement* of S, denoted by $S^\perp$, is defined to be the set of all vectors in $\mathbb{R}^n$ that are orthogonal to *every* vector in S.

Theorem 5.3.1. If S is a nonempty subset of $\mathbb{R}^n$, then $S^\perp$ is a subspace of $\mathbb{R}^n$.

Theorem 5.3.2. If W is a subspace of $\mathbb{R}^n$, then:

i. $W \cap W^{\perp} = \{\mathbf{0}\}$

ii. $(W^{\perp})^{\perp} = W$

Theorem 5.3.3. If S is a nonempty subset of $\mathbb{R}^n$, then $S^{\perp} = \{\text{span}(S)\}^{\perp}$. Therefore, $(S^{\perp})^{\perp} = \text{span}(S)$.

Theorem 5.3.4. If A is a matrix, then:

i. $\text{row}(A)^{\perp} = \text{null}(A)$, $\text{null}(A)^{\perp} = \text{row}(A)$.

ii. $\text{col}(A)^{\perp} = \text{null}(A^T)$, $\text{null}(A^T)^{\perp} = \text{col}(A)$.

Theorem 5.3.5.

i. Elementary row operations do not change the row space or null space of a matrix.

ii. The nonzero row vectors in any row echelon form of a matrix form a basis for the row space of the matrix.

Theorem 5.3.6. If A and B are matrices with the same number of columns, then the following statements are equivalent.

i. A and B have the same row space.

ii. A and B have the same null space.

iii. The row vectors of A are linear combinations of the row vectors of B, and conversely.

5.3.3 Determining Whether a Vector Is in a Given Subspace

Let W be a subspace of $\mathbb{R}^n$. We want to know whether a vector $\mathbf{b} \in \mathbb{R}^n$ is in W. Then we have the following ways to determine it:

1. Solve $A\mathbf{x} = \mathbf{b}$ by row reduction, where the column vectors of A span W.
2. Find a basis for W and find a linear combination of the basis for $\mathbf{b}$. You can use Theorem 5.3.5 to find a basis for W.
3. Find a basis for $W^{\perp}$ and determine whether every vector in the basis is orthogonal to $\mathbf{b}$.

5.4 The Dimension Theorem and Its Implications

5.4.1 The Dimension Theorem for Matrices

Theorem 5.4.1 (The Dimension Theorem for Matrices)**.** If A is an $m \times n$ matrix, then $\text{rank}(A) + \text{nullity}(A) = n$.

Sketch of Proof. The number of nonzero rows in a row echelon form of A is $\text{rank}(A)$. Thus, it follows from Theorem 1.2.2. $\square$

It follows from the dimension theorem that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k, \mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \cdots, \mathbf{v}_n\}$ forms a basis for $\mathbb{R}^n$ if $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ forms a basis for row(A) and $\{\mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \cdots, \mathbf{v}_n\}$ forms a basis for null(A).

Theorem 5.4.2. If the rank of an $m \times n$ matrix A is k, then:

i. nullity$(A) = n - k$.

ii. Every reduced echelon form of A has k nonzero rows and $m-k$ zero rows.

iii. $A\mathbf{x} = \mathbf{0}$ has k pivot (or leading) variables and $n - k$ free variables.

Note that the definition of rank(A) imposes rank$(A) \leq m$.

Theorem 5.4.3 (The Dimension Theorem for Subspaces)**.** If W is a subspace of $\mathbb{R}^n$, then $\dim(W) + \dim(W^{\perp}) = n$.

Theorem 5.4.4 (Characterizations of Invertibility)**.** If A is an $n \times n$ matrix, then the following statements are equivalent.

i. The reduced row echelon form of A is I_n.

ii. A is expressible as a product of elementary matrices.

iii. A is invertible.

iv. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

v. $A\mathbf{x} = \mathbf{b}$ is consistent for every vector $\mathbf{b}$ in $\mathbb{R}^n$.

vi. $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every vector $\mathbf{b}$ in $\mathbb{R}^n$.

vii. The column vectors of A are linearly independent.

viii. The row vectors of A are linearly independent.

ix. $\det(A) \neq 0$.

x. $\lambda = 0$ is not an eigenvalue of A.

xi. T_A is one-to-one.

xii. T_A is onto.

xiii. The column vectors of A span $\mathbb{R}^n$.

xiv. The row vectors of A span $\mathbb{R}^n$.

xv. The column vectors of A form a basis for $\mathbb{R}^n$.

xvi. The row vectors of A form a basis for $\mathbb{R}^n$.

xvii. rank$(A) = n$.

xviii. null$(A) = 0$.

5.4.2 Rank 1 Matrices

Theorem 5.4.5. If A is an $m \times n$ matrix with rank 1, then $A = \mathbf{u}\mathbf{v}^T$ where $\mathbf{u}$ is a nonzero column vector in $\mathbb{R}^m$ and $\mathbf{v}$ is a nonzero column vector in $\mathbb{R}^n$. The converse is also true.

Theorem 5.4.6. If A is a symmetric $n \times n$ matrix with rank 1, then $A = \mathbf{u}\mathbf{u}^T$ or $A = -\mathbf{u}\mathbf{u}^T$ where $\mathbf{u}$ is a nonzero column vector in $\mathbb{R}^n$. The converse is also true.

Sketch of Proof. Let $A = \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ is a nonzero column vector in $\mathbb{R}^n$ and $\mathbf{v}$ is a nonzero column vector in $\mathbb{R}^n$. Since $A^T = A$, we have $\mathbf{u}^T(\mathbf{v}\mathbf{u}^T)\mathbf{v} = \mathbf{u}^T(\mathbf{u}\mathbf{v}^T)\mathbf{v}$. Then this equation becomes $(\mathbf{u} \cdot \mathbf{v})^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2$. Therefore $\mathbf{v}$ is a nonzero scalar multiple of $\mathbf{u}$. $\square$

Theorem 5.4.7. For an $m \times n$ matrix A, $\operatorname{rank}(A)$ is the order of the largest square submatrix of A formed by deleting rows and columns of A, whose determinant is nonzero.

5.5 The Rank Theorem and Its Implications

5.5.1 The Rank Theorem

Theorem 5.5.1 (The Rank Theorem)**.** If A is an $m \times n$ matrix, then $\operatorname{rank}(A) = \operatorname{rank}(A^T)$

By the dimension theorem, we have

$$\begin{gathered} \operatorname{rank}(A) + \operatorname{nullity}(A) = n \\ \operatorname{rank}(A^T) + \operatorname{nullity}(A^T) = m \end{gathered}$$

Let $\operatorname{rank}(A) = k$. Then the rank theorem implies $\operatorname{rank}(A^T) = k$. Therefore, we have the dimension of all the fundamental spaces of A.

Theorem 5.5.2. If $A\mathbf{x} = \mathbf{b}$ is a linear system of m equations in n unknowns, then the following statements are equivalent.

i. $A\mathbf{x} = \mathbf{b}$ is consistent.

ii. $\mathbf{b} \in \operatorname{col}(A)$.

iii. $\operatorname{rank}(A) = \operatorname{rank}([A \mid \mathrm{b}])$.

Definition 5.5.1. An $m \times n$ matrix A is said to have *full column rank* (resp. *full row rank*) if its column vectors (resp. row vectors) are linearly independent. That is, the column vectors of A form a basis for $\operatorname{col}(A)$.

Theorem 5.5.3. Let A be an $m \times n$ matrix. Then,

i. A has full column rank iff $\operatorname{rank}(A) = n$.

ii. A has full row rank iff $\operatorname{rank}(A) = m$.

Theorem 5.5.4. If A is an $m \times n$ matrix, then:

i. $\text{row}(A) = \text{row}(A^T A)$, $\text{null}(A) = \text{null}(A^T A)$, $\text{col}(^T A) = \text{col}(A^T A)$.

ii. $\text{row}(A^T) = \text{row}(AA^T)$, $\text{null}(A) = \text{null}(AA^T)$, $\text{col}(A) = \text{col}(AA^T)$.

Sketch of Proof. Let us prove that $\text{null}(A) = \text{null}(A^T A)$.

(i) Assume $A\mathbf{x} = \mathbf{0}$. Then $A^T A\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$.

(ii) Assume $A^T A\mathbf{x} = \mathbf{0}$. Then $\mathbf{x} \perp \text{row}(A^T A) = \text{col}((A^T A)^T) = \text{col}(A^T A)$. Since $A^T A\mathbf{x} \in \text{col}(A^T A)$, we have

$$\mathbf{x} \cdot (A^T A\mathbf{x}) = 0 \Leftrightarrow A\mathbf{x} \cdot A\mathbf{x} = 0$$

Therefore, $A\mathbf{x} = \mathbf{0}$.

Combining the cases above, we have $A^T A\mathbf{x} = \mathbf{0} \Leftrightarrow A\mathbf{x} = \mathbf{0}$. □

Theorem 5.5.5. If A is an $m \times n$ matrix, then the following statements are equivalent.

i. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

ii. $A\mathbf{x} = \mathbf{b}$ has at most one solution for every vector $\mathbf{b} \in \mathbb{R}^n$.

iii. A has full column rank.

iv. $A^T A$ is invertible.

Theorem 5.5.6. Let A be an $m \times n$ matrix.

i. If $m > n$, then there exists a vector $\mathbf{b} \in \mathbb{R}^m$ such that the linear system $A\mathbf{x} = \mathbf{b}$ is inconsistent.

ii. If $m < n$, then for every vector $\mathbf{b} \in \mathbb{R}^m$, the linear system $A\mathbf{x} = \mathbf{b}$ is either inconsistent or has infinitely many solutions.

