Matrices and Vector Spaces:

A brief introduction to linear algebra

with help from Eileen Melville

originally inspired by text: Linear Algebra and its Applications (5th ed), D.C. Lay, S.R. Lay, J.J. McDonald

Contents

1 Solving Linear Systems

In the first chapter we show how to find the solution set of a system of linear equations.

1.1 Systems of Linear Equations

*A **linear equation** is that the sum of coefficients times variables is some value; for example, $3x - y = 7$. A **linear system** is a collection of linear equations. A **solution** of the system satisfies all the equations. A system is **consistent** if it has at least one solution. The **solution set** is all solutions.*

Consider, for example, a linear equation in the two variables x and y. This represents a line in the plane. If there are two such equations, then the corresponding lines can either intersect, be the same line, or be parallel and different. In the first case the system has a unique solution, in the second case any point on the line is a solution, and in the third case they have no point in common. Here are some examples:

EXAMPLE.

$2x + 3y = 7$	$2x + 3y = 7$	$2x + 3y = 7$
$x - y = 1$	$2x + 3y = 11$	$4x + 6y = 14$
Solution: $x = 2$, $y = 1$	*No Solution*	*Infinite solutions*

We will show that the above behavior captures all possibilities:

Fact 1.1 *There are exactly three possibilities for the solution set of a linear system: no solution, unique solution, or infinitely many solutions.*

A Word About Proofs. Every Fact in this course has a proof. We sometimes give the proof, sometimes sketch the highlights, and sometimes just skip it. Mathematics rests on proof. Proof provides a guarantee that the Fact is true. Proofs use logic, calculation, previous facts, and definitions.

1.2 Triangular Systems

A special case of a linear system is what is called a triangular system.

> *A linear system is **triangular** if the first equation has only one variable, the second equation has only the first variable and another, and so on.*

Triangular linear systems can be solved by an algorithm/process known as **back-substitution**:

ALGOR Back Substitution.

- *Solve the first equation for its variable.*
- *Substitute the result into the second equation, and solve for its remaining variable.*
- *Repeat.*

Example. *Consider the system*

$$\begin{aligned} 2x_1 \quad\quad\quad\quad &= 6 \\ x_1 + x_2 \quad\quad &= 2 \\ -x_1 + 4x_2 + x_3 &= 19 \end{aligned}$$

The first equation implies that $x_1 = 3$. Substituting this into the second equation implies $x_2 = -1$. Substituting both these values into the third equation implies $x_3 = 26$. The solution is unique.

1.3 Matrices

To both represent and solve linear systems, it is convenient to use the notation and ideas of matrices. We will see in the rest of the course that matrices have many many other uses.

> *A **matrix** is a rectangular arrangement of numbers. Its **size** is the number of rows and columns. We use the terminology "$\boldsymbol{m \times n}$ matrix" to mean a matrix with m rows and n columns. (Always rows first.)*

A matrix can represent the left-hand-side of a linear system by recording the coefficients of each variable, one row for each equation. The **augmented** matrix is formed by adding the constants as the last column.

EXAMPLE. *The linear system given in the above example has the augmented matrix*

$$\left[\begin{array}{ccc|c} 2 & 0 & 0 & 6 \\ 1 & 1 & 0 & 2 \\ -1 & 4 & 1 & 19 \end{array}\right]$$

The augmented matrix is a 3×4 *matrix.*

1.4 Elementary Row Operations

Our approach to solving a linear system uses what are called elementary row operations. We apply these to bring the linear system to a simpler-looking system (such as one that is triangular) without changing the solution set.

The three **elementary row operations** we use are: replacement, interchange, and scaling:

- **Replacement**: One can replace a row by the sum of it and a multiple of another row. For example, replace the second row by the sum of it and 3 times the first row. This is often abbreviated to "add 3 times the first row to the second" or "$R2' = R2 + 3R1$" (where the prime means the new value).
- **Interchange**: One can interchange two rows. For example, swap the first and third rows.
- **Scaling**: One can scale a row by a nonzero factor. For example, multiply all entries in the second row by 5.

The crucial point is that these operations do not change the solution set. Hopefully this is obvious for the interchange and the scaling operations. For the replacement operation, note that any solution of the original system remains a solution to the new system. Further, each row operation is reversible. So by the same logic, any solution of the new system remains a solution if we revert to the original system. That is, any solution of the system after the replacement is a solution to the original system. We state this as a fact:

Fact 1.2 *Any row operation preserves the solution set.*

Two matrices are defined to be **row equivalent** if it is possible to get from one matrix to the other by elementary row operations. Note that, since the row operations are reversible, if matrix A is row equivalent to matrix B, then matrix B is also row equivalent to matrix A.

1.5 Echelon Forms

The overall approach to solving a linear system is to perform a series of row operations on the augmented matrix until the solution set is easy to obtain. There are two phases:

- The first phase brings it to a triangular system, or rather a generalization of that;
- The second phase solves for all the variables, or at least as many as one can.

The goal of the first phase of row operations is called echelon form; the goal of the second phase is called reduced row echelon form. These forms are defined by the unaugmented columns.

*An unaugmented matrix is defined to be in **echelon form** if:*

(A) for each row that is not completely zero, the leftmost nonzero entry has zeros below it and below its preceding zeroes; and

(B) the all-zero rows (if any) are at the bottom.

Echelon form is a generalization of triangular system, except that we list the equations in decreasing number of variables.

EXAMPLE. *An example of echelon form is:*

$$\begin{bmatrix} -7 & 5 & 4\frac{1}{3} & -10 \\ 0 & 0 & 1 & \frac{1}{5} \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

*An unaugmented matrix in echelon form is defined to be in **reduced row echelon form** if furthermore*

(C) the first nonzero entry in every row is a 1 and has zeroes above it.

Example. *An example of reduced row echelon form is:*

$$\begin{bmatrix} 0 & 1 & 5^\star & 0 & 0 & -1^\star \\ 0 & 0 & 0 & 1 & 0 & -\frac{1}{2}^\star \\ 0 & 0 & 0 & 0 & 1 & 43^\star \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

It remains in such form if any of the starred values are changed arbitrarily (even to 0*).*

It can be shown that every matrix is row-equivalent to a *unique* matrix in reduced row echelon form.

1.6 Row Reduction

We now present the **row reduction algorithm**. It is often called **Gaussian Elimination** or **Gauss-Jordan Elimination**.

*A **pivot** is the leftmost nonzero entry in a row.*

The first phase of the algorithm proceeds:

ALGOR Achieving Echelon Form.

≻ *Go to first nonzero column (but not the augmented column). If needed, interchange rows so top entry in that column is not zero.*

≻ *Add suitable multiples of the first row to create zeroes below pivot.*

≻ *Repeat, ignoring rows with pivots, until there is no nonzero row.*

Example. *For example, consider the matrix*

$$\left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 6 & -2 & -4 & 10 \end{array}\right]$$

The first nonzero column is the first column. The top entry is nonzero; so no interchange needed. The -1 *is the pivot. The second entry in the column is* 0*. So no work needed there. The third entry is* 6*. The key point is that if one adds* 6 *times the first row to the third row, in the third row the first entry will become zero. That is, one* *gets*

$$\left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 6 & -2 & -4 & 10 \end{array}\right] R_3' = R_3 + 6R_1 \quad \rightsquigarrow \quad \left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{array}\right]$$

Next we work on the second column. We pretend the first row is not there. The column has a nonzero where we need it in the second row: so 2 is the pivot. We just have to get a zero in the third row. This time the row replacement entails subtracting 5 *times the second row from the third row. Thus we get to echelon form:*

$$\left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{array}\right] R_3' = R_3 - 5R_2 \quad \rightsquigarrow \quad \left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 0 & 30 & -30 \end{array}\right]$$

We now present the second phase of the **row reduction algorithm**. The algorithm proceeds:

ALGOR ACHIEVING REDUCED ROW ECHELON FORM.

From right to left:

- ≻ *add suitable multiples of the rows to create zeroes above each pivot; and*
- ≻ *make each pivot* 1 *by scaling.*

(When doing this process by hand, it is sometimes easier to scale first and then zero-out the column, and sometimes it is easier to zero-out the column with the old row. It just depends. . .)

EXAMPLE. *We earlier reached an echelon form of*

$$\left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 0 & 30 & -30 \end{array}\right]$$

We start with the third column. The reduction continues with scaling the third row by dividing by 30*. This is followed by adding* 8 *times the new third row to the second row, and adding* 1 *times the new third row to the first row. Thus we get*

$$\left[\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 0 & 30 & -30 \end{array}\right] \begin{array}{l} R_1' = R_1 + R_3' \\ R_2' = R_2 + 8R_3' \\ R_3' = R_3/30 \end{array} \quad \rightsquigarrow \quad \left[\begin{array}{ccc|c} -1 & 2 & 0 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

There after we do the same for the second column

$$\left[\begin{array}{ccc|c} -1 & 2 & 0 & -1 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right] \begin{array}{l} R_1' = R_1 - 2R_2' \\ R_2' = R_2/2 \\ \end{array} \quad \rightsquigarrow \quad \left[\begin{array}{ccc|c} -1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

and finally, only scaling is needed in the first row to produce reduced row echelon form.

$$\left[\begin{array}{ccc|c} -1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right] \begin{array}{l} R_1' = -R_1 \\ \\ \end{array} \quad \rightsquigarrow \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{array}\right]$$

This system has the unique solution $x_1 = 1$, $x_2 = 0$, and $x_3 = -1$.

1.7 The Solution Set of a Linear Systems

We saw earlier that some linear systems have no solution, some a unique solution, and some infinitely many solutions. The next fact gives the rule to determine whether a linear system has a solution or not, that is, whether it is consistent or not:

Fact 1.3 *A linear system is consistent if and only if the echelon form of the augmented matrix has no row that is all zeroes except for a nonzero in the augmented column.*

We skip the full justification of this fact. We note at least that such a row creates a problem; for example, requiring that $0x + 0y = -5$ is impossible.

EXAMPLE. *In the following echelon forms, the first two correspond to systems that are consistent but the third does not.*

$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & 13 \\ 0 & 1 & -5 & 0 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & 0 & 2 & 13 \\ 0 & 1 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & 0 & 2 & 13 \\ 0 & 1 & -5 & 0 \\ 0 & 0 & 0 & 5 \end{array}\right]$$

Consider a matrix that represents a linear system. If the system is consistent, we can read off the solution set from the reduced row echelon form. We define a **basic** variable as one whose column has a pivot; otherwise it is a **free** variable. If every column has a pivot (equivalently every variable is basic), then the solution is unique.

But what happens in general?

ALGOR *The solution set is obtained by expressing each basic variable in terms of the free variables; this is called a* ***parametric description****.*

With the parametric description, we can think of the free variables as "parameters": each setting of the free variables yields one solution of the original linear system.

EXAMPLE. *Consider the following row reduced echelon form.*

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & -3 & -1 \\ 0 & 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The system is consistent. (There is no bad row.) Assume the variables are x_1, x_2, x_3, x_4*. Then the free variables are* x_2 *and* x_4*. So the system has solution set with parametric description:*

$$\begin{aligned} x_1 &= \quad -1 - 2x_2 + 3x_4 \\ x_3 &= \quad 5 - x_4 \end{aligned}$$

Note that this algorithm proves that Fact 1.1, which we stated on the first page, is true. If a consistent system has a free variable, then there are infinitely many solutions.

Practice

1.1. Solve the following triangular system.

$$\begin{aligned} x - y + z &= 7 \\ 2x - y &= 11 \\ 3y &= 42 \end{aligned}$$

1.2. Consider this matrix

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

Is this matrix B in echelon form? In reduced row echelon form? Explain.

1.3. Solve the following systems.

$$\begin{aligned} a+2b+3c &= 2 \\ 2a \quad -c &= 8 \\ -3a+7b+2c &= -19 \end{aligned} \qquad \begin{aligned} a+2b+3c &= 2 \\ 2a \quad -c &= 8 \\ 4a+4b+5c &= 11 \end{aligned} \qquad \begin{aligned} a+2b+3c &= 2 \\ 2a \quad -c &= 8 \\ 4a+4b+5c &= 12 \end{aligned}$$

Solutions to Practice Exercieses

1.1. In order we find $y = 14$, $x = 25/2$, $z = 17/2$

1.2. Yes to both. The first nonzero entry in each row is a 1, and it has zeroes above, below, and lower-left of it.

1.3. $\begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & 0 & -1 & 8 \\ -3 & 7 & 2 & -19 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -4 & -7 & 4 \\ 0 & 13 & 11 & -13 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -4 & -7 & 4 \\ 0 & 0 & -\frac{47}{4} & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & -4 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

Solution is $a = 4$, $b = -1$, $c = 0$

$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & 0 & -1 & 8 \\ 4 & 4 & 5 & 11 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -4 & -7 & 4 \\ 0 & -4 & -7 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -4 & -7 & 4 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

Inconsistent

$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & 0 & -1 & 8 \\ 4 & 4 & 5 & 12 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -4 & -7 & 4 \\ 0 & -4 & -7 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & -4 & -7 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\frac{1}{2} & 4 \\ 0 & 1 & \frac{7}{4} & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Parametric description: $a = 4 + c/2$, $b = -1 - 7c/4$, c free

2 Vectors and Linear Combinations

In this chapter we define vectors and operations involving vectors.

