Chapter 1: Linear Equations

(Last Updated: August 12, 2020)

The material for these notes is derived primarily from *Linear Algebra and its applications* by David Lay (4ed).

1. Systems of Linear Equations

Before we spend a significant amount of time laying the groundwork for Linear Algebra, let's talk about some linear algebra you already know.

Example. Consider the following systems of equations:

$$\begin{aligned} x + 2y &= 3 \\ x + 2y &= 4 \end{aligned} \qquad \begin{aligned} 2x - y &= 4 \\ x + y &= -1 \end{aligned} \qquad \begin{aligned} x - y &= 1 \\ -2x + 2y &= -2 \end{aligned}$$

The first system has no solutions. One can see this by solving (via elimination or substitution) or by recognizing these (linear) equations as equations of two parallel lines. That is, intersections between the lines correspond to common solutions for the equations.

The second system has exactly one solution. The corresponding lines intersect at one point $(1, -2)$.

The third system has infinitely many solutions as both equations correspond to the same line.

Exercise. Solve the following system of equations in three variables using the elimination method.

$$\begin{aligned} x + y - z &= 4 \\ 2x - y + 3z &= -13 \\ -x + 2y - z &= 8 \end{aligned}$$

Interpret your solution geometrically.

We will now define several of the terms we have already used.

Definition. The equation $a_1x_1 + \cdots + a_nx_n = b$ is called linear with variables x_i, coefficients a_i (real or complex) and constant b. A solution to a linear equations is a set $(s_1, \ldots, s_n)$ such that substituting the s_i for x_i in the left-hand side produces a true statement.

A system of linear equations is a set of linear equations in the same variables and a solution to the system is a common solution to all the equations in the system. A system is consistent if it has at least one solution and inconsistent if it has no solution. Two systems with the same solution set are said to be equivalent.

When confronted with a system, we are most often interested in the following two questions:

(1) (Existence) Is the system consistent?

(2) (Uniqueness) If the system consistent, is there a *unique* solution?

What we will find is that solving a system of equation can be done much more quickly and efficiently using matrix techniques. First we will lay out notation for this process. Subsequently we will outline the process (Gaussian elimination) and explain how it mirrors the elimination method.

An $m \times n$ matrix M is a rectangular array with m rows and n columns. We denote by M_{ij} the entry in the ith row and jth column of the matrix M.

Consider a system of m equations in n variables. The $m \times n$ matrix C formed by setting C_{ij} to be the coefficient of x_j in the ith equation is called the coefficient matrix of this system. The augmented matrix A is an $m \times (n+1)$ matrix formed just as C but whose last column contains the constants of each system.

Example. Consider the system from the exercise above. The coefficient matrix and augmented matrix of this system are

$$\begin{bmatrix} 1 & 1 & -1 \\ 2 & -1 & 3 \\ -1 & 2 & -1 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & -1 & 3 & -13 \\ -1 & 2 & -1 & 8 \end{bmatrix}$$

Exercise. Recall your solution to the previous exercise. In each step, write the augmented matrix of the system. What observations can you make about the final matrix?

Each *action* we take in solving a system (via elimination) corresponds to an operation on the augmented matrix of the system. We will make these operations more precise now.

Elementary Row Operations.

(1) (Replacement) Replace one row by the sum of itself and a multiple of another row.

(2) (Interchange) Interchange two rows.

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

Two matrices are said to be row equivalent if one is obtainable from the other by a series of elementary row operations. It then follows that two linear systems are equivalent if and only if their augmented matrices are row equivalent.

2. Row reduction and echelon forms

The process we lay out in this section is essential to everything we do in this course. In many ways, row reduction of a matrix is just the elimination method for solving systems.

The leading entry of a row in a matrix is the first nonzero entry when read left to right.

Definition. A rectangular matrix is in (row) echelon form if it has the following three properties.

(1) All nonzero rows are above any rows of all zeros.

(2) Each leading entry of a row is in a column to the right of the leading entry of the row above it.

(3) All entries in a column below a leading entry are zeros.

This form is reduced if in addition

(4) The leading entry in each nonzero row is 1.

(5) Each leading 1 is the only nonzero entry in its column.

Example. The following matrix is in row echelon form (REF): $\begin{bmatrix} 3 & 1 & 2 & -1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 5 & -3 \end{bmatrix}$.

The following matrix is in reduced row echelon form (RREF): $\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$.

Any matrix may be row reduced (via the elementary row operations) into a matrix in REF, but this matrix is not unique.

Theorem 1. Every matrix is equivalent to one and only one matrix in RREF form. Thus, two matrices with the same RREF form are row equivalent.

We won't prove this right now, but I hope to later once we have learned about linear independence/dependence. Hence, it makes sense to speak of *the* echelon form of the matrix.

Definition. A pivot position in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A pivot column is a column of A that contains a pivot position.

Row reduction algorithm (Gaussian Elimination)

(1) Begin with the leftmost nonzero column (this is a pivot column).

(2) Interchange rows as necessary so the top entry is nonzero.

(3) Use row operations to create zeros in all positions below the pivot.

(4) Ignoring the row containing the pivot, repeat 1-3 to the remaining submatrix. Repeat until there are no more rows to modify.

(5) Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. Make each pivot 1 by scaling.

Example. Put the following matrix into RREF. $\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{bmatrix}$

Note that this example corresponds to a system with a unique solution, $(x, y, z) = (1, -2, 2)$.

But of course, solutions need not be unique (or exist at all).

Example. The following matrix row reduces as

$$\begin{bmatrix} 1 & 2 & 4 & 5 \\ 1 & 2 & 4 & 6 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The reduced form of the augmented matrix has a pivot in each column. This corresponds to an inconsistent system. The reason for this is that, translating back to a system, we get the equations $x = 0$, $y = 0$, $z = 0$, and $0 = 1$. This last equation is impossible so there is no solution to the system.

Example. The following matrix is in RREF.

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

This corresponds to the system

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

The variables x_1 and x_2 correspond to pivot columns. These are called basic variables. Since x_3 does not correspond to a pivot column, it is called a free variable. This is because there is a solution for *any* choice of x_3.

One way to write the solution is in parametric form with free variables listed as such and basic variables solved for in terms of the free variables.

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

Theorem 2. A (linear) system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column. If a linear system is consistent, then the solution set contains either (i) a unique solution (no free variables) or (ii) infinitely many solutions.

The first condition in the theorem is equivalent to no row of the form $\begin{bmatrix} 0 & \cdots & 0 & b \end{bmatrix}$, $b \neq 0$ in the RREF form of the matrix.

An optimal way to represent our solutions, for reasons that will become clear later, is in vector form. We will study vectors in more detail in the next section. For now, just think of vectors as a $n \times 1$ matrix wherein we list the variables of the system. We substitute so that the free variable(s) is/are the only visible variables. Split the vectors according to the free variables/constants and factor out free variables.

Example. Write the solution to the previous problem in vector form.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5x_3 + 1 \\ -x_3 + 4 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5x_3 \\ -x_3 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} 1 \\ 4 \\ 0 \end{bmatrix}.$$

The part corresponding to the free variables is the homogeneous solution and the part corresponding to constants is the particular solution.

What this means is that we can substitute *any* (real) value for x_3 and after adding the two vectors together (pointwise), we get a solution to the system. More on vectors in the next section.

Exercise. The following matrix is in RREF. Translate this into a system and identify the basic and free variables. Write the solution to this system in parametric form and then vector form.

$$\begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

3. Vector equations

A matrix with one column is said to be a column vector, which for now we will just call a vector and denote it by $\mathbf{v}$ or $\vec{v}$. (There is a corresponding notion of a row vector but columns are more appropriate for our use now.) The dimension of a vector is the number of rows.

Example. $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} e \\ \pi \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, w_i \in \mathbb{R}.$

Two vectors are **equal** if they have the same dimension and all corresponding entries are equal. A vector whose entries are all zero is called a zero vector and denoted $\mathbf{0}$.

We denote the set of n-dimensional vectors with entries in $\mathbb{R}$ (resp. $\mathbb{C}$) by $\mathbb{R}^n$ (resp. $\mathbb{C}^n$). The standard operations on vectors in $\mathbb{R}^n$ (or $\mathbb{C}^n$) are scalar multiplication and addition.

(Scalar Multiplication) For $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$,

$$c\mathbf{v} = c \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}.$$

(Addition) For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

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

Example. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$. Compute $\mathbf{u} + 2\mathbf{v}$.

For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and all scalars $c, d \in \mathbb{R}$, we have the following algebraic properties of vectors.

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

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

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

(iv) $\mathbf{u} + (-\mathbf{u}) = \mathbf{u} + (-1)\mathbf{u} = 0$

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

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

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

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

An aside on $\mathbb{R}^2$. We visualize a vector $\begin{bmatrix} a \\ b \end{bmatrix}$ as an arrow with endpoint at $(0,0)$ and pointing to (a,b). Any scalar multiple by $c \in \mathbb{R}$, $c \neq 0, 1$, of $\begin{bmatrix} a \\ b \end{bmatrix}$ points in the same direction but is longer if $c > 1$ and shorter if $0 < c < 1$. Vector addition can be visualized via the parallelogram rule. It is important to remember this geometric fact: a parallelogram is uniquely determined by three points in the plane.

Parallelogram Rule for Addition. If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ are represented by points in the plan, then $\mathbf{u} + \mathbf{v}$ corresponds to the fourth vertex of a parallelogram whose other vertices are $\mathbf{0}, \mathbf{u}$, and $\mathbf{v}$.

Example. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Find $\mathbf{u} + \mathbf{v}$ geometrically and confirm that it is correct via the rule above for vector addition.

Now we turn to one of the most fundamental concepts in linear algebra: span. The other (linear independence) will be introduced in section 7.

Definition. A linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ with weights $c_1, \ldots, c_p \in \mathbb{R}$ is defined as the vector $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. The set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is called the span of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ (or the subset of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_p$) and is denoted $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

Geometrically, we think of the span of one vector $\mathbf{v}$ as line through the origin since any vector in $\operatorname{Span}\{\mathbf{v}\}$ is of the form $x\mathbf{v}$ for some scalar (weight) x. Similarly, the span of two vectors $\mathbf{v}_1, \mathbf{v}_2$ which are not scalar multiples forms a plane through the three points $\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2$.

A reasonable question is when a given vector is in the span of a particular set of vectors?

Example. Determine whether $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}$ is a linear combination of $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$.

We are asking whether there exists x_1, x_2 such that $x_1\mathbf{u} + x_2\mathbf{v} = \mathbf{w}$, which is equivalent to

$$\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} x_2 = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 3x_1 + 2x_2 \\ -x_1 + x_2 \\ x_1 - 3x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}.$$

This gives a system of three equations and two unknowns. We form the corresponding augmented matrix and row reduce to find the solution $x_1 = -1$ and $x_2 = 2$.

Theorem 3. Let $\mathbf{e}_i \in \mathbb{R}^n$ denote the vector of all zeros except a 1 in the ith spot. Then $\mathbb{R}^n = \operatorname{Span}\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Proof. Clearly any linear combination of the $\mathbf{e}_i$ lives in $\mathbb{R}^n$. Suppose $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n$. Then $\mathbf{a} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n \in \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. $\square$

A vector equation $x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{b}$ has the same solution as the linear system whose augmented matrix is $\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n & \mathbf{b} \end{bmatrix}$. In particular, $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ if and only if there exists a solution to the corresponding linear system. A linear system/vector equation is said to be homogeneous if $\mathbf{b} = \mathbf{0}$. Such a system is *always* consistent since we can take $x_i = 0$ for all i. This solution is known as the trivial solution. A homogeneous system has a nontrivial solution if and only if the equation has at least one free variable.

Example. Let $\mathbf{v}_1 = \begin{bmatrix} 5 \\ -5 \\ 25 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 3 \\ -6 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 2 \\ -2 \\ 10 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ 5 \end{bmatrix}$. Find all solutions to the homogeneous system $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$ and the system $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}$.

We form the augmented matrix of the system and row reduce,

$$\begin{bmatrix} 5 & 0 & 2 & 0 \\ -5 & 3 & -2 & 0 \\ 25 & -6 & 10 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & \frac{2}{5} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We see that x_3 is a free variable. In standard parametric form, the solution is

$$\begin{cases} x_1 = -\frac{2}{5}x_3 \\ x_2 = 0 \\ x_3 \text{ is free.} \end{cases}$$

Hence, all solutions could be represented in parametric vector form $\begin{bmatrix} -2/5 \\ 0 \\ 1 \end{bmatrix} x_3$.

We repeat with the non-homogeneous system. We form the augmented matrix of the system and row reduce,

$$\begin{bmatrix} 5 & 0 & 2 & 1 \\ -5 & 3 & -2 & -1 \\ 25 & -6 & 10 & 5 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & \frac{2}{5} & \frac{1}{5} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

In parametric form the solution is

$$\begin{cases} x_1 = -\frac{2}{5}x_3 + \frac{1}{5} \\ x_2 = 0 \\ x_3 \text{ is free} \end{cases}$$

and in vector form

$$\begin{bmatrix} -2/5 \\ 0 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} 1/5 \\ 0 \\ 0 \end{bmatrix}.$$

This solution is similar to the one before. We call the portion of the solution containing the free variable x_3 the homogeneous solution of the system.

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

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

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

Observe that the result will be a (column) vector with m rows.

Example. Let $A = \begin{bmatrix} -1 & 2 & 0 \\ 3 & 1 & 1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$. Compute $A\mathbf{x}$.

Example. Consider the system from last lecture:

$$\begin{aligned} 3x_1 + 2x_2 &= 1 \\ -x_1 + x_2 &= 3 \\ x_1 - 3x_2 &= -7. \end{aligned}$$

We could write this as the matrix equation

$$\begin{bmatrix} 3 & 2 \\ -1 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ -7 \end{bmatrix}.$$

A solution to the matrix equation $A\mathbf{x} = \mathbf{b}$ is a vector $\mathbf{s} \in \mathbb{R}^n$ such that replacing $\mathbf{x}$ with $\mathbf{s}$ and multiplying produces the vector $\mathbf{b} \in \mathbb{R}^m$. We say the matrix equation is consistent if at least one such $\mathbf{s}$ exists. The equation $A\mathbf{x} = \mathbf{b}$ is said to be homogeneous if $\mathbf{b} = \mathbf{0}$.

Aside on inverses. If $A = [a]$ is a 1×1 matrix and $\mathbf{x}$ a column vector with 1 entry, so $\mathbf{x} = x$, then the equation $A\mathbf{x} = ax = b$ where $b \in \mathbb{R}$. Assuming $a \neq 0$, the solution to this equation is just $x = a^{-1}b$. Can we generalize this to matrix equations? What does A^{-1} mean? We'll return to this question in Chapter 2.

Theorem 4. If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{b} \in \mathbb{R}^m$, the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution as the vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$, which in turn has the same solution set as the system of linear equations whose augmented matrix is $\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$.

Another way to read the previous theorem is this: The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Example. Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and write $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent *for all* choices of $\mathbf{b}$?

(Partial solution) Row reduce $A|\mathbf{b}$ to REF form

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) \end{bmatrix}.$$

If the last column contains a pivot then there is no solution. Hence, the matrix equation has a solution if and only if $b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) = 0$.

Let's summarize what we've got so far and try to put some rigor behind these statements.

Theorem 5. Let A be an $m \times n$ matrix. The following are equivalent.

(1) For each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a solution.

(2) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(3) The columns of A span $\mathbb{R}^m$.

(4) A has a pivot in every row.

Proof. (1) $\Rightarrow$ (2). Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ denote the columns of A. By definition of the matrix product,

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

and so by definition, $\mathbf{b}$ is a linear combination of the $\mathbf{a}_i$.

(2) $\Rightarrow$ (3). If $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A, then $\mathbf{b} \in \text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$. Hence, $\text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \mathbb{R}^m$.

(3) $\Rightarrow$ (4). Suppose A did not have a pivot in every row. Then, in particular, A does not have a pivot in the last row. But then $\mathbf{b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$ is not in the span of RREF(A). But this implies that $\text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \neq \mathbb{R}^m$.

(4) $\Rightarrow$ (1). Since A has a pivot in every row, then the statement is clearly true for RREF(A). Since $[A|b]$ and RREF$([A|b])$ have the same solution space, then the claim holds. $\square$

5. Solution sets of linear systems

Recall, a system of linear equations is said to be homogeneous if it can be written as $A\mathbf{x} = \mathbf{0}$ where A is an $m \times n$ matrix and $x \in \mathbb{R}^m$. Such a system always has one solution, $\mathbf{0}$, called the trivial solution. The homogeneous system has a nontrivial solution if and only if the equation has at least one free variable.

Example. Solve the following homogeneous 'system' with one equation: $3x_1 + 2x_2 - 5x_3 = 0$.

A general solution is $x_1 = -\frac{2}{3}x_2 + \frac{5}{3}x_3$ with x_2 and x_3 free. We can write the solution in parametric vector form as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3}x_2 + \frac{5}{3}x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} x_3.$$

Let $\mathbf{u} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix}$. Then the solution set is $\mathrm{Span}\{\mathbf{u}, \mathbf{v}\}$, which represents a plane through the origin in $\mathbb{R}^3$.

In general, the parametric vector form of a solution set is of the form $\mathbf{x} = \mathbf{u}_1 x_{i_1} + \cdots + \mathbf{u}_k x_{i_k} + \mathbf{p}$ where the x_{i_j} are free variables and $\mathbf{p}$ is a particular solution not contained in $\mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$.

Example. Solve the following 'system' with one equation: $3x_1 + 2x_2 - 5x_3 = -1$.

We recognize that this is almost the same equation as before and, in fact, it should have the same homogeneous solution. We need a *particular solution.* One such (relatively obvious) solution is $(-1/3, 0, 0)$.

As before, we write the solution in parametric vector form as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3}x_2 + \frac{5}{3}x_3 - \frac{1}{3} \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} 5/3 \\ 0 \\ 1 \end{bmatrix} x_3 + \begin{bmatrix} -1/3 \\ 0 \\ 0 \end{bmatrix}.$$

This corresponds to translation of the homogeneous solution by the vector $\mathbf{p} = \begin{bmatrix} -1/3 \\ 0 \\ 0 \end{bmatrix}$.

The next theorem is another version of the fact we discussed previously on choices of solutions.

Theorem 6. Suppose $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$ and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Proof. Suppose $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$ as given above. Then

$$A\mathbf{w} = A(\mathbf{p} + \mathbf{v}_h) = A\mathbf{p} + A\mathbf{v}_h = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

That is, $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$. Now let $\mathbf{y}$ be some solution. We claim $\mathbf{y}$ has the form given in the theorem. Then

$$A(\mathbf{y} - \mathbf{b}) = A\mathbf{y} - A\mathbf{p} = \mathbf{0} - \mathbf{0} = \mathbf{0},$$

so $(\mathbf{y} - \mathbf{p})$ is a solution of $A\mathbf{x} = \mathbf{0}$. Hence, $\mathbf{y} = \mathbf{p} + (\mathbf{y} - \mathbf{p})$ has the desired form. $\square$

6. Applications of linear systems

We will only discuss one type of application right now. We may return to others as time permits.

Chemical equations describe quantities of substances consumed and produced by chemical reactions. (Friendly reminder: atoms are neither destroyed nor created, just changed.)

Example. When propane gas burns, propane C_3H_8 combines with oxygen O_2 to form carbon dioxide CO_2 and water H_2O according to an equation of the form

$$(x_1)\mathrm{C_3H_8} + (x_2)\mathrm{O_2} \longrightarrow (x_3)\mathrm{CO_2} + (x_4)\mathrm{H_2O}.$$

To *balance the equation* means to find x_i such that the total number of atoms on the left equals the total on the right. We translate the chemicals into vectors $\begin{bmatrix} C \\ H \\ O \end{bmatrix}$,

$$\mathrm{C_3H_8} = \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix}, \qquad \mathrm{O_2} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \qquad \mathrm{CO_2} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \qquad \mathrm{H_2O} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Balancing now becomes the linear system,

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}.$$

Equivalently,

$$x_1 \begin{bmatrix} 3 \\ 8 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} - x_3 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - x_4 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \mathbf{0}.$$

We form the augmented matrix and row reduce,

$$\begin{bmatrix} 3 & 0 & -1 & 0 & 0 \\ 8 & 0 & 0 & -2 & 0 \\ 0 & 2 & -2 & -1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & -1/4 & 0 \\ 0 & 1 & 0 & -5/4 & 0 \\ 0 & 0 & 1 & -3/4 & 0 \end{bmatrix}.$$

Thus, the solution is $x_1 = (1/4)x_4$, $x_2 = (5/4)x_4$, $x_3 = (3/4)x_4$ with x_4 free. Since only nonnegative solutions make sense in this context, any solution with $x_4 \geq 0$ is valid. For example, setting $x_4 = 4$ gives

$$\mathrm{C_3H_8} + 5\mathrm{O_2} \longrightarrow 3\mathrm{CO_2} + 4\mathrm{H_2O}.$$

7. Linear Independence

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

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0} \tag{1}$$

has only the trivial solution. Otherwise the set is said to be linearly dependent. That is, there exist weights $c_1, \dots, c_p$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}.$$

Example. Define the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -1 \\ 2 \\ 7 \end{bmatrix}$. Show that the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

We set up the augmented matrix corresponding to the equation (1) and row reduce

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 3 & 5 & 2 & 0 \\ -2 & -1 & 7 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Hence, the only solution is the trivial one and so the set is linearly independent.