5.6 The Pivot Theorem and Its Implications

5.6.1 Finding a Basis for Column Space

Theorem 5.6.1. Let A and B be row equivalent matrices.

i. If some subset of column vectors from A is linearly independent, then the corresponding column vectors from B are linearly independent, and conversely.

ii. If some subset of column vectors from A is linearly dependent, then the corresponding columns vectors from B are linearly dependent and have the same dependency relationship with the column vectors from A, and conversely.

Definition 5.6.1. The column vectors of a matrix A that lie in the column positions where the leading 1's occur in the row echelon forms of A are called the *pivot columns* of A.

Theorem 5.6.2 (The Pivot Theorem)**.** The pivot columns of a nonzero matrix A form a basis for the column space of A.

The number of pivot columns of A are equal to the number of leading 1's of a row echelon form of A. The former is $\text{rank}(A^T)$ and the latter is $\text{rank}(A)$. Therefore, the pivot theorem implies the rank theorem.

5.6.2 Finding Bases for the Fundamental Spaces of a Matrix

Theorem 5.6.3. Suppose that A is an $m \times n$ matrix with rank k, and

$$[A \,|\, I_m] \to [U \,|\, E] = \left[\begin{array}{c|c} V & E_1 \\ \hline 0 & E_2 \end{array}\right]$$

where U is a row echelon form of A. Then the row vectors of E_2 form a basis for $\text{null}(A^T)$.

Sketch of Proof. Since U is a row echelon form of A, $U = EA$ for some invertible matrix E. Since

$$U = \left[\begin{array}{c} V \\ \hline 0 \end{array}\right], \, E = \left[\begin{array}{c} E_1 \\ \hline E_2 \end{array}\right]$$

So $U = EA$ becomes

$$\left[\begin{array}{c} V \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} E_1 A \\ \hline E_2 A \end{array}\right]$$

where V is an $k \times n$ matrix, E_1 is an $k \times m$ matrix and E_2 is an $(m-k) \times m$ matrix. If we rewrite E_2 and A as

$$E_2 = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_{m-k}^T \end{bmatrix} \qquad A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$$

where $\mathbf{u}_1, \cdots, \mathbf{u}_{m-k}$ and $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are column vectors in $\mathbb{R}^m$, then

$$E_2 A = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{v}_1 & \mathbf{u}_1 \cdot \mathbf{v}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{v}_n \\ \mathbf{u}_2 \cdot \mathbf{v}_1 & \mathbf{u}_2 \cdot \mathbf{v}_2 & \cdots & \mathbf{u}_2 \cdot \mathbf{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_{m-k} \cdot \mathbf{v}_1 & \mathbf{u}_{m-k} \cdot \mathbf{v}_2 & \cdots & \mathbf{u}_{m-k} \cdot \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Therefore, $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{m-k}$ are in $\text{col}(A)^\perp = \text{null}(A^T)$. Since E is invertible so the row vectors of E are linearly independent, $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{m-k}\}$ is linearly independent, hence forms a basis for $\text{null}(A^T)$. □

Example. Let's find all the fundamental spaces of a matrix A.

$$[A \,|\, I_4] = \left[\begin{array}{rrrrrr|rrrr} 1 & -3 & 4 & -2 & 5 & 4 & 1 & 0 & 0 & 0 \\ 2 & -6 & 9 & -1 & 8 & 2 & 0 & 1 & 0 & 0 \\ 2 & -6 & 9 & -1 & 9 & 7 & 0 & 0 & 1 & 0 \\ -1 & 3 & -4 & 2 & -5 & -4 & 0 & 0 & 0 & 1 \end{array}\right]$$

Reduce A to a row echelon form:

$$\left[\begin{array}{c|c} V & E_1 \\ \hline 0 & E_2 \end{array}\right] = \left[\begin{array}{rrrrrr|rrrr} 1 & -3 & 4 & -2 & 5 & 4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & -2 & -6 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 5 & 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

Continue to reduce A to the reduced row echelon form separately:

$$U = \begin{bmatrix} 1 & -3 & 0 & -14 & 0 & -37 \\ 0 & 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Then let us form a matrix W from U by the following procedures:[2]

1. Remove the zero rows of U.
2. Reverse the signs of the entries on the columns that are not pivot columns.
3. Insert standard row vectors appropriately so that every leading 1 is on the main diagonal.
4. Remove the column vectors corresponding to the pivot columns of U.

The resulting matrix W for this example is

$$W = \begin{bmatrix} 3 & 14 & 37 \\ 1 & 0 & 0 \\ 0 & -3 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & -5 \\ 0 & 0 & 1 \end{bmatrix}$$

The four fundamental spaces of A are as follows:

1. The row vectors of V form a basis for row(A):
$$\{(1, -3, 4, -2, 5, 4), (0, 0, 1, 3, -2, -6), (0, 0, 0, 0, 1, 5)\}$$
2. The pivot columns of A form a basis for col(A):[3]
$$\{(1, 2, 2, -1), (4, 9, 9, -4), (5, 8, 9, -5)\}$$
3. The row vectors of E_2 form a basis for null(A^T):
$$\{(1, 0, 0, 1)\}$$
4. The column vectors of W form a basis for null(A) :
$$\{(3, 1, 0, 0, 0, 0), (14, 0, -3, 1, 0, 0), (37, 0, -4, 0, -5, 1)\}$$

We can also construct a basis for $\mathbb{R}^6$ by combining 1 and 3; $\mathbb{R}^4$ by combining 2 and 4. This can be applied to extend a linearly independent set to a basis for $\mathbb{R}^n$. Besides, we extracted a basis for col(A) from the column vectors of A in 2. This can be applied to extract a basis for a subspace from a set of spanning vectors of that subspace.

[2]This procedure is equivalent to reading off a general solution of the system $U\mathbf{x} = \mathbf{0}$.

[3]Leading 1's of the row echelon form of A occurs at the first, third and fifth column. So the first, third and fifth column of A are the pivot columns of A.

5.6.3 Column-Row Factorization

Theorem 5.6.4 (Column-Row Factorization)**.** If A is a nonzero $m \times n$ matrix of rank k, then A can be factored as $A = CR$ where C is $m \times k$ matrix whose column vectors are the pivot columns of A and R is the $k \times n$ matrix whose row vectors are the nonzero rows in the *reduced* row echelon form of A.

Sketch of Proof. Let R_0 be the reduced row echelon form of A. Then there exists an invertible $m \times m$ matrix E such that

$$EA = R_0 \quad \text{or} \quad A = E^{-1}R_0 = [C|D]\left[\frac{R}{0}\right] = CR$$

where C consists of the first k column vectors of E^{-1} and R consists of the nonzero row vectors of R_0. Note that the column vectors of R corresponding to the pivot columns of A (i.e., to the column vectors of R_0 where the leading 1's occur) are the standard vectors of $\mathbb{R}^k$. Moreover, if those column positions are $c_1, c_2, \cdots, c_k$, then each column vector of R that lies on each column position $c_1, c_2, \cdots, c_k$ is $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k$, respectively. So we have

$$\mathbf{c}_{c_i}(A) = C\mathbf{c}_{c_i}(R) = C\mathbf{e}_i = \mathbf{c}_i(C)$$

where $i = 1, 2, \cdots, k$. Therefore, the successive column vectors of C are the successive pivot columns of A. $\square$

Theorem 5.6.5 (Column-Row Expansion)**.** If A is a nonzero matrix of rank k, then A can be expressed as

$$A = \mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_k\mathbf{r}_k$$

where $\mathbf{c}_1, \cdots, \mathbf{c}_k$ are the successive pivot columns of A and $\mathbf{r}_1, \cdots, \mathbf{r}_k$ are the successive nonzero row vectors in the reduced row echelon form of A.

5.7 The Projection Theorem and Its Implications

5.7.1 Orthogonal Projections onto Lines in $\mathbb{R}^n$

Theorem 5.7.1. Every vector $\mathbf{x}$ in $\mathbb{R}^n$ can be expressed in exactly one way as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ where

$$\mathbf{x}_1 = \text{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} \qquad \mathbf{x}_2 = \mathbf{x} - \text{proj}_{\mathbf{a}}\mathbf{x} = \text{proj}_{\mathbf{a}^\perp}\mathbf{x}$$

for a nonzero vector $\mathbf{a} \in \mathbb{R}^n$. $\text{proj}_a\mathbf{x}$ is called the *orthogonal projection of* $\mathbf{x}$ *onto span*$\{\mathbf{a}\}$. It is also called the *vector component of* $\mathbf{x}$ *along* $\mathbf{a}$. $\text{proj}_{\mathbf{a}^\perp}\mathbf{x}$ is called the *vector component of* $\mathbf{x}$ *orthogonal to* $\mathbf{a}$.

Theorem 5.7.2. If $\mathbf{a}$ is a nonzero column vector in $\mathbb{R}^n$, then the standard matrix for the linear operator $T(\mathbf{x}) = \text{proj}_{\mathbf{a}}\mathbf{x}$ is

$$P = \frac{1}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T$$

This matrix is symmetric and has rank 1. If $\mathbf{a}$ is unit vector, then $P = \mathbf{a}\mathbf{a}^T$.

5.7.2 Orthogonal Projections onto General Subspaces

Theorem 5.7.3 (The Projection Theorem for Subspaces)**.** If W is a subspace of $\mathbb{R}^n$, then every vector $\mathbf{x}$ in $\mathbb{R}^n$ can be expressed in exactly one way as

$$\mathbf{x} = \mathbf{u} + \mathbf{v}$$

where $\mathbf{u}$ is in W and $\mathbf{v}$ is in $W^\perp$.

