WI1403-LR Linear Algebra

Delft University of Technology

Year 2013 – 2014

Michele Facchinelli

Version 1.0 Last modified on *February 1, 2017*

Preface

This summary was written for the course WI1403-LR Linear Algebra, taught at the Delft University of Technology. All the material treated is taken from [D. Lay. *Linear Algebra and Its Applications*. Pearson, 4th edition, 2014.]

Throughout the summary, references to chapters and sections can be found. These are labelled with the aid of the symbol § and can be found in the aforementioned book, where exercises and more explanations are given.

In case of any comments about the content of the summary, please do not hesitate to contact me at `m.facchinelli@yahoo.it`.

"Matrices are Roman Catholic —
Rows come before Columns!"

– Linear Algebra professor

Changelog

This is version 1.0. Below are listed the changes applied to each version.

Version	Date	Changes
1.0	February 1, 2017	First version

Contents

1 | Linear Equations in Linear Algebra [§1]

1.1 | Equivalence of Notations

Theorem 1.1. NOTATIONS.
The *matrix equation*

$$A\mathbf{x} = \mathbf{b}$$

the *vector equation*

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

and the *linear system*

$$\begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n & | \, \mathbf{b} \end{bmatrix}$$

all share the same solution set.

1.2 | Homogeneous Linear Systems

A homogeneous linear system is one of the type $A\mathbf{x} = \mathbf{0}$, where A is a $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$.

The homogeneous equation

$$A\mathbf{x} = \mathbf{0}$$

has a non trivial solution (i.e. not $\mathbf{x} = \mathbf{0}$) if and only if the equation has at least one free variable.

1.3 | Nonhomogeneous Linear Systems

A nonhomogeneous linear system, as seen before, is of the form

$$A\mathbf{x} = \mathbf{b} \tag{1.1}$$

To describe the solution set of such a system, consider the solution of the same system, but in the case of $\mathbf{b} = \mathbf{0}$. This specific system will give a solution of the type

$$\mathbf{x} = t\mathbf{v} \tag{1.2}$$

where t is the free variable. To get the solution of the system $A\mathbf{x} = \mathbf{b}$, now one simply has to add a vector, $\mathbf{p}$ for instance, to equation (1.2). Hence, the solution for equation (1.1) is

$$\mathbf{x} = t\mathbf{v} + \mathbf{p} \tag{1.3}$$

If equation (1.2) can be seen as a line passing through the origin and the vector v, then equation (1.3) becomes the eqaution of the line through $\mathbf{p}$ parallel to $\mathbf{v}$. *Thus the solution set of $A\mathbf{x} = \mathbf{b}$ is a line through $\mathbf{p}$ parallel to the solution set of $A\mathbf{x} = \mathbf{0}$.*

1.4 | Linear Independence

A set of vectors is $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution.

The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is said to be **linearly dependent** if there exist weights (or coefficients) $c_1, \ldots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

Theorem 1.2. CHARACTERIZATION OF LINEARLY DEPENDENT SETS.
A set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq \mathbf{0}$, then some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.

Theorem 1.3. If a set containes more vectors that there are entries in each vector (or a matrix with more columns than rows), then the set is linearly dependent. That is, any set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.

1.5 | Introduction to Linear Transformations

A matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ can arise in linear algebra in a way that is not directely connected with linear combinations of vectors. This happens when we think of the matrix A as an object that "acts"as a vector $\boldsymbol{x}$ by multiplication to produce a new vector called $A\boldsymbol{x}$.

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

$$T : \mathbb{R}^n \to \mathbb{R}^m$$

indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For $\boldsymbol{x}$ in $\mathbb{R}^n$, the vector $T(\boldsymbol{x})$ in $\mathbb{R}^m$ is called the **image** of $\boldsymbol{x}$ (under the action of T). The set of all images $T(\boldsymbol{x})$ is called the **range** of T.

For each $\boldsymbol{x}$ in $\mathbb{R}^n$, $T(\boldsymbol{x})$ is computed as $A\boldsymbol{x}$, where A is a $m \times n$ matrix. A *matrix transformation* is usually denoted by $\boldsymbol{x} \mapsto A\boldsymbol{x}$.

A transformation (or mapping) of T is **linear** if:

(i) $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v}$ in the domain of T;

(ii) $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$ for all scalars c and all $\boldsymbol{u}$ in the domain of T.

If T is a linear transformation, then

$$T(\boldsymbol{0}) = \boldsymbol{0}$$

and

$$T(c\boldsymbol{u} + d\boldsymbol{v}) = cT(\boldsymbol{u}) + dT(\boldsymbol{v})$$

for all vectors $\boldsymbol{u}, \boldsymbol{v}$ in the domain of T and all scalars c, d.

1.6 | The Matrix of a Linear Transformation

Every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$. The key to finding A is to observe that T is completely determined by what it does to the columns of the $n \times n$ identity matrix I_n.

The columns of

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

are $\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. If for a linear transformation $T(\boldsymbol{e}_1)$ and $T(\boldsymbol{e}_2)$ are given, $\boldsymbol{x}$ can be rewritten as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 \tag{1.4}$$

Since T is a linear transformation

$$T(\boldsymbol{x}) = x_1T(\boldsymbol{e}_1) + x_2T(\boldsymbol{e}_2) \tag{1.5}$$

The step from equation (1.4) to equation (1.5) explains why the knowledge of $T(\boldsymbol{e}_1)$ and $T(\boldsymbol{e}_2)$ is sufficient to determine $T(\boldsymbol{x})$ for any $\boldsymbol{x}$.

Theorem 1.4. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that

$$T(\boldsymbol{x}) = A\boldsymbol{x} \quad \text{for all } \boldsymbol{x} \text{ in } \mathbb{R}^n$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(\boldsymbol{e}_j)$, where $\boldsymbol{e}_j$ is the jth column of the of the identity matrix in $\mathbb{R}^n$:

$$A = \begin{bmatrix} T(\boldsymbol{e}_1) & \cdots & T(\boldsymbol{e}_n) \end{bmatrix} \tag{1.6}$$

This matrix A in equation (1.6) is called **the standard matrix for the linear transformation** T.

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\boldsymbol{b}$ in $\mathbb{R}^m$ is the image of *at least one* $\boldsymbol{x}$ in $\mathbb{R}^n$.

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** $\mathbb{R}^m$ if each $\boldsymbol{b}$ in $\mathbb{R}^m$ is the image of *at most one* $\boldsymbol{x}$ in $\mathbb{R}^n$.

Theorem 1.5. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T in one-to-one if and only if the equation $T(\boldsymbol{x}) = \mathbf{0}$ has only the trivial solution.

Theorem 1.6. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear trasformation and let A be the standard matrix for T. Then:

(i) T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$;

(ii) T is one-to-one if and only if the columns of A are linearly independent.

2 | Matrix Algebra [§2]

2.1 | Matrix Operations

If A is an $m \times n$ matriz, then the scalar entry in the ith row and jth column of A is denoted by a_{ij} and is called the (i, j)-entry of A.

The **diagonal entries** in an $m \times n$ matrix $A = [a_{ij}]$ are $a_{11}, \dots, a_{ii}$ and they form the **main diagonal** of A. A diagonal matrix is a square $n \times n$ matrix whose nondiagonal entries are zero. Two examples are the $n \times n$ identity matrix, I_n, and the $n \times n$ zero matrix, 0_n.

Two matrices are **equal** if they have the same size and if their corresponding columns are equal.

The **sum** $A + B$ is the $m \times n$ matrix whose columns are the sums of the corresponding columns in A and B.

