Linear Algebra Summary

Based on *Linear Algebra and its applications* by David C. Lay

Preface

The goal of this summary is to offer a complete overview of all theorems and definitions introduced in the chapters of *Linear Algebra and its applications by David C. Lay* that are relevant to the Linear Algebra course at the faculty of Aerospace Engineering at Delft University of Technology. All theorems and definitions have been taken over directly from the book, whereas the accompanying explanation is sometimes formulated in my own words.

Linear Algebra might seem more abstract than the sequence of Calculus courses that are also taken in the first year of the Bachelor of Aerospace Engineering. A great part of the course consists of definitions and theorems that follow from these definitions. An analogy might be of help to your understanding of the relevance of this course. Imagine visiting a relative, who has told you about a collection of model airplanes that are stored in his or her attic. The aerospace enthusiast you are, you insist on taking a look. Upon arrival you are exposed to a complex, yet looking systematic, array of boxes and drawers. The amount of boxes and drawers seems endless, yet your relative knows exactly which contain the airplane models. Having bragged about your challenging studies, the relative refuses to tell you exactly where they are and demands that you put in some effort yourself. However, your relative explains you exactly how he has sorted the boxes and also tells you in which box or drawer to look to discover the contents of several other boxes.

A rainy afternoon later, you have completely figured out the system behind the order of the boxes, and find the airplane models in the first box you open. The relative hints at a friend of his, whose father also collected aircraft models which are now stored in his basement. Next Sunday you stand in the friend's basement and to your surprise you figure out that he has used the exact same ordering system as your relative! Within less than a minute you have found the aircraft models and can leave and enjoy the rest of your day. During a family dinner, the first relative has told your entire family about your passion about aerospace, and multiple others approach you about useful stuff lying in their attics and basement. Apparently, the ordering system has spread across your family and you never have to spend a minute too long in a stale attic or basement again!

That is were the power of Linear Algebra lies: a systematic approach to mathematical operations allowing for fast computation.

Enjoy and good luck with your studies.

Contents

1 Linear Equations in Linear Algebra

1.1 Systems of linear equations

A **linear equation** is an equation that can be written in the form:

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

where b and the **coefficients** a_n may be real or complex. Note that the common equation $y = x + 1$ describing a straight line intercepting the y-axis at the point $(0, 1)$, is a simple example of a linear equation of the form $x_2 - x_1 = 1$ where $x_2 = y$ and $x_1 = x$.
A system of one or more linear equations involving the same variables is called a **system of linear equations**. A solution of such a **linear system** is a list of numbers $(s_1, s_2, ..., s_n)$ that satisfies all equations in the system when substituted for variables $x_1, x_2, ..., x_n$. The set of all solution lists is denoted as the **solution set**.

A simple example is finding the intersection of two lines, such as:

$$x_2 = x_1 + 1$$
$$x_2 = 2x_1$$

For consistency we write above equations in the form defined for a linear equation:

$$x_2 - x_1 = 1$$
$$x_2 - 2x_1 = 0$$

Solving gives one the solution set $(1, 2)$. A solution can always be validated by substituting the solution for the variables and find if the equation is satisfied.

To continue on our last example, we also know that besides an unique solution (i.e. the intersection of two or more lines) there also exists the possibility of two ore more lines being parallel or coincident, as shown in figure 1.1. We can extent this theory for a linear system containing 2 variables to any linear system and state that

A system of linear equations has

1. no solution, or
2. exactly one unique solution, or
3. infinitely many solutions

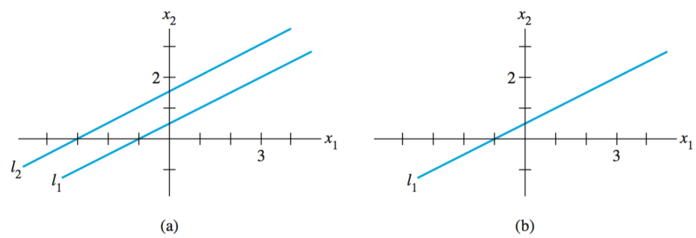

Figure 1.1: A linear system with no solution (a) and infinitely many solutions (b)

A linear system is said to be **consistent** if it has one or infinitely many solutions. If a system does not have a solution it is **inconsistent**.

Matrix notation

It is convenient to record the essential information of a linear system in a rectangular array called a **matrix**. Given the linear system:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ -4x_1 + 5x_2 + 9x_3 &= -9 \end{aligned}$$

We can record the coefficients of the system in a matrix as:

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ -4 & 5 & 9 \end{bmatrix}$$

Above matrix is denoted as the **coefficient matrix**. Adding the constants b from the linear system as an additional column gives us the **augmented matrix** of the system:

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{bmatrix}$$

It is of high importance to know said difference between a coefficient matrix and an augmented matrix for later definitions and theorems.
The size of a matrix is denoted in the format $m \times n$ where m signifies the amount of rows and n the amount of columns.

Solving a linear system

First we define the following 3 **elementary row operations**:

1. (Replacement) Replace a row by the sum of itself and the multiple of another row
2. (Interchange) Interchange two rows

3. (Scale) Scale all entries in a row by a nonzero constant

Two matrices are defined as **row equivalent** if a sequence of elementary row operations transforms the one in to the other.

The following fact is of great importance in linear algebra:

> If the augmented matrices of two linear systems are row equivalent, the two linear systems have the same solution set.

This theorem grants one the advantage of greatly simplifying a linear system using elementary row operations before finding the solution of said system, as the elementary row operations do not alter the solution set.

1.2 Row reduction and echelon forms

For the definitions that follow it is important to know the precise meaning of a **nonzero row or column** in a matrix, that is a a row or column containing at least one nonzero entry. The leftmost nonzero entry in a matrix row is called the **leading entry**.

DEFINITION

A rectangular matrix is in the **echelon** form if it has the following three properties:

1. All nonzero rows are above any rows of all zeros
2. Each leading entry in a row is to the right of the column of the leading entry of the row below it
3. All entries in a a column below a leading entry are zeros

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

Above matrix is an example of a matrix in echelon form. Leading entries are symbolized by $\blacksquare$ and may have any nonzero value whereas the positions $*$ may have any value, nonzero or zero.

DEFINITION

We can build upon the definition of the echelon form to arrive at the **reduced echelon form**. In addition to the three properties introduced above, a matrix must satisfy two other properties being:

1. The leading entry in each row is 1
2. All entries in the column of a leading entry are zero

Extending the exemplary matrix to the reduced echelon form gives us:

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

Where $*$ may be a zero or a nonzero entry. We also find that the following theorem must hold:

THEOREM 1 **Uniqueness of the Reduced Echolon Form**

Each matrix is row equivalent to only one reduced echelon matrix.

Pivot positions

A **pivot position** of a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column containing a pivot position. A square ($\blacksquare$) denotes a pivot position in matrix 1.2.

Solutions of linear systems

A reduced echelon form of an augmented matrix of a linear system leads to an explicit statement of the solution set of this system. For example, row reduction of the augmented matrix of an arbitrary system has led to the equivalent unique reduced echelon form:

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

There are three variables as the augmented matrix (i.e. including the constants b of the linear equations) has four columns, hence the linear system associated with the reduced echelon form above is:

$$\begin{aligned} x_1 - 5x_3 &= 1 \\ x_2 + x_3 &= 4 \\ 0 &= 0 \end{aligned}$$

The variables x_1 and x_2 corresponding to columns 1 and 2 of the augmented matrix are called **basic variables** and are explicitly assigned to a set value by the **free variables** which in this case is x_3. As hinted earlier, a consistent system can be solved for the basic variables in terms of the free variables and constants. Carrying out said operation for the system above gives us:

$$\begin{cases} x_1 = 1 + 5x_{x3} \\ x_2 = 4 - x_3 \\ x_3 \text{ is free} \end{cases}$$

Parametric descriptions of solution sets

The form of the solution set in the previous equation is called a **parametric representation** of a solution set. Solving a linear system amounts to finding the parametric representation of the solution set or finding that it is empty (i.e. the system is inconsistent). The convention is made that the free variables are always used as parameters in such a parametric representation.

Existence and Uniqueness Questions

Using our the previously developed definitions we can introduce the following theorem:

THEOREM 3

Existence and Uniqueness Theorem

A linear system is consistent only if the rightmost column of the augmented matrix is not a pivot column: the reduced echelon form of the of the augmented matrix has no row of the form:

$$[0 \quad \dots \quad 0 \quad b] \text{ with } b \text{ nonzero}$$

If a linear system is indeed consistent it has either one unique solution, if there are no free variables, or infinitely many solutions if there is one or more free variable.

1.3 Vector equations

Vectors in $\mathbb{R}^2$

A matrix with only one column is referred to as a **column vector** or simply a **vector**. An example is:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where u_1 and u_2 are real numbers. The set of all vectors with two entries is denoted by $\mathbb{R}^2$. (similar to the familiar x-y coordinate system)
The sum of two vectors such a $\mathbf{u}$ and $\mathbf{v}$ is obtained as:

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$$

Given a real number c and a vector $\mathbf{u}$, the **scalar multiple** of $\mathbf{u}$ by c is found by:

$$c\mathbf{u} = c \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}$$

The number c is called a **scalar**.

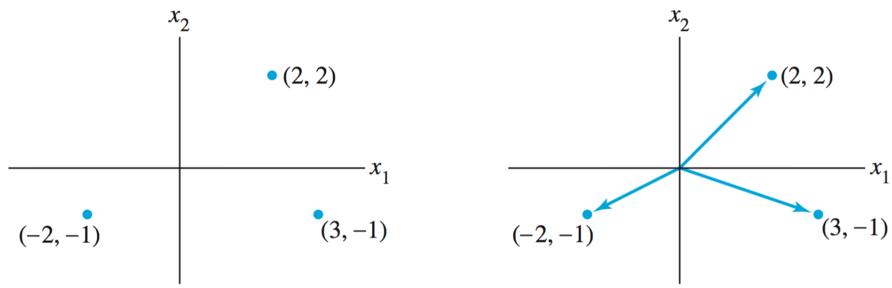

Figure 1.2: Geometrical representation of vectors in $\mathbb{R}^2$ as points and arrows

Vectors in $\mathbb{R}^n$

We can extend the discussion on vectors in $\mathbb{R}^2$ to $\mathbb{R}^n$. If n is a positive integer, $\mathbb{R}^n$ denotes the collection of all ordered lists of n real numbers, usually referred to as $n \times 1$ matrices:

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

The vector whose entries are all zero is called the **zero vector** and is denoted by $\mathbf{0}$. The addition and multiplication operations discussed for $\mathbb{R}^2$ can be extended to $\mathbb{R}^n$.

THEOREM

Algebraic Properties of $\mathbb{R}^n$

For all $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ and all scalars c and d:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{0} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = cd\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

Linear combinations

Given vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ and **weights** $c_1, c_2, \ldots, c_p$ we can define the **linear combination** $\mathbf{y}$ by:

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p$$

Note that we can reverse this situation and determine whether a vector $\mathbf{y}$ exists as a linear combination of given vectors. Hence we would determine if there is a combination of weights $c_1, c_2, \ldots, c_p$ that leads to $\mathbf{y}$. This would amount to us finding the solution of a $n \times (p + 1)$ matrix where n is the length of the vector and p denotes the amount of vectors available to the linear combination. We arrive at the following fact:

A vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n a_n = \mathbf{b}$$

has the same solution as the linear system whose augmented matrix is

$$[\mathbf{a}_1 \quad \mathbf{a}_2 \quad ... \quad \mathbf{a}_n \quad \mathbf{b}]$$

In other words, vector $\mathbf{b}$ can only be generated by a linear combination of $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ if there exists a solution to the linear system corresponding to the matrix above.

A question that often arises during the application of linear algebra is what part of $\mathbb{R}^n$ can be spanned by all possible linear combinations of vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$. The following definition sheds light on this question:

DEFINITION

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 as Span$\{\mathbf{v}_1, ..., \mathbf{v}_p\}$ and is called the **subset of** $\mathbb{R}^n$ **spanned by** $\mathbf{v}_1, ..., \mathbf{v}_p$.That is, Span$\{\mathbf{v}_1, ..., \mathbf{v}_p\}$ is the collection of all vectors that can be written in the form:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_p\mathbf{v}_p$$

with $c_1, ..., c_p$ scalars.

