Linear Algebra Summary

1. Linear Equations in Linear Algebra

1.1 Definitions and Terms

1.1.1 Systems of Linear Equations

A **linear equation** in the variables x_1, x_2, $\ldots$, x_n is an equation that can be written in the form $a_1x_1 + a_2x_2 + \ldots a_nx_n = b$, where a_1, $\ldots$, a_n are the **coefficients**. A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables. A **solution** of a linear system is a list of numbers that makes each equation a true statement. The set of all possible solutions is called the **solution set** of the linear system. Two linear systems are called **equivalent** if they have the same solution set. A linear system is said to be consistent, if it has either one solution or infinitely many solutions. A system is inconsistent if it has no solutions.

1.1.2 Matrices

The essential information of a linear system can be recorded compactly in a rectangular array called a **matrix**. A matrix containing only the coefficients of a linear system is called the **coefficient matrix**, while a matrix also including the constant at the end of a linear equation, is called an **augmented matrix**. The **size** of a matrix tells how many columns and rows it has. An **m** $\times$ **n matrix** has m rows and n columns.

There are three elementary row operations. **Replacement** adds to one row a multiple of another. **Interchange** interchanges two rows. **Scaling** multiplies all entries in a row by a nonzero constant. Two matrices are **row equivalent** if there is a sequence of row operations that transforms one matrix into the other. If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

1.1.3 Matrix Types

A **leading entry** of a row is the leftmost nonzero entry in the row. A rectangular matrix is in **echelon form** (and thus called an **echelon matrix**) if all nonzero rows are above any rows of all zeros, if each leading entry of a row is in a column to the right to the leading entry of the row above it, and all entries in a column below a leading entry are zeros. A matrix in echelon form is in **reduced echelon form** if also the leading entry in each nonzero row is 1, and each leading 1 is the only nonzero entry in its column. If a matrix A is row equivalent to an echelon matrix U, we call U **an echelon form of** A.

A **pivot position** in a matrix A is a location in A the corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot position. Variables corresponding to pivot columns in the matrix are called **basic variables**. The other variables are called **free variables**. A **general solution** of a linear system gives an explicit description of all solutions.

1.1.4 Vectors

A matrix with only one column is called a **vector**. Two vectors are **equal** if, and only if, their corresponding entries are equal. A vector whose entries are all zero is called the **zero vector**, and is denoted by $\mathbf{0}$. If $\mathbf{v_1}$, ..., $\mathbf{v_p}$ are in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v_1}$, ..., $\mathbf{v_p}$ is denoted by Span{$\mathbf{v_1}$, ..., $\mathbf{v_p}$} and is called the **subset of $\mathbb{R}^n$ spanned by $\mathbf{v_1}$, ..., $\mathbf{v_p}$**. So Span{$\mathbf{v_1}$, ..., $\mathbf{v_p}$} is the collection of all vectors that can be written in the form $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}$ with c_1, c_2, ..., c_p scalars.

1.1.5 Matrix Equations

If A is an $m \times n$ matrix, with columns $\mathbf{a_1}$, ..., $\mathbf{a_n}$, and if $\mathbf{x}$ is in $\mathbb{R}^n$, then the **product of A and $\mathbf{x}$**, denoted by $A\mathbf{x}$, is the **linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights**. That is, $A\mathbf{x} = [\mathbf{a_1}\ \ \mathbf{a_2}\ \ \ldots\ \ \mathbf{a_n}]\,\mathbf{x} = x_1\mathbf{a_1} + x_2\mathbf{a_2} + \ldots + x_n\mathbf{a_n}$. $A\mathbf{x}$ is a vector in $\mathbb{R}^m$. An equation in the form $A\mathbf{x} = b$ is called a **matrix equation**.

I is called an **identity matrix**, and has 1's on the diagonal and 0's elsewhere. I_n is the identity matrix of size $n \times n$. It is always true that $I_n\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^n$.

1.1.6 Solution Sets of Linear Systems

A system of linear equations is said to be **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$. Such a system always has the solution $\mathbf{x} = \mathbf{0}$, which is called the trivial solution. The important question is whether there are **nontrivial solutions**, that is, a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = 0$. The total set of solutions can be described by a **parametric vector equation**, which is in the form $\mathbf{x} = a_1\mathbf{u_1} + a_2\mathbf{u_2} + \ldots + a_n\mathbf{u_n}$.

1.1.7 Linear Independence

An indexed set of vectors {$\mathbf{v_1}$, $\mathbf{v_2}$, ..., $\mathbf{v_p}$} in $\mathbb{R}^n$ is said to be **linearly independent** if the vector equation $x_1\mathbf{v_1} + x_2\mathbf{v_2} + \ldots + x_p\mathbf{v_p} = \mathbf{0}$ has only the trivial solution. The set is said to be linearly dependent if there exist weights c_1, c_2, ..., c_p, not all zero, such that $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p} = \mathbf{0}$. This equation is called a **linear dependence relation** among $\mathbf{v_1}$, $\mathbf{v_2}$, ..., $\mathbf{v_p}$. Also, the columns of a matrix A are linearly independent if, and only if, the equation $A\mathbf{x} = 0$ has only the trivial solution.

1.1.8 Linear Transformations

A **transformation** (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$. For $\mathbf{x}$ in $\mathbb{R}^n$, the vector $T(\mathbf{x})$ in $\mathbb{R}^m$ is called the **image** of $\mathbf{x}$. The set $\mathbb{R}^n$ is called the **domain** of T, and $\mathbb{R}^m$ is called the **codomain**. The set of all images $T(\mathbf{x})$ is called the **range** of T.

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of at least one $\mathbf{x}$ in $\mathbb{R}^n$. That is, if the range and the codomain coincide. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of at most one $\mathbf{x}$ in $\mathbb{R}^n$. If a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is both onto $\mathbb{R}^m$ and one-to-one, then for every $\mathbf{b}$ in $\mathbb{R}^m$ $A\mathbf{x} = \mathbf{b}$ has a unique solution. That is, there is exactly 1 $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$.

