IMAGE AND KERNEL OF A LINEAR TRANSFORMATION

MATH 196, SECTION 57 (VIPUL NAIK)

Corresponding material in the book: Section 3.1.

Executive summary

(1) For a function $f : A \to B$, we call A the domain, B the co-domain, $f(A)$ the range, and $f^{-1}(b)$, for any $b \in B$, the fiber (or inverse image or pre-image) of b. For a subset S of B, $f^{-1}(B) = \bigcup_{b \in S} f^{-1}(b)$.
(2) The sizes of fibers can be used to characterize injectivity (each fiber has size at most one), surjectivity (each fiber is non-empty), and bijectivity (each fiber has size exactly one).
(3) Composition rules: composite of injective is injective, composite of surjective is surjective, composite of bijective is bijective.
(4) If $g \circ f$ is injective, then f must be injective.
(5) If $g \circ f$ is surjective, then g must be surjective.
(6) If $g \circ f$ is bijective, then f must be injective and g must be surjective.
(7) Finding the fibers for a function of one variable can be interpreted geometrically (intersect graph with a horizontal line) or algebraically (solve an equation).
(8) For continuous functions of one variable defined on all of $\mathbb{R}$, being injective is equivalent to being increasing throughout or decreasing throughout. More in the lecture notes, sections 2.2-2.5.
(9) A vector $\vec{v}$ is termed a *linear combination* of the vectors $\vec{v_1}, \vec{v_2}, \dots, \vec{v_r}$ if there exist real numbers $a_1, a_2, \dots, a_r \in \mathbb{R}$ such that $\vec{v} = a_1\vec{v_1} + a_2\vec{v_2} + \dots + a_r\vec{v_r}$. We use the term *nontrivial* if the coefficients are not all zero.
(10) A subspace of $\mathbb{R}^n$ is a subset that contains the zero vector and is closed under addition and scalar multiplication.
(11) The span of a set of vectors is defined as the set of all vectors that can be written as linear combinations of vectors in that set. The span of any set of vectors is a subspace.
(12) A spanning set for a subspace is defined as a subset of the subspace whose span is the subspace.
(13) Adding more vectors either preserves or increases the span. If the new vectors are in the span of the previous vectors, it preserves the span, otherwise, it increases it.
(14) The kernel and image (i.e., range) of a linear transformation are respectively subspaces of the domain and co-domain. The kernel is defined as the inverse image of the zero vector.
(15) The column vectors of the matrix of a linear transformation form a spanning set for the image of that linear transformation.
(16) To find a spanning set for the kernel, we convert to rref, then find the solutions parametrically (with zero as the augmenting column) then determine the vectors whose linear combinations are being discussed. The parameters serve as the coefficients for the linear combination. There is a shortening of this method. (See the lecture notes, Section 4.4, for a simple example done the long way and the short way).
(17) The fibers for a linear transformation are translates of the kernel. Explicitly, the inverse image of a vector is either empty or is of the form (particular vector) + (arbitrary element of the kernel).
(18) The dimension of a subspace of $\mathbb{R}^n$ is defined as the minimum possible size of a spanning set for that subspace.
(19) For a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ with $n \times m$ matrix having rank r, the dimension of the kernel is $m - r$ and the dimension of the image is r. Full row rank $r = n$ means surjective (image is all of $\mathbb{R}^n$) and full column rank $r = m$ means injective (kernel is zero subspace).
(20) We can define the intersection and sum of subspaces of $\mathbb{R}^n$.

(21) The kernel of $T_1 + T_2$ contains the intersection of the kernels of T_1 and T_2. More is true (see the lecture notes).
(22) The image of $T_1 + T_2$ is contained in the sum of the images of T_1 and T_2. More is true (see the lecture notes).
(23) The dimension of the inverse image $T^{-1}(X)$ of any subspace X of $\mathbb{R}^n$ under a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ satisfies:

$$\dim(\operatorname{Ker}(T)) \le \dim(T^{-1}(X)) \le \dim(\operatorname{Ker}(T)) + \dim(X)$$

The upper bound holds if X lies inside the image of T.
(24) Please read through the lecture notes thoroughly, since the summary here is very brief and inadequate.

1. Image (range) and inverse images (fibers) for a function

1.1. **Domain, range (image), and co-domain.** Suppose $f : A \to B$ is a function (of any sort). We call A the *domain* of f and we call B the *co-domain* of f. Note that it is not necessary that every element of B occur as the image of an element of A under f. The subset of B that comprises the elements of the form $f(x), x \in A$ is termed the *range* of f or the *image* of f. Note that if the range is all of B, we say that f is *surjective*.

1.2. **Fibers of a function.** As before, $f : A \to B$ is a function. For any $b \in B$, define $f^{-1}(b)$ as the set $\{a \in A \mid f(a) = b\}$. We call $f^{-1}(b)$ the *fiber* of f corresponding to b. Other words used are *pre-image* and *inverse image*. Note the following:

- $f^{-1}(b)$ is non-empty if and only if b is in the range of f. In particular, $f^{-1}(b)$ is non-empty for *every* $b \in B$ if and only if f is surjective.
- $f^{-1}(b)$ has size *at most one* for every $b \in B$ if and only if f is injective.
- $f^{-1}(b)$ has size *exactly one* for every $b \in B$ if and only if f is bijective, i.e., invertible.

We sometimes say that a set map is *uniform* (not a standard term) if all its non-empty fibers have the same size. Note that injective set maps are uniform. However, there may exist uniform non-injective set maps. For instance, a set map where all the fibers have size two is uniform. The idea of uniformity based on cardinality works well for maps of finite sets. For infinite sets, we may use structurally dependent concepts of uniformity since cardinality (set size) is too crude a size measure in the infinite context. We will revisit this idea in the context of linear transformations.

1.3. **Composition, injectivity, surjectivity, and one-sided inverses.** We've discussed in the past that a function has a two-sided inverse (i.e., is invertible) if and only if it is bijective. We now discuss some material that will help us understand one-sided inverses.

First, a slight extension of the terminology introduced previously. Consider a function $f : A \to B$ For a subset S of B, define $f^{-1}(S) = \{a \in A \mid f(a) \in S\}$. It is clear that $f^{-1}(S) = \bigcup_{b \in S} f^{-1}(b)$. Pictorially, it is the union of the fibers over all the points in S.

Now, consider a situation where we are composing two functions:

$$f : A \to B, g : B \to C, \text{composite } g \circ f : A \to C$$

Given $c \in C$, let's try to figure out what $(g \circ f)^{-1}(c)$ would look like. We must find all $a \in A$ such that $g(f(a)) = c$. In order to do that, we first find the $b \in B$ such that $g(b) = c$. In other words, we first find $g^{-1}(c)$. Then, we find, for each $b \in g^{-1}(c)$, the $a \in A$ such that $f(a) = b$. The upshot is that we are computing $f^{-1}(g^{-1}(c))$, where the inner computation yields a subset to which we apply the outer inverse function.

Pictorially, we first take the fiber over c for g. This is a subset of B. For each element here, we take the fiber in A. Then, we take the union of the fibers to get $(g \circ f)^{-1}(a)$.

We note the following facts regarding composition, injectivity and surjectivity:

(1) Suppose $f : A \to B$ and $g : B \to C$ are both injective functions. Then, the composite function $g \circ f : A \to C$ is also an injective function. Here's the proof:
Want to show: If $a_1, a_2 \in A$ satisfy $g(f(a_1)) = g(f(a_2))$, then $a_1 = a_2$.