Example. Define the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$. Show that the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

We set up the augmented matrix corresponding to the equation (1) and row reduce

$$\begin{bmatrix} 1 & 4 & 2 & 0 \\ 2 & 5 & 1 & 0 \\ 3 & 6 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There is a nontrivial solution (in particular, x_3 is free). Hence, the set is linearly dependent.

Example. (1) The set $\{\mathbf{0}\}$ is linearly *dependent* because $c\mathbf{0} = \mathbf{0}$ for all $c \in \mathbb{R}$. Moreover, any set containing the zero vector is linearly dependent. To see this, suppose $\mathbf{v}_1 = \mathbf{0}$, then $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$. A similar argument holds if any $\mathbf{v}_i$ is the zero vector.

(1) A set of one nonzero vector, $\{\mathbf{v}\}$ is linearly *independent* because $c\mathbf{v} = \mathbf{0}$ implies $c = 0$.

(2) A set of two vectors is linearly dependent if and only if one vector is a multiple of the other. (Exercise)

Theorem 7. An indexed set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, $p \geq 2$, is linearly dependent if and only if at least one of the vectors is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq 0$, then some $\mathbf{v}_j$, $j > 1$, is a linear combination of the preceding vectors.

Proof. Suppose S is linearly dependent. Then there exists weights $c_1, \ldots, c_p$ such that

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

If $c_1 \neq 0$, then $\mathbf{v}_1 = \frac{c_2}{c_1}\mathbf{v}_2 + \cdots + \frac{c_p}{c_1}\mathbf{v}_p$, so v_1 is a linear combination of $\mathbf{v}_2, \ldots, \mathbf{v}_p$. Note that it must be true that at least one of $c_2, \ldots, c_p$ is nonzero.

Conversely, if $\mathbf{v}_1$ is a linear combination of $\mathbf{v}_2, \ldots, \mathbf{v}_p$, then there exists weights $c_2, \ldots, c_p$ such that $\mathbf{v}_1 = c_2 v_2 + \cdots + c_p v_p$. Equivalently, $-\mathbf{v}_1 + c_2 v_2 + \cdots + c_p v_p = \mathbf{0}$, so S is linearly dependent.

Both arguments hold with any vector in place of $\mathbf{v}_1$. $\square$

Example. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ with $\mathbf{u}$, $\mathbf{v}$ linearly independent. Then $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent if and only if one of the vectors is in the span of the other two.

Theorem 8. Any set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} \subset \mathbb{R}^n$ is linearly dependent if $p > n$.

Example. The set $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \end{bmatrix} \right\}$ is linearly dependent.

Given an $m \times n$ matrix A, the rule $T(\mathbf{x}) = A\mathbf{x}$ defines a map from $\mathbb{R}^n$ to $\mathbb{R}^m$. We will be almost exclusively interested in these types of maps.

Example. Define $T(\mathbf{x}) = A\mathbf{x}$ where

$$A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}.$$

(1) Find $T(\mathbf{u})$.

(2) Find $\mathbf{x} \in \mathbb{R}^2$ whose image under T is $\mathbf{b}$. Is $\mathbf{x}$ unique?

(3) Is $\mathbf{c}$ in the range of T?

Definition. A transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$ (written $T : \mathbb{R}^n \to \mathbb{R}^m$) is a rule that assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ a vector $T(\mathbf{x}) \in \mathbb{R}^m$, called the image of $\mathbf{x}$. We call $\mathbb{R}^n$ the domain and $\mathbb{R}^m$ the codomain. The set of all images under T is called the range.

A transformation is linear if

(1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

(2) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $c \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^n$.

Note that, in this context, the range of T is the set of all linear combinations of the columns of A.

Example. Let $T(\mathbf{x}) = A\mathbf{x}$ where A is defined as follows.

(1) $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. T is a transformation $\mathbb{R}^3 \to \mathbb{R}^3$ but the range of T is "equivalent" in some way to $\mathbb{R}^2$. We say T is a projection of $\mathbb{R}^3$ onto $\mathbb{R}^2$.

(2) $A = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$. T is called a shear transformation because it leaves fixed the second component of $\mathbf{x}$.

(3) Consider the map $T : \mathbb{R} \to \mathbb{R}^2$ given by $T(\mathbf{x}) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$. This map is not linear (because $(a+b)^2 \neq a^2 + b^2$ in general).

Theorem 9. If T is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.

Proof. Let $\mathbf{y} = T(\mathbf{0})$. By linearity,

$$\mathbf{y} = T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0}) = 2\mathbf{y}.$$

Hence, $0 = 2\mathbf{y} - \mathbf{y} = \mathbf{y} = T(\mathbf{0})$. $\square$

The next theorem says that linear transformations and matrix transformations are the same thing. However, not all transformations are linear. I will leave it as an exercise to give an example of a transformation that is not linear.

Theorem 10. A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there exists a unique $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

Proof. ($\Leftarrow$) Suppose $T(\mathbf{x}) = A\mathbf{x}$. We must show that for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ and $A(c\mathbf{u}) = cA\mathbf{u}$. Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

$$\begin{aligned}
A(\mathbf{u}+\mathbf{v}) &= \mathbf{a}_1(u_1+v_1) + \cdots + \mathbf{a}_n(u_n+v_n) \\
&= (\mathbf{a}_1u_1 + \mathbf{a}_1v_1) + \cdots + (\mathbf{a}_nu_n + \mathbf{a}_nv_n) \\
&= (\mathbf{a}_1u_1 + \cdots + \mathbf{a}_nu_n) + (\mathbf{a}_1v_1 + \cdots + \mathbf{a}_nv_n) \\
&= A\mathbf{u} + A\mathbf{v}. \\
A(c\mathbf{u}) &= \mathbf{a}_1(cu_1) + \cdots + \mathbf{a}_n(cu_n) \\
&= c\mathbf{a}_1u_1 + \cdots + c\mathbf{a}_nu_n \\
&= c(\mathbf{a}_1u_1 + \cdots + \mathbf{a}_nu_n) \\
&= cA\mathbf{u}.
\end{aligned}$$

□

($\Rightarrow$) Suppose T is linear. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis vectors of $\mathbb{R}^n$. For $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

By linearity,

$$\begin{aligned}
T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) \\
&= x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n) \\
&= \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \mathbf{x}.
\end{aligned}$$

Set $A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$. We need only show that A is unique. Suppose $T(\mathbf{x}) = B\mathbf{x}$ for some $m \times n$ matrix B with columns $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$. Then $T(\mathbf{e}_1) = B\mathbf{e}_1 = \mathbf{b}_1$. But $T(\mathbf{e}_1) = A\mathbf{e}_1 = \mathbf{a}_1$, so $\mathbf{a}_1 = \mathbf{b}_1$. Repeating for each column of B gives $B = A$.

We call A the standard matrix of T.

Example. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point $\theta°$ counter-clockwise. Then $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$. Hence, the standard matrix of T is $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Example. Describe the transformations given by the following matrices,

$$\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

9. The matrix of a linear transformation

In this section we investigate some special properties that a linear transformation may possess.

Definition. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be onto if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at least one $\mathbf{x} \in \mathbb{R}^n$. T is one-to-one (1-1) if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at most one $\mathbf{x} \in \mathbb{R}^n$.

Another way to phrase one-to-one is $T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$.

Example. (1) Projections are onto but not 1-1.

(2) The map $T : \mathbb{R} \to \mathbb{R}^2$ given by $T(\mathbf{x}) = \begin{bmatrix} x \\ 2x \end{bmatrix}$ is 1-1 but not onto.

(3) The map $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ is 1-1 and onto.

We will develop criteria for 1-1 and onto based on the standard matrix of a linear transformation.

Theorem 11. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is 1-1 if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Proof. ($\Rightarrow$) Assume T is 1-1. Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. Because T is 1-1, $T(\mathbf{x}) = \mathbf{0} = T(\mathbf{0})$ implies $\mathbf{x} = \mathbf{0}$. Hence $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

($\Leftarrow$) Assume $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution. Suppose $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ so $\mathbf{x} - \mathbf{y} = \mathbf{0}$, so $\mathbf{x} = \mathbf{y}$ and T is 1-1. $\square$

The set $\{\mathbf{x} : T(\mathbf{x}) = \mathbf{0}\}$ is called the kernel of T. Another way to state the previous theorem is to say that T is 1-1 if and only if its kernel contains only $\mathbf{0}$.

Theorem 12. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A.

(1) T is onto if and only if the columns of A span $\mathbb{R}^m$.

(2) T is 1-1 if and only if the columns of A are linearly independent.

Proof. (1) The columns of A span $\mathbb{R}^m$ if and only if for every $\mathbf{b} \in \mathbb{R}^m$ there exists $\mathbf{x} \in \mathbb{R}^n$ such that $x_1\mathbf{a}_1 + \cdots + x_1\mathbf{a}_n = \mathbf{b}$. This is equivalent to $A\mathbf{x} = \mathbf{b}$, which is equivalent to $T(\mathbf{x}) = \mathbf{b}$ and this holds if and only if T is onto.

(2) T is 1-1 if and only if $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution and this holds if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. That is, $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$ has only the trivial solution. This is equivalent to the columns of A being linearly independent. $\square$

In the previous theorem, note that (1) is equivalent to A having a pivot in every row and (2) is equivalent to A having a pivot in every column.

Chapter 2: Matrix Algebra

(Last Updated: September 24, 2020)

These notes are derived primarily from *Linear Algebra and its applications* by David Lay (4ed).

1. Matrix operations

Write $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. Then entry a_{ij} is the ith entry of the jth column (ith row and jth column). The diagonal entries of A are those a_{ij} with $i = j$. A diagonal matrix is one such that $a_{ij} = 0$ if $i \neq j$. The zero matrix 0 is a matrix with $a_{ij} = 0$ for all i, j.

We'll discuss operations on matrices and how these correspond to operations on linear transformations. Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices of the same size $m \times n$ and $c \in \mathbb{R}$.

- Matrix addition: $A + B$ is defined as the $m \times n$ matrix whose i, j entry is $a_{ij} + b_{ij}$.
- Scalar multiplication: cA is defined as the $m \times n$ matrix whose i, j entry is ca_{ij}.

Example 1. Let $A = \begin{bmatrix} 3 & 2 \\ 1 & 5 \\ -1 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 1 & 2 \end{bmatrix}$, and $c = 3$. Compute $A + cB$.

Theorem 2. Let A, B, C be matrices of the same size and $r, s \in \mathbb{R}$ scalars.

(1) $A + B = B + A$
(2) $(A+B)+C = A+(B+C)$
(3) $A + 0 = A$
(4) $r(A + B) = rA + rB$
(5) $(r + s)A = rA + sA$
(6) $r(sA) = (rs)A$

It is worth observing the similarity between these properties and those of vectors. Later in the course we will study *vector spaces* (a generalization of $\mathbb{R}^n$ and $\mathbb{C}^n$). The space of $m \times n$ is another example of a vector space because it obeys the linearity properties of those spaces.

One can prove the above properties directly and without much difficulty. On the other hand, one could prove these by recognizing matrices as linear transformations and using those properties.

If T and S are linear transformations with standard matrices A and B, respectively, then $T + S$ is a linear transformation with standard matrix $A + B$. Similarly, if c is a scalar then the standard matrix of (cT) is just cA. What is the standard matrix associated to the composition of two standard matrices?

If $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^p \to \mathbb{R}^q$, then $T \circ S$ only makes sense of $q = n$. Let A and B again be the standard matrices of T and S, respectively, so A is $m \times n$ and B is $n \times p$. Then

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

Hence, the matrix product AB is the standard matrix of $T \circ S$.

$$\begin{tikzcd}
\mathbb{R}^p \arrow[r, bend left, "S", "B"'] \arrow[rr, bend right, "AB", "T \circ S"'] & \mathbb{R}^n \arrow[r, bend left, "T", "A"'] & \mathbb{R}^m
\end{tikzcd}$$

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

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \cdots \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & \cdots A\mathbf{b}_p \end{bmatrix}.$$

Example 3. Let $A = \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & 4 \\ -3 & 1 & 6 \end{bmatrix}$. Compute AB.

In Example 3, the product BA is not defined. In general, $AB \neq BA$ even if both are defined. Also, cancellation does not work. That is, it is possible for $AB = 0$ even if $A, B \neq 0$.

Another way of phrasing this definition is that each column of AB is a linear combination of the columns of A using weights from the corresponding columns of B. Let $A = (a_{ij})$ be $m \times n$ and $B = (b_{ij})$ $n \times p$. Let $C = (c_{ij}) = AB$. Using the row-column rule,

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}.$$

Example 4. Use the row-column rule to compute AB with A, B as in Example 3.

The $n \times n$ identity matrix is a diagonal matrix I_n with 1 for every diagonal entry.

Theorem 5. Let A, B, C be matrices of appropriate sizes such that each sum/product is defined and let $r \in \mathbb{R}$ be a scalar.

(1) $A(BC) = (AB)C$

(2) $A(B + C) = AB + AC$

(3) $(B + C)A = BA + CA$

(4) $r(AB) = (rA)B = A(rB)$

(5) $I_m A = A = A I_n$

If A is an $n \times n$ matrix and k a positive integer, then the matrix power A^k denotes the product of k copies of A. By definition, $A^0 = I_n$. If $A = (a_{ij})$ is an $m \times n$ matrix, the transpose of A, denoted A^T, is an $n \times m$ matrix whose columns are the rows of A. That is, the i, j entry of A^T is a_{ji}.

Example 6. Let A, B be as in Example 3. Compute A^T, B^T, and $(AB)^T$.

Theorem 7. Let A and B be matrices of appropriate sizes.

(1) $(A^T)^T = A$

(2) $(A + B)^T = A^T + B^T$

(3) $(rA)^T = r(A^T)$

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

2. The inverse of a matrix

In this section we only consider square matrices ($m = n$). Recall that the $n \times n$ identity matrix I_n is the diagonal matrix with 1s along the diagonal. For any $n \times n$ matrix A, we have $I_n A = AI_n = A$.

Definition 1. Let A be an $m \times n$ matrix. A is said to be left invertible if there exists an $n \times m$ matrix C such that $CA = I_n$. A is said to be right invertible if there exists an $n \times m$ matrix D such that $AD = I_m$. A is said to be invertible if it is both left and right invertible. We call C the inverse of A. A matrix which is not invertible is said to be singular.

Suppose A is invertible. Then

$$C = CI_m = C(AD) = (CA)D = I_n D = D.$$

Thus, $C = D$.

Exercise. Show that if A is invertible then it is necessarily $n \times n$.

Example 8. Let $A = \begin{bmatrix} 2 & 5 \\ -3 & 7 \end{bmatrix}$ and $C = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$. Verify that C is the inverse of A.

The next proposition explains why it is proper to refer to C as *the* inverse of A and not *an* inverse.

Proposition 9. Let A be an invertible $n \times n$ matrix. The inverse of A is unique.

Proof. Suppose B, C are inverses of A. Then $B = BI_n = B(AC) = (BA)C = I_n C = C$. Hence, $B = C$. $\square$

If A is invertible, we denote *the* inverse of A as A^{-1}.

Theorem 10. If A is an $n \times n$ invertible matrix, then for each $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof. If A is invertible, then clearly, $A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$, so $A^{-1}\mathbf{b}$ is a solution. We claim it is unique. Suppose $\mathbf{u}$ is another solution, then $A\mathbf{u} = \mathbf{b} = A(A^{-1}\mathbf{b})$. Multiplying both sides by A^{-1} on the left gives,

$$A^{-1}(A\mathbf{u}) = A^{-1}(A(A^{-1}\mathbf{b})) \Rightarrow (A^{-1}A)\mathbf{u} = (A^{-1}A)(A^{-1}\mathbf{b}) \Rightarrow I_n\mathbf{u} = I_n(A^{-1}\mathbf{b}) \Rightarrow \mathbf{u} = A^{-1}\mathbf{b}.$$

Hence, the solution is unique. $\square$

Theorem 11. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. If $ad - bc = 0$, then A is singular.

Proof. Suppose $ad - bc \neq 0$. We claim that A^{-1} satisfies the definition of the inverse of A.

$$AA^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc} \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = I_2,$$

Similarly, $A^{-1}A = I_n$. Hence, A^{-1} is the inverse of A. $\square$

Exercise. Prove that if $ad - bc = 0$, then A is singular. (Hint: Use Theorem 10. You might first consider the case that $a = 0$.)

Example 12. Use matrix inversion to solve the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7. \end{aligned}$$

The solution to the system above is the same as the solution to the matrix equation $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

By Theorem 11,

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{(3)(6)-(4)(5)} \begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \frac{-1}{2} \begin{bmatrix} -10 \\ 6 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

Theorem 13. Let A and B be invertible $n \times n$ matrices.

(1) A^{-1} is invertible and $(A^{-1})^{-1} = A$.
(2) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
(3) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

Proof. (1) follows from the symmetry in the definition of A^{-1}.

For (2), it suffices to show that $B^{-1}A^{-1}$ satisfies the definition of the inverse for AB by Proposition 9. By associativity,

$$(AB)(B^{-1}A^{-1}) = (A(BB^{-1}))A^{-1} = (AI_n)A^{-1} = AA^{-1} = I_n.$$

Similarly, $(B^{-1}A^{-1})(AB) = I_n$.

(3) Follows from Proposition 9 and Theorem 7,

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

Similarly, $(A^{-1})^T A^T = I_n$. $\square$

Next we want to show how one can obtain the inverse of a matrix in general. The method is quite basic, but to explain why it works takes a bit more muscle, so we'll hold off on that momentarily.

Theorem 14. An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms A_n into A^{-1}.

In other words, we form the augmented matrix $[A \mid I_n]$ and row reduce A to I_n (if possible). The result will be the augmented matrix $[I_n \mid A^{-1}]$.

Example 15. Find the inverse of the matrix

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

We form the augmented matrix $[A|I_3]$ and row reduce A to I_3.

$$[A|I_3] = \left[\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{array}\right].$$

Thus,

$$A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$$

An elementary matrix is one obtained by performing a single row operation to an identity matrix.

Example 16. The following are elementary matrices. Identify which row operation has been performed on I_3 to obtain each one.

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

Lemma 17. If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Proof. Let E be the elementary matrix obtained by multiplying row k of I_n by $m \neq 0$. That is,

$$E = \begin{bmatrix} \mathbf{e}_1 & \cdots & m\mathbf{e}_k & \cdots & \mathbf{e}_n \end{bmatrix}.$$

Let A be any $n \times n$ matrix. Then,

$$EA = E \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} E\mathbf{a}_1 & \cdots & E\mathbf{a}_n \end{bmatrix}.$$

It is clear that the result of the multiplication $E\mathbf{a}_i$ is the multiply the kth entry of $\mathbf{a}_i$ by m. Thus, the resulting matrix EA is the same as A, except each entry in the kth row is multiplied by m.

The proofs for the other two row operations are left as an exercise. □

Lemma 18. Each elementary matrix is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

Proof. Let E be an elementary matrix and F the elementary matrix that transforms E back into I. Clearly such an F exists because every row operation is reversible. Moreover, E must reverse F. By Lemma 17, $FE = I$ and $EF = I$, so F is the inverse of E. □

We are now ready to prove Theorem 14.

Proof. Suppose that A is invertible. Then the equation $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b} \in \mathbb{R}^n$. Hence, A has a pivot in each row. Because A is square, A has n pivots and so the reduced row echelon form of A is I_n. That is, $A \sim I_n$.

Conversely, suppose $A \sim I_n$. Then there exists a series of row operations which transform A into I_n. Denote the corresponding elementary matrices by $E_1, \ldots, E_p$. That is, $E_p(E_{p-1} \cdots E_1 A)) = I_n$. That is $E_p \cdots E_1 A = I_n$. Since the product of invertible matrices is invertible, then $A = (E_p \cdots E_1)^{-1}$. Since the inverse of an invertible matrix is invertible, it follows that A is invertible and the inverse is $E_p \cdots E_1$. □

Theorem 19 (The Invertible Matrix Theorem)**.** Let A be an $n \times n$ matrix. The following are equivalent.

(1) A is invertible.
(2) A is row equivalent to I_n.
(3) A has n pivot positions.
(4) The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(5) The columns of A form a linearly independent set.
(6) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is 1-1.
(7) The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b} \in \mathbb{R}^n$.
(8) The columns of A span $\mathbb{R}^n$.
(9) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto.
(10) There exists an $n \times n$ matrix C such that $CA = I_n$.
(11) There exists an $n \times n$ matrix D such that $AD = I_n$.
(12) A^T is invertible.

Proof. (1) $\Leftrightarrow$ (2) by Theorem 14. (2) $\Leftrightarrow$ (3) is clear because I_n has n pivots and row operations do not change the number of pivots. (2) $\Leftrightarrow$ (4) because row operations do not change the solution set and the only solution to $\mathbf{x} = I_n\mathbf{x} = \mathbf{0}$ is $\mathbf{0}$. (4) $\Leftrightarrow$ (6) by Theorem 35 in Chapter 1. (5) $\Leftrightarrow$ (6) and (8) $\Leftrightarrow$ (9) by Theorem 36 in Chapter 1. (3) $\Leftrightarrow$ (7) $\Leftrightarrow$ (8) by Theorem 17 in Chapter 1. Thus, (1) – (9) are all equivalent.