1.2 Theorems

1. Each matrix is row equivalent to one, and only one, reduced echelon matrix.
2. A linear system is consistent if, and only if the rightmost column of the augmented matrix is not a pivot column.
3. If a linear system is consistent, and if there are no free variables, there exists only 1 solution. If there are free variables, the solution set contains infinitely many solutions.
4. A vector equation $x_1\mathbf{a_1} + x_2\mathbf{a_2} + \ldots + x_n\mathbf{a_n} = \mathbf{b}$ has the same solution set as the linear system whose augmented matrix is $[\mathbf{a_1}\ \ \mathbf{a_2}\ \ \ldots\ \ \mathbf{a_n}\ \ \mathbf{b}]$.
5. A vector $\mathbf{b}$ is in Span$\{\mathbf{v_1},\ \ldots,\ \mathbf{v_p}\}$ if, and only if the linear system with augmented matrix $[\mathbf{v_1}\ \ \mathbf{v_2}\ \ \ldots\ \ \mathbf{v_p}\ \ \mathbf{b}]$ has a solution.
6. If A is an $m \times n$ matrix, and if $\mathbf{b}$ is in $\mathbb{R}^m$, the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution set as the linear system whose augmented matrix is $[\mathbf{a_1}\ \ \mathbf{a_2}\ \ \ldots\ \ \mathbf{a_n}\ \ \mathbf{b}]$.
7. The following four statements are equivalent for a particular $m \times n$ coefficient matrix A. That is, if one is true, then all are true, and if one is false, then all are false:
   (a) For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.
   (b) Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.
   (c) The columns of A span $\mathbb{R}^m$.
   (d) A has a pivot position in every row.
8. The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if, and only if the equation has at least one free variable.
9. If the reduced echelon form of A has d free variables, then the solution set consists of a d-dimensional plane (that is, a line is a 1-dimensional plane, a plane is a 2-dimensional plane), which can be described by the parametric vector equation $\mathbf{x} = a_1\mathbf{u_1} + a_2\mathbf{u_2} + \ldots + a_d\mathbf{u_d}$.
10. If $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$, and if $A\mathbf{p} = \mathbf{b}$, then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors $\mathbf{w} = \mathbf{p} + \mathbf{v}$ where $\mathbf{v}$ is any solution of $A\mathbf{x} = \mathbf{0}$.
11. A indexed set $S = \{\mathbf{v_1},\ \mathbf{v_2},\ \ldots,\ \mathbf{v_p}\}$ is linearly dependent if, and only if at least one of the vectors in S is a linear combination of the others.
12. If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v_1},\ \mathbf{v_2},\ \ldots,\ \mathbf{v_p}\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.
13. If a set $S = \{\mathbf{v_1},\ \mathbf{v_2},\ \ldots,\ \mathbf{v_p}\}$ contains the zero vector $\mathbf{0}$, then the set is linearly dependent.
14. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then there exists a unique matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. In fact, $A = [\ T(\mathbf{e_1})\ \ T(\mathbf{e_2})\ \ \ldots\ \ T(\mathbf{e_n})\ ]$.
15. If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and $T(\mathbf{x}) = A\mathbf{x}$, then:
    (a) T is one-to-one if, and only if the equation $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.
    (b) T is one-to-one if, and only if the columns of A are linearly independent.
    (c) T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if, and only if the columns of A span $\mathbb{R}^m$.
16. If A and B are equally sized square matrices, and $AB = I$, then A and B are both invertible, and $A = B^{-1}$ and $B = A^{-1}$.

1.3 Calculation Rules

1.3.1 Vectors

Define the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ as follows:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \tag{1.1}$$

If c is a scalar, then the following rules apply:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} \tag{1.2}$$

$$c\mathbf{u} = \begin{bmatrix} cu_1 \\ cu_2 \\ \vdots \\ cu_n \end{bmatrix} \tag{1.3}$$

1.3.2 Matrices

The product of a matrix A with size $m \times n$ and a vector $\mathbf{x}$ in $\mathbb{R}^n$ is defined as:

$$A\mathbf{x} = [\mathbf{a_1} \ \ \mathbf{a_2} \ \ \ldots \ \ \mathbf{a_n}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a_1} + x_2\mathbf{a_2} + \ldots + x_n\mathbf{a_n} \tag{1.4}$$

Now the following rules apply:

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} \tag{1.5}$$

$$A(c\mathbf{u}) = c(A\mathbf{u}) \tag{1.6}$$

1.3.3 Linear Transformations

If a transformation (or mapping) T is linear, then:

$$T(\mathbf{0}) = \mathbf{0} \tag{1.7}$$

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}) \tag{1.8}$$

Or, more general:

$$T(c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_p\mathbf{v_p}) = c_1T(\mathbf{v_1}) + c_2T(\mathbf{v_2}) + \ldots + c_pT(\mathbf{v_p}) \tag{1.9}$$

2. Matrix Algebra

2.1 Definitions and Terms

2.1.1 Matrix Entries

If A is an $m \times n$ matrix, then the scalar in the ith row and the jth column is denoted by a_{ij}. The **diagonal entries** in a matrix are the numbers a_{ij} where $i = j$. They form the **main diagonal** of A. A **diagonal matrix** is a square matrix whose nondiagonal entries are 0. An example is I_n. A matrix whose entries are all zero is called a **zero matrix**, and denoted as 0. To matrices are **equal** if they have the same size, and all their corresponding entries are equal.

2.1.2 Matrix Operations

If A and B are both $m \times n$ matrices, and $A + B = C$ then C is also an $m \times n$ matrix whose entries are the sum of the corresponding entries of A and B. If r is a scalar, then the **scalar multiple** $C = rA$ is the matrix whose entries are r times the corresponding entries of A.

Two matrices can be multiplied, by multiplying one matrix by the columns of the other matrix. If A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{b_1}$, $\mathbf{b_2}$, ..., $\mathbf{b_p}$, then the product AB is the $m \times p$ matrix $AB = A\,[\,\mathbf{b_1}\ \ \mathbf{b_2}\ \ \ldots\ \ \mathbf{b_p}] = [\,A\mathbf{b_1}\ \ A\mathbf{b_2}\ \ \ldots\ \ A\mathbf{b_p}]$. Note that usually $AB \neq BA$. If $AB = BA$, then we say that A and B **commute** with one another.

Since it is possible to multiply matrices, it is also possible to take their power. If A is a square matrix, then $A^k = A \ldots A$, where there should be k A's. Also A^0 is defined as I_n. Given an $m \times n$ matrix, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A. So $\text{row}_i(A) = \text{col}_i(A^T)$. The transpose should not be confused by a matrix to the power T.