Proof: We "peel off the layers" one by one. Explicitly, we begin with:

$$g(f(a_1)) = g(f(a_2))$$

We use that g is injective as a function from B to C to conclude that we can cancel the g, so we obtain that:

$$f(a_1) = f(a_2)$$

We now use the injectivity of f as a function from A to B and obtain that:

$$a_1 = a_2$$

as desired.

(2) Suppose $f : A \to B$ and $g : B \to C$ are both surjective functions. Then, the composite function $g \circ f : A \to C$ is also a surjective function. The proof follows:

Want to show: For any $c \in C$, there exists $a \in A$ such that $g(f(a)) = c$.

Proof:

- Since g is surjective, the set $g^{-1}(c)$ is non-empty, i.e., there exists $b \in B$ such that $g(b) = c$.
- Since f is surjective, the set $f^{-1}(b)$ is non-empty, i.e., there exists $a \in A$ such that $f(a) = b$.
- The element a obtained this way is the one we desire: $g(f(a)) = g(b) = c$.

In both the injectivity and surjectivity cases, we use the information about the individual functions one step at a time.

(3) Suppose $f : A \to B$ and $g : B \to C$ are both bijective functions. Then, the composite function $g \circ f : A \to C$ is also a bijective function.

This follows by combining the corresponding statements for injectivity and surjectivity, i.e., points (1) and (2) above.

(4) Suppose $f : A \to B$ and $g : B \to C$ are functions such that the composite $g \circ f : A \to C$ is injective. Then, f must be injective. This is easiest to prove by contradiction: we can show that if f is not injective, then $g \circ f$ is not injective. Here's how: if f has a collision, i.e., two different elements $a_1, a_2 \in A$ satisfying $f(a_1) = f(a_2)$, then $g \circ f$ also has a collision for the same pair of inputs, i.e., we also have $g(f(a_1)) = g(f(a_2))$.

In fact, the proof shows something stronger: any collision for f remains a collision for $g \circ f$.

(5) Suppose $f : A \to B$ and $g : B \to C$ are functions such that the composite $g \circ f : A \to C$ is surjective. Then, g is surjective.

To see this, note that if everything in C is hit by the composite, it must be hit by g. Explicitly, if every $c \in C$ is of the form $g(f(a))$ for $a \in A$, then in particular it is $g(b)$ where $b = f(a)$.

In fact, the proof shows something slightly stronger: the range of $g \circ f$ is contained within the range of g.

(6) Suppose $f : A \to B$ and $g : B \to C$ are functions such that the composite function $g \circ f : A \to C$ is bijective. Then, g is surjective and f is injective. This follows from the preceding two observations ((4) and (5) in the list).

(7) Suppose $f : A \to B$ and $g : B \to A$ are functions such that the composite function $g \circ f : A \to A$ is the identity map on A. In other words, g is a left inverse to f, and f is a right inverse to g. Then, f is injective and g is surjective. This is a direct consequence of (6), along with the observation that the identity map from any set to itself is bijective.

There was a lot of confusion about points (4) and (5) of the list in class, suggesting that it is pretty hard to "get" these by yourself. So please review the reasoning carefully and convince yourself of it.

Point (7) raises two interesting converse questions:

- Suppose $f : A \to B$ is injective. Does there exist $g : B \to A$ such that $g \circ f$ is the identity map on A? The answer turns out to be *yes* for set maps. The idea is to construct g as essentially the inverse function to f on the range of f, and define it whatever way we please on the rest of B. Note that if f is injective but not surjective (so that the range of f is not all of B) then this one-sided inverse to f is non-unique, because we have flexibility regarding where we send all the elements that are in B but outside $f(A)$.

- Suppose $g : B \to A$ is surjective. Does there exist $f : A \to B$ such that $g \circ f$ is the identity map on A? The answer turns out to be *yes* for set maps. The idea is to construct f as follows: for every element $a \in A$, let $f(a)$ be any element b satisfying $g(b) = a$. We always *can* pick such an element, because g is surjective.

2. Some warm-up examples involving functions of one variable

2.1. **Real-valued functions of one variable.** Before we proceed to apply the ideas above to linear transformations, let us consider what the ideas tell us in the context of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Since you are quite familiar with single-variable precalculus and calculus, this will be helpful.

For a function f with domain a subset of $\mathbb{R}$ and co-domain $\mathbb{R}$ (so f is a real-valued function of one variable), we can find the fiber $f^{-1}(y_0)$ for a particular $y_0 \in \mathbb{R}$ in either of these ways:

- *Graphically*: We draw the line $y = y_0$, find all its points of intersection with the graph of f, and then look at the x-coordinates of those points of intersection. The set of all the x-coordinates is the inverse image (also called the pre-image or fiber) of y_0.
- *Algebraically*: We set up the equation $f(x) = y_0$ and try to solve for x.

The following turn out to be true:

- A function is injective if and only if every horizontal line intersects its graph at at most one point.
- The range of a function is the set of y-values for which the corresponding horizontal line intersects the graph.
- A function is surjective to $\mathbb{R}$ if and only if every horizontal line intersects its graph at at least one point.
- A function is bijective to $\mathbb{R}$ if and only if every horizontal line intersects its graph at *exactly* one point.

2.2. **The continuous case.** Suppose $f : \mathbb{R} \to \mathbb{R}$ is a continuous function. Note that some of the remarks we make have analogues for functions defined on intervals in $\mathbb{R}$ instead of on all of $\mathbb{R}$, but we will for simplicity consider only the case of f defined on all of $\mathbb{R}$. The following are true:

(1) f is injective if and only if it is increasing throughout $\mathbb{R}$ or decreasing throughout $\mathbb{R}$.
(2) f is surjective if and only if one of these cases is satisfied (*Note: this point was edited to include the oscillatory cases*):
  (a) $\lim_{x\to-\infty} f(x) = -\infty$ and $\lim_{x\to\infty} f(x) = \infty$. For instance, f could be any polynomial with odd degree and positive leading coefficient.
  (b) $\lim_{x\to-\infty} f(x) = \infty$ and $\lim_{x\to\infty} f(x) = -\infty$. For instance, f could be any polynomial with odd degree and negative leading coefficient.
  (c) As $x \to \infty$, $f(x)$ is oscillatory between $-\infty$ and ∞. For instance, $f(x) := e^x \sin x$.
  (d) As $x \to -\infty$, $f(x)$ is oscillatory between $-\infty$ and ∞. For instance, $f(x) := e^{-x} \sin x$.
  (e) As $x \to \infty$, $f(x)$ is oscillatory with the upper end of oscillation going to ∞, and as $x \to -\infty$, $f(x)$ is oscillatory with the lower end of oscillation going to $-\infty$. For instance, $f(x) := x \sin^2 x$.
  (f) As $x \to \infty$, $f(x)$ is oscillatory with the lower end of oscillation going to $-\infty$, and as $x \to -\infty$, $f(x)$ is oscillatory with the upper end of oscillation going to $+\infty$. For instance, $f(x) := -x \sin^2 x$.

