MATH 3A WEEK I
CARTESIAN SPACE

PAUL L. BAILEY

1. Introduction

Linear algebra is the algebraic study of geometric objects which are *flat*, and of the functions between them that preserve this flatness. Often we think of flat as referring to two dimensional; here, we do not mean this. We mean an idealized flatness in the dimension applicable to the object.

Around 300 B.C. in ancient Greece, Euclid set down the fundamental laws of *synthetic geometry*. He starts with *points* and *lines*. Geometric figures such as triangles and circles resided on an abstract notion of *plane*, which stretched indefinitely in two dimensions; the Greeks also analysed solids such as regular tetrahedra, which resided in *space* which stretched indefinitely in three dimensions. Let us think of points, lines, planes, and spaces as flat things of dimension 0, 1, 2, and 3 respectively.

The ancient Greeks had very little algebra, so their mathematics was performed using pictures; no *coordinate system* which gave positions to points was used as an aid in their calculations. We may refer to the uncoordinatized spaces of synthetic geometry as *affine spaces*. The word affine is used in mathematics to indicate lack of a specific preferred origin.

The notion of coordinate system arose in the *analytic geometry* of Fermat and Descartes after the European Renaissance (circa 1630). This technique connected the algebra which was florishing at the time to the ancient Greek geometric notions. We refer to coordinatized lines, planes, and spaces as *cartesian spaces*; these are composed of *ordered n-tuples* of real numbers. The set of real numbers is denoted by $\mathbb{R}$, and the set of ordered n-tuples of real numbers is denoted by $\mathbb{R}^n$. Then $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$ are sets whose geometry replicates Euclid's lines, planes, and spaces.

Since affine spaces and cartesian spaces have essentially the same geometric properties, we refer to either of these types of spaces as *euclidean spaces*.

Just as coordinatizing affine space yields a powerful technique in the understanding of geometric objects, so geometric intuition and the theorems of synthetic geometry aid in the analysis of sets of n-tuples of real numbers.

The concept of *vector* links the geometric world of Euclid to the more algebraic world of Descartes. Vectors may be defined and manipulated entirely in the geometric realm or entirely algebraically; ideally, we use the point of view that best serves our purpose. Typically, this is to understand (geometrically) or to compute (algebraically).

Date: August 17, 1998.

2. Cartesian Space

An *ordered n-tuple* of real numbers is an list $(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ are real numbers, with the defining property that

$$(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \Leftrightarrow x_1 = y_1, \ldots, x_n = y_n.$$

We define n-dimensional *cartesian space* to be the set $\mathbb{R}^n$ of ordered n-tuples of real numbers. The point $(0, \ldots, 0)$ is called the *origin*, and is labeled by O. The numbers $x_1, \ldots, x_n$ are called the *coordinates* of the point $(x_1, \ldots, x_n)$. The set of points of the form $(0, \ldots, 0, x_i, 0, \ldots, 0)$, where x_i is in the i^{th} slot, is known as the i^{th} *coordinate axis*.

In $\mathbb{R}^2$, we often use the standard variables x and y instead of x_1 and x_2. In $\mathbb{R}^3$, we often use x, y, and z instead of x_1, x_2, and x_3.

Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be points in $\mathbb{R}^n$. The *distance* between x and y is defined by

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2};$$

this formula, which is motivated by the Pythagorean Theorem, defines a function

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},$$

called the *distance function*.

3. Loci

We may consider subsets of $\mathbb{R}^n$ such that the coordinates of the points in the subset are related in some specified way. The common way of doing this is to consider *equations* with the coordinates as *variables*. The set of all points which, when their coordinates are plugged into the equation cause the equality to be true, is called the *solution set*, or *locus* of the equation.

Consider the solution set in $\mathbb{R}^3$ of the equation $z = 0$. This is the set of points of the form $(x, y, 0)$. This set is called the xy-plane, and is immediately identified with $\mathbb{R}^2$ in the natural way, via the correspondence $(x, y, 0) \leftrightarrow (x, y)$. Similarly, the solution sets of $x = 0$ and $y = 0$ are called the yz-plane and the xz-plane, respectively. Together, these sets are called *coordinate planes.*

Example 1. Find the locus in $\mathbb{R}^3$ of the equation $xyz = 0$.

Solution. If $xyz = 0$, either $x = 0$, $y = 0$, or $z = 0$. Thus the solution set is the union of the solution sets for these latter equation; that is, the locus of the equation $xyz = 0$ is the union of the coordinate planes. □

Example 2. Find an equation whose solution set in $\mathbb{R}^3$ is the union of the coordinate axes.

Solution. The x-axis is the set of points where $y = 0$ and $z = 0$. We can acheive the x-axis as the solution set of $y^2 + z^2 = 0$. Thus we can see that the solution set of

$$(x^2 + y^2)(x^2 + z^2)(y^2 + z^2) = 0$$

is the union of the coordinate axes. □

Now consider sets of points which simultaneously satisfy all of the equations in a collection of equations. Such sets are merely the intersection of the solution sets. For example, the solution set of $\{x = 0, y = 0\}$ is the z-axis.

If one of the variables is missing from an equation, its locus in $\mathbb{R}^3$ is a *curtain* (or *cylinder*), because the third variable can be anything.

Example 3. The locus in $\mathbb{R}^2$ of the equation $y = 2x + 1$ is a line, but in $\mathbb{R}^3$ it is a plane. The locus in $\mathbb{R}^3$ of the equation $z = \sin y$ is a rippled "plane"; any point of the form $(x, y, \sin y)$ is in the locus.

4. Arrows

An *arrow* is a directed line segment; it is a line segment with one end designated as its *tip* and the other as its *tail*.

Parenthetically, we note that we could be more precise here and define an arrow as a subset (A, x) of $\mathbb{R}^n \times \mathbb{R}$ such that A is a line segment and x is one of its endpoints; then call x the tail and the other endpoint the tip.

A nonzero arrow is determined by three attributes:

(1) Magnitude: its length;
(2) Direction: the line on which it sits, and its orientation on that line;
(3) Position: its tail.

A *zero arrow* is a point; it has zero magnitude and no direction.

The *inverse arrow* of an arrow $\widehat{v}$ is the arrow $-\widehat{v}$, defined to be the same line segment with the tip and tail reversed.

