Math 115A - Week 2
Textbook sections: 1.6-2.1
Topics covered:

- Properties of bases
- Dimension of vector spaces
- Lagrange interpolation
- Linear transformations

* * * * *

Review of bases

- In last week's notes, we had just defined the concept of a *basis.* Just to quickly review the relevant definitions:
- Let V be a vector space, and S be a subset of V. The *span* of S is the set of all linear combinations of elements in S; this space is denoted $\mathrm{span}(S)$ and is a subspace of V. If $\mathrm{span}(S)$ is in fact equal to V, we say that S *spans* V.
- We say that S is *linearly dependent* if there is some non-trivial way to write 0 as a linear combination of elements of S. Otherwise we say that S is *linearly independent.*
- We say that S is a *basis* for V if it spans V and is also linearly independent.
- Generally speaking, the larger the set is, the more likely it is to span, but also the less likely it is to remain linearly independent. In some sense, bases form the boundary between the "large" sets which span but are not independent, and the "small" sets which are independent but do not span.

* * * * *

Examples of bases

- Why are bases useful? One reason is that they give a compact way to describe vector spaces. For instance, one can describe $\mathbf{R}^3$ as the vector space spanned by the basis $\{(1,0,0),(0,1,0),(0,0,1)\}$:

  $$\mathbf{R}^3 = \mathrm{span}(\{(1,0,0),(0,1,0),(0,0,1)\}).$$

  In other words, the three vectors $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ are linearly independent, and $\mathbf{R}^3$ is precisely the set of all vectors which can be written as linear combinations of $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$.

- Similarly, one can describe $P(\mathbf{R})$ as the vector space spanned by the basis $\{1,x,x^2,x^3,\ldots\}$. Or $P_{even}(\mathbf{R})$, the vector space of even polynomials, is the vector space spanned by the basis $\{1,x^2,x^4,x^6,\ldots\}$ (why?).

- Now for a more complicated example. Consider the space

  $$V := \{(x,y,z) \in \mathbf{R}^3 : x+y+z=0\};$$

  in other words, V consists of all the elements in $\mathbf{R}^3$ whose co-ordinates sum to zero. Thus for instance $(3,5,-8)$ lies in V, but $(3,5,-7)$ does not. The space V describes a plane in $\mathbf{R}^3$; if you remember your Math 32A, you'll recall that this is the plane through the origin which is perpendicular to the vector $(1,1,1)$. It is a subspace of $\mathbf{R}^3$, because it is closed under vector addition and scalar multiplication (why?).

- Now let's try to find a basis for this space. A straightforward, but slow, procedure for doing so is to try to build a basis one vector at a time: we put one vector in V into the (potential) basis, and see if it spans. If it doesn't, we throw another (linearly independent) vector into the basis, and then see if it spans. We keep repeating this process until eventually we get a linearly independent set spanning the entire space - i.e. a basis. (Every time one adds more vectors to a set S, the span $\mathrm{span}(S)$ must get larger (or at least stay the same size) - why?).

- To begin this algorithm, let's pick an element of the space V. We can't pick 0 - any set with 0 is automatically linearly dependent (why?), but there are other, fairly simple vectors in V; let's pick $v_1 := (1,0,-1)$. This vector is in V, but it doesn't span V: the linear combinations of v_1 are all of the form $(a,0,-a)$, where $a \in \mathbf{R}$ is a scalar, but this doesn't include all the vectors in V. For instance, $v_2 := (1, -1, 0)$ is clearly not in the span of v_1. So now we take both v_1 and v_2 and see if they span. A typical linear combination of v_1 and v_2 is

$$a_1 v_1 + a_2 v_2 = a_1(1, 0, -1) + a_2(1, -1, 0) = (a_1 + a_2, -a_2, -a_1)$$

and so the question we are asking is: can every element (x, y, z) of V be written in the form $(a_1 + a_2, -a_2, -a_1)$? In other words, can we solve the system

$$\begin{array}{rrl} a_1 & +a_2 & = x \\ & -a_2 & = y \\ -a_1 & & = z \end{array}$$

for every $(x, y, z) \in V$? Well, one can solve for a_1 and a_2 as

$$a_1 := -z, a_2 := -y.$$

The first equation then becomes $-z - y = x$, but this equation is valid because we are assuming that $(x, y, z) \in V$, so that $x + y + z = 0$. (This is not all that of a surprising co-incidence: the vectors v_1 and v_2 were chosen to be in V, which explains why the linear combination $a_1 v_1 + a_2 v_2$ must also be in V). Thus every vector in V can be written as a linear combination of v_1 and v_2. Also, these two vectors are linearly independent (why?), and so $\{v_1, v_2\} = \{(1, 0, -1), (1, -1, 0)\}$ is a basis for V.

- It is clear from the above that this is not the only basis available for V; for instance, $\{(1, 0, -1), (0, 1, -1)\}$ is also a basis. In fact, as it turns out, any two linearly independent vectors in V can be used to form a basis for V. Because of this, we say that V is *two-dimensional.* It turns out (and this is actually a rather deep fact) that many of the vector spaces V we will deal with have some finite dimension d, which means that any d linearly independent vectors in V automatically form a basis; more on this later.

- A philosophical point: we now see that there are (at least) two ways to construct vector spaces. One is to start with a "big" vector space, say $\mathbf{R}^3$, and then impose *constraints* such as $x + y + z = 0$ to cut the vector space down in size to obtain the target vector space, in this case

V. An opposing way to make vector spaces is to start with nothing, and throw in vectors one at a time (in this case, v_1 and v_2) to build up to the target vector space (which is also V). A basis embodies this second, "bottom-up" philosophy.