Sketch of Proof. Suppose that the column vectors of an $n \times m$ matrix M from a basis for W. Then we have the following conditions:

i. $\mathbf{x} = \mathbf{u} + \mathbf{v}$

ii. $\exists!\, \mathbf{u} \in \mathrm{col}(M) \Leftrightarrow \exists!\, \mathbf{a} \in \mathbb{R}^n : \mathbf{u} = M\mathbf{a}$

iii. $\exists!\, \mathbf{v} \in \mathrm{col}(M)^\perp \Leftrightarrow \exists!\, \mathbf{v} \in \mathbb{R}^m : M^T\mathbf{v} = \mathbf{0}$

where $\exists!$ denotes the unique existence. Putting (i) and (ii) to (iii) gives

$$M^T(\mathbf{x} - \mathbf{u}) = 0 \Leftrightarrow M^T(\mathbf{x} - M\mathbf{a}) = 0 \Leftrightarrow M^TM\mathbf{a} = M^T\mathbf{x}$$

Since M has full column rank, M^TM is invertible by Theorem 5.5.5. So the equation above has the unique solution

$$\mathbf{a} = (M^TM)^{-1}M^T\mathbf{x}$$

From (ii), we have the unique vector $\mathbf{u} \in W$, namely, $\mathbf{u} = M\mathbf{a}$; then from (i), we have the unique vector $\mathbf{v} \in W^\perp$, namely, $\mathbf{v} = \mathbf{x} - M\mathbf{a}$. □

Then we *define* the orthogonal projection of $\mathbf{x}$ on W to be $\mathrm{proj}_W\mathbf{x} = \mathbf{u}$. It follows from Theorem 5.7.3 that the orthogonal projection of $\mathbf{x}$ on $W^\perp$ is $\mathrm{proj}_{W^\perp}\mathbf{x} = \mathbf{x} - \mathrm{proj}_W\mathbf{x}$.

Theorem 5.7.4. If W is a nonzero subspace of $\mathbb{R}^n$, and if M is any matrix whose column vectors form a basis for W, then

$$\mathrm{proj}_W\mathbf{x} = M(M^TM)^{-1}M^T\mathbf{x}$$

for every column vector $\mathbf{x}$ in $\mathbb{R}^n$.

Here we can define the linear operator $T(\mathbf{x}) = \mathrm{proj}_W\mathbf{x}$ on $\mathbb{R}^n$ whose standard matrix P is $P = M(M^TM)^{-1}M^T$.[4] This linear operator is called the *orthogonal projection of* $\mathbb{R}^n$ *onto* W. Besides, since $\mathbf{x} = \mathrm{proj}_W\mathbf{x} + \mathrm{proj}_{W^\perp}\mathbf{x}$, we have $\mathrm{proj}_{W^\perp}\mathbf{x} = \mathbf{x} - P\mathbf{x} = (I - P)\mathbf{x}$. Therefore the standard matrix of the orthogonal projection of $\mathbb{R}^n$ onto $W^\perp$ is $I - P = I - M(M^TM)^{-1}M^T$.

Theorem 5.7.5. An $n \times n$ matrix P is the standard matrix for the orthogonal projection of $\mathbb{R}^n$ onto a subspace W iff

i. P is symmetric, i.e., $P^T = P$.

ii. P is idempotent, i.e., $P^2 = P$.

The subspace W is the column space of P.

[4] Note that this is analgous to $P = \mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}$ from Theorem 5.7.2.

Sketch of Proof. Let $\mathbf{x} = P\mathbf{x} + (I - P)\mathbf{x}$ for every vector $\mathbf{x}$ in $\mathbb{R}^n$. Then $P\mathbf{x} \in \mathrm{col}(P)$ and $(I - P)\,\mathbf{x} \in \mathrm{col}(P)^{\perp}$ since

$$P^T(I - P)\,\mathbf{x} = P(I - P)\,\mathbf{x} = (P - P^2)\,\mathbf{x} = \mathbf{0}$$

Therefore putting $W = \mathrm{col}(P)$ completes the proof. □

Theorem 5.7.6. Suppose that A is an $m \times n$ matrix and $\mathbf{b}$ is in $\mathrm{col}(A)$.

i. If A has full column rank, then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution, and that solution is in $\mathrm{row}(A)$.

ii. If A does not have full column rank, then the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions, but there is a unique solution in $\mathrm{row}(A)$. Moreover, this solution has the *smallest norm* among all solutions of the system.

Sketch of Proof.

i. If A has full column rank, then the system $A\mathbf{x} = \mathbf{b}$ has a unique solution by Theorem 2.5.3 and Theorem 5.5.5. Since $\mathrm{rank}(A) = n$ by Theorem 5.5.3, the row vectors of A span $\mathbb{R}^n$. Therefore $\mathbf{x} \in \mathbb{R}^n = \mathrm{row}(A)$.

ii. If A does not have full column rank, then Theorem 5.5.5 implies that the system $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions and so does $A\mathbf{x} = \mathbf{b}$. Let $\mathbf{x}$ be any solution of the system $A\mathbf{x} = \mathbf{b}$. By Theorem 5.7.3, we have

$$\mathbf{x} = \mathbf{x}_{\mathrm{row}(A)} + \mathbf{x}_{\mathrm{null}(A)}$$

where $\mathbf{x}_{\mathrm{row}(A)} \in \mathrm{row}(A)$, $\mathbf{x}_{\mathrm{null}(A)} \in \mathrm{null}(A)$. Then we have $A\mathbf{x}_{\mathrm{row}(A)} = \mathbf{b}$. Suppose that $\mathbf{x}_r$ and $\mathbf{x}'_r$ are two solutions in $\mathrm{row}(A)$. Then $A(\mathbf{x}_r - \mathbf{x}'_r) = 0$, which implies that $\mathbf{x}_r - \mathbf{x}'_r$ is in $\mathrm{row}(A)$ and $\mathrm{row}(A)^{\perp}$. By Theorem 5.3.2, $\mathbf{x}_r = \mathbf{x}'_r$. Therefore we have proved that there is a unique solution of $A\mathbf{x} = \mathbf{b}$ in $\mathrm{row}(A)$. Moreover, the theorem of Pythagoras implies that

$$\|\mathbf{x}\| = \sqrt{\|\mathbf{x}_{\mathrm{row}(A)}\|^2 + \|\mathbf{x}_{\mathrm{null}(A)}\|^2} \geq \|\mathbf{x}_{\mathrm{row}(A)\|}$$

the solution in $\mathrm{row}(A)$ has minimum norm.

So the proof is complete. □

Theorem 5.7.7 (The Double Perp Theorem)**.** If W is a subspace of $\mathbb{R}^n$, then $(W^{\perp})^{\perp} = W$.

Sketch of Proof. Prove first $W \subset (W^{\perp})^{\perp}$ and then $(W^{\perp})^{\perp} \subset W$ using Theorem 5.7.3 □

5.8 Best Approximation and Least Squares

5.8.1 Least Squares Solutions of Linear Systems

Theorem 5.8.1 (Best Approximation Theorem)**.** If W is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a point in $\mathbb{R}^n$, then there is a unique *best approximation to* $\mathbf{b}$ *from* W, i.e., there is a unique vector $\hat{\mathbf{w}}$ in W such that $\|\mathbf{b} - \hat{\mathbf{w}}\| < \|\mathbf{b} - \mathbf{w}\|$ for every vector $\mathbf{w} \neq \hat{\mathbf{w}}$ in W, namely $\hat{\mathbf{w}} = \mathrm{proj}_W \mathbf{b}$.

Sketch of Proof. For every vector $\mathbf{w}$ in W we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \text{proj}_W \mathbf{b}) + (\text{proj}_W \mathbf{b} - \mathbf{w})$$

where the first term is in $W^\perp$ and the second term is in W. □

So we can define the *distance* from a point $\mathbf{b}$ to a subspace W in $\mathbb{R}^n$ to be $d = \|\mathbf{b} - \text{proj}_W \mathbf{b}\| = \|\text{proj}_{W^\perp} \mathbf{b}\|$.

Definition 5.8.1. If A is an $m \times n$ matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^m$, then a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ is called a *best approximate solution* or a *least squares solution* of $A\mathbf{x} = \mathbf{b}$ if $\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$ for every $\mathbf{x} \in \mathbb{R}^n$. $\mathbf{b} - A\hat{\mathbf{x}}$ is called the *least squares error vector*, and $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the *least squares error*.

Theorem 5.8.2.

i. The least squares solutions of $A\mathbf{x} = \mathbf{b}$ are the exact solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.

ii. If A has full column rank, then the normal equation has a unique solution, namely $\hat{\mathbf{x}} = (A^T A)^{-1}\mathbf{b}$.

iii. If A does not have full column rank, then the normal equation has infinitely many solutions, but there is a unique solution in row(A). This soultion has the smallest norm.