If A is an $m \times n$ matrix, and B is an $n \times p$ matrix with columns $\boldsymbol{b}_1, \dots, \boldsymbol{b}_p$, then the **product** AB is the $m \times p$ matrix whose columns are $A\boldsymbol{b}_1, \dots, A\boldsymbol{b}_p$. That is,

$$AB = A \begin{bmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_p \end{bmatrix} = \begin{bmatrix} A\boldsymbol{b}_1 & \cdots & A\boldsymbol{b}_p \end{bmatrix} \tag{2.1}$$

From equation (2.1) it is clear that each column of AB is a linear combination of the columns of A using weights from the corresponding column of B.

If the product AB is defined, then the entry in row i and column j of AB is the sum of the products of the corresponding entries from row i of A and column j of B. If $(A)_{ij}$ denotes the (i, j)-entry in AB, and A is an $m \times n$ matrix, then

$$(AB)_{ij} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj}$$

Warnings:

(i) In general, $AB \neq BA$;

(ii) The cancellation laws do *not* hold for matrix multiplication: i.e. if $AB = AC$, then it is not true in general that $B = C$;

(iii) If a product AB is the zero matrix, you *cannot* conclude in general that either $A = 0$ or $B = 0$.

If A is an $n \times n$ matrix and if k is a positive integer, then A^k denotes the **power** k of the matrix A. That is A^k denoted the product of k copies of A.

Given an $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A. For instance, let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then A^T is

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

Note that the transpose of a product of matrices equals the product of their transposes in the *reverse* order. Hence

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

2.2 | The Inverse of a Matrix

An $n \times n$ matrix A is said to be *invertible* if there is an $n \times n$ matrix C such that

$$AC = I \text{ and } CA = I$$

where $I = I_n$, the $n \times n$ identity matrix. In this case, C is the **inverse** of A. In fact, C is uniquely determined by A. This unique inverse is denoted by A^{-1}, so that

$$AA^{-1} = I \text{ and } A^{-1}A = I$$

A matrix that is *not* invertible is sometimes called a **singular matrix**, and an invertible matrix is called a **nonsingular matrix**.

Theorem 2.1. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

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

If $ad - bc = 0$, then A is not invertible.

The quantity $ad - bc$ is called the **determinant** of A, and we write

$$\det A = ad - bc$$

If A is an invertible $n \times n$ matrix, then for each $\boldsymbol{b}$ in $\mathbb{R}^n$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solutions $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E ia created by performing the same row operation on I_m.

Theorem 2.2. 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 reduce A to I_n also transforms I_n into A^{-1}.

Applying theorem 2.2 to an invertible matrix A gives

$$A \sim E_1 A \sim \cdots \sim E_p(E_{p-1} \cdots E_1 A) = I_n$$

then, the product $E_p \cdots E_1$ consists of the inverse of A. Hence

$$A^{-1} = E_p \cdots E_1$$

If we place A and I side-by-side to form an augmented matrix $[\ A \mid I\ ]$, then row operations on this matrix produce identical operations on A and I. Then, if A is row equivalent to I, $[\ A \mid I\ ]$ is row equivalent to $[\ I \mid A^{-1}\ ]$. Otherwise A does not have an inverse.

2.3 | Characterisation of Invertible Matrices

Theorem. The Invertible Matrix Theorem.
Let A be a square matrix $n \times n$. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false (see appendix A).

(i) A is an invertible matrix;

(ii) A is row equivalent to the $n \times n$ identity matrix;

(iii) A has n pivot points;

(iv) The equation $A\boldsymbol{x} = \mathbf{0}$ has only the trivial solution;

(v) The columns of A form a linearly independent set;

(vi) The linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is one-to-one;

(vii) The equation $A\boldsymbol{x} = \boldsymbol{b}$ has at least one solution for each $\boldsymbol{b}$ in $\mathbb{R}^n$;

(viii) The columns of A span $\mathbb{R}^n$;

(ix) The linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$;

(x) There is an $n \times n$ matrix C such that $CA = I$;

(xi) There is an $n \times n$ matrix D such that $AD = I$;

(xii) A^T is an invertible 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

$$S(T(\boldsymbol{x})) = \boldsymbol{x} \quad \text{for all } \boldsymbol{x} \text{ in } \mathbb{R}^n$$
$$T(S(\boldsymbol{x})) = \boldsymbol{x} \quad \text{for all } \boldsymbol{x} \text{ in } \mathbb{R}^n$$

Hence, T is invertible if and only if its standard matrix A is invertible.

2.4 | Subspaces of $\mathbb{R}^n$

A **subspace** of $\mathbb{R}^n$ is any set $H \in \mathbb{R}^n$ that has three proprieties:

(i) $\mathbf{0} \in H$;

(ii) $\forall\ \boldsymbol{u}, \boldsymbol{v} \in H,\ (\boldsymbol{u} + \boldsymbol{v}) \in H$;

(iii) $\forall\ \boldsymbol{u} \in H$ and $\forall\ c \in \mathbb{R},\ (c\boldsymbol{u}) \in H$.

In words, a subspace is *closed* under addition and scalar multiplication. Examples of subspaces are a plane or a line both through the origin.

The **column space** of a matrix A is the set Col$A$ of all linear combinations of the columns of A.

If $A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{bmatrix}$, with the columns in $\mathbb{R}^m$, then Col$A$ is the same as Span$\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$. The column space of an $m \times n$ matrix is a subspace of $\mathbb{R}^m$. Note that Col$A$ equals $\mathbb{R}^m$ only when the columns of A span $\mathbb{R}^m$.

When a system of linear equations is written in the form $A\boldsymbol{x} = \boldsymbol{b}$, the column space of A is the set of all $\boldsymbol{b}$ for which the system has a solution.

The **null space** of a matrix A is the set Nul$A$ of all solutions of the homogeneous equation $A\boldsymbol{x} = \mathbf{0}$.

Theorem 2.3. The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions of a system $A\boldsymbol{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

Because a subspace typically contains an infinite number of vectors, some problems are handled best by working with a small finite set of vectors that span the subspace. *The smaller the set, the better.* The smallest possible spanning set must be linearly independent.

A **basis** for a subspace H of $\mathbb{R}^n$ is a linearly independent set in H that spans H.

The columns of an invertible $n \times n$ matrix form a basis for all of $\mathbb{R}^n$, because they are linearly independent and span $\mathbb{R}^n$. One such matrix is the $n \times n$ identity matrix. Its columns are denoted by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$:

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix}, \ \ldots\ , \ \boldsymbol{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$$

The set $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$.

The standard procedure for writing the solution set of $A\boldsymbol{x} = \mathbf{0}$ in parametric vector form, actually identifies a basis for Nul$A$.

Suppose you are given a 3×5 matrix A and you are asked to compute the null space of such matrix:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{15} \\ \vdots & \ddots & \vdots \\ a_{31} & \cdots & a_{35} \end{bmatrix} \quad \text{with solution} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix} = x_2\boldsymbol{u} + x_3\boldsymbol{v} + x_5\boldsymbol{w}$$

The general solution shows that Nul$A$ coincides with the set of all linear combinations of $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$. That is, $\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$ generates Nul$A$. So $\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$ is a *basis* for Nul$A$.

Theorem 2.4. The pivot columns of a matrix A form a basis for the column space of A.

Be careful to use *pivot columns* of A itself for the basis of Col$A$. The columns of an echelon form A_R are often not in the column space of A.