(1) $\Leftrightarrow$ (12) by Theorem 13.

Clearly, (1) implies (10) and (11). If (10) holds, then $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. If A has a left inverse C, then

$$\mathbf{x} = I_n\mathbf{x} = (CA)\mathbf{x} = C(A\mathbf{x}) = C\mathbf{0} = \mathbf{0}.$$

Thus, (10) $\Rightarrow$ (4). If (11) holds, then D^T is a left inverse of A^T. Hence, by the above this implies that A^T is invertible and so A is invertible. $\square$

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be invertible if there exists $S : \mathbb{R}^n \Rightarrow \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$. We call $S = T^{-1}$ the inverse of T.

Theorem 20. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix A. Then T is invertible if and only if A is an invertible matrix. In that case, the linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ given by $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique function such that $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. Suppose T is invertible with inverse S. We claim A is invertible. Let $\mathbf{b} \in \mathbb{R}^n$ and set $\mathbf{x} = S(\mathbf{b})$. Then $T(\mathbf{x}) = T(S(\mathbf{b})) = \mathbf{b}$, so T is onto and hence by the Invertible Matrix Theorem, A is invertible.

Conversely, suppose A is invertible. Define a function $S : \mathbb{R}^n \to \mathbb{R}^n$ by $S(\mathbf{x}) = A^{-1}\mathbf{x}$. We claim S is the inverse of T.

$$S(T(\mathbf{x})) = S(A\mathbf{x}) = A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}$$
$$T(S(\mathbf{x})) = T(A^{-1}\mathbf{x}) = A(A^{-1}\mathbf{x}) = (AA^{-1})\mathbf{x} = I_n\mathbf{x} = \mathbf{x}.$$

Hence, the claim holds.

The proof of the uniqueness of S is left as an exercise. $\square$

4. Partitioned Matrices

Definition 2. A partition of a matrix A is a decomposition of A into rectangular submatrices $A_1, \ldots, A_n$ such that each entry in A lies in some submatrix.

Example 21. The matrix

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

can be partitioned as

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{12} & A_{22} \end{array}\right]$$

where $A_{11} = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & 0 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$, $A_{21} = \begin{bmatrix} 4 & 5 & 2 \end{bmatrix}$, and $A_{22} = \begin{bmatrix} 1 \end{bmatrix}$.

Note that there are many other partitions of this matrix and in general a matrix will have several partitions. I leave it as an exercise to write another partition of this matrix.

If A and B are matrices of the same size and partition in the same way, then one may obtain $A+B$ by adding corresponding partitions. Similarly, we can multiply A and B as partitioned matrices so long as multiplication between the blocks makes sense.

Example 22. Let A be as in Example 21 and

$$B = \left[\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right] \text{ where } B_1 = \begin{bmatrix} 1 & 3 \\ 0 & 7 \\ 2 & 0 \end{bmatrix} \text{ and } B_2 = \begin{bmatrix} -1 & -3 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix}.$$

One way to partition a matrix is with only vertical or only horizontal lines. In this way we can denote

$$A = \begin{bmatrix} \text{col}_1(A) & \text{col}_2(A) & \cdots & \text{col}_n(B) \end{bmatrix} \text{ or } B = \begin{bmatrix} \text{row}_1(A) \\ \text{row}_2(A) \\ \cdots \\ \text{row}_n(B) \end{bmatrix}.$$

Theorem 23 (Column-row expansion of AB)**.** If A is $m \times n$ and B is $n \times p$ then

$$AB = \text{col}_1(A)\,\text{row}_1(A) + \cdots + \text{col}_n(A)\,\text{row}_n(B).$$

We say a matrix A is block diagonal if

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & A_n \end{bmatrix}.$$

where each A_i is a square matrix. If each A_i is invertible, then A is invertible and

$$A = \begin{bmatrix} A_1^{-1} & 0 & \cdots & 0 \\ 0 & A_2^{-1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & A_n^{-1} \end{bmatrix}.$$

Example 24. Let $A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. Find A^{-1} without using row reduction.

We say a matrix A is block upper triangular if it is of the form $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{21} \end{bmatrix}$. Suppose A_{11} is $p \times p$ and A_{22} is $q \times q$. If A is invertible, then we can compute the inverse B by setting

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}.$$

Multiplying on the left gives

$$\begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{22}B_{21} & A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}.$$

Since $A_{22}B_{21} = 0$ and A_{22} is invertible, then $B_{21} = 0$. Hence, the (1,1)-entry becomes $A_{11}B_{11} = I_p$, so $B_{11} = A_{11}^{-1}$. Moreover, since $A_{22}B_{22} = I_q$, then $B_{22} = A_{22}^{-1}$. Thus, the above reduces to

$$\begin{bmatrix} I_p & A_{11}B_{12} + A_{12}A_{22}^{-1} \\ 0 & I_q \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}.$$

Since $A_{11}B_{12} + A_{12}A_{22}^{-1} = 0$, then $B_{12} = -A_{11}^{-1}A_{12}A_{22}^{-1}$ so

$$B = A^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

6. The Leontief Input-Output Model

Suppose we divide the economy of the US into n *sectors* that produce or output goods and services. We can then denote the output of each sector as a production vector $\mathbf{x} \in \mathbb{R}^n$. Another part of the economy only consumes. We denote this by a final demand vector $\mathbf{d}$ that lists the values of the goods and services demanded by the nonproductive part of the economy. In between these two are the goods and services needed by the producers to create the products that are in demand. Optimally, one would balance all aspects of the production, but this is very complex. A model for finding this balance is due to Wassily Leontief. Ideally, one would find a production level $\mathbf{x}$ such that the amounts produced will balance the total demand, so that

$$\{ \text{ amt produced } \mathbf{x} \} = \{ \text{ int demand } \} + \{ \text{ final demand } \mathbf{d} \}.$$

The model assumes that for each sector, there is a unit consumption vector in $\mathbb{R}^n$ that lists the inputs needed per unit of output in the sector (in millions of dollars).

Example 25. Suppose the economy consists of three sectors: manufacturing, agriculture, and services, with unit consumption vectors c_1, c_2, and c_3, respectively.

	Inputs consumed per unit of output		
Purchased from:	Manufacturing	Agriculture	Services
Manufacturing	.50	.40	.20
Agriculture	.20	.30	.10
Services	.10	.10	.30
	$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$

We have $100\mathbf{c}_2 = \begin{bmatrix} 50 \\ 20 \\ 10 \end{bmatrix}$. Hence, to produce 100 units of output, manufacturing will order (or *demand*) 50 units from other parts of manufacturing, 20 units from agriculture, and 10 units from services.

If manufacturing produces $\mathbf{x}_1$ units of output, then $\mathbf{x}_1\mathbf{c}_1$ represents the *intermediate demands* of manufacturing, because the amounts in $\mathbf{x}_1\mathbf{c}_1$ will be consumed in the process of creating $\mathbf{x}_1$ units of output. Similarly, $\mathbf{x}_2\mathbf{c}_2$ and $\mathbf{x}_3\mathbf{c}_3$ list the corresponding intermediate demands of agriculture and services, respectively. Hence,

$$\{ \text{ int demand } \} = \mathbf{x}_1\mathbf{c}_1 + \mathbf{x}_2\mathbf{c}_2 + \mathbf{x}_3\mathbf{c}_3 = C\mathbf{x}$$

where C is the consumption matrix $\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \end{bmatrix}$. In this example,

$$C = \begin{bmatrix} .50 & .40 & .20 \\ .20 & .30 & .10 \\ .10 & .10 & .30 \end{bmatrix}$$

The Leontief Input-Output Model

$$\underset{\text{amt produced}}{\mathbf{x}} = \underset{\text{int demand}}{C\mathbf{x}} + \underset{\text{final demand}}{\mathbf{d}}$$

This can be rewritten as $(I - C)\mathbf{x} = \mathbf{d}$ where I is the $n \times n$ identity matrix.

Example 26. In the economy from the previous example, suppose the final demand is 50 units for manufacturing, 30 units for agriculture, and 20 units for services. Find the production level $\mathbf{x}$ that will satisfy this demand.

This amounts to row reducing the augmented matrix $[I - C \mid \mathbf{d}]$ where $\mathbf{d}$ is given in the problem. We can compute $I - C$ directly and we find that

$$C = \begin{bmatrix} .50 & -.40 & -.20 & 50 \\ -.20 & .70 & -.10 & 30 \\ -.10 & -.10 & .70 & 20 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 226 \\ 0 & 1 & 0 & 119 \\ 0 & 0 & 1 & 78 \end{bmatrix}.$$

How else might we solve this problem?

Theorem 27. Let C be the consumption matrix for an economy, and let $\mathbf{d}$ be the final demand. If C and d have nonnegative entries and if each column sum of C is less than 1, then $(I - C)^{-1}$ exists and the production vector

$$\mathbf{x} = (I - C)^{-1}\mathbf{d}$$

has nonnegative entries and is the unique solution of $\mathbf{x} = C\mathbf{x} + \mathbf{d}$.

In a 3×3 system, finding the inverse is not difficult, though a little tedious and we'd repeat the row reduction above. However, in the context of this theorem there is another way to get the inverse, or at least a close approximation of it.

Suppose the initial demand is set at $\mathbf{d}$. This creates an intermediate demand of $C\mathbf{d}$. To meet this intermediate demand, industries will need *more* input. This creates additional intermediate demands of $C(C\mathbf{d}) = C^2\mathbf{d}$. This process repeats (forever?) and in the next round we have $C(C^2\mathbf{d}) = C^3\mathbf{d}$. Hence, the production level that will meet this demand is

$$\mathbf{d} = \mathbf{d} + C\mathbf{d} + C^2\mathbf{d} + C^3\mathbf{d} + \cdots = (I + C + C^2 + C^3 + \cdots)\mathbf{d}.$$

By our hypotheses on C, $C^m \to 0$ as $m \to \infty$. Hence,

$$(I - C)^{-1} \approx I + C + C^2 + \cdots + C^m$$

where the approximation may be made as close to $(I - C)^{-1}$ as desired by taking m sufficiently large. Under what conditions will this ever be exact?

The i, j entry in $(I - C)^{-1}$ is the increased amounts that sector i will have to produce to satisfy an increase of 1 unit in the final demand from sector j.

8. Subspaces of $\mathbb{R}^n$

So far we have focused on linear transformations whose domain and codomain are exclusively $\mathbb{R}^n$. Now we explore *subspaces* and the concept of a *basis*. You should think of a basis as an efficient way of writing the span of a set of vectors.

Definition 3. A subspace of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ such that

(1) $\mathbf{0} \in H$;
(2) If $\mathbf{u}, \mathbf{v} \in H$, then $\mathbf{u} + \mathbf{v} \in H$;
(3) If $\mathbf{u} \in H$ and $c \in \mathbb{R}$, then $c\mathbf{u} \in H$.

$\{0\}$ and $\mathbb{R}^n$ are always subspaces of $\mathbb{R}^n$. Any other subspace is called proper.

Example 28. The set $H = \left\{ k \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} : k \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^3$. More generally, the span of *any* set of vectors in $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$.

Example 29. Let L be a line in $\mathbb{R}^n$ *not* through the origin. Show that L is not a subspace of $\mathbb{R}^n$.

Example 30. Define a set H in $\mathbb{R}^2$ by the following property: $\mathbf{v} \in H$ if $\mathbf{v}$ has exactly one nonzero entry. Is H a subspace of $\mathbb{R}^2$?

Next we'll define two subspaces related to linear transformations. Though they are defined in different ways, we'll see next time how they are connected.

Definition 4. Let A be an $m \times n$ matrix. The column space of a matrix A is the set $\text{Col}(A)$ of all linear combinations of the columns of A. The null space is the set $\text{Nul}(A)$ of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

If $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ with $\mathbf{a}_i \in \mathbb{R}^m$, then $\text{Col}(A) = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ so $\text{Col}(A)$ is a subspace of $\mathbb{R}^m$. We will show momentarily that $\text{Nul}(A)$ is a subspace of $\mathbb{R}^n$.

Example 31. Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Is $\mathbf{b} \in \text{Col}(A)$? What is $\text{Nul}(A)$?

Theorem 32. The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$.

Proof. By definition, $\text{Nul}(A)$ is a set in $\mathbb{R}^n$ and clearly $A\mathbf{0} = \mathbf{0}$, so $\mathbf{0} \in \text{Nul}(A)$.

Let $\mathbf{x}, \mathbf{y} \in \text{Nul}(A)$ so $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$. Then $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Thus, $\mathbf{x} + \mathbf{y} \in \text{Nul}(A)$.

Finally, suppose $\mathbf{x} \in \text{Nul}(A)$ and $c \in \mathbb{R}$. Then $A(c\mathbf{x}) = c(A\mathbf{x}) = c\mathbf{0} = \mathbf{0}$, so $c\mathbf{x} \in \text{Nul}(A)$. Thus, $\text{Nul}(A)$ is a subspace of $\mathbb{R}^n$. $\square$

Definition 5. A basis for a subspace of H of $\mathbb{R}^n$ is a linearly independent set in H that spans H.

If A is $n \times n$ and invertible then the columns of A are a basis for $\mathbb{R}^n$ because they are linearly independent and span $\mathbb{R}^n$.

Recall that $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are called the standard basis vectors for $\mathbb{R}^n$ (now you know why).

Proposition 33. Any basis of $\mathbb{R}^n$ has size n.

Proof. Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$ be a basis for $\mathbb{R}^n$.

Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_p \end{bmatrix}$. Since the columns of A span $\mathbb{R}^n$, then there must be a pivot in every row. Since the columns of A are linearly independent, then there must be a pivot in every column. This implies that A has n rows and columns, so $p = n$. $\square$

Example 34. Find a basis for $\mathrm{Nul}(A)$ in Example 31.

Example 35. Find a basis for the null space of

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & 4 \end{bmatrix}$$

Example 36. Find a basis for the column space of

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

Theorem 37. The pivot columns of a matrix A form a basis for the column space of A.

Proof. Row reduction does not change the linear relations between the columns. Each pivot column corresponds to a standard basis vector in $\mathrm{RREF}(A)$ and hence the pivot columns are linearly independent.

Let $\mathbf{b}$ be a non-pivot column of A. Since $\mathbf{b}$ does not have a pivot, then all of its nonzero entries lie to the right of the pivots. Consequently, it is a linear combination of the pivot columns to its left. Thus, the pivot columns span $\mathrm{Col}(A)$. $\square$

Definition 6. The dimension of a nonzero subspace H of $\mathbb{R}^n$, denoted $\dim H$, is the number of vectors in any basis of H. The dimension of the zero subspace is defined to be zero.

The rank of a matrix A is the dimension of the column space of A ($\operatorname{rank} A = \dim \operatorname{Col} A$). The nullity of a matrix A is the dimension of the null space of A ($\operatorname{nul} A = \dim \operatorname{Nul} A$).

At the end of this section we will prove that dimension is well-defined (that is, all bases have the same dimension). We have already proved this for $\mathbb{R}^n$.

Example 38. Find the rank and nullity of $A = \begin{bmatrix} 1 & 3 & 2 & -6 \\ 3 & 9 & 1 & 5 \\ 2 & 6 & -1 & 9 \\ 5 & 15 & 0 & 14 \end{bmatrix}$.

A row reduces to the following matrix.

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are three pivot colums, so $\operatorname{rank} A = 3$. On the other hand, there is one free variable (x_2), and so $\operatorname{nul} A = 1$.

Theorem 39 (Rank-Nullity Theorem)**.** If a matrix A has n columns, then $\operatorname{rank} A + \operatorname{nul} A = n$.

Proof. We know that $\operatorname{rank} A$ is the number of pivot columns in A and $\operatorname{nul} A$ is the number of columns corresponding to free variables. Since each columns is exactly one of these, the result follows. $\square$

Theorem 40 (The Invertible Matrix Theorem)**.** Let A be an $n \times n$ matrix. The following are equivalent to the statement that A is invertible.

(13) The columns of A form a basis of $\mathbb{R}^n$.
(14) $\operatorname{Col} A = \mathbb{R}^n$.
(15) $\dim \operatorname{Col} A = n$.
(16) $\operatorname{rank} A = n$.
(17) $\operatorname{Nul} A = \{0\}$.
(18) $\operatorname{nul} A = 0$.

Proposition 41. Let H be a subspace of $\mathbb{R}^n$ with basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$. If $\mathbf{x} \in H$, then $\mathbf{x}$ can be written in exactly one way as a linear combination of basis vectors in $\mathcal{B}$.

Proof. Suppose $\mathbf{x}$ can be written in two ways: $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p = d_1\mathbf{b}_1 + \cdots + d_p\mathbf{b}_p$. Taking differences we find $\mathbf{0} = (c_1 - d_1)\mathbf{b}_1 + \cdots + (c_p - d_p)\mathbf{b}_p$. By linear independence, $c_i - d_i = 0$ for all $i = 1, \dots, p$. That is, $c_i = d_i$ for $i = 1, \dots, p$. $\square$

We now want to set up the theorem that says the notion of dimension is well-defined.

Definition 7. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ is a basis for a subspace H of $\mathbb{R}^n$. For each $\mathbf{x} \in H$, the coordinates of $\mathbf{x}$ relative to B are the weights $c_1, \ldots, c_p$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p$. The coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$ (or the $\mathcal{B}$-coordinate vector) is

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

The map $T_{\mathcal{B}} : H \to \mathbb{R}^p$ given by $\mathbf{x} \mapsto [\mathbf{x}]_{\mathcal{B}}$ is called a coordinate mapping of $\mathcal{B}$.

Example 42. Let $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$. Find $[\mathbf{x}]_{\mathcal{B}}$.

Proposition 43. Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ be a basis for a subspace H of $\mathbb{R}^n$. The coordinate mapping $T_{\mathcal{B}}$ is a 1-1 and onto (bijective) linear transformation.

Proof. By Proposition 41, is $T_{\mathcal{B}}$ is well-defined. Linearity is left as an exercise.

Let $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p \in H$ such that $T_{\mathcal{B}}(\mathbf{x}) = \mathbf{0}$. Then $[\mathbf{x}]_{\mathcal{B}} = (c_i) = \mathbf{0}$ so $c_i = 0$ for all i. Hence, $\mathbf{x} = \mathbf{0}$ and T is 1-1. To prove onto, assume $\mathbf{a} = (a_i) \in \mathbb{R}^p$. Then $\mathbf{y} = a_1\mathbf{b}_1 + \cdots + a_p\mathbf{b}_p \in H$ because $\mathcal{B}$ spans H and $T_{\mathcal{B}} = [\mathbf{y}]_{\mathcal{B}} = \mathbf{a}$. □

Theorem 44. Let H be a subspace of $\mathbb{R}^n$ with basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$. Then every basis of H has p elements.

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set in H with $m > p$. We claim the set is linearly dependent (and hence cannot be a basis). Note that the set $\{[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_m]_{\mathcal{B}}\}$ is linearly dependent in $\mathbb{R}^p$ because $m > p$. Hence, there exist scalars $c_1, \ldots, c_m$ (not all zero) such that

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_m[\mathbf{u}_m]_{\mathcal{B}} = \mathbf{0}.$$

Because the coordinate mapping is a linear transformation, this implies

$$[c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m]_{\mathcal{B}} = \mathbf{0}.$$

Moreover, because the coordinate mapping is 1-1 this implies $c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \mathbf{0}$, so the set is linearly dependent. Hence, any basis of H can have at most p elements.

Now suppose $\mathcal{B}'$ is another basis with $k \neq p$ elements. By the above argument, $\mathcal{B}$ can have no more elements than $\mathcal{B}'$, so $p \leq k$. This implies $p = k$ so every basis has p elements. □

Theorem 45 (The Basis Theorem)**.** Let H be a p-dimensional subspace of $\mathbb{R}^n$. Any linearly independent set of exactly p elements in H is automatically a basis for H. Also, any set of p elements of H that spans H is automatically a basis for H.

Proof. Since H has dimension p, then there exists a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$.

Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be a set of p linearly independent elements in H. If $\mathcal{U}$ spans H, then $\mathcal{U}$ is a basis. Otherwise, there exists some element $\mathbf{u}_{p+1} \notin \mathrm{Span}\,\mathcal{U}$. Hence, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{u}_{p+1}\}$ is linearly independent. We can continue this process, but not indefinitely. The number of linearly independent elements in the set cannot exceed n (the dimension of $\mathbb{R}^n$). Thus, our set will eventually span all of H, and hence will be a basis. But a basis for H cannot have more than p elements by Theorem 44. Hence, $\mathcal{U}$ must have been a basis to begin with.

Now suppose $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a set of p elements in H that span H. If the set is not linearly independent, then one of the vectors is a linear combination of the others. Hence, we can remove that element, say $\mathbf{v}_1$ and the span of $\{\mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is the same as the span of $\mathcal{V}$. Continue in this way until the remaining elements are linearly independent. But then the set is linearly independent and spans H, hence is a basis. But by Theorem 45, every basis has exactly p elements, a contradiction unless $\mathcal{V}$ was a basis to begin with. $\square$

Another way of viewing the previous theorem is the following. The first says that any linearly independent set in H can be extended to a basis. The second says that any spanning set contains a basis.

Change of basis (in $\mathbb{R}^n$)

Let $\mathcal{B}$ be a basis for $\mathbb{R}^n$ and denote by $\mathcal{E}$ the standard basis of $\mathbb{R}^n$. The coordinate mapping $T_{\mathcal{B}}$ is as a map from the standard basis to $\mathcal{B}$.