2.1.3 Inverses

An $n \times n$ (square) matrix A is said to be invertible if there is an $n \times n$ matrix C such that $CA = I_n = AC$. In this case C is the **inverse** of A, denoted as A^{-1}. A matrix that is not invertible is called a **singular matrix**. For a 2-dimensional matrix, the quantity $a_{11}a_{22} - a_{12}a_{21}$ is called the determinant, noted as $\det A = ad - bc$. An **elementary matrix** is a matrix that is obtained by performing a single elementary row operation on an identity matrix.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. We call S the **inverse** of T and write it as $S = T^{-1}$ or $S(\mathbf{x}) = T^{-1}(\mathbf{x})$. If $T(\mathbf{x}) = A\mathbf{x}$, then A is called the **standard matrix** of the linear transformation T.

2.1.4 Subspaces

A **subspace** of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ for which three properties apply. The zero vector $\mathbf{0}$ is in H, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H, and for each $\mathbf{u}$ in H, the vector $c\mathbf{u}$ is in H (for every scalar c). Subspaces are always a point (0-dimensional) on the origin, a line (1-dimensional) through the origin, a plane (2-dimensional) through the origin, or any other multidimensional plane through the origin.

The **column space** of a matrix A is the set Col A of all linear combinations of the columns of A. The column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$. The **row space** of a matrix A is the set Row A of all linear combinations of the rows of A. The **null space** of a matrix A is the set Nul A of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$. The null space of an $m \times n$ matrix is a subspace of $\mathbb{R}^n$. A **basis** for a subspace H or $\mathbb{R}^n$ is a linearly independent set in H that spans H.

2.1.5 Dimension and Rank

Suppose the set $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ is a basis for a subspace H. For each $\mathbf{x}$ in H, the **coordinates of x relative to the basis** β are the weights $c_1, \ldots, c_p$ such that $\mathbf{x} = c_1\mathbf{b_1} + \ldots + c_p\mathbf{b_p}$. The vector $[\mathbf{x}]_\beta$ in $\mathbb{R}^p$ with coordinates $c_1, \ldots, c_p$ is called the **coordinate vector of x (relative to** β**)** or the $beta$**-coordinate vector of x**.

The **dimension** of a subspace H, denoted by dim H, is the number of vectors in any basis for H. The zero subspace has no basis, since the zero vector itself forms a linearly dependent set. The **rank** of a matrix A, denoted by rank A, is the dimension of the column space of A. So per definition rank $A =$ dim Col A.

2.1.6 Kernel and Range

Let T be a linear transformation. The **kernel** (or **null space**) of T, denoted as ker T, is the set of all $\mathbf{u}$ such that $T(\mathbf{u}) = \mathbf{0}$. The **range** of T, denoted as range T, is the set of all vectors $\mathbf{v}$ for which $T(\mathbf{x}) = \mathbf{v}$ has a solution. If $T(\mathbf{x}) = A\mathbf{x}$, then the kernel of T is the null space of A, and the range of T is the column space of A.

2.2 Theorems

1. **The Row-Column Rule**. If A is an $m \times n$ matrix, and B is an $n \times p$ matrix, then the entry in the ith row and the jth column of AB is $(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}$.
2. From the Row-Column Rule can be found that $\text{row}_i(AB) = \text{row}_i(A) \cdot B$.
3. If A has size 2×2. If $ad - bc \neq 0$, then A is invertible, and $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.
4. If A is an invertible matrix, then for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.
5. If A is invertible, then A^{-1} is invertible, and $(A^{-1})^{-1} = A$.
6. If A and B are $n \times n$ matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is, $(AB)^{-1} = B^{-1}A^{-1}$. This also goes for any number of matrices. That is, if $A_1, \ldots, A_p$ are $n \times n$ matrices, then $(A_1A_2 \ldots A_p)^{-1} = A_p^{-1} \ldots A_2^{-1}A_1^{-1}$.
7. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is, $(A^T)^{-1} = (A^{-1})^T$.
8. If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ elementary matrix E is created by performing the same row operation on I_m.

9. Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

10. An $n \times n$ matrix A is invertible if, and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

11. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and let A be the standard matrix for T. That is, $T(\mathbf{x}) = A\mathbf{x}$. Then T is invertible if, and only if A is an invertible matrix. In that case, $T^{-1}(\mathbf{x}) = A^{-1}\mathbf{x}$.

12. If $\mathbf{u_1}, \ldots, \mathbf{u_p}$ are in the subspace H, then every vector in Span$\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is in H.

13. If A is an $m \times n$ matrix with column space Col A, then Col A = Span$\{\mathbf{a_1}, \ldots, \mathbf{a_n}\}$. Also Col A is the set of all $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ has a solution.

14. The pivot columns of a matrix A form a basis for the column space of A.

15. The dimension of Nul A is equal to the number of free variables in $A\mathbf{x} = \mathbf{0}$.

16. The dimension of Col A (which is rank A) is equal to the number of pivot columns in A.

17. **The Rank Theorem**. If a matrix A has n columns, then dim Col A + dim Nul A = rank A + dim Nul $A = n$.

18. **The Basis Theorem**. Let H be a p-dimensional subspace of $\mathbb{R}^n$. Any linearly independent set of exactly p elements in H is automatically a basis for H. Also any set of p elements of H that spans H is automatically a basis for H.

19. If the linear transformation $T(\mathbf{x}) = A\mathbf{x}$, then ker T = Nul A and range T = Col A.

20. If $\mathbb{R}^n$ is the domain of T, then dim ker T + dim range $T = n$.

21. If two matrices A and B are row equivalent, then their row spaces are the same. If B is in echelon form, the nonzero rows of B form a basis for the row space of A as well as for that of B.

22. **The Invertible Matrix Theorem**. The following statements are equivalent for a particular square $n \times n$ matrix A (be careful: these statements are not equivalent for rectangular matrices). That is, if one is true, then all are true, and if one is false, then all are false:

    (a) A is an invertible matrix.

    (b) A is row equivalent to the $n \times n$ identity matrix I_n.

    (c) A has n pivot positions.

    (d) The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

    (e) The columns of A form a linearly independent set.