2.5 | Dimension and Rank

The main reason for selecting a basis for a subspace H, is that each vector in H can be written in only one way as a linear combination of the basis vectors of H. To see why, suppose $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$ is a basis for H, and suppose a vector $\boldsymbol{x}$ in H can be generated in two ways, say,

$$\boldsymbol{x} = c_1\boldsymbol{b}_1 + \cdots + c_p\boldsymbol{b}_p \quad \text{and} \quad \boldsymbol{x} = d_1\boldsymbol{b}_1 + \cdots + d_p\boldsymbol{b}_p$$

Then, subtracting gives

$$\mathbf{0} = (c_1 - d_1)\boldsymbol{b}_1 + \cdots + (c_p - d_p)\boldsymbol{b}_p$$

Since $\mathcal{B}$ is linearly independent, the weight in this last equation must all be zero. That is, $c_j = d_j$ for $1 \leq j \leq p$, which shows that the two representations are actually the same.

Suppose the set $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$ is a basis for a subspace H. For each $\boldsymbol{x}$ in H, the **coordinates of $\boldsymbol{x}$ relative to the basis $\mathcal{B}$** are the weights $c_1, \dots, c_p$ such that $\boldsymbol{x} = c_1\boldsymbol{b}_1 + \cdots + c_p\boldsymbol{b}_p$, and the vector in $\mathbb{R}^p$

$$[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

is called the **coordinate vector of $\boldsymbol{x}$ (relative to $\mathcal{B}$)** or the **$\mathcal{B}$-coordinate vector of $\boldsymbol{x}$**. Hence $\boldsymbol{x} = \mathcal{B}[\boldsymbol{x}]_{\mathcal{B}}$.

The mapping $\boldsymbol{x} \mapsto [\boldsymbol{x}]_{\mathcal{B}}$, where $\mathcal{B} = \{\boldsymbol{b}_1, \dots, \boldsymbol{b}_p\}$, is a one-to-one correspondence between H and $\mathbb{R}^p$ that preserves linear combinations. We call such a correspondence as *isomorphism*, and we say that H is *isomorphic* to $\mathbb{R}^p$.

The **dimention** of a nonzero subspace H, denoted by dim$H$, is the number of vectors in any basis for H. The dimention of the zero subspace $\{\mathbf{0}\}$ is defined to be zero.

The **rank** of a matrix A, denoted by rank$A$, is the dimention of the column space of A.

Since the pivot columns of A form a basis for Col$A$, the rank of A is just the number of pivot columns in A.

Theorem 2.5. The Rank Theorem.
Since the nonpivot columns correspond to the free variables in $A\boldsymbol{x} = \mathbf{0}$, if a matrix A has n columns, then rank$A$ + dimNul$A = n$.

Theorem 2.6. The Basis Theorem.
Let H be a p-dimentional (i.e. with dim$H = p$) 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.

If A is an $m \times n$ matrix, each row of A has n entries and thus can be identified with a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vector is called the **row space** of A and is denoted by Row$A$, which is also a subspace of $\mathbb{R}^n$. Since the rows of A are identified with the columns of A^T, we could also write ColA^T in place of Row$A$.

If two matrices A and B are row equivalent ($A \sim B$), 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.

Theorem. The Invertible Matrix Theorem (Continued).
Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement A is an invertible matrix (see appendix A).

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

(xiv) $\text{Col}A = \mathbb{R}^n$;

(xv) $\text{dimCol}A = n$;

(xvi) $\text{rank}A = n$;

(xvii) $\text{Nul}A = \{\mathbf{0}\}$;

(xviii) $\text{dimNul}A = 0$.

3 | Determinants [§3]

3.1 | Introduction to Determinants

If A is a 3×3 matrix, we can write it as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

then its determinant is given by

$$\Delta = \det A = |A| = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}$$

In general, for $n \geq 2$, the **determinant algorithm** for an $n \times n$ matrix $A = [a_{ij}]$ is

$$|A| = \det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j}$$

It is useful to define the **(i, j)-cofactor** of a determinant $A = [a_{ij}]$. The cofactor is

$$C_{ij} = (-1)^{i+j} \det A_{ij} \tag{3.1}$$

Then,

$$\det A = a_{11} C_{11} + \cdots + a_{1n} C_{1n}$$

This formula is called a *cofactor expansion across the first row* of A.

Theorem 3.1. In general, the determinant of an $n \times n$ matrix A can be computed by cofactor expansion across any row or down any column. The expansion across the ith row using cofactor is

$$|A| = \sum_{j=1}^{n} a_{ij} C_{ij} \text{ keeping } i \text{ fixed}$$

The cofactor expansion down the jth column is

$$|A| = \sum_{i=1}^{n} a_{ij} C_{ij} \text{ keeping } j \text{ fixed}$$

Theorem 3.2. If A is a triangular matrix, the $\det A$ is the product of the entries on the main diagonal of A.

3.2 | Proprieties of Determinants

Theorem 3.3. Row Operations.
Let A be a square matrix.

(i) If a multiple of one row of A is added to another row to produce a matrix B, then $\det A = \det B$;

(ii) If two rows of A are interchanged to produce B, then $\det B = -\det A$;

(iii) If one row of A is multiplied by k to produce B, then $\det B = k \cdot \det A$.

Suppose a square matrix has been reduced to an *echelon form* U by row replacements and row interchanges. If there are r interchanges, then, as seen before,

$$\det A = (-1)^r \det U$$

Since U is in echelon form (not the reduced echelon form), it is triangular,

$$U = \begin{bmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \end{bmatrix}$$

and so $\det U$ is the product of the diagonal entries $u_{11}, \ldots, u_{nn}$. If A is invertible, the entries u_{ii} are all pivots. Otherwise, at least one u_{nn} is zero. Thus

$$\det A = \begin{cases} (-1)^r \cdot (\text{product of pivots in } U) & \text{if } A \text{ invertible} \\ 0 & \text{if } A \text{ not invertible} \end{cases}$$

Theorem 3.4. A square matrix A is invertible if and only if $\det A \neq 0$.

Theorem 3.5. If A is an $n \times n$ matrix, then $\det A = \det A^T$.

Because of theorem 3.5 each statement in theorem 3.3 is also true when the word *row* is replaced by the word *column*.

Theorem 3.6. If A and B are $n \times n$ matrices, then $\det AB = \det A \cdot \det B$.

Suppose that the jth column of A is allowed to vary, and write

$$A = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{j-1} & \boldsymbol{x} & \boldsymbol{a}_{j+1} & \cdots & \boldsymbol{a}_n \end{bmatrix}$$

Define a transformation T from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$T(\boldsymbol{x}) = \det \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_{j-1} & \boldsymbol{x} & \boldsymbol{a}_{j+1} & \cdots & \boldsymbol{a}_n \end{bmatrix}$$

Then,

$$\begin{aligned} T(c\boldsymbol{x}) &= cT(\boldsymbol{x}) && \text{for all scalars } c \text{ and all } \boldsymbol{x} \text{ in } \mathbb{R}^n \\ T(\boldsymbol{u}+\boldsymbol{v}) &= T(\boldsymbol{u}) + T(\boldsymbol{v}) && \text{for all } \boldsymbol{u}, \boldsymbol{v} \text{ in } \mathbb{R}^n \end{aligned}$$

3.3 | Cramer's Rule

For any $n \times n$ matrix A and any $\boldsymbol{b}$ in $\mathbb{R}^n$, let $A_i(\boldsymbol{b})$ be the matrix obtained from A by replacing the column i by the vector $\boldsymbol{b}$.

$$A_i(\boldsymbol{b}) = \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{b} & \cdots & \boldsymbol{a}_n \end{bmatrix}$$

