Review of Linear Algebra

Synopsis

A summary of essential concepts, definitions, and theorems in linear algebra used throughout this book.

A.1. SYSTEMS OF LINEAR EQUATIONS

Recall that a system of linear equations can be solved by the process of **Gaussian elimination**.

Example A.1

Consider the following system of equations:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 14 \\ x_1 + 2x_2 + 2x_3 &= 11 \\ x_1 + 3x_2 + 4x_3 &= 19. \end{aligned} \tag{A.1}$$

We eliminate x_1 from the second and third equations by subtracting the first equation from the second and third equations to obtain

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 14 \\ -x_3 &= -3 \\ x_2 + x_3 &= 5. \end{aligned} \tag{A.2}$$

We would like x_2 to appear in the second equation, so we interchange the second and third equations:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 14 \\ x_2 + x_3 &= 5 \\ -x_3 &= -3. \end{aligned} \tag{A.3}$$

Parameter Estimation and Inverse Problems, Second Edition. DOI: 10.1016/B978-0-12-385048-5.00013-6

Next, we eliminate x_2 from the first equation by subtracting two times the second equation from the first equation:

$$\begin{aligned} x_1 + x_3 &= 4 \\ x_2 + x_3 &= 5 \\ -x_3 &= -3. \end{aligned} \tag{A.4}$$

We then multiply the third equation by -1 to get an equation for x_3:

$$\begin{aligned} x_1 + x_3 &= 4 \\ x_2 + x_3 &= 5 \\ x_3 &= 3. \end{aligned} \tag{A.5}$$

Finally, we eliminate x_3 from the first two equations to obtain a solution to the original system of equations:

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 2 \\ x_3 &= 3. \end{aligned} \tag{A.6}$$

Geometrically the constraints specified by the three equations of (A.1) describe three planes that, in this case, intersect in a single point, $x_1 = 1$, $x_2 = 2$, $x_3 = 3$.

In solving (A.1), we used three **elementary row operations**: adding a multiple of one equation to another equation, multiplying an equation by a nonzero constant, and swapping two equations. This process can be extended to solve systems of equations with an arbitrary number of variables.

In performing the elimination process, the actual names of the variables are insignificant. We could have renamed the variables in the above example to a, b, and c without changing the solution in any significant way. Because the actual names of the variables are insignificant, we can save space by writing down the significant coefficients from the system of equations in matrix form as an **augmented matrix**. The augmented matrix form is also useful in solving a system of equations in computer algorithms, where the elements of the augmented matrix are stored in an array.

In augmented matrix form (A.1) becomes

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 1 & 2 & 2 & 11 \\ 1 & 3 & 4 & 19 \end{array}\right]. \tag{A.7}$$

In augmented notation, the elementary row operations become adding a multiple of one row to another row, multiplying a row by a nonzero constant, and interchanging two rows. The Gaussian elimination process is essentially identical to the process used in Example A.1, with the final version of the augmented matrix given by

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]. \tag{A.8}$$

Definition A.1 A matrix is said to be in **reduced row echelon form (RREF)** if it has the following properties:

1. The first nonzero element in each row is a 1. The first nonzero row elements of the matrix are called **pivot elements**. A column in which a pivot element appears is called a **pivot column**.
2. Except for the pivot element, all elements in pivot columns are 0s.
3. Each pivot element is to the right of the pivot elements in previous rows.
4. Any rows consisting entirely of 0s are at the bottom of the matrix.

In solving a system of equations in augmented matrix form, we apply elementary row operations to reduce the augmented matrix to RREF and then convert back to conventional notation to read off the solutions. The process of transforming a matrix into RREF can easily be automated. In MATLAB, this is done by the **rref** command.

It can be shown that any linear system of equations has either no solutions, exactly one solution, or infinitely many solutions [95]. In a system of two dimensions, for example, lines represented by the equations can fail to intersect (no solution), intersect at a point (one solution), or intersect in a line (many solutions). The following example shows how to determine the number of solutions from the RREF of the augmented matrix.

Example A.2

Consider a system of two equations in three variables that has many solutions:

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 + 2x_2 + 2x_3 &= 0. \end{aligned} \tag{A.9}$$

We put this system of equations into augmented matrix form and then find the RREF, which is

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right]. \tag{A.10}$$

We can translate this back into equation form as

$$\begin{aligned} x_1 &= 0 \\ x_1 + x_3 &= 0. \end{aligned} \tag{A.11}$$

Clearly, x_1 must be 0 in any solution to the system of equations. However, x_2 and x_3 are not fixed. We could treat x_3 as a **free variable** and allow it to take on any value. Whatever value x_3 takes on, x_2 must be equal to $-x_3$. Geometrically, this system of equations describes the intersection of two planes, where the intersection consists of points on the line $x_2 = -x_3$ in the $x_1 = 0$ plane.

A linear system of equations may have more equation constraints than variables, in which case the system of equations is **over-determined**. Although over-determined systems often have no solutions, it is possible for an over-determined system of equations to have either many solutions or exactly one solution.

Conversely, a system of equations with fewer equations than variables is **under-determined**. Although in many cases under-determined systems of equations have infinitely many solutions, it is also possible for such systems to have no solutions.

A system of equations with all 0s on the right-hand side is **homogeneous**. Every homogeneous system of equations has at least one solution, the trivial solution in which all of the variables are 0s. A system of equations with a nonzero right-hand side is **nonhomogeneous**.

A.2. MATRIX AND VECTOR ALGEBRA

As we have seen in the previous section, a matrix is a table of numbers laid out in rows and columns. A **vector** is a matrix consisting of a single column or row of numbers (vectors in this text are typically expressed as column vectors). In general, matrices and vectors may contain complex numbers as well as real numbers. With the exception of Chapter 8, all of the vectors and matrices in this book are real.

There are several important notational conventions used here for matrices and vectors. Boldface capital letters such as $\mathbf{A}$, $\mathbf{B}$, . . . are used to denote matrices. Boldface lower-case letters such as $\mathbf{x}$, $\mathbf{y}$, . . . are used to denote vectors. Lower-case letters or Greek letters such as m, n, α, β, . . . will be used to denote scalars.

At times we will need to refer to specific parts of a matrix. The notation $A_{i,j}$ denotes the element of the matrix $\mathbf{A}$ in row i and column j. We denote the *jth* element of the vector $\mathbf{x}$ by x_j. The notation $\mathbf{A}_{\cdot,j}$ is used to refer to column j of the matrix $\mathbf{A}$, while $\mathbf{A}_{i,\cdot}$ refers to row i of $\mathbf{A}$.

We can also construct larger matrices from smaller matrices. The notation $\mathbf{A} = [\mathbf{B}\ \ \mathbf{C}]$ means that the matrix $\mathbf{A}$ is composed of the matrices $\mathbf{B}$ and $\mathbf{C}$, with matrix $\mathbf{C}$ to the right of $\mathbf{B}$.

If $\mathbf{A}$ and $\mathbf{B}$ are two matrices of the *same size*, we can add them by simply adding corresponding elements. Similarly, we can subtract $\mathbf{B}$ from $\mathbf{A}$ by subtracting the corresponding elements of $\mathbf{B}$ from those of $\mathbf{A}$. We can multiply a scalar times a matrix by multiplying the scalar times each matrix element. Because vectors are just n by 1 matrices, we can perform the same arithmetic operations on vectors. A **zero matrix**, $\mathbf{0}$, is a matrix composed of all zero elements. A zero matrix plays the same role in matrix algebra as the scalar 0, with

$$\begin{aligned} \mathbf{A} + \mathbf{0} &= \mathbf{A} && \text{(A.12)} \\ &= \mathbf{0} + \mathbf{A}. && \text{(A.13)} \end{aligned}$$

Using vector notation, we can write a linear system of equations in **vector form**.

Example A.3

Recall the system of equations (A.9),

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 + 2x_2 + 2x_3 &= 0, \end{aligned} \tag{A.14}$$

from Example A.2. We can write this in vector form as

$$x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{A.15}$$

The expression on the left-hand side of (A.15) where vectors are multiplied by scalars and the results are summed together is called a **linear combination**.