    (f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

    (g) The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$. That is, the mapping $\mathbf{x} \mapsto A\mathbf{x}$ is onto $\mathbb{R}^n$.

    (h) The columns of A span $\mathbb{R}^n$.

    (i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

    (j) There is an $n \times n$ matrix C such that $CA = I = AC$.

    (k) The transpose A^T is an invertible matrix.

    (l) The columns of A form a basis of $\mathbb{R}^n$.

    (m) Col $A = \mathbb{R}^n$

(n) dim Col A = rank $A = n$

(o) Nul $A = \mathbf{0}$

(p) dim Nul $A = 0$

(q) det $A \neq 0$ (The definition for determinants will be given in chapter 3.)

(r) The number 0 is not an eigenvalue of A (The definition for eigenvalues will be given in chapter 5.)

2.3 Calculation Rules

2.3.1 Algebraic Definitions

If A, B and C are $m \times n$ matrices, then the addition and multiplication is defined as:

$$A + B = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} = \begin{bmatrix} (a_{11} + b_{11}) & \dots & (a_{1n} + b_{1n}) \\ \vdots & & \vdots \\ (a_{m1} + b_{m1}) & \dots & (a_{mn} + b_{mn}) \end{bmatrix} \tag{2.1}$$

$$rA = r \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ra_{11} & \dots & ra_{1n} \\ \vdots & & \vdots \\ ra_{m1} & \dots & ra_{mn} \end{bmatrix} \tag{2.2}$$

It is also possible to multiply matrices. If A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{b_1}$, $\mathbf{b_2}$, ..., $\mathbf{b_n}$, then the product AB is the $m \times p$ matrix:

$$AB = A[\,\mathbf{b_1} \;\; \mathbf{b_2} \;\; \dots \;\; \mathbf{b_p}] = [\,A\mathbf{b_1} \;\; A\mathbf{b_2} \;\; \dots \;\; A\mathbf{b_p}] \tag{2.3}$$

Note that $AB \neq BA$. Also, their power is:

$$A^k = A \dots A \quad (k \text{ times}) \tag{2.4}$$

The transpose of a matrix is defined as:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \qquad \Rightarrow \qquad A^T = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{nm} \end{bmatrix} \tag{2.5}$$

2.3.2 Algebraic Rules

The following rules apply for matrix addition.

$$A + B = B + A \tag{2.6}$$

$$(A + B) + C = A + (B + C) \tag{2.7}$$

$$A + 0 = A \tag{2.8}$$

$$r(A + B) = rA + rB \tag{2.9}$$

$$(r+s)A = rA + sA \tag{2.10}$$

$$r(sA) = (rs)A \tag{2.11}$$

For matrix multiplication, the following rules apply.

$$A(BC) = (AB)C \tag{2.12}$$

$$A(B+C) = AB + AC \tag{2.13}$$

$$(B+C)A = BA + CA \tag{2.14}$$

$$r(AB) = (rA)B = A(rB) \tag{2.15}$$

$$I_m A = A = AI_n \tag{2.16}$$

$$A^0 = I_n \tag{2.17}$$

$$I\mathbf{u} = \mathbf{u} \tag{2.18}$$

The following rules apply for matrix transposes.

$$(A^T)^T = A \tag{2.19}$$

$$(A+B)^T = A^T + B^T \tag{2.20}$$

$$(rA)^T = r(A^T) \tag{2.21}$$

$$(AB)^T = B^T A^T \tag{2.22}$$

3. Determinants

3.1 Definitions and Terms

3.1.1 Determinants

For any square matrix, let A_{ij} denote the submatrix formed by deleting the ith row and the jth column of A. For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is the sum of n terms of the form $\pm a_{1j}\text{det } A_{1j}$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \ldots, a_{1n}$ are from the first row of A. In symbols:

$$\text{det } A = a_{11}\text{det } A_{11} - a_{12}\text{det } A_{12} + \ldots + (-1)^{n+1}a_{1n}\text{det } A_{1n} = \sum_{j=1}^{n}(-1)^{1+j}a_{1j}\text{det } A_{1j}$$

Next to writing det A to indicate a determinant, it is also often used to write $|A|$.

3.1.2 Cofactors

Given $A = [a_{ij}]$, the **(i, j)-cofactor** of A is the number C_{ij} given by $C_{ij} = (-1)^{i+j}\text{det } A_{ij}$. The determinant of A can be determined using a cofactor expansion. The formula $\text{det } A = a_{11}C_{11} + a_{12}C_{12} + \ldots a_{1n}C_{1n}$ is called a **cofactor expansion across the first row** of A.

A matrix $B = [b_{ij}]$ of cofactors of A, where $b_{ij} = C_{ij}$, is called the **adjugate** (or **classical adjoint**) of A. This is denoted by $\text{adj}A$.

3.2 Theorems

1. The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion. The expansion across the ith row is: $\text{det } A = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$. The expansion down the jth column is: $\text{det } A = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$.
2. If A is a triangular matrix, then det A is the product of the entries on the main diagonal of A.
3. **Row Operations:** Let A be a square matrix.
   - If a multiple of one row of A is added to another row to produce a matrix B, then $\text{det } A = \text{det } B$.
   - If two rows of A are interchanged to produce B, then $\text{detA} = -\text{detB}$.
   - If one row of A is multiplied by k to produce B, then $\text{det } B = k \cdot \text{det } A$.
4. A square matrix A is invertible if, and only if $\text{det } A \neq 0$.
5. If A is an $n \times n$ matrix, then $\text{det } A^T = \text{det } A$.
6. If A and B are $n \times n$ matrices, then $\text{det } AB = (\text{det } A)(\text{det } B)$.
7. **Cramer's Rule:** Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by $x_i = \frac{\text{det } A_i(\mathbf{b})}{\text{det } A}$, where $i = 1, 2, \ldots, n$ and $A_i\mathbf{b}$ is the matrix obtained from A by replacing column i for the vector $\mathbf{b}$.

8. Let A be an invertible $n \times n$ matrix. Then: $A^{-1} = \frac{1}{\det\ A}\text{adj}A$.

9. If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det\ A|$. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det\ A|$.

10. Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is any region in $\mathbb{R}^2$ with finite area, then $\{\text{area of } T(S)\} = |\det\ \text{A}| \cdot \{\text{area of } S\}$. Also, if T is determined by a 3×3 matrix A, and if S is any region in $\mathbb{R}^3$ with finite volume, then $\{\text{volume of } T(S)\} = |\det\ \text{A}| \cdot \{\text{volume of } S\}$.

4. Vector Spaces and Subspaces

4.1 Definitions and Terms

4.1.1 Vector Spaces

A **vector space** is a nonempty set V of objects, called **vectors**, on which are defined two operations, called addition and multiplication by scalars, subject to the ten axioms listed in paragraph 4.3. As was already mentioned in the chapter Matrix Algebra, a **subspace** of a vector space V is a subset H of V that has three properties:

1. The zero vector of V is in H.
2. H is closed under vector addition. That is, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.
3. H is closed under multiplication by scalars. That is, for each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

If $\mathbf{v_1}, \ldots, \mathbf{v_p}$ are in a vector space V, then $\text{Span}\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is called **the subspace spanned** by $\mathbf{v_1}, \ldots, \mathbf{v_p}$. Given any subspace H of V, a **spanning set** for H is a set $\mathbf{v_1}, \ldots, \mathbf{v_p}$ in H such that $H = \text{Span}\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$.

4.1.2 Bases

Let H be a subspace of a vector space V. An indexed set of vectors $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$ in V is a **basis** for H if β is a linearly independent set, and the subspace spanned by β coincides with H, that is, $H = \text{Span}\{\mathbf{b_1}, \ldots, \mathbf{b_p}\}$. The set $\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ is a **standard basis** for $\mathbb{R}^n$. The set $\{1, t, \ldots, t^n\}$ is a **standard basis** for $\mathbb{P}^n$.

4.1.3 Coordinate Systems

Suppose $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ is a basis for V and $\mathbf{x}$ is in V. The **coordinates of x relative to the basis** β (or the **β-coordinates of x**) are the weights $c_1, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b_1} + \ldots + c_n\mathbf{b_n}$. If $c_1, \ldots, c_n$ are the β-coordinates of $\mathbf{x}$, then the vector $[\mathbf{x}]_\beta$ in $\mathbb{R}^n$ (consisting of $c_1, \ldots, c_n$) is the **coordinate vector of x (relative to β)**, or the **β-coordinate vector of x**. The mapping $\mathbf{x} \mapsto [\mathbf{x}]_\beta$ is the **coordinate mapping (determined by β)**.

If $P_\beta = [\,\mathbf{b_1}\ \ldots\ \mathbf{b_n}\,]$, then the vector equation $\mathbf{x} = c_1\mathbf{b_1} + \ldots + c_n\mathbf{b_n}$ is equivalent to $\mathbf{x} = P_\beta[\mathbf{x}]_\beta$. We call P_β the **change-of-coordinates matrix** from β to the standard basis $\mathbb{R}^n$. Since P_β is invertible (invertible matrix theorem), also $[\mathbf{x}]_\beta = P_\beta^{-1}\mathbf{x}$.

4.1.4 Vector Space Dimensions

If V is spanned by a finite set, then V is said to be **finite-dimensional**, and the **dimension** of V, written as dim V, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

4.2 Theorems

1. If $\mathbf{v_1}, \ldots, \mathbf{v_p}$ are in a vector space V, then $\text{Span}\{\mathbf{v_1}, \ldots, \mathbf{v_p}\}$ is a subspace of V.
2. Let $S = \{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ be a set in V, and let $H = \text{Span}\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$. If one of the vectors in S, say, $\mathbf{v_k}$, is a linear combination of the remaining vectors in S, then the set formed from S by removing $\mathbf{v_k}$ still spans H.
3. Let $S = \{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$ be a set in V, and let $H = \text{Span}\{\mathbf{e_1}, \ldots, \mathbf{e_n}\}$. If $H \neq \{\mathbf{0}\}$, some subset of S is a basis for H.
4. Let $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ be a basis for a vector space V. Ten for each $\mathbf{x}$ in V, there exists a unique set of scalars $c_1, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b_1} + \ldots + c_n\mathbf{b_n}$.
5. Let $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ be a basis for a vector space V, and let $P_\beta = [\,\mathbf{b_1} \;\ldots\; \mathbf{b_n}\,]$. Then the coordinate mapping $\mathbf{x} \mapsto [\mathbf{x}]_\beta$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.
6. If a vector space V has a basis $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$, then any set in V containing more than n vectors must be linearly dependent.
7. If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.
8. Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and $\dim H \leq \dim V$.
9. **The Basis Theorem:** Let V be a p-dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

4.3 Vector Space Axioms

The following axioms must hold for all the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in the vector space V and all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} + \mathbf{v}$, is in V.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
4. There is a **zero** vector $\mathbf{0}$ in V such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = 0$.
6. The scalar multiple of $\mathbf{u}$ by c, denoted by $c\mathbf{u}$, is in V.
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$.
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$.
10. $1\mathbf{u} = \mathbf{u}$.

5. Eigenvalues and Eigenvectors

5.1 Definitions and Terms

5.1.1 Introduction to Eigenvectors and Eigenvalues

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$. Such an $\mathbf{x}$ is called an **eigenvector corresponding to** λ. The set of all eigenvectors corresponding to λ is a subspace of $\mathbb{R}^n$ and is called the **eigenspace of** A **corresponding to** λ.

5.1.2 The Characteristic Equation

The scalar equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A. If A is an $n \times n$ matrix, then $\det(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A. A specific eigenvalue λ_s is said to have **multiplicity** r if $(\lambda - \lambda_s)$ occurs r times as a factor of the characteristic polynomial.

If A and B are $n \times n$ matrices, then A and B are **similar** if there is an invertible matrix P such that $P^{-1}AP = B$. Changing A into $P^{-1}AP$ is called a **similarity transformation**.

5.1.3 Diagonalization

A square matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D. The basis formed by all the eigenvectors of a matrix A is called the **eigenvector basis**.