  Note that the cases (c)-(f) are not mutually exclusive. For instance, $f(x) := x \sin x$ satisfies the conditions for cases (c), (d), (e), and (f). Note also that all the cases (c)-(f) are incompatible with injectivity, and also that these cases do not occur for polynomial functions.
(3) f is bijective if and only if one of these cases is satisfied:
  (a) f is increasing, $\lim_{x\to-\infty} f(x) = -\infty$, and $\lim_{x\to\infty} f(x) = \infty$
  (b) f is decreasing, $\lim_{x\to-\infty} f(x) = \infty$, and $\lim_{x\to\infty} f(x) = -\infty$

2.3. **Addition, injectivity, and surjectivity.** We can see from these facts that the following is true:

- The sum of two continuous injective (respectively, surjective or bijective) functions need not be injective (respectively, surjective or bijective). The problem arises if the two functions are of opposite subtypes, for instance, one is increasing and the other is decreasing. If both functions are of a similar type, then the sum is also of that type. Note that for surjective functions, the types (c)-(f) are in general problematic.

- If we take three continuous injective (respectively, bijective) functions, then at least one pair of them has the same subtype, so that sum is also of the same subtype, hence is continuous injective (respectively, bijective). In general, we cannot say more. Note that we cannot make any assertion for surjectivity per se, because of the weird types (c)-(f) for continuous surjective functions. However, if we restrict attention to continuous surjective functions of the subtypes (a) and (b), then adding two functions of the same subtype gives a function of the same subtype.

2.4. **Fibers for polynomial functions.** If $f : \mathbb{R} \to \mathbb{R}$ is a polynomial function of degree $n \geq 1$, finding the inverse image under f of any particular real number amounts to solving a particular polynomial equation of degree n. Explicitly, for an output y_0, finding $f^{-1}(y_0)$ is tantamount to solving the polynomial equation $f(x) - y_0 = 0$ which is a degree n equation in x. The polynomial equations for different points in the co-domain differ only in their constant term. The following are true:

- In the case $n = 1$ (i.e., f is a linear polynomial), every fiber has size exactly 1, since f is bijective. In fact, the inverse function to f is also a linear polynomial.
- In the case $n = 2$ (i.e., f is a quadratic polynomial), every fiber has size 0, 1, or 2. The maximum of the fiber sizes is always 2. If the leading coefficient of f is positive, then f has a unique absolute minimum value. The fiber size for that is 1, the fiber size for bigger values is 2, and the fiber size for smaller values is 0. If the leading coefficient of f is negative, then f has a unique absolute maximum value. The fiber size for that is 1, the fiber size for bigger values is 0, and the fiber size for smaller values is 2.
- In the case $n \geq 3$, the maximum of the fiber sizes is at most n. However, it is *not* necessary that there exist fibers of size n for a particular f (by Rolle's theorem, we know that for this to happen it must be the case that f' has $n - 1$ distinct roots, and this does not happen for all f). An extreme case where we have small fiber sizes is x^n. For n odd, the fiber sizes are all 1 (the function is bijective). For n even, the fiber sizes are all 0, 1, or 2.

2.5. **Fibers for periodic functions, specifically trigonometric functions.** For a periodic function with period h, the inverse image of any point is invariant under translation by h and therefore also by all integer multiples of h. In particular, the inverse image of any point, if non-empty, must be infinite.

For instance, consider the sin function from $\mathbb{R}$ to $\mathbb{R}$. The range is $[-1, 1]$. For any point in this set, the inverse image is invariant under translation by 2π and therefore also by all integer multiples of 2π.

To determine the inverse image for such a function, it makes sense to first determine the inverse image over an interval of length equal to the period, and then note that the inverse image over $\mathbb{R}$ is obtained from this set by translating by multiples of 2π. For instance, to find the inverse image of $1/2$ under sin, we note that on $[0, 2\pi]$, the only solutions to $\sin x = 1/2$ are $x = \pi/6$ and $x = 5\pi/6$. Thus, the fiber is the set:

$$\{2n\pi + \pi/6 \mid n \in \mathbb{Z}\} \cup \{2n\pi + 5\pi/6 \mid n \in \mathbb{Z}\}$$

3. Linear combinations and spans

3.1. **Linear combination.** There is a very important concept that undergirds linear algebra: *linear combination.*

Suppose $\vec{v_1}, \vec{v_2}, \dots, \vec{v_r}$ are vectors in $\mathbb{R}^n$. A vector $\vec{v}$ is termed a *linear combination* of the vectors $\vec{v_1}$, $\vec{v_2}$, $\dots$, $\vec{v_r}$ if there exist real numbers a_1, a_2, $\dots$, a_r such that:

$$\vec{v} = a_1\vec{v_1} + a_2\vec{v_2} + \cdots + a_r\vec{v_r}$$

The real numbers $a_1, a_2, \dots, a_r$ are allowed to be zero and repetitions in values are allowed.

The term "linear combination" indicates the idea of "combining" the vectors in a *linear* fashion. Recall that linearity has to do with addition and scalar multiplication. The values $a_1, a_2, \dots, a_r$ are termed the *coefficients* used in the linear combination.

Geometrically, linear combinations include scaling and addition. If we think of vectors as physical translation, then addition is how we compose the translation operations. Thus, the linear combinations of vectors are the vectors we can obtain by moving along the directions of the individual vectors one after the other. The coefficients $a_1, a_2, \dots, a_r$ describe how much we move along the direction of each vector.

We use the following terms:

- The *trivial* linear combination refers to the situation where all the coefficients are zero. The only vector you can write as a trivial linear combination is the zero vector. Thus, given any collection of vectors, the zero vector is the trivial linear combination of these.
- Any other linear combination is termed a *nontrivial* linear combination. However, a nontrivial linear combination may still have *some* coefficients equal to zero. We may use informal jargon such as "uses a vector nontrivially" to indicate that the linear combination has a nonzero coefficient for that particular vector.

Here are a couple of extreme cases:

- For the empty collection of vectors, we can still define the "trivial linear combination" to be the zero vector. This seems somewhat silly, but there's a deep logic.
- For a set of vectors of size one, the "linear combinations" possible are precisely the scalar multiples of that vector.

Here are a few easy observations:

- If a vector is a linear combination of a certain set of vectors, the given vector is also a linear combination of any bigger set of vectors: just set the coefficients on all the other vectors to be 0. For instance, if $\vec{v} = 2\vec{v_1} + 5\vec{v_2}$, then we also have $\vec{v} = 2\vec{v_1} + 5\vec{v_2} + 0\vec{v_3}$. We don't end up "using" the extra vectors, but it is still a legitimate linear combination.
- For a (possibly infinite) set S of vectors in $\mathbb{R}^n$, writing a vector $\vec{v}$ as a linear combination of S means writing $\vec{v}$ as a linear combination of a finite subset of S. We can think of that linear combination as attaching a coefficient of zero to all but finitely many vectors in $\mathbb{R}^n$.

A priori, there may be more than one way that a given vector can be expressed as a linear combination of a particular collection of vectors. We will later on discuss a concept called *linear independence* to describe a situation where no vector can be written as a linear combination in multiple ways.

3.2. **Definition of subspace.** We will study the concept of subspace more later, but a very brief definition is worthwhile right now. Consider a subset V of $\mathbb{R}^n$. We call V a *subspace* of $\mathbb{R}^n$ if it saitsfies the following three conditions:

- The zero vector is in V.
- V is closed under addition, i.e., for any vectors $\vec{v}, \vec{w}$ (possibly equal, possibly distinct) in V, the sum $\vec{v} + \vec{w}$ is also a vector in V. Note that the choice of leters for the vectors is not important. What matters is that regardless of the vectors chosen, the sum remains in the subspace.
- V is closed under scalar multiplication, i.e., if $\vec{v}$ is a vector in V and a is a real number, then $a\vec{v}$ is also a vector in V.