* * * * *

Rigorous treatment of bases

- Having looked at some examples of how to construct bases, let us now introduce some theory to make the above algorithm rigorous.
- **Theorem 1.** Let V be a vector space, and let S be a linearly independent subset of V. Let v be a vector which does not lie in S.
- (a) If v lies in $\mathrm{span}(S)$, then $S \cup \{v\}$ is linearly dependent, and $\mathrm{span}(S \cup \{v\}) = \mathrm{span}(S)$.
- (b) If v does not lie in $\mathrm{span}(S)$, then $S \cup \{v\}$ is linearly independent, and $\mathrm{span}(S \cup \{v\}) \supsetneq \mathrm{span}(S)$.
- This theorem justifies our previous reasoning: if a linearly independent set S does not span V, then one can make the span bigger by adding a vector outside of $\mathrm{span}(S)$; this will also keep S linearly independent.
- **Proof** We first prove (a). If v lies in $\mathrm{span}(S)$, then by definition of span, v must be a linear combination of S, i.e. there exists vectors $v_1, \ldots, v_n$ in S and scalars $a_1, \ldots, a_n$ such that

  $$v = a_1 v_1 + \ldots + a_n v_n$$

  and thus

  $$0 = (-1)v + a_1 v_1 + \ldots + a_n v_n.$$

  Thus 0 is a non-trivial linear combination of v, $v_1, \ldots, v_n$ (it is non-trivial because the co-efficient -1 in front of v is non-zero. Note that since $v \notin S$, this coefficient cannot be cancelled by any of the v_j). Thus $S \cup \{v\}$ is linearly dependent. Furthermore, since v is a linear combination of $v_1, \ldots, v_n$, any linear combination of v and $v_1, \ldots, v_n$ can be re-expressed as a linear combination just of $v_1, \ldots, v_n$ (why?). Thus $\mathrm{span}(S \cup \{v\})$ does not contain any additional elements which are not already in span(S). On the other hand, every element in span(S) is clearly also in span$(S \cup \{v\})$. Thus span$(S \cup \{v\})$ and span(S) have precisely the same set of elements, i.e. span$(S \cup \{v\}) =$ span(S).

- Now we prove (b). Suppose $v \notin$ span(S). Clearly span$(S \cup \{v\})$ contains span(S), since every linear combination of S is automatically a linear combination of $S \cup \{v\}$. But span$(S \cup \{v\})$ clearly also contains v, which is not in span(S). Thus span$(S \cup \{v\}) \supsetneq$ span(S).

- Now we prove that $S \cup \{v\}$ is linearly independent. Suppose for contradiction that $S \cup \{v\}$ was linearly dependent. This means that there is some non-trivial way to write 0 as a linear combination of v and some vectors $v_1, \ldots, v_n$ in S:

$$0 = av + a_1 v_1 + \ldots + a_n v_n.$$

If a were zero, then we would be writing 0 as a non-trivial linear combination of elements $v_1, \ldots, v_n$ in S, but this contradicts the hypothesis that S is linearly independent. Thus a is non-zero. But then we may divide by a and conclude that

$$v = (-\frac{a_1}{a})v_1 + \ldots + (-\frac{a_n}{a})v_n,$$

so that v is a linear combination of $v_1, \ldots, v_n$, so it is in the span of S, a contradiction. Thus $S \cup \{v\}$ is linearly independent. □

* * * * *

Dimension

- As we saw in previous examples, a vector space may have several bases. For instance, if $V := \{(x, y, z) \in \mathbf{R}^3 : x + y + z = 0\}$, then $\{(1, 0, -1), (1, -1, 0)\}$ is a basis, but so is $\{(1, 0, -1), (0, 1, -1)\}$.

- If V was the line $\{(t, t, t) : t \in \mathbf{R}\}$, then $\{(1, 1, 1)\}$ is a basis, but so is $\{(2, 2, 2)\}$. (On the other hand, $\{(1, 1, 1), (2, 2, 2)\}$ is not a basis because it is linearly dependent).

- If V was the zero vector space $\{0\}$, then the empty set $\{\}$ is a basis (why?), but $\{0\}$ is not (why?).

- In $\mathbf{R}^3$, the three vectors $\{(1,0,0),(0,1,0),(0,0,1)\}$ form a basis, and there are many other examples of three vectors which form a basis in $\mathbf{R}^3$ (for instance, $\{(1,1,0),(1,-1,0),(0,0,1)\}$. As we shall see, any set of two or fewer vectors cannot be a basis for $\mathbf{R}^3$ because they cannot span all of $\mathbf{R}^3$, while any set of four or more vectors cannot be a basis for $\mathbf{R}^3$ because they become linearly dependent.

- One thing that one sees from these examples is that all the bases of a vector space seem to contain the same number of vectors. For instance, $\mathbf{R}^3$ always seems to need exactly three vectors to make a basis, and so forth. The reason for this is in fact rather deep, and we will now give the proof. The first step is to prove the following rather technical result, which says that one can "edit" a spanning set by inserting a fixed linearly independent set, while removing an equal number of vectors from the previous spanning set.

- **Replacement Theorem.** Let V be a vector space, and let S be a finite subset of V which spans V (i.e. $\text{span}(S) = V$). Suppose that S has exactly n elements. Now let L be another finite subset of V which is linearly independent and has exactly m elements. Then m is less than or equal to n. Furthermore, we can find a subset S' of S containing exactly $n-m$ elements such that $S' \cup L$ also spans V.

- This theorem is not by itself particularly interesting, but we can use it to imply a more interesting Corollary, below.

- **Proof** We induct on m. The base case is $m = 0$. Here it is obvious that $n \geq m$. Also, if we just set S' equal to S, then S' has exactly $n-m$ elements, and $S' \cup L$ is equal to S (since L is empty) and so obviously spans V by hypothesis.

- Now suppose inductively that $m > 0$, and that we have already proven the theorem for $m-1$. Since L has m elements, we may write it as

  $$L = \{v_1, \ldots, v_m\}.$$

  Since $\{v_1, \ldots, v_m\}$ is linearly independent, the set $\tilde{L} := \{v_1, \ldots, v_{m-1}\}$ is also linearly independent (why?). We can now apply the induction hypothesis with m replaced by $m-1$ and L replaced by $\tilde{L}$. This tells us that $n \geq m-1$, and also there is some subset $\tilde{S}'$ of S with exactly $n-m+1$ elements, such that $\tilde{S}' \cup \tilde{L}$ spans V.

