Math 115A - Week 1
Textbook sections: 1.1-1.6
Topics covered:

- What is Linear algebra?
- Overview of course
- What is a vector? What is a vector space?
- Examples of vector spaces
- Vector subspaces
- Span, linear dependence, linear independence
- Systems of linear equations
- Bases

* * * * *

Overview of course

- This course is an introduction to *Linear algebra*. Linear algebra is the study of *linear transformations* and their algebraic properties.
- A *transformation* is any operation that transforms an input to an output. A transformation is *linear* if (a) every amplification of the input causes a corresponding amplification of the output (e.g. doubling of the input causes a doubling of the output), and (b) adding inputs together leads to adding of their respective outputs. [We'll be more precise about this much later in the course.]
- A simple example of a linear transformation is the map $y := 3x$, where the input x is a real number, and the output y is also a real number. Thus, for instance, in this example an input of 5 units causes an output of 15 units. Note that a doubling of the input causes a doubling of the output, and if one adds two inputs together (e.g. add a 3-unit input with a 5-unit input to form a 8-unit input) then the respective outputs

(9-unit and 15-unit outputs, in this example) also add together (to form a 24-unit output). Note also that the graph of this linear transformation is a straight line (which is where the term *linear* comes from).

- (Footnote: I use the symbol $:=$ to mean "is defined as", as opposed to the symbol $=$, which means "is equal to". (It's similar to the distinction between the symbols $=$ and $==$ in computer languages such as $C++$, or the distinction between causation and correlation). In many texts one does not make this distinction, and uses the symbol $=$ to denote both. In practice, the distinction is too fine to be really important, so you can ignore the colons and read $:=$ as $=$ if you want.)

- An example of a *non-linear* transformation is the map $y := x^2$; note now that doubling the input leads to quadrupling the output. Also if one adds two inputs together, their outputs do not add (e.g. a 3-unit input has a 9-unit output, and a 5-unit input has a 25-unit output, but a combined $3 + 5$-unit input does not have a $9 + 25 = 34$-unit output, but rather a 64-unit output!). Note the graph of this transformation is very much non-linear.

- In real life, most transformations are non-linear; however, they can often be *approximated* accurately by a linear transformation. (Indeed, this is the whole point of differential calculus - one takes a non-linear function and approximates it by a tangent line, which is a linear function). This is advantageous because linear transformations are much easier to study than non-linear transformations.

- In the examples given above, both the input and output were *scalar* quantities - they were described by a single number. However in many situations, the input or the output (or both) is not described by a single number, but rather by several numbers; in which case the input (or output) is not a *scalar*, but instead a *vector*. [This is a slight oversimplification - more exotic examples of input and output are also possible when the transformation is non-linear.]

- A simple example of a vector-valued linear transformation is given by Newton's second law

$$F = ma, \text{ or equivalently } a = F/m.$$

One can view this law as a statement that a force F applied to an object of mass m causes an acceleration a, equal to $a := F/m$; thus F can be viewed as an input and a as an output. Both F and a are vectors; if for instance F is equal to 15 Newtons in the East direction plus 6 Newtons in the North direction (i.e. $F := (15, 6)N$), and the object has mass $m := 3kg$, then the resulting acceleration is the vector $a = (5, 2)m/s^2$ (i.e. $5m/s^2$ in the East direction plus $2m/s^2$ in the North direction).

- Observe that even though the input and outputs are now vectors in this example, this transformation is still linear (as long as the mass stays constant); doubling the input force still causes a doubling of the output acceleration, and adding two forces together results in adding the two respective accelerations together.