Example 46. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. Compute the standard matrix of $T_{\mathcal{B}}$. Use this to find the $\mathcal{B}$-coordinates of $\mathbf{x}$.

We need to compute $T_{\mathcal{B}}(\mathbf{e}_1)$ and $T_{\mathcal{B}}(\mathbf{e}_2)$. We know how to do this problem in general, Here we'll just observe that $\mathbf{e}_1 = -4\mathbf{b}_1 + 5\mathbf{b}_2$ and $\mathbf{e}_2 = \mathbf{e}_1 - \mathbf{e}_2$. Thus,

$$T_{\mathcal{B}}(\mathbf{e}_1) = [\mathbf{e}_1]_{\mathcal{B}} = \begin{bmatrix} -4 \\ 5 \end{bmatrix} \quad \text{and} \quad T_{\mathcal{B}}(\mathbf{e}_2) = [\mathbf{e}_2]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus, the standard matrix of $T_{\mathcal{B}}$, denoted $P_{\mathcal{B}}$, is

$$P_{\mathcal{B}} = \begin{bmatrix} T_{\mathcal{B}}(\mathbf{e}_1) & T_{\mathcal{B}}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}, \quad \text{so} \quad T_{\mathcal{B}}(\mathbf{x}) = P_{\mathcal{B}}\mathbf{x} = \begin{bmatrix} -13 \\ 16 \end{bmatrix}.$$

We can check this, $\mathbf{x} = -13\mathbf{b}_1 + 17\mathbf{b}_2 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$.

In the last section we proved that $T_{\mathcal{B}}$ is a 1-1 and onto (bijective) linear map and hence it is invertible with inverse $T_{\mathcal{B}}^{-1}$. The standard matrix of $T_{\mathcal{B}}^{-1}$ is $P_{\mathcal{B}}^{-1}$. It's easy to see that $T_{\mathcal{B}}(\mathbf{b}_i) = \mathbf{e}_i$ (see for yourself in the example above then verify this fact in general). Hence, $T_{\mathcal{B}}^{-1}(\mathbf{e}_i) = \mathbf{b}_i$. Thus,

$$P_{\mathcal{B}}^{-1} = \begin{bmatrix} T_{\mathcal{B}}^{-1}(\mathbf{e}_1) & \cdots & T_{\mathcal{B}}(\mathbf{e}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 & \cdots \mathbf{b}_n \end{bmatrix}.$$

This is just the matrix whose columns are the vectors in the basis $\mathcal{B}$!

Example 47. Use the above trick to compute $P_{\mathcal{B}}$ in Example 46.

Since $P_{\mathcal{B}}^{-1} = \begin{bmatrix} 1 & 1 \\ 5 & 4 \end{bmatrix}$. Then $P_{\mathcal{B}} = (P_{\mathcal{B}}^{-1})^{-1} = \begin{bmatrix} -4 & 1 \\ 5 & -1 \end{bmatrix}$.

Now say we have two bases of $\mathbb{R}^n$, $\mathcal{B}$ and $\mathcal{C}$. We have the tools we need to change from the basis $\mathcal{B}$ to the basis $\mathcal{C}$. Consider the following diagram.

$$\begin{tikzcd}
\mathbb{R}^n_{\mathcal{B}} & & \mathbb{R}^n_{\mathcal{C}} \\
& \mathbb{R}^n_{\mathcal{E}} \arrow[ul, "T_{\mathcal{B}}"] \arrow[ur, "T_{\mathcal{C}}"'] &
\end{tikzcd}$$

If we want to go from $\mathcal{B}$ to $\mathcal{C}$, that is, convert coordinates in $\mathcal{B}$ to coordinates in $\mathcal{C}$, we need to first go from $\mathcal{B}$ to $\mathcal{E}$ and then from $\mathcal{E}$ to $\mathcal{C}$.

$$\begin{tikzcd}
\mathbb{R}^n_{\mathcal{B}} \arrow[rr, dashed, "T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}"] \arrow[dr, dashed, bend right, "T_{\mathcal{B}}^{-1}"'] & & \mathbb{R}^n_{\mathcal{C}} \\
& \mathbb{R}^n_{\mathcal{E}} \arrow[ul, "T_{\mathcal{B}}"'] \arrow[ur, "T_{\mathcal{C}}"] &
\end{tikzcd}$$

Definition 8. Suppose we have bases $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ of $\mathbb{R}^n$. Then the change of basis transformation is

$$T_{\mathcal{B}\to\mathcal{C}} : \mathbb{R}^n \to \mathbb{R}^n$$
$$[\mathbf{x}]_{\mathcal{B}} \mapsto [\mathbf{x}]_{\mathcal{C}}$$

is linear and is the composition $T_{\mathcal{C}} \circ T_{\mathcal{B}}^{-1}$ where $T_{\mathcal{B}}$ and $T_{\mathcal{C}}$ are the coordinate mappings of $\mathcal{B}$ and $\mathcal{C}$, respectively. Its standard matrix is $P_{\mathcal{B}\to\mathcal{C}} = P_{\mathcal{C}} P_{\mathcal{B}}^{-1}$ and is called the change of basis matrix.

Example 48. Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}$ be as before and $\mathcal{C} = \left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 5 \end{bmatrix} \right\}$. Find the change of basis matrix $P_{\mathcal{B}\to\mathcal{C}}$. Given a vector $\mathbf{x} \in \mathbb{R}^n$ with $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$, use $P_{\mathcal{B}\to\mathcal{C}}$ to compute $[\mathbf{x}]_{\mathcal{C}}$.

We previously computed $P_{\mathcal{B}}^{-1}$ so it is left to compute $P_{\mathcal{C}}$. We know that

$$P_{\mathcal{C}}^{-1} = \begin{bmatrix} -1 & -2 \\ 3 & 5 \end{bmatrix} \quad \text{so} \quad P_{\mathcal{C}} = (P_{\mathcal{C}}^{-1})^{-1} = \begin{bmatrix} 5 & 2 \\ -3 & -1 \end{bmatrix}.$$

Hence,

$$P_{\mathcal{B}\to\mathcal{C}} = P_{\mathcal{C}} P_{\mathcal{B}}^{-1} = \begin{bmatrix} 5 & 2 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} 15 & 13 \\ -8 & -7 \end{bmatrix}.$$

To find the coordinates $[\mathbf{x}]_{\mathcal{C}}$, we apply $P_{\mathcal{B}\to\mathcal{C}}$ to $[\mathbf{x}]_{\mathcal{B}}$,

$$P_{\mathcal{B}\to\mathcal{C}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 15 & 13 \\ -8 & -7 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -11 \\ 6 \end{bmatrix}.$$

We can check by converting both to the standard basis,

$$\mathbf{x} = (1)\mathbf{b}_1 + (-2)\mathbf{b}_2 = \begin{bmatrix} -1 \\ -3 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = (-11)\mathbf{c}_1 + (6)\mathbf{c}_2 = \begin{bmatrix} -1 \\ -3 \end{bmatrix}.$$

Example 49. Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix} \right\} \quad \text{and} \quad \mathcal{C} = \left\{ \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ 3 \\ 2 \end{bmatrix} \right\}.$$

Both $\mathcal{B}$ and $\mathcal{C}$ are bases for $\mathbb{R}^3$ (check this!). Let $T_\mathcal{B}$ and $T_\mathcal{C}$ be their respective coordinate mappings with standard matrices $P_\mathcal{B}$ and $P_\mathcal{C}$, respectively. Then

$$P_B^{-1} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix}, P_B = \begin{bmatrix} 15 & 2 & -6 \\ 5 & 1 & -2 \\ -7 & -1 & -3 \end{bmatrix}, \quad P_\mathcal{C}^{-1} = \begin{bmatrix} 5 & 0 & 7 \\ 2 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}, P_C = \begin{bmatrix} 3 & -7 & 0 \\ 1 & -3 & 1 \\ -2 & 5 & 0 \end{bmatrix}.$$

Thus,

$$P_{\mathcal{B}\to\mathcal{C}} = P_\mathcal{C} P_\mathcal{B}^{-1} = \begin{bmatrix} 10 & -21 & 6 \\ 6 & -8 & 7 \\ -7 & 15 & -4 \end{bmatrix}$$

Write $\mathbf{x} \in \mathbb{R}^3$ with $[\mathbf{x}]_\mathcal{B} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$. Applying our map above gives $[\mathbf{x}]_\mathcal{C} = P_{\mathcal{B}\to\mathcal{C}}[\mathbf{x}]_\mathcal{B} = \begin{bmatrix} -18 \\ -5 \\ 13 \end{bmatrix}$. We check by writing both in terms of the standard basis,

$$\mathbf{x} = 3\mathbf{b}_1 + 2\mathbf{b}_2 - \mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = -18\mathbf{c}_1 - 5\mathbf{c}_2 + 13\mathbf{c}_3.$$

The above generalizes somewhat. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be any linear transformation and $\mathcal{B}, \mathcal{C}$ bases of $\mathbb{R}^n$. The map

$$\begin{aligned} [T]_{\mathcal{B},\mathcal{C}} : \mathbb{R}^n &\to \mathbb{R}^n \\ [\mathbf{x}]_\mathcal{B} &\mapsto [T(\mathbf{x})]_\mathcal{C} \end{aligned}$$

is linear. In fact, it is just the composition $T_\mathcal{C} \circ T \circ T_\mathcal{B}^{-1}$. If A_T is the standard matrix of T, then the standard matrix of $[T]_{\mathcal{B},\mathcal{C}}$ is $P_\mathcal{C} \circ A_T \circ P_\mathcal{B}^{-1}$. Hence, we have the commutative diagrams.

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "T_\mathcal{B}"] \arrow[d, "T"'] & \mathbb{R}^n \arrow[d, "T_{\mathcal{B},\mathcal{C}}"] \\ \mathbb{R}^n \arrow[r, "T_\mathcal{C}"'] & \mathbb{R}^n \end{tikzcd} \qquad \begin{tikzcd} \mathbb{R}^n \arrow[r, "P_\mathcal{B}"] \arrow[d, "A_T"'] & \mathbb{R}^n \arrow[d, "P_{\mathcal{B},\mathcal{C}}"] \\ \mathbb{R}^n \arrow[r, "P_\mathcal{C}"'] & \mathbb{R}^n \end{tikzcd}$$

When $T = \mathrm{id}_{\mathbb{R}^n}$, then this reduces to the above case.

Chapter 3: Determinants

(Last Updated: October 16, 2017)

These notes are derived primarily from *Linear Algebra and its applications* by David Lay (4ed).

1. Introduction to Determinants

The determinant of a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad - bc$. The matrix is invertible if and only if this value is nonzero. In this chapter we'll develop a similar criteria for $n \times n$ matrices.

For an $n \times n$ matrix $A = (a_{ij})$, we denote by A_{ij} the submatrix obtained by deleting the ith row and jth column.

Example 1. Let $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$. Then $A_{23} = \begin{bmatrix} 2 & -4 \\ 1 & 4 \end{bmatrix}$ and $A_{33} = \begin{bmatrix} 2 & -4 \\ 3 & 1 \end{bmatrix}$.

The definition of determinant is *recursive*. That means, in order to compute the determinant of an $n \times n$ matrix we first need to know how to compute the determinant of an $(n-1) \times (n-1)$ matrix. This is ok because we already know how to compute the determinant of a 2×2 matrix.

We won't fully derive the formula for an $n \times n$ matrix. However, I will try to give you an idea of where the formula comes from in the 2×2 and 3×3 cases.

2×2 **case:** Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be invertible, so $A \sim I_2$. Suppose $a \neq 0$[1]. Row reduction gives

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \sim \begin{bmatrix} 1 & b/a \\ c & d \end{bmatrix} \sim \begin{bmatrix} 1 & b/a \\ 0 & d - cb/a \end{bmatrix}.$$

Thus, $A \sim I_n$ if and only if $d - cb/a \neq 0 \Leftrightarrow ad - bc \neq 0$.

3×3 **case:** Let $A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$ We'll assume that $a_{11} \neq 0$ along with some other nonzero assumptions. Again, A is invertible if and only if $A \sim I_3$. Row reduction gives

$$A \sim \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} - \frac{a_{2,1}a_{1,2}}{a_{1,1}} & a_{2,3} - \frac{a_{1,3}a_{2,1}}{a_{1,1}} \\ 0 & a_{3,2} - \frac{a_{3,1}a_{1,2}}{a_{1,1}} & a_{3,3} - \frac{a_{3,1}a_{1,3}}{a_{1,1}} \end{bmatrix} \sim \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2}a_{1,1} - a_{2,1}a_{1,2} & a_{1,1}a_{2,3} - a_{1,3}a_{2,1} \\ 0 & a_{1,1}a_{3,2} - a_{1,2}a_{3,1} & a_{1,1}a_{3,3} - a_{1,3}a_{3,1} \end{bmatrix}$$

$$\sim \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2}a_{1,1} - a_{2,1}a_{1,2} & a_{1,1}a_{2,3} - a_{1,3}a_{2,1} \\ 0 & 0 & a_{2,2}a_{1,1}a_{3,3} - a_{1,1}a_{2,3}a_{3,2} - a_{2,1}a_{1,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} - a_{1,3}a_{2,2}a_{3,1} \end{bmatrix}$$

[1] If $a = 0$, the just switch row 1 and row 2. If $a = 0$ and $c = 0$, then A is not invertible.

We can rewrite this a bit to get

$$\begin{aligned}\det(A) &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{12} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})\\ &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}).\end{aligned}$$

Definition 1. For $n \geq 2$, the determinant of an $n \times n$ matrix $A = (a_{ij})$ is the sum of n terms of the form $\pm a_{1j}\det(A_{ij})$ with alternating $\pm$ signs:

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

Example 2. Compute the determinant of the matrix in Example 1.

The definition of the determinant we gave is one of many (equivalent) definitions. In particular, the choice of using the first row is completely arbitrary.

The (i, j)-cofactor matrix is $C_{ij} = (-1)^{i+1}\det(A_{ij})$. Our earlier formula is then

$$\det(A) = \sum_{j=1}^{n} a_{1j}C_{1j}.$$

Theorem 3 (Cofactor Expansion)**.** The determinant of any $n \times n$ matrix can be determined by cofactor expansion along any row or column. In particular,

$$\begin{aligned}&(i\text{th row}) && \det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}C_{ij}\\ &(j\text{th col}) && \det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \sum_{i=1}^{n} a_{ij}C_{ij}.\end{aligned}$$

Example 4. Compute the determinant of the matrix A in Example 1 using cofactor expansion along the first column.

Example 5. Computer the determinant of the matrix A below.

$$A = \begin{bmatrix} 6 & 0 & 0 & 5 \\ 1 & 7 & 2 & -5 \\ 2 & 0 & 0 & 0 \\ 8 & 3 & 1 & 8 \end{bmatrix}$$

Recall that a matrix is upper-triangular (resp. lower-triangular) if all entries below (resp. above) the main diagonal are zero. Both of the following facts can be proven using the principle of mathematical induction.

(1) If A is triangular, then $\det(A)$ is the product of the entries on the main diagonal.
(2) If A is a square matrix, then $\det(A) = \det(A^T)$.

The (first) principle of mathematical induction says the following. Let $S(n)$ be a statement about the integers for $n \in \mathbb{N}$ and suppose $S(n_0)$ is true for some integer n_0. If for all integers k with $k \geq n_0$, $S(k)$ implies that $S(k+1)$ is true, then $S(n)$ is true for all integers greater than or equal to n_0.

Example 6. For all $n \in \mathbb{N}$, $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.

We prove this first for the smallest possible value in which the claim is true (n_0). This is typically referred to as the *base case*. Often this is done simply by checking. Note that if $n_0 = 1$, then we have $\frac{1(1+1)}{2} = 1$, so the claim is true.

Now assume the claim is true for some $k \geq n_0 = 1$. This is known as the *inductive hypothesis*. More precisely, our hypothesis says that for this k, we have $1+2+\cdots+k = \frac{k(k+1)}{2}$. We must show that it is then true for $k+1$. That is, we must show that $1 + 2 + \cdots + (k+1) = \frac{(k+1)(k+2)}{2}$. But we need only observe that

$$\begin{aligned} 1 + 2 + \cdots + (k+1) &= (1 + 2 + \cdots + k) + (k+1) \\ &= \frac{k(k+1)}{2} + (k+1) \qquad \text{(by the inductive hypothesis)} \\ &= \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \frac{k^2+3k+2}{2} = \frac{(k+1)(k+2)}{2}. \end{aligned}$$

Thus, the claim holds for all $n \geq n_0 = 1$ by the principle of mathematical induction.

What have we really done? Think of this like a staircase. We proved that the first step is there, then we proved that we can step up. Because we did this *stepping* in an arbitrary way, this means we can then step up again, and again, and so on.

Theorem 7. If A is triangular, then $\det(A)$ is the product of the entries on the main diagonal.

Proof. We will prove this for A upper-triangular. A lower-triangular is similar.

Let A be an upper-triangular 2×2 matrix. Then

$$\det(A) = \begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad - 0 = ad.$$

Hence, the claim holds for all 2×2 matrices.

Suppose the claim holds for all upper-triangular $k \times k$ matrices, with k some integer, $k \geq 2$. Let A be a $(k+1) \times (k+1)$ upper-triangular matrix. By cofactor expansion along the first column, we find, $\det(A) = a_{11} \det(A_{11})$. The submatrix A_{11} is an upper-triangular $k \times k$ matrix and hence by the inductive hypothesis, $\det(A_{11})$ is a product of its diagonal entries. That is, $\det(A_{11}) = a_{22}a_{33} \cdots a_{(k+1)(k+1)}$ so

$$\det(A) = a_{11} \det(A_{11}) = a_{11}a_{22}a_{33} \cdots a_{(k+1)(k+1)}.$$

Thus, the claim follows by the principle of mathematical induction. $\square$

Theorem 8. If A is a square matrix, then $\det(A) = \det(A^T)$.

Proof. Let A be a 2×2 matrix. Then

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

$$\det(A^T) = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - cb = ad - bc.$$

Hence, the claim holds for all 2×2 matrices.

Suppose the claim holds for all $k \times k$ matrices, with k some integer, $k \geq 2$. Let A be a $(k+1) \times (k+1)$ matrix. We use cofactor expansion along the first column of A^T.

$$\det(A^T) = a_{11} \det(A_{11}^T) + a_{12} \det(A_{12}^T) + \cdots + a_{1(k+1)} \det(A_{1(k+1)}^T).$$

Each submatrix A_{1j}^T is a $k \times k$ matrix and hence by the inductive hypothesis, $\det(A_{1j}^T) = \det(A_{1j})$. Thus,

$$\begin{aligned} \det(A^T) &= a_{11} \det(A_{11}^T) + a_{12} \det(A_{12}^T) + \cdots + a_{1(k+1)} \det(A_{1(k+1)}^T) \\ &= a_{11} \det(A_{11}) + a_{12} \det(A_{12}) + \cdots + a_{1(k+1)} \det(A_{1(k+1)}) = \det(A). \end{aligned}$$

Therefore, the claim follows by the principle of mathematical induction. $\square$

2. Properties of determinants

Today we'll see how the elementary row operations affect the determinant. By keeping track of these rules, we are able to use row reduction to compute the determinant. One consequence of these results is the general theorem that $\det A \neq 0$ is equivalent to A invertible. This also leads to a powerful theorem on the multiplicativity of the determinant.

Theorem 9. Let A be a square matrix.

(1) If a multiple of one row of A is added to another row to produce a matrix B, then $\det A = \det B$.
(2) If two rows of A are interchanged to produce B, then $\det A = -\det B$.
(3) If one row of A is multiplied by k to produce B, then $\det B = k \cdot \det A$.

These rules can be proved using elementary matrices which we discussed back in Chapter 1.

Example 10. Use row reduction to compute the determinant of $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$.

Before attempting a proof of the above theorem, let's look at some of the consequences.

Theorem 11. A square matrix A is invertible if and only if $\det A \neq 0$.

Proof. Let U be the row echelon form of A obtained by row replacement and row interchange. That is, we have not used scaling. Suppose r row interchanges were used to produce U. By the Invertible Matrix Theorem, U is invertible if and only if A is invertible.

Since U is upper triangular, then $\det U$ is just the product of the pivots in A. Hence,

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

Thus, $\det A \neq 0$ if and only if $\det U \neq 0$ and $\det U \neq 0$ if and only if U is invertible if and only if A is invertible. □

Our next goal is to prove that for $n \times n$ matrices A and B, we have the rule $\det(AB) = \det(A)\det(B)$. The next lemma is the first step in that. Recall that an elementary matrix is obtained from the identity matrix by performing a single row operation.

If E is an elementary matrix, then

$$\det(E) = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r. \end{cases} \tag{1}$$

The verification of this is left as an exercise.

Lemma 12. If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix. Then $\det(EA) = \det(E)\det(A)$.

Proof. The proof is by induction on the size of A.

Suppose A is 2×2, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. We check each case of E directly.

If E is scaling row 1 by r, then

$$\det(EA) = \begin{vmatrix} ra & rb \\ c & d \end{vmatrix} = rad - rbc = r(ad - bc) = r\det(A) = \det(E)\det(A) \quad \text{by (1).}$$

If E is scaling row 2 by r, the proof is similar.