2.1 Vectors

You might have encountered vectors elsewhere. Some books define both a column vector and a row vector, but we will only use the column vector.

> *A (column)* ***vector*** *is a column of numbers.*

Thus we may think of the columns of a matrix as vectors. (Indeed, some books define vectors first and then define a matrix in this way.) We will in general use bold letters for vector variables, such as $\boldsymbol{x}$ and $\boldsymbol{v}$. To save space, we sometimes write the column vector $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ as $(3, 5)$. To be equal, two vectors must have the same size and the same entries in order.

Several vector operations can be defined.

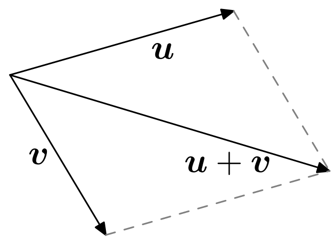
> *Vector* ***addition*** *is performed by adding the corresponding entries. In algebraic notation:*
>
> $$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$$
>
> *Similarly,* ***scalar multiplication*** *is performed by scaling each entry. In algebraic notation:*
>
> $$c \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} c\,u_1 \\ c\,u_2 \end{bmatrix}$$

Example.

$$3 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 4 \begin{bmatrix} -1 \\ 7 \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \end{bmatrix} + \begin{bmatrix} -4 \\ 28 \end{bmatrix} = \begin{bmatrix} 2 \\ 40 \end{bmatrix}$$

We use $\mathbb{R}^d$ for the set of all d-entry vectors whose entries are real numbers. (Complex numbers will make an appearance later.) One can associate vectors in $\mathbb{R}^d$ with the corresponding point in space to give geometric descriptions. For example, $\mathbb{R}^2$ is the 2-dimensional plane. The addition of two vectors in $\mathbb{R}^2$ can be interpreted geometrically:

for example, $(7,2)+(3,-5)=(10,-3)$ can be viewed as “if you go 7 blocks east and 2 blocks north, then 3 blocks east and 5 blocks south, then you are 10 blocks east and 3 blocks south of where you started. Equivalently, vector addition in $\mathbb{R}^2$ can be interpreted as the diagonal of the parallelogram they create:



These two vector operations obey standard properties of arithmetic. These properties include the commutative law (order of addition does not matter), associative law (brackets do not matter), and distributive laws (for example, $c(\boldsymbol{u}+\boldsymbol{v})=c\boldsymbol{u}+c\boldsymbol{v}$).

2.2 Linear Combinations and Spans

*A **linear combination** of vectors is formed by summing some multiple of each vector. The multipliers are called the **weights**.*

For example, $3\boldsymbol{u}+4\boldsymbol{v}$ is a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$.

*The **span** of a collection of vectors is the set of all possible linear combinations. If S is a set, we will denote its span by $Span(S)$.*

EXAMPLE. *The span of a single (nonzero) vector is a line. The span of two vectors is usually a plane.*

We define next the product of a matrix with a vector. We use the notation $A=[\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n]$ to mean that the columns of matrix A are the vectors $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n$.

*If A is an $m\times n$ matrix and vector $\boldsymbol{x}$ is in $\mathbb{R}^n$, then the **matrix-vector product** $A\boldsymbol{x}$ is defined to be the linear combination of the columns of A specified by $\boldsymbol{x}$. That is, if $A=[\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n]$ and $\boldsymbol{x}=(x_1,\ldots,x_n)$ then*

$$A\boldsymbol{x}=x_1\boldsymbol{a}_1+x_2\boldsymbol{a}_2+\ldots+x_n\boldsymbol{x}_n$$

EXAMPLE.

$$\begin{bmatrix} 2 & -1 \\ 4 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 5 \begin{bmatrix} -1 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \\ 47 \end{bmatrix}$$

Every linear system is equivalent to a matrix equation. From the definition of matrix-vector multiplication we immediately obtain the following fact:

Fact 2.1 *The matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution if and only if $\boldsymbol{b}$ is a linear combination of the columns of A.*

In particular: testing whether a vector $\boldsymbol{b}$ is in the span of some collection of vectors, is equivalent to asking whether the augmented matrix with those columns is consistent.

The next fact addresses the question of when a matrix equation for fixed A has a solution for **ALL** possible $\boldsymbol{b}$:

Fact 2.2 *For matrix A, the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b}$*
$\iff$ *the span of the columns of A is $\mathbb{R}^m$*
$\iff$ *A has a pivot in each row.*

Note that this fact says that three conditions are **equivalent**: either all three conditions hold or none of them hold.

PROOF. *The first and second conditions are equivalent, because the product $A\boldsymbol{x}$ is by definition a linear combination of the columns of A. That is, the first and second conditions are different ways of saying the same thing.*
The main part of the proof is to show that the first and third conditions are equivalent. Assume first that one can pivot in every row. This means that the echelon form cannot have a row that is all-zero outside the augmented column; thus the echelon form will always pass the test for consistent system, and the matrix equation is always consistent.
On the other hand, assume that one cannot pivot in every row. Then the echelon form has zeroes in the bottom row outside the augmented column. Thus one can choose the original constant in that row to ensure that in the augmented column there is a nonzero entry at the end. That is, one can create an inconsistent system by choosing $\boldsymbol{b}$ suitably, which has no solution.

2.3 Homogenous Systems and Parametric Vector Form

A ***homogeneous system*** *is one like* $A\boldsymbol{x} = \mathbf{0}$*. It always has at least the* ***trivial solution*** $\boldsymbol{x} = \mathbf{0}$.

Solving the homogeneous system is part of our approach to solving the general system. Indeed, one feature of vectors is that they provide another way to express the solution to a linear system.

ALGOR *The solution set to a general linear system can be written in* ***parametric vector form*** *as: one vector plus an arbitrary linear combination of vectors satisfying the corresponding homogeneous system.*

For example, the solution set to the homogeneous system could be a plane through the origin, while the solution set to a general linear system could be a plane shifted.

EXAMPLE. *Consider the following reduced row echelon form*

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & -3 & -1 \\ 0 & 0 & 1 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Before, we saw that x_2 *and* x_4 *were free variables, and wrote the solution as*

$$\begin{aligned} x_1 &= -1 - 2x_2 + 3x_4 \\ x_3 &= 5 - x_4 \end{aligned}$$

This can be reformulated in vector notation by adding the ("silly") equations that say the free variables equal themselves. Here this means

$$\begin{aligned} x_1 &= -1 - 2x_2 + 3x_4 \\ x_2 &= x_2 \\ x_3 &= 5 - x_4 \\ x_4 &= x_4 \end{aligned}$$

And thus this system has solution:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 5 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

2.4 Linear Independence of Vectors

We consider next a crucial, but slippery, concept.

A collection of vectors is ***linearly independent*** *if the only linear combination of them that equals* $\mathbf{0}$ *is the trivial combination (all weights zero). Otherwise it is said to be* ***linearly dependent****.*

Example. *The trio* $\{(0,1,1),(1,2,-3),(-2,4,14)\}$ *is linearly dependent since* $6(0,1,1) - 2(1,2,-3) - 1(-2,4,14) = (0,0,0)$. *On the other hand, the trio* $\{(1,0,0),(0,1,0),(0,0,1)\}$ *is linearly independent, as the only linear combination that produces the zero-vector is* 0 *of each vector.*

We express two important examples as a fact:

Fact 2.3 *(a) A pair of vectors is linearly dependent if and only if one vector is a multiple of the other.*
(b) If a set contains the zero vector, then it is linearly dependent.

More generally, a collection is linearly dependent if at least one vector in the collection can be written as a linear combination of the other vectors.

Example. *If vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are linearly independent, then they span a plane through the origin. Further, inserting* $\boldsymbol{w}$ *into the collection produces a linearly independent set if and only if* $\boldsymbol{w}$ *is not in* $Span(\{\boldsymbol{u},\boldsymbol{v}\})$.

We will use the concept of linear independence often in the course. The first application is to note that this concept captures when a homogenous system has a unique solution.

Fact 2.4 *The columns of matrix* A *are linearly independent*
$\iff$ $A\boldsymbol{x} = \mathbf{0}$ *has only the trivial solution*
$\iff$ *there is no free variable*

Proof. *The first and second condition are equivalent by the definitions of matrix-vector multiplication and linear independence. Our algorithm for solving linear systems shows that the second and third condition are equivalent.*

One consequence of the above fact, is that if there are more columns than rows, then the columns must be linearly dependent.

Practice

2.1. Consider the matrix

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

(a) How many rows of B contain a pivot?

(b) Is the span of the columns of B all of $\mathbb{R}^4$?

(c) Give one nonzero vector $\boldsymbol{d}$ such that $B\boldsymbol{x} = \boldsymbol{d}$ has a solution.

(d) Give one nonzero vector $\boldsymbol{d}$ such that $B\boldsymbol{x} = \boldsymbol{d}$ does not have a solution.

2.2. Give the solution set in parametric vector form for the following systems:

$$\begin{aligned} a + 2b - c + d &= 2 \\ 2a \qquad - c + 2d &= 8 \\ 4a + 4b + 5c - 4d &= 12 \end{aligned} \qquad\qquad \begin{aligned} x + 2y + 3z + 4t &= 3 \\ 2x + 4y + 7z \qquad &= -5 \\ 3x + 6y + 10z + 4t &= -2 \end{aligned}$$

2.3. For each triple of vectors, find all value(s) h for which the triple of vectors is linearly dependent:

(a) $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, h)$

(b) $(1, -1, 1)$, $(2, 3, 4)$, $(5, h, h)$

(c) $(0, 0, 0)$, $(1, h, 3)$, $(4, -5, -2)$

Solutions to Practice Exercieses

2.1. (a) 3 (2nd, 4th, 5th)

(b) No.

(c) Any vector where last entry is zero.

(d) Any vector where last entry is nonzero.

2.2. The reductions proceed:

$$\begin{bmatrix}1&2&-1&1&2\\2&0&-1&2&8\\4&4&5&-4&12\end{bmatrix}\begin{bmatrix}1&2&-1&1&2\\0&-4&1&0&4\\0&-4&9&-8&4\end{bmatrix}\begin{bmatrix}1&2&-1&1&2\\0&-4&1&0&4\\0&0&8&-8&0\end{bmatrix}$$

$$\rightsquigarrow\begin{bmatrix}1&2&0&0&2\\0&-4&0&1&4\\0&0&1&-1&0\end{bmatrix}\begin{bmatrix}1&0&0&1/2&4\\0&1&0&-1/4&-1\\0&0&1&-1&0\end{bmatrix}$$

$$\begin{bmatrix}1&2&3&4&3\\2&4&7&0&-5\\3&6&10&4&-2\end{bmatrix}\begin{bmatrix}1&2&3&4&3\\0&0&1&-8&-11\\0&0&1&-8&-11\end{bmatrix}\begin{bmatrix}1&2&3&4&3\\0&0&1&-8&-11\\0&0&0&0&0\end{bmatrix}\begin{bmatrix}1&2&0&28&36\\0&0&1&-8&-11\\0&0&0&0&0\end{bmatrix}$$

So the solution sets are:

$$\begin{bmatrix}a\\b\\c\\d\end{bmatrix}=\begin{bmatrix}4\\-1\\0\\0\end{bmatrix}+d\begin{bmatrix}-\frac{1}{2}\\\frac{1}{4}\\1\\1\end{bmatrix}\qquad\begin{bmatrix}x\\y\\z\\t\end{bmatrix}=\begin{bmatrix}36\\0\\-11\\0\end{bmatrix}+y\begin{bmatrix}-2\\1\\0\\0\end{bmatrix}+t\begin{bmatrix}-28\\0\\8\\1\end{bmatrix}$$

2.3. (a) $h=0$

(b) If third vector is a times first plus b times second, need $a+2b=5$ and $-a+3b=a+4b$. Solves to $a=-5/3$ and $b=10/3$, and so $h=35/3$.

(c) All h.

3 Matrix Operations

There are several operations one can apply to a matrix. Addition and scalar multiplication behave as you would expect (just like in vectors), but matrix multiplication and its counterpart matrix inverses are more interesting.

3.1 Basic Matrix Operations

We need some notation. For matrix A, the notation a_{ij} means the entry in row i and column j of A. (Always row index first.)