- One can write Newton's second law in co-ordinates. If we are in three dimensions, so that $F := (F_x, F_y, F_z)$ and $a := (a_x, a_y, a_z)$, then the law can be written as
$$F_x = \frac{1}{m}a_x + 0a_y + 0a_z$$
$$F_y = 0a_x + \frac{1}{m}a_y + 0a_z$$
$$F_z = 0a_x + 0a_y + \frac{1}{m}a_z.$$
This linear transformation is associated to the *matrix*
$$\begin{pmatrix} \frac{1}{m} & 0 & 0 \\ 0 & \frac{1}{m} & 0 \\ 0 & 0 & \frac{1}{m} \end{pmatrix}.$$

- Here is another example of a linear transformation with vector inputs and vector outputs:
$$y_1 = 3x_1 + 5x_2 + 7x_3$$
$$y_2 = 2x_1 + 4x_2 + 6x_3;$$
this linear transformation corresponds to the matrix
$$\begin{pmatrix} 3 & 5 & 7 \\ 2 & 4 & 6 \end{pmatrix}.$$

As it turns out, every linear transformation corresponds to a matrix, although if one wants to split hairs the two concepts are not quite the same thing. [Linear transformations are to matrices as concepts are to words; different languages can encode the same concept using different words. We'll discuss linear transformations and matrices much later in the course.]

- Linear algebra is the study of the *algebraic properties* of linear transformations (and matrices). *Algebra* is concerned with how to manipulate symbolic combinations of objects, and how to equate one such combination with another; e.g. how to simplify an expression such as $(x-3)(x+5)$. In linear algebra we shall manipulate not just scalars, but also vectors, vector spaces, matrices, and linear transformations. These manipulations will include familiar operations such as addition, multiplication, and reciprocal (multiplicative inverse), but also new operations such as span, dimension, transpose, determinant, trace, eigenvalue, eigenvector, and characteristic polynomial. [Algebra is distinct from other branches of mathematics such as *combinatorics* (which is more concerned with counting objects than equating them) or *analysis* (which is more concerned with estimating and approximating objects, and obtaining qualitative rather than quantitative properties).]

* * * * *

Overview of course

- Linear transformations and matrices are the focus of this course. However, before we study them, we first must study the more basic concepts of vectors and vector spaces; this is what the first two weeks will cover. (You will have had some exposure to vectors in 32AB and 33A, but we will need to review this material in more depth - in particular we concentrate much more on concepts, theory and proofs than on computation). One of our main goals here is to understand how a small set of vectors (called a *basis*) can be used to describe all other vectors in a vector space (thus giving rise to a *co-ordinate system* for that vector space).

- In weeks 3-5, we will study linear transformations and their *co-ordinate representation* in terms of matrices. We will study how to multiply two transformations (or matrices), as well as the more difficult question of how to invert a transformation (or matrix). The material from weeks 1-5 will then be tested in the midterm for the course.

- After the midterm, we will focus on matrices. A general matrix or linear transformation is difficult to visualize directly, however one can understand them much better if they can be *diagonalized*. This will force us to understand various statistics associated with a matrix, such as determinant, trace, characteristic polynomial, eigenvalues, and eigenvectors; this will occupy weeks 6-8.

- In the last three weeks we will study *inner product spaces*, which are a fancier version of vector spaces. (Vector spaces allow you to add and scalar multiply vectors; inner product spaces also allow you to compute lengths, angles, and inner products). We then review the earlier material on bases using inner products, and begin the study of how linear transformations behave on inner product spaces. (This study will be continued in 115B).

- Much of the early material may seem familiar to you from previous courses, but I definitely recommend that you still review it carefully, as this will make the more difficult later material much easier to handle.

* * * * *

What is a vector? What is a vector space?

- We now review what a vector is, and what a vector space is. First let us recall what a *scalar* is.

- Informally, a *scalar* is any quantity which can be described by a single number. An example is mass: an object has a mass of m kg for some real number m. Other examples of scalar quantities from physics include charge, density, speed, length, time, energy, temperature, volume, and pressure. In finance, scalars would include money, interest rates, prices, and volume. (You can think up examples of scalars in chemistry, EE, mathematical biology, or many other fields).

- The set of all scalars is referred to as the *field of scalars*; it is usually just **R**, the field of real numbers, but occasionally one likes to work with other fields such as **C**, the field of complex numbers, or **Q**, the field of rational numbers. However in this course the field of scalars will almost always be **R**. (In the textbook the scalar field is often denoted **F**, just to keep aside the possibility that it might not be the reals **R**; but I will not bother trying to make this distinction.)