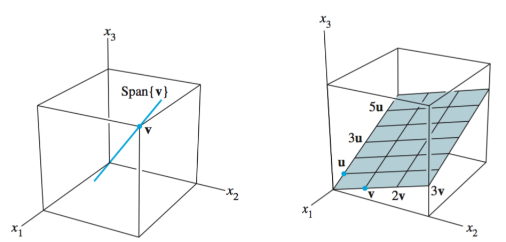

Figure 1.3: Geometric interpretation of Span in $\mathbb{R}^3$

Let $\mathbf{v}$ be a nonzero vector in $\mathbb{R}^3$. Then Span$\{\mathbf{v}\}$ is the set of all scalar multiples of $\mathbf{v}$, which is the set of points on the line through $\mathbf{0}$ and $\mathbf{v}$ in $\mathbb{R}^3$. If we consider another vector $\mathbf{u}$ which is not the zero vector or a multiple of $\mathbf{v}$, Span$\{\mathbf{u}, \mathbf{v}\}$ is a plane in $\mathbb{R}^3$ containing $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{0}$.

1.4 The matrix equation $A\mathbf{x} = \mathbf{b}$

We can link the ideas developed in sections 1.1 and 1.2 on matrices and solution sets to the theory on vectors from section 1.3 with the following definition:

DEFINITION

If A is a $m \times n$ matrix, with columns $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ and if $\mathbf{x}$ is in $\mathbb{R}^n$ then the product of A and $\mathbf{x}$, denoted as $A\mathbf{x}$, is the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights; that is,

$$A\mathbf{x} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n$$

An equation of the form $A\mathbf{x} = \mathbf{b}$ is called a **matrix equation***. Note that such a matrix equation is only defined if the number of columns of A equals the number of entries of $\mathbf{x}$. Also note how we area able to write any system of linear equations or any vector equation in the form $A\mathbf{x} = \mathbf{b}$. We use the following theorem to link these concepts:

THEOREM 3

If A is a $m \times n$ matrix with columns $\mathbf{a}_1, ..., \mathbf{a}_n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, then the matrix equation

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

has the same solution set as the vector equation

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

which has the same solution set as the system of linear equations whose augmented matrix is

$$[\mathbf{a}_1 \quad ... \quad \mathbf{a}_n \quad \mathbf{b}]$$

The power of the theorem above lies in the fact that we are now able to see a system of linear equations in multiple ways: as a vector equation, a matrix equation and simply as a linear system. Depending on the nature of the physical problem one would like to solve, one can use any of the three views to approach the problem. Solving it will always amount to finding the solution set to the augmented matrix.

Another theorem is introduced, composed of 4 logically equivalent statements:

THEOREM 4

Let A be a $m \times n$ matrix, then the following 4 statements are logically equivalent (i.e. all true or false for matrix A):

1. For each b in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution
2. Each $\mathbf{b}$ is a linear combination of the columns of A
3. The columns of A span $\mathbb{R}^m$
4. A has a pivot position in every row

PROOF Statements 1, 2 and 3 are equivalent due to the definition of $\mathbb{R}^m$ and the matrix equation. Statement 4 requires some additional explanation. If a matrix A has a pivot position in every row, we have excluded the possibility that the last column of the augmented matrix of the linear system involving A has a pivot position (one row cannot have

2 pivot positions by its definition). If there would be a pivot position in the last column of the augmented matrix of the system, we induce a possible inconsistency for certain vectors **b**, meaning that the first three statements of above theorem are false: there are possible vectors **b** that are in $\mathbb{R}^m$ but not in the span of the columns of A.

The following properties hold for the matrix-vector product $A\mathbf{x} = \mathbf{b}$:

THEOREM 5

If A is a $m \times n$ matrix, **u** and **v** are vectors in $\mathbb{R}^n$ and c is a scalar:

a. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$

b. $Ac(\mathbf{u}) = c(A\mathbf{u})$

1.5 Solution sets of linear systems

Homogeneous Linear Systems

A linear system is said to be **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$ where A is a $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Systems like these always have at least one solution, namely $\mathbf{x} = 0$, which is called the **trivial solution**. An important question regarding these homogeneous linear systems is whether they have a **nontrivial solution**. Following the theory developed in earlier sections we arrive at the following fact:

The homogeneous equation $A\mathbf{x} = \mathbf{0}$ only has a nontrivial solution of it has at least one free variable.

If there is no free variable (i.e. the coefficient matrix has a pivot position in every column) the solution $\mathbf{x}$ would always amount to $\mathbf{0}$ as the last column in the augmented matrix consists entirely of zeros, which does not change during elementary row operations.

We can also note how every solution set of a homogeneous linear system can be written as a parametric representation of n vectors where n is the amount of free variables. Lets give an illustration with the following homogeneous system:

$$x_1 - 3x_2 - 2x_3 = 0$$

Solving this system can be done without any matrix operations, the solution set is $x_1 = 3x_2 + 2x_3$ .Rewriting this final solution as a vector gives us:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3x_2 + 2x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

Hence we can interpret the solution set as all possible linear combinations of two vectors. The solution set is the span of the two vectors above.

Parametric vector form

The representation of the solution set of above example is called the **parametric vector form**. Such a solution the matrix equation $A\mathbf{x} = 0$ can be written as:

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \quad (s, t \text{ in } \mathbb{R})$$

Solutions of nonhomogeneous systems

THEOREM 6

Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some b, and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_p$, where $\mathbf{v}_p$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

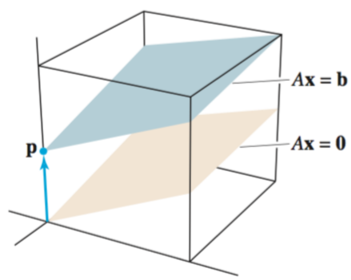

Figure 1.4: Geometrical interpretation of the solution set of equations $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$

Why does this make sense? Let's come up with an analogy. We have a big field of grass and a brand-new autonomous electric car. The electric car is being tested and always drives the same pattern, that is, for a specified moment in time it always goes in a certain direction. The x and y position of the car with respect to one of the corners of the grass field are its fixed variables, whereas the time t is its free variable: it is known how x and y vary with t but t has to be specified! One of the companies' employees observes the pattern the car drives on board of a helicopter: after the car has reached the other end of the field he has identified the pattern and knows how x and y vary with t.

Now, we would like to have the car reach the end of the field at the location of a pole, which we can achieve by displacing the just observed pattern such that the pattern intersects with the the pole at the other end of the field. Now each point in time satisfies the trajectory leading up to the pole, and we have found our solution. Notice how this is similar? The behaviour of the solution set does not change, the boundary condition and thus the positions passed by the car do change!

1.6 Linear Independence

We shift the knowledge applied on homogeneous and nonhomogeneous equations of the form $A\mathbf{x} = \mathbf{b}$ to that of vectors. We start with the following definition:

DEFINITION

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

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

has only the trivial solution. The set $\{\mathbf{v}_1, ..., \mathbf{v}_p\}$ is said to be **linearly dependent** if there exists weights $c_1, ..., c_p$, not all zero, such that:

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

Using this theorem we can also find that:

The columns of matrix A are linearly independent only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Sets of vectors

In case of a set of only two vectors we have that:

A set of two vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.

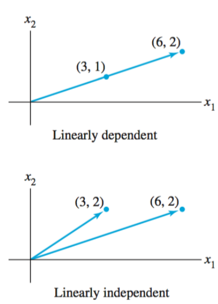

Figure 1.5: Geometric interpretation of linear dependence and independence of a set of two vectors.

We can extend to sets of more than two vectors by use of the following theorem on the characteriziation of linearly dependent sets:

THEOREM 7

Characterization of Linearly Dependent Sets

An indexed set $S = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ of more than two 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 $\mathbf{v}_1, ..., \mathbf{v}_{j-1}$.

We also have theorems describing special cases of vector sets, for which the linear dependence is automatic:

THEOREM 8 If a set contains more vectors than the number of entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v}_1, ..., \mathbf{v}_p\}$ is linearly dependent if $p > n$.

PROOF Say we have a matrix $A = [\mathbf{v}_1 \quad ... \quad \mathbf{v}_p]$. Then A is a $n \times p$ matrix. As $p > n$ we know that the coefficient matrix of A cannot have a pivot position in every column, thus there must be free variables. Now we know that the equation $A\mathbf{x} = \mathbf{0}$ also has a nontrivial solution, thus the set of vectors is linearly dependent.
The second special case is the following:

THEOREM 9 If a set $S = \{\mathbf{v}_1, ..., \mathbf{v}_p\}$ in $\mathbb{R}^n$ contains the zero vector $\mathbf{0}$, then the set is linearly dependent.

PROOF Note that if we assume that $\mathbf{v}_1 = \mathbf{0}$ we can write a linear combination as follows:

$$1\mathbf{v}_1 + 0\mathbf{v}_2 + ... + 0\mathbf{v}_p = \mathbf{0}$$

As not all weights are zero, we have a nontrivial solution and the set is linearly dependent.

1.7 Introduction to 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 $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the domain of T and the set $\mathbb{R}^m$ is called the codomain of T. The notation $T : \mathbb{R}^n \to \mathbb{R}^m$ indicates that $\mathbb{R}^n$ is the domain and $\mathbb{R}^m$ is the codomain of T. 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 of all images $T(\mathbf{x})$ is called the **range** of T.

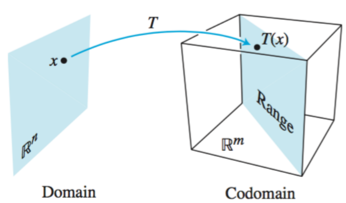

Figure 1.6: Visualization of domain, codomain and range of a transformation

Matrix Transformations

For matrix transformations, $T(\mathbf{x})$ is computed as $A\mathbf{x}$. Note that A is a $m \times n$ matrix: the domain of T is thus $\mathbb{R}^n$ as the number of entries in $\mathbf{x}$ must be equal to the amount of columns n. The codomain is $\mathbb{R}^m$ as the amount of entries (i.e. rows) in the columns of A is m. The range of T is the set of all linear combinations of the columns of A.

Linear Transformations

Recall from section 1.4 the following two algebraic properties of the matrix equation:

$$A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} \quad \text{and} \quad A(c\mathbf{u}) = cA\mathbf{u}$$

Which hold for $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ and c scalar. We arrive at the following definition for linear transformations:

DEFINITION

A transformation T is **linear** if:

- (i) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in the domain of T
- (ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars and $\mathbf{u}$ in the domain of T

Note how every matrix transformation is a linear transformation by the algebraic properties recalled from section 1.4. These two properties lead to the following useful facts:

If T is a linear transformation, then

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

and

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

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

Extending this last property to linear combinations gives us:

$$T(c_1\mathbf{v}_1 + ... + c_p\mathbf{v}_p) = cT(\mathbf{v}_1) + ... + c_pT(\mathbf{v}_p)$$

1.8 The Matrix of a Linear Transformation

The discussion that follows shows that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$. We start with the following theorem:

THEOREM 10

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

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

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

$$A = [T(\mathbf{e}_1) \quad ... \quad T(\mathbf{e}_n)]$$

The matrix A is called the **standard matrix for the linear transformation** T. We now know that every linear transformation is a matrix transformation, and vice versa. The term *linear transformation* is mainly used when speaking of mapping methods, whereas the term *matrix transformation* is a means of describing how such mapping is done.

Existence and Uniqueness Questions

The concept of linear transformations provides a new way of interpreting the existence and uniqueness questions asked earlier. We begin with the following definition:

DEFINITION

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 one or more $\mathbf{x}$ in $\mathbb{R}^n$.

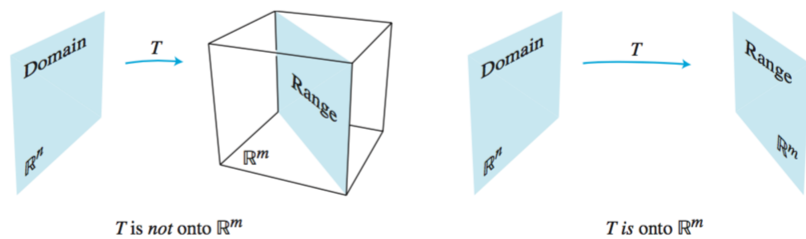

Figure 1.7: Geometric interpretation of existence and uniqueness questions in linear transformations

Note how the previous definition is applicable if each vector $\mathbf{b}$ has *at least* one solution. For the special case where each vector $\mathbf{b}$ has *only* one solution we have the definition:

DEFINITION

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 **only one** $\mathbf{x}$ in $\mathbb{R}^n$.

Note that for above definition, T does not have to be *onto* $\mathbb{R}^m$. This uniqueness question is simple to answer with this theorem:

THEOREM 11

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

PROOF Assume that our transformation T is not one-to-one. Hence there are 2 distinct vectors in $\mathbb{R}^n$ which have the same image $\mathbf{b}$ in $\mathbb{R}^m$, lets call these vectors $\mathbf{u}$ and $\mathbf{v}$. As $T(\mathbf{u}) = T(\mathbf{v})$, we have that:

$$T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0}$$

Hence $T(\mathbf{x}) = \mathbf{0}$ has a nontrivial solution, excluding the possibility of T being one-to-one. We can also state that:

THEOREM 12

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

a. T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ only if the columns of A span $\mathbb{R}^m$.

b. T is one-to-one only if the columns of A are linearly independent.