Two extreme examples of subspaces are:

- The zero subspace, i.e., the subspace comprising only the zero vector.
- The whole space, in this case $\mathbb{R}^n$.

Pictorially, a subspace is something flat, closed both under scaling and addition. Lines through the origin are subspaces. Planes through the origin are subspaces. There are other higher-dimensional concepts of subspaces. I'd love to say more, but that would be jumping the gun.

3.3. **Spanning sets and spans.** Consider a subset S of $\mathbb{R}^n$. For convenience, you can imagine S to be finite, though this is not necessary for the definition. The *span* of S is defined as the set of all vectors that arise as linear combinations of vectors in S. The span of S is a *subspace* of $\mathbb{R}^n$. Explicitly:

- The zero vector is in the span of any set, on account of being the trivial linear combination.
- *Addition*: If $\vec{v}$ and $\vec{w}$ are vectors in the span of S, then $\vec{v} + \vec{w}$ is also in the span of S: To see this, note that we can add up the coefficients for each vector. For instance, if we are looking at the span of three vectors $\vec{u_1}$, $\vec{u_2}$, $\vec{u_3}$, and we have:

$$\begin{aligned} \vec{v} &= 4\vec{u_1} - 7\vec{u_2} + 6\vec{u_3} \\ \vec{w} &= 11\vec{u_1} + 9\vec{u_2} + 3\vec{u_3} \end{aligned}$$

Then we get:

$$\vec{v} + \vec{w} = (4 + 11)\vec{u_1} + (-7 + 9)\vec{u_2} + (6 + 3)\vec{u_3}$$

The same idea applies in general.

Note that when we add, some of the coefficients may become zero, so the sum may end up "not using" all the vectors used in the descriptions of the individual vectors. However, *a priori*, we do not know whether any vectors will become unused after adding.

- *Scalar multiplication*: If $\vec{v}$ is in the span of S, and $a \in \mathbb{R}$, then $a\vec{v}$ is also in the span of S: Whatever linear combination we use to describe $\vec{v}$, we can multiply each of the coefficients there with a to describe $a\vec{v}$ as a linear combination of the same set.

Suppose W is a sub*space* of $\mathbb{R}^n$, i.e., it is a nonempty subset of $\mathbb{R}^n$ closed under addition and scalar multiplication. Then, a subset S of W is termed a *spanning set* for W if W is the span of S. Note that if $S_1 \subseteq S_2 \subseteq W$ with W a subspace, and S_1 is a spanning set for W, then S_2 is also a spanning set for W.

3.4. **Why do we care about spanning sets and spans?** Vector spaces such as $\mathbb{R}^n$, and their subspaces, are *infinite* (with the exception of the zero space, which has only one element, namely the zero vector). Describing a subspace of $\mathbb{R}^n$ by listing out all the elements will take infinite time. Thus, it is useful to have a way to describe subspaces of $\mathbb{R}^n$ using finite data. Finite spanning sets offer a convenient method.

Note, however, that there can be many different candidates for a spanning set for a vector space. We will later on consider a concept of *basis*, which can be defined as a *minimal* spanning set. Even if we use basis instead of spanning set, however, we do not have any form of uniqueness. While this is a drawback, we will later see ways to figure out whether two sets have the same span.

4. Image and kernel

4.1. **Image of a linear transformation is a subspace.** Suppose $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation and A is the matrix of T. The *image* of T (also called the *range* of T) is the subset of $\mathbb{R}^n$ comprising those vectors that can be written in the form $T(\vec{v})$ where $\vec{v} \in \mathbb{R}^m$. The question is: *what does the image of T look like*?

First, it is easy to see from abstract considerations that the image of T is a subspace of $\mathbb{R}^n$. We show this in three parts:

- The zero vector is in the image of any linear transformation, since it is the image of the zero vector.
- *Addition*: Suppose $\vec{u_1}$ and $\vec{u_2}$ are vectors in the image (range) of T. Our goal is to show that $\vec{u_1} + \vec{u_2}$ is in the image of T.

  Proof sketch: By the definition of what it means to be in the image, there exist vectors $\vec{v_1}, \vec{v_2} \in \mathbb{R}^m$ such that $\vec{u_1} = T(\vec{v_1})$ and $\vec{u_2} = T(\vec{v_2})$. Thus, $\vec{u_1} + \vec{u_2} = T(\vec{v_1}) + T(\vec{v_2}) = T(\vec{v_1} + \vec{v_2})$ (using the linearity of T). Thus, $\vec{u_1} + \vec{u_2}$ is in the image of T. Hence, the image of T is closed under addition.
- *Scalar multiplication*: Suppose $\vec{u}$ is a vector in the image of T and a is a real number. Our goal is to show that the vector $a\vec{u}$ is in the image of T.

  Proof sketch: By the definition of what it means to be in the image, there exists a vector $\vec{v} \in \mathbb{R}^m$ such that $T(\vec{v}) = \vec{u}$. Then, $a\vec{u} = aT(\vec{v}) = T(a\vec{v})$. In particular, $a\vec{u}$ is in the image of T.

Note that both these proofs rely on the *abstract* conception of linearity. In particular, the proofs above also work for "infinite-dimensional vector spaces" – something we shall return to after a while.

4.2. **Column vectors span the image.** The goal now is to *explicitly describe* the image as a vector space. The explicit description is as the span of a finite set of vectors, namely, the column vectors of the matrix. This explicit description crucially relies on the concrete description of the linear transformation.

Recall that the matrix A for $T : \mathbb{R}^m \to \mathbb{R}^n$ is a $n \times m$ matrix. The columns of A describe the values $T(\vec{e_1})$, $T(\vec{e_2})$, and so on, right till $T(\vec{e_m})$ (the last column). Thus, each column of A, viewed as a column vector, is in the image of T.

There are other vectors in the image of T, but these are essentially described as *linear combinations* of these column vectors. By *linear combination*, I mean a sum of scalar multiples of the given vectors (as described a little while back). Explicitly, consider a generic vector of the domain $\mathbb{R}^m$:

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{bmatrix} = x_1\vec{e_1} + x_2\vec{e_2} + \cdots + x_m\vec{e_m}$$

The image of this is:

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{bmatrix}\right) = T(x_1\vec{e_1} + x_2\vec{e_2} + \cdots + x_m\vec{e_m}) = x_1T(\vec{e_1}) + x_2T(\vec{e_2}) + \cdots + x_mT(\vec{e_m})$$

In particular, the image is a linear combination of $T(\vec{e_1})$, $T(\vec{e_2})$, $\ldots$, $T(\vec{e_m})$. Conversely, every linear combination of these vectors is in the image. In other words, the image of the linear transformation T is *precisely* the set of linear combinations of $T(\vec{e_1})$, $T(\vec{e_2})$, $\ldots$, $T(\vec{e_m})$.

As per the terminology introduced in a previous section, a set of vectors such that every vector in the image can be expressed as a linear combination of vectors in that set is said to be a *spanning set* for the image or is said to *span* the image. The upshot is that the image of a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ and that the column vectors of the matrix form a spanning set for the image. Note that its being a subspace follows both directly and from the fact that the span of any set of vectors is a subspace.

For instance, consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ with matrix:

$$\begin{bmatrix} 3 & 4 \\ 2 & 6 \\ 1 & 5 \end{bmatrix}$$

Note that this is a 3×2 matrix, as expected, since the linear transformation goes from $\mathbb{R}^2$ to $\mathbb{R}^3$. The image of $\vec{e_1}$ under the linear transformation is the column vector $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ and the image of $\vec{e_2}$ under the linear transformation is the column vector $\begin{bmatrix} 4 \\ 6 \\ 5 \end{bmatrix}$.