If E is switching row 1 and row 2, then

$$\det(EA) = \begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - ad = -(ad - bc) = -\det(A) = \det(E)\det(A) \quad \text{by (1).}$$

If E is adding r times row 1 to row 2, then

$$\begin{aligned} \det(EA) &= \begin{vmatrix} a & b \\ c + ra & d + rb \end{vmatrix} = a(d + rb) - b(c + ra) = (ad - bc) + r(ab - ba) \\ &= ad - bc = \det(A) = \det(E)\det(A) \quad \text{by (1).} \end{aligned}$$

Hence, the claim holds for 2×2 matrices.

Suppose the claim holds for $k \times k$ matrices, $k \geq 2$. That is, if A is $k \times k$ then $\det(EA) = \alpha \det(A)$ where $\alpha = 1, -1, r$ depending on the case of E.

We must show that the claim holds for $(k+1) \times (k+1)$ matrices. Since $k + 1 \geq 3$, there is some row, say row i of A that is unchanged by E. Let $B = EA$. We will use cofactor expansion along row i of B. Note that the B_{ij} are $k \times k$ matrices obtained from A by applying E. Thus, by the inductive hypothesis, $\det(B_{ij}) = \alpha \det(A_{ij}$ where $\alpha = 1, -1, r$ depending on the case of E. Thus,

$$\begin{aligned} \det(EA) &= \det(B) \\ &= a_{i1}(-1)^{i+1}\det(B_{i1}) + a_{i2}(-1)^{i+2}\det(B_{i2}) + \cdots + a_{i(k+1)}(-1)^{i+(k+1)}\det(B_{i(k+1)}). \\ &= \alpha a_{i1}(-1)^{i+1}\det(A_{i1}) + \alpha a_{i2}(-1)^{i+2}\det(A_{i2}) + \cdots + \alpha a_{i(k+1)}(-1)^{i+(k+1)}\det(A_{i(k+1)}) \\ &= \alpha\det(A) = \det(E)\det(A). \end{aligned}$$

Thus, the claim holds for $(k+1) \times (k+1)$ matrices and so the theorem is true by the principle of mathematical induction. □

Theorem 9 is now an immediate consequence of the preceding lemma.

Theorem 13. If A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.

Proof. Suppose A is not invertible, then neither is AB (this was a homework exercise). Hence, $\det(AB) = \det(A)\det(B)$ because both sides are zero by Theorem 11. A similar argument holds if B is not invertible.

Suppose A is invertible. Then $A \sim I_n$ by a series of elementary operations. That is, $A = E_p E_{p-1} \cdots E_1 I_n = E_p E_{p-1} \cdots E_1$. Thus, by Lemma 12,

$$\begin{aligned} |AB| &= |E_p E_{p-1} \cdots E_1 B| \\ &= |E_p E_{p-1} \cdots E_1 B| \\ &= |E_p||E_{p-1}| \cdots |E_1||B| \\ &= |E_p E_{p-1} \cdots E_1||B| = |A||B|. \end{aligned}$$

□

Corollary 14. If A is an invertible $n \times n$ matrix, then $\det(A^{-1}) = \frac{1}{\det(A)}$.

3. Cramer's Rule

For a $n \times n$ matrix A and any $\mathbf{b} \in \mathbb{R}^n$, denote by $A_i(\mathbf{b})$ the matrix obtained by replacing column i in A with $\mathbf{b}$.

Theorem 15 (Cramer's Rule)**.** Let A be an invertible $n \times n$ matrix. For any $\mathbf{b} \in \mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by

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

Proof. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$. If $A\mathbf{x} = \mathbf{b}$, then

$$\begin{aligned} A \cdot I_i(\mathbf{x}) &= A \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} = A_i(\mathbf{b}). \end{aligned}$$

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

$$\det(A) \det I_i(\mathbf{x}) = \det A_i(\mathbf{b}). \tag{2}$$

But $\det(I_i(\mathbf{x})) = x_i$, so (2) redues to $\det(A)x_i = \det A_i(\mathbf{b})$. The result now follows. $\square$

Example 16. Use Cramer's Rule to solve the following system.

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7. \end{aligned}$$

Let A be an $n \times n$ invertible matrix. Cramer's Rule also leads to a formula for the inverse of A. Denote by $\mathbf{x}$ the jth column of A^{-1}. This satisfies $A\mathbf{x} = \mathbf{e}_j$. By Cramer's Rule, $x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A}$. This is the (i, j)-entry of A^{-1}. Cofactor expansion down column i of $A_i(e_j)$ shows that

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

Define the adjugate matrix of A, denoted $\operatorname{adj} A$, as an $n \times n$ matrix with i, j entry C_{ji}.

Theorem 17 (Cramer's Inverse Formula)**.** Let A be an invertible $n \times n$ matrix. Then

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

Example 18. Use Cramer's Inverse Formula to find the inverse of $A = \begin{bmatrix} 2 & -4 & 3 \\ 3 & 1 & 2 \\ 1 & 4 & -1 \end{bmatrix}$.

The determinant has a geometric interpretation as well.

Theorem 19. If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix,, then $|\det A|$ is the volume of the parallelepiped determined by the columns of A.

The key point to the above theorem is that if $\mathbf{a}_1, \mathbf{a}_2 \neq \mathbf{0}$, then for any scalar c the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ is equal to that of $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$.

Example 20. Calculate the area of the parallelogram determined by the points

$$(-2,-2), (0,3), (4,-1), (6,4).$$

We first translate the parallelogram by adding $(2,2)$ to each vertex. Then the points are

$$(0,0), (2,5), (6,1), (8,6).$$

This is the parallelogram determined by the vectors through the points $(2,5)$ and $(6,1)$. Thus,

$$\text{Area } = \begin{vmatrix} 2 & 6 \\ 5 & 1 \end{vmatrix} = |-28| = 28.$$

Theorem 21. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation with standard matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

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

If $T : \mathbb{R}^3 \to \mathbb{R}^3$ and S is a parallelepiped in $\mathbb{R}^3$, then

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

Chapter 5: Eigenvectors and Eigenvalues

(Last Updated: November 29, 2021)

These notes are derived primarily from *Linear Algebra and its applications* by David Lay (4ed). A few theorems have been moved around.

1. Introduction to Eigenstuff

Eigenvectors and eigenvalues are important tools in both pure and applied mathematics. One place we'll see them applied is in solutions to differential equations.

Example 1. Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Observe that $A\mathbf{u} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = 4\mathbf{u}$. Geometrically, we can interpret this as A stretching the vector $\mathbf{u}$. This does not happen with $\mathbf{v}$.

Definition 1. An eigenvector of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution of $A\mathbf{x} = \lambda\mathbf{x}$. We say $\mathbf{x}$ is the eigenvector corresponding to λ.

In the previous example, $\mathbf{u}$ is an eigenvector of A corresponding to the eigenvalue 4.

Example 2. Show that 7 is an eigenvalue of $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$ and find the corresponding eigenvector.

Observe that λ is an eigenvalue of A if the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. The solution set is the null space the matrix $A - \lambda I$.

Definition 2. For an eigenvalue λ of a matrix A, the eigenspace $E_\lambda(A)$ of A corresponding to λ is the null space of the matrix $A - \lambda I$. The dimension of the eigenspace corresponding to an eigenvalue λ is called the geometric multiplicity of λ, denoted $\mathrm{geomult}_\lambda(A)$.

Any eigenspace of a matrix A is automatically a subspace of $\mathbb{R}^n$ because null spaces are subspaces.

Example 3. Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. Find a basis for the eigenspace $E_2(A)$. What is $\mathrm{geomult}_2(A)$?

By a problem on Exam 2, 0 is an eigenvalue of A if and only if A is not invertible.

Theorem 4 (Invertible Matrix Theorem (cont.))**.** Let A be an $n \times n$ matrix. Then A is invertible if and only if

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

(20) The determinant of A is *not* 0.

2. The characteristic equation

Suppose λ is an eigenvalue of an $n \times n$ matrix A, then $A\mathbf{v} = \lambda\mathbf{v}$ for some nonzero $\mathbf{v}$. Said another way, the matrix equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. That is, $(A - \lambda I)$ is **not** invertible. Hence, $\det(A - \lambda I) = 0$. The roots of the this equation are the eigenvalues of A.

Example 5. Find the eigenvalues of each matrix below.

$$(1)\ \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} \qquad (2)\ \begin{bmatrix} 1 & 5 \\ 1 & 2 \end{bmatrix} \qquad (3)\ \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \qquad (4)\ \begin{bmatrix} 2 & 1 & 7 & 0 \\ 0 & 5 & 2 & 4 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Recall that the multiplicity of a root λ in an equation $p(t)$ is the number of times $(\lambda - t)$ appears in the factorization of $p(t)$. That is, $p(t) = (\lambda - t)^{\alpha} q(t)$ such that $(\lambda - t)$ does not divide $q(t)$, then the multiplicity of λ is α.

Definition 3. For an $n \times n$ matrix A, the characteristic polynomial of A is $\chi_A(t) = \det(A - tI) = 0$. If λ is a root of the characteristic equation, then the multiplicity of the root is called the algebraic multiplicity of λ, denoted $\text{algmult}_{\lambda}(A)$.

Theorem 6. The eigenvalues of a triangular matrix are the entries along the main diagonal.

Proof. If A is triangular, then so is $A - \lambda I$ and hence $\det(A - tI) = (a_{11} - t)(a_{22} - t) \cdots (a_{nn} - t)$. Thus, the roots of the characteristic equation are the diagonal entries of A. □

Definition 4. If A and B are $n \times n$ matrices, then A is said to be similar to B if there exists an invertible matrix P such that $B = P^{-1}AP$.

Theorem 7. If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues.

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

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

Thus,

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

□

3. Diagonalization

Diagonal matrices are some of the easiest to work with.

Example 8. Let $D = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}$. Find a formula for D^k.

We'll see in this section how to use eigenvalues and eigenvectors to diagonalize matrices.

Definition 5. An $n \times n$ matrix is diagonalizable if it is similar to a diagonal matrix.

Example 9. Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for A^k given that $A = PDP^{-1}$ with $P = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$.

To diagonalize a matrix is to find P invertible and D diagonal such that $D = P^{-1}AP$.

Theorem 10. An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the (linearly independent) eigenvectors of A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Set $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ and D to be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$. By IMT, P is invertible and

$$AP = A \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} = PD.$$

Hence, $D = P^{-1}AD$ so A is diagonalizable.

Conversely, suppose A is diagonalizable. Then there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$. Write $P = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}$ and let $c_1, \ldots, c_n$ be the diagonal entries of D. As above, $AP = PD$ implies $A\mathbf{x}_i = c_1\mathbf{x}_i$ for each i. But then each $\mathbf{x}_i$ is an eigenvector for A. Because P is invertible, the vectors $\mathbf{x}_i$ are linearly independent. □

Under the hypotheses of Theorem **??**, the eigenvectors of A form a basis of $\mathbb{R}^n$ called an eigenbasis of $\mathbb{R}^n$ with respect to A.

By Theorem **??**, the following steps diagonalize an $n \times n$ matrix A.

(1) Find the eigenvalues of A.
(2) Find the linearly independent eigenvectors of A.
(3) Construct P with columns from eigenvectors.
(4) Construct D from eigenvalues in order corresponding to P.
(5) Check!

Example 11. Diagonalize the following matrix.

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

The eigenvalues for this matrix are 1 and -2. We compute a basis for $E_\lambda(A)$ for each eigenvalue.

$$A - (1)I \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\},$$

$$A - (-2)I \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Thus, A is diagonalizable by the matrices

$$P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

When is it that a matrix is diagonalizable?

Theorem 12. If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

Proof. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly dependent. Because $\mathbf{v}_1 \neq \mathbf{0}$, then there exists $p \in \{2, \ldots, r\}$ such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent and $\mathbf{v}_{p+1}$ is a linear combination of those vectors. That is, there exist $c_1, \ldots, c_p$ not all zero such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1}. \tag{1}$$

Multiplying both sides of (**??**) by A gives

$$\begin{aligned} c_1 A\mathbf{v}_1 + \cdots + c_p A\mathbf{v}_p &= A\mathbf{v}_{p+1} \\ c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}\mathbf{v}_{p+1} \\ c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) \\ c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p &= \mathbf{0}. \end{aligned}$$

This contradicts the linear independence of $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ because none of the $\lambda_i - \lambda_{p+1}$ are zero by hypothesis. $\square$

The next theorem follows almost directly from Theorem **??**.

Theorem 13. An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Here is the more general case. It says that A is diagonalizable if the dimension of the eigenspace for each eigenvalue matches multiplicity of that eigenvalue as a root of the characteristic polynomial.

Theorem 14. For every eigenvalue λ of an $n \times n$ matrix A,

$$\text{geomult}_\lambda(A) \leq \text{algmult}_\lambda(A).$$

Proof. Let λ be an eigenvalue of A and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for $E_\lambda(A)$. We extend[1] $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$. Note that none of the $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ are eigenvectors for λ (if they were, they would be basis elements of $E_\lambda(A)$).

Let $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$, so P is invertible by the IMT. Set $B = P^{-1}AP$. Then B is a block matrix of the form

$$B = \left[\begin{array}{cccc|c} \lambda & 0 & \cdots & 0 & \\ 0 & \lambda & \cdots & 0 & \\ \vdots & & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & \lambda & \\ \hline & & 0 & & \star \end{array}\right]$$

Since B is similar to A, they have the same characteristic equation and hence the same algebraic multiplicity for each eigenvalue. It is clear that the algebraic multiplicity of λ in B is k. $\square$

Theorem 15. An $n \times n$ matrix A is diagonalizable if and only if $\text{geomult}_\lambda(A) = \text{algmult}_\lambda(A)$ for every eigenvalue λ.

Proof. The statement $\text{geomult}_\lambda(A) = \text{algmult}_\lambda(A)$ for every eigenvalue λ is equivalent to $\mathbb{R}^n$ having an eigenbasis with respect to A, which is equivalent to A having n linearly independent eigenvectors. $\square$

Example 16. Is the following matrix diagonalizable?

$$A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

The eigenvalues of this matrix are 1 and -2. The algebraic multiplicity of -2 is 2. However, the geometric multiplicity of -2 is 1 (check). Hence, A is *not* diagonalizable.

Example 17. Diagonalize the following matrix.

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

[1]As discussed in Chapter 2, we add linearly independent vectors of our original set until it spans $\mathbb{R}^n$.

4. Orthogonal diagonalization of symmetric matrices

We have seen already that it is quite time intensive to determine whether a matrix is diagonalizable. We'll see that there are certain cases when a matrix is always diagonalizable.

Recall that a matrix is *symmetric* if $A^T = A$. It is (reasonably) easy to show that the product of symmetric matrices is symmetric, and the inverse of a symmetric

Let $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$.

The eigenvalues of A are $\lambda_1 = -2$ and $\lambda_2 = 4$ and corresponding eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Hence, A is diagonalizable. Set $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$, then $P^{-1}AP = D$ where $D = \begin{bmatrix} -2 & 0 \\ 0 & 4 \end{bmatrix}$.

But note that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. Normalizing these vectors we have

$$\mathbf{u}_1 = \frac{1}{\mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \qquad \mathbf{u}_2 = \frac{1}{\mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

Set $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$. Since Q is an orthogonal matrix, then $Q^{-1} = Q^T$, so we have $Q^T AQ = D$. This is adventageous since computing the transpose is much easier than computing the inverse.

Orthogonally diagonlizable An square matrix A is *orthogonally diagonalizable* if there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$.

The next theorem is easy to prove. Its converse is also true, but this will take much more work to prove.

Orthogonally diagonalizable implies symmetric If A is orthogonally diagonalizable, then A is symmetric.

Proof. Since A is orthogonally diagonalizable, then $Q^T AQ = D$ for some orthogonal matrix Q and diagonal matrix D. But then $A = (Q^T)^{-1}DQ^T = QDQ^T$. Hence,

$$A^T = (QDQ^T) = (Q^T)^T D^T Q^T = QDQ^T = A. \qquad \square$$

Recall that matrices like $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, though all the entries are real, have complex eigenvalues. It turns out that this never happens with symmetric matrices.

We write complex numbers in $\mathbb{C}$ as $a + bi$ where a and b are real numbers and $i = \sqrt{-1}$. The *complex conjugate* of $z = a + bi$ is $\overline{z} = \overline{a + bi} = a - bi$. We can extend this notion to complex matrices. The conjugate of a matrix $A = [a_{ij}]$ with complex entries is $\overline{A} = [\overline{a_{ij}}]$.

Eigenvalues of symmetric matrices are real If A is a real symmetric matrix, then the eigenvalues of A are real.

Proof. Suppose λ is an eigenvalue of A with corresponding eigenvector $\mathbf{v}$. Then $A\mathbf{v} = \lambda\mathbf{v}$ and so

$$A\overline{\mathbf{v}} = \overline{A}\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

Now we take transposes. Using the fact that A is symmetric we have

$$\overline{\mathbf{v}}^T A = \overline{\mathbf{v}}^T A^T = (A\overline{\mathbf{v}})^T = (\overline{\lambda}\overline{\mathbf{v}})^T = \overline{\lambda}\overline{\mathbf{v}}^T.$$

Therefore,

$$\lambda(\overline{\mathbf{v}}^T\mathbf{v}) = \overline{\mathbf{v}}^T(\lambda\mathbf{v}) = \overline{\mathbf{v}}^T(A\mathbf{v}) = (\overline{\mathbf{v}}^T A)\mathbf{v} = (\overline{\lambda}\overline{\mathbf{v}}^T)\mathbf{v} = \overline{\lambda}(\overline{\mathbf{v}}^T\mathbf{v}).$$

This shows that $(\lambda - \overline{\lambda})(\overline{\mathbf{v}}^T\mathbf{v}) = 0$. Write

$$\mathbf{v} = \begin{bmatrix} a_1 + b_1 i \\ \vdots \\ a_n + b_n i \end{bmatrix} \quad \text{so} \quad \overline{\mathbf{v}} = \begin{bmatrix} a_1 - b_1 i \\ \vdots \\ a_n - b_n i \end{bmatrix}.$$

Then

$$\overline{\mathbf{v}}^T\mathbf{v} = \overline{\mathbf{v}} \cdot \overline{\mathbf{v}} = (a_1^2 + b_1)^2 + \cdots + (a_n^2 + b_n^2) \neq 0$$

because $\mathbf{v} \neq 0$. We conclude that $\lambda - \overline{\lambda} = 0$. That is $\lambda = \overline{\lambda}$ so λ is real. □

The next theorem is stronger than a previous result for general matrices.

Eigenvectors of symmetric matrices If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof. Let $\mathbf{v}_1, \mathbf{v}_2$ be eigenvectors for A with corresponding eigenvalues λ_1, λ_2, $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1\mathbf{v}_1)^T\mathbf{v}_2 = (A\mathbf{v}_1)^T\mathbf{v}_2 = \mathbf{v}_1^T A^T\mathbf{v}_2 = \mathbf{v}_1^T A\mathbf{v}_2 = \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2).$$

Hence, $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, then we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Based on the previous theorem, we say that the eigenspaces of A are *mutually orthogonal.*

Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$.

The eigenvalues of A are -2 and 7. The eigenspaces have bases,

$$E_7 : \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix} \right\}, \qquad E_{-2} : \left\{ \begin{bmatrix} -1 \\ -1/2 \\ 1 \end{bmatrix} \right\}.$$

It is easy now to verify that the basis vector for E_{-2} is orthogonal to those of E_7.

However, the two basis vectors for E_7 *are not* orthogonal. In order to orthgonally diagonalize A, we need an orthogonal basis for E_7. To do this, we use Gram-Schmidt:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/4 \\ 1 \\ 1/4 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Note for $\mathbf{v}_3$ we only scaled the vector. Finally, we normalize each vector,

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Now the matrix $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ is orthogonal and so $Q^T Q = I$.

The set of eigenvalues of a matrix A is called the *spectrum of* A and is denoted σ_A.

The (real) Spectral Theorem Let A be an $n \times n$ real matrix. Then A is symmetric if and only if it is orthogonally diagonalizable.

Proof. We already proved that orthogonally diagonalizable implies symmetric. Now we assume A is symmetric and prove that it is orthgonally diagonalizable. Clearly the result holds when A is 1×1. Assume $(n-1) \times (n-1)$ symmetric matrices are orthogonally diagonalizable.

Let A be $n \times n$ and let λ_1 be an eigenvalue of A and $\mathbf{u}_1$ a (unit) eigenvector for λ_1. Set $W = \text{Span}\{\mathbf{u}_1\}$. By the Gram-Schmidt process we may extend $\mathbf{u}_1$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbb{R}^n$ where $\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis for $W^\perp$. Set $Q_1 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$Q_1^T A Q_1 = \begin{bmatrix} \mathbf{u}_1^T A \mathbf{u}_1 & \cdots & \mathbf{u}_1^T A \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n^T A \mathbf{u}_1 & \cdots & \mathbf{u}_n^T A \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}.$$

The first column is as indicated because $\mathbf{u}_i^T A \mathbf{u}_1 = \mathbf{u}_i^T (\lambda \mathbf{u}_1) = \lambda(\mathbf{u}_i \cdot \mathbf{u}_1) = \lambda \delta_{ij}$. As $Q_1^T A Q_1$ is symmetric, $* = 0$ and B is a symmetric $(n-1) \times (n-1)$ matrix that is orthogonally diagonalizable with eigenvalues $\lambda_2, \dots, \lambda_n$ (by the inductive hypothesis). Because A and $Q_1^T A Q_1$ are similar, then the eigenvalues of A are $\lambda_1, \dots, \lambda_n$.