5.1.4 Eigenvectors and Linear Transformations

Let V be an n-dimensional vector space, W an m-dimensional vector space, T any linear transformation from V to W, β a basis for V and γ a basis for W. Now the image of any vector $[\mathbf{x}]_\beta$ (the vector $\mathbf{x}$ relative to the base β) to $[T(\mathbf{x})]_\gamma$ is given by $[T(\mathbf{x})]_\gamma = M[\mathbf{x}]_\beta$, where $M = [[T(\mathbf{b_1})]_\gamma \; [T(\mathbf{b_2})]_\gamma \; \dots \; [T(\mathbf{b_n})]_\gamma]$. The $m \times n$ matrix M is a matrix representation of T, called the **matrix for T relative to the bases** β **and** γ.

In the common case when W is the same as V, and the basis γ is the same as β, the matrix M is called the **matrix for** T **relative to** β or simply the β**-matrix for** T and is denoted by $[T]_\beta$. Now $[T(\mathbf{x})]_\beta = [T]_\beta[\mathbf{x}]_\beta$.

5.1.5 Complex Eigenvalues

A complex scalar λ satisfies $\det(A - \lambda I) = 0$ if, and only if there is a nonzero vector $\mathbf{x}$ in $\mathbb{C}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. We call λ a **(complex) eigenvalue** and $\mathbf{x}$ a **(complex) eigenvector** corresponding to λ.

5.1.6 Dynamical Systems

Many dynamical systems can be described or approximated by a series of vectors $\mathbf{x_k}$ where $\mathbf{x_{k+1}} = A\mathbf{x_k}$. The variable k often indicates a certain time variable. If A is a diagonal matrix, having n eigenvalues forming a basis for $\mathbb{R}^n$, any vector $\mathbf{x_k}$ can be described by $\mathbf{x_k} = c_1(\lambda_1)^k\mathbf{v_1} + \ldots + c_n(\lambda_n)^k\mathbf{v_n}$. This is called the **eigenvector decomposition** of $\mathbf{x_k}$.

The graph $\mathbf{x_0}, \mathbf{x_1}, \ldots$ is called a **trajectory** of the dynamical system. If, for every $\mathbf{x_0}$, the trajectory goes to the origin $\mathbf{0}$ as k increases, the origin is called an **attractor** (or sometimes **sink**). If, for every $\mathbf{x_0}$, the trajectory goes away from the origin, it is called a **repellor** (or sometimes **source**). If $\mathbf{0}$ attracts for certain $\mathbf{x_0}$ and repels for other $\mathbf{x_0}$, then it is called a **saddle point**. For matrices having complex eigenvalues/eigenvectors, it often occurs that the trajectory spirals inward to the origin (attractor) or outward (repellor) from the origin (the origin is then called a **spiral point**).

5.1.7 Differential Equations

Linear algebra comes in handy when differential equations take the form $\mathbf{x}' = A\mathbf{x}$. The **solution** is then a vector-valued function that satisfies $\mathbf{x}' = A\mathbf{x}$ for all t in some interval. There is always a **fundamental set of solutions**, being a basis for the set of all solutions. If a vector $\mathbf{x_0}$ is specified, then the **initial value problem** is to construct the function such that $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{x}(0) = \mathbf{x_0}$.

5.2 Theorems

1. The solution set of $A\mathbf{x} = \lambda\mathbf{x}$ is the null space of $A - \lambda I$. This is the eigenspace corresponding to λ.
2. The eigenvalues of a triangular matrix are the entries on its main diagonal.
3. 0 is an eigenvalue of A if, and only if A is not invertible.
4. If $\mathbf{v_1}, \ldots, \mathbf{v_r}$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v_1}, \ldots, \mathbf{v_r}\}$ is linearly independent.
5. A scalar λ is an eigenvalue of an $n \times n$ matrix A if, and only if λ satisfies the characteristic equation $\det(A - \lambda I) = 0$.
6. If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).
7. **The Diagonalization Theorem:** An $n \times n$ matrix A is diagonalizable if, and only if A has n linearly independent eigenvectors. In fact, $A = PDP^{-1}$, with D a diagonal matrix, if, and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.
8. If a $n \times n$ matrix has n distinct eigenvalues, then it is diagonalizable. (Note that the opposite is not always true.)
9. Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_p$.
   (a) For $1 \leq k \leq p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.
   (b) The matrix A is diagonalizable if, and only if the sum of the dimensions of the distinct eigenspaces equals n, and this happens if, and only if the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

(c) If A is diagonalizable and β_k is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\beta_1, \ldots, \beta_p$ forms an eigenvector basis for $\mathbb{R}^n$.

10. **Diagonal Matrix Representation:** Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\beta = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the β-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

11. If P is the matrix whose columns come from the vectors in β (that is, $P = [\mathbf{b_1} \ \ldots \ \mathbf{b_n}]$), then $[T]_\beta = P^{-1}AP$.

12. If λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector in $\mathbb{C}^n$, then the complex conjugate $\bar{\lambda}$ is also an eigenvalue of A, with $\bar{\mathbf{x}}$ as the corresponding eigenvector.

13. Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ ($b \neq 0$) and an associated eigenvector $\mathbf{v}$ in $\mathbb{C}^2$. Then $A = PCP^{-1}$, where $P = [\mathrm{Re}\,\mathbf{v} \ \ \mathrm{Im}\,\mathbf{v}]$ and $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$.

14. For discrete dynamical systems, multiple possibilities are present:

    (a) $|\lambda_i| < 1$ for every i. In that case the system is an attractor.

    (b) $|\lambda_i| > 1$ for every i. In that case the system is a repellor.

    (c) $|\lambda_i| > 1$ for some i and $|\lambda_i| < 1$ for all other i. In that case the system is a saddle point.

    (d) $|\lambda_i| = 1$ for some i. In that case the trajectory can converge to any vector in the eigenspace corresponding to the eigenvalue 1. However, it can also diverge.

15. For linear differential equations, each eigenvalue-eigenvector pair provides a solution of $\mathbf{x}' = A\mathbf{x}$. This solution is $\mathbf{x}(t) = \mathbf{v}e^{\lambda t}$.

16. For linear differential equations, any linear combination of solutions is also a solution for the differential equation. So if $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are solutions, then $c\mathbf{u}(t) + d\mathbf{v}(t)$ is also a solution for any scalars c and d.