The image is thus the set of all vectors that can be expressed as linear combinations of these two vectors. Explicitly, it is the set of all vectors of the form (secretly, the input vector is $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$):

$$x_1 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 6 \\ 5 \end{bmatrix}, x_1, x_2 \in \mathbb{R}$$

The generic vector in the image could alternatively be written in the form:

$$\begin{bmatrix} 3x_1 + 4x_2 \\ 2x_1 + 6x_2 \\ x_1 + 5x_2 \end{bmatrix}, x_1, x_2 \in \mathbb{R}$$

We can therefore use the two column vectors as our spanning set:

$$\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 6 \\ 5 \end{bmatrix}$$

4.3. **Kernel of a linear transformation is a subspace.** The kernel of a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is the set of those vectors in the domain that get mapped to the zero vector. In other words, the kernel is the fiber (inverse image) corresponding to the zero vector.

It turns out that the kernel of a linear transformation is also a subspace:

- The zero vector is in the kernel of any linear transformation, because its image is the zero vector.
- *Addition*: We want to show that if $\vec{v}$ and $\vec{w}$ are both vectors in the kernel of T, then $\vec{v} + \vec{w}$ is also in the kernel of T. This follows from the additivity of T:

$$T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w}) = 0 + 0 = 0$$

- *Scalar multiplication*: If $\vec{v}$ is in the kernel of T and a is a real number, then $a\vec{v}$ is also in the kernel of T. This follows from the scalar multiples aspect of T:

$$T(a\vec{v}) = aT(\vec{v}) = a(0) = 0$$

Thus, the kernel is a subspace. Note that the argument used relied only on the *abstract* definition of linearity and therefore applies to vector spaces in the abstract setting, and also to infinte-dimensional spaces.

4.4. **Finding a spanning set for the kernel.** Finding the kernel is effectively the same as solving the system with the matrix as coefficient matrix and the augmenting column the zero column. We solve this the usual way: first, convert the matrix to reduced row echelon form. We can now describe the kernel the way we have done in the past: parametrically, as a solution set. Alternatively, we can try to describe the kernel using a spanning set. To describe it in the latter way, do the following: for each non-leading variable, take the case where that variable takes the value 1 and all the other non-leading variables take the value 0. Find the values of the leading variables, and construct a vector from these. Thus, for each non-leading variable, we get a corresponding vector. This set of vectors spans the kernel.

We will illustrate both the longer and shorter approach using a worked example in the next section.

4.4.1. *A worked-out example, the long way.* Suppose that the reduced row-echelon form of a matrix turns out to be:

$$\begin{bmatrix} 1 & 3 & 8 & 0 & 10 \\ 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

There are three non-leading variables: the second, third, and fifth variable. Let's first use the parametric solution description for the kernel. Suppose t is the fifth variable, u is the third variable, and v is the second variable. Then, the general solution can be written as:

$$\begin{bmatrix} -3v - 8u - 10t \\ v \\ u \\ -7t \\ t \end{bmatrix}$$

Let us separate out the multiples of t, u, and v here. We can write this as:

$$\begin{bmatrix} -3v \\ v \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -8u \\ 0 \\ u \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -10t \\ 0 \\ 0 \\ -7t \\ t \end{bmatrix}$$

If we now "pull out" the variables from their respective expressions, we obtain:

$$v \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + u \begin{bmatrix} -8 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -10 \\ 0 \\ 0 \\ -7 \\ 1 \end{bmatrix}$$

Note that t, u, v all vary freely over all of $\mathbb{R}$. Thus, the kernel is spanned by the vectors:

$$\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -8 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -10 \\ 0 \\ 0 \\ -7 \\ 1 \end{bmatrix}$$

Setting up names for the parameters, writing the general expression, and then pulling the variables out is a somewhat time-consuming process. Let's now consider a shorter way of doing the same example.

4.4.2. *The same worked example, a shorter way.*

- Our spanning set will include one vector corresponding to each non-leading variable, i.e., each column without a pivotal 1.
- The vector corresponding to a particular non-leading variable is obtained as follows: set the value of that variable to equal 1, and set the values of all the other non-leading variables to equal 0. Now, calculate the values of the *leading* variables based on the linear system. The value of a particular leading variable is simply the negative of the entry for the pivotal row of the leading variable in the column of the non-leading variable.

Let's first take the case where the second variable is 1 while the third and fifth variable are 0. In this case, the first variable is -3 and the fourth variable is 0, so we get the vector:

$$\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The -3 out there comes from the fact that the entry in the first row and second column is 3. The fourth variable is 0 because it comes after the second variable.

The vector corresponding to the third variable being 1 while the second and fifth are 0 is:

$$\begin{bmatrix} -8 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

The -8 out there comes from the fact that the entry in the first row and third column is 8. The fourth variable is 0 because it comes after the third variable.

The vector corresponding to the fifth variable being 1 while the second and third are 0 is:

$$\begin{bmatrix} -10 \\ 0 \\ 0 \\ -7 \\ 1 \end{bmatrix}$$

The first variable is -10 because the entry in the first row and fifth column is 10. The fourth variable is -7 because the entry in the *second* row (the row where the pivotal variable is the fourth variable) and fifth column is 7.

If you compare this shorter approach with the longer approach that involves writing the parameters explicitly, you should after a while be able to see just how the shorter approach is a shortening of the longer approach.

5. Fibers of a linear transformation

5.1. **Fibers are translates of the kernel.** The fibers of a linear transformation are the inverse images of specific points in the co-domain. To find the fiber (inverse image) for a particular point, we need to solve the linear system with that as the augmenting column. The solution set is either empty or is described parametrically.

Thus, the fibers over points not in the image are empty. For points that are in the image, the corresponding fiber looks like a translate of the kernel. Explicitly, any nonempty fiber looks like:

(particular solution vector) + (arbitrary element of the kernel)

This is extremely crucial to understand, so let's go over it explicitly.

Suppose $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation. We have a vector $\vec{v} \in \mathbb{R}^n$. We will show the following two things:

- If $\vec{u}$ is a vector in $\mathbb{R}^m$ such that $T(\vec{u}) = \vec{v}$, and $\vec{w}$ is in the kernel of T, then $T(\vec{u} + \vec{w}) = \vec{v}$. In other words, adding any element of the kernel to an element in the fiber keeps one within the fiber.

  Proof: We have that:

$$T(\vec{u} + \vec{w}) = T(\vec{u}) + T(\vec{w}) = \vec{v} + 0 = \vec{v}$$

- If there are two elements in the same fiber, they differ by an element in the kernel. Explicitly, if $\vec{u_1}, \vec{u_2} \in \mathbb{R}^m$ are such that $T(\vec{u_1}) = T(\vec{u_2}) = \vec{v}$, then $\vec{u_1} - \vec{u_2}$ is in the kernel of T. In other words, each fiber must involve only a *single* translate of the kernel.

  Proof: We have:

$$T(\vec{u_1} - \vec{u_2}) = T(\vec{u_1}) - T(\vec{u_2}) = \vec{v} - \vec{v} = 0$$

In other words, each non-empty fiber is a *single* translate of the kernel. For instance, imagine the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix:

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

The kernel of this linear transformation is the subspace generated by the vector $\vec{e_2}$. The image of this linear transformation is the space generated by the vector $\vec{e_2}$.