- Write $\tilde{S}' = \{w_1, \ldots, w_{n-m+1}\}$. To prove that $n \geq m$, we have to exclude the possibility that $n = m-1$. We do this as follows. Consider the vector v_m, which is in L but not in $\tilde{L}$. Since the set

  $$\tilde{S}' \cup \tilde{L} = \{v_1, \ldots, v_{m-1}, w_1, \ldots, w_{n-m+1}\}$$

  spans V, we can write v_m as a linear combination of $\tilde{S}' \cup \tilde{L}$. In other words, we have

  $$v_m = a_1 v_1 + \ldots + a_{m-1} v_{m-1} + b_1 w_1 + \ldots + b_{n-m+1} w_{n-m+1} \qquad (0.1)$$

  for some scalars $a_1, \ldots, a_{m-1}, b_1, \ldots, b_{n-m+1}$.

- Suppose for contradiction that $n = m-1$. Then $\tilde{S}'$ is empty, and there are no vectors $w_1, \ldots, w_{n-m+1}$. We thus have

  $$v_m = a_1 v_1 + \ldots + a_{m-1} v_{m-1} \qquad (0.2)$$

  so that

  $$0 = a_1 v_1 + \ldots + a_{m-1} v_{m-1} + (-1) v_m$$

  but this contradicts the hypothesis that $\{v_1, \ldots, v_m\}$ is linearly independent. Thus n cannot equal $m-1$, and so must be greater than or equal to m.

- We now have $n \geq m$, so that there is at least one vector in $w_1, \ldots, w_{n-m+1}$. Since we know the set

  $$\tilde{S}' \cup \tilde{L} = \{v_1, \ldots, v_{m-1}, w_1, \ldots, w_{n-m+1}\}$$

  spans V, it is clear that

  $$\tilde{S}' \cup L = \{v_1, \ldots, v_m, w_1, \ldots, w_{n-m+1}\}$$

  also spans V (adding an element cannot decrease the span). To finish the proof we need to eliminate one of the vectors w_j, to cut $\tilde{S}'$ down to a set S' of size $n-m$, while still making $S' \cup L$ span V.

- We first observe that the $b_1, \ldots, b_{n-m+1}$ cannot all be zero, otherwise we would be back to equation (0.2) again, which leads to contradiction. So at least one of the b's must be non-zero; since the order of the vectors w_j is irrelevant, let's say that b_1 is the one which is non-zero. But then we can divide by b_1, and use (0.1) to solve for w_1:

$$w_1 = \frac{1}{b_1} v_m - \frac{a_1}{b_1} v_1 - \ldots - \frac{a_{m-1}}{b_1} v_{m-1} - \frac{b_2}{b_1} w_2 - \ldots - \frac{b_{n-m+1}}{b_1} w_{n-m+1}.$$

Thus w_1 is a linear combination of $\{v_1, \ldots, v_m, w_2, \ldots, w_{n-m+1}\}$. In other words, if we write $S' := \{w_2, \ldots, w_{n-m+1}\}$, then w_1 is a linear combination of $S' \cup L$. Thus by Theorem 1,

$$\mathrm{span}(S' \cup L) = \mathrm{span}(S' \cup L \cup \{w_1\}).$$

But $S' \cup L \cup \{w_1\}$ is just $\tilde{S}' \cup L$, which spans V. Thus $S' \cup L$ spans V. Since S' has exactly $n - m$ elements, we are done. □

- **Corollary 1** Suppose that a vector space V contains a finite basis B which consists of exactly d elements. Then:

- (a) Any set $S \subseteq V$ consisting of fewer than d elements cannot span V. (In other words, every spanning set of V must contain at least d elements).

- (b) Any set $S \subset V$ consisting of more than d elements must be linearly dependent. (In other words, every linearly independent set in V can contain at most d elements).

- (c) Any basis of V must consist of exactly d elements.

- (d) Any spanning set of V with exactly d elements, forms a basis.

- (e) Any set of d linearly independent elements of V forms a basis.

- (f) Any set of linearly independent elements of V is contained in a basis.

- (g) Any spanning set of V contains a basis.

- **Proof** We first prove (a). Let S have d' elements for some $d' < d$. Suppose for contradiction that S spanned V. Since B is linearly independent, we may apply the Replacement Theorem (with B playing the role of L) to conclude that $d' \geq d$, a contradiction. Thus S cannot span V.

- Now we prove (b). First suppose that S is finite, so that S has d' elements for some $d' > d$. Suppose for contradiction that S is linearly independent. Since B spans V, we can apply the Replacement theorem (with B playing the role of S, while S instead plays the role of L) to conclude that $d \geq d'$, a contradiction. So we've proven (b) when S is finite. When S is infinite, we can find a finite subset S' of S with, say, $d + 1$ elements; since we've already proven (b) for finite subsets, we know that S' is linearly dependent. But this implies that S is also linearly dependent.

- Now we prove (c). Let B' be any basis of V. Since B' spans, it must contain at least d elements, by (a). Since B' is linearly independent, it must contain at most d elements, by (b). Thus it must contain exactly d elements.

- Now we prove (d). Let S be a spanning set of V with exactly d elements. To show that S is a basis, we need to show that S is linearly independent. Suppose for contradiction that S was linearly dependent. Then by a theorem in page 34 of last week's notes, there exists a vector v in S such that $\text{span}(S - \{v\}) = \text{span}(S)$. Thus $S - \{v\}$ also spans V, but it has fewer than d elements, contradicting (a). Thus S must be linearly independent.

- Now we prove (e). Let L be a linearly independent set in V with exactly d elements. To show that L is a basis, we need to show that L spans. Suppose for contradiction that L did not span, then there must be some vector v which is not in the span of L. But by Theorem 1 in this week's notes, $L \cup \{v\}$ is linearly independent. But this set has more than d elements, contradicting (b). Thus L must span V.