Sketch of Proof.

i. The least squares solutions of $A\mathbf{x} = \mathbf{b}$ are the solution $A\mathbf{x} = \text{proj}_{\text{col}(A)}\mathbf{b}$. Then this system is equivalent to

$$\mathbf{b} - A\mathbf{x} = \mathbf{b} - \text{proj}_{\text{col}(A)}\mathbf{b} \quad \text{or} \quad \mathbf{b} - A\mathbf{x} = \text{proj}_{\text{null}(A^T)}\mathbf{b}$$

From Theorem 5.7.3, $\mathbf{b}$ can be expressed in exactly one way as

$$\mathbf{b} = A\mathbf{x} + (\mathbf{b} - A\mathbf{x})$$

where $A\mathbf{x} \in \text{col}(A)$ and $(\mathbf{b} - A\mathbf{x}) \in \text{null}(A^T)$. Therefore, finding all vectors $\mathbf{x} \in \mathbb{R}^n$ that satisfy the equation $\mathbf{b} - A\mathbf{x} = \text{proj}_{\text{null}(A^T)}\mathbf{b}$ is equivalent to finding all vectors that satisfy $\mathbf{b} - A\mathbf{x} \in \text{null}(A^T)$. This is equivalent to $A^T(\mathbf{b} - A\mathbf{x}) = 0$, or, $A^T A\mathbf{x} = A^T\mathbf{b}$.

The other statements can be proved with Theorem 5.7.6 and Theorem 5.5.4. □

Theorem 5.8.3. A vector $\hat{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ iff the error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of A.

5.8.2 Fitting a Curve to Experimental Data

A common problem in experimental work is to obtain a mathematical relationship between two variables x and y by "fitting" a curve $y = f(x)$ of a specified form to set of points in the plane that correspond to experimentally determined values of x and y. The curve $y = f(x)$ is called a *mathematical model* for the data. Here we will find a polynomial of the form

$$y = P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

whose graph comes *as close as possible* to passing through n known data points

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

That is, we want to find the least squares solutions of $M\mathbf{v} = \mathbf{y}$ where

$$M = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

which minimizes the *residue* S

$$S = \{y_1 - P(x_1)\}^2 + \{y_2 - P(x_2)\}^2 + \cdots + \{y_n - P(x_n)\}^2$$

So we can find the least squares solutions by solving the normal system $M^T M\mathbf{v} = M^T\mathbf{y}$. The polynomial $P(x)$ that results from the least squares solution is called the *least squares polynomial of best fit* or the *regression curve.*

5.9 Orthonormal Bases and the Gram-Schmidt Process

5.9.1 Orthogonal Bases and Orthonormal Bases

Definition 5.9.1. A set of vectors in $\mathbb{R}^n$ is said to be *orthogonal* if each pair of distinct vectors in the set is orthogonal, and it is said to be *orthonormal* if it is orthogonal and each vector has length 1.

Theorem 5.9.1. An orthogonal set of nonzero vectors in $\mathbb{R}^n$ is linearly independent.

5.9.2 Orthogonal Projections Using Orthonormal Bases

The standard matrix of the orthogonal projection P onto a subspace W is

$$P = M(M^T M)^{-1} M^T = MM^T$$

where the column vectors of M are orthonormal and span W, since $M^T M = I$ by Theorem 4.2.3.

Theorem 5.9.2. If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an orthogonal basis for a subspace W of $\mathbb{R}^n$, then the orthogonal projection of a vector $\mathbf{x}$ in $\mathbb{R}^n$ onto W can be expressed as

$$\text{proj}_W \mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\mathbf{x} \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_k + \cdots + \frac{\mathbf{x} \cdot \mathbf{v}_k}{\|\mathbf{v}_1\|^2}\mathbf{v}_k$$

Theorem 5.9.3. If P is the standard matrix for an orthogonal projection of $\mathbb{R}^n$ onto a subspace $W \subset \mathbb{R}^n$, then $\text{tr}(P) = \text{rank}(P) = \dim(W)$.

Sketch of Proof. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an orthonormal basis for W.

$$P = \mathbf{v}_1\mathbf{v}_1^T + \mathbf{v}_2\mathbf{v}_2^T + \cdots + \mathbf{v}_k\mathbf{v}_k^T$$

Then, $\text{tr}(P) = \text{tr}(\mathbf{v}_1\mathbf{v}_1^T) + \cdots + \text{tr}(\mathbf{v}_k\mathbf{v}_k^T) = \mathbf{v}_1 \cdot \mathbf{v}_1 + \cdots + \mathbf{v}_k \cdot \mathbf{v}_k = k = \dim(W)$, since each of $\mathbf{v}_1, \cdots, \mathbf{v}_k$ has length 1. By Theorem 5.7.5, $\dim(W) = \text{rank}(P)$. □

Theorem 5.9.4. If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, and if $\mathbf{w}$ is a vector in W, then the linear combination of the orthonormal basis vectors for $\mathbf{w}$ is

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{w} \cdot \mathbf{v}_2)\mathbf{v}_2 + \cdots + (\mathbf{w} \cdot \mathbf{v}_k)\mathbf{v}_k$$

5.9.3 Finding Orthonormal Bases

Theorem 5.9.5. Every nonzero subspace of $\mathbb{R}^n$ has an orthonormal basis.

Sketch of Proof. Let W be a nonzero subspace of $\mathbb{R}^n$, $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ be any basis for W. Then the following procedure produces an orthogonal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ for W.

1. Let $\mathbf{v}_1 = \mathbf{w}_1$.
2. Let $\mathbf{v}_2 = \mathbf{w}_2 - \text{proj}_{\text{span}\{\mathbf{v}_1\}}\mathbf{w}_2 = \mathbf{w}_2 - \dfrac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1$.
3. Let $\mathbf{v}_3 = \mathbf{w}_3 - \text{proj}_{\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}}\mathbf{w}_3 = \mathbf{w}_3 - \left[\dfrac{\mathbf{w}_3 \cdot \mathbf{v}_1}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \dfrac{\mathbf{w}_3 \cdot \mathbf{v}_2}{\|\mathbf{v}_2\|^2}\mathbf{v}_2\right]$.
4. $\cdots$
5. Let $\mathbf{v}_k = \mathbf{w}_k - \text{proj}_{\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{k-1}\}}\mathbf{w}_k = \mathbf{w}_k - \displaystyle\sum_{i=1}^{k-1} \frac{\mathbf{w}_k \cdot \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i$.

Therefore,

$$\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \cdots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}$$

forms an orthonormal basis for W. This algorithm is called the *Gram-Schmidt process.* □

Theorem 5.9.6. Every orthonormal set in a subspace $W \subset \mathbb{R}^n$ can be enlarged to an orthonormal basis for W.

5.10 QR-Decomposition; Householder Transformations

5.10.1 QR-Decomposition

Theorem 5.10.1 (QR-Decomposition)**.** If A is an $m \times k$ matrix with full column rank, then A can be factored as $A = QR$ where Q is an $m \times k$ matrix whose column vectors form an orthonormal basis for the column space of A and R is a $k \times k$ invertible upper triangular matrix.

Sketch of Proof.

$$A = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_k \end{bmatrix} \xrightarrow{\text{Gram-Schmidt}} Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k \end{bmatrix}$$

Let

$$R = \begin{bmatrix} \mathbf{w}_1 \cdot \mathbf{q}_1 & \mathbf{w}_2 \cdot \mathbf{q}_1 & \cdots & \mathbf{w}_k \cdot \mathbf{q}_1 \\ 0 & \mathbf{w}_2 \cdot \mathbf{q}_2 & \cdots & \mathbf{w}_k \cdot \mathbf{q}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{w}_k \cdot \mathbf{q}_k \end{bmatrix}$$

then $A = QR$ by Theorem 5.9.4. □

Since $Q^TQ = I$, $Q^TA = Q^TQR = R$.

Theorem 5.10.2. If A is an $m \times k$ matrix with full column rank, and if $A = QR$ is a QR-decomposition of A, then the normal system for $A\mathbf{x} = \mathbf{b}$ can be expressed as $R\mathbf{x} = Q^T\mathbf{b}$, and the least squares solution can be expressed as $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.[5]

5.10.2 Householder Reflections

Definition 5.10.1. The *reflection of* $\mathbf{x}$ *about the hyperplane* $\mathbf{a}^\perp$ is defined as

$$\text{refl}_{\mathbf{a}^\perp}\mathbf{x} = \mathbf{x} - 2\text{proj}_{\mathbf{a}}\mathbf{x}$$

where $\mathbf{a}$ is a nonzero vector in $\mathbb{R}^n$ and $\mathbf{x}$ is any vector in $\mathbb{R}^n$.

The standard matrix $H_{\mathbf{a}^\perp}$ of the reflection about the hyperplane $\mathbf{a}^\perp$ is

$$H_{\mathbf{a}^\perp} = I - \frac{2}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T = I - 2\mathbf{u}\mathbf{u}^T$$

where $\mathbf{u} = \mathbf{a}/\|\mathbf{a}\|$. This matrix is called a *Householder matrix*.

Theorem 5.10.3. Householder matrix is symmetric and orthogonal.

Theorem 5.10.4. If $\mathbf{v}$ and $\mathbf{w}$ are distinct vectors in $\mathbb{R}^n$ with the same length, then the Householder reflection about the hyperplane $(\mathbf{v} - \mathbf{w})^\perp$ maps $\mathbf{v}$ into $\mathbf{w}$, and conversely.