If $\mathbf{A}$ is an m by n matrix, and $\mathbf{x}$ is an n element vector, we can multiply $\mathbf{A}$ times $\mathbf{x}$, where the product is defined by

$$\mathbf{A}\mathbf{x} = x_1 \mathbf{A}_{\cdot,1} + x_2 \mathbf{A}_{\cdot,2} + \cdots + x_n \mathbf{A}_{\cdot,n}. \tag{A.16}$$

Example A.4

Given

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \tag{A.17}$$

and

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \tag{A.18}$$

then

$$\mathbf{Ax} = 1\begin{bmatrix}1\\4\end{bmatrix} + 0\begin{bmatrix}2\\5\end{bmatrix} + 2\begin{bmatrix}3\\6\end{bmatrix} = \begin{bmatrix}7\\16\end{bmatrix}. \tag{A.19}$$

The formula (A.16) for $\mathbf{Ax}$ is a linear combination much like the one that occurred in the vector form of a system of equations. It is possible to write any linear system of equations in the form $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a matrix containing the coefficients of the variables in the equations, $\mathbf{b}$ is a vector containing the coefficients on the right-hand sides of the equations, and $\mathbf{x}$ is a vector containing the variables.

Definition A.2 If $\mathbf{A}$ is a matrix of size m by n, and $\mathbf{B}$ is a matrix of size n by r, then the product $\mathbf{C} = \mathbf{AB}$ is obtained by multiplying $\mathbf{A}$ times each of the columns of $\mathbf{B}$ and assembling the matrix vector products in $\mathbf{C}$:

$$\mathbf{C} = \left[\mathbf{AB}_{\cdot,1} \;\; \mathbf{AB}_{\cdot,2} \;\; \ldots \;\; \mathbf{AB}_{\cdot,r}\right]. \tag{A.20}$$

This approach given in (A.20) for calculating a matrix-matrix product will be referred to as the **matrix-vector method**.

Note that the product (A.20) is only possible if the two matrices are of compatible sizes. If $\mathbf{A}$ has m rows and n columns, and $\mathbf{B}$ has n rows and r columns, then the product $\mathbf{AB}$ exists and is of size m by r. In some cases, it is thus possible to multiply $\mathbf{AB}$ but not $\mathbf{BA}$. It is important to note that when both $\mathbf{AB}$ and $\mathbf{BA}$ exist, $\mathbf{AB}$ is not generally equal to $\mathbf{BA}$!

An alternate way to compute the product of two matrices is the **row-column expansion method**, where the product element $C_{i,j}$ is calculated as the matrix product of row i of $\mathbf{A}$ and column j of $\mathbf{B}$.

Example A.5

Let

$$\mathbf{A} = \begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix} \tag{A.21}$$

and

$$\mathbf{B} = \begin{bmatrix}5 & 2\\3 & 7\end{bmatrix}. \tag{A.22}$$

The product matrix $\mathbf{C} = \mathbf{AB}$ will be of size 3 by 2. We compute the product using both methods. First, using the matrix–vector approach (A.20), we have

$$\mathbf{C} = \begin{bmatrix} \mathbf{AB}_{\cdot,1} & \mathbf{AB}_{\cdot,2} \end{bmatrix} \tag{A.23}$$

$$= \begin{bmatrix} 5\begin{bmatrix}1\\3\\5\end{bmatrix} + 3\begin{bmatrix}2\\4\\6\end{bmatrix} & 2\begin{bmatrix}1\\3\\5\end{bmatrix} + 7\begin{bmatrix}2\\4\\6\end{bmatrix} \end{bmatrix} \tag{A.24}$$

$$= \begin{bmatrix} 11 & 16 \\ 27 & 34 \\ 43 & 52 \end{bmatrix}. \tag{A.25}$$

Next, we use the row–column approach:

$$\mathbf{C} = \begin{bmatrix} 1\cdot5+2\cdot3 & 1\cdot2+2\cdot7 \\ 3\cdot5+4\cdot3 & 3\cdot2+4\cdot7 \\ 5\cdot5+6\cdot3 & 5\cdot2+6\cdot7 \end{bmatrix} \tag{A.26}$$

$$= \begin{bmatrix} 11 & 16 \\ 27 & 34 \\ 43 & 52 \end{bmatrix}. \tag{A.27}$$

Definition A.3 The n by n **identity matrix** $\mathbf{I}_n$ is composed of 1s in the diagonal and 0s in the off-diagonal elements.

For example, the 3 by 3 identity matrix is

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{A.28}$$

We often write $\mathbf{I}$ without specifying the size of the matrix in situations where the size of matrix is obvious from the context. It is easily shown that if $\mathbf{A}$ is an m by n matrix, then

$$\mathbf{AI}_n = \mathbf{A} \tag{A.29}$$

$$= \mathbf{I}_m\mathbf{A}. \tag{A.30}$$

Thus, multiplying by $\mathbf{I}$ in matrix algebra is similar to multiplying by 1 in conventional scalar algebra.

We have not defined matrix division, but it is possible at this point to define the matrix algebra equivalent of the reciprocal.

Definition A.4 If $\mathbf{A}$ is an n by n matrix, and there is a matrix $\mathbf{B}$ such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}, \tag{A.31}$$

then $\mathbf{B}$ is the **inverse** of $\mathbf{A}$. We write $\mathbf{B} = \mathbf{A}^{-1}$.

How do we compute the inverse of a matrix? If $\mathbf{AB} = \mathbf{I}$, then

$$\begin{bmatrix} \mathbf{AB}_{\cdot,1} & \mathbf{AB}_{\cdot,2} & \ldots & \mathbf{AB}_{\cdot,n} \end{bmatrix} = \mathbf{I}. \tag{A.32}$$

Since the columns of the identity matrix and $\mathbf{A}$ are known, we can solve

$$\mathbf{AB}_{\cdot,1} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{A.33}$$

to obtain $\mathbf{B}_1$. We can find the remaining columns of the inverse in the same way. If any of these systems of equations are inconsistent, then $\mathbf{A}^{-1}$ does not exist.

The inverse matrix can be used to solve a system of linear equations with n equations and n variables. Given the system of equations $\mathbf{Ax} = \mathbf{b}$, and $\mathbf{A}^{-1}$, we can multiply $\mathbf{Ax} = \mathbf{b}$ on both sides by the inverse to obtain

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b}. \tag{A.34}$$

Because

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{Ix} \tag{A.35}$$

$$= \mathbf{x}, \tag{A.36}$$

this gives the solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}. \tag{A.37}$$

This argument shows that if $\mathbf{A}^{-1}$ exists, then for any right-hand side $\mathbf{b}$, a system of equations has a *unique* solution. If $\mathbf{A}^{-1}$ does not exist, then the system of equations may have either many solutions or no solution.

Definition A.5 When $\mathbf{A}$ is an n by n matrix, $\mathbf{A}^k$ is the product of k copies of $\mathbf{A}$. By convention, we define $\mathbf{A}^0 = \mathbf{I}$.

Definition A.6 The **transpose** of a matrix $\mathbf{A}$, denoted $\mathbf{A}^T$, is obtained by taking the columns of $\mathbf{A}$ and writing them as the rows of the transpose. We will also use the notation $\mathbf{A}^{-T}$ for $(\mathbf{A}^{-1})^T$.

Example A.6
Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}. \tag{A.38}$$

Then

$$\mathbf{A}^T = \begin{bmatrix} 2 & 5 \\ 1 & 2 \end{bmatrix}. \tag{A.39}$$

Definition A.7 A matrix is **symmetric** if $\mathbf{A} = \mathbf{A}^T$.

Although many elementary textbooks on linear algebra consider only square diagonal matrices, we will have occasion to refer to m by n matrices that have nonzero elements only on the diagonal.

Definition A.8 An m by n matrix $\mathbf{A}$ is **diagonal** if $A_{i,j} = 0$ whenever $i \neq j$.

Definition A.9 An m by n matrix $\mathbf{R}$ is **upper-triangular** if $R_{i,j} = 0$ whenever $i > j$. A matrix $\mathbf{L}$ is **lower-triangular** if $\mathbf{L}^T$ is upper-triangular.

Example A.7
The matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \end{bmatrix} \tag{A.40}$$

is diagonal, and the matrix

$$\mathbf{R} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix} \tag{A.41}$$

is upper-triangular.

Theorem A.1 *The following statements are true for any scalars s and t and matrices* $\mathbf{A}$, $\mathbf{B}$, *and* $\mathbf{C}$. *It is assumed that the matrices are of the appropriate size for the operations involved and that whenever an inverse occurs, the matrix is invertible.*

1. $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$.
2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.
3. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$.
4. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.
5. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.
6. $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.
7. $(st)\mathbf{A} = s(t\mathbf{A})$.
8. $s(\mathbf{AB}) = (s\mathbf{A})\mathbf{B} = \mathbf{A}(s\mathbf{B})$.
9. $(s + t)\mathbf{A} = s\mathbf{A} + t\mathbf{A}$.
10. $s(\mathbf{A} + \mathbf{B}) = s\mathbf{A} + s\mathbf{B}$.
11. $(\mathbf{A}^T)^T = \mathbf{A}$.
12. $(s\mathbf{A})^T = s(\mathbf{A}^T)$.
13. $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
14. $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.
15. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
16. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
17. $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.
18. If $\mathbf{A}$ and $\mathbf{B}$ are n by n matrices, and $\mathbf{AB} = \mathbf{I}$, then $\mathbf{A}^{-1} = \mathbf{B}$ and $\mathbf{B}^{-1} = \mathbf{A}$.