Let P and Q be points in $\mathbb{R}^n$ and let $\widehat{PQ}$ denote the arrow whose tail is P and whose tip is Q; this is the arrow *from P to Q*. We may add two arrows if the tip of the first equals to the tail of the second. Thus

$$\widehat{PQ} + \widehat{QR} = \widehat{PR}.$$

The arrow $\widehat{PR}$ forms the third side of a triangle.

We would like to be able to add any two arrows, but the dependence on the positioning of the arrows in our definition prevents us. Thus we eliminate this attribute of an arrow and retain the attributes of magnitude and direction; this leads us to the concept of vector.

5. Vectors

We say that two arrows are *equivalent* if they have the same magnitude and direction, but not necessarily the same position. If $\widehat{v}$ is an arrow, define

$$\vec{v} = \{\widehat{w} \mid \widehat{w} \text{ is an arrow which is equivalent to } \widehat{v}\};$$

such a set is called an *equivalence class of arrows*, or a *vector*. If $\widehat{w}$ is equivalent to $\widehat{v}$, we say that $\widehat{w}$ *represents* $\vec{v}$. Technically, $\widehat{w}$ represents $\vec{v}$ means that $\widehat{w} \in \vec{v}$. Since any arrow is equivalent to itself, we see that in particular $\widehat{v}$ represents $\vec{v}$.

If P is the tail and Q is the tip of an arrow, we write $\vec{PQ}$ for the vector represented by the arrow $\widehat{PQ}$.

We can show that $\widehat{w} \in \vec{v}$ if and only if $\vec{w} = \vec{v}$. Thus a vector is determined by two attributes:

(1) Magnitude;
(2) Direction.

All zero arrows are equivalent; thus there is a unique zero vector.

The *inverse vector* of a vector $\vec{v}$ is the vector $-\vec{v}$, defined to be the vector represented by any arrow $-\widehat{v}$, where $\widehat{v}$ represents $\vec{v}$.

For any vector $\vec{v}$ and any point $P \in \mathbb{R}^n$, there is a unique arrow $\widehat{w}$ such that $\widehat{w} \in \vec{v}$ and the tail of $\widehat{w}$ is equal to P. It is now possible to add the vectors $\vec{PQ}$ and $\vec{RS}$; let $\widehat{QT}$ be the unique arrow with the same magnitude and direction as $\widehat{RS}$, and define the geometric sum by $\vec{PQ} + \vec{RS} = \vec{PT}$. Note that $-\vec{PQ} = \vec{QP}$ and that $\vec{PQ} + \vec{QP}$ is the point P; thus adding the inverse vector produces the zero vector.

If $P = O$ is the origin, there is a unique arrow representing $\vec{v}$ whose tail is O. We refer to this arrow as the *standard position* of the vector v. The tip of the vector v in standard position is a point in $\mathbb{R}^n$; this creates a one to one correspondence between vectors (which are equivalence classes of arrows) and points in $\mathbb{R}^n$. If $\vec{v}$ is a vector, we define the coordinates of $\vec{v}$ to be the coordinates of the corresponding point; that is, the coordinates of $\vec{v}$ are the coordinates of the tip of $\vec{v}$ when its tail is placed at the origin.

This correspondence allows us to switch between the concepts of points in n-space and vectors in n-space at will, blurring the distinction. We consider points in $\mathbb{R}^n$ and vectors in $\mathbb{R}^n$ as interchangeable; the point of view we adopt depends on the situation. Thus we may use the notation $\mathbb{R}^n$ to denote the set of all vectors in n-space.

We no longer put arrows over the vectors; we will specify in each case what set an element is coming from (in particular, whether it comes from $\mathbb{R}$ or from $\mathbb{R}^n$). We may denote the zero vector (the origin) either by O, or by 0 (when it cannot be confused with the zero scalar).

6. Vector Addition and Scalar Multiplication

Let $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ be vectors in $\mathbb{R}^n$. We define the *vector sum* of these vectors algebraically by adding the corresponding coordinates:

$$v + w = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n).$$

Let $v = (v_1, v_2, \ldots, v_n)$ and let a be a real number; we often refer to real numbers as *scalars*. We define the *scalar multiplication* of a times v algebraically by multiplying each coordinate of v by a:

$$a \cdot v = (av_1, av_2, \ldots, av_n).$$

The dot is usually omitted from the notation, so $a \cdot v$ is written as av.

Primary Properties of Vector Addition and Scalar Multiplication
Let $x, y, z \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Then

- **(a)** $x + y = y + x$; (Commutativity)
- **(b)** $(x + y) + z = x + (y + z)$; (Associativity)
- **(c)** $x + 0 = x$; (Existence of an Additive Identity)
- **(d)** $x + (-x) = 0$; (Existence of Additive Inverses)
- **(e)** $1 \cdot x = x$; (Scalar Identity)
- **(f)** $(ab)x = a(bx)$; (Scalar Associativity)
- **(g)** $a(x + y) = ax + ay$; (Distributivity of Scalar Mult over Vector Add)
- **(h)** $(a + b)x = ax + bx$. (Distributivity of Scalar Mult over Scalar Add)

Remark. These properties are derived directly from the definition. □

Secondary Properties of Vector Addition and Scalar Multiplication
Let $x, y, z \in \mathbb{R}^n$ and let $a, b \in \mathbb{R}$. Let $O \in \mathbb{R}^n$ be the origin. Then

- **(a)** $0 \cdot x = O$;
- **(b)** $a \cdot O = O$;
- **(c)** $-1 \cdot x = -x$;
- **(d)** $(-a)x = -(ax)$.

Remark. These properties may be derived from the primary properties. □

Geometrically, the vector sum $v + w$ corresponds to sliding an arrow representing w over so that its tail is equal to the tip of v. That is, there is a unique arrow which represents the vector w whose tail equals the tip of the vector v. We interpret $v + w$ geometrically to be the tip of this arrow. It is the endpoint of the diagonal of the parallelogram determined by v and w.

Geometrically, the scalar multiple av is interpreted as the vector whose direction is that of v but whose length is $|a||v|$. If $a < 0$, then the orientation of av is opposite the orientation of v. Thus multiplying a vector by negative one reverses its orientation, and produces its negative.