- Now we prove (f). Let L be a linearly independent set in V; by (a), we know it has d' elements for some $d' \leq d$. Applying the Replacement theorem (with B playing the role of the spanning set S), we see that there is some subset S' of B with $d-d'$ elements such that $S' \cup L$ spans V. Since S' has $d-d'$ elements and L has d' elements, $S' \cup L$ can have at most d elements; actually it must have exactly d, else it would not span by (a). But then by (d) it must be a basis. Thus L is contained in a basis.

- Now we prove (g). Let S be a spanning set in V. To build a basis inside S, we see by (e) that we just need to find d linearly independent vectors in S. Suppose for contradiction that we can only find at most d' linearly independent vectors in S for some $d' < d$. Let $v_1, \ldots, v_{d'}$ be d' such linearly independent vectors in S. Then every other vector v in S must be a linear combination of $v_1, \ldots, v_{d'}$, otherwise we could add v to $\{v_1, \ldots, v_{d'}\}$ and obtain a larger collection of linearly independent vectors in S (see Theorem 1). But if every vector in S is a linear combination of $v_1, \ldots, v_{d'}$, and S spans V, then $v_1, \ldots, v_{d'}$ must span V. By (a) this means that $d' \geq d$, contradiction. Thus we must be able to find d linearly independent vectors in S, and so S contains a basis. □

- **Definition** We say that V has *dimension d* if it contains a basis of d elements (and so that all the consequences of the Corollary 1 follow). We say that V is *finite-dimensional* if it has dimension d for some finite number d, otherwise we say that V is *infinite-dimensional.*

- From Corollary 1 we see that all bases have the same number of elements, so a vector space cannot have two different dimensions. (e.g. a vector space cannot be simultaneously two-dimensional and three-dimensional). We sometimes use $\dim(V)$ to denote the dimension of V. One can think of $\dim(V)$ as the number of degrees of freedom inherent in V (or equivalently, the number of possible linearly independent vectors in V).

- **Example** The vector space $\mathbf{R}^3$ has a basis $\{(1,0,0),(0,1,0),(0,0,1)\}$, and thus has dimension 3. Thus any three linearly independent vectors in $\mathbf{R}^3$ will span $\mathbf{R}^3$ and form a basis.

- **Example** The vector space $P_n(\mathbf{R})$ of polynomials of degree $\leq n$ has basis $\{1, x, x^2, \ldots, x^n\}$ and thus has dimension $n+1$.

- **Example** The zero vector space $\{0\}$ has a basis $\{\}$ and thus has dimension zero. (It is the only vector space with dimension zero - why?)

- **Example** The vector space $P(\mathbf{R})$ of all polynomials is infinite dimensional. To see this, suppose for contradiction that it had some finite dimension d. But then one could not have more than d linearly independent elements. But the set $\{1, x, x^2, \ldots, x^d\}$ contains $d+1$ elements which are linearly independent (why?), contradiction. Thus $P(\mathbf{R})$ is infinite dimensional.

- As we have seen, every finite dimensional space has a basis. It is also true that infinite-dimensional spaces also have bases, but this is significantly harder to prove and beyond the scope of this course.

* * * * *

Subspaces and dimension

- We now prove an intuitively obvious statement about subspaces and dimension:

- **Theorem 2.** Let V be a finite-dimensional vector space, and let W be a subspace of V. Then W is also finite-dimensional, and $\dim(W) \leq \dim(V)$. Furthermore, the only way that $\dim(W)$ can equal $\dim(V)$ is if $W = V$.

- **Proof.** We first construct a finite basis for W via the following algorithm. If $W = \{0\}$, then we can use the empty set as a basis. Now suppose that $W \neq \{0\}$. Then we can find a non-zero vector v_1 in W. If v_1 spans W, then we have found a basis for W. If v_1 does not span W, then we can find a vector v_2 which does not lie in $\mathrm{span}(\{v_1\})$; by Theorem 1, $\{v_1, v_2\}$ is linearly independent. If this set spans W, then we can found a basis for W. Otherwise, we can find a vector v_3 which does not lie in $\mathrm{span}(\{v_1, v_2\})$. By Theorem 1, $\{v_1, v_2, v_3\}$ is linearly independent. We continue in this manner until we finally span W. Note that we must stop before we exceed $\dim(V)$ vectors, since from part (b) of the dimension theorem we cannot make a linearly independent set with more than $\dim(V)$ vectors. Thus this algorithm must eventually generate a basis of W which consists of at most $\dim(V)$ vectors, which implies that W is finite-dimensional with $\dim(W) \leq \dim(V)$.

- Now suppose that $\dim(W) = \dim(V)$. Then W has a basis B which consists of $\dim(V)$ estimates; B is of course linearly independent. But then by part (e) of Corollary 1, B is also a basis for V. Thus $\mathrm{span}(B) = V$ and $\mathrm{span}(B) = W$, which implies that $W = V$ as desired. □

* * * * *

Lagrange interpolation

- We now give an application of this abstract theory to a basic problem: how to fit a polynomial to a specified number of points.

- Everyone knows that given two points in the plane, one can find a line joining them. A more precise way of saying this is that given two data points (x_1, y_1) and (x_2, y_2) in $\mathbf{R}^2$, with $x_1 \neq x_2$, then we can find a line $y = mx + b$ which passes through both these points. (We need $x_1 \neq x_2$ otherwise the line will have infinite slope).