Theorem 3.7. Cramer's Rule.
Let A be an invertible $n \times n$ matrix. For any $\boldsymbol{b}$ in $\mathbb{R}^n$, the unique solution $\boldsymbol{x}$ of $A\boldsymbol{x} = \boldsymbol{b}$ has entries given by

$$x_i = \frac{\det A_i(\boldsymbol{b})}{\det A}, \qquad i = 1, \ldots, n$$

Cramer's rule leads easily to a general formula for the inverse of an $n \times n$ matrix A. The jth column of A^{-1} is a vector $\boldsymbol{x}$ that satisfies

$$A\boldsymbol{x} = \boldsymbol{e}_j$$

where $\boldsymbol{e}_j$ is the jth column of the identity matrix, and the ith entry of $\boldsymbol{x}$ is the (i, j)-entry of A^{-1}. By Cramer's rule,

$$\left\{(i, j)\text{-entry of } A^{-1}\right\} = x_i = \frac{\det A_i(\boldsymbol{e}_j)}{\det A} \tag{3.2}$$

Then recalling the definition of cofactor of A, equation (3.1), $\det A_i(\boldsymbol{e}_j)$ can be written as the cofactor expansion C_{ji}

$$\det A_i(\boldsymbol{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji}$$

Thus

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix} \tag{3.3}$$

The matrix of cofactors on the right side of (3.3) is called the **adjugate** of A, denoted by adj$A$.

Theorem 3.8. An Inverse Formula.
Let A be an invertible $n \times n$ matrix. Then

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

4 | Eigenvalues and Eigenvectors [§5]

4.1 | Eigenvectors and Eigenvalues

An **eigenvecotr** of an $n \times n$ matrix A is a nonzero vector $\boldsymbol{x}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\boldsymbol{x}$ of $A\boldsymbol{x} = \lambda\boldsymbol{x}$; such an $\boldsymbol{x}$ is called an *eigenvector corresponding to* λ.

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation

$$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0} \tag{4.1}$$

has a non trivial solution. The set of *all* solutions of (4.1) is just the null space of the matrix $A - \lambda I$. So this set is a *subspace* of $\mathbb{R}^n$ and is called the **eigenspace** of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

Theorem 4.1. The eigenvalues of a triangular matrix are the entries on its main diagonal.

Zero is an eigenvalue of A if and only if A is *not invertible*.

Theorem 4.2. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ is linearly independent.

Theorem. The Invertible Matrix Theorem (Continued).
Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement A is an invertible matrix (see appendix A).

(xix) The determinant of A is *not* zero;

(xx) The number 0 is *not* an eigenvalue of A.

4.2 | The Characteristic Equation

To find the eigenvalues of an $n \times n$ matrix A, one has to find all scalars λ such that the matrix equation (4.1)

$$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$$

has a nontrivial solution. By the Invertible Matrix Theorem (appendix A), this problem is equivalent to finding all λ such that the matrix $A - \lambda I$ is *not* invertible. Recalling that the determinant of a singular (not invertible) matrix is always zero, the following fact will result very useful for solving exercises.

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the **characteristic equation**

$$\det(A - \lambda I) = 0 \tag{4.2}$$

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.

The **algebraic multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

4.3 | Similarity

If A and B are $n \times n$ matrices, then A **is similar to** B if there is an invertible matrix P such that $P^{-1}AP = B$, or, equivalentely, $A = PBP^{-1}$. Writing Q for P^{-1}, we have $Q^{-1}BQ = A$. So B is also similar to A, and we say simply that A and B **are similar**. Changing A into $P^{-1}AP$ is called a **similarity transformation**.

Theorem 4.3. 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 algebraic multiplicities).

Warnings:

(i) Having the same eigenvalues does not mean being similar;

(ii) Similarity is not the same a row equivalence; row operations on a matrix usually changes its eigenvalues.

4.4 | Diagonalization

In many cases, the eigenvalue-eigenvector information contained within a matrix A can be displayed in a useful factorizartion of the form $A = PDP^{-1}$ where D is a diagonal matrix.

A 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.

Theorem 4.4. 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.

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such basis an **eigenvector basis** of $\mathbb{R}^n$.

Diagonalizing Matrices:
To diagonalize A, an $n \times n$ matrix:

Step 1 Find the eigenvalues of A using equation (4.2).

Step 2 Find n linearly independent eigenvectors of A.

Step 3 Construct P from the vectors just found.

Step 4 Construct D from the corresponding eigenvalues.

Theorem 4.5. An $n \times n$ matrix with n distrinct eigenvalues is diagonalizable.

Notice that the statement in theorem 4.5 is *not* necessary.

Theorem 4.6. Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \dots, \lambda_p$.

(i) For $1 \leq k \leq p$, the dimention of the eigenspace for λ_k is less than or equal to the algebraic multiplicity of the eigenvalue λ_k;

(ii) The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if (1) the characteristic polynomial factors completely into linear factors, and (2) the dimension of the eigenspace for each λ_k, the **geometric multiplicity**, equals the algebraic multiplicity of λ_k;

(iii) If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the set $\mathcal{B}_1, \dots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

4.5 | Eigenvectors and Linear Transformations

Let V be an n-dimentional vector space, let W be an m-dimensional vector space, and let T be any linear transformation from V to W. To associate a matrix with T, choose bases $\mathcal{B}$ and $\mathcal{C}$ for V and W, respectively.

Given any $\boldsymbol{x}$ in V, the coordinate vector $[\boldsymbol{x}]_{\mathcal{B}}$ is in $\mathbb{R}^n$ and the coordinate vector of its image, $[T(\boldsymbol{x})]_{\mathcal{C}}$, is in $\mathbb{R}^m$.

The connection between $[\boldsymbol{x}]_{\mathcal{B}}$ and $[T(\boldsymbol{x})]_{\mathcal{C}}$ can be found in the following way. Let $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_n\}$ be the basis $\mathcal{B}$ for V. If $\boldsymbol{x} = r_1\boldsymbol{b}_1 + \cdots + r_n\boldsymbol{b}_n$, then

$$[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

and

$$T(\boldsymbol{x}) = T(r_1\boldsymbol{b}_1 + \cdots + r_n\boldsymbol{b}_n) = r_1T(\boldsymbol{b}_1) + \cdots + r_nT(\boldsymbol{b}_n) \tag{4.3}$$

because T is linear. Since the coordinate mapping from W to $\mathbb{R}^m$ is linear, equation (4.3) leads to:

$$[T(\boldsymbol{x})]_{\mathcal{C}} = r_1[T(\boldsymbol{b}_1)]_{\mathcal{C}} + \cdots + r_n[T(\boldsymbol{b}_n)]_{\mathcal{C}} \tag{4.4}$$

Since $\mathcal{C}$-coordinate vectors are in $\mathbb{R}^m$, the vector equation (4.4) can be written as a matrix equation, namely

$$[T(\boldsymbol{x})]_{\mathcal{C}} = M[\boldsymbol{x}]_{\mathcal{B}} \tag{4.5}$$

where

$$M = \begin{bmatrix} [T(\boldsymbol{b}_1)] & \cdots & [T(\boldsymbol{b}_n)] \end{bmatrix} \tag{4.6}$$

The matrix M is a matrix representation of T, called the **matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$**.

Equation (4.5) says that the action of T on $\boldsymbol{x}$ may be viewed as left-multiplication by M.

In the common case where W is the same as V and the basis $\mathcal{C}$ is the same as $\mathcal{B}$, the matrix M in equation (4.6) is called the **matrix for T relative to $\mathcal{B}$**, or simply the **$\mathcal{B}$-matrix for T**, and is denoted by $[T]_{\mathcal{B}}$.