The vector which proceeds from the tip of v to the tip of w is $w - v$. This is clear, since $v + (w - v) = w$.

7. Norm of a Vector

The *norm* of a vector is its magnitude; if we pick an arrow representing the vector whose tail is at the origin, then its norm is the distance between its tip and the origin. Thus if $x = (x_1, \ldots, x_n)$, the norm of x is denoted $|x|$ and is given by

$$|x| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

A *unit vector* is a vector whose norm is 1. In some sense, a unit vector represents pure direction (without length); if u is a unit vector and a is a scalar, then au is a vector in the direction of u with norm a.

Let v be any nonzero vector. We obtain a unit vector in the direction of v simply by dividing by the length of v. Thus the *unitization* of v is

$$u = \frac{1}{|v|} v.$$

8. Dot Product

Let $v = (v_1, v_2, \ldots, v_n)$ and $w = (w_1, w_2, \ldots, w_n)$ be vectors in $\mathbb{R}^n$. We define the *dot product* of v and w by the rule

$$v \cdot w = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

There is no ambiguity caused by using a dot for scalar multiplication and vector dot product, because their definitions agree in the only case where there is overlap (namely, if $n = 1$). We usually drop the dot from the notation for scalar multiplication anyway (unless the vector is a known constant). Note that $v + w \in \mathbb{R}^n$ and $av \in \mathbb{R}^n$, but $v \cdot w \in \mathbb{R}$.

Properties of Dot Product and Norm
Let $x, y, z \in \mathbb{R}^n$ and $a \in \mathbb{R}$. Then

- **(a)** $x \cdot x = |x|^2$;
- **(b)** $x \cdot y = y \cdot x$; (Commutativity)
- **(c)** $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$; (Distributivity over Vector Addition)
- **(d)** $a(x \cdot y) = (ax) \cdot y = x \cdot (ay)$;
- **(e)** $x \cdot O = 0$;
- **(f)** $|ax| = |a||x|$.

Remark. These properties are proved directly from the algebraic definitions. □

Remark. Properties **(a)** through **(f)** are derived directly from the algebraic definitions. Properties **(c)** and **(d)** together are called *linearity of dot product.* □

The geometric interpretation of dot product is as useful as it is unanticipated from the definition. To understand it, we first need to understand the concept of projection.

Given a line L in $\mathbb{R}^n$ and a point P in $\mathbb{R}^n$ not on the line, there is a unique point Q on the line which is closest to the point. The lines between L and $\overline{PQ}$ are perpendicular. The point Q is the *projection* of P onto L.

Let v and w be vectors in $\mathbb{R}^n$. There is a unique point on the line through w which is the projection of the tip of v onto this line. The vector whose tail is the origin and whose tip is this projected point is called the *vector projection* of v onto w. The norm of this vector projection is the distance from the origin to this projected point and is called the *scalar projection* of v onto w. Let $\mathrm{proj}_w(v)$ denote the scalar projection of v onto w.

Drop a perpendicular from the tip of v onto the line through w to obtain a right triangle. If θ is the angle between the vectors v and w, we see that $\mathrm{proj}_w(v) = |v| \cos\theta$.

Proposition 1. *Let $v, w \in \mathbb{R}^n$ and let θ be the angle between v and w. Then*

$$v \cdot w = |v||w| \cos\theta.$$

Proof. To prove this result, we use the Law of Cosines, a generalization of the Pythagorean Theorem. The Law of Cosines states that for any triangle whose sides have lengths a, b, and c and whose angle opposite the side of length c has angle θ, then $c^2 = a^2 + b^2 - 2ab\cos\theta$.

To use this, consider the triangle whose vertices are the tips of v and w. The vector from v to w is $w - v$, so the lengths of the sides of this triangle are $|v|$, $|w|$, and $|w - v|$. The Law of Cosines now gives us

$$|w - v|^2 = |v|^2 + |w|^2 - 2|v||w|\cos\theta.$$

Since the square of the modulus of a vector is its dot product with itself, we have

$$(w - v) \cdot (w - v) = v \cdot v + w \cdot w - 2|v||w|\cos\theta.$$

By distributativity of dot product over vector addition and other properties,

$$w \cdot w - 2v \cdot w + v \cdot v = v \cdot v + w \cdot w - 2|v||w|\cos\theta.$$

Cancelling like terms on both sides and then dividing by -2 yields

$$v \cdot w = |v||w|\cos\theta.$$

$\square$

Corollary 1. *Let $v, w \in \mathbb{R}^n$ and let θ be the angle between v and w. Then*

$$v \cdot w = |w|\mathrm{proj}_w(v).$$

If u is of unit length, then

$$v \cdot u = \mathrm{proj}_w(v).$$

We say that v is *orthogonal* (or *perpendicular*) to w, and write $v \perp w$, if the angle θ between them is a right angle. This happens exactly when the cosine of this angle is zero: $\cos\theta = 0$. Also, by the definition of projection, this happens exactly when the vector projection of v onto w is the zero vector.

Dot product gives us a test for perpendicularity:

$$v \perp w \Leftrightarrow v \cdot w = 0.$$

Note that from this point of view, any vector is perpendicular to the zero vector.

9. Lines in $\mathbb{R}^n$

A line in $\mathbb{R}^n$ is determined by a point Q on the line and a direction vector v; the points on the line are those we encounter if we proceed from Q in the direction of v. Each such point is of the form $Q + tv$, where we think of the real number t as being the time spent traveling in that direction. Thus the line is the set of points P of the form

$$P = tv + Q.$$

Note that the distance between P and Q is equal to $|t||v|$; we may think of $|v|$ as the velocity with which we proceed away from the point Q.