5.10.3 QR-Decomposition Using Householder Reflections

Example. Let us find a QR-decomposition of

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

The standard matrix for the Householder reflection that maps $(1, 1, 1, 1)$ into $(2, 0, 0, 0)$ is

$$Q_1 = I - \frac{2}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

[5] Remember as a *plausible* formula $\hat{\mathbf{x}} = (QR)^{-1}\mathbf{b}$.

where $\mathbf{a} = (2,0,0,0) - (1,1,1,1)$. Then we have

$$Q_1 A = \begin{bmatrix} 2 & 4 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 3 \end{bmatrix}$$

The standard matrix for the Householder reflection that maps $(0,0,2)$ into $(2,0,0)$ is

$$H_2 = I - \frac{2}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

where $\mathbf{a} = (2,0,0) - (0,0,2)$. Let us form the *orthogonal* matrix Q_2 as follows:

$$Q_2 = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & H_2 & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right]$$

in order to introduce zeros without destroying the zeros already created. Then,

$$Q_2 Q_1 A = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right] \left[\begin{array}{c|cc} 2 & 4 & 5 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 3 \end{array}\right] = \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

If you want to continue to introduce zeros, you can form the orthogonal matrix Q_3 of the form (which can be used to find a QR-decomposition of a 4×4 invertible matrix)

$$Q_3 = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & * & * \\ 0 & 0 & * & * \end{array}\right]$$

So we have the following decomposition of A:

$$A = Q_1^T Q_2^T (Q_2 Q_1 A) = \left[\begin{array}{ccc|c} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right] \left[\begin{array}{ccc} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \\ \hline 0 & 0 & 0 \end{array}\right]$$

Therefore, a QR-decomposition of A is

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

5.11 Coordinates with Respect to a Basis

5.11.1 Coordinates with Respect to a Basis

Definition 5.11.1. If $B = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an ordered basis for a subspace $W \subset \mathbb{R}^n$, and if $\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$, then:

- *Coordinates of* $\mathbf{w}$ *with respect to* B: $a_1, a_2, \cdots, a_k$
- v_i*-coordinate of* $\mathbf{w}$: a_i
- *Coordinate vector for* $\mathbf{w}$ *with respect to* B: $(\mathbf{w})_B = (a_1, a_2, \cdots, a_k)$
- *Coordinate matrix for* $\mathbf{w}$ *with respect to* B: $[\mathbf{w}]_B = [\, a_1 \ a_2 \ \cdots \ a_k \,]^T$

Example.

- If S is the standard basis for $\mathbb{R}^n$, then $(\mathbf{w})_S = \mathbf{w}$.
- If $B = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an orthonormal basis for a subspace $W \subset \mathbb{R}^n$, and if $\mathbf{w}$ is a vector in W, then

$$(\mathbf{w})_B = (\mathbf{w} \cdot \mathbf{v}_1, \mathbf{w} \cdot \mathbf{v}_2, \cdots, \mathbf{w} \cdot \mathbf{v}_k)$$

  by Theorem 5.9.4.

Theorem 5.11.1. If B is an orthonormal basis for a subspace $W \subset \mathbb{R}^n$, and if $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ are vectors in W, then:

i. $\|\mathbf{w}\| = \|(\mathbf{w})_B\|$

ii. $\mathbf{u} \cdot \mathbf{v} = (\mathbf{u})_B \cdot (\mathbf{v})_B$

Sketch of Proof. An orthogonal linear operator on $\mathbb{R}^n$ is norm-preserving and dot product preserving. (Theorem 4.2.1) □

5.11.2 Change of Basis for $\mathbb{R}^n$

Theorem 5.11.2. If W is a vector in $\mathbb{R}^n$, and if $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}'_1, \cdots, \mathbf{v}'_n\}$ are bases for $\mathbb{R}^n$, then $[\mathbf{w}]_{B'} = P_{B \to B'}[\mathbf{w}]_B$ where

$$P_{B \to B'} = \Big[\, [\mathbf{v}_1]_{B'} \;\Big|\; [\mathbf{v}_2]_{B'} \;\Big|\; \cdots \;\Big|\; [\mathbf{v}_n]_{B'} \,\Big]$$

which is called the *transition matrix* (or the *change of coordinates matrix*) from B to B'.

Sketch of Proof. Let $[B]$ and $[B']$ be matrices where

$$[B] = [\, \mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n \,], \quad [B'] = [\, \mathbf{v}'_1 \quad \mathbf{v}'_2 \quad \cdots \quad \mathbf{v}'_n \,]$$

Then we have $\mathbf{w} = [B][\mathbf{w}]_B$ and

$$\mathbf{v}_1 = [B'][\mathbf{v}_1]_{B'}, \ \mathbf{v}_2 = [B'][\mathbf{v}_2]_{B'}, \ \cdots, \ \mathbf{v}_n = [B'][\mathbf{v}_n]_{B'}$$

Therefore $[B]$ and $[B']$ are related by

$$[B] = [B']\Big[\, [\mathbf{v}_1]_{B'} \;\Big|\; [\mathbf{v}_2]_{B'} \;\Big|\; \cdots \;\Big|\; [\mathbf{v}_n]_{B'} \,\Big]$$

Putting this formula into $\mathbf{w} = [B][\mathbf{w}]_B = [B'][\mathbf{w}]_{B'}$ yields the desired result. □

As we can see in Theorem 5.11.2, it is useful to define an $n \times n$ matrix $[B]$ for an ordered base B for $\mathbb{R}^n$ to be

$$[B] = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$$

if $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. Let's call the matrix $[B]$ as the *standard matrix* for the base B.[6] Note that from Definition 5.11.1, we have $\mathbf{w} = [B][\mathbf{w}]_B$ if

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \quad \text{or} \quad [\mathbf{w}]_B = [\, a_1 \; a_2 \; \cdots \; a_k \,]^T$$

Corollary. If P is an invertible $n \times n$ matrix, then P is the transition matrix from the basis $B = \{\mathbf{c}_1(P), \mathbf{c}_2(P), \cdots, \mathbf{c}_n(P)\}$ for $\mathbb{R}^n$ to the standard basis $S = \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. That is, $[B] = P_{B\to S}$.

Sketch of Proof. Use $\mathbf{w} = [B][\mathbf{w}]_B$ and $\mathbf{w} = [\mathbf{w}]_S$. Apply Theorem 5.11.2. □

Theorem 5.11.3. If B_1, B_2, and B_3 are bases for $\mathbb{R}^n$, then:

i. $P_{B_2\to B_3} P_{B_1\to B_2} = P_{B_1\to B_3}$

ii. $(P_{B_1\to B_2})^{-1} = P_{B_2\to B_1}, \quad (P_{B_2\to B_1})^{-1} = P_{B_1\to B_2}$

Corollary. If B and B' are bases for $\mathbb{R}^n$, then $P_{B\to B'} = [B']^{-1}[B]$.

From this corollary, we have the following algorithm for computing $P_{B\to B'}$:

$$\Big[\, [B'] \mid [B] \,\Big] \xrightarrow{\text{elem. row. op.}} \Big[\, I \mid P_{B\to B'} \,\Big]$$

Theorem 5.11.4. If B and B' are orthonormal bases for $\mathbb{R}^n$, then the transition matrices $P_{B\to B'}$ and $P_{B'\to B}$ are orthogonal.

5.11.3 Coordinate Maps

If B is a basis for $\mathbb{R}^n$, then the transformation $\mathbf{x} \to (\mathbf{x})_B$ is called the *coordinate map* for B. Since the coordinate map is a linear transformation,

$$(c\mathbf{v})_B = c(\mathbf{v})_B, \quad (\mathbf{v} + \mathbf{w})_B = (\mathbf{v})_B + (\mathbf{w})_B$$

for any vectors $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ and for any scalar c.

Theorem 5.11.5. The coordinate map for a basis B for $\mathbb{R}^n$ is a one-to-one linear operator on $\mathbb{R}^n$. If B is an orthonormal basis for $\mathbb{R}^n$, then the coordinate map for B is orthogonal.

Sketch of Proof. The standard matrix $[T]$ of $\mathbf{x} \xrightarrow{T} (\mathbf{x})_B$ is $[T] = [B]^{-1}$. □

[6] This notion is not in the textbook.

Chapter 6

Diagonalization

6.1 Matrix Representations of Linear Transformations

6.1.1 Matrix of a Linear Operator with Respect to a Basis

Theorem 6.1.1. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator and $B = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$, then $[T(\mathbf{x})]_B = [T]_B[\mathbf{x}]_B$ for every vector $\mathbf{x}$ in $\mathbb{R}^n$ where

$$[T]_B = \Big[\, [T(\mathbf{v}_1)]_B \;\big|\; [T(\mathbf{v}_2)]_B \;\big|\; \cdots \;\big|\; [T(\mathbf{v}_n)]_B \,\Big]$$

which is called the *matrix for T with respect to the basis B*.

Sketch of Proof. The superposition principle (Theorem 4.1.1) implies

$$T(\mathbf{x}) = T([B][\mathbf{x}]_B) = \begin{bmatrix} T(\mathbf{v}_1) & T(\mathbf{v}_2) & \cdots & T(\mathbf{v}_n) \end{bmatrix} [\mathbf{x}]_B$$

Since $T(\mathbf{v}_1) = [B][T(\mathbf{v}_1)]_B,\ T(\mathbf{v}_2) = [B][T(\mathbf{v}_2)]_B,\ \cdots,\ T(\mathbf{v}_n) = [B][T(\mathbf{v}_n)]_B$,

$$T(\mathbf{x}) = [B]\Big[\, [T(\mathbf{v}_1)]_B \;\big|\; [T(\mathbf{v}_2)]_B \;\big|\; \cdots \;\big|\; [T(\mathbf{v}_n)]_B \,\Big][\mathbf{x}]_B$$

From $T(\mathbf{x}) = [B][T(\mathbf{x})]_B$, we have the desired result. $\square$

Example. Let T be the linear operator on $\mathbb{R}^3$ whose standard matrix is

$$[T] = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Then the matrix for T with respect to an orthonormal basis

$$B = \left\{ \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(-\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right), \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) \right\}$$

for $\mathbb{R}^3$ is

$$[T]_B = \begin{bmatrix} \cos\frac{2\pi}{3} & -\sin\frac{2\pi}{3} & 0 \\ \sin\frac{2\pi}{3} & \cos\frac{2\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that the 3rd coordinate of $(T(\mathbf{x}))_B$ is the same as that of $(\mathbf{x})_B$.