The first 10 rules in this list are identical to rules of conventional algebra, and you should have little trouble in applying them. The rules involving transposes and inverses are new, but they can be mastered without too much trouble.

Some students have difficulty with the following statements, which would appear to be true on the surface, but that are in fact **false** for at least some matrices.

1. $\mathbf{AB} = \mathbf{BA}$.
2. If $\mathbf{AB} = \mathbf{0}$, then $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.
3. If $\mathbf{AB} = \mathbf{AC}$ and $\mathbf{A} \neq 0$, then $\mathbf{B} = \mathbf{C}$.

It is a worthwhile exercise to construct examples of 2 by 2 matrices for which each of these statements is false.

A.3. LINEAR INDEPENDENCE

Definition A.10 The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **linearly independent** if the system of equations

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0} \tag{A.42}$$

has only the trivial solution $\mathbf{c} = \mathbf{0}$. If there are multiple solutions, then the vectors are **linearly dependent**.

Determining whether a set of vectors is linearly independent is simple. Just solve the above system of equations (A.42).

Example A.8
Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}. \tag{A.43}$$

Are the columns of $\mathbf{A}$ linearly independent vectors? To determine this we set up the system of equations $\mathbf{Ax} = \mathbf{0}$ in an augmented matrix, and then find the RREF,

$$\left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]. \tag{A.44}$$

The solutions are

$$\mathbf{x} = x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}. \tag{A.45}$$

We can set $x_3 = 1$ and obtain the nonzero solution,

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}. \tag{A.46}$$

Thus, the columns of $\mathbf{A}$ are linearly dependent.

There are a number of important theoretical consequences of linear independence. For example, it can be shown that if the columns of an n by n matrix $\mathbf{A}$ are linearly independent, then $\mathbf{A}^{-1}$ exists, and the system of equations $\mathbf{Ax} = \mathbf{b}$ has a unique solution for every right-hand side $\mathbf{b}$ [95].

A.4. SUBSPACES OF R^n

So far, we have worked with vectors of real numbers in the n-dimensional space R^n. There are a number of properties of R^n that make it convenient to work with vectors. First, the operation of vector addition always works. We can take any two vectors in R^n and add them together and get another vector in R^n. Second, we can multiply any vector in R^n by a scalar and obtain another vector in R^n. Finally, we have the $\mathbf{0}$ vector, with the property that for any vector $\mathbf{x}$, $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$.

Definition A.11 A **subspace** W of R^n is a subset of R^n that satisfies the three properties:

1. If $\mathbf{x}$ and $\mathbf{y}$ are vectors in W, then $\mathbf{x}+\mathbf{y}$ is also a vector in W.
2. If $\mathbf{x}$ is a vector in W and s is any real scalar, then $s\mathbf{x}$ is also a vector in W.
3. The $\mathbf{0}$ vector is in W. A subspace of R^n is **nontrivial** if it contains vectors other than the $\mathbf{0}$ vector.

Example A.9

In R^3, the plane P defined by the equation

$$x_1 + x_2 + x_3 = 0 \tag{A.47}$$

is a subspace of R^n. To see this, note that if we take any two vectors in the plane and add them together, we get another vector in the plane. If we take a vector in this plane and multiply it by any scalar, we get another vector in the plane. Finally, $\mathbf{0}$ is a vector in the plane.

Subspaces are important because they provide an environment within which all of the rules of matrix-vector algebra apply. An especially important subspace of R^n that we will work with is the **null space** of an m by n matrix.

Definition A.12 Let $\mathbf{A}$ be an m by n matrix. The null space of $\mathbf{A}$, written $N(\mathbf{A})$, is the set of all vectors $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{0}$.

To show that $N(\mathbf{A})$ is actually a subspace of R^n, we need to show that:

1. If $\mathbf{x}$ and $\mathbf{y}$ are in $N(\mathbf{A})$, then $\mathbf{Ax} = \mathbf{0}$ and $\mathbf{Ay} = \mathbf{0}$. By adding these equations, we find that $\mathbf{A}(\mathbf{x}+\mathbf{y}) = \mathbf{0}$. Thus $\mathbf{x}+\mathbf{y}$ is in $N(\mathbf{A})$.
2. If $\mathbf{x}$ is in $N(\mathbf{A})$ and s is any scalar, then $\mathbf{Ax} = \mathbf{0}$. We can multiply this equation by s to get $s\mathbf{Ax} = \mathbf{0}$. Thus $\mathbf{A}(s\mathbf{x}) = \mathbf{0}$, and $s\mathbf{x}$ is in $N(\mathbf{A})$.
3. $\mathbf{A0} = \mathbf{0}$, so $\mathbf{0}$ is in $N(\mathbf{A})$.

Computationally, the null space of a matrix can be determined by solving the system of equations $\mathbf{Ax} = \mathbf{0}$.

Example A.10

Let

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 9 & 4 \\ 2 & 1 & 7 & 3 \\ 5 & 2 & 16 & 7 \end{bmatrix}. \tag{A.48}$$

To find the null space of $\mathbf{A}$, we solve the system of equations $\mathbf{Ax} = \mathbf{0}$. To solve the equations, we put the system of equations into an augmented matrix,

$$\left[\begin{array}{cccc|c} 3 & 1 & 9 & 4 & 0 \\ 2 & 1 & 7 & 3 & 0 \\ 5 & 2 & 16 & 7 & 0 \end{array}\right], \tag{A.49}$$

and find the RREF,

$$\left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{A.50}$$

From the augmented matrix, we find that

$$\mathbf{x} = x_3 \begin{bmatrix} -2 \\ -3 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}. \tag{A.51}$$

Any vector in the null space can be written as a linear combination of the above vectors, so the null space is a two-dimensional plane within R^4.

Now, consider the problem of solving $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{b} = \begin{bmatrix} 22 \\ 17 \\ 39 \end{bmatrix} \tag{A.52}$$

and one particular solution is

$$\mathbf{p} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}. \tag{A.53}$$

We can take any vector in the null space of $\mathbf{A}$ and add it to this solution to obtain another solution. Suppose that $\mathbf{x}$ is in $N(\mathbf{A})$. Then

$$\begin{aligned} \mathbf{A}(\mathbf{x} + \mathbf{p}) &= \mathbf{Ax} + \mathbf{Ap} \\ \mathbf{A}(\mathbf{x} + \mathbf{p}) &= \mathbf{0} + \mathbf{b} \\ \mathbf{A}(\mathbf{x} + \mathbf{p}) &= \mathbf{b}. \end{aligned}$$

For example,

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} -2 \\ -3 \\ 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \tag{A.54}$$

is also a solution to $\mathbf{Ax} = \mathbf{b}$.

In the context of inverse problems, the null space is critical because the presence of a nontrivial null space leads to nonuniqueness in the solution to a linear system of equations.

Definition A.13 A **basis** for a subspace W is a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ such that:

1. Any vector in W can be written as a linear combination of the basis vectors.
2. The basis vectors are linearly independent.

A particularly simple and useful basis is the **standard basis**.

Definition A.14 The **standard basis** for R^n is the set of vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ such that the elements of $\mathbf{e}_i$ are all zero except for the *ith* element, which is 1.

Any nontrivial subspace W of R^n will have many different bases. For example, we can take any basis and multiply one of the basis vectors by 2 to obtain a new basis. It is possible to show that all bases for a subspace W have the same number of basis vectors [95].

Theorem A.2 *Let W be a subspace of R^n with basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. Then all bases for W have p basis vectors, and p is the* **dimension** *of W.*

It can be shown that the procedure used in the above example always produces a basis for $N(\mathbf{A})$ [95]. A basis for the null space of a matrix can be found in MATLAB using the **null** command.

Definition A.15 Let $\mathbf{A}$ be an m by n matrix. The **column space** or **range** of $\mathbf{A}$ (written $R(\mathbf{A})$) is the set of all vectors $\mathbf{b}$ such that $\mathbf{Ax} = \mathbf{b}$ has at least one solution. In other words, the column space is the set of all vectors $\mathbf{b}$ that can be written as a linear combination of the columns of $\mathbf{A}$.