Matrix ***addition*** *requires that the two matrices have the same dimensions. The sum is defined by adding the corresponding entries. For example,*

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+a_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+a_{22} & a_{23}+b_{23} \end{bmatrix}$$

Similarly, ***scalar multiplication*** *is defined entry-wise. For example,*

$$c \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}$$

Example.

$$3 \begin{bmatrix} 2 & 2 \\ 1 & 4 \end{bmatrix} + 4 \begin{bmatrix} 0 & -1 \\ -5 & 7 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 3 & 12 \end{bmatrix} + \begin{bmatrix} 0 & -4 \\ -20 & 28 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ -17 & 40 \end{bmatrix}$$

Another matrix operation is the transpose:

The ***transpose*** *of a matrix A, denoted A^T, exchanges rows and columns. That is, $(A^T)_{ij} = A_{ji}$.*

Example.

$$\text{The transpose of } \begin{bmatrix} 3 & 4 & 7 \\ -2 & 5 & -3 \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} 3 & -2 \\ 4 & 5 \\ 7 & -3 \end{bmatrix}$$

*A **square** matrix has an equal number of rows and columns. The **diagonal** of a square matrix runs from top-left to bottom-right. A **symmetric matrix** is a square matrix that is symmetric around its diagonal. In other words, $A = A^T$.*

3.2 Matrix Multiplication

An important operation is matrix multiplication. It looks intimidating to start with, but you'll get used to it. Matrix **multiplication** produces a matrix. Only matrices of compatible sizes can be multiplied. There are multiple ways to present the definition. One way to define matrix-matrix multiplication is in terms of matrix-vector multiplication:

If matrix A is $m \times n$ and matrix B is $r \times s$, then for the product AB to be valid it must be that $n = r$. If valid, the product AB has size $m \times s$. The columns of the product are the results of multiplying the first matrix by the columns of the second. That is,

$$AB = \begin{bmatrix} A\boldsymbol{b}_1 & A\boldsymbol{b}_2 & \cdots & A\boldsymbol{b}_s \end{bmatrix}$$

where $\boldsymbol{b}_j$ is the j^{th} column of B.

Example. *Here is the product of a 2×3 and 3×4 matrix:*

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & -2 \end{bmatrix} \begin{bmatrix} 3 & 1 & -1 & 5 \\ -2 & 0 & 3 & -4 \\ 1 & -2 & 2 & -1 \end{bmatrix} = \begin{bmatrix} -2 & 3 & 3 & -2 \\ -8 & 4 & 5 & -10 \end{bmatrix}$$

An example detail: the 3rd column of the result is given by $-\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 2\begin{bmatrix} -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

It is more usual to define matrix multiplication without reference to vectors. There is a formula for each entry in the product; namely the sum

$$(AB)_{ij} = \sum_k a_{ik} b_{kj}$$

That is, to calculate the entry in row i and column j of the product AB, look at row i of the first matrix A and column j of the second matrix B; then multiply the corresponding entries and add. This calculation is illustrated here:

Example. *The entry in the 2nd row 3rd column of the previous example is calculated as* $0 \times (-1) + 3 \times 3 + (-2) \times 2 = 5$.

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

$$(0 \times -1) + (3 \times 3) + (-2 \times 2)$$

3.3 Properties of Matrix Multiplication

It is very important to note two fundamental properties about multiplication:

1. Matrix multiplication is **associative**. That is, brackets don't matter. For example, the two products $(AB)C$ and $A(BC)$ are equal (and the one product is valid whenever the other one is).

2. However, matrix multiplication is not **commutative**. That is, order matters. There is no guarantee that $AB = BA$. Indeed, the one product might be valid when the other one is not. Even if A and B are square matrices of the same size, so that both products are defined and the results have the same size, there is no guarantee (and indeed it is unlikely that) the two products are the same.

We repeat this for emphasis:

Fact 3.1 *Matrix multiplication is associative but not commutative.*

In particular, this means that if we multiply by a matrix, we must specify whether we are multiplying it on the left or the right.

A ***diagonal matrix*** *is a square matrix that has zeros off the diagonal (and might or might not have zeroes on the diagonal). The* ***identity matrix*** I_n *is the* $n \times n$ *diagonal matrix with* 1*'s on the diagonal. (We sometimes write just* I*.) Its columns are the vectors* $\mathbf{e}_i$*: these have* 0*'s in every position except for a* 1 *in the* i^{th} *position.*

Example. *If* A *is a square matrix, then* $IA = AI = A$*, where* I *is the identity matrix of the same size. (This is why* I *is called the identity matrix.)*

Using the above formula for the entries of the product, the following fact about transposes can be shown. Note that the order is swapped!

Fact 3.2 $(AB)^T = B^T A^T$

We conclude this section with the following fact.

Fact 3.3 *Each elementary row operation is equivalent to multiplying on the left by a matrix called an* ***elementary matrix****.*

Instead of providing a proof of this fact, we just give an example of each type.

EXAMPLE. *If A is a 2×2 matrix, then:*
left-multiplication by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *interchanges the first and second row;*
left-multiplication by $\begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}$ *divides the second row by 3; and*
left-multiplication by $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ *subtracts twice the first row from the second.*

3.4 Matrix Transforms

We can also think of the matrix multiplication $A\boldsymbol{x}$ as transforming the vector $\boldsymbol{x}$.

If A is an $m \times n$ matrix, then the matrix transform $\boldsymbol{x} \mapsto A\boldsymbol{x}$ takes a vector in $\mathbb{R}^n$ and produces a vector (called its ***image****) in $\mathbb{R}^m$. That is, it is a function with domain $\mathbb{R}^n$ and range contained in $\mathbb{R}^m$.*

EXAMPLE. *For example, if A is* $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, *then the transform maps* $(5, 3)$ *to* $(3, 5)$.

Some matrix transforms have physical or geometric meaning:

EXAMPLE. *Examples of transforms include:*

- *projections, such as* $P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

- *shears, such as* $S = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$
- *contractions/dilations, such as* $C = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$
- *rotations, such as* $R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

The reasons for the names can be seen by observing the effect of these transforms on sets of points in the plane.

One useful property when applying matrix transforms is that the composition of transforms (meaning applying one transform after another) is equivalent to matrix multiplication. For example, in $\mathbb{R}^2$ to rotate by θ and then contract by a factor of 3, transform by the matrix product CR, where the matrices C and R are as in the above example.

One can also think of applying the same transform multiple times. For matrix A, we use A^p to mean the product of p copies of A. (This needs A to be square.) That is,

$$A^p = \underbrace{A\,A\,A \cdots A}_{p \text{ copies}}$$

3.5 The Inverse of a Matrix

We next define the inverse of a matrix.

The ***inverse*** *of a square matrix A, denoted A^{-1}, is the matrix such that $A\,A^{-1} = A^{-1}\,A = I$. The inverse is not guaranteed to exist. If it exists, then A is said to be* ***invertible****; otherwise A is said to be* ***not invertible*** *or* ***singular****.*

It can be shown that, if the inverse exists, it is unique. If matrix A is invertible, then $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution, namely $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. However, perhaps surprisingly, the inverse is not often calculated, though its existence is crucial.

The inverse of a 2×2 matrix has a formula. Note that the formula also captures when the inverse exists: the matrix is invertible if and only if $ad - bc \neq 0$.

Fact 3.4 *The inverse of a 2×2 matrix is given by:*

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

Example.

$$C = \begin{bmatrix} 2 & -5 \\ -3 & 9 \end{bmatrix} \quad \textit{has inverse} \quad C^{-1} = \frac{1}{2 \times 9 - (-5) \times (-3)} \begin{bmatrix} 9 & 5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 5/3 \\ 1 & 2/3 \end{bmatrix}$$

One general way to find the inverse, albeit slow, is to solve the collection of n vector equations $A\boldsymbol{x} = \boldsymbol{e}_1, \ldots, A\boldsymbol{x} = \boldsymbol{e}_n$ (where the $\boldsymbol{e}_j$ are the columns of I_n as before). These n linear systems can be solved simultaneously by augmenting the matrix with the identity matrix I_n, and bringing the result to reduced row echelon form.

Example.

$$C = \begin{bmatrix} 2 & -5 \\ -3 & 9 \end{bmatrix} \quad \textit{is augmented to} \quad \left[\begin{array}{cc|cc} 2 & -5 & 1 & 0 \\ -3 & 9 & 0 & 1 \end{array}\right]$$

$$\textit{This reduces to} \quad \left[\begin{array}{cc|cc} 1 & 0 & 3 & 5/3 \\ 0 & 1 & 1 & 2/3 \end{array}\right] \quad \textit{so that} \quad C^{-1} = \begin{bmatrix} 3 & 5/3 \\ 1 & 2/3 \end{bmatrix}$$

The above discussion shows that a matrix is invertible if and only if it is row equivalent to the identity. Indeed, it can be shown that the inverse is the product of the elementary matrices that reduce A to the identity.

Here are some useful formulas for inverses:

Fact 3.5 *If A and B are square matrices of the same size:*
(a) $(A^{-1})^{-1} = A$
(b) $(AB)^{-1} = B^{-1}A^{-1}$ *(Note the reversal!)*
(c) $(A^T)^{-1} = (A^{-1})^T$.

Proof. *In each case, to prove that one has the inverse, one can just calculate the product and check that this yields the identity matrix. We do part (b); the other parts are similar. This proof uses the fact that, by the associative law, one can move* *the brackets around. We have*

$$\begin{aligned}
(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} && \text{by associative law} \\
&= AIA^{-1} && \text{since } B \text{ and } B^{-1} \text{ multiply to } I \\
&= AA^{-1} && \\
&= I && \text{since } A \text{ and } A^{-1} \text{ multiply to } I
\end{aligned}$$

Multiplying on the left similarly yields the identity matrix.

3.6 Characterization of Invertible Matrices

Let us collect in one place the big theorem that pulls many things together. (We will define the determinant in the next chapter.)

Fact 3.6 *For an $n \times n$ matrix A, the following conditions are* ***equivalent****:*

- *A is invertible*
- *A has n pivots*
- *A is row equivalent to I_n*
- *$A\boldsymbol{x} = \mathbf{0}$ has a unique solution*
- *the columns of A are linearly independent*
- *the span of the columns of A is all of $\mathbb{R}^n$*
- *the determinant of A is nonzero*

We will use this fact repeatedly.

3.7 Block and Diagonal Matrices

A ***partitioned*** *matrix has the rows and columns partitioned, dividing the matrix up into blocks. A* ***block-diagonal*** *matrix is one where all blocks off the diagonal are zero.*

If they are the correct size, the blocks of a partitioned matrix can be treated as elements for formulas.

Example. *If two $n \times n$ matrices are partitioned into $n/2 \times n/2$ blocks and both* *matrices have a zero block as the bottom left, then*

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \begin{bmatrix} D & E \\ 0 & F \end{bmatrix} = \begin{bmatrix} AD & AE+BF \\ 0 & CF \end{bmatrix}$$

It can be shown that a block-diagonal matrix is invertible if and only if all the diagonal blocks are invertible. Moreover, its inverse is the block-diagonal matrix with the inverses of the diagonal blocks.

Example.

$$\begin{bmatrix} 6 & 0 & 0 \\ 0 & 3 & -5 \\ 0 & -5 & 9 \end{bmatrix}^{-1} = \begin{bmatrix} 1/6 & 0 & 0 \\ 0 & 9/2 & 5/2 \\ 0 & 5/2 & 3/2 \end{bmatrix}$$

A **lower triangular** matrix is one whose entries above the main diagonal are zero. An **upper triangular** matrix is defined similarly. For example, a diagonal matrix is both lower and upper triangular.

Fact 3.7 *A square triangular matrix is invertible if and only if every entry on the diagonal is nonzero.*

This fact follows from the Characterization of Invertible Matrices (Fact 3.6), since there will be a full set of pivots if and only if every entry on the diagonal is nonzero.

Practice

3.1. Compute XY, $X+Y$, YZ, $Y+Z$, and ZX, when they exist, for the following the matrices:

$$X = \begin{bmatrix} -3 & 1 & 4 \\ 1 & 2 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} 5 & -4 \\ 3 & -1 \end{bmatrix} \qquad Z = \begin{bmatrix} -2 & 7 \\ 5 & -1 \end{bmatrix}$$

3.2. Assume A, B, and C are 2×2 matrices.

(a) Show that if $AB = AC$ and A is invertible, then $B = C$.

(b) Give an example such that $AB = AC$ but $B \neq C$.

3.3. Calculate the inverses of the following matrices:

$$E = \begin{bmatrix} 2 & -2 \\ 3 & 7 \end{bmatrix} \qquad F = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad G = \begin{bmatrix} 2 & -2 & 0 & 0 \\ 3 & 7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

Solutions to Practice Exercieses

3.1. XY and $X+Y$ do not exist.

$$YZ = \begin{bmatrix} -30 & 39 \\ -11 & 22 \end{bmatrix} \qquad Y+Z = \begin{bmatrix} 3 & 3 \\ 8 & -2 \end{bmatrix} \qquad ZX = \begin{bmatrix} 13 & 12 & -8 \\ -16 & 3 & 20 \end{bmatrix}$$

3.2. (a) If A^{-1} exists, then we have $A^{-1}(AB) = A^{-1}(AC)$ so that, using the associative law, we have $(A^{-1}A)B = (A^{-1}A)C$ so that $B = C$.

(b) e.g. Take A to be the all-zero matrix.

3.3.

$$E^{-1} = \begin{bmatrix} 7/20 & 1/10 \\ -3/20 & 1/10 \end{bmatrix} \qquad F^{-1} = F \qquad G^{-1} = \begin{bmatrix} E^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{9} \end{bmatrix}$$

4 Determinants

Associated with each square matrix is a number called its determinant. The most important property is that the determinant is zero precisely when the matrix is not invertible. We give below a general definition of a determinant. For calculations, however, the formulas and results in the later sections are more useful. (And indeed, some books use those formulas as the definition.)