If we use x to denote the first coordinate and y to denote the second coordinate, the map can be written as:

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ x \end{bmatrix}$$

The kernel is the y-axis and the image is also the y-axis.

Now, consider a vector, say:

$$\vec{v} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$$

We know that the vector is in the image of the linear transformation. What's an example of a vector that maps to it? Clearly, since this transformation sends $\vec{e_1}$ to $\vec{e_2}$, the following vector $\vec{u}$ gets mapped to $\vec{v}$ under the transformation:

$$\vec{u} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

Now, however, note that since the kernel comprises the scalar multiples of the vector $\vec{e_2}$, we can add any scalar multiple of $\vec{e_2}$ to the vector $\vec{u}$ and get a new vector that maps to $\vec{v}$. Explicitly, the fiber over $\vec{v}$ is the set:

$$\left\{ \begin{bmatrix} 5 \\ y \end{bmatrix} \mid y \in \mathbb{R} \right\}$$

Thus, the fiber over the point $(0,5)$ in the image is the line $x = 5$. Notice that this is a line parallel to the kernel of the linear transformation. In fact, all the fibers are lines parallel to the kernel of the linear transformation. In this case, they are all vertical lines.

Let's look at another, somewhat more complicated example:

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

Explicitly, the map is:

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x + y \\ 0 \end{bmatrix}$$

The map sends both $\vec{e_1}$ and $\vec{e_2}$ to $\vec{e_1}$. Hence, the matrix is idempotent.

The kernel is the line $y = -x$, and it is spanned by the vector:

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

The image is the line $y = 0$, i.e., the x-axis, so it is spanned by the vector:

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

The inverse image of the vector:

$$\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

is the empty set, because this vector is not in the image of the linear transformation. On the other hand, the inverse image of the vector:

$$\vec{w} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

is the set of all vectors on the line $x + y = 2$. Note that this is a parallel line to the kernel $x + y = 0$. Also note that the linear transformation is idempotent, and this explains the fact that every point in the image is in its own fiber.

5.2. **In terms of solving linear systems.** Let's think a bit more about this in terms of solving linear systems. If the linear system has full row rank, then row reducing gives a consistent system and we can read off the solutions after row reducing. Thus, *the linear transformation is surjective*. Each fiber looks like a translate of the kernel.

If the system does not have full row rank, there are rows that, after converting to rref, become zero in the coefficient matrix. Start with an augmenting column describing a vector in the co-domain. If the corresponding augmenting entry after doing the row operations is nonzero, the system is inconsistent, so the augmenting column is not in the image, or equivalently, the fiber is empty. If, however, after row reduction, all zero rows in the coefficient matrix have zero as the augmenting entry, then the system is consistent. We can take the particular solution vector as a vector where each leading variable is the corresponding augmenting column value and the non-leading variables are all zero. This is clearly a solution. We can add anything from the kernel to it and still get a solution.

6. Dimension computation

6.1. **Dimension of subspace.** The *dimension* of a subspace is defined as the minimum possible size of a spanning set. The dimension of the zero subspace is zero, because by convention we consider the span of the empty set to be zero (this may be unclear right now but will hopefully become clearer as we proceed). The dimension of $\mathbb{R}^p$ itself is p.

6.2. **Dimension of kernel and image.** Suppose $T : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation. The matrix A of T is a $n \times m$ matrix. We have seen the following in the past:

- T is injective if and only if A has full column rank, i.e., rank m. This implies $m \leq n$.
  Explanation: For T to be injective, we need that the linear system $A\vec{x} = \vec{y}$ always has at most one solution. This requires that there be no non-leading variables, i.e., the rref of A should have all its columns as pivotal columns. This forces full column rank.
- T is surjective if and only if A has full row rank, i.e., rank n. This implies $n \leq m$.
  Explanation: For T to be surjective, we need that the linear system $A\vec{x} = \vec{y}$ always has at least one solution, regardless of $\vec{y}$. In particular, this means that we cannot have any zero rows in the rref of A, because a zero row would open up the possibility that the corresponding augmenting entry is nonzero and the system could therefore be inconsistent for some $\vec{y}$. This possibility needs to be avoided, so we need that all rows of A be nonzero, i.e., all of them contain pivotal 1s. This forces A to have full row rank.

Also note that:

- *Injectivity* is the same as the kernel being the zero subspace: This was noted at the end of our discussion on fibers.
- *Surjectivity* is the same as the image being the whole target space $\mathbb{R}^n$.

Thus, we get that:

- The kernel of T is the zero subspace if and only if T has full column rank, i.e., rank m. This implies $m \leq n$.
- The image of T is all of $\mathbb{R}^n$ if and only if T has full row rank, i.e., rank n. This implies $n \leq m$.

We can generalize these to give formulas for the dimension of the kernel and the image:

- The dimension of the kernel is (number of columns) - (rank). In terms of systems of linear equations, it is the number of non-leading variables. Intuitively, this is because the non-leading variables are the parameters in an equational description. Our explicit construction of a spanning set for the kernel gave one vector for each non-leading variable. In fact, a spanning set constructed this way is minimal.
- The dimension of the image is the rank: Although our original choice of spanning set for the image was the set of *all* column vectors, we may not need all column vectors. If the map is not injective, then some of the column vectors are *redundant* (we will discusss this term later). It will turn out that it suffices to use the column vectors (in the original matrix) corresponding to the leading variables. Note that determining which variables are leading and which variables are non-leading requires us to convert the matrix to rref, but we need to use the columns of the original matrix, *not* those of the rref. We will discuss this more after introducing the concept of basis.

6.3. **Facts about dimension we will prove later.** We will show later that if A is a vector subspace of a vector space B, then the dimension of A is less than or equal to the dimension of B. In fact, we will show that any minimal spanning set of A (also called a *basis* for A) can be extended (non-uniquely) to a minimal spanning set of B.

7. Kernel and image for some easy and useful types of linear transformations

7.1. **Definition of projection.** A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is termed a *projection* if $T^2 = T$. Here, T^2 denotes the composite $T \circ T$ of T with itself. Another term used to describe a projection is *idempotent*, and the term *idempotent* is typically used for the matrix corresponding to a projection (the term "idempotent" means "equal to its own square" and is used in all algebraic structures, whereas projection is a term specific to thinking of transformations).

If T is a projection, then the intersection of the kernel of T and the image of T comprises only the zero vector. However, this is not a sufficient condition to describe projections.

Here is an example of a projection (discussed earlier in the notes, when talking about fibers):

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

This linear transformation sends $\vec{e_1}$ to itself and sends the vector $\vec{e_2}$ to the vector $\vec{e_1}$. Doing it twice is the same as doing it once.

Note that the matrix is already in rref, so we can read off the kernel as well as the image:

- The kernel is the one-dimensional space spanned by the vector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. To see this, note that the second variable is non-leading, and setting that to 1 gives the first variable as -1.
- The image is the one-dimensional space spanned by $\vec{e_1}$. We can see this from the fact that the image is spanned by the column vectors, and that in this case, both the column vectors are the same.

Pictorially, both the kernel and the image are lines. If we call the coordinate directions x and y, then the kernel is the line $y = -x$ and the image is the line $y = 0$ (i.e., the x-axis). These lines intersect at a unique point (the origin).