The range is important in the context of discrete linear inverse problems, because $R(\mathbf{G})$ consists of all vectors $\mathbf{d}$ for which there is a model $\mathbf{m}$ such that $\mathbf{Gm} = \mathbf{d}$.

To find the column space of a matrix, we consider what happens when we compute the RREF of $[\mathbf{A} \mid \mathbf{b}]$. In the part of the augmented matrix corresponding to the left-hand side of the equations we always get the same result, namely the RREF of $\mathbf{A}$. The solution to the system of equations may involve some free variables, but we can always set these free variables to 0. Thus when we are able to solve $\mathbf{Ax} = \mathbf{b}$, we can solve the system of equations by using only variables corresponding to the pivot columns in the RREF of $\mathbf{A}$. In other words, if we can solve $\mathbf{Ax} = \mathbf{b}$, then we can write $\mathbf{b}$ as a linear combination of the pivot columns of $\mathbf{A}$. Note that these are columns from the original matrix $\mathbf{A}$, not columns from the RREF of $\mathbf{A}$. An orthonormal (see below) basis for the range of a matrix can be found in MATLAB using the **orth** command.

Example A.11

As in the previous example, let

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 9 & 4 \\ 2 & 1 & 7 & 3 \\ 5 & 2 & 16 & 7 \end{bmatrix}. \tag{A.55}$$

To find the column space of $\mathbf{A}$, note that the RREF of $\mathbf{A}$ is

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{A.56}$$

Thus, whenever we can solve $\mathbf{Ax} = \mathbf{b}$, we can find a solution in which x_3 and x_4 are 0. In other words, whenever there is a solution to $\mathbf{Ax} = \mathbf{b}$, we can write $\mathbf{b}$ as a linear combination of the first two columns of $\mathbf{A}$:

$$\mathbf{b} = x_1 \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}. \tag{A.57}$$

Since these two vectors are linearly independent and span $R(\mathbf{A})$, they form a basis for $R(\mathbf{A})$. The dimension of $R(\mathbf{A})$ is 2.

In finding the null space and range of a matrix $\mathbf{A}$, we found that the basis vectors for $N(\mathbf{A})$ corresponded to nonpivot columns of $\mathbf{A}$, while the basis vectors for $R(\mathbf{A})$ corresponded to pivot columns of $\mathbf{A}$. Since the matrix $\mathbf{A}$ had n columns, we obtain the following theorem.

Theorem A.3

$$dim\ N(\mathbf{A}) + dim\ R(\mathbf{A}) = n. \tag{A.58}$$

In addition to the null space and range of a matrix $\mathbf{A}$, we will often work with the null space and range of the transpose of $\mathbf{A}$. Since the columns of $\mathbf{A}^T$ are rows of $\mathbf{A}$, the column space of $\mathbf{A}^T$ is also called the **row space** of $\mathbf{A}$. Since each row of $\mathbf{A}$ can be written as a linear combination of the nonzero rows of the RREF of $\mathbf{A}$, the nonzero rows of the RREF form a basis for the row space of $\mathbf{A}$. There are exactly as many nonzero rows in the RREF of $\mathbf{A}$ as there are pivot columns. Thus we have the following theorem.

Theorem A.4

$$dim(R(\mathbf{A}^T)) = dim\ R(\mathbf{A}). \tag{A.59}$$

Definition A.16 The **rank** of an m by n matrix $\mathbf{A}$ is the dimension of $R(\mathbf{A})$. If $\text{rank}(\mathbf{A}) = \min(m, n)$, then $\mathbf{A}$ has **full rank**. If $\text{rank}(\mathbf{A}) = m$, then $\mathbf{A}$ has **full row rank**. If $\text{rank}(\mathbf{A}) = n$, then $\mathbf{A}$ has **full column rank**. If $\text{rank}(\mathbf{A}) < \min(m, n)$, then $\mathbf{A}$ is **rank deficient**.

The rank of a matrix is readily found in MATLAB by using the **rank** command.

A.5. ORTHOGONALITY AND THE DOT PRODUCT

Definition A.17 Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in R^n. The **dot product** of $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n. \tag{A.60}$$

Definition A.18 Let $\mathbf{x}$ be a vector in R^n. The **2-norm** or **Euclidean length** of $\mathbf{x}$ is

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{A.61}$$

Later we will introduce two other ways of measuring the "length" of a vector. The subscript 2 is used to distinguish this 2-norm from the other norms.

You may be familiar with an alternative definition of the dot product in which $\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos(\theta)$, where θ is the angle between the two vectors. The two definitions are equivalent. To see this, consider a triangle with sides $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x} - \mathbf{y}$. See Figure A.1.

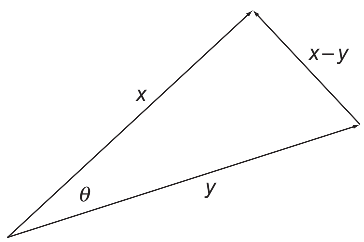

Figure A.1 Relationship between the dot product and the angle between two vectors.

The angle between sides $\mathbf{x}$ and $\mathbf{y}$ is θ. By the law of cosines,

$$\|\mathbf{x}-\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos(\theta)$$
$$(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y}) = \mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} - 2\|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos(\theta)$$
$$\mathbf{x}^T\mathbf{x} - 2\mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{y} = \mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} - 2\|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos(\theta)$$
$$-2\mathbf{x}^T\mathbf{y} = -2\|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos(\theta)$$
$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos(\theta).$$

We can also use this formula to compute the angle between two vectors:

$$\theta = \cos^{-1}\left(\frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\right). \tag{A.62}$$

Definition A.19 Two vectors $\mathbf{x}$ and $\mathbf{y}$ in R^n are **orthogonal**, or equivalently, **perpendicular** (written $\mathbf{x} \perp \mathbf{y}$), if $\mathbf{x}^T\mathbf{y} = 0$.

Definition A.20 A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is **orthogonal** if each pair of vectors in the set is orthogonal.

Definition A.21 Two subspaces V and W of R^n are **orthogonal** if every vector in V is perpendicular to every vector in W.

If $\mathbf{x}$ is in $N(\mathbf{A})$, then $\mathbf{A}\mathbf{x} = \mathbf{0}$. Since each element of the product $\mathbf{A}\mathbf{x}$ can be obtained by taking the dot product of a row of $\mathbf{A}$ and $\mathbf{x}$, $\mathbf{x}$ is perpendicular to each row of $\mathbf{A}$. Since $\mathbf{x}$ is perpendicular to all of the columns of $\mathbf{A}^T$, it is perpendicular to $R(\mathbf{A}^T)$. We have the following theorem.

Theorem A.5 *Let $\mathbf{A}$ be an m by n matrix. Then*

$$N(\mathbf{A}) \perp R(\mathbf{A}^T). \tag{A.63}$$

Furthermore,

$$N(\mathbf{A}) + R(\mathbf{A}^T) = R^n. \tag{A.64}$$

That is, any vector $\mathbf{x}$ in R^n can be written uniquely as $\mathbf{x} = \mathbf{p} + \mathbf{q}$, where $\mathbf{p}$ is in $N(\mathbf{A})$ and $\mathbf{q}$ is in $R(\mathbf{A}^T)$.

Definition A.22 A basis in which the basis vectors are orthogonal is an **orthogonal basis**. A basis in which the basis vectors are orthogonal and have length 1 is an **orthonormal basis**.

Definition A.23 An n by n matrix $\mathbf{Q}$ is **orthogonal** if the columns of $\mathbf{Q}$ are orthogonal and each column of $\mathbf{Q}$ has length 1.

With the requirement that the columns of an orthogonal matrix have length 1, using the term "orthonormal" would make logical sense. However, the definition of "orthogonal" given here is standard and we will not try to change standard usage.

Orthogonal matrices have a number of useful properties.

Theorem A.6 *If* $\mathbf{Q}$ *is an orthogonal matrix, then:*

1. $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$. *In other words,* $\mathbf{Q}^{-1} = \mathbf{Q}^T$.
2. *For any vector* $\mathbf{x}$ *in* R^n, $\|\mathbf{Q}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$.
3. *For any two vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *in* R^n, $\mathbf{x}^T\mathbf{y} = (\mathbf{Q}\mathbf{x})^T(\mathbf{Q}\mathbf{y})$.