The equation $P = tv + Q$ is a *parametric equation*; here we have a parameter t which is allowed to range throughout the entire set of real numbers. The line itself is not the locus of this equation; it is the set

$$L = \{P \in \mathbb{R}^n \mid P = tv + Q \text{ for some } t \in \mathbb{R}\}.$$

10. Planes in $\mathbb{R}^3$

A plane in $\mathbb{R}^3$ is determined by a point Q on the plane and a normal vector n; here, "normal" means "perpendicular to the plane". That is, given a plane (in $\mathbb{R}^3$), there is a unique direction which is perpendicular to the plane. If P is another point on the plane, then the vector from Q to P lies on the plane. This vector is $P - Q$, which is perpendicular to the normal vector n, so its dot product with n is zero. This gives the vector equation of the plane to be

$$(P - Q) \cdot n = 0.$$

By properties of dot product, we may rewrite this as

$$P \cdot n = Q \cdot n.$$

The plane is the locus of this equation; thus it is the set

$$L = \{P \in \mathbb{R}^n \mid (P - Q) \cdot n = 0\}.$$

Suppose $Q = (x_0, y_0, z_0)$ is the fixed point, $n = (n_1, n_2, n_3)$ is the normal vector and $P = (x, y, z)$ is the variable point. In this case, our equation becomes

$$(x, y, z) \cdot (n_1, n_2, n_3) = (x_0, y_0, z_0) \cdot (n_1, n_2, n_3),$$

or

$$n_1 x + n_2 y + n_3 z = n_1 x_0 + n_2 y_0 + n_3 z_0.$$

We know that three points in $\mathbb{R}^3$ determine a plane, which leads us to an example.

Example 4. Let $Q = (0, 1, 4)$, $R = (1, 0, 3)$, and $S = (-2, 6, 0)$ be three points in $\mathbb{R}^3$. Find the equation of the plane which passes through these points.

Solution. Let $v = R - Q = (1, -1, -1)$ and $w = S - Q = (-2, 5, -4)$. There is an entire plane's worth of vectors which are perpendicular to v, and a different plane's worth of vectors which are perpendicular to w; there intersection is a line which is perpendicular to both. Let $n = (n_1, n_2, n_3)$ be a direction vector for this line. Then n is a normal vector for the plane we seek. Note that any nonzero vector along this line is a normal vector, so we anticipate some choice in our eventual solution.

Now n is perpendicular to both v and w, so

$$v \cdot n = 0 \quad \text{and } w \cdot n = 0.$$

Multiplying this out and thinking of the n_i's as variables, this gives two equations in three variables:

$$\begin{aligned} n_1 - n_2 - n_3 &= 0 \\ -2n_1 + 5n_2 - 4n_3 &= 0 \end{aligned}$$

Multiply the first equation by 2, add the resulting equations, and simplify to see that $n_2 = 2n_3$. Plug this into the first equation and simplify to get $n_1 = 3n_3$.

Thus any vector of the form $n = (3n_3, 2n_3, n_3)$ is a normal vector. Set $n_3 = 1$ to get $n = (3, 2, 1)$. The equation of the plane in $P \cdot n = Q \cdot n$, or

$$3x + 2y + z = 6.$$

□

11. Hyperplanes in $\mathbb{R}^n$

The construction of a plane in $\mathbb{R}^3$ is easily generalized to any dimension. For example, the set of all points in $\mathbb{R}^2$ perpendicular to a given vector is a line.

Example 5. Find the line in $\mathbb{R}^2$ passing through the point $Q = (1, 5)$ in the direction of the vector $v = (3, 7)$.

Solution. We seek an equation for which this line is the locus. Although there are others ways to do this, we demonstrate that the above idea suffices.

First we want a vector $n = (n_1, n_2)$ which is perpendicular to the line. Then $v \perp n$, so $v \cdot n = 0$, i.e., $3n_1 + 7n_2 = 0$. Let $n = (7, -3)$, which is solution to the above equation.

The line is given by $(P - Q) \cdot n = 0$. If $P = (x, y)$ and $Q = (1, 5)$, we have $7x - 3y = 7 - 35 = -28$. So $7x - 3y = -28$ is an equation of the line. □

Let us define a *hyperplane* in $\mathbb{R}^n$ to be the set of all points perpendicular to a given vector x and passing through a given point Q. If $H \subset \mathbb{R}^n$ is such a hyperplane, then

$$H = \{P \in \mathbb{R}^n \mid (P - Q) \cdot x = 0\}.$$

A hyperplane in $\mathbb{R}$ is a point; a hyperplane in $\mathbb{R}^2$ is a line, and a hyperplane in $\mathbb{R}^3$ is a plane in the standard sense. In general, a hyperplane in $\mathbb{R}^n$ is geometrically identical to a copy of $\mathbb{R}^{n-1}$ embedded in $\mathbb{R}^n$.

There are two types of hyperplanes; those that pass through the origin and those that do not. We will see that hyperplanes which pass through the origin have the additional property that they are "closed" under vector addition and scalar multiplication. In this way, they are both geometrically and algebraically identical to a copy of $\mathbb{R}^{n-1}$.

12. Linear Combinations and Spans

Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$ be a finite set of vectors from $\mathbb{R}^n$. A *linear combination* of the vectors in A is an element of $\mathbb{R}^n$ of the form

$$a_1 v_1 + \cdots + a_r v_r,$$

where $a_1, \ldots, a_r \in \mathbb{R}$. We may also call this a *linear combination from A*. We do not place any restrictions in our definitions regarding the relative size of r and n; however, this relative size will play a role in what we will be able to conclude.

The *span* of A is the subset $\mathrm{span}(A) \subset \mathbb{R}^n$ defined by

$$\mathrm{span}(A) = \{w \in \mathbb{R}^n \mid w \text{ is a linear combination from } A\}.$$

Let $X \subset \mathbb{R}^n$ be an arbitrary subset, not necessarily finite. Then define the span of X to be the union of all spans of finite subsets of X:

$$\mathrm{span}(X) = \{a_1 v_1 + \cdots + a_r v_r \mid a_i \in \mathbb{R} \text{ and } v_i \in X \text{ for } i = 1, \ldots, r\}.$$

The span of X is the set of all finite linear combinations of vectors in X; we do not have a definition for a linear combination of an infinite number of vectors (one could try to use limits here to get a definition in some cases).

Proposition 2. *Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$. Then*

(a) $A \subset \mathrm{span}(A)$*;*
(b) $B \subset A \Rightarrow \mathrm{span}(B) \subset \mathrm{span}(A)$*;*
(c) $X \subset \mathrm{span}(A) \Rightarrow \mathrm{span}(X) \subset \mathrm{span}(A)$*.*

Proof. Since the vector v_i is a linear combination of the vectors in A simply by taking $a_i = 1$ and $a_j = 0$ for $i \neq j$, we get **(a)**.