PROOF

a. The columns of A span $\mathbb{R}^m$ if $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$, hence every $\mathbf{b}$ has at least one vector $\mathbf{x}$ for which $T(\mathbf{x}) = \mathbf{b}$

b. This theorem is just another notation of the theorem that $T(\mathbf{x}) = \mathbf{0}$ only having the trivial solution means that it is one-to-one. Linear independence of columns of A suggests no nontrivial solution.

Matrix algebra

2

2.1 Matrix operations

Once again we refer to the definition of a matrix, allowing us to precisely define the matrix operations that follow.

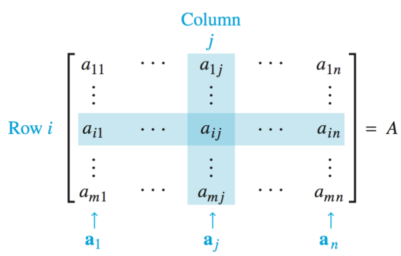

Figure 2.1: Matrix notation

If A is an $m \times n$ matrix, then the scalar entry in the ith row and jth column is denoted as a_{ij} and is is referred to as the (i, j)-entry of A. The **diagonal entries** in an $m \times n$ matrix A are $a_{11}, a_{22}, a_{33}, ...$ and they form the **main diagonal** of A. A **diagonal matrix** is a $n \times n$, thus square, matrix whose nondiagonal entries are zero. An $m \times n$ matrix whose entries are all zero is called the **zero matrix** and is written as 0.

Sums and scalar multiples

We can extend the arithmetic used for vectors to matrices. We first define two matrices to be **equal** if they are of the same size and their corresponding columns are equal (i.e. all entries are the same). If A and B are $m \times n$ matrices, then their **sum** is computed as the sum of the corresponding columns, which are simply vectors! For example, let A and B be 2×2 matrices, then the sum is:

$$A + B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

More general, we have the following algebraic properties of matrix addition:

THEOREM 1

Let A,B and C be matrices of the same size and let r and s be scalars:

a. $A + B = B + A$

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

c. $A + 0 = A$

d. $r(A + B) = rA + rB$

e. $(r + s)A = rA + sA$

f. $r(sA) = (rs)A$

Matrix multiplication

When a matrix B multiplies a vector $\mathbf{x}$, the result is a vector $B\mathbf{x}$, if this vector is in turn multiplied by a matrix A the result is the vector $A(B\mathbf{x})$. It is essentially a composition of two mapping procedures. We would like to represent this process as one multiplication of the vector $\mathbf{x}$ with a given matrix so that:

$$A(B\mathbf{x}) = (AB)\mathbf{x}$$

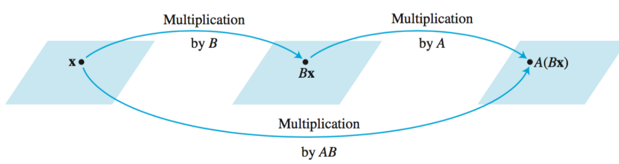

Figure 2.2: Matrix multiplication

As shown in figure 2.2. We can easily find an answer to this question, if we assume A to be a $m \times n$ matrix, B a $n \times p$ matrix and $\mathbf{x}$ in $\mathbb{R}^p$. Then:

$$B(\mathbf{x}) = x_1\mathbf{b}_1 + ... + x_p\mathbf{b}_p$$

By the linearity of the matrix transformation by A:

$$A(B\mathbf{x}) = x_1A\mathbf{b}_1 + ... + x_pA\mathbf{b}_p$$

Note however that the vector $A(\mathbf{b}x)$ is simply a linear combination of the vectors $A\mathbf{b}_1, ..., A\mathbf{b}_p$ using the entries of $\mathbf{x}$ as weights. In turn, $A\mathbf{b}_1$ is simply a linear combination of the columns of matrix A using the entries of $\mathbf{b}_1$ as weights! Now it becomes simple to define the following:

DEFINITION

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

$$AB = A[\mathbf{b}_1 \quad ... \quad \mathbf{b}_p] = [A\mathbf{b}_1 \quad ... \quad A\mathbf{b}_p]$$

Note how 2.1 is now true for all $\mathbf{x}$ using above definition. We can say that: *matrix multiplication corresponds to composition of linear transformations.*

It is important to have some intuition and knowledge about above definition. However, in practice it is convenient to use the following computation rule:

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 in row i of matrix A and column j in matrix B. Let A be an $m \times n$ matrix, then:

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{mj}$$

Properties of matrix multiplication

For the following properties it is important to recall that I_m denotes the $m \times m$ identity matrix and $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

THEOREM 2

Let A be an $m \times m$ matrix and let B and C be matrices of appropriate sizes for the products and sums defined:

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B + C) = AB + AC$ (left distributive law)

c. $(B + C)A = BA + CA$ (right distributive law)

d. $r(AB) = (rA)B = A(rB)$ for any scalar r

e. $I_m A = A = AI_n$ (identity for matrix multiplication)

Note how properties b and c might be confusing. The importance of these definitions is that BA and AB are usually not the same, as BA uses the columns of B to form a linear combination with columns of A as weights. The product AB however uses the columns of A to form a linear combination with columns of B as weights. If $AB = BA$, we say that A and B **commute** with each other.

Powers of a Matrix

If A is an $n \times n$ matrix and k is a positive integer, then A^k denotes the the product of k copies of A:

$$A^k = \underbrace{A...A}_{k}$$

The Transpose of a Matrix

Let A be an $m \times n$ matrix, then the transpose of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A. If

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

Then the transpose of A is:

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

THEOREM 3

Let A and B denote matrices of size appropriate for the following operations, then:

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

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

c. For any scalar r, $(rA)^T = rA^T$

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

2.2 The Inverse of a Matrix

Recall that the multiplicative inverse of a number such as 4 is $\frac{1}{4}$ or 4^{-1}. This inverse satisfies the equations:

$$5\frac{1}{5} = 1 \quad 55^{-1} = 1$$

Note how both equations are needed if we offer a generalization for a matrix inverse, as matrix multiplication is not commutative (i.e. in general $AB \neq BA$). The matrices involved in this generalization must be square.
An $n \times n$ matrix A is said to be **invertible** if there exists a $n \times n$ matrix C such that:

$$AC = I \quad CA = I$$

Where I denotes the $n \times n$ identity matrix. In this case C is an **inverse** of A. Such an inverse is unique and is commonly denoted as A^{-1}. A matrix that is not invertible is commonly called a **singular matrix**, and an invertible matrix is called a **nonsingular matrix**.

THEOREM 4

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

The quantity $ad - bc$ is referred to as the **determinant** of a matrix, and we write:

$$\det A = ad - bc$$

The definition of the inverse of a matrix also allows us to find solutions to a matrix equation in another way, namely:

THEOREM 5

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

This theorem is easily proved by multiplying the matrix equation $A\mathbf{x} = \mathbf{b}$ with A^{-1}. The next theorem provides three useful facts about invertible matrices.

THEOREM 6

a. If A is an invertible matrix, then A^{-1} is invertible and:

$$(A^{-1})^{-1} = A$$

b. If A and B are invertible $n \times n$ matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in reverse order. That is:

$$(AB)^{-1} = B^{-1}A^{-1}$$

c. 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:

$$(A^T)^-1 = (A^{-1})^T$$

Elementary Matrices

An **elementary matrix** is one that can be 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 is created by performing the same elementary row operation on the identity matrix I_m.

Since row operations are reversible, for each elementary matrix E, that is produced by a row operation on I,there must be a reverse row operation changing E back to I. This reverse row operation can be represented by another elementary matrix F such that $FE = I$ and $EF = I$.

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

The following theorem is another way to visualize the inverse of a matrix:

THEOREM 7 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}.

The exact proof of this theorem is omitted here for brevity. It relies mainly on the insight that any elementary row operation can be represented as an elementary matrix E which always has an inverse. As A is essentially $(E_p...E_1)I$ then there exists a sequence of matrix multiplications $(E_p...E_1)^{-1}$ which transforms I into A. Thus A is $(E_p...E_1)^{-1}I = (E_p...E_1)^{-1}$ and is thus the inverse of an invertible matrix, meaning that it is invertible itself. From this theorem it is easy to create the following algorithm for finding the inverse of a matrix A:

Row reduce the augmented matrix $[A \quad I]$. If A is row equivalent to I, then $[A \quad I]$ is row equivalent to $[I \quad A^{-1}]$. Otherwise A does not have an inverse.

2.3 Characterizations of Invertible Matrices

In this section we relate the theory developed in chapter 1 to systems of n linear equations with n unknowns and thus square matrices. The main result is given in the next theorem.

THEOREM 8

The Invertible Matrix Theorem

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

a. A is an invertible matrix.

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

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 every $\mathbf{b}$ in $\mathbb{R}^n$.

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 $CA = I$.

k. There is an $n \times n$ matrix D such that $AD = I$.

l. A^T is an invertible matrix.

This theorem is named the *invertible matrix theorem* and its power lies in the fact that the negation of one statement allows one to include that the matrix is singular (i.e. not invertible) and thus all above statements are false for this matrix.

Invertible Linear Transformations

Recall that any matrix multiplication corresponds to the composition of linear mappings. When a matrix A is invertible, the equation $A^{-1}A\mathbf{x} = \mathbf{x}$ can be viewed as a statement about linear transformations as seen in figure 2.3.
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:

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

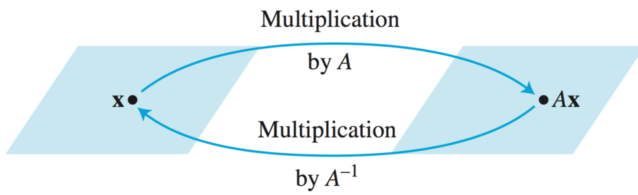

Figure 2.3: Illustration of how multiplication by A^{-1} transforms $A\mathbf{x}$ back to $\mathbf{x}$.

The next theorem shows that if such an S exists, it is unique and a linear transformation. We call S the **inverse** of T and denote it as T^{-1}.

THEOREM 9

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and let A be the standard matrix of T. Then T is invertible if and only if A is invertible. In that case, the linear transformation $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique function satisfying the previous 2 equations.

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

We start with the following definition:

DEFINITION

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

a. The zero vector is in H.

b. For each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.

c. For each $\mathbf{u}$ in H and scalar c, the vector $c\mathbf{u}$ is in H.

We can say that a *subspace* is closed by addition and scalar multiplication. A standard visualization of such a subspace in $\mathbb{R}^3$ is a plane through the origin. In this case Span$\{\mathbf{v}_1, \mathbf{v}_2\}$ is H.
Note how $\mathbb{R}^n$ is also a subspace as it satisfies all three properties defined above. Another special example is the subspace only containing the zero vector in $\mathbb{R}^n$. This set is called the **zero subspace** .

Column Space and Null Space of a Matrix

We can relate subspaces to matrices as follows:

DEFINITION

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

When a system of linear equations is written in the form $A\mathbf{x} = \mathbf{b}$, the column space of A is the set of all $\mathbf{b}$ for which the system has a solution (which is another way of writing above definition). Another definition is:

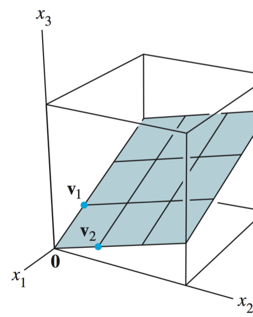

Figure 2.4: A subspace H as Span$\{\mathbf{v}_1, \mathbf{v}_2\}$

THEOREM 12

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

When A has n columns, the solutions of the equation $A\mathbf{x} = \mathbf{0}$ belong to $\mathbb{R}^n$, such that Nul A is a subspace of $\mathbb{R}^n$. Hence:

DEFINITION

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 $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

We can easily test if a vector is in Nul A by checking whether $A\mathbf{v} = \mathbf{0}$. This is an *implicit* definition of Nul A as a condition must be checked every time. In contrast, Col A is defined explicitly as we can make a linear combination of the columns of A and find a vector $\mathbf{b}$ in Col A. To create an explicit description of Nul A, we must write the solution of $A\mathbf{x} = \mathbf{0}$ in parametric vector form.

Basis for a subspace

A subspace of $\mathbb{R}^n$ typically contains infinitely many vectors. Hence it is sometimes more convenient to work with a 'smaller' part of a subspace, that consists of a finite number of vectors that span the subset. We can show that the smallest possible set of a subspace must be linearly independent.