6. Orthogonality and Least Squares

6.1 Definitions and Terms

6.1.1 Basics of Vectors

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ can be multiplied with each other, using the **dot product**, also called the **inner product**, which produces a scalar value. It is denoted as $\mathbf{u} \cdot \mathbf{v}$, and defined as $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \ldots + u_n v_n$. The **length** of a vector $\mathbf{u}$, sometimes also called the **norm**, is denoted by $\|\mathbf{u}\|$. It is defined as $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + \ldots + u_n^2}$.

A **unit vector** is a vector whose length is 1. For any nonzero vector $\mathbf{u}$, the vector $\frac{\mathbf{u}}{\|\mathbf{u}\|}$ is a unit vector in the direction of $\mathbf{u}$. This process of creating unit vectors is called **normalizing**. The **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{v} - \mathbf{u}$. That is, $\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{v} - \mathbf{u}\|$.

6.1.2 Orthogonal Sets

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$. If $\mathbf{z}$ is orthogonal to every vector in a subspace W, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the **orthogonal complement** of W, and is denoted by $W^{\perp}$.

A set of vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct pair of vectors is orthogonal, that is, if $\mathbf{u_i} \cdot \mathbf{u_j} = 0$ whenever $i \neq j$. An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

6.1.3 Orthonormal Sets

A set of vectors $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ in $\mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u_1}, \ldots, \mathbf{u_n}\}$ is an **orthonormal basis** for W. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. Such a matrix always has orthonormal columns.

6.1.4 Decomposing Vectors

If $\mathbf{u}$ is any nonzero vector in $\mathbb{R}^n$, then it is possible to decompose any vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one being a multiple of $\mathbf{u}$, and one being orthogonal to it. The projection $\hat{\mathbf{y}}$ (being the multiple of $\mathbf{u}$) is called the **orthogonal projection of y onto u**, and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is, surprisingly, called the **component of y orthogonal to u**.

Just like it is possible to project vectors on a vector, it is also possible to project vectors on a subspace. The projection $\hat{\mathbf{y}}$ onto the subspace W is called the **orthogonal projection of y onto** W. $\hat{\mathbf{y}}$ is sometimes also called **the best approximation to y by elements of** W.

6.1.5 The Gram-Schmidt Process

The **Gram-Schmidt Process** is an algorithm for producing an orthogonal or orthonormal basis $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ for any nonzero subspace of $\mathbb{R}^n$. Let W be the subspace, having basis $\{\mathbf{x_1}, \ldots, \mathbf{x_p}\}$. Let $\mathbf{u_1} = \mathbf{x_1}$ and $\mathbf{u_i} = \mathbf{x_i} - \mathbf{\hat{x}_i}$ for $1 < i \leq n$, where $\mathbf{\hat{x}_i}$ is the projection of $\mathbf{x_i}$ on the subspace with basis $\{\mathbf{u_1}, \ldots, \mathbf{u_{i-1}}\}$. In formula: $\mathbf{u_1} = \mathbf{x_1}$ and $\mathbf{u_i} = \mathbf{x_i} - (\frac{x_i \cdot v_1}{v_1 \cdot v_1} v_1 + \ldots + \frac{x_i \cdot v_{i-1}}{v_{i-1} \cdot v_{i-1}} v_{i-1})$.

6.1.6 Least-Squares Problem

The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible. If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\mathbf{\hat{x}}$ in $\mathbb{R}^n$ such that $\|\mathbf{b} - A\mathbf{\hat{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$ for all $\mathbf{x}$ in $\mathbb{R}^n$. When a least-squares solution $\mathbf{\hat{x}}$ is used to produce $A\mathbf{\hat{x}}$ as an approximation of $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\mathbf{\hat{x}}$ is called the **least-squares error** of this approximation.

6.1.7 Linear Models

In statistical analysis of scientific and engineering data, there is commonly a different notation used. Instead of $A\mathbf{x} = \mathbf{b}$, we write $X\beta = \mathbf{y}$ and refer to X as the **design matrix**, β as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

Suppose we have a certain amount of measurement data which, when plotted, seem to lie close to a straight line. Let $y = \beta_0 + \beta_1 x$. The difference between the observed value (from the measurements) and the predicted value (from the line) is called a **residual**. The **least-squares line** is the line that minimizes the sum of the squares of the residuals. This line is also called a **line of regression of** y **on** x. The coefficients β_0 and β_1 are called (linear) **regression coefficients**.

The previous system is equivalent to solving the least-squares solution of $X\beta = \mathbf{y}$ if $X = [\ \mathbf{1}\ \ \mathbf{x}\ ]$ (where $\mathbf{1}$ has entries 1, 1, …, 1, and $\mathbf{x}$ has entries $x_1, \ldots, x_n$), β has entries β_0 and β_1 and $\mathbf{y}$ has entries $y_1, \ldots, y_n$. A common practice before computing a least-squares line is to compute the average $\bar{x}$ of the original x-values, and form a new variable $x^* = x - \bar{x}$. The new x-data are said to be in **mean-deviation form**. In this case, the two columns of X will be orthogonal.

The **residual vector** ϵ is defined as $\epsilon = \mathbf{y} - X\beta$. So $\mathbf{y} = X\beta + \epsilon$. Any equation in this form is referred to as a **linear model**, in which ϵ should be minimized.

6.1.8 Inner Product Spaces

An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in U, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in V and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if, and only if $\mathbf{u} = \mathbf{0}$

A vector space with an inner product is called an **inner product space**.

6.2 Theorems

1. Consider the vectors $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices. Then, $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$.