In light of this, **(b)** follows from **(c)**, so we prove **(c)**. Suppose that $X \subset \mathrm{span}(A)$. Let $B = \{w_1, \ldots, w_s\} \subset X$ be a finite subset. It suffices to show that $\mathrm{span}(B) \subset \mathrm{span}(A)$. Pick an arbitrary vector $w \in \mathrm{span}(B)$; it suffices to show that $w \in \mathrm{span}(A)$.

Now $w = \sum_{j=1}^{s} b_j w_j$ for some real numbers b_j. Also, each vector w_j is a linear combination of the v_i, that is, $w_j = \sum_{i=1}^{r} a_{ij} v_i$ for some real numbers a_{ij}. Thus

$$w = \sum_{j=1}^{s} b_j w_j = \sum_{j=1}^{s} b_j (\sum_{i=1}^{r} a_{ij} v_i) = \sum_{j=1}^{s} \sum_{i=1}^{r} b_j a_{ij} v_i = \sum_{i=1}^{r} (\sum_{j=1}^{s} a_{ij} b_j) v_i$$

We have expressed w as a linear combination of the v_is, thus $w \in \mathrm{span}(A)$. $\square$

Proposition 3. *Let $A = \{v_1, \ldots, v_r\} \subset \mathbb{R}^n$. Let $x, y \in \mathrm{span}(A)$ and let L be the line through x and y. Then $L \subset \mathrm{span}(A)$.*

Exercise Hint. Pick an arbitrary point on the line. It suffices to show that this point is in $\mathrm{span}(A)$. First show that the point is in $\mathrm{span}\{x, y\}$. $\square$

Remark 1. The two propositions above remain true if A is replaced by an infinite subset of $\mathbb{R}^n$.

13. Subspaces

A subset $W \subset \mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ if

(S0) W is nonempty;
(S1) $x, y \in W \Rightarrow x + y \in W$;
(S2) $a \in \mathbb{R}, x \in X \Rightarrow ax \in W$.

If W is a subspace of $\mathbb{R}^n$, this fact is denoted by $W \leq \mathbb{R}^n$.

Property **(S1)** says that W is *closed under vector addition*, and property **(S2)** says that W is *closed under scalar multiplication*. In the presence of these properties, property **(S0)** is equivalent to the assertion that the origin is an element of W. For if $0 \in W$, then W is certainly nonempty; on the other hand, suppose that W is nonempty and let $w \in W$. Then $-1w = -w \in W$ by property **(S2)**, so $0 = w + (-w) \in W$ by property **(S1)**.

Example 6. The set $\{0\}$, which contains only the origin, is a subspace, called the *trivial* subspace. Also, $\mathbb{R}^n$ is a subspace of itself.

Example 7. Let $v, w \in \mathbb{R}^3$ and let $W = \operatorname{span}(v, w) = \{av + bw \mid a, b \in \mathbb{R}\}$. Then W is a subspace of $\mathbb{R}^3$. To see this, first note that $0 = 0v + 0w \in W$, so **(S0)** is satisfied. Next select arbitrary vectors $a_1v + b_1w$ and $a_2v + b_2w$ from V and note that their sum is $(a_1 + a_2)v + (b_1 + b_2)w$, which is also in W; thus **(S1)** is satisfied. Moreover, if $av + bw \in W$ and $c \in \mathbb{R}$, we have $cav + cbw \in W$; thus **(S2)** is satisfied.

The subspace W is a plane through the origin in $\mathbb{R}^3$.

Proposition 4. *Let* $A \subset \mathbb{R}^n$. *Then* $\operatorname{span}(A) \leq \mathbb{R}^n$.

Reason. Sums and scalar products of linear combinations from A are linear combinations from A. □

Proposition 5. *Let* $X \subset \mathbb{R}^n$. *Then* $X \leq \mathbb{R}^n$ *if and only if* $\operatorname{span}(X) = X$.

Reason. Suppose X is a subspace of $\mathbb{R}^n$. We wish to show that $\operatorname{span}(X) = X$. Since we already know that $X \subset \operatorname{span}(X)$, it suffices to show that $\operatorname{span}(X) \subset X$. Let $w \in \operatorname{span}(X)$. It suffices to show that $w \in X$. Now w is a finite linear combination of vectors from X. Since X is a subspace, it is closed under addition and scalar multiplication, so all sums and scalar multiples of vectors in X are also in X. Thus linear combinations of vectors from X are also in X; thus $w \in X$.

Suppose that $\operatorname{span}(X) = X$. Let $x, y \in X$ and $a \in \mathbb{R}$. Then $x + y$ is a linear combination of vectors from X, so $x + y \in \operatorname{span}(X) = X$. Also ax is a linear combination of vectors from X, so $ax \in \operatorname{span}(X) = X$. Thus X is closed under vector addition and scalar multiplication, i.e., X is a subspace of $\mathbb{R}^n$. □

14. Bases

Let W be a subspace of $\mathbb{R}^n$. A *basis* for W is a subset $B \subset W$ such that

(B1) $\mathrm{span}(B) = W$;

(B2) $C \subsetneq B \Rightarrow \mathrm{span}(C) \subsetneq \mathrm{span}(B)$.

Together, these properties state that B is a *minimal* spanning set. Later, we will show that every subspace has a basis, and that all bases have the same number of elements; we will call this number the *dimension* of the subspace.

Example 8. Let $v = (1, 1)$ and $w = (1, -1)$. Then $\{v, w\}$ is a basis for $\mathbb{R}^2$. Indeed, let $p = (x, y) \in \mathbb{R}^2$ be an arbitrary point; we wish to write p as a linear combination of v and w. This means that we wish to find real numbers $a, b \in \mathbb{R}$ such that $p = av + bw$, or $(x, y) = (a, a) + (b, -b)$. This leads to a pair of equations $x = a + b$ and $y = a - b$. Manipulate these to get $a = \frac{1}{2}(x + y)$ and $b = \frac{1}{2}(x - y)$. We have found a and b in terms of the coordinates of the point p, which shows that $p \in \mathrm{span}(v, w)$.

Neither v nor w span $\mathbb{R}^2$ by themselves, so $\{v, w\}$ is a minimal spanning set, so it is a basis.