DEFINITION

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

A great example of such a smaller part of a subset is that of the $n \times n$ identity matrix which forms the basis of $\mathbb{R}^n$. Its columns are denoted as $[\mathbf{e}_1, ..., \mathbf{e}_n]$. We like to call the set $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ the **standard basis** for $\mathbb{R}^n$.
The base of the subspace Nul A is easily found by solving the matrix equation $A\mathbf{x} = \mathbf{0}$ in parametric vector form.

The base of a subspace Col A is however, more complex to find. Let's start with the following matrix B:

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

Note how B is in the reduced echolon form. Also note that if we denote the columns of B as vectors that: $\mathbf{b}_3 = -3\mathbf{b}_1 + 2\mathbf{b}_2$ and $\mathbf{b}_4 = 5\mathbf{b}_1 + -\mathbf{b}_2$. As we are able to express all columns of B in terms of 3 vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$, we can express any linear combination of the columns of B as a linear combination of 3 vectors! Hence we have found a basis to Col B.

We can extend above discussion to the general matrix A. Note how the linear dependence relationship between the columns of A and its reduced echolon form B do not change, as both are the set of solutions to $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$, which share a solution set. This brings us to the following theorem:

THEOREM 13 The pivot columns of a matrix A form the basis for the column space of A.

The Row Space

If A is an $m \times n$ matrix, each row of A has n entries and thus can be seen as a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vectors of A is called the **row space** and is denoted by Row A. This also implies that Row A is a subspace of $\mathbb{R}^n$. Since each row of A is a column in A^T, we could also write Col A^T.

Note however, that we cannot find the basis of the row space of A by simply finding the pivot columns of A: row reduction of A changes the linear dependence relationships between the rows of A. However, we can make us of the following theorem:

THEOREM 14 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 for that of B.

PROOF If B is obtained by row operations on A, than the rows of B are linear combinations of the rows of A. Thus, any linear combination of the rows of B is actually also a linear combination of the rows of A, meaning that Row B is in Row A. Also, as row operations are reversible, the rows of A are linear combinations of the rows of B: Row A is also in Row B. That means the row spaces are the same. Then, if B is in echelon form, the nonzero rows of B cannot be a linear combination of those below it. Thus the nonzero rows of B form the basis of both Row A as Row B.

2.5 Dimension and Rank

Coordinate Systems

The main reason for our definition of a subspace H is that each vector in H can be written in only one way as the linear combination of the basis vectors. Suppose we have the subspace $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_p\}$, and a vector $\mathbf{x}$ which can be generated in two ways:

$$\mathbf{x} = c_1\mathbf{b}_1 + ... + c_p\mathbf{b}_p \quad \textbf{and} \quad \mathbf{x} = d_1\mathbf{b}_1 + ... + d_p\mathbf{b}_p \tag{2.1}$$

Then subtracting gives:

$$\mathbf{0} = \mathbf{x} - \mathbf{x} = (c_1 - d_1)\mathbf{b}_1 + ... + (c_p - d_p)\mathbf{b}_p \tag{2.2}$$

And since $\mathcal{B}$ is linearly independent, all weights must be zero. That is, $c_j = d_j$ for $1 \leq j \leq p$. Hence both representations are the same.

DEFINITION

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

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \tag{2.3}$$

is called the **coordinate vector of x (relative to** $\mathcal{B}$**)** or the $\mathcal{B}$**-coordinate vector of x**.

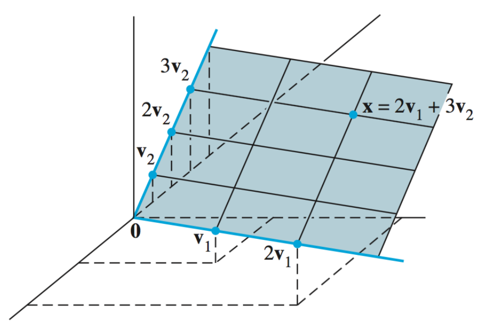

Figure 2.5: A coordinate system of a subspace H in $\mathbb{R}^3$

Figure 2.5 illustrates how such a basis $\mathcal{B}$ determines a coordinate system of subspace H. Note how the grid on the plane in the figure makes H 'look' like $\mathbb{R}^2$ while each vector is still in $\mathbb{R}^3$. We can see this as a mapping $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$, which is one-to-one between H and $\mathbb{R}^2$ and preserves linear combinations. We call such a correspondence *isomorphism*, and say that H is *isomorphic* to $\mathbb{R}^2$. In general, if $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_p\}$ is basis for H, then the mapping $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is a one-to-one correspondence that makes H look and act like $\mathbb{R}^p$.

The Dimension of a Subspace

DEFINITION

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

This makes sense, as in $\mathbb{R}^3$, a plane through $\mathbf{0}$ will be two-dimensional and a line through $\mathbf{0}$ is one-dimensional.

DEFINITION

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

Since the pivot columns form the basis for Col A, the number of pivot columns is simply the rank of matrix A.
We find the following useful connection between the dimensions of Nul A and Col A:

THEOREM 15

The Rank Theorem

If matrix A has n columns, then dim A + dim Nul $A = n$

The following theorem will be important in the applications of Linear Algebra:

THEOREM 16

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 that spans H is automatically a basis for H.

Recall the definition of the basis of a subspace: it is a set of vectors that both spans the subspace and is linearly independent. Let us take $\mathbb{R}^3$ as a subspace. If we would like to compose a set of 4 vectors, for which we know that they together span the subspace, we would already know that the coefficient matrix of the systems composed of these 4 vectors cannot contain a pivot position in every column, thus has a nontrivial solution to $A\mathbf{x} = \mathbf{0}$ and thus is linearly dependent. It does not satisfy the requirements of a subspace. However, removing the vector that is a linear combination of other vectors out of the set gives us a set of 3 elements and a set that spans the subspace, and corollary is linearly independent as there are no free variables ! The key to this theorem is that you cannot have one without other, always ending up a basis consisting of p elements for a p-dimensional subspace.

Rank and the Invertible Matrix Theorem

We can use the developed definitions on rank and dimensions to extend upon the Invertible Matrix Theorem introduced earlier:

THEOREM **The Invertible Matrix Theorem (continued)**

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.

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

n. Col $A = \mathbb{R}^n$

o. dim Col $A = n$

p. rank $A = n$

q. Nul $A = \{\mathbf{0}\}$

r. dim Nul $A = 0$

PROOF The first statement (m) is easy to prove. By, the basis theorem, any set of n vectors (i.e. the columns of A) that are linearly independent serves as a basis for $\mathbb{R}^n$. The columns of A are linearly independent by multiple previous statements of the Invertible Matrix Theorem: one is that an invertible matrix only has the trivial solution to $A\mathbf{x} = \mathbf{0}$ as there are no free variables. Hence, the columns are linearly independent and satisfy the basis theorem. Statement (n) is another way of stating statement (m). Statements (o) and (p) are equivalent and are a consequence of (m) and (n). Statement (q) and statement (r) are equivalent and follow from the rank theorem.

Determinants

3

3.1 Introduction to Determinants

Recall from section 2.2 that a 2×2 matrix is invertible if and only if its determinant is nonzero. To extend this useful fact to larger matrices, we must define the determinant for an $n \times n$ matrix.
Consider 3×3 matrix $A = [a_{ij}]$ with $a_{11} \neq 0$. By use of the appropriate elementary row operations it is always possible to find that:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{23} & a_{33} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{13}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

where

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Since A must be invertible, it must be row equivalent to the 3×3 identity matrix and hence Δ must be nonzero. We call Δ the **determinant** of the 3×3 matrix A. We can also group the terms and write:

$$\Delta = a_{11} \cdot \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \cdot \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \cdot \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

Note how the determinant of the smaller 2×2 matrices is computed as discussed in section 2.2. For brevity we can also write:

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

Where A_{ij} can be obtained by deleting the ith row and the jth column of matrix A. We can extend the example above to a more general definition of the **determinant** of an $n \times n$ matrix.

DEFINITION

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$, is the sum of n terms of the form $\pm a_{1j} \det A_{1j}$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, ..., a_{1n}$ are from the first row of A. In symbols,

$$\begin{aligned} \det A &= a_{11}\det A_{11} - a_{12}\det A_{12} + ... + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{ij} \end{aligned}$$

If we define the (i,j)**-cofactor** of A to be the number C_{ij} given by

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

Then

$$\det A = a_{11}C_{11} + a_{12}C_{12} + ... + a_{1n}C_{1n}$$

Which is called the **cofactor expansion across the first row** of A. The following theorem builds upon this definition.

THEOREM 1

The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the ith row using the cofactor definition is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + ... + a_{in}C_{in}$$

The cofactor expansion down the jth column is

$$\det A = A_{ij}C_{1j} + a_{2j}C_{2j} + ... + a_{nj}C_{nj}$$

Above theorem is very useful for computing the determinant of a matrix containing many zeros. By choice of the row or column with the least nonzero entries the computation of the determined can be shortened.

THEOREM 2

If A is a triangular matrix, then det A is the product of the entries on the main diagonal of A.

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

In case of a triangular matrix, such as the simple example above, a cofactor expansion across the first column leads to another triangular 3×3 matrix whose determinant can be computed by a cofactor expansion across its first column, leading to a 2×2 matrix etc. The result is always the product of the main diagonal's entries.

3.2 Properties of Determinants

THEOREM 3

Row Operations

Let A be a square matrix.

a. If a multiple of one row of A is added to another row to produce a matrix B, then det B = det A.

b. If two rows of A are interchanged to produce B, then det $B = -$det A.

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

Suppose square matrix A has been reduced to an echelon form U by row replacements and row interchanges. If we denote the number of interchanges as r, then by use of theorem 3 we find that:

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

Since U is in echelon form, it must be triangular, and thus det U is simply the product of its diagonal entries $u_{11}, ..., u_{nn}$. If A is indeed invertible, all diagonal entries of U must be nonzero. Otherwise, at least one entry is zero and the product is also zero. Thus for the determinant of A:

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

Above generalization allows us to conclude that:

THEOREM 4

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

Column Operations

Recall from section 2.8 the discussion on row spaces and the accompanying operations on the columns of a matrix, analogous to row operations. We can show the following:

THEOREM 5

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

Above theorem is simple to prove. For an $n \times n$ matrix A, the cofactor of a_{1j} equals the cofactor of a_{i1} in A^T. Hence the cofactor expansion along the first row of A equals that along the first column of A^T, and thus their determinants are equal.

Determinants and Matrix Products

The following theorem on the multiplicative property of determinants shall proof to be useful later on.

THEOREM 6

Multiplicative Property

If A and B are $n \times n$ matrices, then det $AB = (\det A)(\det B)$.

A Linearity Property of the Determinant Function

For an $n \times n$ matrix A, we can consider det$A$ as a function of the n column vectors in A. We can show that if all columns in A except for one are held fixed, then det$A$ is a *linear function* of this one vector variable. Suppose that we chose the jth column of A to vary, then:

$$A = [\mathbf{a}_1 \quad ...\mathbf{a}_{j-1} \quad \mathbf{x} \quad \mathbf{a}_{j+1} \quad ... \quad \mathbf{a}_n]$$

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

$$T(\mathbf{x}) = \det[\mathbf{a}_1 \quad ...\mathbf{a}_{j-1} \quad \mathbf{x} \quad \mathbf{a}_{j+1} \quad ... \quad \mathbf{a}_n]$$

Then we can prove that,

$$T(c\mathbf{x}) = cT(\mathbf{x}) \quad \text{for all scalars } c \text{ and all } \mathbf{x} \text{ in } \mathbb{R}^n$$
$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ in } \mathbb{R}^n$$

And thus is a linear transformation.

3.3 Cramer's rule, Volume and Linear Transformations

Cramer's Rule

Let A be an $n \times n$ matrix and $\mathbf{b}$ any vector in $\mathbb{R}^n$, then $A_i(\mathbf{b})$ is the matrix obtained by replacing the ith column in A by vector $\mathbf{b}$.

$$A_i(\mathbf{b}) = [\mathbf{a}_1 \quad ... \quad \mathbf{b} \quad ... \quad \mathbf{a}_n]$$

THEOREM 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

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \quad i = 1, 2, ..., n$$

Denote the columns of A by $\mathbf{a}_1, ..., \mathbf{a}_n$ and the columns of the $n \times n$ identity matrix I by $\mathbf{e}_1, ..., \mathbf{e}_n$. If $A\mathbf{x} = \mathbf{b}$, the definition of matrix multiplication shows that

$$A \cdot I_i(\mathbf{x}) = A[\mathbf{e}_1 \quad .... \quad \mathbf{x} \quad \mathbf{e}_n] = [A\mathbf{e}_1 \quad ...\mathbf{x} \quad .... \quad ...A\mathbf{e}_n] = [\mathbf{a}_1 \quad ... \quad \mathbf{b} \quad ... \quad \mathbf{a}_n] = A_i(\mathbf{b})$$

The use of above computation becomes clear when we use the multiplicative property of determinants and find that:

$$(\det A)(\det I_i(\mathbf{x})) = \det A_i(\mathbf{b})$$

The second determinant on the left is simply x_i, as cofactor expansion along the ith row always leads to the multiplication of x_i and another identity matrix of size $(n-1) \times (n-1)$ and thus is equal to x_i. Hence $(\det A) \cdot x_i = \det A_i(\mathbf{b})$, proving above theorem.