A problem that we will often encounter in practice is projecting a vector $\mathbf{x}$ onto another vector $\mathbf{y}$ or onto a subspace W to obtain a projected vector $\mathbf{p}$. See Figure A.2. We know that

$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|_2\|\mathbf{y}\|_2 \cos(\theta), \tag{A.65}$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. Also,

$$\cos(\theta) = \frac{\|\mathbf{p}\|_2}{\|\mathbf{x}\|_2}. \tag{A.66}$$

Thus

$$\|\mathbf{p}\|_2 = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|_2}. \tag{A.67}$$

Since $\mathbf{p}$ points in the same direction as $\mathbf{y}$,

$$\mathbf{p} = \mathrm{proj}_{\mathbf{y}}\mathbf{x} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}. \tag{A.68}$$

The vector $\mathbf{p}$ is called the **orthogonal projection** or simply the **projection** of $\mathbf{x}$ onto $\mathbf{y}$.

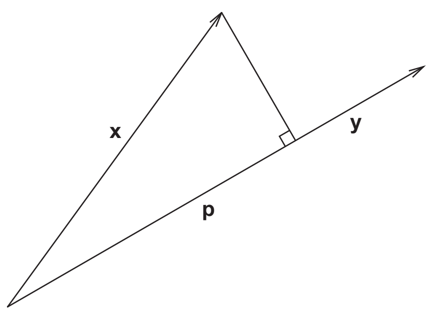

Figure A.2 The orthogonal projection of $\mathbf{x}$ onto $\mathbf{y}$.

Similarly, if W is a subspace of R^n with an orthogonal basis $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p$, then the **orthogonal projection of x onto** W is

$$\mathbf{p} = \mathrm{proj}_W \mathbf{x} = \frac{\mathbf{x}^T\mathbf{w}_1}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 + \frac{\mathbf{x}^T\mathbf{w}_2}{\mathbf{w}_2^T\mathbf{w}_2}\mathbf{w}_2 + \cdots + \frac{\mathbf{x}^T\mathbf{w}_p}{\mathbf{w}_p^T\mathbf{w}_p}\mathbf{w}_p. \tag{A.69}$$

Note that this equation can be simplified considerably if the orthogonal basis vectors are also orthonormal. In that case, $\mathbf{w}_1^T\mathbf{w}_1, \mathbf{w}_2^T\mathbf{w}_2, \ldots, \mathbf{w}_p^T\mathbf{w}_p$ are all 1.

It is inconvenient that the projection formula requires an orthogonal basis. The **Gram-Schmidt orthogonalization process** can be used to turn any basis for a subspace of R^n into an orthogonal basis. We begin with a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. The process recursively constructs an orthogonal basis by taking each vector in the original basis and then subtracting off its projection on the space spanned by the previous vectors. The formulas are

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_1^T\mathbf{v}_2}{\mathbf{v}_1^T\mathbf{v}_1}\mathbf{v}_1 = \mathbf{v}_2 - \frac{\mathbf{w}_1^T\mathbf{v}_2}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 \\
&\vdots \\
\mathbf{w}_p &= \mathbf{v}_p - \frac{\mathbf{w}_1^T\mathbf{v}_p}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 - \cdots - \frac{\mathbf{w}_p^T\mathbf{v}_p}{\mathbf{w}_p^T\mathbf{w}_p}\mathbf{w}_p.
\end{aligned} \tag{A.70}$$

Unfortunately, the Gram-Schmidt process is numerically unstable when applied to large bases. In MATLAB the command **orth** provides a numerically stable way to produce an orthogonal basis from a nonorthogonal basis. An important property of orthogonal projection is that the projection of $\mathbf{x}$ onto W is the point in W which is closest to $\mathbf{x}$. In the special case that $\mathbf{x}$ is in W, the projection of $\mathbf{x}$ onto W is $\mathbf{x}$.

Given an inconsistent system of equations $\mathbf{Ax} = \mathbf{b}$, it is often desirable to find an approximate solution. A natural measure of the quality of an approximate solution is the norm of the difference between $\mathbf{Ax}$ and $\mathbf{b}$, $\|\mathbf{Ax} - \mathbf{b}\|$. A solution that minimizes the 2-norm, $\|\mathbf{Ax} - \mathbf{b}\|_2$, is called a **least squares solution**, because it minimizes the sum of the squares of the residuals.

The least squares solution can be obtained by projecting $\mathbf{b}$ onto $R(\mathbf{A})$. This calculation requires us to first find an orthogonal basis for $R(\mathbf{A})$. There is an alternative approach that does not require finding an orthogonal basis. Let

$$\mathbf{Ax}_{ls} = \mathrm{proj}_{R(\mathbf{A})}\mathbf{b}. \tag{A.71}$$

Then, the difference between the projection (A.71) and $\mathbf{b}$, $\mathbf{Ax}_{ls} - \mathbf{b}$, will be perpendicular to $R(\mathbf{A})$ (Figure A.3). This orthogonality means that each of the columns of $\mathbf{A}$ will be orthogonal to $\mathbf{Ax}_{ls} - \mathbf{b}$. Thus

$$\mathbf{A}^T(\mathbf{Ax}_{ls} - \mathbf{b}) = \mathbf{0} \tag{A.72}$$

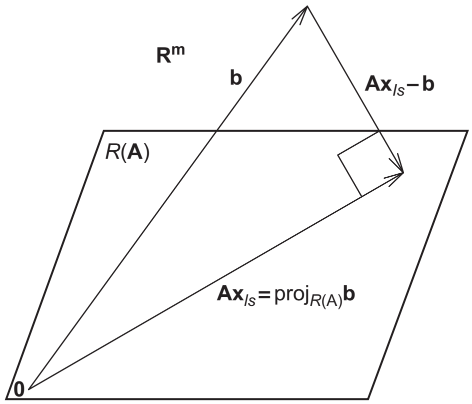

Figure A.3 Geometric conceptualization of the least squares solution to $\mathbf{Ax} = \mathbf{b}$. $\mathbf{b}$ generally lies in R^m, but $R(\mathbf{A})$ is generally a subspace of R^m. The least squares solution $\mathbf{x}_{ls}$ minimizes $\|\mathbf{Ax} - \mathbf{b}\|_2$.

or

$$\mathbf{A}^T\mathbf{A}\mathbf{x}_{ls} = \mathbf{A}^T\mathbf{b}. \tag{A.73}$$

This last system of equations is referred to as the **normal equations** for the least squares problem. It can be shown that if the columns of $\mathbf{A}$ are linearly independent, then the normal equations have exactly one solution for $\mathbf{x}_{ls}$ and this solution minimizes the sum of squared residuals [95].

A.6. EIGENVALUES AND EIGENVECTORS

Definition A.24 An n by n matrix $\mathbf{A}$ has an eigenvalue λ with an associated eigenvector $\mathbf{x}$ if $\mathbf{x}$ is not $\mathbf{0}$, and

$$\mathbf{Ax} = \lambda\mathbf{x}. \tag{A.74}$$

To find eigenvalues and eigenvectors, we rewrite the eigenvector equation (A.74) as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}. \tag{A.75}$$

To find nonzero eigenvectors, the matrix $\mathbf{A} - \lambda\mathbf{I}$ must be singular. This leads to the **characteristic equation**,

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{A.76}$$

where det denotes the determinant. For small matrices (e.g., 2 by 2 or 3 by 3), it is relatively easy to solve (A.76) to find the eigenvalues. The eigenvalues can then be substituted into (A.75), and the resulting system can then be solved to find corresponding eigenvectors. Note that the eigenvalues can, in general, be complex. For larger matrices, solving the characteristic equation becomes impractical and more sophisticated numerical methods are used. The MATLAB command **eig** can be used to find eigenvalues and eigenvectors of a matrix.

Suppose that we can find a set of n linearly independent eigenvectors, $\mathbf{v}_i$, of an n by n matrix $\mathbf{A}$ with associated eigenvalues λ_i. These eigenvectors form a basis for R^n. We can use the eigenvectors to **diagonalize** the matrix as

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \tag{A.77}$$

where

$$\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}, \tag{A.78}$$

and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues

$$\Lambda_{ii} = \lambda_i. \tag{A.79}$$

To see that this works, simply compute $\mathbf{AP}$:

$$\begin{aligned} \mathbf{AP} &= \mathbf{A} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \ldots & \lambda_n\mathbf{v}_n \end{bmatrix} \\ &= \mathbf{P}\mathbf{\Lambda}. \end{aligned}$$

Thus, $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$. Not all matrices are diagonalizable, because not all matrices have n linearly independent eigenvectors. However, there is an important special case in which matrices can always be diagonalized.

Theorem A.7 *If* $\mathbf{A}$ *is a real symmetric matrix, then* $\mathbf{A}$ *can be written as*

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \tag{A.80}$$

where $\mathbf{Q}$ *is a real orthogonal matrix of eigenvectors of* $\mathbf{A}$*, and* $\mathbf{\Lambda}$ *is a real diagonal matrix of the eigenvalues of* $\mathbf{A}$.