6.1.2 Changing Bases

Theorem 6.1.2. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, and if $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}'_1, \cdots, \mathbf{v}'_n\}$ are bases for $\mathbb{R}^n$, then $[T]_B$ and $[T]_{B'}$ are related by the equation

$$[T]_{B'} = P[T]_B P^{-1} \quad \text{where} \quad P = P_{B \to B'}$$

Moreover, if B and B' are orthonormal bases, then $[T]_{B'} = P[T]_B P^T$.

Sketch of Proof. An alternative way for $[\mathbf{x}]_{B'} \xrightarrow{[T]_{B'}} [T(\mathbf{x})]_{B'}$ is

$$[\mathbf{x}_{B'}] \xrightarrow{P_{B' \to B}} [\mathbf{x}]_B \xrightarrow{[T]_B} [T(\mathbf{x})]_B \xrightarrow{P_{B \to B'}} [T(\mathbf{x})]_{B'}$$

Therefore $[T]_{B'} = P_{B \to B'}[T]_B P_{B' \to B}$. □

We can see that $[T]$ and $[T]_B$ are related by the equation by $[T]_B = [B]^{-1}[T][B]$.

6.1.3 Matrix of a Linear Transformation with Respect to a Pair of Basis

Theorem 6.1.3. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and if $B = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $B' = \{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ are bases for $\mathbf{R}^n$ and $\mathbf{R}^m$, respectively, then $[T(\mathbf{x})]_{B'} = [T]_{B',B}[\mathbf{x}]_B$ for every vector $\mathbf{x}$ in $\mathbb{R}^n$ where

$$[T]_{B',B} = \Big[\, [T(\mathbf{v}_1)]_{B'} \;\Big|\; [T(\mathbf{v}_2)]_{B'} \;\Big|\; \cdots \;\Big|\; [T(\mathbf{v}_n)]_{B'} \,\Big]$$

which is called the *matrix for T with respect to the basis B and B'*.

Suppose that B_1, B_2 are bases for $\mathbb{R}^n$ and B'_1, B'_2 are bases for $\mathbb{R}^m$. Then,

$$[T]_{B'_2,B_2} = \left(P_{B'_2 \to B'_1}\right)^{-1} [T]_{B'_1,B_1} P_{B_2 \to B_1}$$

6.2 Similarity and Diagonalizability

6.2.1 Similar Matrices

Definition 6.2.1. If A and C are square matrices with the same size, then we say C *is similar to* A if there is an invertible matrix P such that $C = P^{-1}AP$.

Theorem 6.2.1. Two square matrices are similar iff there exist bases with respect to which the matrices represent the same linear operator.

Theorem 6.2.2 (Similarity Invariants)**.** Similar matrices have the same determinant, rank, nullity, trace, and characteristic polynomial (hence the same eigenvalues and algebraic multiplicities).

Sketch of Proof. Suppose that $C = P^{-1}AP$ for an invertible matrix P.

i. Since P is invertible, the row space of A is the same as that of PA.

$$\text{rank}(P^{-1}AP) = \text{rank}(AP) = \text{rank}(P^TA^T) = \text{rank}(A^T) = \text{rank}(A)$$

ii. Commuting the matrix product does not change the trace.

$$\operatorname{tr}(P^{-1}AP) = \operatorname{tr}(APP^{-1}) = \operatorname{tr}(A)$$

iii. $\lambda I - C = P^{-1}(\lambda I - A)P$

The other ones are straightforward. □

Note that if any of them is different, then the matrices are not similar.

6.2.2 Eigenvalues and Eigenvectors of Similar Matrices

Definition 6.2.2. The dimension of the eigenspace of A corresponding to an eigenvalue λ of A is called the *geometric multiplicity* of λ.

The geometric multiplicity can be easily determined using the dimension theorem (Theorem 5.4.1).

Theorem 6.2.3. Similar matrices have the same geometric multiplicity.

Theorem 6.2.4. Suppose that $C = P^{-1}AP$ and that λ is an eigenvalue of A.

i. If $\mathbf{x}$ is an eigenvector of C corresponding to λ, then $P\mathbf{x}$ is an eigenvector of A corresponding to λ.

ii. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, then $P^{-1}\mathbf{x}$ is an eigenvector of C corresponding to λ.

6.2.3 Diagonalization

Definition 6.2.3. If there exists an invertible matrix P such that $P^{-1}AP$ is diagonal, then A is said to be *diagonalizable*, and P is said to *diagonalize* A.

Theorem 6.2.5. An $n \times n$ matrix A is diagonalizable iff A has n linearly independent eigenvectors. Moreover,

$$P^{-1}AP = D \quad \text{where} \quad P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix},\ D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

in which $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n$ are n linearly independent eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively.

Sketch of Proof. $D = P^{-1}AP \Leftrightarrow PD = AP \wedge P$ is invertible. That is,

$$\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix} D = A \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix}$$

and $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n$ are linearly independent. This equation becomes

$$\begin{bmatrix} \lambda_1\mathbf{p}_1 & \lambda_2\mathbf{p}_2 & \cdots & \lambda_n\mathbf{p}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{p}_1 & A\mathbf{p}_2 & \cdots & A\mathbf{p}_n \end{bmatrix}$$

So $P^{-1}AP = D$ is equivalent to $A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, A\mathbf{p}_2 = \lambda_2\mathbf{p}_2, \cdots, A\mathbf{p}_n = \lambda_n\mathbf{p}_n$ where $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n$ are linearly independent, which means A has n linearly independent eigenvectors, namely, $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n$ corresponding to the eigenvalues of A, $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively. □

Theorem 6.2.6. The union of bases for all eigenspaces of a matrix A is linearly independent.

Sketch of Proof. Suppose that A has k distinct eigenvalues, namely, $\lambda_1, \lambda_2, \cdots, \lambda_k$, and that $\mathbf{v}_{i1}\mathbf{v}_{12}, \cdots, \mathbf{v}_{ir_i}$ form a basis for the eigenspace of A corresponding to λ_i. So we have to prove that the only scalars that meet the equation

$$\sum_{i=1}^{r_1} c_{1i}\mathbf{v}_{1i} + \sum_{i=1}^{r_2} c_{2i}\mathbf{v}_{2i} + \cdots + \sum_{i=1}^{r_k} c_{ki}\mathbf{v}_{ki} = 0$$

are $c_{11} = \cdots = c_{1r_1} = c_{21} = \cdots = c_{2r_2} = \cdots = c_{kr_k} = 0$. Multiplying both sides of the equation by A and λ_1 yields

$$\sum_{i=1}^{r_1} \lambda_1 c_{1i}\mathbf{v}_{1i} + \sum_{i=1}^{r_2} \lambda_2 c_{2i}\mathbf{v}_{2i} + \cdots + \sum_{i=1}^{r_k} \lambda_k c_{ki}\mathbf{v}_{ki} = 0$$

$$\sum_{i=1}^{r_1} \lambda_1 c_{1i}\mathbf{v}_{1i} + \sum_{i=1}^{r_2} \lambda_1 c_{2i}\mathbf{v}_{2i} + \cdots + \sum_{i=1}^{r_k} \lambda_1 c_{ki}\mathbf{v}_{ki} = 0$$

Therefore we have

$$\sum_{i=1}^{r_2}(\lambda_2 - \lambda_1)c_{2i}\mathbf{v}_{2i} + \sum_{i=1}^{r_3}(\lambda_3 - \lambda_1)c_{3i}\mathbf{v}_{3i} + \cdots + \sum_{i=1}^{r_k}(\lambda_k - \lambda_1)c_{ki}\mathbf{v}_{ki} = 0$$

Since $\lambda_1, \lambda_2, \cdots, \lambda_k$ are all distinct, this equation is equivalent to[1]

$$\sum_{i=1}^{r_2} c_{2i}\mathbf{v}_{2i} + \sum_{i=1}^{r_3} c_{3i}\mathbf{v}_{3i} + \cdots + \sum_{i=1}^{r_k} c_{ki}\mathbf{v}_{ki} = 0$$

In the same way, this equation can be reduced to

$$\sum_{i=1}^{r_3} c_{3i}\mathbf{v}_{3i} + \sum_{i=1}^{r_4} c_{4i}\mathbf{v}_{4i} + \cdots + \sum_{i=1}^{r_k} c_{ki}\mathbf{v}_{ki} = 0$$

By induction, we eventually have

$$\sum_{i=1}^{r_k} c_{ki}\mathbf{v}_{ki} = 0$$

Since $\{\mathbf{v}_{k1}, \cdots, \mathbf{v}_{kr_k}\}$ is linearly independent, $c_{k1} = \cdots = c_{kr_k} = 0$. Substituting this backward yields the desired result. □

Corollary. An $n \times n$ matrix with n distinct real eigenvalues is diagonalizable.

Corollary. An $n \times n$ matrix A is diagonalizable iff the sum of the geometric multiplicities of its eigenvalues is n.

Theorem 6.2.7. If A is a square matrix, then the geometric multiplicity of an eigenvalue of A is less than or equal to its algebraic multiplicity.

[1]Consider the elementary row operations.