A Formula for A^{-1}

By use of Cramer's rule we can easily find a general formula for the inverse of an $n \times n$ matrix A. Note how the jth column of A^{-1} is vector $\mathbf{x}$ that satisfies

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

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

$$\{(i,j) - \text{entry of} A^{-1}\} = x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A}$$

Recall from previous sections that A_{ji} denotes the submatrix of A formed by deleting row j and column I. A cofactor expansion down column i of $A_I(\mathbf{e}_j)$ shows that

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

where C_{ji} is a cofactor of A. We can then find that the (i, j)-entry of A^{-1} is the cofactor C_{ji} divided by $\det A$. Thus

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$$

The matrix on the right side of above equation is called the **adjugant** of A, denoted by adj A. Above derivation can then be simply stated as:

THEOREM 8

An Inverse Formula

Let A be an invertible $n \times n$ matrix, then: $\mathrm{A}_{-1} = \frac{1}{\det A}\mathrm{adj} A$

Determinants as Area or Volume

We can now go on and give geometric interpretations of the determinant of a matrix.

THEOREM 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 $|\det A|$.

PROOF Recall that row interchanges and addition of the multiple of one row to the other do not change the *absolute* value of the determinant. These two operations suffice to change any matrix A into a diagonal matrix, such that the determinant is simply the product of the diagonal entries and thus the area or volume of the shape defined by the columns of A.

We can also show that column operations do not change the parallelogram or parallelepiped at all. The proof is simple and is geometrically intuitive, thus omitted for brevity:

THEOREM 10

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be nonzero vectors. Then for any scalar c the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ equals the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$

Linear Transformations

Determinants can be used to describe an important geometric property of linear transformations in the plane and in $\mathbb{R}^3$. If T is a linear transformation and the set S is the domain of T, let $T(S)$ denote the set of images of points in S.

THEOREM 11

Let $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\{\text{area of } T(S)\} = |\text{det}A| \cdot \{\text{area of } S\}$$

If T is determined by a 3×3 matrix A, and if S is a paralellepiped in $\mathbb{R}^3$, then

$$\{\text{volume of } T(S)\} = |\text{det}A| \cdot \{\text{volume of } S\}$$

PROOF For our proof, we consider a transformation T determined by 2×2 matrix A. The set S bounded by a parallelogram at the origin in $\mathbb{R}^2$ has the form:

$$S = \{s_1\mathbf{b}_1 + s_2\mathbf{b}_2 : 0 \leq s_1 \leq 1, 0 \leq s_2 \leq 1\}$$

As the transformation T is linear as it is a matrix transformation we know that every point in the set $T(S)$ is of the form:

$$T(s_1\mathbf{b}_1 + s_2\mathbf{b}_2) = s_1T(b_1) + s_2T(b_2) = s_1A\mathbf{b}_1 + s_2A\mathbf{b}_2$$

It follows that the set $T(S)$ is bounded by the parallelogram determined by the columns of $[A\mathbf{b}_1 \quad A\mathbf{b}_2]$, which can be written as AB where $B = [\mathbf{b}_1 \quad \mathbf{b}_2]$. Then we find that:

$$\{\text{area of } T(S)\} = |\text{det}AB| = |\text{det}A| \cdot |\text{det}B| = |\text{det}A| \cdot \{\text{area of } S\}$$

And above theorem is proven. Note how any translation of the set S by a vector $\mathbf{p}$ simply results in the translation of $T(S)$ by $T(\mathbf{p})$ and is thus not influential on the size of the resulting set.

Orthogonality and Least Squares 4

4.1 Inner Product, Length, and Orthogonality

We can extend the well-known geometrical concepts of length, distance and perpendicularity for $\mathbb{R}^2$ and $\mathbb{R}^3$ to that of $\mathbb{R}^n$.

The Inner Product

The **inner product** of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is written as:

DEFINITION

$$[u_1 \quad u_2 \quad \dots \quad u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \dots + u_nv_n$$

Note how this is simply a matrix multiplication of the form $\mathbf{u}^T\mathbf{v}$. Using this fact we can easily deduce the following properties:

THEOREM 1

Let $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and c be scalar, then

a. $\mathbf{u} \cdot \mathbf{v}=\mathbf{v} \cdot \mathbf{u}$

b. $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$

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

d. $\mathbf{u} \cdot \mathbf{u} \geq 0, \quad$ and$\mathbf{u} \cdot \mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u} = \mathbf{0}$

Properties (b) and (c) can be combined several times to produce the following useful fact with regards to linear combinations:

$$(c_1\mathbf{u}_1 + \dots + c_p\mathbf{u}_p) \cdot \mathbf{w} = c_1(\mathbf{u}_1 \cdot \mathbf{w}) + \dots + c_p(\mathbf{u}_p \cdot \mathbf{w})$$

The Length of a Vector

If $\mathbf{v}$ is in $\mathbb{R}^n$ then the square root of $\mathbf{v} \cdot \mathbf{v}$ is defined as it is never negative.

DEFINITION

The **length** (or **norm**) of $\mathbf{v}$ is n the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + ... + v_n^2}, \quad \text{and } \|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$$

Note how this coincides with the standard notion of length in $\mathbb{R}^2$ and $\mathbb{R}^3$.

A vector whose length is 1 is called a **unit vector**. If we divide a nonzero vector $\mathbf{v}$ by its own length - that is multiply by $1/\|\mathbf{v}\|$ - we obtain a unit vector $\mathbf{u}$ because the length of $\mathbf{u}$ is equal to $\|\mathbf{v}\|(1/\|\mathbf{v}\|)$. The process of creating $\mathbf{u}$ from $\mathbf{v}$ is often called **normalizing** $\mathbf{v}$.

Distance in $\mathbb{R}^n$

DEFINITION

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

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

Above definition coincides with the usual formulas for the Eucledian distance between points in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Orthogonal Vectors

Note how the vectors in figure 4.1 can only be perpendicular if the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u}$ to $-\mathbf{v}$. Note how this is the same as asking the square of

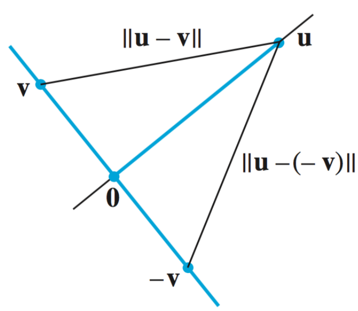

Figure 4.1: Two perpendicular vectors $\mathbf{u}$ and $\mathbf{v}$

the distances to be the same. Using this fact it is possible to rewrite this equality as:

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u} - (-\mathbf{v})\|^2$$
$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u}\cdot\mathbf{v}$$
$$-\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{v}$$

Which can only be satisfied if $\mathbf{u}\cdot\mathbf{v} = 0$, bringing us to the following definition:

DEFINITION

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

By the derivation above we also arrive at the following (familiar) theorem:

THEOREM 2

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}+\mathbf{v})\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$

Orthogonal Complements

If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, 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 $W^\perp$. We introduce the following two facts about $W^\perp$.

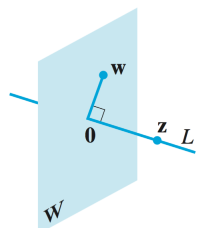

Figure 4.2: Illustration of vector $\mathbf{z}$ orthogonal to subspace W

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

1. Every vector in W can be written as a linear combination of vectors in basis of W, note how weights of linear combination do not affect whether the linear combination and vector $\mathbf{x}$ are orthogonal as the outcome is always zero due to the orthogonality of the basis vectors in $\mathbf{x}$.
2. Inner products and thus the question of orthogonality is only defined for vectors with the same amount of entries. Hence any orthogonal vector $\mathbf{x}$ must be in the same $\mathbb{R}^n$ as the vectors in W.

THEOREM 3

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:

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

The row-column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul A, then $\mathbf{x}$ is orthogonal to each row of A (if we treat each row as a vector in $\mathbb{R}^n$). Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A. This statement also holds for A^T and hence the null space of A^T is the orthogonal complement to Col A as Row A^T = Col A.

4.2 Orthogonal Sets

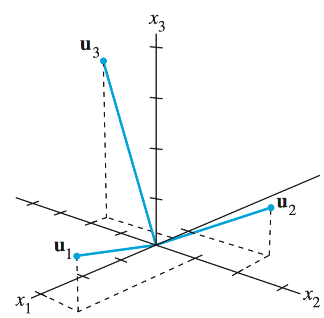

Figure 4.3: Orthogonal set in $\mathbb{R}^3$

A set of vectors $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$. An orthogonal set is shown in figure 4.3.

THEOREM 4

> If $S = \{\mathbf{u}_1, ..., \mathbf{u}_p\}$ 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.

Albeit we can argue about the geometrical intuition of above theorem (all vectors are orthogonal and hence can be seen as a separate orthogonal coordinate system with the set's vectors as basis vectors), we can proof it mathematically. If there is a nontrivial solution such that $\mathbf{0} = c_1\mathbf{u}_1 + ... + c_p\mathbf{u}_p$ for some weights $c_1, ..., c_p$ not all zero, then the following must hold:

$$0 = \mathbf{0} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + ... + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

because $\mathbf{u}_1$ is orthogonal to all vectors in the set S all other terms cancel. As $\mathbf{u}_1$ is nonzero, c_1 must be zero. We can repeat above logic for all vectors in S and find that all weights must be zero, meaning that there is no nontrivial solution and the set is linearly independent. We now arrive at the following definition

DEFINITION

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

Now we can find why an orthogonal basis to a subspace W is convenient:

THEOREM 5 Let $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + ... + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} \quad (j = 1, ..., p)$$

PROOF As in the preceding proof of the linear independence of an orthogonal set, we find that:

$$\mathbf{y} \cdot \mathbf{u} = (c_1\mathbf{u}_1 + ... + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

Since $\mathbf{u}_1 \cdot \mathbf{u}_1$ is nonzero, the equation can be solved for c_1. Geometrically speaking, we are finding that part of vector $\mathbf{y}$ that lies on the line L formed by $c\mathbf{u}_1$ and then dividing this part of $\mathbf{y}$ by the length of the vector $\mathbf{u}_1$ to find the scalar weight c. Note how this computation is more convenient than solving a system of linear combinations as used in previous chapters to find the weights of a given linear combination.

An Orthogonal Projection

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

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

Where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See figure 4.4. Given any scalar α we know that $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$ to satisfy the above equation. Then $\mathbf{y} - \mathbf{y}'$ is

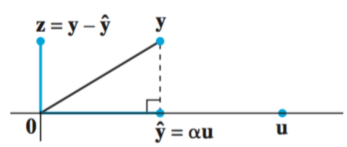

Figure 4.4: Orthogonal projection of vector $\mathbf{y}$ on line spanned by vector $\mathbf{u}$

orthogonal to $\mathbf{u}$ if and only if

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

That is, the decomposition of $\mathbf{y}$ is satisfied with $\mathbf{z}$ orthogonal to $\mathbf{u}$ if and only if $\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$ and $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$. The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto u** and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**. The geometrical interpretation of finding the weights of a linear combination as explained in the previous section is once again applicable here.
If c is any nonzero scalar and if $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$ then the orthogonal projection $\hat{\mathbf{y}}$ does not change. We can say that the projection is determined by subspace L spanned by $\mathbf{u}$. Sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L\mathbf{y}$ and is called the **orthogonal projection of y onto** L.

Orthonormal Sets

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

THEOREM 6

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

PROOF Let us suppose that U has three columns where each is a vector in $\mathbb{R}^m$. Or, $U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3]$. Then we can compute:

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix}$$

The entries in the matrix on the right are all scalars resulting from inner products using the transpose notation. The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0$$

And the columns of U all have unit length, as required by the definition of orthonormal columns, if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \quad \mathbf{u}_2^T\mathbf{u}_2 = 1, \quad \mathbf{u}_3^T\mathbf{u}_3 = 1$$

And hence the theorem is proved. More over, we find that

THEOREM 7

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 $\mathbf{x} \cdot \mathbf{y} = 0$

Above theorems are particularly useful when applied to *square* matrices. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$: it has orthonormal columns. More on these type of matrices will follow.