- Now suppose we have three points (x_1, y_1), (x_2, y_2), (x_3, y_3) in the plane, with x_1, x_2, x_3 all distinct. Then one usually cannot fit a line which goes exactly through these three data points. (One can still do a *best fit* to these data points by a straight line, e.g. by using the least squares fit; this is an important topic but not one we will address now). However, it turns out that one can still fit a parabola $y = ax^2 + bx + c$ to these three points. With four points, one cannot always fit a parabola, but one can always fit a cubic. More generally:

- **Theorem 3 (Lagrange interpolation formula)** Let $n \geq 1$, and let $(x_1, y_1), \ldots, (x_n, y_n)$ be n points in $\mathbf{R}^2$ such that $x_1, x_2, \ldots, x_n$ are all distinct. Then there exists a unique polynomial $f \in P_P n(\mathbf{R})$ of degree $\leq n - 1$ such that the curve $y = f(x)$ passes through all n points $(x_1, y_1), \ldots, (x_n, y_n)$. In other words, we have $y_j = f(x_j)$ for all $j = 1, \ldots, n$. Furthermore, f is given by the formula

$$f(x) = \sum_{j=1}^{n} \frac{\prod_{1 \leq k \leq n : k \neq j}(x - x_k)}{\prod_{1 \leq k \leq n : k \neq j}(x_j - x_k)} y_j.$$

- The polynomial f is sometimes called the *interpolating polynomial* for the points $(x_1, y_1), \ldots (x_n, y_n)$; in some sense it is the simplest object that can pass through all n points. These interpolating polynomials have several uses, for instance in taking a sequence of still images, and finding a smooth sequence of intermediate images to fit between these images.

- To prove this theorem, we first proceed by considering some simple examples.

- First suppose that $y_1 = y_2 = \ldots = y_n = 0$. Then the choice of interpolating polynomial is obvious: just take the zero polynomial $f(x) = 0$.

- Now let's take the next simplest case, when $y_1 = 1$ and $y_2 = y_3 = \ldots = y_n = 0$. The interpolating polynomial f that we need here must obey the conditions $f(x_1) = 1$, and $f(x_2) = \ldots = f(x_n) = 0$.

- Since f has zeroes at $x_2, \ldots, x_n$, it must have factors of $(x - x_2), (x - x_3), \ldots, (x - x_n)$. So it must look like

  $$f = Q(x)(x - x_2) \ldots (x - x_n).$$

  Since $(x - x_2) \ldots (x - x_n)$ has degree $n - 1$, and we want f to have degree at most $n - 1$, $Q(x)$ must be constant, say $Q(x) = c$:

  $$f = c(x - x_2) \ldots (x - x_n).$$

  To find out what c is, we use the extra fact that $f(x_1) = 1$, so

  $$1 = c(x_1 - x_2) \ldots (x_1 - x_n).$$

  Thus the interpolating polynomial is given by f_1, where

  $$f_1(x) := \frac{(x - x_2) \ldots (x - x_n)}{(x_1 - x_2) \ldots (x_1 - x_n)}$$

  or equivalently

  $$f_1(x) := \frac{\prod_{k=2}^n (x - x_k)}{\prod_{k=2}^n (x_1 - x_k)}$$

  One can see by inspection that indeed $f_1(x_1)$ is equal to 1, while $f_1(x_2) = \ldots = f_1(x_n) = 0$.

- Now consider the case when $y_j = 1$ for some $1 \leq j \leq n$, and $y_k = 0$ for all other $k \neq j$ (the earlier case being the special case when $j = 1$). Then a similar argument gives that f must equal f_j, where f_j is the polynomial
$$f_j(x) := \frac{\prod_{1 \leq k \leq n : k \neq j}(x - x_k)}{\prod_{1 \leq k \leq n : k \neq j}(x_j - x_k)}.$$
For instance, if $n = 4$ and $j = 2$, then
$$f_2(x) := \frac{(x - x_1)(x - x_3)(x - x_4)}{(x_2 - x_1)(x_2 - x_3)(x_2 - x_4)}.$$

- To summarize, for each $1 \leq j \leq n$, we can find a polynomial $f_j \in P_{n-1}(\mathbf{R})$ such that $f_j(x_j) = 1$ and $f_j(x_k) = 0$ for $k \neq j$. Thus, for instance, when $n = 4$, then we have
$$f_1(x_1) = 1, f_1(x_2) = f_1(x_3) = f_1(x_4) = 0$$
$$f_2(x_2) = 1, f_2(x_1) = f_2(x_3) = f_2(x_4) = 0$$
$$f_3(x_3) = 1, f_3(x_1) = f_3(x_2) = f_3(x_4) = 0$$
$$f_4(x_4) = 1, f_4(x_1) = f_4(x_2) = f_4(x_3) = 0.$$

- To proceed further we need a key lemma.

- **Lemma 4.** The set $\{f_1, f_2, \ldots, f_n\}$ is a basis for $P_{n-1}(\mathbf{R})$.

- **Proof.** We already know that P_{n-1} is n-dimensional, since it has a basis $\{1, x, x^2, \ldots, x^{n-1}\}$ of n elements. Since $\{f_1, \ldots, f_n\}$ also has n elements, to show that it is a basis it will suffice by part (e) of Corollary 1 to show that $\{f_1, \ldots, f_n\}$ is linearly independent.

- Suppose for contradiction that $\{f_1, \ldots, f_n\}$ was linearly dependent. This means that there exists scalars $a_1, \ldots, a_n$, not all zero, such that $a_1 f_1 + a_2 f_2 + \ldots + a_n f_n$ is the zero polynomial i.e.
$$a_1 f_1(x) + a_2 f_2(x) + \ldots + a_n f_n(x) = 0 \text{ for all } x.$$
In particular, we have
$$a_1 f_1(x_1) + a_2 f_2(x_1) + \ldots + a_n f_n(x_1) = 0.$$

But since $f_1(x_1) = 1$ and $f_2(x_1) = \ldots = f_n(x_1) = 0$, we thus have

$$a_1 \times 1 + a_2 \times 0 + \ldots + a_n \times 0 = 0,$$

i.e. $a_1 = 0$. A similar argument gives that $a_2 = 0$, $a_3 = 0, \ldots$ - contradicting the assumption that the a_j were not all zero. Thus $\{f_1, \ldots, f_n\}$ is linearly independent, and is thus a basis by Corollary 1. □

- From Lemma 4 we know that $\{f_1, \ldots, f_n\}$ spans P_{n-1}. Thus every polynomial $f \in P_{n-1}$ can be written in the form

  $$f = a_1 f_1 + \ldots + a_n f_n \qquad (0.3)$$

  for some scalars $a_1, \ldots, a_n$. In particular, the interpolating polynomial between the data points $(x_1, y_1), \ldots, (x_n, y_n)$ must have this form. So to work out what the interpolating polynomial is, we just have to work out what the scalars $a_1, \ldots, a_n$ are.