Sketch of Proof. Suppose that λ_0 is an eigenvalue of an $n \times n$ matrix A with geometric multiplicity k. Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a basis for the eigenspace of A corresponding to λ_0, and $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be an extended basis for $\mathbb{R}^n$. Then let P be the matrix whose successive column vectors are $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. Since P is invertible, we have

$$C = \begin{bmatrix} \lambda_0 I_k & X \\ 0 & Y \end{bmatrix} = P^{-1}AP$$

where X is an $k \times (n-k)$ matrix and Y is an $(n-k) \times (n-k)$ matrix. Therefore C is similar to A hence has the same characteristic polynomial as that of A. Since $\lambda I - C$ is of the block upper triangular form,

$$\lambda I - C = \begin{bmatrix} (\lambda - \lambda_0) I_k & X \\ 0 & \lambda I_{n-k} - Y \end{bmatrix}$$

we have $\det(\lambda I - C) = (\lambda - \lambda_0)^k \det(\lambda I_{n-k} - Y)$. Therefore the algebraic multiplicity of λ is at least k. □

Corollary. *A is diagonalizable iff the geometric multiplicity of each eigenvalue of A is the same as its algebraic multiplicity.*

Theorem 6.2.8 (Characterizations of Diagonalizability)**.** *If A is an $n \times n$ matrix, then the following statements are equivalent.*

i. *A is diagonalizable.*

ii. *A has n linearly independent eigenvectors.*

iii. *The sum of the geometric multiplicities of the eigenvalues of A is n.*

iv. *$\mathbb{R}^n$ has a basis consisting of eigenvectors of A.*

v. *The geometric multiplicity of each eigenvalue of A is the same as its algebraic multiplicity.*

6.3 Orthogonal Diagonalizability; Functions of a Matrix

6.3.1 Orthogonal Similarity

Definition 6.3.1. If A and C are square matrices with the same size, then we say C *is orthogonally similar to* A if there is an orthogonal matrix P such that $C = P^T AP$.

Theorem 6.3.1. *Two matricies are orthogonally similar iff there exist orthonormal bases with respect to which the matrices represent the same linear operator.*

6.3.2 Orthogonal Diagonalization

Definition 6.3.2. If there exists an orthogonal matrix P such that P^TAP is diagonal, then A is said to be *orthogonally diagonalizable*, and P is said to *orthogonally diagonalize* A.

Theorem 6.3.2. An $n \times n$ matrix A is orthogonally diagonalizable iff A has n orthonormal eigenvectors. Moreover,

$$P^TAP = D \quad \text{where} \quad P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix},\ D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

in which $\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n$ are n orthonormal eigenvectors of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectively.

Apply the Gram-Schmidt process to each of bases for the eigenspaces of A to produce an orthonormal set of n eigenvectors of A.

Theorem 6.3.3.

i. If A is a symmetric matrix, then eigenvectors from different eigenspaces are orthogonal.

ii. A matrix A is orthogonally diagonalizable iff A is symmetric.

Sketch of Proof for (i). Let $\mathbf{v}_1$ and $\mathbf{v_2}$ be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2. Then, $\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\mathbf{v}_1^T A^T)\mathbf{v}_2 = \mathbf{v}_1^T(A\mathbf{v}_2) = \lambda_2 \mathbf{v}_1^T \mathbf{v}_2$. So we have $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$ hence $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Sketch of Proof for (ii). Let us prove that an $n \times n$ matrix A is orthogonally diagonalizable if A is symmetric. Suppose that λ_0 is an eigenvalue of A with geometric multiplicity k. Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be an orthonormal basis for the eigenspace of A corresponding to λ_0, and $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be an orthonormally extended basis for $\mathbb{R}^n$. Then let P be the matrix whose successive column vectors are $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. Since P is orthogonal, we have

$$C = \begin{bmatrix} \lambda_0 I_k & 0 \\ 0 & Y \end{bmatrix} = P^TAP$$

where Y is an $(n-k) \times (n-k)$ matrix. Therefore C is orthogonally similar to A hence has the same characteristic polynomial as that of A. Since $\lambda I - C$ is of the block diagonal form,

$$\lambda I - C = \begin{bmatrix} (\lambda - \lambda_0) I_k & 0 \\ 0 & \lambda I_{n-k} - Y \end{bmatrix}$$

we have $\det(\lambda I - C) = (\lambda - \lambda_0)^k \det(\lambda I_{n-k} - Y)$. Since Y cannot have eigenvalue λ_0, the algebraic multiplicity of λ is k. Therefore A is diagonalizable. By (i), A is orthogonally diagonalizable. □

6.3.3 Spectral Decomposition

Theorem 6.3.4 (Spectral Decomposition)**.** If A is a symmetric matrix that is orthogonally diagonalized by $P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$ where $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are eigenvectors of A that correspond to the eigenvalues of A, namely, $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then, $A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$. This is called a *spectral decomposition* or an *eigenvalue decomposition* of A.

Sketch of Proof.

$$A = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_n^T \end{bmatrix}$$

□

Theorem 6.3.5. Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$. If A can be expressed as $A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T$, then A is symmetric and has eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$.

Note that if $\mathbf{u}$ is a unit vector in $\mathbb{R}^n$, then $\mathbf{u}\mathbf{u}^T$ is the standard matrix for the orthogonal projection onto span$\{\mathbf{u}\}$. The spectal decomposition of A can be interpreted as the sum of the standard matrices of the projection onto span of each $\mathbf{u}_i$ multiplied by λ_i.

6.3.4 Powers of a Diagonalizable Matrix

If a matrix A is diagonalized by an invertible matrix P and $D = P^{-1}AP$, then $A^k = PD^kP^{-1}$ for any positive integer k. In the special case where A is a symmetric matrix and is orthogonally diagonalized by $P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$, we have $A^k = PD^kP^T$, or

$$A = \lambda_1^k \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2^k \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n^k \mathbf{u}_n \mathbf{u}_n^T$$

for any positive integer k.

Theorem 6.3.6 (The Cayley-Hamilton Theorem)**.** Every square matrix satisfies its characteristic equation. That is, if A is a square matrix whose characteristic equation is $\chi_A(\lambda) = 0$, then $\chi_A(A) = 0$.

The Cayley-Hamilton thereom makes it possible to express all positive integer powers of an $n \times n$ matrix A in terms of $I, A, \cdots, A^{n-1}$. Moreover, if A is invertible, then it also makes it possible to express A^{-1} (hence all negative powers of A) in terms of $I, A, \cdots, A^{n-1}$.

6.3.5 Functions of a Matrix

Suppose that f is a real-valued function whose Maclaurin series converges on some proper interval. If D is an $n \times n$ diagonal matrix, then

$$f((D)_{ii}) = f(0) + f'(0)(D)_{ii} + \frac{f''(0)}{2!}(D)_{ii}^2 + \cdots + \frac{f^{(k)}(0)}{k!}(D)_{ii}^k + \cdots$$

So we *define* the function f of the diagonal matrix D as

$$f(D) = \begin{bmatrix} f((D)_{11}) & 0 & \cdots & 0 \\ 0 & f((D)_{22}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f((D)_{nn}) \end{bmatrix}$$

Theorem 6.3.7. Suppose that A is an $n \times n$ diagonalizable matrix that is diagonalized by P and that $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of A corresponding to the successive column vectors of P. If f is a real-valued function whose Maclaurin series converges on some interval containing the eigenvalues of A, then

$$f(A) = P \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix} P^{-1}$$

6.4 Quadratic Forms

6.4.1 Quadratic Forms

Definition 6.4.1. The *quadratic form* associated with an $n \times n$ symmetric matrix A is

$$Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = a_{11}x_1^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \cdots + a_{22}x_2^2 + \cdots + a_{nn}x_n^2$$

where

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

6.4.2 Change of Variable in Quadratic Forms

Theorem 6.4.1 (The Principal Axes Theorem)**.** Suppose that A is an $n \times n$ symmetric matrix and is orthogonally diagonalized by P. Then making the change of variable $\mathbf{x} = P\mathbf{y}$ in $\mathbf{x}^T A\mathbf{x}$ yields the quadratic form

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

where $D = P^T AP$ and $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of A corresponding to the eigenvectors that form the successive columns of P.

Example. $5x_1^2 - 4x_1x_2 + 8x_2^2 = 36$ becomes $4y_1^2 + 9y_2^2 = 36$ if

$$A = \begin{bmatrix} 5 & -2 \\ -2 & 8 \end{bmatrix}, \quad P = \begin{bmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

As you can easily see, if A is a 2×2 symmetric matrix and λ_1, λ_2 are its eigenvalues, then:

- $\mathbf{x}^T A\mathbf{x} = 1$ represents an ellipse if $\lambda_1, \lambda_2 > 0$.
- $\mathbf{x}^T A\mathbf{x} = 1$ represents a hyperbola if $\lambda_1\lambda_2 < 0$.
- $\mathbf{x}^T A\mathbf{x} = 1$ has no graph if $\lambda_1, \lambda_2 < 0$.

6.4.3 Positive Definite Quadratic Forms

Definition 6.4.2. A quadratic form $\mathbf{x}^T A\mathbf{x}$ is said to be

- *positive definite* if $\mathbf{x}^T A\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$.
- *positive semidefinite* if $\mathbf{x}^T A\mathbf{x} \geq 0$ for $\mathbf{x} \neq \mathbf{0}$.
- *negaitve definite* if $\mathbf{x}^T A\mathbf{x} < 0$ for $\mathbf{x} \neq \mathbf{0}$.
- *negative semidefinite* if $\mathbf{x}^T A\mathbf{x} \leq 0$ for $\mathbf{x} \neq \mathbf{0}$.
- *indefinite* if $\mathbf{x}^T A\mathbf{x}$ has both positive and negative values.