4.3 Orthogonal Projections

This section covers the analogue to the orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin for that in $\mathbb{R}^n$. Given a vector $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$ there is a vector $\hat{\mathbf{y}}$ in W such that $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W, and $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$.

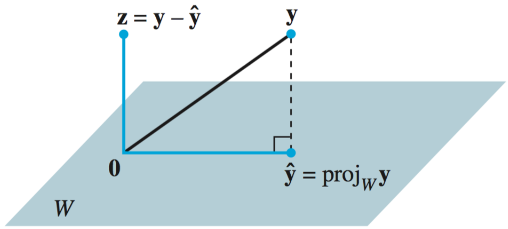

Figure 4.5: Geometric interpretation of orthogonal projection of $\mathbf{y}$ onto subspace W.

THEOREM 8

The Orthogonal Decomposition Theorem

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 $\{\mathbf{u}_1, ..., \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + ... + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p$$

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

The vector $\hat{\mathbf{y}}$ in above theorem is called the **orthogonal projection of y onto** W and is often written as $\text{proj}_W\mathbf{y}$.

Properties of Orthogonal Projections

THEOREM 9

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be 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

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

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

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

If the basis to subspace W is an orthonormal set:

THEOREM 10

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

$$\text{proj}_W\mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + ... + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p$$

if $U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad ... \quad \mathbf{u}_p]$, then

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

PROOF The first equation follows directly from the orthogonal decomposition theorem as $\mathbf{u}_j \cdot \mathbf{u}_j = 1$ for any vector $\mathbf{u}_j$ in an orthonormal set. The second equation is another way of writing the first. Note how the first equation indicates that $\text{proj}_W \mathbf{y}$ is essentially a linear combination of the vectors in the orthonormal basis with weights $\mathbf{y} \cdot \mathbf{u}_j$. The equation $U^T\mathbf{y}$ results in a vector containing the result of all inner products of $\mathbf{y}$ with the rows of U^T, and hence the columns of U. Thus, if this resulting vector is used to form a linear combination with the columns of U, we arrive back at the first equation.

4.4 The Gram-Schmidt Process

THEOREM 11

The Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, ..., \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_2 &= \mathbf{x}_3 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - ... - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

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

$$\text{Span}\{\mathbf{v}_1, ..., \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, ..., \mathbf{x}_k\} \quad \text{for } 1 \leq k \leq p$$

Suppose for some $k < p$ we have constructed $\mathbf{v}_1, ..., \mathbf{v}_k$ so that $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is an orthogonal basis for W_k (note that this always possible as we start out with $\mathbf{v}_1 = \mathbf{x}_1$). If we define the next vector to be added to the orthogonal set as:

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{proj}_{W_k}\mathbf{x}_{k+1}$$

By the orthogonal decomposition theorem we konw that $\mathbf{v}_{k+1}$ is orthogonal to W_k. We also know that $\text{proj}_{W_k}\mathbf{x}_{k+1}$ is in W_k and is thus also in W_k. Now we know that $\mathbf{v}_{k+1}$ is in W_{k+1} as the addition of two vectors in a given subspace always leads to another vector in that subspace (i.e. closed by addition and subtraction). We have now proven that $\mathbf{v}_1, ..., \mathbf{v}_{k+1}$ is an orthogonal set of nonzero vectors (making them all linearly independent) in the $(k+1)$-dimensional subspace W_{k+1}. By the basis theorem it follows that this set is a basis, and thus an orthogonal basis to W_{k+1}.

Note how we can use the Gram-Schmidt process to find an orthonormal basis by normalizing the orthogonal basis resulting from the algorithm.

4.5 Least-Squares Problems

When a solution for an equation $A\mathbf{x} = \mathbf{b}$ is demanded, yet the system is consistent (i.e. no solution exists), the best one can do is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$. The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible.

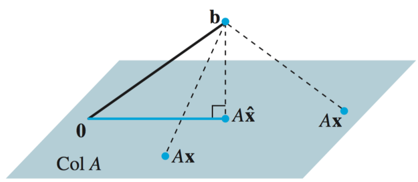

Figure 4.6: Vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for all other $\mathbf{x}$

DEFINITION

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 $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

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

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

Solution of the General Least-Squares Problem

We can apply the best approximation theorem to the subspace Col A in order to find an approximation to the vector $\mathbf{b}$. Let

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

Because $\hat{\mathbf{b}}$ is in the column space of A, there must be an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that:

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

Hence if a vector $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ satisfies above equation, it is a least-squares solution to $A\mathbf{x} = \mathbf{b}$. See figure 4.7 for a geometrical interpretation. Note that by the Orthogonal Decomposition Theorem the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to the column space of A. Then, if we use the fact that $\mathbf{a}_j \cdot \mathbf{u} = \mathbf{a}_j^T\mathbf{u}$, where $\mathbf{a}_j$ is a column of A, we can state that:

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$$

Which is simply matrix notation for the fact that the vector $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to the column space of A. Above equation also brings us to the fact that:

$$A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$$

For every least-squares solution to $A\mathbf{x} = \mathbf{b}$. Above matrix equation represents a system called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.

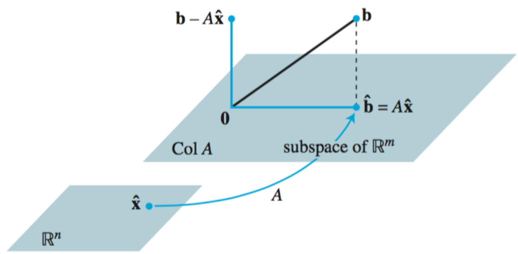

Figure 4.7: A least squares solution

THEOREM 13 The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.

Solving the augmented matrix of the normal equations leads to a general expression of the set of least-squares solutions.

The next theorem gives useful criteria for determining when there is only one or more least-squares solution of $A\mathbf{x} = \mathbf{b}$ (note that the projection of $\mathbf{b}$ on Col A is always unique).

THEOREM 14 Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix $A^T A$ is invertible.

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

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

PROOF If statement (b) is true, we know that there are no free variables in any equation $A\mathbf{x} = \mathbf{v}$ and there is one unique solution for every $\mathbf{v}$ in Col A, hence also for the least-squares solution of $A\mathbf{x} = \mathbf{b}$ and thus statement (a) is true. Statement (c) follows from the fact that there is an unique solution $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.
If a least-squares solution $\hat{\mathbf{x}}$ is used to produce an approximation $A\hat{\mathbf{x}}$ to $\mathbf{b}$, the distance between $A\hat{\mathbf{x}}$ and $\mathbf{b}$ is called the **least-squares error** of this approximation.

4.6 Applications to Linear Models

In this section, we denote the matrix equation $A\mathbf{x} = \mathbf{b}$ as $X\boldsymbol{\beta} = \mathbf{y}$ and refer to X as the **design matrix**, $\boldsymbol{\beta}$ as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

Least-squares Lines

The simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), ..., (x_n, y_n)$ that, when graphed, seem to lie close to a line. We would like to find parameters β_0 and β_1 that make the line as close as possible to the measured data. Suppose we have chosen a combination of β_0 and β_1. Then for each data point (x_i, y_i) there is a corresponding point $(x_i, \beta_0 + \beta_1 x_i)$ on the line with the same x-coordinate. We call y_i the *observed value* of y and $\beta_0 + \beta_1 x_i$ the *predicted value* of y. The difference between the two is called a *residual*. The **least-squares line** is the line

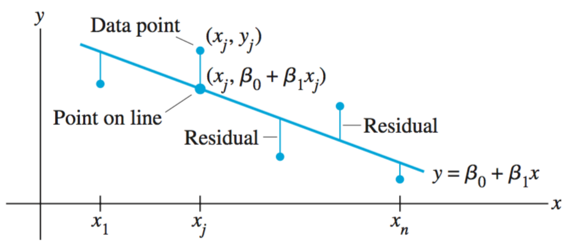

Figure 4.8: Line fit to experimental data.

$y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. It is also called a **line of regression of y on x**, because any error is assumed to be only in the y-coordinates. The coeficients β_0, β_1 of the line are called **regression coefficients**. Note how, if all data points were satisfied by our choice of regression coefficients that the equation:

$$\beta_0 + \beta_1 x_i = y_j$$

would be satisfied for $1 \leq i \leq n$. We can write this system of linear equations as:

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Note how computing the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ is equivalent to finding the line that determines the least-squares line in figure 4.8. Common practice is to compute the average $\bar{x}$ of the original x-values and form a new variable $x^* = x - \bar{x}*$. This is referred to as the **mean-deviation form**. In this case, the columns of design matrix X will be orthogonal and simplify the computation of the least-squares solutions.

The General Linear Model

It is also possible to fit data points with something other than a straight line. The matrix equation used for the linear model is still $X\boldsymbol{\beta} = \mathbf{y}$ but the form of X may change. Also, a **residual vector $\boldsymbol{\epsilon}$** is introduced and defined as $\boldsymbol{\epsilon} = \mathbf{y} - X\boldsymbol{\beta}$ such that

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

A data fit by curves has the general form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + ... + \beta_n f_n(x)$$

Which can be extended to the familiar matrix equation $X\boldsymbol{\beta} = \mathbf{y}$. Note that $f_i(x)$ may be any function of x, and is still used in a linear model as the coefficients β_i are unknown.

Eigenvalues and Eigenvectors

5

5.1 Eigenvectors and Eigenvalues

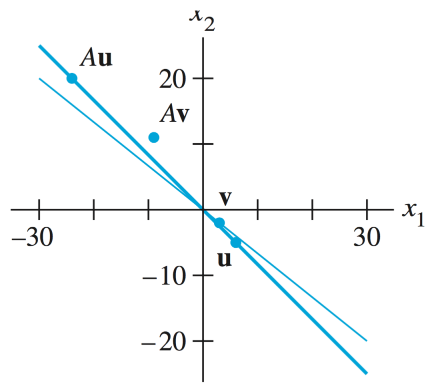

Figure 5.1: Note how $\mathbf{u}$ is an eigenvector of A as $A\mathbf{u} = -4\mathbf{u}$. Vector $\mathbf{v}$ is not an eigenvector

DEFINITION

An **eigenvector** of an $n \times n$ matrix A is the 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.

Note that we can also say that an eigenvector must satisfy:

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

For $\mathbf{x} \neq \mathbf{0}$. Hence:

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

Must be satisfied. Thus the set of all eigenvectors for a given λ and matrix A is the null space of $A - \lambda I$. This subspace of $\mathbb{R}^n$ is denoted as the **eigenspace**.

THEOREM 1

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

PROOF Consider the case of a 3×3 matrix A. If A is upper triangular, then:

$$A - \lambda I = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix}$$

Note that if A would have eigenvalues, then $A - \lambda I$ must have one or more free variables as otherwise the equation would only be satisfied for the zero vector. Hence, one of the diagonal entries of $A - \lambda I$ must be zero (i.e. the matrix does not have a pivot position in every column). Thus, for each diagonal entry of A, there exists an eigenvalue λ and accompanying eigenvectors.

THEOREM 2

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

Eigenvectors and Difference Equations

A first-order difference equation is of the form:

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad (k = 0, 1, 2, ...)$$

If A is $n \times n$, then this a *recursive* (i.e. depending on the preceding term) description of a sequence $\{\mathbf{x}_k\}$ in $\mathbb{R}^n$. We define a **solution** to be an explicit description of $\{\mathbf{x}_k\}$ which does not depend directly on A or the preceding terms in the sequence other than the original term $\mathbf{x}_0$. The simplest way to such a solution is

$$\mathbf{x}_k = \lambda^k \mathbf{x}_0$$

Where $\mathbf{x}_0$ is an eigenvector and λ is its corresponding eigenvalue. Note that this holds since

$$A\mathbf{x}_k = A(\lambda^k \mathbf{x}_0) = \lambda^k (A\mathbf{x}_0) = \lambda^k (\lambda \mathbf{x}_0) = \lambda^{k+1} \mathbf{x}_0 = \mathbf{x}_{k+1}$$

This way of forming a solution makes sense, as we find an eigenvector with respective eigenvalue λ such that $A\mathbf{x}_0 = \lambda \mathbf{x}_0$ which is also in the respective eigenspace as it is simply a multiple of $\mathbf{x}_0$! Hence we can once again use scalar multiplication by λ to represent $A\mathbf{x}$, which may continue infinitely.