- In order for f to be an interpolating polynomial, we need $f(x_1) = y_1$, $f(x_2) = y_2$, etc. Let's look at the first condition $f(x_1) = y_1$. Using (0.3), we have

  $$f(x_1) = a_1 f_1(x_1) + \ldots + a_n f_n(x_1) = y_1.$$

  But by arguing as in the lemma, we have

  $$a_1 f_1(x_1) + \ldots + a_n f_n(x_1) = a_1 \times 1 + a_2 \times 0 + \ldots + a_n \times 0 = a_1.$$

  Thus we must have $a_1 = y_1$. More generally, we see that $a_2 = y_2$, $a_3 = y_3, \ldots$. Thus the only possible choice of interpolating polynomial is

  $$f := y_1 f_1 + \ldots + y_n f_n = \sum_{j=1}^{n} y_j f_j \qquad (0.4)$$

  which is the Lagrange interpolation formula. Conversely, it is easy to check that if we define f by the formula (0.4), then $f(x_1) = y_1$, $f(x_2) = y_2$, etc. so f is indeed the unique interpolating polynomial between the data points $(x_1, y_1), \ldots, (x_3, y_3)$. □

- As an example, suppose one wants to interpolate a quadratic polynomial between the points $(1,0)$, $(2,2)$, and $(3,1)$, so that $x_1 := 1$, $x_2 := 2$, $x_3 := 3$, $y_1 := 0$, $y_2 := 2$, $y_3 := 1$. The formulae for f_1, f_2, f_3 are

$$f_1 := \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} = \frac{(x-2)(x-3)}{(1-2)(1-3)}$$

$$f_2 := \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} = \frac{(x-1)(x-3)}{(2-1)(2-3)}$$

$$f_3 := \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} = \frac{(x-1)(x-2)}{(3-1)(3-2)}$$

and so the interpolating polynomial is

$$f = 0f_1 + 2f_2 + 1f_3 = 2\frac{(x-1)(x-3)}{(2-1)(2-3)} + \frac{(x-1)(x-2)}{(3-1)(3-2)}.$$

You can check by direct substitution that $f(1) = 0$, $f(2) = 2$, and $f(3) = 1$ as desired. After a lot of algebra one can simplify f to a more standard form

$$f = -3x^2/2 + 13x/2 - 5.$$

- If one were to interpolate a single point (x_1, y_1), one would just get the constant polynomial $f = y_1$, which is of course the only polynomial of degree 0 which passes through (x_1, y_1).

- The Lagrange interpolation formula says that there is exactly one polynomial of degree at most $n-1$ which passes through n given points. However, if one is willing to use more complicated polynomials (i.e. polynomials of degree higher than $n-1$) then there are infinitely many more ways to interpolate those data points. For instance, take the points $(0,0)$ and $(1,1)$. There is only one linear polynomial which interpolates these points - the polynomial $f(x) := x$. But there are many quadratic polynomials which also interpolate these two points: $f(x) = x^2$ will work, as will $f(x) = \frac{1}{2}x^2 + \frac{1}{2}x$, or in fact any polynomial of the form $(1-\theta)x^2 + \theta x$. And with cubic polynomials there are even more possibilities. The point is that each degree you add to the polynomial adds one more degree of freedom (remember that the dimension of $P_n(\mathbf{R})$ is $n+1$), and is it comes increasingly easier to satisfy a fixed number of constraints (in this example there are only two constraints, one for each data point). This is part of a more general principle: when the number of degrees of freedom exceeds the number of constraints, then usually one has many solutions to a problem. When the number of constraints exceeds the number of degrees of freedom, one usually has no solutions to a problem. When the number of constraints exactly equals the number of degrees of freedom, one usually has exactly one solution to a problem. We will make this principle more precise later in this course.

* * * * *

Linear transformations

- Up until now we have studied each vector space in isolation, and looked at what one can do with the vectors in that vector space. However, this is only a very limited portion of linear algebra. To appreciate the full power of linear algebra, we have to not only understand each vector space individually, but also all the various *linear transformations* between one vector space and another.

- A *transformation* from one set X to another set Y is just a function $f : X \to Y$ whose domain is X and whose range is in Y. The set of all possible transformations is extremely large. In linear algebra, however, we are not concerned with all types of transformations, but only a very special type known as *linear transformations.* These are transformations from one vector space to another which preserves the additive and scalar multiplicative structure:

- **Definition.** Let X, Y be vector spaces. A *linear transformation T from X to Y* is any transformation $T : X \to Y$ which obeys the following two properties:

- (T preserves vector addition) For any $x, x' \in X$, $T(x + x') = Tx + Tx'$.

- (T preserves scalar multiplication) For any $x \in X$ and any scalar $c \in \mathbf{R}$, $T(cx) = cTx$.

- Note that there are now two types of vectors: vectors in X and vectors in Y. In some cases, X and Y will be the same space, but other times they will not. So one should take a little care; for instance one cannot necessarily add a vector in X to a vector in Y. In the above definition, x and x' were vectors in X, so $x+x'$ used the X vector addition rule, but Tx and Tx' were vectors in Y, so $Tx+Tx'$ used the Y vector addition rule. (An expression like $x+Tx$ would not necessarily make sense, unless X and Y were equal, or at least contained inside a common vector space).

- The two properties of a linear transformation can be described as follows: if you combine two inputs, then the outputs also combine (the whole is equal to the sum of its parts); and if you amplify an input by a constant, the output also amplifies by the same constant (another way of saying this is that the transformation is homogeneous).