This **orthogonal diagonalization** of a real symmetric matrix $\mathbf{A}$ will be useful later on when we consider orthogonal factorizations of general matrices.

The eigenvalues of symmetric matrices are particularly important in the analysis of quadratic forms.

Definition A.25 A **quadratic form** is a function of the form

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}, \tag{A.81}$$

where $\mathbf{A}$ is a symmetric $n \times n$ matrix. The quadratic form $f(\mathbf{x})$ is **positive definite** if $f(x) \geq 0$ for all $\mathbf{x}$ and $f(\mathbf{x}) = 0$ only when $\mathbf{x} = \mathbf{0}$. The quadratic form is **positive**

semidefinite if $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. Similarly, a symmetric matrix $\mathbf{A}$ is positive definite if the associated quadratic form $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ is positive definite. The quadratic form is **negative semidefinite** if $-f(\mathbf{x})$ is positive semidefinite. If $f(\mathbf{x})$ is neither positive semidefinite nor negative semidefinite, then $f(\mathbf{x})$ is **indefinite**.

Positive definite quadratic forms have an important application in analytic geometry. Let $\mathbf{A}$ be a symmetric and positive definite matrix. Then the region defined by the inequality

$$(\mathbf{x}-\mathbf{c})^T\mathbf{A}(\mathbf{x}-\mathbf{c}) \leq \delta \tag{A.82}$$

is an ellipsoidal volume, with its center at $\mathbf{c}$. We can diagonalize $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \tag{A.83}$$

where the columns of $\mathbf{P}$ are normalized eigenvectors of $\mathbf{A}$, and $\mathbf{\Lambda}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{A}$. It can be shown that the ith eigenvector of $\mathbf{A}$ points in the direction of the ith semimajor axis of the ellipsoid, and the length of the ith semimajor axis is given by $\sqrt{\delta/\lambda_i}$ [95].

An important connection between positive semidefinite matrices and eigenvalues is the following theorem.

Theorem A.8 *A symmetric matrix* $\mathbf{A}$ *is positive semidefinite if and only if its eigenvalues are greater than or equal to* 0. $\mathbf{A}$ *is positive definite if and only if its eigenvalues are greater than* 0.

This provides a convenient way to check whether a symmetric matrix is positive semidefinite or positive definite.

The **Cholesky factorization** provides another way to determine whether a symmetric matrix is positive definite.

Theorem A.9 *Let* $\mathbf{A}$ *be an an* n *by* n *positive definite and symmetric matrix. Then* $\mathbf{A}$ *can be written uniquely as*

$$\mathbf{A} = \mathbf{R}^T\mathbf{R} = \mathbf{L}\mathbf{L}^T, \tag{A.84}$$

where $\mathbf{R}$ *is a nonsingular upper-triangular matrix and* $\mathbf{L} = \mathbf{R}^T$ *is a nonsingular lower-triangular matrix. Note that* $\mathbf{A}$ *can be factored in this way if and only if it is positive definite.*

The MATLAB command **chol** can be used to compute the Cholesky factorization.

A.7. VECTOR AND MATRIX NORMS

Although the conventional Euclidean length (A.61) is most commonly used, there are alternative ways to measure the length of a vector.

Definition A.26 Any measure of vector length satisfying the following four conditions is called a **norm**.

1. For any vector $\mathbf{x}$, $\|\mathbf{x}\| \geq 0$.
2. For any vector $\mathbf{x}$ and any scalar s, $\|s\mathbf{x}\| = |s|\,\|\mathbf{x}\|$.
3. For any vectors $\mathbf{x}$ and $\mathbf{y}$, $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.
4. $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

If $\|\,\|$ satisfies conditions 1, 2, and 3, but does not satisfy condition 4, then $\|\,\|$ is called a **seminorm**.

Definition A.27 The p**-norm** of a vector in R^n is defined for $p \geq 1$ by

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}. \tag{A.85}$$

It can be shown that for any $p \geq 1$, (A.85) satisfies the conditions of Definition A.26 [53]. The conventional Euclidean length is just the 2-norm, but two other p-norms are also commonly used. The **1-norm** is the sum of the absolute values of the elements in $\mathbf{x}$. The ∞**-norm** is obtained by taking the limit as p goes to infinity. The ∞-norm is the maximum of the absolute values of the elements in $\mathbf{x}$. The MATLAB command **norm** can be used to compute the norm of a vector, and has options for the 1, 2, and infinity norms.

The 2-norm is particularly important because of its natural connection with dot products and projections. The projection of a vector onto a subspace is the point in the subspace that is closest to the vector as measured by the 2-norm. We have also seen in (A.73) that the problem of minimizing $\|\mathbf{Ax}-\mathbf{b}\|_2$ can be solved by computing projections or by using the normal equations. In fact, the 2-norm can be tied directly to the dot product by the formula

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}}. \tag{A.86}$$

The 1- and ∞-norms can also be useful in finding approximate solutions to overdetermined linear systems of equations. To minimize the maximum of the residuals, we minimize $\|\mathbf{Ax}-\mathbf{b}\|_\infty$. To minimize the sum of the absolute values of the residuals, we minimize $\|\mathbf{Ax}-\mathbf{b}\|_1$. Unfortunately, these minimization problems are generally more difficult to solve than least squares problems.

Definition A.28 Any measure of the size or length of an m by n matrix that satisfies the following five properties can be used as a **matrix norm**.

1. For any matrix $\mathbf{A}$, $\|\mathbf{A}\| \geq 0$.
2. For any matrix $\mathbf{A}$ and any scalar s, $\|s\mathbf{A}\| = |s|\,\|\mathbf{A}\|$.
3. For any matrices $\mathbf{A}$ and $\mathbf{B}$, $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.
4. $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = \mathbf{0}$.
5. For any two matrices $\mathbf{A}$ and $\mathbf{B}$ of compatible sizes, $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\,\|\mathbf{B}\|$.

Definition A.29 The p**-norm** of a matrix $\mathbf{A}$ is

$$\|\mathbf{A}\|_p = \max_{\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_p, \tag{A.87}$$

where $\|\mathbf{x}\|_p$ and $\|\mathbf{A}\mathbf{x}\|_p$ are vector p-norms, while $\|\mathbf{A}\|_p$ is the matrix p-norm of $\mathbf{A}$.

Solving the maximization problem of (A.87) to determine a matrix p-norm could be extremely difficult. Fortunately, there are simpler formulas for the most commonly used matrix p-norms. See Exercises A.11, A.12, and C.4:

$$\|\mathbf{A}\|_1 = \max_j \sum_{i=1}^{m} |A_{i,j}| \tag{A.88}$$

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})} \tag{A.89}$$

$$\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^{n} |A_{i,j}|, \tag{A.90}$$

where $\lambda_{\max}(\mathbf{A}^T\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$.

Definition A.30 The **Frobenius norm** of an m by n matrix is given by

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{i,j}^2}. \tag{A.91}$$

Definition A.31 A matrix norm $\| \ \|_M$ and a vector norm $\| \ \|_V$ are **compatible** if

$$\|\mathbf{A}\mathbf{x}\|_V \leq \|\mathbf{A}\|_M \|\mathbf{x}\|_V. \tag{A.92}$$

The matrix p-norm is by its definition compatible with the vector p-norm from which it was derived. It can also be shown that the Frobenius norm of a matrix is compatible with the vector 2-norm [109]. Thus the Frobenius norm is often used with the vector 2-norm.

In practice, the Frobenius norm, 1-norm, and ∞-norm of a matrix are easy to compute, while the 2-norm of a matrix can be difficult to compute for large matrices. The MATLAB **norm** command has options for computing these matrix norms.