5.2 The Characteristic Equation

We start this section with an extension of the Invertible Matrix Theorem. The use of this extension will become evident soon after.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then A is invertible if and only if:

a. The number 0 is *not* an eigenvalue of A.

b. The determinant of A is *not* zero.

The Characteristic Equation

The following scalar equation allows one to directly find the eigenvalues of a square matrix A.

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

$$\det(A - \lambda I) = 0$$

This makes sense, since a determinant equal to zero indicates that the matrix is not invertible which indicates that it does contain free variables, which means that there are nontrivial solutions to $(A - \lambda I)\mathbf{x} = \mathbf{0}$) and hence means that λ is an eigenvalue of A.

The n-degree polynomial in λ resulting from the computation of the determinant of an $n \times n$ matrix is called the **characteristic polynomial** of A. The eigenvalues of A are the roots of this polynomial. The **multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation (i.e. how many times the scalar occurs as root).

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, equivalently, $A = PBP^{-1}$. We can also write $Q = P^{-1}$ and thus $Q^{-1}BQ = A$. Then A and B **are similar**. If we change A into $P^{-1}AP$ it is called a **similarity transformation**.

THEOREM 4 **I**

f $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

PROOF If $B = P^{-1}AP$, then

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(AP - \lambda P) = P^{-1}(A - \lambda I)P$$

Then using the multiplicative property of determinants:

$$\det(B - \lambda I) = \det[P^{-1}(A - \lambda I)P] = \det(P^{-1}) \cdot \det(A - \lambda I) \cdot \det(P)$$

Since $\det(P^{-1}) \cdot \det(P) = \det(P^{-1}P) = \det(I) = 1$, we see that $\det(B - \lambda I) = \det(A - \lambda I)$ and hence they have the same characteristic polynomial.

5.3 Diagonalization

This section discusses the useful concept of a diagonal matrix. If we have such a diagonal matrix D

$$D = \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \end{bmatrix}$$

with d_{11} and d_{22} nonzero, we have that:

$$D^k = \begin{bmatrix} d_{11}^k & 0 \\ 0 & d_{22}^k \end{bmatrix}$$

If we recall from section 5.2 that a matrix A similar to D can be written as $A = PDP^{-1}$ with P invertible, then we can easily find an expression to the k-th power of A. We call such a matrix A that is similar to a diagonal matrix, **diagonalizable**.

THEOREM 5

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, repsectively, to the eigenvectors in P.

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

PROOF If P is an $n \times n$ matrix with columns $\mathbf{v}_1, ..., \mathbf{v}_n$, and if D is any diagonal matrix with diagonal entries $\lambda_1, ..., \lambda_n$, then

$$AP = A[\mathbf{v}_1 \quad \mathbf{v}_2 \quad ... \quad \mathbf{v}_n] = [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad ... \quad A\mathbf{v}_n]$$

and

$$PD = P \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & ... & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{v}_1 \quad \lambda_2 \mathbf{v}_2 \quad ... \quad \lambda_n \mathbf{v}_n]$$

Now suppose that A is indeed diagonalizable and $A = PDP^{-1}$. Then right-multipliying this equality with matrix P gives that $AP = PD$, implying that:

$$[A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad ... \quad A\mathbf{v}_n] = [\lambda_1 \mathbf{v}_1 \quad \lambda_2 \mathbf{v}_2 \quad ... \quad \lambda_n \mathbf{v}_n]$$

If we equate columns we find that:

$$A\mathbf{v}_j = \lambda_j \mathbf{v}_j$$

for $1 \leq j \leq n$. As P is invertible, all columns are linearly independent and thus are nonzero. Hence, the equation above implies that $\lambda_1, \lambda_2, ..., \lambda_n$ are indeed eigenvalues with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, which proves the theorem above.

The following theorem is a more brief condition than the previous theorem to see if a matrix is indeed diagonalizable.

THEOREM 6

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

PROOF Let $\lambda_1, ..., \lambda_n$ be distinct eigenvalues of a $n \times n$ matrix A, then $\mathbf{v}_1, ..., \mathbf{v}_n$ are the corresponding eigenvectors. By theorem 2 in section 5.1, these eigenvectors are linearly independent, hence there exists an invertible matrix P composed of n eigenvectors thus satisfying the requirement for A to be diagonalizable.

Matrices Whose Eigenvalues Are Not Distinct

When A is diagonalizable but has fewer than n distinct eigenvalues and thus fewer than n linearly independent eigenvectors to construct matrix P, it is still possible to build P in a way that makes P automatically invertible.

THEOREM 7

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

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 eigenspaces equals n, and this happens if and only if (i) the characteristic polynomial factors completely into linear factors and (ii) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for th eigenspace corresponding to λ_k for each K, then the total collection of vectors in the sets $\lfloor_1, ..., \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

Note how (b) and (c) follow from (a), it basically states that certain distinct eigenvalues may have a basis consisting of more than one eigenvectors (which are linearly independent) which thus can be used to form the invertible matrix P, if the total amount of linearly independent eigenvectors is n or greater.

Linear Transformations on $\mathbb{R}^n$

Sometimes, it is more convenient to represent a transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ with respect to a different coordinate system in $\mathbb{R}^n$. Note how we can represent a linear transformation as $\mathbf{x} \mapsto A\mathbf{x}$. We would like to find a manner in which we can represent a linear transformation of a vector $\mathbf{u}$ defined relative to a basis $\mathcal{B}$ in $\mathbb{R}^n$ that is represented as $\mathbf{u} \mapsto D\mathbf{u}$.

THEOREM 8

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 $\mathbf{x} \mapsto A\mathbf{x}$.

PROOF Denote the columns of$P$ by $\mathbf{b}_1, ..., \mathbf{b}_n$, so that $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ and $P = [\mathbf{b}_1, ..., \mathbf{b}_n]$. In this case, we can represent the coordinate transformation relative to basis $\mathcal{B}$ as:

$$P[\mathbf{x}]_\mathcal{B} = \mathbf{x} \quad \text{and} \quad [\mathbf{x}]_\mathcal{B} = P^{-1}\mathbf{x}$$

Note how this makes sense as $P[\mathbf{x}]_\mathcal{B}$ is a linear combination of the columns of P, which is basis $\mathcal{B}$, with weights of coordinate vector $[\mathbf{x}]_\mathcal{B}$, which is relative to basis $\mathcal{B}$ and thus results in the 'original' vector $\mathbf{x}$.

If $T(\mathbf{x}) = A\mathbf{x}$ for $\mathbf{x}$ in $\mathbb{R}^n$, then

$$\begin{aligned}
[T]_{\mathcal{B}} &= [[T(\mathbf{b}_1)]_{\mathcal{B}} \quad ... \quad ...[T(\mathbf{b}_n)]_{\mathcal{B}}] \\
&= [[A\mathbf{b}_1]_{\mathcal{B}}, \quad ... \quad [A\mathbf{b}_n]_{\mathcal{B}}] \\
&= [P^{-1}A\mathbf{b}_1 \quad ... \quad P^{-1}A\mathbf{b}_n] \\
&= P^{-1}A[\mathbf{b}_1 \quad ... \quad \mathbf{b}_n] \\
&= P^{-1}AP
\end{aligned}$$

Since $A = PDP^{-1}$, we have $[T]_{\mathcal{B}} = P^{-1}AP = D$, which proves above theorem.

5.4 Complex Eigenvalues

Since the characteristic equation of an $n \times n$ matrix involves a polynomial of degree n, the equation always has exactly n roots, counting multiplicities, *provided that possibly complex roots are included.* The key is to let A also act on the space $\mathbb{C}^n$ of n-tuples of complex numbers.

The matrix eigenvalue-eigenvector theory already developed for $\mathbb{R}^n$ applies equally well to $\mathbb{C}^n$. So a complex scalar λ if and only if there is a nonzero vector $\mathbf{x}$ in $\mathbb{C}^n$ such that $A\mathbf{x} = \mathbf{x}$.

Real and Imaginary parts of Vectors

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

For example, if $\mathbf{x} = \begin{bmatrix} 3-i \\ i \\ 2+5i \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} + i \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}$, then

$$\text{Re } \mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, \quad \text{Im } \mathbf{x} = \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix}, \quad \text{and} \quad \bar{\boldsymbol{x}} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} - i \begin{bmatrix} -1 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 3+i \\ -i \\ 2-5i \end{bmatrix}$$

If B is an $m \times n$ matrix with possibly complex entries, then $\bar{B}$ denotes the matrix whose entries are the complex conjugates of the entries in B. Properties of conjugates for complex numbers carry over to complex matrix algebra:

$$\overline{r\boldsymbol{x}} = \bar{r}\bar{\boldsymbol{x}}, \quad \overline{B\boldsymbol{x}} = \bar{B}\bar{\boldsymbol{x}}, \quad \overline{BC} = \bar{B}\bar{C}, \quad \text{and} \quad \overline{rB} = \bar{r}\bar{B}$$

Eigenvalues and Eigenvectors of a Real matrix That Acts on $\mathbb{C}^n$

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

$$A\bar{\mathbf{x}} = \overline{Ax} = \overline{\lambda x} = \bar{\lambda}\bar{\mathbf{x}}$$

Hence $\bar{\lambda}$ is also an eigenvalue of A, with $\bar{\mathbf{x}}$ a corresponding eigenvector. This shows that *when A is real, its complex eigenvalues occur in conjugate pairs.* A **complex eigenvalue** refers to an eigenvalue of the form $\lambda = a + bi$, with $b \neq 0$.

If $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ where a and b are real and not both zero, then the eigenvalues of C are

5.5 Applications to Differential Equations

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

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

Where $x_1, ..., x_n$ are differentiable functions of t, with derivatives $x_1', ..., x_n'$ and constants a_{ij}. The crucial feature of the system above is that it is indeed *linear*. In order to see this we write the system in matrix notation and find:

$$\mathbf{x}'(t) = A\mathbf{x}(t)$$

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} a_{11} & ... & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & ... & a_{nn} \end{bmatrix}$$

A **solution** of equation 5.5 is a vector-valued function that satisfies the equation for all t in some intervfal of real numbers, such as $t \geq 0$. Th equation is linear because both differentiation of a function and multiplication of vectors by a matrix are linear transformations. If $\mathbf{u}$ and $\mathbf{v}$ are both solutions of $\mathbf{x}' = A\mathbf{x}$, then $c\mathbf{u} + d\mathbf{v}$ is also a solution because

$$\begin{aligned} (c\mathbf{u} + d\mathbf{v})' &= c\mathbf{u}' + c\mathbf{v}' \\ &= cA\mathbf{u} + dA\mathbf{v} = A(c\mathbf{u} + d\mathbf{v}) \end{aligned}$$

Note that the zero function is a trivial solution of $\mathbf{x}' = A\mathbf{x}$. It can be shown that there always exists what is called a **fundamental set of solutions** to 5.5. If A is $n \times n$, then there are n linearly independent functions in a fundamental set, and each solution of 5.5 is a unique linear combination of these n functions. If a vector $\mathbf{x}_0$ is specified, then the **initial value problem** is to construct the function $\mathbf{x}$ such that $\mathbf{x}' = A\mathbf{x}$ and $\mathbf{x}(0) = \mathbf{x}_0$. When A is diagonal, the solution of such a system of differential equations is simple, consider:

$$\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

Which represents the system

$$x_1'(t) = 3x_1(t)$$
$$x_2'(t) = -5x_2(t)$$

The system is said to be *decoupled* as each derivative only depends on its own initial function. The solution to the system is

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} c_1 e^{3t} \\ c_2 e^{-5t} \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{3t} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} e^{-5t}$$

Which suggests that for the general equation $\mathbf{x}' = A\mathbf{x}$, a solution might be a linear combination of functions of the form

$$\mathbf{x}(t) = \mathbf{v}e^{\lambda t}$$

for some scalar λ and some fixed nonzero vector $\mathbf{v}$. Observe that, by this suggested solution we have that

$$\mathbf{x}'(t) = \lambda \mathbf{v} e^{\lambda t}$$
$$A\mathbf{x}(t) = A\mathbf{v}e^{\lambda t}$$

As $e^{\lambda t}$ is never zero, $\mathbf{x}'(t)$ will equal $A\mathbf{x}(t)$ if and only if $\lambda \mathbf{v} = A\mathbf{v}$, that is, if and only if λ is an eigenvalue of A and $\mathbf{v}$ is a corresponding eigenvector. Thus each eigenvalue-eigenvector pair provides a solution of $\mathbf{x}' = A\mathbf{x}$.