The $\mathcal{B}$-matrix for $T : V \to V$ satisfies

$$[T(\boldsymbol{x})]_{\mathcal{B}} = [T]_{\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}} \qquad \text{for all } \boldsymbol{x} \text{ is } V$$

Theorem 4.7. Diagonal Matrix Representation.
Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$.

4.6 | Complex Eigenvalues

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

The complex conjugate of a complex vector $\boldsymbol{x}$ in $\mathbb{C}^n$ is the vector $\overline{\boldsymbol{x}}$ in $\mathbb{C}^n$ whose entries are the complex conjugates of the entries in $\boldsymbol{x}$. The **real** and **imaginary parts** of a complex vector $\boldsymbol{x}$ are the vectors $\Re\boldsymbol{x}$ and $\Im\boldsymbol{x}$ formed from the real and imaginary parts of the entries of $\boldsymbol{x}$.

Let A be an $n \times n$ matrix whose entries are real. Then $\overline{A\boldsymbol{x}} = \overline{A}\overline{\boldsymbol{x}} = A\overline{\boldsymbol{x}}$. If λ is an eigenvalue of A and $\boldsymbol{x}$ is a corresponding eigenvector in $\mathbb{C}^n$, then

$$A\overline{\boldsymbol{x}} = \overline{A\boldsymbol{x}} = \overline{\lambda\boldsymbol{x}} = \overline{\lambda}\overline{\boldsymbol{x}}$$

Hence $\overline{\lambda}$ is also an eigenvalue of A, with $\overline{\boldsymbol{x}}$ a corresponding eigenvector. This shows that, when A is real, *its complex eigenvalues occur in conjugate pairs.*

Theorem 4.8. Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$ ($b \neq 0$) and an associated eigenvector $\boldsymbol{v}$ in $\mathbb{C}^2$. Then

$$A = PCP^{-1}, \text{ where } P = \begin{bmatrix} \Re\boldsymbol{v} & \Im\boldsymbol{v} \end{bmatrix} \text{ and } C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

In theorem 4.8, the matrix P provides a change of variable. The action of A amounts to a change of variables (P), followed by a rotation (C) and then a return to the original variable (P^{-1}).

The matrix C in theorem 4.8 can also be written as

$$C = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

where r is given by $\sqrt{a^2 + b^2}$ and φ represents the rotation.

4.7 | Application to Differential Equations

In many applied problems, several quantities are varying continuously in time, and they are related by a system of differential equations:

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned}$$

Here $x_1, \ldots, x_n$ are differentiable functions of t, with derivatives $x_1', \ldots, x_n'$, and the a_{ij} are constants. The crucial feature of this system is that it is *linear*. To see this, write the system as a matrix differential equation

$$\boldsymbol{x}'(t) = A\boldsymbol{x}(t) \tag{4.7}$$

where

$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \qquad \boldsymbol{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

A **solution** of equation (4.7) is a vector valued function that satisfies (4.7) for all t in some interval of real numbers.

Equation (4.7) is *linear* because both differentiation of functions and multiplication of vectors by a matrix are linear transformations.

For the general solution of equation (4.7), a solution might be a linear combination of functions of the form

$$\boldsymbol{x}(t) = \boldsymbol{v}e^{\lambda t} \tag{4.8}$$

for some scalar λ and some fixed nonzero vector $\boldsymbol{v}$. Observe that

$$\left.\begin{aligned} \boldsymbol{x}'(t) &= \lambda \boldsymbol{v}e^{\lambda t} \\ A\boldsymbol{x}(t) &= A\boldsymbol{v}e^{\lambda t} \end{aligned}\right\} \lambda\boldsymbol{v}e^{\lambda t} = A\boldsymbol{v}e^{\lambda t}$$

Since $e^{\lambda t}$ is never zero, $\boldsymbol{x}'(t)$ will equal $A\boldsymbol{x}(t)$ if and only if $\lambda\boldsymbol{v} = A\boldsymbol{v}$, that is, if and only if λ is an eigenvalue of A and $\boldsymbol{v}$ is a corresponding eigenvector. Thus each eigenvalue-eigenvector pair provides a solution (4.8) of $\boldsymbol{x}' = A\boldsymbol{x}$. Such solutions are sometimes called **eigenfunctions** of the differential equation.

For any dynamical system described by $\boldsymbol{x}'(t) = A\boldsymbol{x}$ with A an $n \times n$ matrix with n linearly independent eigenvectors (i.e. with A diagonalizable), a solution can be found in the following way.

Suppose the eigenfunctions for A are

$$\boldsymbol{v}_1e^{\lambda_1 t}, \ldots, \boldsymbol{v}_ne^{\lambda_n t}$$

with $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ linearly independent eigenvectors. Let $P = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix}$, and let D be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$ so that $A = PDP^{-1}$. Now make a *change of variable*, defining a new function $\boldsymbol{y}$ by

$$\boldsymbol{y}(t) = P^{-1}\boldsymbol{x}(t) \qquad \text{or, equivalentely} \qquad \boldsymbol{x}(t) = P\boldsymbol{y}(t)$$

The equation $\boldsymbol{x}(t) = P\boldsymbol{y}(t)$ says that $\boldsymbol{y}(t)$ is the coordinate vector of $\boldsymbol{x}(t)$ relative to the eigenvector basis. Substitution of $P\boldsymbol{y}$ for $\boldsymbol{x}$ in the equation $\boldsymbol{x}' = A\boldsymbol{x}$ gives

$$\frac{d}{dt}(P\boldsymbol{y}) = A(P\boldsymbol{y}) = (PDP^{-1})P\boldsymbol{y} = PD\boldsymbol{y} \tag{4.9}$$

Since P is a constant matrix, the left side of equation (4.9) is $P\boldsymbol{y}'$. Left-multiply both sides of (4.9) by P^{-1} and obtain $\boldsymbol{y}' = D\boldsymbol{y}$, or

$$\begin{bmatrix} y_1'(t) \\ \vdots \\ y_n'(t) \end{bmatrix} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{bmatrix}$$

The change of variables from $\boldsymbol{x}$ to $\boldsymbol{y}$ has *decoupled* the system of differential equations. Since $y_1' = \lambda_1 y_1$, we have $y_1(t) = c_1 e^{\lambda_1 t}$, with similar formulas for $y_2, \ldots, y_n$. Thus

$$\boldsymbol{y}(t) = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix}$$

the general solution $\boldsymbol{x}$ is

$$\boldsymbol{x}(t) = P\boldsymbol{y}(t) = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix} \boldsymbol{y}(t) =$$

$$= c_1 \boldsymbol{v}_1 e^{\lambda_1 t} + \cdots + c_n \boldsymbol{v}_n e^{\lambda_n t} \tag{4.10}$$

Equation (4.10) is known as the eigenfunction expansion.

In case a real matrix A has a pair of complex eigenvalues λ and $\overline{\lambda}$, with associated eigenvectors $\boldsymbol{v}$ and $\overline{\boldsymbol{v}}$, two solutions of $\boldsymbol{x}' = A\boldsymbol{x}$ are

$$\boldsymbol{x}_1(t) = \boldsymbol{v} e^{\lambda t} \qquad \text{and} \qquad \boldsymbol{x}_2(t) = \overline{\boldsymbol{v}} e^{\overline{\lambda} t}$$

where it can be shown that $\boldsymbol{x}_2(t) = \overline{\boldsymbol{x}_1(t)}$. This two solutions will obvioulsy involve complex numbers.