A.8. THE CONDITION NUMBER OF A LINEAR SYSTEM

Suppose that we want to solve a system of n equations in n variables

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{A.93}$$

Suppose further that because of measurement errors in $\mathbf{b}$, we actually solve

$$\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}. \tag{A.94}$$

Can we get a bound on $\|\mathbf{x} - \hat{\mathbf{x}}\|$ in terms of $\|\mathbf{b} - \hat{\mathbf{b}}\|$? Starting with (A.93) and (A.94), we have

$$\mathbf{A}(\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{b} - \hat{\mathbf{b}} \tag{A.95}$$

$$(\mathbf{x} - \hat{\mathbf{x}}) = \mathbf{A}^{-1}(\mathbf{b} - \hat{\mathbf{b}}) \tag{A.96}$$

$$\|\mathbf{x} - \hat{\mathbf{x}}\| = \|\mathbf{A}^{-1}(\mathbf{b} - \hat{\mathbf{b}})\| \tag{A.97}$$

$$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \|\mathbf{A}^{-1}\|\|\mathbf{b} - \hat{\mathbf{b}}\|. \tag{A.98}$$

This formula provides an absolute bound on the error in the solution. It is also worthwhile to compute a relative error bound:

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{b}\|} \leq \frac{\|\mathbf{A}^{-1}\|\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|} \tag{A.99}$$

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{A}\mathbf{x}\|} \leq \frac{\|\mathbf{A}^{-1}\|\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|} \tag{A.100}$$

$$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \|\mathbf{A}\mathbf{x}\|\|\mathbf{A}^{-1}\|\frac{\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|} \tag{A.101}$$

$$\|\mathbf{x} - \hat{\mathbf{x}}\| \leq \|\mathbf{A}\|\|\mathbf{x}\|\|\mathbf{A}^{-1}\|\frac{\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|} \tag{A.102}$$

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|}. \tag{A.103}$$

The relative error in $\mathbf{b}$ is measured by

$$\frac{\|\mathbf{b} - \hat{\mathbf{b}}\|}{\|\mathbf{b}\|}. \tag{A.104}$$

The relative error in $\mathbf{x}$ is measured by

$$\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|}{\|\mathbf{x}\|}. \tag{A.105}$$

The constant

$$\mathrm{cond}(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| \tag{A.106}$$

is called the **condition number** of $\mathbf{A}$.

Note that nothing that we did in the calculation of the condition number depends on which norm we used. The condition number can be computed using the 1-norm, 2-norm, ∞-norm, or Frobenius norm. The MATLAB **cond** command can be used to find the condition number of a matrix using each of these norms.

The condition number provides an upper bound on how inaccurate the solution to a system of equations might be because of errors in the right-hand side. In some cases, the condition number greatly overestimates the error in the solution. As a practical matter, it is wise to assume that the error is of roughly the size predicted by the condition number. In practice, double-precision floating point arithmetic only allows us to store numbers to about 16 digits of precision. If the condition number is greater than 10^{16}, then by the above inequality, there may be no accurate digits in the computer solution to the system of equations. Systems of equations with very large condition numbers are called **ill-conditioned**.

It is important to understand that ill-conditioning is a property of the system of equations and not of the algorithm used to solve the system of equations. Ill-conditioning cannot be fixed simply by using a better algorithm. Instead, we must either increase the precision of our floating point representation or find a different, better conditioned system of equations to solve.

A.9. THE QR FACTORIZATION

Although the theory of linear algebra can be developed using the reduced row echelon form, there is an alternative computational approach that works better in practice. The basic idea is to compute factorizations of matrices that involve orthogonal, diagonal, and upper-triangular matrices. This alternative approach leads to algorithms that can quickly compute accurate solutions to linear systems of equations and least squares problems. In this section we introduce the QR factorization, which is one of the most widely used orthogonal matrix factorizations. Another factorization, the singular value decomposition (SVD), is introduced in Chapter 3.

Theorem A.10 *Let* $\mathbf{A}$ *be an* m *by* n *matrix.* $\mathbf{A}$ *can be written as*

$$\mathbf{A} = \mathbf{QR}, \tag{A.107}$$

where $\mathbf{Q}$ *is an* m *by* m *orthogonal matrix, and* $\mathbf{R}$ *is an* m *by* n *upper-triangular matrix. This is called the* **QR factorization of A**.

The MATLAB command **qr** can be used to compute the QR factorization of a matrix. In a common situation, $\mathbf{A}$ will be an m by n matrix with $m > n$, and the rank of $\mathbf{A}$ will be n. In this case, we can write

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}, \tag{A.108}$$

where $\mathbf{R}_1$ is n by n, and

$$\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2], \tag{A.109}$$

where $\mathbf{Q}_1$ is m by n and $\mathbf{Q}_2$ is m by $m - n$. In this case, the QR factorization has some important properties.

Theorem A.11 *Let* $\mathbf{Q}$ *and* $\mathbf{R}$ *be the QR factorization of an* m *by* n *matrix* $\mathbf{A}$ *with* $m > n$ *and rank*$(\mathbf{A}) = n$. *Then,*

1. *The columns of* $\mathbf{Q}_1$ *are an orthonormal basis for* $R(\mathbf{A})$.
2. *The columns of* $\mathbf{Q}_2$ *are an orthonormal basis for* $N(\mathbf{A}^T)$.
3. *The matrix* $\mathbf{R}_1$ *is nonsingular.*

Now, suppose that we want to solve the least squares problem,

$$\min \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2. \tag{A.110}$$

Since multiplying a vector by an orthogonal matrix does not change its length, this is equivalent to

$$\min \|\mathbf{Q}^T(\mathbf{A}\mathbf{x} - \mathbf{b})\|_2. \tag{A.111}$$

But

$$\mathbf{Q}^T\mathbf{A} = \mathbf{Q}^T\mathbf{Q}\mathbf{R} = \mathbf{R}. \tag{A.112}$$

So, we have

$$\min \|\mathbf{R}\mathbf{x} - \mathbf{Q}^T\mathbf{b}\|_2 \tag{A.113}$$

or

$$\min \left\| \begin{matrix} \mathbf{R}_1\mathbf{x} - \mathbf{Q}_1^T\mathbf{b} \\ \mathbf{0}\mathbf{x} - \mathbf{Q}_2^T\mathbf{b} \end{matrix} \right\|_2 . \tag{A.114}$$

Whatever value of $\mathbf{x}$ we pick, we will probably end up with nonzero residuals because of the $\mathbf{0}\mathbf{x} - \mathbf{Q}_2^T\mathbf{b}$ part of the least squares problem. We cannot minimize the norm of this part of the vector. However, we can find an $\mathbf{x}$ that exactly solves $\mathbf{R}_1\mathbf{x} = \mathbf{Q}_1^T\mathbf{b}$. Thus, we can minimize the least squares problem by solving the square system of equations,

$$\mathbf{R}_1\mathbf{x} = \mathbf{Q}_1^T\mathbf{b}. \tag{A.115}$$

The advantage of solving this system of equations instead of the normal equations (A.73) is that the normal equations are typically much more badly conditioned than (A.115).

A.10. COMPLEX MATRICES AND VECTORS

Although nearly all of the mathematics in this textbook involves real numbers, complex numbers do appear in Chapter 8 when we consider the Fourier transform. We assume that the reader is already familiar with arithmetic involving complex numbers including addition, subtraction, multiplication, division, and complex exponentials. Most theorems of linear algebra extend trivially from real to complex vectors and matrices. In this section we briefly discuss our notation and some important differences between the real and complex cases.

Given a complex number $z = a + b\imath$, where $\imath$ is the $\sqrt{-1}$, the **complex conjugate** of z is $z^* = a - b\imath$. Note that the absolute value of z is

$$|z| = \sqrt{a^2 + b^2} = \sqrt{z^* z}. \tag{A.116}$$

The main difference between linear algebra on real vectors and complex vectors is in the definition of the dot product of two vectors. We define the dot product of two complex vectors $\mathbf{x}$ and $\mathbf{y}$ to be

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^{*^T} \mathbf{y}. \tag{A.117}$$

The advantage of this definition is that

$$\mathbf{x}^{*^T} \mathbf{x} = \sum_{k=1}^{n} x_k^* x_k = \sum_{k=1}^{n} |x_k|^2. \tag{A.118}$$

Thus we can then define the 2-norm of a complex vector by

$$\|x\|_2 = \sqrt{\mathbf{x}^{*^T} \mathbf{x}}. \tag{A.119}$$

The combination of taking the complex conjugate and transpose, called the **Hermitian transpose**, occurs so frequently that we denote this by

$$\mathbf{x}^H = \mathbf{x}^{*^T}. \tag{A.120}$$

Note that for a real vector, $\mathbf{x}$, the conjugate is simply $\mathbf{x}^* = \mathbf{x}$, so $\mathbf{x}^H = \mathbf{x}^T$. In MATLAB, the apostrophe denotes the Hermitian transpose.

In general, you will almost never go wrong by using the Hermitian transpose in any linear algebra computation involving complex numbers that would normally involve a transpose when working with real vectors and matrices. For example, if we want to minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$, where $\mathbf{G}$, $\mathbf{m}$, and $\mathbf{d}$ are complex, we can solve the normal equations,

$$\mathbf{G}^H \mathbf{G} \mathbf{m} = \mathbf{G}^H \mathbf{d}. \tag{A.121}$$

A.11. LINEAR ALGEBRA IN SPACES OF FUNCTIONS

So far, we have considered only vectors in R^n. The concepts of linear algebra can be extended to other contexts. In general, as long as the objects that we want to consider can be multiplied by scalars and added together, and as long as they obey the laws of vector algebra, then we have a **vector space** in which we can practice linear algebra. If we can also define a vector product similar to the dot product, then we have what is called an **inner product space**, and we can define orthogonality, projections, and the 2-norm.