4.1 Introduction to Determinants

Determinants are defined for square matrices. The determinant of square matrix A is denoted $\det A$, or indicated by the use of vertical lines replacing the square brackets of the matrix.

Before we give the full definition, let us note that the **determinant** of a 2×2 matrix has a famous formula:

Fact 4.1

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

We saw the expression $ad-bc$ earlier in the formula for the inverse of a 2×2 matrix. Indeed, like in the 2×2 case, we will see that the determinant being nonzero captures when a matrix is invertible.

4.2 A Definition of Determinant

The determinant of a general matrix can be defined in terms of permutation matrices.

> *A* ***permutation matrix*** *is a square matrix that contains only* 0*'s and* 1*'s with exactly one* 1 *in each row and column*

Example. *The identity matrix is a permutation matrix. So too is*

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

There are exactly $n!$ permutation matrices of size $n \times n$.

The ***sign*** *of a permutation matrix is $(-1)^k$, where k is the number of row interchanges needed to change the matrix to be the identity.*

Example. *The identity matrix has sign $+1$. The matrix P of the previous example has sign of $+1$, since it can be made the identity by first interchanging the first and third rows, then interchanging the third and fourth row.*

With this machinery, we are now able to define the determinant.

The ***determinant*** *of an $n \times n$ matrix A is defined by:*
- *consider each $n \times n$ permutation matrix P*
- *for each P, multiply together the corresponding entries of A (to obtain what we call a* ***transversal****)*
- *finally sum each transversal after multiplying it by the sign of its P*

For example, in the 2×2 case, there are two permutation matrices $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then P corresponds to the transversal ad and has positive sign, while Q corresponds to the transversal bc and has negative sign. Thus we get the formula for the determinant of a 2×2 matrix above (Fact 4.1.).

Though we will usually not evaluate determinants using this formula, it does reveal some properties. For example, if every entry in some row is zero, then each of the transversals contains a zero, and so is zero. It follows that:

Fact 4.2 *If matrix A has an all-zero row or column, then* $\det A = 0$.

Here is another useful property:

Fact 4.3 *The determinant of a triangular matrix is the product of the diagonal entries.*

PROOF. *Every transversal except the main diagonal is guaranteed to contain a* 0*-entry and thus be* 0*. The diagonal comes from the identity permutation matrix, which has positive sign (since* $k = 0$*).*

In particular, the determinant of the identity matrix I is $\det I = 1$.

4.3 Recursive Formula: Cofactor Expansion

It can be shown that the definition of determinant implies the following formula.

Fact 4.4 *Assume A is an $n \times n$ matrix. Let A_{ij} denote the matrix formed by removing row i and column j. Then* ***expansion*** *across the first row of A gives the formula*

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

The formula is called **recursive** because it give the value in terms of the value for smaller versions of the same problem. The idea behind the proof of the formula (which we omit) is that if one considers a transversal of A that contains a_{11}, then the rest of it is a transversal of A_{11}; and thus the contribution to the overall determinant involving a_{11} is given by $a_{11} \det A_{11}$.

EXAMPLE. *Let B be the following matrix:*

$$\begin{bmatrix} 3 & 6 & 0 \\ 2 & 7 & -1 \\ 0 & 4 & -8 \end{bmatrix}$$

Then B has determinant

$$3 \begin{vmatrix} 7 & -1 \\ 4 & -8 \end{vmatrix} - 6 \begin{vmatrix} 2 & -1 \\ 0 & -8 \end{vmatrix} + 0 \begin{vmatrix} 2 & 7 \\ 0 & 4 \end{vmatrix} = 3 \times (-52) - 6 \times (-16) + 0 = -60.$$

One can expand across other rows, but note that the sign is always $(-1)^{i+j}$. Each term $C_{ij} = (-1)^{i+j} A_{ij}$ in the expansion is called a **cofactor**.

4.4 Properties of Determinants

As we mentioned at the start, the most important aspect of determinants is the fundamental fact:

Fact 4.5 *A square matrix is invertible if and only if its determinant is not zero.*

This is equivalent to saying that the determinant is zero if and only if the columns (and rows) are linearly dependent.

This result is a consequence of the more general fact that the elementary row operations do not change the determinant much. We omit the proof of the following:

Fact 4.6
- *Adding a multiple of a row to another row does not change* det
- *Interchanging two rows flips the sign of* det
- *Multiplying a row by a scalar does the same to* det

By the result about triangular matrices earlier (Fact 4.3), the above result gives us another method to calculate the determinant:

ALGOR *Assume A is an $n \times n$ matrix. If one obtains pivots in every row/column when reducing A to echelon form, without using an interchange, then the determinant of A is the product of the pivots.*

Example. *Consider the matrix B from earlier.*

$$\begin{bmatrix} 3 & 6 & 0 \\ 2 & 7 & -1 \\ 0 & 4 & -8 \end{bmatrix}$$

This reduces to

$$\begin{bmatrix} 3 & 6 & 0 \\ 0 & 3 & -1 \\ 0 & 0 & -\frac{20}{3} \end{bmatrix}$$

without interchanges. Thus the determinant of B is $3 \times 3 \times (-\frac{20}{3}) = -60$.

Other properties of determinants include the following. We omit the proofs:

Fact 4.7 *(a)* $\det(A^T) = \det A$
(b) $\det(AB) = (\det A)(\det B)$
(c) $\det(A^{-1}) = \frac{1}{\det A}$

4.5 Applications of Determinants

Though we don't give it or use it, a famous idea is called **Cramer's rule**. This says that the solution $\boldsymbol{x}$ to the matrix equation $A\boldsymbol{x} = \boldsymbol{b}$ can be expressed in terms of the determinants of A and matrices built from A. There is similarly a formula for the inverse A^{-1}.

There is also a geometric interpretation of the determinant. The **volume** of a box whose sides are vectors is given by the absolute value of the associated determinant. For example, the area of the parallelogram determined by vectors (x_1, y_1) and (x_2, y_2) is $|x_1y_2 - x_2y_1|$.

Fact 4.8 *In $\mathbb{R}^2$, if one applies a matrix transform M to some shape, then the area of the shape changes by a factor of* $\det M$*.*

A similar result holds for the change in volumes under matrix transforms in $\mathbb{R}^3$.

Practice

4.1. Calculate the determinants of the following matrices using cofactors.

$$C = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 0 & 0 & 5 & 4 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & -1 & 3 \\ 3 & 4 & 1 \\ 0 & 2 & -1 \end{bmatrix} \qquad L = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 2 & 0 \\ 4 & 1 & 5 \end{bmatrix}$$

4.2. Calculate the determinants of the matrices in the previous question using row reduction.

4.3. Suppose A is a 3×3 matrix such that $\det A = 5$. Give the determinant of:

(a) A^T

(b) $(A^2)^{-1}$

(c) The matrix that results if one takes A and **replaces** the 2nd row by the sum of the 1st and 3rd rows.

(d) The matrix that results if one takes A and **increases** the 2nd row by the sum of the 1st and 3rd rows.

(e) $A + A$.

(f) $A + A^T$.

Solutions to Practice Exercieses

4.1. $\det C = 30$
$\det D = 9$
$\det L = -10$

4.2. $\det C = 30$
$\det D = 9$
$\det L = -10$

4.3. (a) 5 (by Fact 4.7)
(b) $\frac{1}{25}$ (by Fact 4.7)
(c) 0 (the rows of resultant matrix are linearly dependent so it is not invertible)
(d) 5 (by Fact 4.6)
(e) 40 (equivalent to multiplying each row by 2)
(f) Could be anything.

5 Subspaces, Bases, and Dimension

In this chapter we introduce a subspace, which is a collection of vectors such that if one adds two vectors or scales a vector, one gets a vector already in the collection. We then consider a basis, with which one can construct every vector in a subspace in a unique way; and the size of the basis is the dimension of the subspace.

5.1 Sets Closed Under Operations

A key mathematical idea that we need is the idea of a set closed under an operation.

> *A set S is* ***closed*** *under some operation if applying that operation to elements of S always produces an element of S.*

Note that the definition says nothing about what happens if you start with two elements outside S; it only says that if you start inside S you cannot escape S using that operation.

EXAMPLE. *The integers are closed under addition: adding two integers always produces an integer. The integers are also closed under subtraction and multiplication: subtracting or multiplying two integers always produces an integer. But the integers are not closed under division.*

EXAMPLE. *The set of positive real numbers is not closed under subtraction: for example, $2 - \pi$ is not positive. This set is closed under addition, multiplication, division, and exponentiation.*

Fact 5.1 *The set of solutions to the homogeneous equation $A\boldsymbol{x} = \mathbf{0}$ is closed under both addition and scalar multiplication.*

In other words, if one takes two solutions and adds them, then the result is again a solution; and if one takes a solution and scales it, then the result is again a solution.

PROOF. *Assume $\boldsymbol{x}$ and $\boldsymbol{x}'$ are solutions to the homogeneous equation. Then $A(\boldsymbol{x} + \boldsymbol{x}') = A\boldsymbol{x} + A\boldsymbol{x}' = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so that $\boldsymbol{x} + \boldsymbol{x}'$ is a solution. Similarly, $A(c\boldsymbol{x}) = c(A\boldsymbol{x}) = c\mathbf{0} = \mathbf{0}$, so that $c\boldsymbol{x}$ is a solution.*

5.2 Subspaces

*A **subspace** of $\mathbb{R}^n$ is a subset S such that*
(0) S contains the zero vector;
(1) S is closed under addition;
(2) S is closed under scalar multiplication.

Consider $\mathbb{R}^2$.

Example. *Let S be the set of points (x, y) in $\mathbb{R}^2$ on the line $y = 3x$. This is a subspace. Check the three conditions. But the set of points on the line $y = 3x + 1$ is not a subspace (as it satisfies none of the three conditions!)*

Example. *Let T be the set of points (x, y) in $\mathbb{R}^2$ such that $|x| = |y|$. This is not a subspace. Though the set T satisfies Conditions (0) and (2) (check!), it is not closed under addition: for example $(2, -2)$ and $(3, 3)$ are in T but their sum $(5, 1)$ is not in T.*

Example. *Let U be the set of points (x, y) in $\mathbb{R}^2$ such that $x, y \geq 0$. This is not a subspace. Though the set U satisfies Conditions (0) and (1) (check!), it is not closed under scalar multiplication: for example $(2, 3)$ is in U, but scaling by -1 produces $(-2, -3)$, which is not in U.*

By the matrix-inversion theorem, we know that in $\mathbb{R}^2$ that if two vectors are linearly independent, then the span of the two vectors in all of $\mathbb{R}^2$. Thus we get

Fact 5.2 *Every subspace of $\mathbb{R}^2$ is either $\{0\}$, a line through the origin, or $\mathbb{R}^2$ itself.*

Example. *Consider $\mathbb{R}^3$ and let S be the set of all vectors whose third coordinate is 0. This is a subspace. Because, the zero vector is in S; adding two vectors in S produces a vector in S; and scaling a vector in S produces a vector in S. On the other hand, let T be the set of all vectors whose third coordinate is 1; this is not a subspace as it fails all three conditions.*

There is a result for $\mathbb{R}^3$ that is similar to the result for $\mathbb{R}^2$:

Fact 5.3 *Every subspace of $\mathbb{R}^3$ is either $\{0\}$, a line through the origin, a plane through the origin, or $\mathbb{R}^3$ itself.*

Here is an important example:

EXAMPLE. *The set containing just the zero-vector is always a subspace.* $\mathbb{R}^n$ *itself is also a subspace.*

One useful fact is the following:

Fact 5.4 *If S is a finite set of vectors in $\mathbb{R}^n$, then Span(S) is a subspace.*

PROOF. *Say $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$. Then $\mathbf{0}$ is the linear combination $0\boldsymbol{v}_1 + \ldots + 0\boldsymbol{v}_k$. So Condition (0) holds. If $\boldsymbol{x} = a_1\boldsymbol{v}_1 + \ldots + a_k\boldsymbol{v}_k$ and $\boldsymbol{y} = b_1\boldsymbol{v}_1 + \ldots + b_k\boldsymbol{v}_k$, then $\boldsymbol{x} + \boldsymbol{y} = (a_1 + b_1)\boldsymbol{v}_1 + \ldots + (a_k + b_k)\boldsymbol{v}_k$. Thus Span$(S)$ is closed under addition. Also, $c\boldsymbol{x} = (ca_1)\boldsymbol{v}_1 + \ldots + (ca_k)\boldsymbol{v}_k$. So Span$(S)$ is also closed under scalar multiplication.*

(The above fact is also true about infinite S; indeed the proof is the same except one has to use different notation.)

5.3 Bases

*A **basis** B of a subspace S is a set such that*

(i) B is linearly independent, and

(ii) the span of B is all of S.

The set $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ forms what is called the **standard basis** of $\mathbb{R}^n$. (Recall that $\boldsymbol{e}_i$ denotes the i^{th} column of the identity matrix.)