Note that it is possible to have a linear transformation whose kernel and image intersect only in the zero vector, but such that the linear transformation is not a projection. In fact, any linear automorphism other than the identity automorphism is of this form: the kernel is the zero subspace, and the image is the whole space, so the intersection is the zero subspace, so the kernel and image intersect in the zero subspace, but the map is not a projection. As a concrete example, consider the linear transformation with matrix:

$$\begin{bmatrix} 2 \end{bmatrix}$$

The kernel of this is the zero subspace, and the image is all of $\mathbb{R}$. It is not idempotent. To see this, note that $2^2 = 4$ which differs from 2.

7.2. **Orthogonal projection.** We define $T : \mathbb{R}^n \to \mathbb{R}^n$ to be an *orthogonal projection* if it is a projection and every vector in the kernel of T is orthogonal to every vector in the image of T. Note that the projection described above, namely:

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

is *not* an orthogonal projection, i.e., it does *not* have orthogonal kernel and image.

Suppose an orthogonal projection $T : \mathbb{R}^n \to \mathbb{R}^n$ has rank m. That means that the image is m-dimensional and the kernel is $(n - m)$-dimensional. The kernel and image are orthogonal subspaces.

For now, we consider a special case of orthogonal projection: *the orthogonal projection to the subspace of the first few coordinates.* Explicitly, consider a vector space $\mathbb{R}^n$ and consider the projection map that preserves the first m coordinates, with $0 < m < n$, and sends the remaining coordinates to zero. The matrix for such a linear transformation has the following block form:

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

where I_m is the $m \times m$ identity matrix. Explicitly, the first m diagonal entries are 1. The remaining $n - m$ entries on the diagonal are 0. All the entries that are not on the diagonal are also zero.

In the case $n = 2$, $m = 1$, the projection is given by the matrix:

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

Physically, this is the orthogonal projection onto the x-axis.

If P is an orthogonal projection matrix, then $P^2 = P$. This makes sense, because projecting something and then projecting again should have the same effect as doing one projection.

7.3. **Linear transformations that project, permute, and pad.** A linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ could be constructed that combines aspects from many different types of linear transformations:

- It could look at only r of the m coordinates and forget the remaining $m - r$.
- It could then map these r coordinates into separate coordinates of $\mathbb{R}^n$.
- It could then fill in zeros for the remaining coordinates.

The matrix for this linear transformation has rank r. There are r columns with one 1 and the other $(n-1)$ entries 0. The remaining $m-r$ columns are all zeros. Similarly, there are r rows with one 1 and the other $(m-1)$ entries zero. The remaining $n-r$ rows are all zeros.

Here's what we can say about the kernel and image:

- The kernel comprises those vectors that have zeros in all the r coordinates that matter, and can have arbitrary entries in the other coordinates. It thus has dimension $m-r$, and has a spanning set comprising those $m-r$ standard basis vectors.
- The image has dimension r, and has a spanning set comprising the standard basis vectors for the coordinates that get hit.

Here's an example of a linear transformation of this sort from $\mathbb{R}^3$ to $\mathbb{R}^4$:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} x_3 \\ 0 \\ x_2 \\ 0 \end{bmatrix}$$

The kernel here is spanned by $\vec{e_1}$ in $\mathbb{R}^3$ and the image is spanned by $\vec{e_1}$ and $\vec{e_3}$ in $\mathbb{R}^4$. The kernel is one-dimensional and the image is two-dimensional.

8. Intersection and sum of vector spaces

8.1. **Intersection of vector subspaces.** Suppose U and V are vector subspaces of $\mathbb{R}^n$. The intersection $U \cap V$ of U and V as sets is then also a vector subspace of $\mathbb{R}^n$. This is fairly easy to see: if both U and V individually contain the zero vector, so does the intersection. If both U and V are closed under addition, so is the intersection. Finally, if both U and V are closed under scalar multiplication, so is the intersection.

8.2. **Sum of vector subspaces.** Suppose U and V are vector subspaces of $\mathbb{R}^n$. We define $U+V$ as the subset of $\mathbb{R}^n$ comprising vectors expressible as the sum of a vector in U and a vector in V. The sum $U+V$ is a vector space: clearly it contains the 0 vector (since $0 = 0 + 0$). In addition, it is closed under addition and scalar multiplication:

- *Addition*: Suppose $\vec{w_1}$, $\vec{w_2}$ are vectors in $U+V$. Our goal is to show that $\vec{w_1}+\vec{w_2}$ is also in $U+V$. We know that for $\vec{w_1}$, we can find vectors $\vec{u_1} \in U$ and $\vec{v_1} \in V$ such that $\vec{w_1} = \vec{u_1} + \vec{v_1}$. Similarly, for $\vec{w_2}$, we can find vectors $\vec{u_2} \in U$ and $\vec{v_2} \in V$ such that $\vec{w_2} = \vec{u_2} + \vec{v_2}$. Thus, we have that:

$$\vec{w_1} + \vec{w_2} = (\vec{u_1} + \vec{v_1}) + (\vec{u_2} + \vec{v_2}) = (\vec{u_1} + \vec{u_2}) + (\vec{v_1} + \vec{v_2})$$

  Since U is a vector space, we know that $\vec{u_1} + \vec{u_2} \in U$. And since V is a vector space, we know that $\vec{v_1} + \vec{v_2} \in V$. Thus, the vector $\vec{w_1} + \vec{w_2}$ is in $U + V = W$.
- *Scalar multiplication*: Suppose $\vec{w} \in W$ and $a \in \mathbb{R}$. We want to show that $a\vec{w} \in W$. Note that by the definition of W as $U+V$, there exists a vector $\vec{u} \in U$ and $vecv \in V$ such that $\vec{w} = \vec{u} + \vec{v}$. Then:

$$a\vec{w} = a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$$

9. Sum of linear transformations, kernel, and image

In general, adding two linear transformations can play havoc with the structure, particularly if the linear transformations look very different from each other. We explore the kernel and image separately.

9.1. **Effect of adding linear transformations on the kernel.** Suppose $T_1, T_2 : \mathbb{R}^m \to \mathbb{R}^n$ are linear transformations. Note that although for convenience we describe these as maps from $\mathbb{R}^m$ to $\mathbb{R}^n$, whatever we say generalizes to maps between arbitrary vector spaces, including infinite-dimensional ones (with the caveat that statements about dimension would have to accommodate the possibility of the dimension being infinite). The sum $T_1 + T_2$ is also a linear transformation. What is the relation between the kernels of T_1, T_2, and $T_1 + T_2$? The kernels satisfy a "flower arrangement" which means that the following are true:

- If a vector $\vec{u} \in \mathbb{R}^m$ is in the kernels of both T_1 and T_2, then $\vec{u}$ is also in the kernel of $T_1 + T_2$. To see this, note that:

$$(T_1 + T_2)(\vec{u}) = T_1(\vec{u}) + T_2(\vec{u}) = 0 + 0 = 0$$

- If a vector $\vec{u} \in \mathbb{R}^m$ is in the kernels of both T_1 and $T_1 + T_2$, then $\vec{u}$ is also in the kernel of T_2. To see this, note that:

$$T_2(\vec{u}) = (T_1 + T_2)(\vec{u}) - T_1(\vec{u}) = 0 - 0 = 0$$

- If a vector $\vec{u} \in \mathbb{R}^m$ is in the kernels of both T_2 and $T_1 + T_2$, then $\vec{u}$ is also in the kernel of T_1. To see this, note that:

$$T_1(\vec{u}) = (T_1 + T_2)(\vec{u}) - T_2(\vec{u}) = 0 - 0 = 0$$

- In particular, this means that the intersection of the kernels of T_1, T_2, and $T_1 + T_2$ equals the intersection of any two of the three.