Since B is orthogonally diagonalizable, there exists an orthogonal matrix Q_2 such that $Q_2^T B Q_2 = D$, where the diagonal entries of D are $\lambda_2, \dots, \lambda_n$. Now

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}^T \begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q_2^T B Q_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D \end{bmatrix}.$$

Note that $\begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}$ is orthogonal. Set $Q = Q_1 \begin{bmatrix} 1 & 0 \\ 0 & Q_2 \end{bmatrix}$. As the product of orthogonal matrices is orthogonal, Q is itself orthogonal and $Q^T A Q$ is diagonal. $\square$

Let A be orthogonally diagonalizable. Then $A = QDQ^T$ where

$$Q = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$$

and D is the diagonal matrix whose diagonal entries are the eigenvalues of A: $\lambda_1, \ldots, \lambda_n$. Then

$$A = QDQ^T = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T.$$

This is known as the *spectral decomposition of* A, or *projection form of the Spectral Theorem.* Each $\mathbf{u}_i \mathbf{u}_i^T$ is called a *projection matrix* because $(\mathbf{u}_i \mathbf{u}_i^T)\mathbf{x}$ is the projection of $\mathbf{x}$ onto $\mathrm{Span}\{\mathbf{u}_i\}$.

Example 18. Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$. An orthonormal basis of the column space is

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\sqrt{2}/6 \\ 2\sqrt{2}/3 \\ \sqrt{2}/6 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Setting $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ gives $Q^T AQ = D = (7, 7, -2)$. The projection matrices are

$$\mathbf{u}_1 \mathbf{u}_1^T = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \mathbf{u}_2 \mathbf{u}_2^T = \frac{1}{18} \begin{bmatrix} 1 & -4 & -1 \\ -4 & 16 & 4 \\ -1 & 4 & 1 \end{bmatrix}, \mathbf{u}_3 \mathbf{u}_3^T = \frac{1}{9} \begin{bmatrix} 4 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 4 \end{bmatrix}.$$

The spectral decomposition is

$$7\mathbf{u}_1 \mathbf{u}_1^T + 7\mathbf{u}_2 \mathbf{u}_2^T - 2\mathbf{u}_3 \mathbf{u}_3^T = A.$$

7. Applications to Differential Equations

Example 19. Let $x(t)$ be a differentiable function of t. Find all solutions to the differential equation

$$x'(t) = 2x(t).$$

This is calculus. Write $x = x(t)$. By separation of variables,

$$\frac{dx}{x} = 2dt \Rightarrow \ln x = 2t + C.$$

Hence, the general solution is $x = ce^{2t}$ for any constant c.

A system of linear first-order differential equations (DEs) is a set of equations

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n. \end{aligned}$$

The x_i are all differentiable functions in some variable (say t), so we are actually using shorthand here with x_i in place of $x_i(t)$ where x_i' represents the derivative (with respect to t) and the a_{ij} are all constants. We can represent the entire system in matrix-vector form as

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

where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}, \qquad A = (a_{ij}).$$

A solution to the system is a vector-valued function $\mathbf{x}(t)$ that satisfies (**??**). An initial value problem is a system of DEs along with an initial condition $\mathbf{x}_0 = \mathbf{x}(0)$.

Also note that (**??**) is linear. If $c, d \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v}$ are solutions to (**??**), then

$$(c\mathbf{u} + d\mathbf{v})' = c\mathbf{u}' + d\mathbf{v}' = c(A\mathbf{u}) + d(A\mathbf{v}) = A(c\mathbf{u} + d\mathbf{v}).$$

Moreover, $\mathbf{0}$ is (trivially) a solution and so the set of solutions is a subspace of $\mathbb{R}^n$.

Example 20. Let $\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t)) \end{bmatrix}$ and $A = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}$.

It follows that our solutions are $x_1 = c_1e^{3t}$ and $x_2 = c_2e^{-5t}$. Thus, the solution space is

$$\begin{bmatrix} x_1(t) \\ x_2(t)) \end{bmatrix} = \begin{bmatrix} c_1e^{3t} \\ c_2e^{-5t} \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} e^{3t} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} c_2e^{-5t}.$$

This example suggests that solutions to a system of linear equations might have a simple presentation as a linear combination of functions of the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$.

Suppose a solution to (**??**) has the form $\mathbf{x} = \mathbf{v}e^{\lambda t}$. By calculus, $\mathbf{x}'(t) = \lambda \mathbf{v}e^{\lambda t}$. But then $A\mathbf{x} = A\mathbf{v}e^{\lambda t}$ (this uses the fact that $e^{\lambda t} \neq 0$ for all t). Hence, $A\mathbf{v} = \lambda\mathbf{v}$, that is, $\mathbf{v}$ is an eigenvector with eigenvalue λ. Solutions of this form are called eigenfunctions.

Example 21. Solve the initial value problem $\mathbf{x}'(t) = A\mathbf{x}(t)$ given

$$\mathbf{x} = \begin{bmatrix} x_1(t) \\ x_2(t)) \end{bmatrix}, \qquad A = \begin{bmatrix} -1.5 & .5 \\ 1 & -1 \end{bmatrix}, \qquad \mathbf{x}(0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}.$$

Eigenvalues are $-.5, -2$ with corresponding eigenvectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Thus, the eigenfunctions are

$$x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t}, \qquad x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t},$$

and any solution is a linear combination of these two. That is,

$$\mathbf{x} = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} e^{-.5t} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} e^{-2t}$$

and we can use the initial conditions to solve for c_1, c_2.

Chapter 6: Orthogonality

(Last Updated: November 27, 2017)

These notes are derived primarily from *Linear Algebra and its applications* by David Lay (4ed). A few theorems have been moved around.

1. Inner products

We now return to a discussion on the geometry of vectors. There are many applications of the notion of orthogonality, some of which we will discuss. A basic (geometric) question that we will address shortly is the following. Suppose you are given a plane P and a point p (in $\mathbb{R}^3$). What is the distance from p to P? That is, what is the length of the shortest possible line segment that one could draw from p to P.

Definition 1. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The inner product of u and v is defined as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots + u_n v_n.$$

The inner product is also referred to as the dot product. Another product, the cross product, will be discussed at a later time.

Example 1. Let $\mathbf{a} = \begin{bmatrix} -1 \\ 3 \\ 5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Compute $\mathbf{a} \cdot \mathbf{b}$ and $\mathbf{b} \cdot \mathbf{a}$.

Theorem 2. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and let $c \in \mathbb{R}$. Then

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

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

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

(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

The inner product is a useful tool to study the geometry of vectors.

Definition 2. The length (or norm) of $\mathbf{v} \in \mathbb{R}^n$ is the non-negative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \qquad \text{and} \qquad \|\mathbf{v}\| = \mathbf{v} \cdot \mathbf{v}.$$

A unit vector is a vector of length 1. Note that $\|c\mathbf{v}\| = |c| \, \|\mathbf{v}\|$ for $c \in \mathbb{R}$.

If $\mathbf{v} \neq 0$, then $\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ is the unit vector in the same direction as $\mathbf{v}$.

Example 3. Find the lengths of $\mathbf{a}$ and $\mathbf{b}$ in Example 1 and find their associated unit vectors.

Recall that the distance between two points (a_1, b_1) and (a_2, b_2) in $\mathbb{R}^2$ is determined by the well-known distance formula $\sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}$. We can similarly define distance between vectors.

Definition 3. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the distance between $\mathbf{u}$ and $\mathbf{v}$, written $d(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is, $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Example 4. Find the distance between $\mathbf{a}$ and $\mathbf{b}$ in Example 1.

Definition 4. Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be orthogonal (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.

Orthogonality generalizes the idea of perpendicular lines in $\mathbb{R}^2$. Two lines (represented as vectors $\mathbf{u}, \mathbf{v}$) are perpendicular if and only if the distance from $\mathbf{u}$ to $\mathbf{v}$ equals the distance from $\mathbf{u}$ to $-\mathbf{v}$.

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

Hence, these two quantities are equal if and only if $2\mathbf{u} \cdot \mathbf{v} = -2\mathbf{u} \cdot \mathbf{v}$. Equivalently, $\mathbf{u} \cdot \mathbf{v} = 0$. The next theorem now follows directly.

Theorem 5 (The Pythagorean Theorem)**.** Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal if and only if $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Definition 5. Let $W \subset \mathbb{R}^n$ be a subspace. The set $W^\perp = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{z} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}$ is called the orthogonal complement of W.

Part of your homework will be to show that $W^\perp$ is a subspace of $\mathbb{R}^n$.

Example 6. Let W be a plane through the origin in $\mathbb{R}^3$ and let L be a line through $\mathbf{0}$ perpendicular to W. If $\mathbf{z} \in L$ and $\mathbf{w} \in W$ are nonzero, then the line segment from $\mathbf{0}$ to $\mathbf{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\mathbf{w}$. In fact, $L = W^\perp$ and $W = L^\perp$.

Theorem 7. Let A be an $n \times n$ matrix. Then $(\text{Row}A)^\perp = \text{Nul}A$ and $(\text{Col}A)^\perp = \text{Nul}A^T$.

Proof. If $\mathbf{x} \in \text{Nul}A$, then $A\mathbf{x} = 0$ by definition. Hence, $\mathbf{x}$ is perpendicular to each row of A. Since the rows of A span $\text{Row}A$, then $\mathbf{x} \in (\text{Row}A)^\perp$. Conversely, if $\mathbf{x} \in (\text{Row}A)^\perp$, then $\mathbf{x}$ is orthogonal to each row of A and $A\mathbf{x} = 0$, so $\mathbf{x} \in \text{Nul}A$. The proof of the second statement is similar. □

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ be nonzero. By the Law of Cosines, $\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$. Rearranging gives

$$\begin{aligned}\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2\right] = \frac{1}{2}\left[(u_1^2 + u_2^2) + (v_1^2 + v_2^2) - (u_1 - v_1)^2 - (u_2 - v_2)^2\right] \\ &= u_1v_1 + u_2v_2 = \mathbf{u} \cdot \mathbf{v}.\end{aligned}$$

Hence,

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

2. Orthogonal Sets

Definition 6. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an orthogonal if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$. If, in addition, each $\mathbf{u}_i$ is a unit vector, then the set is said to be orthonormal.

Example 8. Show that the following set is orthogonal. Is it orthonormal? If not, find a set of orthonormal vectors with the same span.

$$\left\{ \begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix} \right\}$$

Theorem 9. If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence a basis for a subspace spanned by S.

Proof. Write $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$. Then

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

Since $\mathbf{u}_1 \cdot \mathbf{u}_1 \neq 0$ (because $\mathbf{u}_1 \neq 0$), then $c_1 = 0$. Repeating this argument with $\mathbf{u}_2, \ldots, \mathbf{u}_p$ gives $c_2 = \cdots = c_p = 0$. Hence, S is linearly independent. □

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Let $\mathbf{u} \in W$ and write $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$. Then $\mathbf{y} \cdot \mathbf{u}_i = c_i(\mathbf{u}_i \cdot \mathbf{u}_i)$, and so

$$c_i = \frac{y \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}, \qquad i = 1, \ldots, p.$$

Example 10. Show that the set S below is an orthogonal basis of $\mathbb{R}^3$ and express the given vector $\mathbf{x}$ as a linear combination of these vectors.

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \right\}, \qquad \mathbf{x} = \begin{bmatrix} 8 \\ -4 \\ -3 \end{bmatrix}.$$

Here is an easier version of the problem hinted at in the beginning of this chapter. Given a point p and a line L (in $\mathbb{R}^2$), what is the distance from p to L. The solution of this uses orthogonal projections.

Let $L = \text{Span}\{\mathbf{u}\}$ and p given by the vector $\mathbf{y}$. We need to know the length of the vector orthogonal to $\mathbf{u}$ through $\mathbf{y}$. By translation, this is equivalent to the length of the vector $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α. Then

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

Hence, $\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$ and so $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$. Note that if we replace $\mathbf{u}$ by $c\mathbf{u}$ for any scalar c this definition does not change and thus we have defined the projection for all of L.

Definition 7. Given vectors $\mathbf{y}, \mathbf{u} \in \mathbb{R}^n$, and $L = \operatorname{Span}\{\mathbf{u}\}$, the orthogonal projection of $\mathbf{y}$ onto L is defined as

$$\hat{\mathbf{y}} = \operatorname{proj}_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}.$$

Note that this gives a decomposition of the vector $\mathbf{y}$ as $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}} \in L$ and $\mathbf{z} \in L^\perp$. Hence, every vector in $\mathbb{R}^n$ can be written (uniquely) as the sum of an element in L and an element in $L^\perp$. Since $L \cap L^\perp = \{\mathbf{0}\}$, then it follows that $\dim L^\perp = n - 1$. In the next section we will generalize this to larger subspaces.

Example 11. Compute the orthogonal projection of $\begin{bmatrix} 1 \\ 7 \end{bmatrix}$ onto the line L through $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ and the origin. Use this to find the distance from $\mathbf{y}$ to L.

Definition 8. If W is a subspace of $\mathbb{R}^n$ spanned by an orthonormal set $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then we say S is an orthonormal basis of W.

Example 12. The standard basis is an orthonormal basis of $\mathbb{R}^n$.

Theorem 13. An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

Proof. Write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \cdots \\ \mathbf{u}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \cdots & \mathbf{u}_1^T \mathbf{u}_n \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \cdots & \mathbf{u}_2^T \mathbf{u}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{u}_n^T \mathbf{u}_1 & \mathbf{u}_n^T \mathbf{u}_2 & \cdots & \mathbf{u}_n^T \mathbf{u}_n \end{bmatrix}.$$

Hence, $U^T U = I$ if and only if $\mathbf{u}_i \cdot \mathbf{u}_i = 1$ for all i and $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$. □

Theorem 14. Let U be an $m \times n$ matrix with orthonormal columns and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

(1) $\|U\mathbf{x}\| = \|\mathbf{x}\|$
(2) $(U\mathbf{x}) \cdot (U\mathbf{x}) = \mathbf{x} \cdot \mathbf{y}$
(3) $(U\mathbf{x}) \cdot (U\mathbf{x}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

Proof. We will prove (1). The rest are left as an exercise.

Write $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$\begin{aligned} \|U\mathbf{x}\| &= U\mathbf{x} \cdot U\mathbf{x} = (\mathbf{u}_1 x_1 + \cdots \mathbf{u}_n x_n) \cdot (\mathbf{u}_1 x_1 + \cdots \mathbf{u}_n x_n) \\ &= \sum_{i,j} (\mathbf{u}_i x_i) \cdot (\mathbf{u}_j x_j) = \sum_{i,j} x_i x_j (\mathbf{u}_i \cdot \mathbf{u}_j) = \sum_i x_i^2 (\mathbf{u}_i \cdot \mathbf{u}_i) = \sum_i x_i^2 = \|\mathbf{x}\|. \end{aligned}$$

□

3. Orthogonal Projections

The next definition generalizes projections onto lines.

Definition 9. Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. For $\mathbf{y} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{y}$ onto W is given by

$$\operatorname{proj}_W \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_n}{\mathbf{u}_n \cdot \mathbf{u}_n}\mathbf{u}_n.$$

This definition matches our previous one when W is 1-dimensional. Note that $\operatorname{proj}_W \mathbf{y} \in W$ because it is a linear combination of basis elements.

Also note that the definition simplifies when the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is orthonormal. In this case, if we let $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{bmatrix}$, then $\operatorname{proj}_W \mathbf{y} = UU^T\mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$.

Theorem 15 (Orthogonal Decomposition Theorem)**.** Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact $\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

Proof. Note that if $W = \{\mathbf{0}\}$, then this theorem is trivial.

As noted above, $\operatorname{proj}_W \mathbf{y} \in W$. We claim $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} \in W^\perp$.

$$\mathbf{z} \cdot \mathbf{u}_1 = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \hat{\mathbf{y}} \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \left(\frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right) \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \mathbf{y} \cdot \mathbf{u}_1 = 0.$$

It is clear that this holds similarly for $\mathbf{u}_2, \cdots, \mathbf{u}_p$. By linearity, $\mathbf{z} \cdot \mathbf{y} = 0$, so $\mathbf{z} \in W^\perp$.

To prove uniqueness, let $\mathbf{y} = \mathbf{w} + \mathbf{x}$ be another decomposition with $\mathbf{w} \in W$ and $\mathbf{x} \in W^\perp$. Then $\mathbf{w}+\mathbf{x} = \mathbf{y} = \hat{\mathbf{y}}+\mathbf{z}$, so $(\mathbf{w}-\hat{\mathbf{y}}) = (\mathbf{z}-\mathbf{x})$. But $(\mathbf{w}-\hat{\mathbf{y}}) \in W$ and $(\mathbf{z}-\mathbf{x}) \in W^\perp$. Since $W \cap W^\perp = \{\mathbf{0}\}$, then $\mathbf{w} - \hat{\mathbf{y}} = \mathbf{0}$ so $\mathbf{w} = \hat{\mathbf{y}}$. Similarly, $\mathbf{z} = \mathbf{x}$. $\square$

We will show in the next section that every subspace has an orthogonal basis.

Corollary 16. Let W be a subspace of $\mathbb{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Then $\mathbf{y} \in W$ if and only if $\operatorname{proj}_W \mathbf{y} = \mathbf{y}$.

Example 17. Let $W = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ below. Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. Write $\mathbf{y}$ (below) as a vector $\hat{\mathbf{y}} \in W$ and $z \in W^\perp$.

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 6 \end{bmatrix}.$$

Theorem 18 (Best Approximation Theorem)**.** Let W be a subspace of $\mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$. Then $\hat{\mathbf{y}} = \operatorname{proj}_W \mathbf{y}$ is the closest point to W in the sense that $\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$ for all $\mathbf{v} \in W$, $\mathbf{v} \neq \mathbf{y}$.

4. The Gram-Schmidt Process

Orthogonal projections give us a way to find an orthogonal basis for any W of $\mathbb{R}^n$.

Example 19. Let $W = \operatorname{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with $\mathbf{x}_1, \mathbf{x}_2$ below. Construct an orthogonal basis for W.

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}.$$

Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \operatorname{Span}\{\mathbf{v}_1\}$. It suffices to find a vector $\mathbf{v}_2 \in W$ orthogonal to W_1. Let $p = \operatorname{proj}_{W_1} \mathbf{x}_2 \in W_1$. Then $\mathbf{x}_2 = p + (\mathbf{x}_2 - p)$ where $\mathbf{x}_2 - p \in W_1^{\perp}$.

$$\mathbf{v}_2 = \mathbf{x}_2 - p = x_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} - \begin{bmatrix} 9 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}.$$

Now $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ and $\mathbf{v}_1, \mathbf{v}_2 \in W$. Hence, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. Note that if we wanted an orthonormal basis for W then we can just take the unit vectors associated to $\mathbf{v}_1$ and $\mathbf{v}_2$.

This process could continue. Say W was three-dimensional. We could then let $W_2 = \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and find the projection of x_3 onto W_2. We'll prove the next theorem using this idea.

Theorem 20 (The Gram-Schmidt Process)**.** Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace $W \subset \mathbb{R}^n$, define

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

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition, $\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ for all $1 \leq k \leq p$.

Proof. For $1 \leq k \leq p$, set $W_k = \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and $V_k = \operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$. Since $\mathbf{v}_1 = \mathbf{x}_1$. Then it (trivially) holds that $W_1 = V_1$ and $\{\mathbf{v}_1\}$ is orthogonal.

Suppose for some k, $1 \leq k < n$, that $W_k = V_k$ and that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal set. Define

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \operatorname{proj}_{W_k} \mathbf{x}_{k+1} \in W_k \subset W_{k+1}.$$

By the Orthogonal Decomposition Theorem, $\mathbf{v}_{k+1}$ is orthogonal to W_k. Since $\mathbf{x}_{k+1} \in W_{k+1}$, then $\mathbf{v}_{k+1} \in W_{k+1}$. Hence, $\{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ is an orthogonal set of $k+1$ nonzero vectors in W_{k+1} and hence a basis of W_{k+1}. Hence, $W_{k+1} = V_{k+1}$. The result now follows by induction. $\square$

Example 21. Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with $\mathbf{x}_i$ below. Construct an orthogonal basis for W.

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

Set $\mathbf{v}_1 = \mathbf{x}_1$. Then

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0 \\ 2/3 \\ 1/3 \end{bmatrix}.$$

Now,

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} - \frac{5}{2}\begin{bmatrix} 1/3 \\ 0 \\ 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix}.$$

Hence, an orthogonal basis for W is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

5. Least-squares problems

In data science, one often wants to be able to approximate a set of data by a curve. Possibly, one might hope to construct the line that best fits the data. This is known (by one name) as *linear regression.* In this section we'll study the linear algebra approach to this problem.

Suppose the system $A\mathbf{x} = \mathbf{b}$ is inconsistent. Previously, we gave up all hope then of "solving" this system because no solution existed. However, if we give up the idea that we must find an *exact* solution and instead focus on finding an *approximate* solution, then we may have hope of solving.

Definition 10. If A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$, a least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that for all $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$.