Theorem 6.4.2. If A is a symmetric matrix, then:

i. $\mathbf{x}^T A\mathbf{x}$ is positive (resp. negative) definite iff all eigenvalues of A are positive (resp. negative).

ii. $\mathbf{x}^T A\mathbf{x}$ is positive (resp. negative) semidefinite iff all eigenvalues of A are nonnegative (resp. nonpositive).

iii. $\mathbf{x}^T A\mathbf{x}$ is indefinite iff A has at least on positive and negative eigenvalues.

Corollary. If A is a 2×2 symmetric matrix, then:

- $\mathbf{x}^T A\mathbf{x} = 1$ represents an ellipse if A is positive definite.
- $\mathbf{x}^T A\mathbf{x} = 1$ represents a hyperbola if A is indefinite.
- $\mathbf{x}^T A\mathbf{x} = 1$ has no graph if A is negative definite.

Theorem 6.4.3 (Sylvester's Criterion)**.** A symmetric matrix A is positive definite iff the determinant of every principal submatrix[2] is positive.

Theorem 6.4.4. If A is a symmetric matrix, then the following statements are equivalent:

i. A is positive definite.

ii. There is a symmetric positive defnite matrix B such that $A = B^2$.

iii. There is an invertible matrix C such that $A = C^T C$.

Sketch of Proof. Take $B = P\sqrt{D}P^T = C$. □

[2] *k-th principal submatrix* of a square matrix A is the $k \times k$ submatrix consisting of the first k rows and columns of A.

6.5 Singular Value Decomposition

6.5.1 Singular Value Decomposition

Theorem 6.5.1 (Singular Value Decomposition)**.** If A is an $m \times n$ matrix of rank k, then A can be factored as $A = U\Sigma V^T$, called the *singular value decomposition* of A, where U is an $m \times m$ orthogonal matrix,

$$U = \left[\begin{array}{cccc|ccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_m \end{array}\right]$$

Σ is an $m \times n$ matrix,

$$\Sigma = \left[\begin{array}{cccc|ccc} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_k & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right]$$

and V is an $n \times n$ orthogonal matrix.

$$V = \left[\begin{array}{cccc|ccc} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k & \mathbf{v}_{k+1} & \cdots & \mathbf{v}_n \end{array}\right]$$

Sketch of Proof.

1. Find the $n \times n$ matrix V that orthogonally diagonalizes A^TA.

$$A^TA = VDV^T,\ V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix},\ D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

   with nonincreasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

   i. $\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x} \cdot (A^TA\mathbf{x}) = \mathbf{x} \cdot \lambda\mathbf{x} = \lambda\|\mathbf{x}\|^2 \quad \therefore \lambda \geq 0.$

   ii. $\text{rank}(A) = \text{rank}(A^TA) = \text{rank}(D) = k$,
   $\therefore \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$, $\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = 0$.

   iii. $\{\mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \cdots, \mathbf{v}_n\}$ is an orthonormal set of $n-k$ nonzero vectors in null(A), hence forms an orthonormal basis foor null(A). So $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ forms an orthonormal basis for null$(A)^\perp$ = row(A).

2. Normalize $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_k\}$ to obtain orthonormal vectors $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$.

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}, \text{ or equivalently, } \sqrt{\lambda_i}\mathbf{u}_i = A\mathbf{v}_i$$

   i. $\|A\mathbf{v}_i\|^2 = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i \quad \therefore \|A\mathbf{v}_i\| = \sqrt{\lambda_i}$

   ii. $(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \mathbf{v}_i \cdot (A^TA\mathbf{v}_j) = \mathbf{v}_i \cdot (\lambda_j\mathbf{v}_j) = 0$ for $i \neq j$. Therefore, $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_k\}$ is an orthogonal set of k nonzero vectors in col(A) so $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ forms an orthonormal basis for col(A).

3. Extend $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\}$ to form an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ for $\mathbb{R}^m$.

   i. $\{\mathbf{u}_{k+1}, \cdots, \mathbf{u}_m\}$ forms an orthonormal basis for $\text{col}(A)^\perp = \text{null}(A^T)$. Now we have bases for all the fundamental spaces of A.

4. Let

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix}, \; \Sigma = \left[\begin{array}{cccc|ccc} \sqrt{\lambda_1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_k} & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right]$$

   i. $U\Sigma = \left[\sqrt{\lambda_1}\mathbf{u}_1 \quad \sqrt{\lambda_2}\mathbf{u}_2 \quad \cdots \quad \sqrt{\lambda_k}\mathbf{u}_k \;\middle|\; \mathbf{0} \quad \cdots \quad \mathbf{0} \right] = AV$.

Therefore, $U\Sigma V^T = A(VV^T) = A$. □

From the transformation point of view, $A = U\Sigma V^T$ is the matrix for T_A with respect to bases $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

Definition 6.5.1. From the proof of Theorem 6.5.1,

- $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \cdots, \sigma_k = \sqrt{\lambda_k}$ are the *singular values* of A.
- $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$ are the *left singular vectors* of A.
- $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ are the *right singular vectors* of A.

We can find a singular value decomposition of a symmetric matrix A from an eigenvalue decomposition of A by shifting the negative signs from the diagonal factor to the first (or second) orthogonal factor.

6.5.2 Polar Decomposition

Theorem 6.5.2 (Polar Decomposition)**.** If A is an $n \times n$ matrix with rank k, then A can be factored as $A = PQ$ where P is an $n \times n$ positive semidefinite matrix of rank k and Q is an $n \times n$ orthogonal matrix.

Sketch of Proof. Let the singular value decomposition of A be $A = U\Sigma V^T$. Then take $P = U\Sigma U^T, Q = UV^T$. □

6.5.3 Reduced Singular Value Decomposition

Theorem 6.5.3 (Reduced Singular Value Decomposition)**.** If A is an $m \times n$ matrix of rank k, then A can be factored as $A = U_1\Sigma_1V_1^T$, called the *reduced singular value decompsotion* of A, where U_1 is an $m \times k$ orthogonal matrix,

$$U_1 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{bmatrix}$$

Σ_1 is a $k \times k$ diagonal matrix,

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix}$$

and V_1 is a $k \times n$ orthogonal matrix.

$$V_1 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{bmatrix}$$

From the Theorem 6.5.3, we have the *reduced singular value expansion* of A.

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

Since the singular values of A are nonnegative and in nonincreasing order, we can approximate A by

$$A \approx \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

for some $r < k$. This is called the *rank r approximation* of A.

6.6 The Pseudoinverse

6.6.1 The Pseudoinverse

The inverse of an $n \times n$ invertible matrix A whose the reduced singular value decomposition is $A = U_1 \Sigma_1 V_1^T$ is $A^{-1} = V_1 \Sigma_1^{-1} U_1^T$. However, $V_1 \Sigma_1^{-1} U_1^T$ is well-defined even if A is neither invertible nor square.

Definition 6.6.1. If A is a nonzero $m \times n$ matrix, the *pseudoinverse* of A is defined as the following $n \times m$ matrix.

$$A^+ = V_1 \Sigma_1^{-1} U_1^T$$

If $A = 0$, we define $A^+ = 0$.

Theorem 6.6.1. If A is an $m \times n$ matrix with full column rank, then:

$$A^+ = (A^T A)^{-1} A^T$$

6.6.2 Properties of the Pseudoinverse

Theorem 6.6.2. If A^+ is the pseudoinverse of an $m \times n$ matrix A, then:

i. $AA^+A = A$

ii. $A^+AA^+ = A^+$

iii. $(AA^+)^T = AA^+$

iv. $(A^+A)^T = A^+A$

v. $(A^T)^+ = (A^+)^T$

vi. $A^{++} = A$

(v) is useful to compute A^+ when A^T has full column rank.

Theorem 6.6.3. If $A^+ = V_1\Sigma_1^{-1}U_1^T$ is the pseudoinverse of an $m \times n$ matrix A of rank k, and if the column vectors of U_1 and V_1 are $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$ and $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$, respectively, then:

i. $\forall \mathbf{y} \in \mathbb{R}^m \ A^+\mathbf{y} \in \text{row}(A)$

ii. $A^+\mathbf{u}_i = \sigma_i^{-1}\mathbf{v}_i$

iii. $\forall \mathbf{y} \in \text{null}(A^T) \ A^+\mathbf{y} = 0$

iv. $\forall \mathbf{x} \in \mathbb{R}^m \ AA^+\mathbf{x} = \text{proj}_{\text{col}(A)}\mathbf{x}$

v. $\forall \mathbf{x} \in \mathbb{R}^n \ A^+A\mathbf{x} = \text{proj}_{\text{row}(A)}\mathbf{x}$

Sketch of Proof.

i. $A^+\mathbf{y} = V_1(\Sigma_1^{-1}U_1^T\mathbf{y})$

ii. $A^+U_1 = V_1\Sigma_1^{-1}$

iii. $\text{col}(U_1) = \text{col}(A) \Rightarrow U_1^T\mathbf{y} = \mathbf{0}$

iv. $AA^+ = U_1U_1^T$

v. $A^+A = V_1V_1^T$

$\square$

6.6.3 Pseudoinverse and Least Squares

Theorem 6.6.4. If A is an $m \times n$ matrix, and $\mathbf{b}$ is any vector in $\mathbb{R}^m$, then $\mathbf{x} = A^+\mathbf{b}$ is the least squares solution of $A\mathbf{x} = \mathbf{b}$ that has minimum norm.