Proposition 6. *Let $B \subset \mathbb{R}^n$. Then B is a basis for $\mathbb{R}^n$ if and only if every vector in $\mathbb{R}^n$ can be written as a linear combination from B in a unique way.*

Remark. We wish to show that the "minimality" property can be exchanged for the "uniqueness" property. We will show this later; think about why it is true. $\square$

The i^{th} *standard basis vector* for $\mathbb{R}^n$ is denoted e_i and is defined to be the vector with 1 in the i^{th} coordinate and zero in every other coordinate. The set of all such vectors is called the *standard basis* for $\mathbb{R}^n$.

For example, the standard basis for $\mathbb{R}^4$ is

$$\{e_1, e_2, e_3, e_4\} = \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)\}.$$

This is indeed a basis; for example, we can write

$$(1, -3, \pi, \sqrt{2}) = e_1 - 3e_2 + \pi e_3 + \sqrt{2}e_4.$$

15. Linear Transformations

A *linear transformation* from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a function

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

which satisfies

(L1) $T(v + w) = T(v) + T(w)$ for all $v, w \in \mathbb{R}^n$;
(L2) $T(av) = aT(v)$ for all $v \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

Example 9. The function $P_i : \mathbb{R}^n \to \mathbb{R}$ given by $T(x_1, \dots, x_n) = x_i$ is linear; this is called *projection* onto the i^{th} coordinate.

Example 10. Let $a, b \in \mathbb{R}$ be arbitrary constants. The function $T : \mathbb{R}^2 \to \mathbb{R}^1$ given by $T(x, y) = ax + by$ is linear. To see this, let $v = (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$. Then $v + w = (x_1 + x_2, y_1 + y_2)$, so

$$\begin{aligned} T(v + w) &= T((x_1 + x_2, y_1 + y_2) \\ &= a(x_1 + x_2) + b(y_1 + y_2) \\ &= (ax_1 + by_1) + (ax_2 + by_2) \\ &= T(v) + T(w). \end{aligned}$$

Now let $v = (x, y) \in \mathbb{R}^2$ and $c \in \mathbb{R}$; then

$$T(cv) = T(cx, cy) = acx + acy = c(ax + by) = cT(v).$$

Thus T is linear.

Example 11. Fix an arbitrary vector $w \in \mathbb{R}^n$. Then the function $T : \mathbb{R}^n \to \mathbb{R}$ given by $T(v) = v \cdot w$ is linear.

Proposition 7. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then*

(a) $T(0) = 0$*;*
(b) $T(\operatorname{span}(A)) = \operatorname{span}(T(A))$*, where $A \subset \mathbb{R}^n$.*

Proof. Let O_n denote the origin in $\mathbb{R}^n$ and let O_m denote the origin in $\mathbb{R}^m$, to distinguish them from the 0 scalar. Then $T(O_n) = T(0 \cdot O_n) = 0 \cdot T(O_n) = O_m$, since 0 times anything in $\mathbb{R}^m$ is O_m.

Let $A \subset \mathbb{R}^n$; for simplicity assume that $A = \{v_1, \dots, v_r\}$ is a finite set. Then

$$\begin{aligned} T(\operatorname{span}(A)) &= T(\{\sum_{i=1}^{r} a_i v_i \mid a_i \in \mathbb{R}\}) && \text{by definition of span} \\ &= \{T(\sum_{i=1}^{r} a_i v_i) \mid a_i \in \mathbb{R}\} && \text{by definition of image} \\ &= \{\sum_{i=1}^{r} a_i T(v_i) \mid a_i \in \mathbb{R}\} && \text{since } T \text{ is linear} \\ &= \operatorname{span}(\{T(v_1), \dots, T(v_r)\}) && \text{by definition of span} \\ &= \operatorname{span}(T(A)) && \text{by definition of image} \end{aligned}$$

$\square$

Proposition 8. *A linear transformation is completely determined by its effect on the standard basis.*

Proof. This means that if we know the effect of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ on the standard basis, then we know its effect on all of $\mathbb{R}^n$. This follows from the fact that if $v \in \mathbb{R}^n$, then $v = (a_1, \ldots, a_n)$ for some real numbers $a_i \in \mathbb{R}$. This is the same as saying that $v = \sum_{i=1}^n a_i e_i$; but since T is linear, we have

$$\begin{aligned} T(v) &= T(\sum_{i=1}^n a_i e_i) \\ &= \sum_{i=1}^n T(a_i e_i) \\ &= \sum_{i=1}^n a_i T(e_i). \end{aligned}$$

□

Remark 2. The above argument shows that every vector in the image of a linear transformation is a linear combination of the images of the basis vectors.

Remark 3. The above argument proceeds without change if we replace the standard basis by any spanning set.

Proposition 9. *Let $w_1, \ldots, w_n \in \mathbb{R}^m$. Then there exists a unique linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that $T(e_i) = w_i$ for $i = 1, \ldots, n$.*

Proof. Define T by $T(v) = \sum_{i=1}^n a_i T(e_i)$, where $v = (a_1, \ldots, a_n)$. This is linear and sends e_i to the vector w_i. It is unique by the previous proposition. □

Remark 4. The above argument proceeds without change if we replace the standard basis by any finite spanning set.

Example 12. Define a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(e_1) = (1, 2, 0)$, $T(e_2) = (0, 1, 2)$, and $T(e_3) = (2, 0, 1)$. Let $v = (1, 2, 3)$. What is $T(v)$?

Solution. Note that $v = e_1 + 2e_2 + 3e_3$. Thus

$$T(v) = T(e_1) + 2T(e_2) + 3T(e_3) = (1, 2, 0) + (0, 2, 4) + (6, 0, 3) = (1, 4, 7).$$

□

16. Images and Preimages under Linear Transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let V be a subspace of $\mathbb{R}^n$. The *image* of V under T is denoted by $T(V)$ and is defined to be the set of all vectors in $\mathbb{R}^m$ which are "hit" by an element of V under the transformation T:

$$T(V) = \{w \in \mathbb{R}^m \mid w = T(v) \text{ for some } v \in V\}.$$

Then $T(V)$ is actually a subspace of $\mathbb{R}^m$.

Proposition 10. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $V \leq \mathbb{R}^n$. Then $T(V) \leq \mathbb{R}^m$.*

Proof. In order to show that something is a subspace, we need to verify properties **(S0)**, **(S1)**, and **(S2)**.