For a real matrix A with an eigenvalue $\lambda = a + bi$, a corresponding eigenvector $\boldsymbol{v}$ and a complex solution $\boldsymbol{x}_1(t) = \boldsymbol{v} e^{\lambda t}$ of $\boldsymbol{x}' = A\boldsymbol{x}$, two real solutions of the same $\boldsymbol{x}' = A\boldsymbol{x}$ are given by

$$\boldsymbol{y}_1(t) = \Re \boldsymbol{x}_1(t) = [\, (\Re \boldsymbol{v}) \cos bt - (\Im \boldsymbol{v}) \sin bt \,]\, e^{at}$$

$$\boldsymbol{y}_2(t) = \Im \boldsymbol{x}_1(t) = [\, (\Re \boldsymbol{v}) \sin bt + (\Im \boldsymbol{v}) \cos bt \,]\, e^{at}$$

Hence, the general solution is

$$\boldsymbol{x}(t) = c_1 \boldsymbol{y}_1(t) + c_2 \boldsymbol{y}_2(t)$$

4.8 | Trajectories of Eigenfunctions

When the matrix A in $\boldsymbol{x}'(t) = A\boldsymbol{x}(t)$ is 2×2, algebraic calculations can be supplemented by a geometric description of a system's evolution. We can plot the graph of the two eigenfunctions $\boldsymbol{x}_1(t)$ and $\boldsymbol{x}_2(t)$ as a description of what happens to the system as $t \to \infty$. The graph of a single solution $\boldsymbol{x}_1(t)$ is called a **trajectory** of the dynamical system.

The origin of the graph of a dynamical system may be:

(i) an **attractor**, or sink,

(ii) a **repeller**, or source, or

(iii) a **saddle point**.

The origin is called an *attractor* of the dynamical system when all trajectories tend toward $\mathbf{0}$. This occurs whenever both eigenvalues are negative:

$$\lambda_1 < 0 \qquad \text{and} \qquad \lambda_2 < 0$$

The direction of greatest attraction is along the line through $\mathbf{0}$ and the eigenfunction corresponding to the smaller eigenvalue.

The origin is called a *repeller* of the dynamical system when trajectories tend away from **0**. This occurs whenever both eigenvalues are positive:

$$\lambda_1 > 0 \quad \text{and} \quad \lambda_2 > 0$$

The direction of greatest repulsion is the line though **0** and the eigenfunction corresponding to the eigenvalue of larger magnitude.

The origin is called a *saddle point* of the dynamical system when some trajectories approach the origin at first and then change direction and move away from the origin. This occurs whenever one eigenvalue is positive and the other is negative:

$$\lambda_1 < 0 \quad \text{and} \quad \lambda_2 > 0$$

The direction of greatest attraction is determined by the eigenfunction for the eigenvalue of smaller magnitude. The direction of greatest repulsion is determined by the eigenfunction for the eigenvalue of greater magnitude.

In case the matrix A has complex eigenvalues given by $\lambda = a \pm bi$, the origin may be a **spiral point**. The rotation is caused by the sine and cosine functions that arise from a complex eigenvalue. When the real part of the complex eigenvalue is positive ($a > 0$), the trajectories spiral outward. When the real part of the complex eigenvalue is negative ($a < 0$), the trajectories spiral inward.

If the real part of the eigenvalue is zero ($a = 0$), the trajectories form **ellipses** around the origin.

5 | Orthogonality and Least Squares [§6]

5.1 | Inner Product and Orthogonality

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$

$$\boldsymbol{u}^T\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{u} = \boldsymbol{u}\cdot\boldsymbol{v}$$

is called the **inner product** or *dot product*. If

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

the inner product of $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$u_1v_1 + \cdots + u_nv_n$$

The **length** or **norm** of $\boldsymbol{v}$ is the nonnegative scalar $||\boldsymbol{v}||$ defined by

$$||\boldsymbol{v}|| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \quad \text{and} \quad ||\boldsymbol{v}||^2 = \boldsymbol{v}\cdot\boldsymbol{v}$$

A vector whose length is 1 is called a **unit vector**. If we divide a nonzero vector $\boldsymbol{v}$ by its length, we obtain a unit vector $\boldsymbol{u}$. This process is sometimes called **normalising**.

For $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$, the **distance** between $\boldsymbol{u}$ and $\boldsymbol{v}$, written as $\text{dist}(\boldsymbol{u},\boldsymbol{v})$, is the length of the vector $\boldsymbol{u}-\boldsymbol{v}$. That is,

$$\text{dist}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{u}-\boldsymbol{v}$$

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal** (to each other) if $\boldsymbol{u}\cdot\boldsymbol{v} = 0$.

Theorem 5.1. The Pythagorean Theorem.
Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if and only if $||\boldsymbol{u}+\boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2$.

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

A vector $\boldsymbol{x}$ is in $W^\perp$ if and only if $\boldsymbol{x}$ is orthogonal to every vector is a set that spans W. Moreover, $W^\perp$ is a subspace of $\mathbb{R}^n$.

Theorem 5.2. Let A be an $m\times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row}A)^\perp = \text{Nul}A \quad \text{and} \quad (\text{Col}A)^\perp = \text{Nul}A^T$$

For $\boldsymbol{u}$ and $\boldsymbol{v}$ in either $\mathbb{R}^2$ or $\mathbb{R}^3$, the inner product of the two vectors can be written as

$$\boldsymbol{u}\cdot\boldsymbol{v} = ||\boldsymbol{u}||\,||\boldsymbol{v}||\cos\vartheta$$

5.2 | Orthogonal Sets

A set of vectors $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vector from the set is orthogonal, that is, if $\boldsymbol{u}_i\cdot\boldsymbol{u}_j = 0$ whenever $i \neq j$.

Theorem 5.3. If $S = \{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_p\}$ 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.

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Theorem 5.4. Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\boldsymbol{y}$ in W, the weights in the linear combination

$$\boldsymbol{y} = c_1\boldsymbol{u}_1 + \cdots + c_p\boldsymbol{u}_p$$

are given by

$$c_j = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_j}{\boldsymbol{u}_j \cdot \boldsymbol{u}_j} \qquad (j = 1, \ldots, p)$$

Given a nonzero vector $\boldsymbol{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\boldsymbol{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\boldsymbol{u}$ and the other orthogonal to $\boldsymbol{u}$. We wish to write

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z} \tag{5.1}$$

where $\hat{\boldsymbol{y}} = \alpha\boldsymbol{u}$ for some scalar α and $\boldsymbol{z}$ is some vector orthogonal to $\boldsymbol{u}$. Let $\boldsymbol{z} = \boldsymbol{y} - \alpha\boldsymbol{u}$. Then $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to $\boldsymbol{u}$ if and only if

$$0 = (\boldsymbol{y} - \alpha\boldsymbol{u}) \cdot \boldsymbol{u} = \boldsymbol{y} \cdot \boldsymbol{u} - (\alpha\boldsymbol{u}) \cdot \boldsymbol{u} = \boldsymbol{y} \cdot \boldsymbol{u} - \alpha(\boldsymbol{u} \cdot \boldsymbol{u})$$

That is, equation (5.1) is satisfied if and only if $\alpha = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}$ and $\hat{\boldsymbol{y}} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$. The vector $\hat{\boldsymbol{y}}$ is called the **orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$**, and the vector $\boldsymbol{z}$ is called the **component of $\boldsymbol{y}$ orthogonal to $\boldsymbol{u}$**.

Sometimes $\hat{\boldsymbol{y}}$ is denoted by $\text{proj}_L\boldsymbol{y}$ and is called the **orthogonal projection of $\boldsymbol{y}$ onto L**. That is

$$\hat{\boldsymbol{y}} = \text{proj}_L\boldsymbol{y} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}}\boldsymbol{u}$$

Theorem 5.4 decomposes each $\boldsymbol{y}$ in $\text{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ into the sum of p projections onto one-dimensional subspaces that are mutually orthogonal.