Geometrically, we think of $A\hat{\mathbf{x}}$ as the projection of $\mathbf{b}$ onto $\mathrm{Col}A$. That is, if $\hat{\mathbf{b}} = \mathrm{proj}_{\mathrm{Col}A}\,\mathbf{b}$, then the equation $A\mathbf{x} = \hat{\mathbf{b}}$ is consistent. Let $\hat{\mathbf{x}} \in \mathbb{R}^n$ be a solution (there may be several). By the Best Approximation Theorem, $\hat{\mathbf{b}}$ is the point on $\mathrm{Col}A$ closest to $\mathbf{b}$ and so $A\hat{\mathbf{x}}$ is a least-squares solution to $A\mathbf{x} = \mathbf{b}$

By the Orthogonal Decomposition Theorem, $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to $\mathrm{Col}A$. Hence, if $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - \hat{\mathbf{b}}) = 0$. That is, $\mathbf{a}_j^T(\mathbf{b} - \hat{\mathbf{b}}) = 0$. But $\mathbf{a}_j^T$ is a row of A^T and so $A^T(\mathbf{b} - \hat{\mathbf{b}})0\mathbf{0}$. Replacing $\hat{\mathbf{b}}$ with $A\hat{\mathbf{x}}$ and expanding we get $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. The equations corresponding to this system are the normal equations for $A\mathbf{x} = \mathbf{b}$. We have now essentially proven the following theorem.

Theorem 22. The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

Example 23. Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 5 & 1 \\ 4 & 1 \\ 3 & 0 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

We will use normal equations. First we compute

$$A^TA = \begin{bmatrix} 50 & 9 \\ 9 & 50 \end{bmatrix}, \qquad A^T\mathbf{b} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

To solve the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ we invert A^TA.

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{19}\begin{bmatrix} 2 & -9 \\ -9 & 50 \end{bmatrix}\begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/19 \\ 5/19 \end{bmatrix}.$$

Hence, when A^TA is invertible then the least-squares solution $\hat{\mathbf{x}}$ is unique and

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

As an application of this, we'll see how to fit data to a line using least-squares. To match notation commonly used in statistical analysis, we denote the equation $A\mathbf{x} = \mathbf{b}$ by $X\boldsymbol{\beta} = \mathbf{y}$. The matrix X is referred to as the design matrix, $\boldsymbol{\beta}$ as the parameter vector, and $\mathbf{y}$ as the observation vector.

Suppose we have a a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, perhaps from some experiments. We would like to model this data be a line to predict outcomes that did not appear in our experiment. Say this line is written $y = \beta_0 + \beta_1 x$. The residual of a point (x_i, y_i) is the distance from that point to the line. The least-squares line is the line that minimizes the sum of the squares of the residuals.

Suppose the data was all on the line. Then they would all satisfy,

$$\begin{aligned} \beta_0 + \beta_1 x_1 &= y_1 \\ \beta_0 + \beta_1 x_2 &= y_2 \\ &\vdots \\ \beta_0 + \beta_1 x_n &= y_n. \end{aligned}$$

We could write this system as $X\boldsymbol{\beta} = \mathbf{y}$ where

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

If the data does not lie on the line (and this is likely) then we want the vector $\boldsymbol{\beta}$ to be the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ that minimizes the distance between $X\boldsymbol{\beta}$ and $\mathbf{y}$.

Example 24. Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(4, 1)$, $(1, 2)$, $(3, 3)$, $(5, 5)$.

We build the matrix X and vector $\mathbf{y}$ from the data,

$$X = \begin{bmatrix} 1 & 4 \\ 1 & 1 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}.$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, we have the normal equation $X^T X\boldsymbol{\beta} = X^T\mathbf{y}$ where

$$X^T X = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix}, \quad X^T\mathbf{y} = \begin{bmatrix} 11 \\ 37 \end{bmatrix}.$$

Hence,

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (X^T X)^{-1} X^T\mathbf{y} = \begin{bmatrix} 41/35 \\ 17/35 \end{bmatrix}.$$

7.1 Diagonalization of Symmetric Matrices

We have seen already that it is quite time intensive to determine whether a matrix is diagonalizable. We'll see that there are certain cases when a matrix is always diagonalizable.

Definition 11. A matrix A is symmetric if $A^T = A$.

Example 25. Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$.

Note that $A^T = A$, so A is symmetric. The characteristic polynomial of A is $\chi_A(t) = (t+2)(t-7)^2$ so the eigenvalues are -2 and 7. The corresponding eigenspaces have bases,

$$\lambda = -2, \left\{ \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \right\}, \qquad \lambda = 7, \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Hence, A is diagonalizable. Now we use Gram-Schmidt to find an orthogonal basis for $\mathbb{R}^3$. Note that the eigenvector for $\lambda = -2$ is already orthogonal to both eigenvectors for $\lambda = 7$.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 2 \\ 1/2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Finally, we normalize each vector,

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/3\sqrt{2} \\ 2\sqrt{2}/3 \\ 1/3\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Now the matrix $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ is orthogonal and so $U^T U = I$.

Theorem 26. If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof. Let $\mathbf{v}_1, \mathbf{v}_2$ be eigenvectors for A with corresponding eigenvalues λ_1, λ_2, $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2).$$

Hence, $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, then we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. $\square$

Based on the previous theorem, we say that the eigenspaces of A are mutually orthogonal.

Definition 12. An $n \times n$ matrix A is orthogonally diagonalizable if there exists an orthogonal $n \times n$ matrix P and a diagonal matrix D such that $A = PDP^T$.

Theorem 27. If A is orthogonally diagonalizable, then A is symmetric.

Proof. Since A is orthogonally diagonalizable, then $A = PDP^T$ for some orthogonal matrix P and diagonal matrix D. A is symmetric because $A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$. $\square$

It turns out the converse of the above theorem is also true! The set of eigenvalues of a matrix A is called the spectrum of A and is denoted σ_A.

Theorem 28 (The Spectral Theorem for symmetric matrices)**.** Let A be a (real) $n \times n$ symmetric matrix. Then the following hold.

(1) A has n real eigenvalues, counting multiplicities.
(2) For each eigenvalue λ of A, $\operatorname{geomult}_\lambda(A) = \operatorname{algmult}_\lambda(A)$.
(3) The eigenspaces are mutually orthogonal.
(4) A is orthogonally diagonalizable.

Proof. Every eigenvalue of a symmetric matrix is real[1]. The second part of (1) as well as (2) are immediate consequences of (4). We proved (3) in Theorem 26. Note that (4) is trivial when A has n *distinct* eigenvalues by (3).

We prove (4) by induction. Clearly the result holds when A is 1×1. Assume $(n-1) \times (n-1)$ symmetric matrices are orthogonally diagonalizable.

Let A be $n \times n$ and let λ_1 be an eigenvalue of A and $\mathbf{u}_1$ a (unit) eigenvector for λ_1. By the Gram-Schmidt process, we may extend $\mathbf{u}_1$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbb{R}^n$ where $\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis for $W^\perp$.

Set $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$U^T A U = \begin{bmatrix} \mathbf{u}_1^T A \mathbf{u}_1 & \cdots & \mathbf{u}_1^T A \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n^T A \mathbf{u}_1 & \cdots & \mathbf{u}_n^T A \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}.$$

The first column is as indicated because $\mathbf{u}_i^T A \mathbf{u}_1 = \mathbf{u}_i^T(\lambda \mathbf{u}_1) = \lambda(\mathbf{u}_i \cdot \mathbf{u}_1) = \lambda \delta_{ij}$. As $U^T A U$ is symmetric, $* = 0$ and B is a symmetric $(n-1) \times (n-1)$ matrix that is orthogonally diagonalizable with eigenvalues $\lambda_2, \dots, \lambda_n$ (by the inductive hypothesis). Because A and $U^T A U$ are similar, then the eigenvalues of A are $\lambda_1, \dots, \lambda_n$.

Since B is orthogonally diagonalizable, there exists an orthogonal matrix Q such that $Q^T B Q = D$, where the diagonal entries of D are $\lambda_2, \dots, \lambda_n$. Now

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}^T \begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q^T B Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D \end{bmatrix}.$$

---

[1] This is one of the problems on the extra credit homework assignment.

Note that $\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal. Set $V = U \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$. As the product of orthogonal matrices is orthogonal, V is itself orthogonal and $V^T AV$ is diagonal. □

Suppose A is orthogonally diagonalizable, so $A = UDU^T$ where $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ and D is the diagonal matrix whose diagonal entries are the eigenvalues of A, $\lambda_1, \ldots, \lambda_n$. Then

$$A = UDU^T = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T.$$

This is known as the spectral decomposition of A. Each $\mathbf{u}_i \mathbf{u}_i^T$ is called a projection matrix because $(\mathbf{u}_i \mathbf{u}_i^T)\mathbf{x}$ is the projection of $\mathbf{x}$ onto $\text{Span}\{\mathbf{u}_i\}$.

Example 29. Construct a spectral decomposition of the matrix A in Example 25

Recall that $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$ and our orthonormal basis of $\text{Col}(A)$ was

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/3\sqrt{2} \\ 2\sqrt{2}/3 \\ 1/3\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Setting $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ gives $U^T AU = D = \text{diag}(-2, 7, 7)$. The projection matrices are

$$\mathbf{u}_1 \mathbf{u}_1^T = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}, \mathbf{u}_2 \mathbf{u}_2^T = \begin{bmatrix} 1/18 & -2/9 & -1/18 \\ -2/9 & \frac{8}{9} & 2/9 \\ -1/18 & 2/9 & 1/18 \end{bmatrix}, \mathbf{u}_3 \mathbf{u}_3^T = \begin{bmatrix} 4/9 & 2/9 & -4/9 \\ 2/9 & 1/9 & -2/9 \\ -4/9 & -2/9 & 4/9 \end{bmatrix}.$$

The spectral decomposition is

$$7\mathbf{u}_1 \mathbf{u}_1^T + 7\mathbf{u}_2 \mathbf{u}_2^T - 2\mathbf{u}_3 \mathbf{u}_3^T = A.$$

Chapter 7: Symmetric Matrices and Quadratic Forms

(Last Updated: December 5, 2016)

These notes are derived primarily from *Linear Algebra and its applications* by David Lay (4ed). A few theorems have been moved around.

1. Diagonalization of Symmetric Matrices

We have seen already that it is quite time intensive to determine whether a matrix is diagonalizable. We'll see that there are certain cases when a matrix is always diagonalizable.

Definition 1. A matrix A is symmetric if $A^T = A$.

Example 1. Let $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$.

Note that $A^T = A$, so A is symmetric. The characteristic polynomial of A is $\chi_A(t) = (t+2)(t-7)^2$ so the eigenvalues are -2 and 7. The corresponding eigenspaces have bases,

$$\lambda = -2, \left\{ \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \right\}, \qquad \lambda = 7, \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Hence, A is diagonalizable. Now we use Gram-Schmidt to find an orthogonal basis for $\mathbb{R}^3$. Note that the eigenvector for $\lambda = -2$ is already orthogonal to both eigenvectors for $\lambda = 7$.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/2 \\ 2 \\ 1/2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix}.$$

Finally, we normalize each vector,

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/3\sqrt{2} \\ 2\sqrt{2}/3 \\ 1/3\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Now the matrix $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ is orthogonal and so $U^T U = I$.

Theorem 2. If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

Proof. Let $\mathbf{v}_1, \mathbf{v}_2$ be eigenvectors for A with corresponding eigenvalues λ_1, λ_2, $\lambda_1 \neq \lambda_2$. Then

$$\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) = (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 = (A\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T A \mathbf{v}_2 = \mathbf{v}_1^T(\lambda_2 \mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2).$$

Hence, $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, then we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Based on the previous theorem, we say that the eigenspaces of A are mutually orthogonal.

Definition 2. An $n \times n$ matrix A is orthogonally diagonalizable if there exists an orthogonal $n \times n$ matrix P and a diagonal matrix D such that $A = PDP^T$.

Theorem 3. If A is orthogonally diagonalizable, then A is symmetric.

Proof. Since A is orthogonally diagonalizable, then $A = PDP^T$ for some orthogonal matrix P and diagonal matrix D. A is symmetric because $A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$. □

It turns out the converse of the above theorem is also true! The set of eigenvalues of a matrix A is called the spectrum of A and is denoted σ_A.

Theorem 4 (The Spectral Theorem for symmetric matrices)**.** Let A be a (real) $n \times n$ matrix. Then the following hold.

(1) A has n real eigenvalues, counting multiplicities.
(2) For each eigenvalue λ of A, $\text{geomult}_\lambda(A) = \text{algmult}_\lambda(A)$.
(3) The eigenspaces are mutually orthogonal.
(4) A is orthogonally diagonalizable.

Proof. We proved in HW9, Exercise 6 that every eigenvalue of a symmetric matrix is real. The second part of (1) as well as (2) are immediate consequences of (4). We proved (3) in Theorem 2. Note that (4) is trivial when A has n *distinct* eigenvalues by (3).

We prove (4) by induction. Clearly the result holds when A is 1×1. Assume $(n-1) \times (n-1)$ symmetric matrices are orthogonally diagonalizable.

Let A be $n \times n$ and let λ_1 be an eigenvalue of A and $\mathbf{u}_1$ a (unit) eigenvector for λ_1. By the Gram-Schmidt process, we may extend $\mathbf{u}_1$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbb{R}^n$ where $\{\mathbf{u}_2, \dots, \mathbf{u}_n\}$ is a basis for $W^\perp$.

Set $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$. Then

$$U^T A U = \begin{bmatrix} \mathbf{u}_1^T A \mathbf{u}_1 & \cdots & \mathbf{u}_1^T A \mathbf{u}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_n^T A \mathbf{u}_1 & \cdots & \mathbf{u}_n^T A \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ 0 & B \end{bmatrix}.$$

The first column is as indicated because $\mathbf{u}_i^T A \mathbf{u}_1 = \mathbf{u}_i^T (\lambda \mathbf{u}_1) = \lambda(\mathbf{u}_i \cdot \mathbf{u}_1) = \lambda \delta_{ij}$. As $U^T A U$ is symmetric, $* = 0$ and B is a symmetric $(n-1) \times (n-1)$ matrix that is orthogonally diagonalizable with eigenvalues $\lambda_2, \dots, \lambda_n$ (by the inductive hypothesis). Because A and $U^T A U$ are similar, then the eigenvalues of A are $\lambda_1, \dots, \lambda_n$.

Since B is orthogonally diagonalizable, there exists an orthogonal matrix Q such that $Q^T BQ = D$, where the diagonal entries of D are $\lambda_2, \ldots, \lambda_n$. Now

$$\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}^T \begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & Q^T BQ \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & D \end{bmatrix}.$$

Note that $\begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$ is orthogonal. Set $V = U \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$. As the product of orthogonal matrices is orthogonal, V is itself orthogonal and $V^T AV$ is diagonal. □

Suppose A is orthogonally diagonalizable, so $A = UDU^T$ where $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ and D is the diagonal matrix whose diagonal entries are the eigenvalues of A, $\lambda_1, \ldots, \lambda_n$. Then

$$A = UDU^T = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T.$$

This is known as the spectral decomposition of A. Each $\mathbf{u}_i \mathbf{u}_i^T$ is called a projection matrix because $(\mathbf{u}_i \mathbf{u}_i^T)\mathbf{x}$ is the projection of $\mathbf{x}$ onto $\mathrm{Span}\{\mathbf{u}_i\}$.

Example 5. Construct a spectral decomposition of the matrix A in Example 1

Recall that $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$ and our orthonormal basis of $\mathrm{Col}(A)$ was

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/3\sqrt{2} \\ 2\sqrt{2}/3 \\ 1/3\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}.$$

Setting $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ gives $U^T AU = D = \mathrm{diag}(-2, 7, 7)$. The projection matrices are

$$\mathbf{u}_1 \mathbf{u}_1^T = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}, \mathbf{u}_2 \mathbf{u}_2^T = \begin{bmatrix} 1/18 & -2/9 & -1/18 \\ -2/9 & \frac{8}{9} & 2/9 \\ -1/18 & 2/9 & 1/18 \end{bmatrix}, \mathbf{u}_3 \mathbf{u}_3^T = \begin{bmatrix} 4/9 & 2/9 & -4/9 \\ 2/9 & 1/9 & -2/9 \\ -4/9 & -2/9 & 4/9 \end{bmatrix}.$$

The spectral decomposition is

$$7\mathbf{u}_1 \mathbf{u}_1^T + 7\mathbf{u}_2 \mathbf{u}_2^T - 2\mathbf{u}_3 \mathbf{u}_3^T = A.$$

2. Quadratic Forms

Definition 3. A quadratic form is a function Q on $\mathbb{R}^n$ given by $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ where A is an $n \times n$ symmetric matrix, called the matrix of the quadratic form.

Example 6. The function $\mathbf{x} \mapsto \|\mathbf{x}\|$ is a quadratic form given by setting $A = I$.

Quadratic forms appear in differential geometry, physics, economics, and statistics.

Example 7. Let $A = \begin{bmatrix} 5 & -1 \\ -1 & 2 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The corresponding quadratic form is

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = 5x_1^2 - 2x_1x_2 + 2x_2^2.$$

Example 8. Find the matrix of the quadratic form $Q(\mathbf{x}) = 8x_1^2 + 7x_2^2 - 3x_3^2 - 6x_1x_2 + 4x_1x_3 - 2x_2x_3$. By inspection we see that

$$A = \begin{bmatrix} 8 & -3 & 2 \\ -3 & 7 & -1 \\ 2 & -1 & -3 \end{bmatrix}.$$

Theorem 9 (Principal Axes Theorem)**.** Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ such that D is diagonal.

Proof. By the Spectral Theorem, there exists an orthogonal matrix P such that $P^T AP = D$ with D diagonal. For all $\mathbf{x} \in \mathbb{R}^n$, set $\mathbf{y} = P^T\mathbf{x}$. Then $\mathbf{x} = P\mathbf{y}$ and

$$\mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T(P^T AP)\mathbf{y} = \mathbf{y}^T D\mathbf{y}. \qquad \square$$

Example 10. Let $Q(\mathbf{x}) = 3x_1^2 - 4x_1x_2 + 6x_2^2$. Make a change of variable that transforms Q into a quadratic form with no cross-product terms.

We have $A = \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix}$ with eigenvalues 7 and -2, and eigenbases $\left\{\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right\}, \left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$, respectively. We normalize each to determine our diagonalizing matrix P, so that $P^T AP = D$ where

$$P = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix}, \text{ and } D = \begin{bmatrix} 7 & 0 \\ 0 & 2 \end{bmatrix}.$$

Our change of variable is $\mathbf{x} = P\mathbf{y}$ and the new form is $Q'(\mathbf{x}) = \mathbf{y}^T D\mathbf{y} = 7y_1^2 + 2y_2^2$.

If A is a symmetric $n \times n$ matrix, the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a real-valued function with domain $\mathbb{R}^n$. If $n = 2$, then the graph of $Q(\mathbf{x})$ is the set of points (x_1, x_2, z) with $z = Q(\mathbf{x})$. For example, if $Q(\mathbf{x}) = 3x_1^2 + 7x_2^2$, then $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$. Such a form is called positive definite, and it is possible to determine this property from the eigenvalues of A.

Definition 4. A quadratic form Q is

(1) positive definite if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
(2) negative definite if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
(3) indefinite if $Q(\mathbf{x})$ assumes positive and negative values.

Theorem 11. Let A be an $n \times n$ symmetric matrix. Then $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is

(1) positive definite if and only if the eigenvalues of A are all positive.
(2) negative definite if and only if the eigenvalues of A are all negative.
(3) indefinite otherwise.

Proof. By the Principal Axes Theorem, there exists an orthogonal matrix P such that $\mathbf{x} = P\mathbf{y}$ and

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 \mathbf{y}_1^2 + \cdots + \lambda_n \mathbf{y}_n^2$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. Since P is invertible, there is a 1-1 correspondence between all nonzero $\mathbf{x}$ and nonzero $\mathbf{y}$. The values above for $\mathbf{y} \neq 0$ are clearly determined by the signs of the λ_i. Hence so are the corresponding values of $\mathbf{x} \neq 0$. $\square$

We can also determine the maximum and minimum of a quadratic form when evaluated on a *unit vector*. This is known as *constrained optimization*

Theorem 12. Let A be a symmetric matrix. Set

$$m = \min\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\} \quad \text{and} \quad M = \max\{\mathbf{x}^T A\mathbf{x} : \|\mathbf{x}\| = 1\}$$

The M is the greatest eigenvalue of A and m is the least eigenvalue of A. Moreover, $\mathbf{x}^T A\mathbf{x} = M$ (resp. m) when $\mathbf{x}$ is the unit eigenvector corresponding to M (resp. m).

Example 13. Let $Q(\mathbf{x}) = 7x_1^2 + x_2^2 + 7x_3^2 - 8x_1x_2 - 4x_1x_3 - 8x_2x_3$. Find a vector $\mathbf{x}$ such that $Q(\mathbf{x})$ is maximimized (mimimized) subject to $\mathbf{x}^T\mathbf{x} = 1$.

The matrix of Q is $A = \begin{bmatrix} 7 & -4 & -2 \\ -4 & 1 & -4 \\ -2 & -4 & 7 \end{bmatrix}$ and the eigenvalues are $-3, 9$. Hence, the maximum (resp. minimum) values of Q subject to the constraint is 9 (resp. 3).