- To test whether a transformation is linear, you have to check separately whether it is closed under vector addition, and closed under scalar multiplication. It is possible to combine the two checks into one: if you can check that for every scalar $c \in \mathbf{R}$ and vectors $x, x' \in X$, that $T(cx + x') = cTx + Tx'$, then you are automatically a linear transformation (See homework)

- **Scalar multiplication as a linear transformation.** A very simple example of a linear transformation is the map $T : \mathbf{R} \to \mathbf{R}$ defined by $Tx := 3x$ - it maps a scalar to three times that scalar. It is clear that this map preserves addition and multiplication. An example of a non-linear transformation is the map $T : \mathbf{R} \to \mathbf{R}$ defined by $Tx := x^2$.

- **Dilations as a linear transformation** As a variation of this theme, given any vector space V, the map $T : V \to V$ given by $Tx := 3x$ is a linear transformation (why?). This transformation takes vectors and dilates them by 3.

- **The identity as a linear transformation** A special case of dilations is the dilation by 1: $Ix = x$. This is a linear transformation from V to V, known as the *identity transformation*, and is usually called I or I_V.

- **Zero as a linear transformation** Another special case is dilation by 0: $Tx = 0$. This is a linear transformation from V to V, called the *zero transformation.*

- Another example of a linear transformation is the map $T : \mathbf{R}^2 \to \mathbf{R}^3$ defined by
$$Tx := \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} x,$$
where we temporarily think of the vectors in $\mathbf{R}^2$ and $\mathbf{R}^3$ as column vectors. In other words,
$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{pmatrix}.$$

- Let's check that T preserves vector addition. If x, x' are two vectors in $\mathbf{R}^2$, say
$$x := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \quad x' := \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}$$
then
$$T(x + x') = T \begin{pmatrix} x_1 + x_1' \\ x_2 + x_2' \end{pmatrix}$$
$$= \begin{pmatrix} (x_1 + x_1') + 2(x_2 + x_2') \\ 3(x_1 + x_1') + 4(x_2 + x_2') \\ 5(x_1 + x_1') + 6(x_2 + x_2') \end{pmatrix}$$
while
$$Tx + Tx' = T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + T \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}$$
$$= \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{pmatrix} + \begin{pmatrix} x_1' + 2x_2' \\ 3x_1' + 4x_2' \\ 5x_1' + 6x_2' \end{pmatrix}.$$
One can then see by inspection that $T(x + x')$ and $Tx + Tx'$ are equal. A similar computation shows that $T(cx) = cTx$; we leave this as an exercise.

- More generally, any $m \times n$ matrix (m rows and n columns) gives rise to a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$. Later on, we shall see that the converse is true: every linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$ is given by a $m \times n$ matrix. For instance, the transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ given by $Tx := 5x$ corresponds to the matrix

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

(why?), while the identity transformation on $\mathbf{R}^3$ corresponds to the *identity matrix*

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

(why?). (What matrix does the zero transformation correspond to?)

- Thus matrices provide a good example of linear transformations; but they are not the only type of linear transformation (just as row and column vectors are not the only type of vectors we study). We now give several more examples.

- **Reflections as linear transformations** Let $\mathbf{R}^2$ be the plane, and let $T : \mathbf{R}^2 \to \mathbf{R}^2$ denote the operation of reflection through the x-axis:

  $$T(x_1, x_2) := (x_1, -x_2).$$

  (Now we once again view vectors in $\mathbf{R}^n$ as row vectors). It is straightforward to verify that this is a linear transformation; indeed, it corresponds to the matrix

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

  (why? - note we are confusing row and column vectors here. We will clear this confusion up later.). More generally, given any line in $\mathbf{R}^2$ through the origin (or any plane in $\mathbf{R}^3$ through the origin), the operation of reflection through that line (resp. plane) is a linear transformation from $\mathbf{R}^2$ to $\mathbf{R}^2$ (resp. $\mathbf{R}^3$ to $\mathbf{R}^3$), as can be seen by elementary geometry.

- **Rotations as linear transformations** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ denote the operation of rotation anticlockwise by 90 degrees. A little geometry shows that

  $$T(x_1, x_2) := (-x_2, x_1).$$

This is a linear transformation, corresponding to the matrix

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

More generally, given any angle θ, the rotation anticlockwise or clockwise around the origin gives rise to a linear transformation from $\mathbf{R}^2$ to $\mathbf{R}^2$. In $\mathbf{R}^3$, it doesn't quite make sense to rotate around the origin (which way would it spin?), but given any line through the origin (called the *axis of rotation*), one can rotate around that line by an angle θ (though there are two ways one can do it, clockwise or anticlockwise). We will not cover rotation and reflection matrices in detail here - that's a topic for 115B.

- **Permutation as a linear transformation** Let's take a standard vector space, say $\mathbf{R}^4$, and consider the operation of switching the first and third components:

$$T(x_1, x_2, x_3, x_4) = (x_3, x_2, x_1, x_4).$$

This is a linear transformation (why?) It corresponds to the matrix

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

(why?). This type of operation - the rearranging of the co-ordinates - is known as a *permutation*, and the corresponding matrix is known as a *permutation matrix.* One property of permutation matrices is that every row and column contains exactly one 1, with the rest of the entries being 0.

- **Differentiation as a linear transformation** Here's a more interesting transformation: Consider the transformation $T : P_n(\mathbf{R}) \to P_{n-1}(\mathbf{R})$ defined by differentiation:

$$Tf := \frac{df}{dx}.$$

Thus, for instance, if $n = 3$, then T would send the vector $x^3 + 2x + 4 \in P_3(\mathbf{R})$ to the vector $3x^2 + 2 \in P_2(\mathbf{R})$. To show that T preserves vector addition, pick two polynomials f, g in $\mathbf{R}$. We have to show that $T(f + g) = Tf + Tg$, i.e.

$$\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}.$$

But this is just the sum rule for differentiation. A similar argument shows that T preserves scalar multiplication.