(S0) Since $0 \in V$ and $T(0) = 0$, we see that $0 \in T(V)$.

(S1) Let $w_1, w_2 \in T(V)$. Then there exist vectors $v_1, v_2 \in V$ such that $w_1 = T(v_1)$ and $w_2 = T(v_2)$. We have $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$. Since V is a subspace, $v_1 + v_2 \in V$; thus $w_1 + w_2 \in T(U)$.

(S2) Let $w \in T(V)$ and $a \in \mathbb{R}$. Then there exists $v \in V$ such that $T(v) = w$. We have $aw = aT(v) = T(av)$. Since V is a subspace, $av \in U$; thus $aw \in T(V)$. □

Example 13. Let V be the subspace of $\mathbb{R}^2$ spanned by the vector $v = (1, 1)$; that is, $V = \{(t, t) \mid t \in \mathbb{R}\}$ is a line through the origin of slope 1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x, y) = (x + y, x - y)$; this is linear. Then $T(V)$ is the subspace of $\mathbb{R}^2$ spanned by $T(v) = (1 + 1, 1 - 1) = (2, 0)$; that is, $T(V)$ is the x-axis. Thus T rotates V by $-\frac{\pi}{6}$ degrees and expands it by a factor of $\sqrt{2}$. In fact, this is the effect of T on the entire plane.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let W be a subspace of $\mathbb{R}^m$. The *preimage* of W under T is denoted by $T^{-1}(W)$ and is defined to be the set of all vectors in $\mathbb{R}^n$ which "hit" elements in W under the transformation T:

$$T^{-1}(W) = \{v \in \mathbb{R}^n \mid T(v) = w \text{ for some } w \in W\}.$$

Then $T^{-1}(W)$ is actually a subspace of $\mathbb{R}^n$.

Proposition 11. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $W \leq \mathbb{R}^m$. Then $T^{-1}(W) \leq \mathbb{R}^n$.*

Proof. We verify properties **(S0)**, **(S1)**, and **(S2)**.

(S0) Since $0 \in W$ and $T(0) = 0$, we see that $0 \in T^{-1}(W)$.

(S1) Let $v_1, v_2 \in T^{-1}(W)$; then $T(v_1)$ and $T(v_2)$ are elements of W. Now $T(v_1 + v_2) = T(v_1) + T(v_2)$, and since W is a subspace, this sum is also in W. Thus $v_1 + v_2 \in T^{-1}(W)$.

(S2) Let $v \in T^{-1}(W)$ and $a \in \mathbb{R}$. Then $T(av) = aT(v)$; since $T(v)$ is in W and W is a subspace, $aT(v) \in W$. Thus $av \in T^{-1}(W)$. □

Example 14. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation given by $T(x, y) = (x - y, y - z, z - x)$. Let $W = \{0\} \subset \mathbb{R}^m$ be the trivial subspace of $\mathbb{R}^m$; here, 0 means the point $(0, 0, 0)$. The preimage is given by solving the equations

$$x - y = 0; \quad y - z = 0; \quad z - x = 0.$$

Any point of the form (t, t, t), where $t \in \mathbb{R}$, is a solution. Thus $T^{-1}(W)$ is the line in $\mathbb{R}^3$ spanned by the vector $(1, 1, 1)$.

17. Kernels of Linear Transformations

The *kernel* of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the set of all vectors in the domain $\mathbb{R}^n$ which are sent to the origin in the range $\mathbb{R}^m$. We denote this set by $\ker(T)$:

$$\ker(T) = \{v \in \mathbb{R}^n \mid T(v) = 0\}.$$

Proposition 12. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $\ker(T) \leq \mathbb{R}^n$.*

Proof. We verify properties **(S0)**, **(S1)**, and **(S2)**.

(S0) We know that $T(0) = 0$; thus $0 \in \ker(T)$.

(S1) Let $v_1, v_2 \in \ker(T)$; this means that $T(v_1) = T(v_2) = 0$. Then $T(v_1 + v_2) = T(v_1) + T(v_2) = 0 + 0 = 0$, so $v_1 + v_2 \in \ker(T)$.

(S2) Let $v \in \ker(T)$ and $a \in \mathbb{R}$. Then $T(av) = aT(v) = a \cdot 0 = 0$; thus $av \in \ker(T)$. $\square$

Alternate Proof. Since $W = \{0\}$ is a subspace of $\mathbb{R}^m$ and $\ker(T)$ is the preimage of W, we know that W is a subspace by a Proposition 11. $\square$

Example 15. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by $T(x, y, z) = (x, y, 0)$. This is projection onto the xy-plane, and is linear. The kernel is the z-axis.

Proposition 13. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $\ker(T) = \{0\}$ if and only if T is injective.*

Proof. We must show both sides of the implication. Recall that T is injective means that whenever $T(v_1) = T(v_2)$, we must have $v_1 = v_2$.

($\Rightarrow$) Suppose that $\ker(T) = \{0\}$. Let $v_1, v_2 \in \mathbb{R}^n$ such that $T(v_1) = T(v_2)$; we wish to show that $v_1 = v_2$. Then $T(v_1) - T(v_2) = 0$, so $T(v_1 - v_2) = 0$, so $v_1 - v_2 \in \ker(T)$. Since $\ker(T) = \{0\}$, we have $v_1 - v_2 = 0$, so $v_1 = v_2$. Therefore T is injective.

($\Leftarrow$) Suppose that T is injective. Let $v \in \ker(T)$; we wish to show that $v = 0$. But $T(v) = 0$ and $T(0) = 0$, and since T is injective, we must have $v = 0$. $\square$

If $W \leq \mathbb{R}^n$ is a subspace and $v \in \mathbb{R}^n$, the *translate* of W by v is the set

$$v + W = \{v + w \mid w \in W\}.$$

Proposition 14. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $w \in \mathbb{R}^m$ be in the image of T and let $v \in \mathbb{R}^n$ such that $T(v) = w$. Then*

$$T^{-1}(w) = v + \ker(T).$$

Proof. To show that two sets are equal, we show that each is contained in the other.