A set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an **orthonormal basis** for W.

Theorem 5.5. An $m \times n$ matrix U has orthonormal columns if and only if $U^TU = I$.

An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$.

The reflection of a vector $\boldsymbol{y} \in \mathbb{R}^n$ in $L = \text{Span}\{\boldsymbol{u}\}$, where $\boldsymbol{u} \neq \boldsymbol{0} \in \mathbb{R}^n$, is the point $\text{refl}_L\boldsymbol{y}$ defined by

$$\text{refl}_L\boldsymbol{y} = 2 \cdot \text{proj}_L\boldsymbol{y} - \boldsymbol{y}$$

5.3 | Orthogonal Projections

Theorem 5.6. The Orthogonal Decomposition Theorem.
Let W be a subspace of $\mathbb{R}^n$. Then each $\boldsymbol{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z} \tag{5.2}$$

where $\hat{\boldsymbol{y}}$ is in W and $\boldsymbol{z}$ is in $W^\perp$. In fact, if $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} \cdot \boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p} \cdot \boldsymbol{u}_p$$

and $\boldsymbol{z} = \boldsymbol{y} - \hat{\boldsymbol{y}}$.

The vector $\hat{\boldsymbol{y}}$ in (5.2) is called the **orthogonal projection of $\boldsymbol{y}$ onto** W and is often written as $\text{proj}_W\boldsymbol{y}$.

If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an orthogonal basis for W and if $\boldsymbol{y}$ happens to be in W, then $\text{proj}_W\boldsymbol{y} = \boldsymbol{y}$.

Theorem 5.7. THE BEST APPROXIMATION THEOREM.
Let W be a subspace of $\mathbb{R}^n$, let $\boldsymbol{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\boldsymbol{y}}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Then $\hat{\boldsymbol{y}}$ is the closest point in W to $\boldsymbol{y}$, in the sense that

$$||y - \hat{y}|| < ||y - v||$$

for all $\boldsymbol{v}$ in W distinct from $\hat{\boldsymbol{y}}$.

The vector $\hat{\boldsymbol{y}}$ in theorem 5.7 is called the **best approximation to $\boldsymbol{y}$ by elements of** W.

Theorem 5.8. If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W \boldsymbol{y} = (\boldsymbol{y} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{y} \cdot \boldsymbol{u}_p)\boldsymbol{u}_p$$

If $U = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_p \end{bmatrix}$, then

$$\text{proj}_W \boldsymbol{y} = UU^T \boldsymbol{y} \qquad \text{for all } \boldsymbol{y} \text{ in } \mathbb{R}^n$$

Suppose U is an $n \times p$ matrix with orthonormal columns, and let W be the column space of U. Then

$$U^T U\boldsymbol{x} = I_p \boldsymbol{x} = \boldsymbol{x} \qquad \text{for all } \boldsymbol{x} \text{ in } \mathbb{R}^p$$

$$UU^T \boldsymbol{y} = \text{proj}_W \boldsymbol{y} \qquad \text{for all } \boldsymbol{y} \text{ in } \mathbb{R}^n$$

If U is an $n \times n$ matrix with orthonormal columns, then U is *orthogonal*, the columns space W is all of $\mathbb{R}^n$ and $UU^T\boldsymbol{y} = I\boldsymbol{y} = \boldsymbol{y}$ for all $\boldsymbol{y} \in \mathbb{R}^n$.

5.4 | The Gram-Schmidt Process

The Gram-Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$. Considering a basis $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p\}$, when the Gram-Schmidt process is applied, on any vector of the basis $\boldsymbol{s}_k$ with $1 < k \leq p$, the components of the vectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{k-1}$ parallel to $\boldsymbol{s}_k$ are removed from $\boldsymbol{s}_k$. In this way only the perpendicular part of $\boldsymbol{s}_k$ stays in the basis, making it orthogonal. The following theorem explains this process.

Theorem 5.9. THE GRAM-SCHMIDT PROCESS.
Given a basis $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\boldsymbol{v}_1 = \boldsymbol{x}_1$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1$$

$$\vdots$$

$$\boldsymbol{v}_p = \boldsymbol{x}_p - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \cdots - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_{p-1}}{\boldsymbol{v}_{p-1} \cdot \boldsymbol{v}_{p-1}}\boldsymbol{v}_{p-1}$$

Then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} = \text{Span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \qquad \text{for } 1 \leq k \leq p$$

An orthonormal basis is constructed easily from an orthogonal basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$: simply normalize all the $\boldsymbol{v}_k$.

Theorem 5.10. THE QR FACTORIZATION.
If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col$A$ and R is an $n \times n$ upper triagular invertible matrix with positive entries on its diagonal.

To find R, observe that $Q^T Q = I$, because the columns fo Q are orthonormal. Hence

$$R = Q^T A$$

5.5 | Least-square Problems

When a solution to a linear system $A\boldsymbol{x} = \boldsymbol{b}$ is demanded and none exists, the best one can do is to find an $\boldsymbol{x}$ that makes $A\boldsymbol{x}$ as close as possible to $\boldsymbol{b}$.

Think of $A\boldsymbol{x}$ as an *approximation* to $\boldsymbol{b}$. The smaller the distance between $\boldsymbol{b}$ and $A\boldsymbol{x}$, given by $||\boldsymbol{b} - A\boldsymbol{x}||$, the better the approximation. The **general least-squares problem** is to find an $\boldsymbol{x}$ that makes $||\boldsymbol{b} - A\boldsymbol{x}||$ as small as possible.

If A is $m \times n$ and $\boldsymbol{b}$ is in $\mathbb{R}^m$, a **least-square solution** of $A\boldsymbol{x} = \boldsymbol{b}$ is an $\hat{\boldsymbol{x}}$ in $\mathbb{R}^n$ such that

$$||\boldsymbol{b} - A\hat{\boldsymbol{x}}|| \leq ||\boldsymbol{b} - A\boldsymbol{x}||$$

for all $\boldsymbol{x}$ in $\mathbb{R}^n$.

The most important aspect of the least-square problem is that no matter what $\boldsymbol{x}$ we select, the vector $A\boldsymbol{x}$ will necessarely be in the column space, Col$A$. So we can seek an $\boldsymbol{x}$ that makes $A\boldsymbol{x}$ the closest point in Col$A$ to $\boldsymbol{b}$.

Given A and $\boldsymbol{b}$, let

$$\hat{\boldsymbol{b}} = \text{proj}_{\text{Col}A}\boldsymbol{b}$$

Because $\hat{\boldsymbol{b}}$ is in the column space of A, the equation $A\boldsymbol{x} = \hat{\boldsymbol{b}}$ is consistent, and there is an $\hat{\boldsymbol{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}} \tag{5.3}$$

Since $\hat{\boldsymbol{b}}$ is the closest point in Col$A$ to $\boldsymbol{b}$, a vector $\hat{\boldsymbol{x}}$ is a least-squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ if and only if $\hat{\boldsymbol{x}}$ satisfies equation (5.3).

Each least-squares solution of $A\boldsymbol{x} = \boldsymbol{b}$ satisfies the equation

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b} \tag{5.4}$$

The matrix equation (5.4) represents a system of equations called the **normal equations** for $A\boldsymbol{x} = \boldsymbol{b}$. A solution of (5.4) is often denoted by $\hat{\boldsymbol{x}}$.

Theorem 5.11. The set of least-squares solutions of $A\boldsymbol{x} = \boldsymbol{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\boldsymbol{x} = A^T\boldsymbol{b}$.