EXAMPLE. *Consider $\mathbb{R}^2$. The standard basis is $\{(1,0),(0,1)\}$. But $\mathbb{R}^2$ has many bases. In fact, if one takes any two vectors that are linearly independent, then their span is $\mathbb{R}^2$ and thus the pair of vectors is a basis. On the other hand, any trio of vectors is linearly dependent, while no single vector has a span that is $\mathbb{R}^2$.*

By the Matrix-Inversion Theorem (Fact 3.6) we get the following important fact:

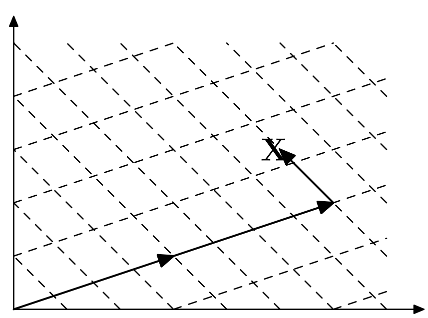
EXAMPLE. *If A is an invertible $n \times n$ matrix, then its columns form a basis of $\mathbb{R}^n$.*

The usefulness of a basis is the following fundamental result:

Fact 5.5 *If B is a finite basis for subspace S, then every element in S is* ***uniquely*** *expressible as a linear combination of vectors of B.*

PROOF. *Say $B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k\}$. Let $\boldsymbol{w}$ be any vector in S. Because the span of B is S, we know that $\boldsymbol{w}$ can be expressed as a linear combination of vectors of B. Consider any two linear combinations that give $\boldsymbol{w}$; say $\boldsymbol{w} = c_1\boldsymbol{b}_1 + \ldots + c_k\boldsymbol{b}_k$ and $\boldsymbol{w} = d_1\boldsymbol{b}_1 + \ldots + d_k\boldsymbol{b}_k$. Then, by subtracting the one equation from the other, we get that $\mathbf{0} = (c_1 - d_1)\boldsymbol{b}_1 + \ldots + (c_k - d_k)\boldsymbol{b}_k$. Because B is linearly independent, it must be that $(c_1 - d_1) = \cdots = (c_k - d_k) = 0$. That is, $c_1 = d_1$, $c_2 = d_2$, etc. It follows that the two linear combinations for $\boldsymbol{w}$ are the same. Or in other words, the linear combination is unique.*

In $\mathbb{R}^2$, one can think of the linear combination as identifying a point in the plane: with the standard basis, $(5, 3)$ means go 5 units in $\boldsymbol{e}_1$ direction and (then) 3 units in $\boldsymbol{e}_2$ direction. But one can also uniquely identify the point given any two linearly independent vectors $\boldsymbol{u}$ and $\boldsymbol{v}$: go so many units in the direction of $\boldsymbol{u}$ and then so many units in the direction of $\boldsymbol{v}$. For example, $(5, 3)$ can also be reached by going 2 units of $(3, 1)$ and -1 units of $(1, -1)$. Instead of a rectangular grid, one can think of grid lines parallel to the vectors:



5.4 Dimension

The big theorem is the following result. We omit the proof.

Fact 5.6 *All bases of a subspace have the same number of elements.*

Thus we can define:

*The **dimension** of a subspace is defined to be the number of elements in a basis.*

EXAMPLE. $\mathbb{R}^d$ *has dimension* d.
A line (through the origin) has dimension 1.
A place (through the origin) has dimension 2.

An important but silly example is:

EXAMPLE. *The subspace* $\{\mathbf{0}\}$ *has dimension* 0 *and an empty basis. (We noted earlier that any set containing the zero-vector is linearly dependent. The claim that* $\mathbf{0}$ *is in the span of the empty set is mathematical prestidigitation: one could take it simply as a definition, but the argument is that if you add up nothing you get the zero-vector. . .)*

It can be shown that:

Fact 5.7 *Assume subspace* S *has dimension* p.
(a) Any linearly independent set of p *vectors forms a basis of* S.
(b) Any set of p *vectors whose span is* S *forms a basis of* S.

In other words, a basis is simultaneously a spanning set that is as small as possible, and a linearly independent set that is as large as possible.

Practice

5.1. Consider the following subsets of $\mathbb{R}^3$. Explain why each is not a subspace.

(a) The points in the xy-plane in the first quadrant.

(b) All integer solutions to the equation $x^2 + y^2 = z^2$.

(c) All points on the line $x + z = 5$.

(d) All vectors where the three coordinates are the same in absolute value.

5.2. What is the dimension of:

(a) The set of all vectors in $\mathbb{R}^{26}$ whose last coordinate is 0.

(b) The set of all vectors in $\mathbb{R}^{26}$ that are a multiple of $(1, 1, \ldots, 1)$.

(c) The set of all solutions to the equation $a + b + c + \ldots + z = 0$.

5.3. Show that the intersection of two subspaces is always a subspace.

Solutions to Practice Exercieses

5.1. (a) Not closed under scalar multiplication: $(1,1,0)$ in set but not $-1(1,1,0)$

(b) Not closed under scalar multiplication: $(3,4,5)$ in set but not $\frac{1}{2}(3,4,5)$

(c) Does not contain zero.

(d) Not closed under addition: $(1,1,1)$ and $(1,1,-1)$ in set but not their sum $(2,2,0)$.

5.2. (a) 25

(b) 1

(c) 25

5.3. Consider subspaces S and T. Since the zero vector is in both, it's in their intersection. Consider vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in the intersection. Since both S and T are closed under addition, the sum $\boldsymbol{u}+\boldsymbol{v}$ is in both and therefore in their intersection. Similarly one gets closure under scalar multiplication.

6 The Three Matrix Spaces and Coordinate Systems

In this chapter we consider three standard subspaces associated with a matrix. We also examine how a basis for a subspace can be used to provide a coordinate system.

6.1 The Three Matrix Spaces

For a matrix A, we define three fundamental sets as follows.

≻ *The* ***null space*** *of matrix A, denoted Nul A, is the set of all solutions to the homogeneous system $A\boldsymbol{x} = \mathbf{0}$. That is, all vectors mapped to $\mathbf{0}$ by the matrix transform $x \mapsto Ax$.*

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

≻ *The* ***row space*** *of matrix A, denoted Row A, is the set of linear combinations of rows of A.*

EXAMPLE. *Consider the matrix* $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. *The column space consists of* $(1,2)$, $(3,4)$, *and* $(5,6)$ *and every linear combination thereof (which turns out to be all of* $\mathbb{R}^2$*). The row space consists of* $(1,2,3)$ *and* $(4,5,6)$ *and every linear combination thereof. The null space contains vectors such as* $(1,-2,1)$.

All three of these sets are subspaces:

Fact 6.1 *If A is an $m \times n$ matrix, then*
(a) Nul A is a subspace of $\mathbb{R}^n$.
(b) Col A is a subspace of $\mathbb{R}^m$.
(c) Row A is a subspace of $\mathbb{R}^n$.

PROOF. *Parts (b) and (c) follow from the earlier observation that any span is a subspace (Fact 5.4). To prove (a), by the definition we need to check the three conditions. The two closure conditions were noted in Fact 5.1. Further, the zero vector* $\mathbf{0}$ *is in the null space, since* $A\mathbf{0} = \mathbf{0}$.

Our earlier discussion about row operations showed that:

Fact 6.2 *If two matrices are row equivalent, then they have the same row space.*

6.2 Bases for the Three Matrix Spaces

We have actually already seen how to construct a basis for the null space. We did that when finding the general solution to the homogeneous system $A\boldsymbol{x} = \mathbf{0}$ in parametric vector form. That is,

ALGOR BASIS FOR NULL SPACE.
A basis for Nul A is obtained by creating one vector for each free variable after row reduction. The dimension of the null space is the number of free variables.

The situation for the column space is the following. We need a set of linearly independent columns that cannot be increased. This set can be determined by row reduction, even though row reduction changes the column space:

ALGOR BASIS FOR COLUMN SPACE.
*A basis for Col A is obtained by taking each **original** column vector that corresponds to a pivot column. The dimension of the column space is the number of basic variables.*

Since row operations do not change the row space, we get that

ALGOR BASIS FOR ROW SPACE.
A basis for the row space is obtained by taking the nonzero rows of echelon form. The dimension of the row space is the number of pivots.

EXAMPLE.

$$\textit{Matrx} \quad \begin{bmatrix} 1 & -1 & 4 & 7 & -1 \\ 5 & -5 & 0 & 15 & 11 \\ 2 & -2 & -2 & 4 & 6 \end{bmatrix} \quad \textit{reduces to} \quad \begin{bmatrix} 1 & -1 & 0 & 3 & 11/5 \\ 0 & 0 & 1 & 1 & -4/5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So the null space has dimension 3*. Further, by our discussion of parametric vector description, the null space has a basis of* $(1, 1, 0, 0, 0)$, $(-3, 0, -1, 1, 0)$, *and* $(-11/5, 0, 4/5, 0, 1)$.

The column space has dimension 2 *with a basis of* $(1, 5, 2)$ *and* $(4, 0, -2)$.

The row space has dimension 2 with a basis of $(1, -1, 0, 3, 11/5)$ *and* $(0, 0, 1, 1, -4/5)$.

EXAMPLE. *For matrix* $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, *the null space is* $\{\mathbf{0}\}$, *the column space is all of* $\mathbb{R}^2$, *as is the row space.*

In particular, the above discussion shows:

Fact 6.3 *For any matrix, its row and column spaces have the same dimension.*

The ***rank*** *of a matrix* A *is the dimension of Col* A *or equivalently the dimension of Row* A. *A square matrix has* ***full rank*** *if its rank is equal to the number of rows/columns.*

We can add another characterization of invertible matrices.

Fact 6.4 *Square matrix* A *is invertible* $\iff$ *Nul* $A = \{0\}$ $\iff$ A *has full rank.*

6.3 Coordinate Systems

In Fact 5.5 we noted that if B is a basis for a subspace, then every vertex in the subspace is uniquely expressed as a linear combination of B. These coefficients provide a way to specify the vector:

If B *is a basis of* $\mathbb{R}^n$, *the notation* $[\boldsymbol{x}]_B$ *is defined to be the coefficients used when expressing* $\boldsymbol{x}$ *as a linear combination of vectors in* B. *This vector is called the* ***coordinates*** *of* $\boldsymbol{x}$ *relative to* B.

EXAMPLE. *Consider* $\mathbb{R}^2$ *and basis* B *is* $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\}$. *If* $\boldsymbol{x} = (5, -6)$, *then* $[\boldsymbol{x}]_B = (7, -2)$. *This can be checked by calculating* $7 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ *and seeing that this equals* $\begin{bmatrix} 5 \\ -6 \end{bmatrix}$.

6.4 Changing the Basis

One common task is to convert between coordinate systems.

*In $\mathbb{R}^n$, the **change-of-coordinates matrix** P_B has B as its columns.*

Fact 6.5 *If B is a basis of $\mathbb{R}^n$ then*

$$\boldsymbol{x} = P_B\,[\boldsymbol{x}]_B$$

That is, one can convert from coordinates relative to B to standard coordinates by multiplying by the matrix P_B. On the other hand, to convert from standard coordinates to those relative to B, one multiplies by P_B^{-1}.

More generally, if we have bases B and C, then there is a matrix that allows one to convert between the two coordinate systems:

*If B and C are bases of $\mathbb{R}^n$, then the **change-of-coordinates** matrix, denoted $\underset{C \leftarrow B}{P}$, allows one to convert between the two coordinate systems. That is*

$$[\boldsymbol{x}]_C = \underset{C \leftarrow B}{P}\,[\boldsymbol{x}]_B$$

The columns of $\underset{C \leftarrow B}{P}$ express each vector of B in terms of C.

A cute fact is that $\underset{C \leftarrow B}{P} = P_C^{-1} P_B$. This is not surprising: it says that to convert from B to C, one can convert from B to standard coordinates, and then from there to C.

EXAMPLE. *If $B = \{(1,0),(1,3)\}$ and $C = \{(1,-2),(3,-5)\}$, then* $\underset{C \leftarrow B}{P} = \begin{bmatrix} -5 & -14 \\ 2 & 5 \end{bmatrix}$

Practice

6.1. Consider the matrix

$$F = \begin{bmatrix} 2 & -1 & 0 & -3 \\ 12 & -6 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) Give the dimension and a basis of the column space of F.

(b) Give the dimension and a basis of the null space of F.

(c) Give the dimension and a basis of the row space of F.

6.2. Assume A and B are 2×2 matrices. For each of the following, either prove or disprove. (That is, either explain why we know it's always true, or give an example that shows it is not always true.)

(a) If both A and B have rank 2, then so does the product AB.

(b) If both A and B have rank 1, then so does the product AB.

(c) If both A and B have rank 0, then so does the product AB.

6.3. Consider basis $B = \{(1,3), (-4,1)\}$ of $\mathbb{R}^2$.

(a) If $[\boldsymbol{x}]_B = (7,-2)$, what is $\boldsymbol{x}$?

(b) If $\boldsymbol{y} = (-5,5)$, what is $[\boldsymbol{y}]_B$?

(c) If $\boldsymbol{z} = [\boldsymbol{z}]_B$, what is $\boldsymbol{z}$?

Solutions to Practice Exercieses

6.1. (a) Dimension 2. Example basis: the first and third columns of F.