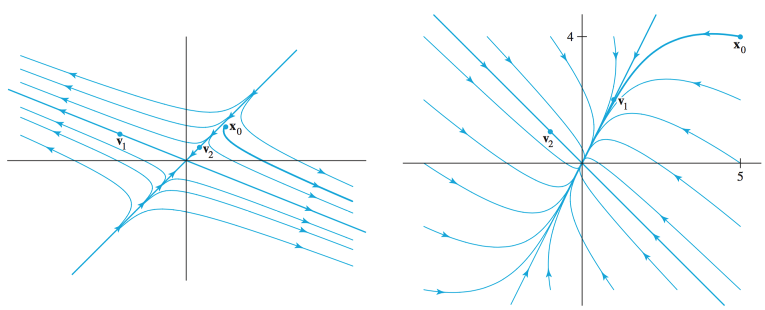

Figure 5.2: Origin functioning as a saddle point on the left and as an attractor on the right

Note that if we have two solutions to a system $\mathbf{x}'(t) = Ax(t)$, namely $\mathbf{x}_1$ and $\mathbf{x}_2$, that any linear combination of these is also a solution of the system by the linearity of differentiation and matrix transformation. Note that the behavior of the solution depends on the exponent λ. If all exponents are negative, the origin functions as an attractor as for large values of t the solution $\mathbf{x}(t)$ will approach zero, which is opposite to the behavior of a solution with only positive exponents. If one exponent is negative and the other is positive, the origin functions as a saddle point. Both are shown in figure 5.2.

Decoupling a Dynamical System

The following discussion shows that the method treated in the previous section produces a fundamental set of solutions for any dynamical system described by $\mathbf{x}' = A\mathbf{x}$ when A is $n \times n$ and has n linearly independent eigenvectors (i.e. A is diagonalizable). Suppose the eigenfunctions for A are

$$\mathbf{v}_1 e^{\lambda_1 t}, ..., \mathbf{v}_n e^{\lambda_n t}$$

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

$$\mathbf{y}(t) = P^{-1}\mathbf{x}(t) \quad \text{or, equivalently,} \quad \mathbf{x}(t) = P\mathbf{y}(t)$$

Note how $\mathbf{y}(t)$ is the coordinate vector of $\mathbf{x}(t)$ relative to the eigenvector basis. If we substitute $P\mathbf{y}$ in $\mathbf{x}' = A\mathbf{x}$

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

Since P is a constant matrix, the left side is differentiated to $P\mathbf{y}'$, hence if we left multiply both equations with P^{-1} we have that $\mathbf{y}' = D\mathbf{y}$, or

$$\begin{bmatrix} y_1'(t) \\ y_2'(t) \\ \vdots \\ y_n'(t) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & ... & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{bmatrix}$$

By change of variable from $\mathbf{x}$ to $\mathbf{y}$ we have *decoupled* the system of differential equations. Now the solution is simply:

$$\mathbf{y}(t) = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix}, \quad \text{where} \quad \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{y}_0 = P^{-1}\mathbf{x}_0$$

To obtain the general solution $\mathbf{x}$ of the original system

$$\begin{aligned} \mathbf{x}(t) &= P\mathbf{y}(t) = [\mathbf{v}_1 \quad ... \quad \mathbf{v}_n]\mathbf{y}(t) \\ &= c_1\mathbf{v}_1 e^{\lambda_1 t} + ... + c_n \mathbf{v}_n e^{\lambda_n t} \end{aligned}$$

Complex Eigenvalues

First recall that the complex eigenvalues of a real matrix always come in conjugate pairs. Then, we have that two solutions of $\mathbf{x}' = A\mathbf{x}$ are

$$\mathbf{x}_1(t) = \mathbf{v}e^{\lambda t} \quad \text{and} \quad \mathbf{x}_2(t) = \bar{\mathbf{v}}e^{\bar{\lambda} t}$$

Note that above eigenfunctions are complex, and that most applications require real eigenfunctions. Fortunately, the real and imaginary parts of $\mathbf{x}_1$ are real solutions of $\mathbf{x}' = A\mathbf{x}$, because they are linear combinations of the solutions above:

$$\text{Re}(\mathbf{v}e^{\lambda t}) = \frac{1}{2}[\mathbf{x}_1(t) + \overline{\mathbf{x}_1(t)}], \quad \text{Im}(\mathbf{v}e^{\lambda t}) = \frac{1}{2i}[\mathbf{x}_1(t) - \overline{\mathbf{x}_1(t)}]$$

So let us find an expression for the real and imaginary parts of $\mathbf{x}_1$. Recall that it is possible, by the use of Maclaurin series, to show that

$$e^{\lambda t} = e^{(a+bi)t} = e^{at} \cdot e^{ibt} = e^{at}(\cos bt + i \sin bt)$$

By use of this fact the two real solutions of $\mathbf{x}' = A\mathbf{x}$ are

$$\begin{aligned} \mathbf{y}_1(t) &= \text{Re } \mathbf{x}_1(t) = [(\text{Re } \mathbf{v})\cos bt - (\text{Im } \mathbf{v})\sin bt]e^{at} \\ \mathbf{y}_2(t) &= \text{Im } \mathbf{x}_1(t) = [(\text{Re } \mathbf{v})\sin bt + (\text{Im } \mathbf{v})\cos bt]e^{at} \end{aligned}$$

Moreover, it can be shown that $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ are linearly independent functions (when $b \neq 0$).

6 Symmetric Matrices and Quadratic Forms

6.1 Diagonalization of Symmetric Matrices

A **symmetric** matrix is a matrix A such that $A^T = A$, which means that A must be square. Note that its main diagonal entries may be arbitrary, whereas other entries occur in pairs on opposite sides of the main diagonal.

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

THEOREM 1 If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, compute

$$\begin{aligned} \lambda \mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 \\ &= (\mathbf{v}_1^T A^T)\mathbf{v}_2 = \mathbf{v}_1^T (A\mathbf{v}_2) \\ &= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) \\ &= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2 \end{aligned}$$

Hence $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. But $\lambda_1 - \lambda_2 \neq 0$, so $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.
An $n \times n$ matrix A is said to be **orthogonally diagonalizable** if is an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

The diagonalization of above requires n linearly independent orthonormal eigenvectors. When is this possible? If A is orthogonally diagonalizable, then

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A$$

Thus it is possible if A is symmetric. We can then reason the following:

THEOREM 2 An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is a symmetric matrix.

The Spectral Theorem

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

THEOREM 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.

PROOF Property (a) can be proven by the fact that any vector $\mathbf{x}$ in $\mathbb{C}^n$ satisfies that $q = \bar{\mathbf{x}}^T A\mathbf{x}$ is real, hence λ must be real as $\bar{\mathbf{x}}^T A\mathbf{x} = \bar{\mathbf{x}}\lambda\mathbf{x}$ and $\bar{\mathbf{x}}^T\mathbf{x}$ is real. Part (c) is Theorem 1. Statement (b) follows from (d).

Spectral Decomposition

Suppose $A = PDP^{-1}$, where the columns of P are orthonormal eigenvectors $\mathbf{u}_1, ..., \mathbf{u}_n$ of A and the corresponding eigenvalues $\lambda_1, ..., \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$\begin{aligned} A = PDP^T &= [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_2] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \\ &= [\lambda_1\mathbf{u}_1 \quad \dots \quad {}_n\mathbf{u}_n] \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_2^T \end{bmatrix} \end{aligned}$$

Then we can represent A as follows:

$$A = \lambda_1\mathbf{u}_!\mathbf{u}_1^T + \lambda_2\mathbf{u}_2\mathbf{u}_2^T + ... + \lambda_n\mathbf{u}_n\mathbf{u}_n^T$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into terms determined by the spectrum (i.e. eigenvalues) of A. Note how every term is an $n \times n$ matrix of rank 1. Why? Each column of matrix $\lambda_j\mathbf{u}_j\mathbf{u}_j^T$ is a multiple of $\mathbf{u}_j$.

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 $\mathbf{x}$ 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. Note how the simplest example of a nonzero quadratic form is the inner product $Q(\mathbf{x}) = \mathbf{x}^T I\mathbf{x} = \|\mathbf{x}\|^2$.

Change of Variable in a Quadratic Form

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

$$\mathbf{x} = P\mathbf{y}, \quad \text{or equivalently,} \quad \mathbf{y} = P^{-1}\mathbf{x}$$

where P is an invertible matrix and $\mathbf{y}$ is a new variable vector in $\mathbb{R}^n$. Here $\mathbf{y}$ is the coordinate vector of $\mathbf{x}$ relative to the basis of $\mathbb{R}^n$ determined by the columns of P.

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

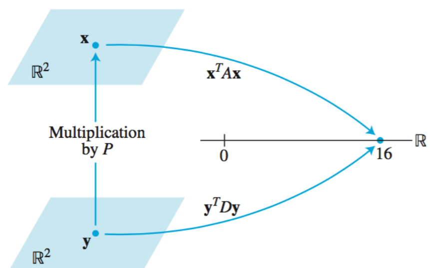

Figure 6.1: Change of variable in $\mathbf{x}^T A\mathbf{x}$

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T AP\mathbf{y} = \mathbf{y}^T (P^T AP)\mathbf{y}$$

and the new matrix of the quadratic form is $P^T AP$. Since A is symmetric, Theorem 2 guarantees that there is an orthogonal matrix P such that $P^T AP$ is a diagonal matrix D, and we can write the quadratic form as $\mathbf{y}^T D\mathbf{y}$. Why would we do this? By ensuring that the matrix of the quadratic form is diagonal, we avoid cross-product terms. Hence we have the following theorem:

THEOREM 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.

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

A Geometric View of Principal Axes

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

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

either corresponds to an ellipse (or circle), a hyperbola, two intersecting lines, or a single point, or contains no points at all. If A is a diagonal matrix, the graph is in *standard position* as in figure 6.2. If A is not diagonal, the graph is rotated out of standard position as cross-

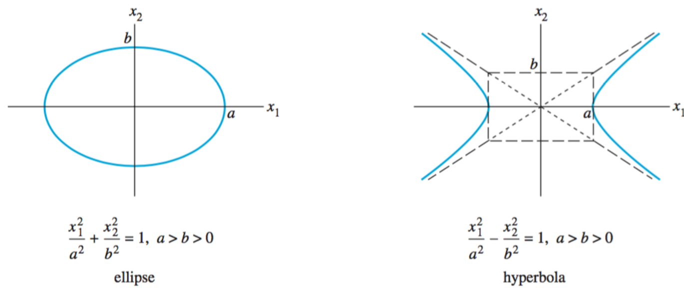

Figure 6.2: An ellipse and hyperbola in standard position

product terms are introduced. Finding the *principal axes* (determined by the eigenvectors of A) amounts to finding a new coordinate system with respect to which the graph is in standard position, as shown in figure 6.3.

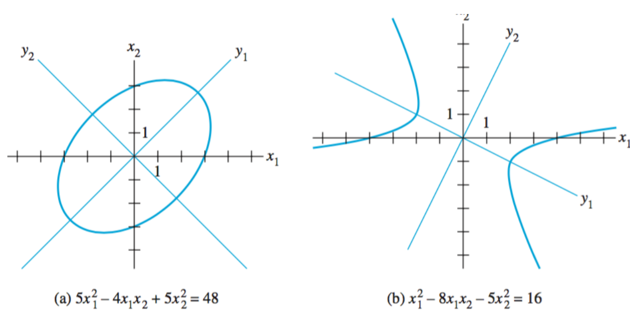

Figure 6.3: An ellipse and hyperbola *not* in standard position

Classifying Quadratic Forms

When A is an $n \times n$ matrix, the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a real-valued function with domain $\mathbb{R}^n$. Figure 6.4 shows four quadratic forms with domain $\mathbb{R}^2$, where the points (x_1, x_2, z) with $z = Q(\mathbf{x})$ are plotted.

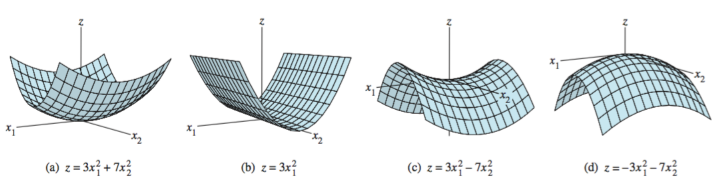

Figure 6.4: Graphs of quadratic forms. Classifications: (a) positive definite, (b) positive semidefinite, (c) indefinite, (d) negative definite

DEFINITION

A quadratic form Q is:

a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,

b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,

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

In addition to this definition, Q is said to be **positive semidefinite** if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x}$, and to be **negative semidefinite** if $Q(\mathbf{x}) \leq 0$ for all $\mathbf{x}$. Note how both (a) and (b) in figure 6.4 are positive semidefinite, but the form in (a) is also positive definite.

THEOREM 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, or

c. indefinite if and only if A has both positive and negative eigenvalues.

PROOF By the Principal Axes Theorem, there exists an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ such that

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

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.