Theorem 5.12. Let A be an $m \times n$ matrix. The following statements are logically equivalent:

(i) The equation $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least-squares solution for each $\boldsymbol{b}$ in $\mathbb{R}^m$;

(ii) The columns of A are linearly independent;

(iii) The matrix $A^T A$ is invertible.

When these statements are true, the least-squares solution $\hat{\boldsymbol{x}}$ is given by

$$\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T\boldsymbol{b}$$

When a least-squares solution $\hat{\boldsymbol{x}}$ is used to produce $A\hat{\boldsymbol{x}}$ as an approximation to $\boldsymbol{b}$, the distance from $\boldsymbol{b}$ to $A\hat{\boldsymbol{x}}$, $||\boldsymbol{b} - \hat{\boldsymbol{b}}|| = ||\boldsymbol{b} - A\hat{\boldsymbol{x}}||$, is called the **least-squares error** of this approximation.

Theorem 5.13. Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A, as in theorem 5.10. Then, for each $\boldsymbol{b}$ in $\mathbb{R}^m$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a unique least-squares solution, given by

$$\hat{\boldsymbol{x}} = R^{-1} Q^T\boldsymbol{b}$$

6 | Symmetric Matrices [§7]

6.1 | Diagonalization of Symmetric Matrices

A **symmetric** matrix is a matrix A such that $A^T = A$. Such a matrix is necessarily square. Its main diagonal entries may be arbitrary, but its other entries occur in pairs.

Theorem 6.1. If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1} \tag{6.1}$$

Such a diagonalization requires n linearly independent and orthonormal eigenvectors. If A is orthogonally diagonalizable as in (6.1), then

$$A^T = (PDP^T)^T = PDP^T = A$$

Thus A is symmetric. Note that the diagonal entries of D are the eigenvalues $\lambda_1, \dots, \lambda_n$.

Theorem 6.2. An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.

The set of eigenvalues of a matrix A is sometimes called the *spectrum* of A, and the following description of the eigenvalues is called a *spectral theorem*.

Theorem 6.3. The Spectral Theorem for Symmetric Matrices.
An $n \times n$ symmetric matrix A has the following proprieties:

(i) A has n real eigenvalues, counting algebraic multiplicities;

(ii) The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation;

(iii) The eigenspaces are mutually orthogonal;

(iv) A is orthogonally diagonalizable.

Using the definition of orthogonal diagonalization given by equation (6.1) we can rewrite a matrix A as

$$A = \lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T + \cdots + \lambda_n \boldsymbol{u}_n \boldsymbol{u}_n^T \tag{6.2}$$

This representation of A is called **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum of A. Each term in equation (6.2) is an $n \times n$ matrix of rank 1. Furthermore, each matrix $\boldsymbol{u}_j \boldsymbol{u}_j^T$ is a **projection matrix** in the sense that for each $\boldsymbol{x}$ in $\mathbb{R}^n$, the vector $(\boldsymbol{u}_j \boldsymbol{u}_j^T)\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}$ onto the subspace spanned by $\boldsymbol{u}_j$.

6.2 | Quadratic Forms

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

If $\boldsymbol{x}$ represents a variable vector in $\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$\boldsymbol{x} = P\boldsymbol{y} \qquad \text{or equivalentely,} \qquad \boldsymbol{y} = P^{-1}\boldsymbol{x} \tag{6.3}$$

where P is an invertible matrix and $\boldsymbol{y}$ is a new variable vector in $\mathbb{R}^n$.

If the change of variable (6.3) is made in a quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$, then

$$\boldsymbol{x}^T A\boldsymbol{x} = \boldsymbol{y}^T (P^T AP)\boldsymbol{y} \tag{6.4}$$

and the new quadratic form is $P^T AP$. Since A is symmetric, by theorem 6.2, there is an *orthogonal* matrix P such that $P^T AP$ is a diagonal matrix D, and the quadratic form in (6.4) becomes $\boldsymbol{y}^T D\boldsymbol{y}$.

Theorem 6.4. The Principal Axes Theorem.
Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable $\boldsymbol{x} = P\boldsymbol{y}$, that transforms the quadratic form from $\boldsymbol{x}^T A\boldsymbol{x}$ into a quadratic form $\boldsymbol{y}^T D\boldsymbol{y}$ with no cross-product term.

The columns of P in theorem 6.4 are called the **principal axes** of the quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$. The vector $\boldsymbol{y}$ is the coordinate vector of $\boldsymbol{x}$ relative to the orthonormal basis of $\mathbb{R}^n$ given by these principal axes.

Suppose $Q(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$ where A is an invertible 2×2 symmetric matrix, and let c be a constant. The set of all $\boldsymbol{x}$ in $\mathbb{R}^2$ that satisfy

$$\boldsymbol{x}^T A\boldsymbol{x} = c$$

either corresponds to an ellipse, a hyperbola, two intersecting lines, a single point or contains no points at all. If A is a diagonal matrix, the graph is in standard position, otherwise, if A is not diagonal, the graph is rotated out of strandard position.

When A is an $n \times n$ matrix, the quadratic form $Q(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$ is a real-valued function with domain $\mathbb{R}^n$.

A quadratic form Q is:

(i) **positive definite** if $Q(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$,

(ii) **negative definite** if $Q(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$,

(iii) **indefinite** if $Q(\boldsymbol{x})$ assumes both positive and negative values.

Also Q is said to be *positive semidefinite* if $Q(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$, and to be *negative semidefinite* if $Q(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x}$.

Theorem 6.5. Quadratic Forms and Eigenvalues.
Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$ is:

(i) positive definite if and only if the eigenvalues of A are all positive,

(ii) negative definite if and only if the eigenvalues of A are all negative, or

(iii) indefinite if and only if A has both positive and negative eigenvalues.

A | Invertible Matrix Theorem

Theorem. The Invertible Matrix Theorem.
Let A be a square matrix $n \times n$. Then the following statements are equivalent. That is, for a given A, the statements are either all true or all false.

(i) A is an invertible matrix;

(ii) A is row equivalent to the $n \times n$ identity matrix;

(iii) A has n pivot points;

(iv) The equation $A\boldsymbol{x} = \mathbf{0}$ has only the trivial solution;

(v) The columns of A form a linearly independent set;

(vi) The linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is one-to-one;

(vii) The equation $A\boldsymbol{x} = \boldsymbol{b}$ has at least one solution for each $\boldsymbol{b}$ in $\mathbb{R}^n$;

(viii) The columns of A span $\mathbb{R}^n$;

(ix) The linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$;

(x) There is an $n \times n$ matrix C such that $CA = I$;

(xi) There is an $n \times n$ matrix D such that $AD = I$;

(xii) A^T is an invertible matrix;

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

(xiv) $\text{Col}A = \mathbb{R}^n$;

(xv) $\text{dimCol}A = n$;

(xvi) $\text{rank}A = n$;

(xvii) $\text{Nul}A = \{\mathbf{0}\}$;

(xviii) $\text{dimNul}A = 0$;

(xix) The determinant of A is *not* zero;

(xx) The number 0 is *not* an eigenvalue of A.

This was:

WI1403-LR Linear Algebra

Version 1.0

More summaries are available for the following courses:

- AE1110 Intro to AE I - Module E: Aerodynamics
- AE1110 Intro to AE I - Module F: Flight Mechanics
- AE1108-I Aerospace Materials
- AE1240-I Physics I
- AE1240-II Physics II
- AE3211-II Production of Aerospace Systems
- AE4447 Aircraft Performance Optimization
- WI1402-LR Calculus II
- WI1403-LR Linear Algebra