2. If $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$ then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if, and only if $\mathbf{u} \cdot \mathbf{v} = 0$.
3. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if, and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.
4. A vector $\mathbf{z}$ is in $W^{\perp}$ if, and only if $\mathbf{z}$ is orthogonal to every vector in a set that spans W.
5. $W^{\perp}$ is a subspace of $\mathbb{R}^n$.
6. If A is an $m \times n$ matrix, then $(\text{ Row}A)^{\perp} = \text{ Nul}A$ and $(\text{ Col}A)^{\perp} = \text{ Nul}A^T$.
7. If A is an $m \times n$ matrix, then $\text{Row}A = \text{ Col}A^T$.
8. If $S = \{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.
9. Let $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination $\mathbf{y} = c_1\mathbf{u_1} + \ldots + c_p\mathbf{u_p}$ are given by $c_j = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\mathbf{u_j}\cdot\mathbf{u_j}} = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\|\mathbf{u_j}\|^2}$.
10. An $m \times n$ matrix U has orthonormal columns if, and only if $U^TU = I$.
11. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$, then:
    (a) $\|U\mathbf{x}\| = \|\mathbf{x}\|$
    (b) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
    (c) $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if, and only if $\mathbf{x} \cdot \mathbf{y} = 0$.
12. If U is a square matrix, then U is an orthogonal matrix if, and only if its columns are orthonormal columns. The rows of an orthogonal matrix are also orthonormal rows.
13. If $\mathbf{y}$ and $\mathbf{u}$ are any nonzero vectors in $\mathbb{R}^n$, then the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is $\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\mathbf{u_j}\cdot\mathbf{u_j}}\mathbf{u} = \frac{\mathbf{y}\cdot\mathbf{u_j}}{\|\mathbf{u_j}\|^2}\mathbf{u}$, and the component $\mathbf{z}$ of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.
14. Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is any orthogonal basis of W, then $\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u_1}}{\mathbf{u_1}\cdot\mathbf{u_1}}\mathbf{u_1} + \ldots + \frac{\mathbf{y}\cdot\mathbf{u_p}}{\mathbf{u_p}\cdot\mathbf{u_p}}\mathbf{u_p}$, and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.
15. **The Best Approximation Theorem**. Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y}$ any vector in $\mathbb{R}^n$, and $\hat{\mathbf{y}}$ the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that $\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{u}\|$ for all $\mathbf{u} \neq \hat{\mathbf{y}}$ in W.
16. If $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then $\hat{\mathbf{y}} = (\mathbf{y}\cdot\mathbf{u_1})\mathbf{u_1} + \ldots + (\mathbf{y}\cdot\mathbf{u_p})\mathbf{u_p}$. If $U = [\ \mathbf{u_1}\ \ \ldots\ \ \mathbf{u_p}\ ]$, then $\hat{\mathbf{y}} = UU^T\mathbf{y}$ for all $\mathbf{y}$ in $\mathbb{R}^n$.
17. The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.
18. The matrix A^TA is invertible if, and only if the columns of A are linearly independent. In that case, the equation $A\mathbf{x} = \mathbf{b}$ has only one least-squares solution $\hat{\mathbf{x}}$, and it is given by $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$.

6.3 Calculation Rules

6.3.1 Algebraic Definitions

The dot product of vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is defined as:

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \ldots u_n v_n \tag{6.1}$$

The length of a vector is defined as:

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + \ldots + u_n^2} \tag{6.2}$$

6.3.2 Algebraic Rules

The following rules apply for the dot product:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \tag{6.3}$$

$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \tag{6.4}$$

$$(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v}) \tag{6.5}$$

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} \tag{6.6}$$

The following rules apply for vector lengths:

$$\|c\mathbf{u}\| = |c| \|\mathbf{u}\| \tag{6.7}$$

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta \tag{6.8}$$

Where θ is the angle between vectors $\mathbf{v}$ and $\mathbf{u}$.

7. Symmetric Matrices and Quadratic Forms

7.1 Definitions and Terms

7.1.1 Diagonalization of Symmetric Matrices

A **symmetric matrix** is a square matrix such that $A^T = A$. A matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P (so $P^{-1} = P^T$) and a diagonal matrix D such that $A = PDP^T = PDP^{-1}$. An orthogonally diagonalizable matrix A with orthonormal eigenvectors $\mathbf{u_1} \dots \mathbf{u_n}$ can be written as $A = \lambda_1 \mathbf{u_1}\mathbf{u_1}^T + \dots + \lambda_n \mathbf{u_n}\mathbf{u_n}^T$. This representation of A is called a **spectral decomposition** of A. Furthermore, each matrix $\mathbf{u_j}\mathbf{u_j}^T$ is a **projection matrix**.

7.1.2 Quadratic Forms

A **quadratic form** on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ whose value at a vector $\mathbf{x}$ in $\mathbb{R}^n$ can be computed by an expression of the form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is an $n \times n$ symmetric matrix. The matrix A is called the **matrix of the quadratic form**.

If $\mathbf{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form $\mathbf{x} = P\mathbf{y}$, where P is an invertible matrix and $\mathbf{y}$ is a new variable vector in $\mathbb{R}^n$. Now $\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T(P^T AP)\mathbf{y}$. If P diagonalizes A, then $P^T AP = P^{-1}AP = D$, in which D is a diagonal matrix. The columns op P are called the **principal axes** of the quadratic form $\mathbf{x}^T A\mathbf{x}$.

A quadratic form Q is per definition:

- **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.

The classification of a quadratic form is often carried over to the matrix of the form. Thus a **positive definite matrix** A is a symmetric matrix for which the quadratic form $\mathbf{x}^T A\mathbf{x}$ is positive definite.

7.1.3 Geometric View of Principal Axes

When $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where A is an invertible 2×2 symmetric matrix, and c is a constant, then the set of all $\mathbf{x}$ such that $\mathbf{x}^T A\mathbf{x} = c$ corresponds to an ellipse (or circle) or a hyperbola. An ellipse is described by the following equation in standard form: $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$ $(a > b > 0)$, where a is the semi-mayor axes and b is the semi-minor axes. A hyperbola is described by the following equation in standard form: $\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$ $(a > b > 0)$, where the asymptotes are given by the equations $x_2 = \pm\frac{b}{a}x_1$.

7.2 Theorems

1. If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.
2. An $n \times n$ matrix A is orthogonally diagonalizable if, and only if A is a symmetric matrix.

3. **The Spectral Theorem for Symmetric Matrices:** An $n \times n$ symmetric matrix A has the following properties:
   (a) A has n real eigenvalues, counting multiplicities.
   (b) The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.
   (c) The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.
   (d) A is orthogonally diagonalizable.

4. **The Principal Axes Theorem:** Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross-product term.

5. **Quadratic Forms and Eigenvalues:** Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^T A\mathbf{x}$ is:
   (a) positive definite if, and only if the eigenvalues of A are all positive.
   (b) negative definite if, and only if the eigenvalues of A are all negative.
   (c) positive semidefinite if, and only if one eigenvalue of A is 0, and the others are positive.
   (d) negative semidefinite if, and only if one eigenvalue of A is 0, and the others are negative.
   (e) indefinite if, and only if A has both positive and negative eigenvalues.