($\subset$) Let $x \in T^{-1}(w)$. Then $T(x) = w$, so $T(x) - w = 0$. Since $T(v) = w$, we have $T(x) - T(v) = T(x - v) = 0$. Thus $x - v \in \ker(T)$, so $x = v + (x - v) \in v + \ker(T)$.

($\supset$) Let $x \in v + \ker(T)$. Then $x = v + y$, where $y \in \ker(T)$. Thus $T(x) = T(v + y) = T(v) + T(y) = w + 0 = w$, so $x \in T^{-1}(w)$. $\square$

18. Sums and Scalar Products of Linear Transformations

Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations identical domains and identical ranges. We define the sum of these linear transformations to be the function $S + T$ given by adding pointwise:

$$S + T : \mathbb{R}^n \to \mathbb{R}^m \quad \text{given by} \quad (S + T)(v) = S(v) + T(v).$$

Proposition 15. *Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Then $S + T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

Proof. We verify properties **(L1)** and **(L2)**.

(L1) Let $v_1, v_2 \in \mathbb{R}^n$. Then

$$\begin{aligned}(S + T)(v_1 + v_2) &= S(v_1 + v_2) + T(v_1 + v_2) \\ &= S(v_1) + S(v_2) + T(v_1) + T(v_2) \\ &= S(v_1) + T(v_1) + S(v_2) + T(v_2) \\ &= (S + T)(v_1) + (S + T)(v_2).\end{aligned}$$

(L2) Let $v \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$\begin{aligned}(S + T)(av) &= S(av) + T(av) \\ &= aS(v) + aT(v) \\ &= a(S(v) + T(v)) \\ &= a(S + T)(v).\end{aligned}$$

$\square$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $a \in \mathbb{R}$ be a scalar. We define the scalar product of b and T to be the function bT given by multiplying pointwise:

$$bT : \mathbb{R}^n \to \mathbb{R}^m \quad \text{given by } (bT)(v) = bT(v).$$

Proposition 16. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformations and let $a \in \mathbb{R}$. Then $bT : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

Proof. We verify properties **(L1)** and **(L2)**.

(L1) Let $v_1, v_2 \in \mathbb{R}^n$. Then

$$bT(v_1 + v_2) = b(T(v_1) + T(v_2)) = bT(v_1) + aT(v_2).$$

(L2) Let $v \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$bT(av) = baT(v) = abT(v) = a(bT(v)).$$

$\square$

19. Compositions of Linear Transformations

Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. The *composition* of S and T is the function

$$T \circ S : \mathbb{R}^p \to \mathbb{R}^m \quad \text{given by} \quad (T \circ S)(v) = T(S(v)).$$

Then $T \circ S$ is actually a linear transformation.

Proposition 17. *Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Then $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ is a linear transformation.*

Proof. We verify properties **(L1)** and **(L2)**.

(L1) Let $v_1, v_2 \in \mathbb{R}^p$. Then

$$T(S(v_1 + v_2)) = T(S(v_1) + S(v_2)) = T(S(v_1)) + T(S(v_2)).$$

(L2) Let $v \in \mathbb{R}^p$ and $a \in \mathbb{R}$. Then

$$T(S(av)) = T(aS(v)) = aT(S(v)).$$

□

The *identity* transformation on $\mathbb{R}^n$ is the function $J_n = J : \mathbb{R}^n \to \mathbb{R}^n$ which sends every element to itself; that is, $J(v) = v$ for all $v \in \mathbb{R}^n$. This is clearly linear.

Actually, given any arbitrary set A, we can define the identity function on it.

Let A be a set. The *identity function* on A is the function

$$\mathrm{id}_A : A \to A \quad \text{given by } \mathrm{id}_A(a) = a.$$

Let $f : A \to B$ be a function. We say that f is *invertible* if there exists a function $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$. The function g is called the *inverse* of f, and is denoted by f^{-1}.

Proposition 18. *Let $f : A \to B$. Then f is invertible if and only if f is bijective.*

Proof. To show an if and only if statement, we show implication in both directions.

($\Rightarrow$) Suppose that f is invertible. Then there exists a function $f^{-1} : B \to A$ such that $f^{-1}(f(a)) = a$ for every $a \in A$, and $f(f^{-1}(b)) = b$ for every $b \in B$.

We wish to show that f is injective and surjective.

To show injectivity, we select arbitrary elements of A which go to the same place under f and show that they must have been the same element in the first place.

Let $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$. Then $f^{-1}(f(a_1)) = f^{-1}(f(a_2))$, so $a_1 = a_2$. Therefore f is injective.

To show surjectivity, we select an arbitrary element of B and find an element $a \in A$ such that $f(a) = b$.

Let $b \in B$. Let $a = f^{-1}(b)$. Then $f(a) = f(f^{-1}(b)) = b$. Therefore f is surjective.

($\Leftarrow$) Suppose that f is bijective. The for every $b \in B$ there exists a unique element $a \in A$ such that $f(a) = b$. Define $f^{-1} : B \to A$ by $f^{-1}(b) = a$. Clearly f^{-1} is the inverse of f. □

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is called invertible if it is invertible as a function. If T is invertible, we have a function $S : \mathbb{R}^m \to \mathbb{R}^n$ such that $T \circ S = J_m$ and $S \circ T = J_n$. We will see that this implies that m must equal n. For now, we content ourselves to be reassured that if T is invertible, its inverse is also linear.

Proposition 19. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a bijective linear transformation and let $S : \mathbb{R}^m \to \mathbb{R}^n$ be its inverse. Then S is a linear transformation.*

Proof. We verify properties **(L1)** and **(L2)**.

(L1) Let $w_1, w_2 \in \mathbb{R}^m$. Since T is surjective, there exist $v_1, v_2 \in \mathbb{R}^n$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Then $S(w_1) = v_1$ and $S(w_2) = v_2$.

Now $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$, so $S(w_1 + w_2) = v_1 + v_2 = S(w_1) + S(w_2)$.

(L2) Let $w \in \mathbb{R}^m$ and $a \in \mathbb{R}$. There exists $v \in \mathbb{R}^n$ such that $T(v) = w$. Then $S(w) = v$.

Now $T(av) = aT(v) = aw$, so $S(aw) = av = aS(v)$. □

Department of Mathematics, University of California, Irvine
E-mail address: pbailey@math.uci.edu