There are many different vector spaces used in various areas of science and mathematics. For our work in inverse problems, a very commonly used vector space is the space of functions defined on an interval $[a, b]$.

Multiplying a scalar times a function or adding two functions together clearly produces another function. In this space, the function $z(x) = 0$ takes the place of the $\mathbf{0}$ vector, since $f(x) + z(x) = f(x)$. Two functions $f(x)$ and $g(x)$ are linearly independent if the only solution to

$$c_1\, f(x) + c_2\, g(x) = z(x) \tag{A.122}$$

is $c_1 = c_2 = 0$.

We can define the dot product of two functions f and g to be

$$f \cdot g = \int_a^b f(x)g(x)\ dx. \tag{A.123}$$

Another commonly used notation for this dot product or **inner product** of f and g is

$$f \cdot g = \langle f, g \rangle. \tag{A.124}$$

It is easy to show that this inner product has all of the algebraic properties of the dot product of two vectors in R^n. A more important motivation for defining the dot product in this way is that it leads to a useful definition of the 2-norm of a function. Following our earlier formula that $\|x\|_2 = \sqrt{x^T x}$, we have

$$\|f\|_2 = \sqrt{\int_a^b f(x)^2\ dx}. \tag{A.125}$$

Using this definition, the distance between two functions f and g is

$$\|f - g\|_2 = \sqrt{\int_a^b (f(x) - g(x))^2\ dx}. \tag{A.126}$$

This measure is obviously 0 when $f(x) = g(x)$ everywhere, but can also be 0 when $f(x)$ and $g(x)$ differ at a finite or countably infinite set of points. The measure is only nonzero if $f(x)$ and $g(x)$ differ on an interval.

Using this inner product and norm, we can reconstruct the theory of linear algebra from R^n in our space of functions. This includes the concepts of orthogonality, projections, norms, and least squares solutions.

Definition A.32 Given a collection of functions $f_1(x), f_2(x), \ldots, f_m(x)$ in an inner product space, the **Gram matrix** of the functions is the $m \times m$ matrix $\mathbf{\Gamma}$, whose elements are given by

$$\mathbf{\Gamma}_{i,j} = f_i \cdot f_j. \tag{A.127}$$

The Gram matrix has several important properties. It is symmetric and positive semidefinite. If the functions are linearly independent, then the Gram matrix is also positive definite. Furthermore, the rank of $\mathbf{\Gamma}$ is equal to the size of the largest linearly independent subset of the functions $f_1(x), \ldots, f_m(x)$.

A.12. EXERCISES

1. Let $\mathbf{A}$ be an m by n matrix with n pivot columns in its RREF. Can the system of equations $\mathbf{Ax} = \mathbf{b}$ have infinitely many solutions?
2. If $\mathbf{C} = \mathbf{AB}$ is a 5 by 4 matrix, then how many rows does $\mathbf{A}$ have? How many columns does $\mathbf{B}$ have? Can you say anything about the number of columns in $\mathbf{A}$?
3. Suppose that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are three vectors in R^3 and that $\mathbf{v}_3 = -2\mathbf{v}_1 + 3\mathbf{v}_2$. Are the vectors linearly dependent or linearly independent?
4. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 1 & 3 \\ 4 & 6 & 7 & 11 \end{bmatrix}. \tag{A.128}$$

   Find bases for $N(\mathbf{A})$, $R(\mathbf{A})$, $N(\mathbf{A}^T)$, and $R(\mathbf{A}^T)$. What are the dimensions of the four subspaces?
5. Let $\mathbf{A}$ be an n by n matrix such that $\mathbf{A}^{-1}$ exists. What are $N(\mathbf{A})$, $R(\mathbf{A})$, $N(\mathbf{A}^T)$, and $R(\mathbf{A}^T)$?
6. Let $\mathbf{A}$ be any 9 by 6 matrix. If the dimension of the null space of $\mathbf{A}$ is 5, then what is the dimension of $R(\mathbf{A})$? What is the dimension of $R(\mathbf{A}^T)$? What is the rank of $\mathbf{A}$?
7. Suppose that a nonhomogeneous system of equations with four equations and six unknowns has a solution with two free variables. Is it possible to change the right-hand side of the system of equations so that the modified system of equations has no solutions?

8. Let W be the set of vectors $\mathbf{x}$ in R^4 such that $x_1 x_2 = 0$. Is W a subspace of R^4?

9. Let $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ be a set of three nonzero orthogonal vectors. Show that the vectors are also linearly independent.

10. Show that if $\mathbf{x} \perp \mathbf{y}$, then

$$\|\mathbf{x}+\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2. \tag{A.129}$$

11. In this exercise, we will derive the formula (A.88) for the 1-norm of a matrix. Begin with the optimization problem

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1. \tag{A.130}$$

(a) Show that if $\|\mathbf{x}\|_1 = 1$, then

$$\|\mathbf{A}\mathbf{x}\|_1 \leq \max_j \sum_{i=1}^{m} |A_{i,j}|. \tag{A.131}$$

(b) Find a vector $\mathbf{x}$ such that $\|\mathbf{x}\|_1 = 1$, and

$$\|\mathbf{A}\mathbf{x}\|_1 = \max_j \sum_{i=1}^{m} |A_{i,j}|. \tag{A.132}$$

(c) Conclude that

$$\|\mathbf{A}\|_1 = \max_{\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1 = \max_j \sum_{i=1}^{m} |A_{i,j}|. \tag{A.133}$$

12. Derive the formula (A.90) for the infinity norm of a matrix.

13. Let $\mathbf{A}$ be an m by n matrix.

(a) Show that $\mathbf{A}^T\mathbf{A}$ is symmetric.

(b) Show that $\mathbf{A}^T\mathbf{A}$ is positive semidefinite. Hint: Use the definition of positive semidefinite rather than trying to compute eigenvalues.

(c) Show that if $\text{rank}(\mathbf{A}) = n$, then the only solution to $\mathbf{A}\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.

(d) Use part c to show that if $\text{rank}(\mathbf{A}) = n$, then $\mathbf{A}^T\mathbf{A}$ is positive definite.

(e) Use part d to show that if $\text{rank}(\mathbf{A}) = n$, then $\mathbf{A}^T\mathbf{A}$ is nonsingular.

(f) Show that $N(\mathbf{A}^T\mathbf{A}) = N(\mathbf{A})$.

14. Show that

$$\text{cond}(\mathbf{AB}) \leq \text{cond}(\mathbf{A})\text{cond}(\mathbf{B}). \tag{A.134}$$

15. Let $\mathbf{A}$ be a symmetric and positive definite matrix with Cholesky factorization,

$$\mathbf{A} = \mathbf{R}^T\mathbf{R}. \tag{A.135}$$

Show how the Cholesky factorization can be used to solve $\mathbf{Ax} = \mathbf{b}$ by solving two systems of equations, each of which has $\mathbf{R}$ or $\mathbf{R}^T$ as its matrix.

16. Let $P_3[0, 1]$ be the space of polynomials of degree less than or equal to 3 on the interval $[0, 1]$. The polynomials $p_1(x) = 1$, $p_2(x) = x$, $p_3(x) = x^2$, and $p_4(x) = x^3$ form a basis for $P_3[0, 1]$, but they are not orthogonal with respect to the inner product,

$$f \cdot g = \int_0^1 f(x)g(x)\ dx. \tag{A.136}$$

Use the Gram-Schmidt orthogonalization process to construct an orthogonal basis for $P_3[0, 1]$. Once you have your basis, use it to find the third-degree polynomial that best approximates $f(x) = e^{-x}$ on the interval $[0, 1]$ in the least squares sense.

A.13. NOTES AND FURTHER READING

Much of this material is typically covered in sophomore-level linear algebra courses, and there are an enormous number of textbooks at this level. One good introductory linear algebra textbook is [95]. At a slightly more advanced level, [109] and [152] are both excellent. The book by Strang and Borre [153] reviews linear algebra in the context of geodetic problems.

Fast and accurate algorithms for linear algebra computations are a somewhat more advanced and specialized topic. A classic reference is [53]. Other good books on this topic include [38] and [164].

The extension of linear algebra to spaces of functions is a topic in the subject of functional analysis. Unfortunately, most textbooks on functional analysis assume that the reader has a considerable mathematical background. One book that is reasonably accessible to readers with limited mathematical backgrounds is [102].