An eigenvector for 9 is $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Hence, setting $\mathbf{u} = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$ gives $Q(\mathbf{u}) = 9$.

An eigenvector for -3 is $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. Hence, setting $\mathbf{v} = \begin{bmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$ gives $Q(\mathbf{v}) = 6$.

4. Singular Value Decomposition

The key question in this section is whether it is possible to diagonalize a non-square matrix.

Example 14. Let $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$. A linear transformation with matrix A maps the unit sphere $\{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\}$ onto an ellipse in $\mathbb{R}^2$. Find a unit vector $\mathbf{x}$ at which the length $\|\mathbf{x}\|$ is maximized and compute its length.

The key observation here is that $\|A\mathbf{x}\|$ is maximized at the same $\mathbf{x}$ that maximizes $\|A\mathbf{x}\|^2$ and

$$\|A\mathbf{x}\|^2 = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^T(A^TA)\mathbf{x}.$$

Thus, we want to maximize the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T(A^TA)\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$. The eigenvalues of

$$A^TA = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 220 \end{bmatrix}$$

are $\lambda_1 = 360$, $\lambda_2 = 90$, $\lambda_3 = 0$ with corresponding (unit) eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}.$$

The maximum value is 360 obtained when $\mathbf{x} = \mathbf{v}_1$. That is, the vector $A\mathbf{v}_1$ corresponds to the point on the ellipse furthest from the origin. Then

$$A\mathbf{v}_1 = \begin{bmatrix} 18 \\ 6 \end{bmatrix} \quad \text{and} \quad \|A\mathbf{v}_1\| = \sqrt{360} = 6\sqrt{10}.$$

The trick we utilized above is a handy one. That is, even though A is not symmetric (it wasn't even square), A^TA is symmetric and we can extract information about A from A^TA.

Let A be an $m \times n$ matrix. Then A^TA can be orthogonally diagonalized. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal A^TA-eigenbasis for $\mathbb{R}^n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. For $1 \leq i \leq n$,

$$0 \leq \|A\mathbf{v}_i\|^2 = (A\mathbf{v}_i)^T(A\mathbf{v}_i) = \mathbf{v}_i^T(A^TA)\mathbf{v}_i = \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) = \lambda_i\mathbf{v}_i^T\mathbf{v}_i = \lambda_i.$$

Hence, $\lambda_i \geq 0$ for all i. Arrange the eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$ and define $\sigma_i = \sqrt{\lambda_i}$. That is, the σ_i represent the lengths of the vectors $A\mathbf{v}_i$. These are called the singular values of A.

Example 15. In Example 14, the singular values are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \quad \sigma_2 = \sqrt{90} = 3\sqrt{10}, \quad \sigma_3 = 0.$$

Theorem 16. Let A be an $m \times n$ matrix. Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA with corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$. Suppose A has r singular values. Then $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for $\mathrm{Col}(A)$ and $\mathrm{rank}A = r$.

Proof. Suppose $i \neq j$, then

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = (A\mathbf{v}_i)^T(A\mathbf{v}_j) = \mathbf{v}_i^T(A^TA)\mathbf{v}_j = \lambda(\mathbf{v}_i^T\mathbf{v}_j) = 0.$$

Hence, $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal set and hence linearly independent.

It is left to show that the given set spans $\mathrm{Col}(A)$. Since there are exactly r nonzero singular values, $A\mathbf{v}_i \neq 0$ if and only if $1 \leq i \leq r$. Let $\mathbf{y} \in \mathrm{Col}(A)$, so $\mathbf{y} = A\mathbf{x} \in \mathrm{Col}(A)$ for some $\mathbf{x} \in \mathbb{R}^n$. Then $\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ for some scalars $c_i \in \mathbb{R}$ and so

$$\mathbf{y} = A\mathbf{x} = c_1A\mathbf{v}_1 + \cdots + c_rA\mathbf{v}_r + c_{r+1}A\mathbf{v}_{r+1} + \cdots + c_n\mathbf{v}_n = c_1A\mathbf{v}_1 + \cdots + c_rA\mathbf{v}_r + 0 + \cdots + 0.$$

Thus, $\mathbf{y} \in \mathrm{Span}\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ and so $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ is an orthogonal basis for $\mathrm{Col}(A)$ and $\mathrm{rank} A = \dim \mathrm{Col}(A) = r$. $\square$

Theorem 17 (Singular Value Decomposition)**.** Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ where D is an $r \times r$ diagonal matrix for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. and there exists an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that $A = U\Sigma V^T$.

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ By Theorem 16, there exists an orthogonal basis $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_r\}$ of $\mathrm{Col}(A)$. For each i, set $\mathbf{u}_i = A\mathbf{v}_i / \|A\mathbf{v}_i\| = \frac{1}{\sigma_i}A\mathbf{v}_i$. Then $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$.

Extend $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of $\mathbb{R}^m$. Let $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}$ and $V = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$. Then both U and V are orthogonal and

$$AV = \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_r & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & 0 & \cdots & 0 \end{bmatrix} = U\Sigma.$$

Since V is orthogonal, $A = AVV^T = U\Sigma V^T$. $\square$

The columns of U in the preceding theorem are called the left singular vectors of A and the columns of V are the right singular vectors of A.

We summarize the method for SVD below. Let A be an $m \times n$ matrix of rank r. Then $A = U\Sigma V^T$ where U, Σ, V^T are as below.

(1) Find an orthonormal basis of A^TA, $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.
(2) Arrange the eigenvalues of A^TA in decreasing order. The matrix V is $\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ *in this order.*
(3) The matrix Σ is obtained by placing the r singular values along the diagonal in decreasing order.
(4) Set $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|}$ for $i = 1, \ldots, r$. Extend the orthogonal set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis of $\mathbb{R}^m$, $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ to a basis of $\mathbb{R}^m$ by adding vectors not in the set and applying Gram-Schmidt. The matrix U is $\begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix}$.

Example 18. Consider A as in Example 14. We have already that

$$V = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}.$$

The nonzero singular values are $\sigma_1 = \sqrt{360} = 6\sqrt{10}$ and $\sigma_2 = 3\sqrt{10}$ so

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} D & 0 \end{bmatrix} = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}.$$

Now

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}.$$

Note that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is already a basis for $\mathbb{R}^2$ and so $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$. Now we check that $A = U\Sigma V^T$.

Example 19. Construct the Singular Value Decomposition of $A = \begin{bmatrix} 7 & 1 \\ 0 & 0 \\ 5 & 5 \end{bmatrix}$.

We compute $A^TA = \begin{bmatrix} 74 & 32 \\ 32 & 26 \end{bmatrix}$. The eigenvalues of A^TA are $\lambda_1 = 90$ and $\lambda_2 = 10$ (note that $\lambda_1 \geq \lambda_2 > 0$). The corresponding eigenvectors are $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$, respectively. Normalizing gives

$$\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \quad \text{so} \quad V = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}.$$

The singular values are now $\sigma_1 = \sqrt{90} = 2\sqrt{10}$ and $\sigma_2 = \sqrt{10}$. Hence,

$$D = \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} D \\ 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{10} & 0 \\ 0 & \sqrt{10} \\ 0 & 0 \end{bmatrix}.$$

A has rank 2 and

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}.$$

Choose $\mathbf{u}_3 = cv3010$ so that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis of $\mathbb{R}^3$. Then $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ and $A = U\Sigma V^T$.

Example 20. Construct the Singular Value Decomposition of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$.

We compute $A^TA = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$. The eigenvalues of A^TA are $\lambda_1 = 18$ and $\lambda_2 = 0$ with normalized eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad \text{so} \quad V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

The singular values are now $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$. Hence,

$$\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}.$$

To find $\mathbf{u}_2, \mathbf{u}_3$ such that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, we find a basis for $\mathrm{Nul}(\mathbf{u}_1^T)$. Set $\mathbf{u}_1^T\mathbf{x} = 0$. Then we have

$$\frac{1}{3}x_1 - \frac{2}{3}x_2 + \frac{2}{3}x_3 = 0,$$

which implies $x_1 - 2x_2 + 2x_3 = 0$. Hence, the parametric solution is given by

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

Set $\mathbf{w}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{w}_3 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$. By construction, $\underline{1}w_2$ and $\mathbf{w}_3$ are orthogonal to $\mathbf{u}_1$ but not to each other. We apply Gram-Schmidt. Set $\widetilde{\mathbf{u}}_21 = \mathbf{w}_2$. Then

$$\widetilde{\mathbf{u}}_3 = \mathbf{w}_3 - \mathrm{proj}_{\mathrm{Span}\{\widetilde{\mathbf{u}}_2\}}\, \widetilde{\mathbf{u}}_2 = \mathbf{w}_3 - \frac{\widetilde{\mathbf{u}}_2 \cdot \mathbf{w}_3}{\widetilde{\mathbf{u}}_2 \cdot \widetilde{\mathbf{u}}_2}\widetilde{\mathbf{u}}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} - \left(-\frac{4}{5}\right)\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 4/5 \\ 1 \end{bmatrix}.$$

Normalizing gives,

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \qquad \mathbf{u}_3 = \frac{\sqrt{5}}{3}\begin{bmatrix} -2/5 \\ 4/5 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/3\sqrt{5} \\ 4/3\sqrt{5} \\ \sqrt{5}/3 \end{bmatrix}.$$

Now $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis of $\mathbb{R}^3$. Then $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ and $A = U\Sigma V^T$.

Vector Spaces

(Last Updated: October 22, 2020)

These notes are derived primarily from *Linear Algebra and its applications* by David Lay (4ed). A few theorems have been moved around.

Throughout we will work over $\mathbb{R}$, however, everything we will do extends naturally to other fields like $\mathbb{C}$. One of the main takeaways from this section should be that that these "abstract" vector spaces really just behave like $\mathbb{R}^n$.

Definition 1. A vector space (over $\mathbb{R}$) is a nonempty set V of objects (called vectors) along with two operations: addition and multiplication by scalars in $\mathbb{R}$, subject to the following axioms.

(1) If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u} + \mathbf{v} \in V$;
(2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v} \in V$;
(3) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$;
(4) There exists an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$;
(5) For all $\mathbf{v} \in V$, there exists an element $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$;
(6) For all $\mathbf{v} \in V$ and $c \in \mathbb{R}$, $c\mathbf{v} \in V$;
(7) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in V$, $c \in \mathbb{R}$;
(8) $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$ for all $c, d \in \mathbb{R}$, $\mathbf{v} \in V$;
(9) $c(d\mathbf{v}) = (cd)\mathbf{v}$ for all $c, d \in \mathbb{R}$, $\mathbf{v} \in V$;
(10) $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$.

Example 1. The following are vector spaces under the standard operations.

(1) $\mathbb{R}^n$;
(2) $\mathbb{C}^n$;
(3) $\mathcal{M}_n$, $n \times n$ matrices (with entries in $\mathbb{R}$);
(4) $\mathbb{S}$, the space of doubly infinite sequences:

$$\{y_k\} = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots)$$

(5) For $[a, b] \subset \mathbb{R}$, the set $\mathcal{C}([a, b])$ of continuous real-valued functions with domain $[a, b]$.

Checking the vector space axioms is an easy (but tiresome) exercise in most cases. We will check one example below and accept for now that the examples above are vectors spaces. Our primary focus will be on subspaces.

Example 2. Let $\mathcal{P}_n$ denote polynomials (with coefficients in $\mathbb{R}$) of degree at most n. That is, $\mathcal{P}_n$ consists of polynomials of the form

$$p(t) = a_0 + a_1 t + \cdots + a_n t^n,$$

with $a_i \in \mathbb{R}$. If all coefficients are zero, then $p(t)$ is the *zero polynomial*[1].

Let $p(t), q(t) \in \mathcal{P}_n$. Write $p(t)$ as above and $q(t) = b_0 + b_1 t + \cdots + b_n t^n$, $b_i \in \mathbb{R}$. Then

$$(p+q)(t) = p(t) + q(t) = (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n \in \mathcal{P}_n.$$

For $c \in \mathbb{R}$,

$$(cp)(t) = cp(t) = (ca_0) + (ca_1)t + \cdots + c(a_n)t \in \mathcal{P}_n.$$

Thus, axioms (1) and (6) are satisfied by definition. Axioms (2), (3), and (4) are clear. The additive inverse of $p(t)$ is $-p(t) = (-a_0) + (-a_1)t + \cdots + (-a_n)t^n$, so axiom 5) is satisfied. Axioms (7)-(10) are also easy to check.

Definition 2. A subspace of a vector space V is a subset H of V that has three properties:

(1) $\mathbf{0} \in V \Rightarrow \mathbf{0} \in H$ (zero vector);
(2) closed under addition;
(3) closed under scalar multiplication.

Example 3. (1) The zero subspace.
(2) The space itself.
(3) Subspaces of $\mathbb{R}^n$.
(4) Polynomials of even degree are a subspace of $\mathcal{P}_n$.

Definition 3. The trace of an $n \times n$ matrix M is defined as the sum of the diagonal entries, denoted $\operatorname{tr}(M)$.

Example 4. Let $S \subset \mathcal{M}_n$ denote the set of $n \times n$ matrices of trace zero. We will show that S is a subspace of $\mathcal{M}_n$.

Clearly, the zero matrix has trace zero. Let $A, B \in S$ with $A = (a_{ij})$ and $B = (b_{ij})$. Then

$$\operatorname{tr}(A+B) = \sum_{i=1}^{n} a_{ii} + b_{ii} = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = 0 + 0 = 0.$$

Thus, S is closed under addition. Now let $c \in \mathbb{R}$, then

$$\operatorname{tr}(cA) = \sum_{i=1}^{n} (ca_{ii}) = c \sum_{i=1}^{n} a_{ii} = c \cdot 0 = 0.$$

Thus, S is closed under scalar multiplication and is therefore a subspace.

One can extend the notion of a linear transformation to abstract vector spaces.

Definition 4. A linear transformation $T : V \to W$ between vector spaces V and W is a rule that assigns to each vector $\mathbf{x} \in V$ a unique vector $T(\mathbf{x}) \in W$ such that

[1] Your text states that the degree of the zero polynomial is not not defined. Another convention is to set its degree to be $-\infty$. This is so degree behaves correctly with respect to multiplication, which is not of to us right now.

(1) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$;
(2) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{v} \in V$, $\mathbf{u} \in \mathbb{R}$.

The kernel of T is the set

$$\ker T = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$$

and the image of T is the set

$$\operatorname{im} T = \{\mathbf{w} \in W : T(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\}.$$

A one-to-one and onto linear transformation is called an isomorphism and in this case we way V and W are isomorphic.

Example 5. Let $[a, b] \subset \mathbb{R}$ be an interval and set $V = \mathcal{C}([a, b])$, the set of continuous real-valued functions on $[a, b]$. Let $D : V \to V$ be the linear transformation defined by $D(f) = f'$, the derivative of f. By elementary calculus, $D(f + g) = D(f) + D(g)$ for all $f, g \in V$ and $D(cf) = cD(f)$ for all $c \in \mathbb{R}$ and $v \in V$. Thus, the differentiation operator D is a linear transformation on V.

The kernel of D is the set of constant functions and the image is the set of all continuous functions on $[a, b]$.

Proposition 6. Let $T : V \to W$ between vector spaces V and W. Then $\ker T$ is a subspace of V and $\operatorname{im} T$ is a subspace of W.

Proof. First we will prove that $\ker T$ is a subspace of V. Note that $T(\mathbf{0})+T(\mathbf{0}) = T(\mathbf{0}+\mathbf{0}) = T(\mathbf{0})$, so $T(\mathbf{0}) = \mathbf{0}$. Thus, $\mathbf{0} \in \ker T$. Now suppose $\mathbf{v}_1, \mathbf{v}_2 \in \ker T$. Then

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

so $\mathbf{v}_1 + \mathbf{v}_2 \in \ker T$. Finally, let $\mathbf{v} \in \ker T$ and $c \in \mathbb{R}$. Then

$$T(c\mathbf{v}) = cT(\mathbf{v}) = c\mathbf{0} = \mathbf{0},$$

so $c\mathbf{v} \in V$ and so $\ker T$ is a subspace of V.

Next we will show that $\operatorname{im} T$ is a subspace of W. As above, $T(\mathbf{0}) = \mathbf{0}$, so $\mathbf{0} \in \operatorname{im} T$. Let $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{im} T$. Then there exists $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Thus,

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2,$$

so $\mathbf{w}_1 + \mathbf{w}_2 \in \operatorname{im} T$. Finally, let $\mathbf{w} \in \operatorname{im} T$ and $c \in \mathbb{R}$. Then there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$ so

$$T(c\mathbf{v}) = cT(\mathbf{v}) = c\mathbf{w}.$$

Thus, $c\mathbf{w} \in \operatorname{im} T$ and $\operatorname{im} T$ is a subspace of W. $\square$

We will be particularly interested in the coordinate mapping $T : H \to \mathbb{R}^n$. This will be used to show that abstract vector spaces are *isomorphic* to subspaces of $\mathbb{R}^n$.

Ideas of linear dependence/independence and spanning extend (with the same definitions) to abstract vector spaces.

Definition 5. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a subset of the vector space V.

The set S is linearly dependent if there exist scalars $c_1, \ldots, c_p$ not all zero such that

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

Otherwise, S is linearly independent.

The set S spans a subspace H of a vector space V if for every $\mathbf{y} \in H$ there exist $c_1, \ldots, c_p$ such that $\mathbf{y} = c_1\mathbf{v}_1 + \cdots c_p\mathbf{v}_p$. The set S is a basis of H if S is linearly independent and spans H.

Note that the span of a set $S \subset V$ is a subspace of V.

Example 7. The set $\{1, t, \ldots, t^n\}$ is a basis of $\mathcal{P}_n$.

The next result we already know for $\mathbb{R}^n$.

Theorem 8. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a basis for a vector space V and let $\mathbf{y} \in V$. There exist unique scalars $c_1, \ldots, c_p \in \mathbb{R}$ such that $c_1\mathbf{v}_1 + \cdots c_p\mathbf{v}_p = \mathbf{y}$.

Proof. Existence follows from the fact that S spans V. Suppose there exist two sets of scalars, $c_1, \ldots, c_p \in \mathbb{R}$ and $d_1, \ldots, d_p \in \mathbb{R}$ such that

$$\begin{aligned} \mathbf{y} &= c_1\mathbf{v}_1 + \cdots c_p\mathbf{v}_p \quad \text{and} \\ \mathbf{y} &= d_1\mathbf{v}_1 + \cdots d_p\mathbf{v}_p. \end{aligned}$$

Subtracting these two equations gives

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

By linear independence, $c_i = d_i$ for $i = 1, \ldots, p$. □

Definition 6. Let H be a subspace of a vector space V and let $\mathcal{B} = \mathbf{b}_1, \ldots, \mathbf{b}_p$ be a basis of H. For $\mathbf{y} \in H$, the coordinates of $\mathbf{y}$ are the *unique* scalars $c_1, \ldots, c_p$ such that $c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_p = \mathbf{y}$. The coordinate vector is

$$[\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}.$$

The map $T_{\mathcal{B}} : H \to \mathbb{R}^n$ defined by $\mathbf{y} \mapsto [\mathbf{y}]_{\mathcal{B}}$ is the coordinate mapping (determined by $\mathcal{B}$).

The next theorem was proved previously for subspaces of $\mathbb{R}^n$. The proof here is no different and so we omit it.

Theorem 9. Let H be a subspace of a vector space V and let $\mathcal{B} = \mathbf{b}_1, \ldots, \mathbf{b}_p$ be a basis of H. The coordinate mapping $T_{\mathcal{B}} : H \to \mathbb{R}^p$ is an isomorphism.

Example 10. Since $\{1, t, \ldots, t^n\}$ is a basis of $\mathcal{P}_n$, then by the theorem $\mathcal{P}_n$ is isomorphic of $\mathbb{R}^{n+1}$.

We also had the next theorem, though we do recall this proof.

Theorem 11. Let V be a (finite-dimensional) vector space with basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$. Then every basis of V has p elements.

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a set in V with $m > p$. We claim the set is linearly dependent (and hence cannot be a basis). Note that the set $\{[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_m]_{\mathcal{B}}\}$ is linearly dependent in $\mathbb{R}^p$ because $m > p$. Hence, there exist scalars $c_1, \ldots, c_m$ (not all zero) such that

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_m[\mathbf{u}_m]_{\mathcal{B}} = \mathbf{0}.$$

Because the coordinate mapping is a linear transformation, this implies

$$[c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m]_{\mathcal{B}} = \mathbf{0}.$$

Moreover, because the coordinate mapping is 1-1 this implies $c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \mathbf{0}$, so the set is linearly dependent. Hence, any basis of H can have at most p elements.

Now suppose $\mathcal{B}'$ is another basis with $k \neq p$ elements. By the above argument, $\mathcal{B}$ can have no more elements than $\mathcal{B}'$, so $p \leq k$. This implies $p = k$ so every basis has p elements. □

Note that this implies that the notion of *dimension* is well-defined for abstract vector spaces. We can extend this to subspaces of V in an obvious way, or we can simply note that any subspace of V is also a vector space and then apply that theorem.

Definition 7. The dimension of a subspace H of a vector space V is the number of vectors in a basis of H.

The next theorem is now clear.

Theorem 12. Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and $\dim H \leq \dim V$.

Theorem 13 (The Basis Theorem)**.** Let V be a p-dimensional vector subspace, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Also, any set of p elements that spans V is automatically a basis for V.