(b) Dimension 2. Example basis: $\begin{bmatrix} 1/2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3/2 \\ 0 \\ -25 \\ 1 \end{bmatrix}$

(c) Dimension 2. Example basis: the first two rows of F.

6.2. (a) True. For 2×2 matrices, having rank 2 is equivalent to being invertible, and we know from Chapter 3 that the product of invertible matrices is invertible.

(b) Not necessarily true. Consider $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. They have rank 1 but their product has rank 0.

(c) True. The only matrix that has rank 0 is the all-zero matrix.

6.3. (a) $(15, 19)$

(b) $(15/13, 20/13)$

(c) $(0, 0)$

7 Vector Spaces and Linear Transforms

In this chapter we show that the idea of a vector and of a matrix transform can be generalized.

7.1 Vector Spaces

A ***vector space*** *is a collection of vector-like-objects with operations addition and scalar multiplication defined that obey the "usual" vector laws.*

The full list of laws, also known as the **axioms** of a vector space, is given on the next page. In short, they say that addition and scalar multiplication behave like they do for ordinary vectors.

More specifically, the axioms start with the requirement that the vector space is closed under the two operations: that is, if you take two objects in the space and add them, the sum is in the space; if you take an object in the space and scale it by some real number, the scaled object is in the space. Other **axioms** include that:

- addition is **commutative**, **associative**, and has **negation**;
- the **0** vector and 1 scalar behave as **identities**; and
- addition and scalar multiplication **distribute**.

EXAMPLE. $\mathbb{R}^n$, *with addition and scalar multiplication as we've been doing, is a vector space.*

To provide some examples, let us define:

- $\mathbb{P}_n$ *is the set of all polynomials in* t *of degree at most* n
- $\mathbb{P}$ *is the set of all polynomials in* t
- $C[t]$ *is the set of all continuous functions in variable* t *with domain all of* $\mathbb{R}$
- M_n *is the set of all* $n \times n$ *matrices*

Example. *To see that $\mathbb{P}_n$ is a vector space, note, for example, that if one adds two polynomials the result is a polynomial, and its degree cannot be larger than both summands, and thus $\mathbb{P}_n$ is closed under addition. The zero of the vector space is the* 0 *polynomial. Similarly $\mathbb{P}$, the set of all polynomials, is a vector space.*
From calculus we know that if you add continuous functions then the result is a continuous function. So $C[t]$ is a vector space. (The zero constant function plays the role of the zero vector.) And we saw earlier that one can add and scale matrices; thus M_n is a vector space.

Here are the promised axioms of a vector space V. As is standard, we will just talk about "vectors" rather than "vector-like objects". Hopefully this is not too confusing.

AXIOMS *For all vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in V and all (real) scalars c and d:*

1) The sum $\boldsymbol{u}+\boldsymbol{v}$ is in V
2) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$
3) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$
4) There is a vector $\mathbf{0}$ such that $\boldsymbol{u}+\mathbf{0}=\boldsymbol{u}$
5) There is a vector $-\boldsymbol{u}$ such that $\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0}$
6) The scalar multiple $c\boldsymbol{u}$ is in V
7) $c(\boldsymbol{u}+\boldsymbol{v})=c\boldsymbol{u}+c\boldsymbol{v}$
8) $(c+d)\boldsymbol{u}=c\boldsymbol{u}+d\boldsymbol{u}$
9) $c(d\boldsymbol{u})=(cd)\boldsymbol{u}$
10) $1\boldsymbol{u}=\boldsymbol{u}$

It should be noted that what we have defined as a vector space is sometimes called a **real vector space**, because the scalars are restricted to being real numbers. A **complex vector space** would be one where the scalars are allowed to be complex numbers.

7.2 Spans, Linear Independence, and Subspaces

Most of the concepts that we introduced for $\mathbb{R}^n$ carry over to an abstract vector space. In particular, one can talk about the span of a set of vector-like-objects as before and a subspace. A consequence of the definition of a subspace is that a subspace is a vector space in its own right, using the same operations.

EXAMPLE. *Recall that* $\mathbb{P}_n$ *is the set of all polynomials of degree at most* n. *This is a subspace of the space* $\mathbb{P}$ *of all polynomials. And* $\mathbb{P}$ *is a subspace of the space* $C[t]$ *of continuous functions. Another subspace of subspace of* $C[t]$ *is the set of continuous functions such that* $\int_{-\infty}^{\infty} f(t)\,dt = 0$.

We can also define linearly independent, basis, and dimension as before. So for example, a **basis** B of a vector space is a set such that (i) B is linearly independent, and (ii) the span of B is the whole vector space. The dimension of the vector space is the size of a basis.

EXAMPLE. $\mathbb{P}_n$ *has dimension* $n+1$. *A basis is the set* $\{1, t, t^2, \ldots, t^n\}$.

EXAMPLE. *The set* $C[t]$ *of continuous functions has infinite dimension.*

EXAMPLE. *The set* M_2 *of* 2×2 *matrices has dimension* 4. *A basis is, for example, the quartet of* 2×2 *matrices that have three* 0*'s and one* 1. *(Equivalently, to specify a* 2×2 *matrix one has to give* 4 *numbers.)*

EXAMPLE. *In calculus one might consider the equation* $f'' = -f$; *that is, all functions whose second derivative is the negative of the function. One solution to this equation is* $\sin t$; *another is* $\cos t$. *Indeed, any linear combination of* $\sin t$ *and* $\cos t$ *is a solution. It can be shown that this gives all the solutions. That is, the set of all functions* $f(t)$ *that obey the differential equation has basis* $\{\sin t, \cos t\}$.

7.3 Linear Transforms

We saw matrix transforms earlier. If A is an $m \times n$ matrix, then the matrix transform $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. This can be generalized:

A ***linear transform*** T *is a function from one vector space to another vector space. It is required to obey two rules. For all vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in the domain vector space and all reals* c:

(1) $T(\boldsymbol{u}+\boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$, *and*

(2) $T(c\boldsymbol{u}) = cT(\boldsymbol{u})$.

That is, a linear transform has the property: if you add first and then transform, you get the same result as if you transform first and then add; if you scale first and then transform, you get the same result as if you transform first and then scale. One can easily check that matrix multiplication obeys these two rules; that is:

Fact 7.1 *Every matrix transform is a linear transform.*

The null space of a linear transform is the set of all vectors that are mapped to $\mathbf{0}$*; it is often called the* ***kernel*** *of the transform.*

EXAMPLE. *For example, differentiation is a linear transform from the polynomial space* $\mathbb{P}_n$ *to the polynomial space* $\mathbb{P}_{n-1}$*. Its kernel is the set of all constants.*

It can be shown that:

Fact 7.2 *For any linear transform from vector space* V *to subspace* W*:*
1) The kernel is a subspace of V*.*
2) The range is a subspace of W*.*
3) The dimension of the kernel and the dimension of the range sum to the dimension of V*.*

Practice

7.1. In each of the following, state whether it is a vector space. Justify your answer. For the cases where the set is a space, give the dimension and a basis.

(a) the set of all polynomials with degree exactly 1

(b) the set of all 2×2 matrices with determinant 2

(c) the set of all diagonal 3×3 matrices

(d) the set of all vectors in $\mathbb{R}^4$ whose entries sum to 0

(e) the set of all antiderivatives of $f(x) = x^5$

7.2. Show that the axioms of a vector space imply that: for any vector $\boldsymbol{u}$ its additive inverse is unique.

7.3. Show that for any linear transform T it holds that $T\mathbf{0} = \mathbf{0}$.

Solutions to Practice Exercieses

7.1. (a) No; does not include zero

(b) No; does not include zero

(c) Vector space. Dimension 3. The trio of matrices with 1 in one diagonal position and 0's elsewhere.

(d) Vector space. Dimension 3. For example, $(1, -1, 0, 0)$, $(1, 0, -1, 0)$, $(1, 0, 0, -1)$.

(e) No; does not include zero

7.2. The standard way to show that something is unique is to suppose there are two different possibilities and prove them equal. So, suppose that $\boldsymbol{u}$ has additive inverses $\boldsymbol{a}$ and $\boldsymbol{b}$. This means that $\boldsymbol{u} + \boldsymbol{a} = \mathbf{0}$ and $\boldsymbol{u} + \boldsymbol{b} = \mathbf{0}$. Consider $X = (\boldsymbol{a} + \boldsymbol{u}) + \boldsymbol{b}$. We know $X = (\boldsymbol{u} + \boldsymbol{a}) + \boldsymbol{b}$ by Axiom 2, and so $X = \mathbf{0} + \boldsymbol{b} = \boldsymbol{b}$ by Axiom 4. On the other hand, $X = \boldsymbol{a} + (\boldsymbol{u} + \boldsymbol{b})$ by Axiom 3. So $X = \boldsymbol{a} + \mathbf{0} = \boldsymbol{a}$. Thus we have shown that $\boldsymbol{a} = \boldsymbol{b}$. That is, the inverse is unique.

7.3. One way to see this is that $T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0})$ by simplification, but $T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$ by the rules of a linear transform. So $T(\mathbf{0}) + T(\mathbf{0}) = T(\mathbf{0})$, which means that $T(\mathbf{0}) = \mathbf{0}$.

8 Eigenvalues

In this chapter we consider the eigenvalues and eigenvectors of a matrix.

8.1 Eigenvalues and Eigenvectors

*An **eigenvector** of a square matrix A is a nonzero vector $\boldsymbol{x}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some scalar λ called an **eigenvalue**.*

We will soon discuss how to find them and indeed whether they always exist. But first an example:

EXAMPLE. *Consider matrix* $A = \begin{bmatrix} 13 & 15 \\ -10 & -12 \end{bmatrix}$. *Then* $\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *is an eigenvector for* $\lambda_1 = -2$, *since* $A\boldsymbol{x}_1 = (2, -2)$; *and* $\boldsymbol{x}_2 = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$ *is an eigenvector for* $\lambda_2 = 3$, *since* $A\boldsymbol{x}_2 = (-9, 6)$.

Now, in general the equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$ is equivalent to $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$. That is, the eigenvector $\boldsymbol{x}$ is in the null space of $A - \lambda I$. So it matters when the null space of $A - \lambda I$ is nontrivial, or equivalently:

Fact 8.1 *The value λ is an eigenvalue if and only if the matrix $A - \lambda I$ is not invertible.*

In particular, it follows that 0 is an eigenvalue of A if and only if A is not invertible.

*The **characteristic polynomial** of a matrix A is defined as the determinant of the matrix $A - \lambda I$.*

It can be shown that if A is $n \times n$ then the characteristic polynomial is a polynomial of degree n in variable λ. Further, the above discussion shows that:

Fact 8.2 *The eigenvalues are the roots of the characteristic polynomial.*

Example. *Consider again the matrix* $A = \begin{bmatrix} 13 & 15 \\ -10 & -12 \end{bmatrix}$. *This matrix has characteristic polynomial* $(13-\lambda)(-12-\lambda)-(15)(-10) = \lambda^2 - \lambda - 6$. *This polynomial has roots* $\lambda = 3$ *and* $\lambda = -2$.

The characteristic polynomial approach gives us a method to find the eigenvalues, though for large matrices potentially the method is inefficient and/or requires numerical methods to find the roots.

The ***eigenspace*** *of* λ *is defined as all eigenvectors corresponding to* λ *along with the zero vector.*

Equivalently, the eigenspace of eigenvalue λ is the null space of $A - \lambda I$.

Example. *Consider again* $A = \begin{bmatrix} 13 & 15 \\ -10 & -12 \end{bmatrix}$. *The eigenvalues are* $\lambda_1 = 3$ *and* $\lambda_2 = -2$. *For* λ_1, *we need a vector in the null space of* $\begin{bmatrix} 15 & 15 \\ -10 & -10 \end{bmatrix}$; *for example* $(-1, 1)$. *For* λ_2, *we need a vector in the null space of* $\begin{bmatrix} 10 & 15 \\ -10 & -15 \end{bmatrix}$; *for example* $(-3, 2)$.

Fact 8.3 *The eigenvalues of a diagonal matrix are its diagonal entries. The vectors* $\mathbf{e}_i$ *are its eigenvectors.*

We used ad hoc techniques in the above examples, but in general the problem then is to find the null space of $A - \lambda I$. And we know from previous chapters how to find null spaces!

Example. *Consider the matrix*

$$\begin{bmatrix} 5 & -4 & -4 \\ -8 & 9 & 8 \\ 10 & -10 & -9 \end{bmatrix}$$

It can be calculated that this matrix has characteristic polynomial $-\lambda^3 + 5\lambda^2 - 7\lambda + 3$. *The roots of the characteristic polynomial are (according to software)* $3, 1, 1$. *Finally, it can be calculated that the eigenspaces have bases as follows:*

$$\lambda = 1\text{:} \quad \begin{bmatrix} 4 & -4 & -4 \\ -8 & 8 & 8 \\ 10 & -10 & -10 \end{bmatrix} \quad \text{has null space with basis} \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda = 3: \quad \begin{bmatrix} 2 & -4 & -4 \\ -8 & 6 & 8 \\ 10 & -10 & -12 \end{bmatrix} \textit{ has null space with basis } \quad \begin{bmatrix} 2/5 \\ -4/5 \\ 1 \end{bmatrix}$$

8.2 Properties of Eigenvalues and Eigenvectors

One useful property of eigenvectors is the following observation. If one has a vector $\boldsymbol{v}$ expressed as a linear combination of eigenvectors of matrix A, then multiplication by A is quick. For, say vector $\boldsymbol{v} = \sum_i a_i \boldsymbol{x}_i$, where $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$; then

$$A\boldsymbol{v} = \sum_i a_i \lambda_i \boldsymbol{x}_i$$

This property is especially useful if there is a basis for the space consisting of eigenvectors.