Note that the kernel of $T_1 + T_2$ could be a lot *bigger* than the kernels of T_1 and T_2. For instance, if T_1 is the identity transformation and $T_2 = -T_1$, then both T_1 and T_2 have zero kernel but the kernel of $T_1 + T_2$ is the entire domain.

9.2. **Effect of adding linear transformations on the image.** Using the same notation as for the preceding section, the following results hold regarding the relationship between the image of T_1, the image of T_2, and the image of $T_1 + T_2$:

- The image of $T_1 + T_2$ is contained in the sum of the image of T_1 and the image of T_2.
- The image of T_1 is contained in the sum of the image of T_2 and the image of $T_1 + T_2$.
- The image of T_2 is contained in the sum of the image of T_1 and the image of $T_1 + T_2$.
- Combining, we get that the sum of the images of all three of T_1, T_2, and $T_1 + T_2$ equals the sum of the images of any two of the three.

10. Composition, kernel, and image

10.1. **Dimensions of images of subspaces.** Suppose $T : U \to V$ is a linear transformation and W is a subspace of U. Then, $T(W)$ is a subspace of V, and the dimension of $T(W)$ is less than or equal to the dimension of W. It's fairly easy to see this: if S is a spanning set for W, then $T(S)$ is a spanning set for $T(W) = \operatorname{Im}(T)$.

In fact, there is a stronger result, which we will return to later:

$$\dim(W) = \dim(\operatorname{Im}(T)) + \dim(\ker(T))$$

10.2. **Dimensions of inverse images of subspaces.** Suppose $T : U \to V$ is a linear transformation and X is a linear subspace of V. Then, $T^{-1}(X)$, defined as the set of vectors in U whose image is in X, is a subspace of U. We have the following lower and upper bounds for the dimension of $T^{-1}(X)$:

$$\dim(\operatorname{Ker}(T)) \le \dim(T^{-1}(X)) \le \dim(\operatorname{Ker}(T)) + \dim(X)$$

The second inequality becomes an equality if and only if $X \subseteq \operatorname{Im}(T)$.

The first inequality is easy to justify: since X is a vector space, the zero vector is in X. Thus, everything in $\operatorname{Ker}(T)$ maps inside X, so that:

$$\operatorname{Ker}(T) \subseteq T^{-1}(X)$$

Based on the observation made above about dimensions, we thus get the first inequality:

$$\dim(\operatorname{Ker}(T)) \le \dim(T^{-1}(X))$$

The second inequality is a little harder to show directly. Suppose $Y = T^{-1}(X)$. Now, you may be tempted to say that $X = T(Y)$. That is not quite true. What is true is that $T(Y) = X \cap \operatorname{Im}(T)$. Based on the result of the preceding section:

$$\dim(Y) = \dim(\text{Ker}(T)) + \dim(T(Y))$$

Using the fact that $T(Y) \subseteq X$ we get the second inequality.

These generic statements are easiest to justify using the case of projections.

10.3. **Composition, injectivity, and surjectivity in the abstract linear context.** The goal here is to adapt the statements we made earlier about injectivity, surjectivity, and composition to the context of linear transformations.

Suppose $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations. The composite $T_2 \circ T_1 : U \to W$ is also a linear transformation. The following are true:

(1) If T_1 and T_2 are both injective, then $T_2 \circ T_1$ is also injective. This follows from the corresponding statement for arbitrary functions.
    In linear transformation terms, if the kernel of T_1 is the zero subspace of U and the kernel of T_2 is the zero subspace of V, then the kernel of $T_2 \circ T_1$ is the zero subspace of U.
(2) If T_1 and T_2 are both surjective, then $T_2 \circ T_1$ is also surjective. This follows from the corresponding statement for arbitrary functions.
    In linear transformation terms, if the image of T_1 is all of V and the image of T_2 is all of W, then the image of $T_2 \circ T_1$ is all of W.
(3) If T_1 and T_2 are both bijective, then $T_2 \circ T_1$ is also bijective. This follows from the corresponding statement for arbitrary functions. Bijective linear transformations between vector spaces are called *linear isomorphisms*, a topic we shall return to shortly.
(4) If $T_2 \circ T_1$ is injective, then T_1 is injective. This follows from the corresponding statement for functions.
    In linear transformation terms, if the kernel of $T_2 \circ T_1$ is the zero subspace of U, the kernel of T_1 is also the zero subspace of U.
(5) If $T_2 \circ T_1$ is surjective, then T_2 is surjective. This follows from the corresponding statement for functions.
    In linear transformation terms, if $T_2 \circ T_1$ has image equal to all of W, then the image of T_2 is also all of W.

10.4. **Composition, kernel and image: more granular formulations.** The above statements about injectivity and surjectivity can be modified somewhat to keep better track of the *lack* of injectivity and surjectivity by using the dimension of the kernel and image. Explicitly, suppose $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, so that the composite $T_2 \circ T_1 : U \to W$ is also a linear transformation. We then have:

(1) The dimension of the kernel of the composite $T_2 \circ T_1$ is at *least* equal to the dimension of the kernel of T_1 and at *most* equal to the sum of the dimensions of the kernels of T_1 and T_2 respectively. This is because the kernel of $T_2 \circ T_1$ can be expressed as $T_1^{-1}(T_2^{-1}(0))$, and we can apply the upper and lower bounds discussed earlier. Note that the lower bound is realized through an actual subspace containment: the kernel of T_1 is contained in the kernel of $T_2 \circ T_1$.
(2) The dimension of the image of the composite $T_2 \circ T_1$ is at *most* equal to the minimum of the dimensions of the images of T_1 and T_2. The bound in terms of the image of T_2 is realized interms of an explicit subspace containment: the image of $T_2 \circ T_1$ is a subspace of the image of T_2. The bound in terms of the dimension of T_1 is more subtle: the image of $T_2 \circ T_1$ is an image under T_2 of the image of T_1, therefore its dimension is at most equal to the dimension of the image of T_1.

10.5. **Consequent facts about matrix multiplication and rank.** Suppose A is a $m \times n$ matrix and B is a $n \times p$ matrix, where m, n, p are (possibly equal, possibly distinct) positive integers. Suppose the rank of A is r and the rank of B is s. Recall that the rank equals the dimension of the image of the associated linear transformation. The difference between the dimension of the domain and the rank equals the dimension of the kernel of the associated linear transformation. We will later on use the term *nullity* to describe the dimension of the kernel. The nullity of A is $n - r$ and the nullity of B is $p - s$.

The matrix AB is a $m \times p$ matrix. We can say the following:

- The rank of AB is at *most* equal to the minimum of the ranks of A and B. In this case, it is at most equal to $\min\{r, s\}$. This follows from the observation about the dimension of the image of a composite, made in the preceding section.
- The nullity of AB is at *least* equal to the nullity of B. In particular, in this case, it is at least equal to $p - s$. Note that the information about the maximum value of the rank actually allows us to improve this lower bound to $p - \min\{r, s\}$, though that is not directly obvious at all.
- In particular, if the product AB has full column rank p, then B has full column rank p, $s = p$, and $r \geq p$ (so in particular $m \geq p$ and $n \geq p$).

  Note that this is the linear algebra version of the statement that if the composite $g \circ f$ is injective, then f must be injective.
- If the product AB has full row rank m, then A has full row rank m, $r = m$, and $s \geq m$ (so in particular $n \geq m$ and $p \geq m$).

  Note that this is the linear algebra version of the statement that if $g \circ f$ is surjective, then g is surjective.