- **The right-shift as a linear transformation** Recall that $\mathbf{R}^\infty$ is the space of all sequences, e.g. R^∞ contains

  $$(x_1, x_2, x_3, x_4, \ldots)$$

  as a typical vector. Define the *right-shift* operator $U : \mathbf{R}^\infty \to \mathbf{R}^\infty$ by

  $$U(x_1, x_2, x_3, x_4, \ldots) := (0, x_1, x_2, x_3, x_4, \ldots)$$

  i.e. we shift all the entries right by one, and add a zero at the beginning. This is a linear transformation (why?). However, it cannot be represented by a matrix since $\mathbf{R}^\infty$ is infinite dimensional (unless you are willing to consider infinite-dimensional matrices, but that is another story).

- **The left-shift as a linear transformation** There is a companion operator to the right-shift, namely the *left-shift* operator $U^* : \mathbf{R}^\infty \to \mathbf{R}^\infty$ defined by

  $$U^*(x_1, x_2, x_3, x_4, \ldots) := (x_2, x_3, x_4, \ldots),$$

  i.e. we shift all the entries left by one, with the x_1 entry disappearing entirely. It is almost, but not quite, the inverse of the right-shift operator; more on this later.

- **Inclusion as a linear transformation** Strictly speaking, the spaces $\mathbf{R}^3$ and $\mathbf{R}^2$ are not related: $\mathbf{R}^2$ is not a subspace of $\mathbf{R}^3$, because two-dimensional vectors are not three-dimensional vectors. Nevertheless, we can "force" $\mathbf{R}^2$ into $\mathbf{R}^3$ by adding an extra zero on the end of each two-dimensional vector. The formal way of doing this is introducing the linear transformation $\iota : \mathbf{R}^2 \to \mathbf{R}^3$ defined by

$$\iota(x_1, x_2) := (x_1, x_2, 0).$$

Thus $\mathbf{R}^2$ is not directly contained in $\mathbf{R}^3$, but we can make a linear transformation which embeds $\mathbf{R}^2$ into $\mathbf{R}^3$ anyway via the transformation ι, which is often called an "inclusion" or "embedding" transformation. The transformation ι corresponds to the matrix

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

- **Projection as a linear transformation** Conversely, we can squish a three-dimensional vector into a two-dimensional one by leaving out the third component. More precisely, we may consider the linear transformation $\pi : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

  $$\pi(x_1, x_2, x_3) := (x_1, x_2).$$

  This is a linear transformation (why?). It is almost, but not quite, the inverse of ι; more on this later.

- **Conversions as a linear transformation** Linear transformations arise naturally when *converting* one type of unit to another. A simple example is, say, converting yards to feet: x yards becomes $3x$ feet, thus demonstrating the linear transformation $Tx = 3x$. A more sophisticated example comes from converting a number of atoms - let's take hydrogen, carbon, and oxygen - to elementary particles (electrons, protons, and neutrons). Let's say that the vector (N_H, N_C, N_O) represents the number of hydrogen, carbon, and oxygen atoms in a compound, and (N_e, N_p, N_n) represents the number of electrons, protons, and neutrons. Since hydrogen consists of one proton and one electron, carbon consists of six protons, six neutrons, and six electrons, and oxygen consists of eight protons, eight neutrons, and eight electrons, the conversion formula is

  $$\begin{aligned} N_e &= N_H + 6N_C + 8N_O \\ N_p &= N_H + 6N_C + 8N_O \\ N_n &= 6N_C + 8N_O \end{aligned}$$

or in other words

$$\begin{pmatrix} N_e \\ N_p \\ N_n \end{pmatrix} = \begin{pmatrix} 1 & 6 & 8 \\ 1 & 6 & 8 \\ 0 & 6 & 8 \end{pmatrix} \begin{pmatrix} N_H \\ N_C \\ N_O \end{pmatrix}.$$

The matrix $\begin{pmatrix} 1 & 6 & 8 \\ 1 & 6 & 8 \\ 0 & 6 & 8 \end{pmatrix}$ is thus the *conversion matrix* from the hydrogen-carbon-oxygen vector space to the electron-proton-neutron vector space. (A philosophical question: why are conversions always linear?)

- **Population growth as a linear transformation** Linear transformations are well adapted to handle the growth of heterogeneous populations - populations consisting of more than one type of species or creature. A basic example is that of Fibonacci's rabbits. These are pairs of rabbits which reach maturity after one year, and then produce one pair of juvenile rabbits for every year after that. Thus, if at one year there are A pairs of juvenile rabbits and B pairs of adult rabbits, in the next year there will be B pairs of juvenile rabbits (because each pair of adult rabbits gives birth to a juvenile pair), and $A + B$ pairs of adult rabbits. Thus one can describe the passage of one year by a linear transformation:

  $$T(A, B) := (B, A + B).$$

  Thus, for instance, if in the first year there is one pair of juvenile rabbits, $(1, 0)$, in the next year the population vector will be $T(1, 0) = (0, 1)$. Then in the year after that it will be $T(0, 1) = (1, 1)$. Then $T(1, 1) = (1, 2)$, then $T(1, 2) = (2, 3)$, then $T(2, 3) = (3, 5)$, and so forth. (We will return to this example and analyze it more carefully much later in this course).

- **Electrical circuits as a linear transformation** Many examples of analog electric circuits, such as amplifiers, capacitors and filters, can be thought of as linear transformations: they take in some input (either a voltage or a current) and return an output (also a voltage or a current). Often the input is not a scalar, but is a function of time (e.g. for AC circuits), and similarly for the output. Thus a circuit can be viewed as a transformation from $\mathcal{F}(\mathbf{R}, \mathbf{R})$ (which represents the input as a function or time) to $\mathcal{F}(\mathbf{R}, \mathbf{R})$ (which represents the output as a function of time). Usually this transformation is linear, provided that your input is below a certain threshhold. (Too much current or voltage and your circuit might blow out or short-circuit - both very non-linear effects!). To actually write down what this transformation is mathematically, though, one usually has to solve a differential equation; this is important stuff, but is beyond the scope of this course.

- As you can see, linear transformations exist in all sorts of fields. (You may amuse yourself by finding examples of linear transformations in finance, physics, computer science, etc.)