An $n \times n$ matrix is defined to have a ***full set of eigenvectors*** *if it has n linearly independent eigenvectors.*

It turns out that if all n eigenvalues are different, then such a full set is guaranteed. This follows from the more general fact that we do not prove:

Fact 8.4 *Eigenvectors for distinct eigenvalues are linearly independent.*

Applying a matrix transform repeatedly is equivalent to transforming with the power of the matrix. There is a simple relationship between the eigenvalues of matrix A and matrix A^k:

Fact 8.5 *If matrix A has eigenvalues λ_i, then the power A^k has eigenvalues λ_i^k. Moreover, the eigenvectors are the same.*

PROOF. *Assume $A\boldsymbol{x} = \lambda_i \boldsymbol{x}$. Then $A^k\boldsymbol{x} = A^{k-1}\lambda_i\boldsymbol{x} = \cdots = \lambda_i^k \boldsymbol{x}$. So $\boldsymbol{x}$ is also an eigenvector of A^k, but with eigenvalue λ_i^k.*

We finish this section with some (remarkable!) facts that can be deduced from properties of the sum or product of the roots of a polynomial.

The ***trace*** *of a matrix is defined as the sum of the diagonal entries.*

Fact 8.6 *For any matrix A,*
(a) the determinant of A equals the product of its eigenvalues.
(b) the trace of A equals the sum of its eigenvalues.

8.3 Similarity and Diagonalization

Matrices A and B are defined to be ***similar*** *if there is an invertible matrix P such that multiplying A on the one side by P and on the other side by P^{-1} gives B. Note that the definition is symmetric: if $B = P^{-1}AP$ then $A = PBP^{-1}$.*

Using the fact that the product of determinants equals the determinant of the product, it can be shown that:

Fact 8.7 *If matrices A and B are similar, then they have the same characteristic polynomial.*

However, while similar matrices have the same eigenvalues, they need not (and usually do not) have the same eigenvectors.

A matrix A is defined to be ***diagonalizable*** *if it is similar to a diagonal matrix.*

By the above fact, if A is similar to diagonal matrix D, then the diagonal entries of D must be the eigenvalues of A. The following result shows how to diagonalize a matrix if it has a full set of eigenvectors:

Fact 8.8 *If matrix A has a full set of eigenvectors, then $A = PDP^{-1}$, where matrix P has the eigenvectors of A as its columns, and diagonal matrix D has the eigenvalues of A on its diagonal (in the same order).*

Proof. *Say $n \times n$ matrix A has eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct) with eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ that are linearly independent. Then, because each column of P is an eigenvector of A, we have that the columns of the product AP can be written as:*

$$AP = [\lambda_1 \boldsymbol{v}_1 \ \lambda_2 \boldsymbol{v}_2 \ \cdots \ \lambda_k \boldsymbol{v}_k]$$

If one multiplies any matrix on the right by a diagonal matrix, it multiplies the columns *by those diagonal entries. That is,*

$$PD = [\lambda_1 \boldsymbol{v}_1 \ \lambda_2 \boldsymbol{v}_2 \ \cdots \ \lambda_k \boldsymbol{v}_k]$$

That is, we have shown that

$$AP = PD.$$

Since we were told the eigenvectors are linearly independent, the matrix P is invertible. Thus $(AP)P^{-1} = (PD)P^{-1}$, and so $A = PDP^{-1}$, as required.

EXAMPLE. *Consider $A = \begin{bmatrix} 13 & 15 \\ -10 & -12 \end{bmatrix}$. Then $P = \begin{bmatrix} -1 & -3 \\ 1 & 2 \end{bmatrix}$ and $D = \begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix}$. The reader should check the product PDP^{-1}.*

Diagonalization is very useful in applications. One little application is

Fact 8.9 *If $A = PDP^{-1}$ then $A^k = PD^kP^{-1}$.*

The proof is left as an exercise.

8.4 Complex Eigenvalues

Recall that i denotes the square-root of -1. Complex numbers have the form $a + bi$, where a and b are real numbers. If $\lambda = a+bi$, then its **(complex) conjugate** is defined to be $\bar{\lambda} = a - bi$.

We do not prove the following result; but, for example, in a 2×2 matrix, its truth follows from the quadratic formula.

Fact 8.10 *If λ is a complex eigenvalue of A, then so is its conjugate $\bar{\lambda}$.*

EXAMPLE. *Consider the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. The characteristic polynomial is given by $(a-\lambda)^2 + b^2$; thus the matrix has eigenvalues $\lambda = a \pm bi$.*
As a matrix transform, the above matrix represents scaling by $|\lambda| = \sqrt{a^2+b^2}$ and rotation through the angle $\arctan b/a$.

An important fact is the following. We omit the proof:

Fact 8.11 *A real symmetric matrix has only real eigenvalues.*

Practice

8.1. For the following matrices, find all eigenvalues and a basis for each eigenspace.

$$J = \begin{bmatrix} 1 & 3 \\ -1 & 5 \end{bmatrix} \qquad K = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad L = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 5 & 1 \\ -1 & 3 & 3 \end{bmatrix}$$

8.2. Prove that: if matrix A is similar to matrix B, and matrix B is similar to matrix C, then A is similar to C.

8.3. Diagonalize the matrix J from the first question by computing P, D, and P^{-1}.

Solutions to Practice Exercieses

8.1. J has eigenvalue 2 with eigenvector $(3, 1)$ and eigenvalue 4 with eigenvector $(1, 1)$.
K has eigenvalue 3 with eigenvector $(1, 1, 1)$ and eigenvalue 0 with basis $(1, -1, 0)$, $(1, 0, -1)$.
L has eigenvalue 2 with eigenvector $(0, -\frac{1}{3}, 1)$ and eigenvalue 6 with basis $(0, 1, 1)$.

8.2. Assume $A = P^{-1}BP$ and $B = Q^{-1}CQ$. Thus $A = P^{-1}Q^{-1}CQP = (QP)^{-1}C(QP)$.

8.3. One solution is given by

$$P = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \qquad P^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 3/2 \end{bmatrix}$$

Another solution is obtained by interchanging the columns of P, interchanging the entries of D, and interchanging the rows of P^{-1}.

9 Orthogonality and Projections

In this section we discuss how to test if two vectors are orthogonal and how to construct vectors that are orthogonal.

9.1 Dot Products and Orthogonality

*The **dot product** (or **inner product**) of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is denoted by $\boldsymbol{u} \cdot \boldsymbol{v}$ and defined as the sum of the product of corresponding entries: that is*

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum_i u_i v_i.$$

Example. *For example,*

$$\begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 4 \\ 7 \end{bmatrix} = 12 - 4 - 14 = -6.$$

If we view the two vectors as matrices, then the dot product $\boldsymbol{u} \cdot \boldsymbol{v}$ is the entry in the 1×1 matrix given by $\boldsymbol{u}^T \boldsymbol{v}$.

There are some immediate properties of the dot product. These include the following:

Fact 9.1 *(Commutative law)* $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
(Distributive law) $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$

*Two vectors are **orthogonal** if their dot product is zero. Orthogonal vectors are sometimes called **perpendicular** vectors.*

Example. *Find vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ that are orthogonal to each other and to $\boldsymbol{w} = (0, 1, 1)$. There are systematic ways to do this. But one way to proceed is to note that the vectors orthogonal to $(0, 1, 1)$ form the null space of the matrix $\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$. Using that, or just observations, we can see that one vector orthogonal to $\boldsymbol{w}$ is $(1, 0, 0)$. Another vector is to take $(0, 1, -1)$.*

*The **length** (or **norm**) of vector $\boldsymbol{v}$ is defined as $||\boldsymbol{v}|| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$. A **unit vector** has length 1.*

ALGOR *To obtain a unit vector in the same direction, divide by the length (called **normalization**).*

EXAMPLE. *For example, the vector $(3, -1, 2)$ has norm $\sqrt{14}$; a unit vector in the same direction as it is $\frac{1}{\sqrt{14}}(3, -1, 2)$.*

We next state Pythagoras' theorem as it appears in vectors. Note the converse.

Fact 9.2 PYTHAGORAS' THEOREM. *Vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if and only if $||\boldsymbol{u}+\boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2$.*

PROOF. *Consider the following computation:*

$$\begin{aligned}
||\boldsymbol{u}+\boldsymbol{v}||^2 &= (\boldsymbol{u}+\boldsymbol{v}) \cdot (\boldsymbol{u}+\boldsymbol{v}) \qquad \text{(by defn. of norm)} \\
&= \boldsymbol{u} \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{v} \qquad \text{(by distr. law)} \\
&= \boldsymbol{u} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{v} + 2\boldsymbol{u} \cdot \boldsymbol{v} \qquad \text{(by comm. law)} \\
&= ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 + 2\boldsymbol{u} \cdot \boldsymbol{v}.
\end{aligned}$$

This means that $||\boldsymbol{u}+\boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2$ if and only if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

9.2 Orthogonal Complements

*For a set W, the **orthogonal complement** is denoted by $W^\perp$ and is defined as the set of all vectors that are orthogonal to **all** of W.*

Fact 9.3 *For any subset W, the orthogonal complement $W^\perp$ is a subspace.*

This fact can be shown using the standard recipe for a subspace. For example, suppose both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal to every $\boldsymbol{w}$ in W. Then so is the sum of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, since $(\boldsymbol{x}_1 + \boldsymbol{x}_2) \cdot \boldsymbol{w} = \boldsymbol{x}_1 \cdot \boldsymbol{w} + \boldsymbol{x}_1 \cdot \boldsymbol{w} = 0 + 0 = 0$ for each $\boldsymbol{w}$. Thus $W^\perp$ is closed under addition.

EXAMPLE. *Consider in $\mathbb{R}^3$ the plane P given by $3x + 4y - z = 0$. Then if we take any vector (x, y, z) in the plane P, it is orthogonal to the vector $(3, 4, -1)$: just computes their dot product and note that it is zero. This means that the orthogonal complement of the plane P contains all multiples of the vector $(3, 4, -1)$. Indeed, these are the only elements: the orthogonal complement to plane P in $\mathbb{R}^3$ is a line.*

The concept of orthogonal complement connects our three spaces of a matrix:

Fact 9.4 *For any matrix A:*

$$(\textit{Row } A)^\perp = \textit{Nul } A \qquad \textit{and} \qquad (\textit{Col } A)^\perp = \textit{Nul } A^T$$

9.3 Orthogonal and Orthonormal Sets

*An **orthogonal set** is a collection of vectors that are pairwise orthogonal. An **orthonormal set** is an orthogonal set of unit vectors.*

(The "pairwise" comment means that for every pair of distinct vectors, the two vectors are orthogonal to each other.)

A key result is that orthogonality implies independence. We omit the proof:

Fact 9.5 *If S is an orthogonal set of nonzero vectors, then S is linearly independent.*

If matrix U has orthonormal columns, then $U^T U = I$. As a matrix transform, such a matrix U preserves lengths and orthogonality. It can be shown that such a matrix must also have orthonormal rows. Thus we can speak of an **orthonormal matrix**.

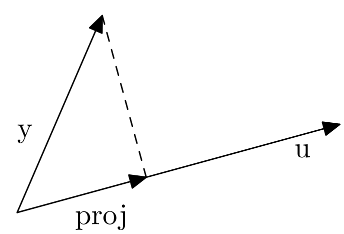
Fact 9.6 *If $B = \{\mathbf{w}_i\}$ is an orthonormal basis, then the coordinates of vector $\mathbf{v}$ relative to B are the dot-products of $\mathbf{v}$ with each $\mathbf{w}_i$.*

We will eventually show that:

Fact 9.7 *Every vector space has an orthonormal basis.*

9.4 Projections

The (orthogonal) ***projection*** *of vector $\boldsymbol{y}$ onto vector $\boldsymbol{u}$ is its "shadow". It is denoted by* $\text{proj}_{\boldsymbol{u}}(\boldsymbol{y})$.



If one lets the projection be $\alpha\boldsymbol{u}$ and requires that $y - \text{proj}_{\boldsymbol{u}}(\boldsymbol{y})$ is orthogonal to $\boldsymbol{u}$, some algebra produces the following formula:

Fact 9.8 *For vectors $\boldsymbol{y}$ and $\boldsymbol{u}$, the projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$ is given by:*

$$\text{proj}_{\boldsymbol{u}}(\boldsymbol{y}) = \frac{\boldsymbol{y} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}} \boldsymbol{u}$$

Example. *Calculate* $\text{proj}_{\boldsymbol{b}}(\boldsymbol{a})$ *and* $\text{proj}_{\boldsymbol{a}}(\boldsymbol{b})$ *for* $\boldsymbol{a} = (3, 4)$ *and* $\boldsymbol{b} = (-5, 2)$. *Then* $\text{proj}_{\boldsymbol{b}}(\boldsymbol{a}) = (35/29, -14/29)$ *and* $\text{proj}_{\boldsymbol{a}}(\boldsymbol{b}) = (-21/25, -28/25)$.

One can also define the (orthogonal) projection $\text{proj}_W(\boldsymbol{y})$ of the vector $\boldsymbol{y}$ onto the vector space W. If we think of $\boldsymbol{y}$ as a point, then the projection of it onto W is the closest point of W to it.

It can be shown that, if W has an orthonormal basis, then the projection of $\boldsymbol{y}$ onto W is given by a simple formula: the coefficients of the basis are just the dot-product of $\boldsymbol{y}$ with each of them.

Fact 9.9 *If W is a subspace with orthonormal basis $\{\boldsymbol{w}_i\}$, then*

$$\text{proj}_W(\boldsymbol{y}) = \sum_i (\boldsymbol{y} \cdot \boldsymbol{w}_i)\, \boldsymbol{w}_i$$

We omit the proof of the following result:

Fact 9.10 *If W is a subspace of V, then every vector $\boldsymbol{y}$ in V can be written uniquely in the form $\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z}$, where $\hat{\boldsymbol{y}}$ in W and $\boldsymbol{z}$ in $W^\perp$.*

In the above result, $\hat{\boldsymbol{y}} = \text{proj}_W(\boldsymbol{y})$.

Example. *Consider in* $\mathbb{R}^3$ *the plane* P *given by* $3x + 4y - z = 0$. *Write the vector* $\boldsymbol{v} = (9, 9, 11)$ *as the sum of vector in* P *and vector in* $P^\perp$.

Well, one way to proceed is to calculate the vector in P, *which by the above result is* $\text{proj}_P(\boldsymbol{v})$; *if we use the above method we need an orthonormal basis of* P *(which we don't yet know how to find).*

An alternative approach is that we noted that any vector in $P^\perp$ *is a multiple of* $\boldsymbol{w} = (3, 4, -1)$. *So if we assume the requisite vector in* $P^\perp$ *is* $a\boldsymbol{w}$, *then we need* $(\boldsymbol{v} - a\boldsymbol{w}) \cdot \boldsymbol{v} = 0$. *This solves to* $a = 2$; *so* $\boldsymbol{v} = 2\boldsymbol{w} + (3, 1, 13)$.

9.5 Iterative Orthogonalization

There is a famous process (usually called Gram-Schmidt) that can be used to take a set of vectors and produce a set of vectors with the same span as the original but whose vectors are pairwise orthogonal. The process works through the set of vectors one at a time. Each time it "straightens" the vector with respect to the previous ones. The key point is that it suffices to subtract the projection of the vector onto the previous vectors.

ALGOR *Input: collection* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ *of linearly independent vectors.*
Output: collection $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ *of orthogonal vectors that span the same space.*
Process: Generate vectors $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, $\boldsymbol{y}_3$, $\ldots$ *by*

$$\boldsymbol{y}_i = \boldsymbol{x}_i - \sum_{j=1}^{i-1} \text{proj}_{\boldsymbol{y}_j}(\boldsymbol{x}_i)$$

The process can be justified by checking that each requisite dot product is zero.

Example. *Find an orthonormal basis of the subspace spanned by the following three vectors.*

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{x}_2 = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 3 \end{bmatrix} \qquad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ 1 \\ 1 \\ -7 \end{bmatrix}$$

The above process produces an orthogonal basis:

$$\boldsymbol{y}_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \qquad \boldsymbol{y}_2 = \boldsymbol{x}_2 - \frac{2}{2}\boldsymbol{y_1} = \begin{bmatrix}1\\-1\\-1\\3\end{bmatrix} \qquad \boldsymbol{y}_3 = \boldsymbol{x}_3 - \frac{4}{2}\boldsymbol{y}_1 - \frac{-20}{12}\boldsymbol{y}_2 = \begin{bmatrix}8/3\\-8/3\\-2/3\\-2\end{bmatrix}$$

So after normalization we have the vectors
$\frac{1}{\sqrt{2}}(1,1,0,0)$, $\frac{1}{\sqrt{12}}(1,-1,-1,3)$, *and* $\frac{1}{\sqrt{42}}(4,-4,-1,-3)$

Since one can take an orthogonal basis and normalize each vector, the above algorithm provides a constructive proof of the earlier claim that every vector space has an orthonormal basis.

Practice

9.1. For each of the following triples, determine h such that the triple is orthogonal.
   (a) $(1,0,0)$, $(0,1,1)$, $(0,1,h)$
   (b) $(1,2,-2)$, $(2,3,4)$, $(0,1,h)$
   (c) $(h,h,0)$, $(h,0,h)$, $(0,h,h)$

9.2. Prove that the formula in Fact 9.8 for the orthogonal projection of one vector onto another is correct.

9.3. (a) Find an orthonormal basis for the plane through the points $(1,0,-1)$, $(3,2,5)$, and $(0,0,0)$.
   (b) Find the closest point on the plane to $(2,1,3)$.

Solutions to Practice Exercieses

9.1. (a) $h=-1$
   (b) DNE
   (c) $h=0$

9.2. Say projection $\boldsymbol{P} = \alpha\boldsymbol{u}$. Then we need $(\boldsymbol{y}-\boldsymbol{P})\cdot\boldsymbol{u} = 0$. But $(\boldsymbol{y}-\boldsymbol{P})\cdot\boldsymbol{u} = \boldsymbol{y}\cdot\boldsymbol{u} - \alpha\boldsymbol{u}\cdot\boldsymbol{u}$. Solve for α.

9.3. (a) The plane is generated by the vectors $(1, 0, -1)$ and $(3, 2, 5)$. From this we calculate that an orthogonal basis for this plane is $(1, 0, -1)$ and $(4, 2, 4)$ and normalized this is $\frac{1}{\sqrt{2}}(1, 0, -1)$ and $\frac{1}{6}(4, 2, 4)$.
(b) Using Fact 9.9 we get: $(\frac{35}{18}, \frac{11}{9}, \frac{53}{18})$.

10 A Taste of Further Applications

Further applications of linear algebra include: (i) ranking sports teams and websites; (ii) binary vector spaces and error-correcting codes; and (iii) principal component analysis. We give here an extremely brief introduction.

10.1 Ranking Sports Teams and Websites

10.1.1 Massey's Original Rating Idea

Consider trying to assign ratings to sports teams. By **rating** we mean a numerical value. (From this one can produce ordinal values, if desired.) The following is the original idea for rating put forward by Massey. Suppose there was a rating such that: *the result (net score) between any teams was always given by the difference in their ratings.* Then one obtains a system of equations of the form

$$r_i - r_j = g_{ij}$$

where r_i and r_j are the (unknown) ratings of teams i and j, and g_{ij} is the actual net score of the game between them.

So the task is, given the actual results g_{ij}, to determine the r_i. There are of course probably many more equations than variables; so this system of linear equations almost surely has no solution (it is "over-determined"). So instead, we try to find the "best" vector $\boldsymbol{r}$. The standard approach is to minimize the sum of the squares of the errors, where here the error is $g_{ij} - (r_i - r_j)$. With a few lines of (vector) calculus, one obtains a system of equations:

$$\text{for each } i\text{:} \quad \text{net rating for team } i = \text{net result for team } i$$

where net rating is the sum of rating differences (the $r_i - r_j$) and net result is the net score (sum of g_{ij}). That is, we need to solve the matrix equation

$$S\boldsymbol{r} = \boldsymbol{N}$$

where S is square matrix such that each diagonal entry is the number of games that team played; and off the diagonal the value is -1 if the corresponding teams played, and 0 otherwise. And vector $\boldsymbol{N}$ is the net score for each team: points-for minus points-against.

However, matrix S is not invertible. There are two ways to see this. The first is that each row of S sums to 0, and thus the all-1 vector is in its null-space. The second way to see this is that the solution for $\boldsymbol{r}$ is not unique: one can add the same constant to all entries and still have a solution. So the way to resolve this is to discard one row of S (since it is redundant), and add a row that normalizes the ratings. For example, add the equation that the ratings sum to 0. Then go ahead and solve, using some software.

10.1.2 Page's Idea

Consider ranking a collection of websites. Suppose there is a bot that does a **random walk** on this collection: at each site it randomly chooses one outgoing link and follows it. (Sometimes called a Markov process.) The idea behind what is called PageRank is that maybe the "importance" of a website is determined by the average proportion of time the bot spends there. Note that one needs to deal with websites with no outgoing links.

Assuming this random-walk proportion converges to some vector π, then you get system of equations saying

for each i: π_i equals sum of $\frac{1}{c_j}\pi_j$

where the sum is over all j that have a link to i, and c_j is how many outgoing links j has in total. This can be written as the matrix equation $\pi = A\pi$ for some matrix A. So what we want is the eigenvector of A for the eigenvalue 1.

10.2 An Introduction to Error-Correcting Codes

10.2.1 Binary Vector Spaces

For this section the arithmetic will only use the numbers `0` and `1`. That is, arithmetic will be "modulo 2". This is defined as the remainder when the ordinary result is divided by 2. So the arithmetic tables are

$+$	0	1
0	0	1
1	1	0

$\times$	0	1
0	0	0
1	0	1

Define $\mathcal{Z}^n$ as the set of all vectors with n components each of which is either a 0 or a 1. (So there are 2^n vectors in $\mathcal{Z}^n$.) One can view $\mathcal{Z}^n$ as a vector space, where the only allowed scalars are 0 and 1, and all calculations are in binary arithmetic. Most of the definitions we made and theorems we proved about vector spaces are still valid.

10.2.2 Error-Detecting and Error-Correcting Codes

Suppose one has a binary message to send. Then the goal is to "enhance" the message (that is, add redundancy) so that even if the message is somewhat corrupted in transit, the original message can be recovered. (Note that this is different to encryption, where the goal is to send data that is private.)

EXAMPLE. *One idea is to take the message data and append a* ***check-bit****: the sum of the message bits in binary arithmetic. For example, if we have two message bits, then what is sent is either 000, 101, 011, 110. The claim is that: the receiver can detect if exactly one bit is changed in transit. But the message has to be resent.*

EXAMPLE. *Another idea is to repeat oneself. For example, send each message bit three times. So if we have two message bits, then what is sent is either 000000, 000111, 111000, 111111. The claim is that: the receiver can detect and correct a change in one bit.*

The above examples can be generalized. We define a **code** as a collection of binary vectors. For two vectors of the same length, their **Hamming distance** is the number of places the corresponding entries are different. The **distance** of a code is the minimum Hamming distance between any two vectors in it. It can be shown that:

Fact 10.1 *(a) A code can detect k errors if and only if its distance is at least $k+1$.*
(b) A code can correct k errors if and only if its distance is at least $2k+1$.

What one would like is a code that (a) is efficient (the sent message is not much longer than the original), and (b) is easy to decode (if the code has the property that one can correct some number of errors, then these corrections can be done quickly).

10.2.3 Linear Codes

A **linear code** is a subspace of $\mathcal{Z}^n$. The generator matrix M has rows that are the basis of the linear code. Hence, for message data d (treated as a row-vector or matrix with single row), what is sent is

$$s = dM$$

EXAMPLE. *We saw above the codes 000, 101, 011, 110 and 000000, 000111, 111000, 111111. Their generator matrices are:*

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

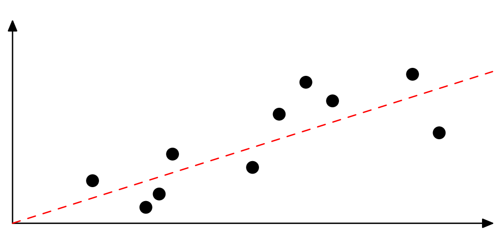
Here is the generator matrix of what is called the **Hamming code**. It can be thought of as adding three check-bits to four data-bits.

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

For example, suppose the data is 1101. Then the sent message is 1010101.

10.3 Principal Component Analysis

Consider a typical scatter plot. One can shift it so that regression line goes through origin. Then treat the regression line as new axis. In 2D, the projection onto regression line gives the best dispersion of data.



More generally, suppose data matrix A has rows of observations (where columns represent statistical variables). Consider $V = A^T A$. This matrix is square and symmetric, so its eigenvalues are real. In fact it turns out all its eigenvalues are nonnegative, and eigenvectors from different eigenvalues are orthogonal. The **principal components** of data A are the (unit) eigenvectors of $A^T A$. This is related to what is called the covariance matrix in statistics.

We know these eigenvectors form an eigen-basis of $\mathbb{R}^n$. So one can represent each vector in terms of the coordinates. Because the eigenvectors are orthogonal, the projections are independent. We write the eigenvectors in decreasing order of the eigenvalues. If the eigenvalues decay quickly, then one can approximate each vector by only a few entries in this coordinate system. This transforms a high-dimension system into a low-dimension system.

10.4 Some References

Statistical Models Applied to the Rating of Sports Teams, Kenneth Massey, 1997.

Who's #1? The science of rating and ranking. A.N. Langville and C.D. Meyer, Princeton University Press, 2013.