- Any two scalars can be added, subtracted, or multiplied together to form another scalar. Scalars obey various rules of algebra, for instance $x + y$ is always equal to $y + x$, and $x * (y + z)$ is equal to $x * y + x * z$.

- Now we turn to vectors and vector spaces. Informally, a vector is any member of a vector space; a vector space is any class of objects which can be added together, or multiplied with scalars. (A more popular, but less mathematically accurate, definition of a vector is any quantity with both direction and magnitude. This is true for some common kinds of vectors - most notably physical vectors - but is misleading or false for other kinds). As with scalars, vectors must obey certain rules of algebra.

- Before we give the formal definition, let us first recall some familiar examples.

- The vector space $\mathbf{R}^2$ is the space of all vectors of the form (x, y), where x and y are real numbers. (In other words, $\mathbf{R}^2 := \{(x, y) : x, y \in \mathbf{R}\}$). For instance, $(-4, 3.5)$ is a vector in $\mathbf{R}^2$. One can add two vectors in $\mathbf{R}^2$ by adding their components separately, thus for instance $(1, 2)+(3, 4) = (4, 6)$. One can multiply a vector in $\mathbf{R}^2$ by a scalar by multiplying each component separately, thus for instance $3 * (1, 2) = (3, 6)$. Among all the vectors in $\mathbf{R}^2$ is the *zero vector* $(0, 0)$. Vectors in $\mathbf{R}^2$ are used for many physical quantities in two dimensions; they can be represented graphically by arrows in a plane, with addition represented by the parallelogram law and scalar multiplication by dilation.

- The vector space $\mathbf{R}^3$ is the space of all vectors of the form (x, y, z), where x, y, z are real numbers: $\mathbf{R}^3 := \{(x, y, z) : x, y, z \in \mathbf{R}\}$. Addition and scalar multiplication proceeds similar to $\mathbf{R}^2$: $(1, 2, 3) + (4, 5, 6) = (5, 7, 9)$, and $4 * (1, 2, 3) = (4, 8, 12)$. However, addition of a vector in $\mathbf{R}^2$ to a vector in $\mathbf{R}^3$ is undefined; $(1, 2) + (3, 4, 5)$ doesn't make sense.

Among all the vectors in $\mathbf{R}^3$ is the *zero vector* $(0,0,0)$. Vectors in $\mathbf{R}^3$ are used for many physical quantities in three dimensions, such as velocity, momentum, current, electric and magnetic fields, force, acceleration, and displacement; they can be represented by arrows in space.

- One can similarly define the vector spaces $\mathbf{R}^4$, $\mathbf{R}^5$, etc. Vectors in these spaces are not often used to represent physical quantities, and are more difficult to represent graphically, but are useful for describing populations in biology, portfolios in finance, or many other types of quantities which need several numbers to describe them completely.

* * * * *

Definition of a vector space

- **Definition.** A *vector space* is any collection V of objects (called *vectors*) for which two operations can be performed:

- *Vector addition*, which takes two vectors v and w in V and returns another vector $v + w$ in V. (Thus V must be closed under addition).

- *Scalar multiplication*, which takes a scalar c in $\mathbf{R}$ and a vector v in V, and returns another vector cv in V. (Thus V must be closed under scalar multiplication).

- Furthermore, for V to be a vector space, the following properties must be satisfied:

- (I. Addition is commutative) For all $v, w \in V$, $v + w = w + v$.

- (II. Addition is associative) For all $u, v, w \in V$, $u+(v+w) = (u+v)+w$.

- (III. Additive identity) There is a vector $0 \in V$, called the *zero vector*, such that $0 + v = v$ for all $v \in V$.

- (IV. Additive inverse) For each vector $v \in V$, there is a vector $-v \in V$, called the *additive inverse* of v, such that $-v + v = 0$.

- (V. Multiplicative identity) The scalar 1 has the property that $1v = v$ for all $v \in V$.

- (VI. Multiplication is associative) For any scalars $a, b \in \mathbf{R}$ and any vector $v \in V$, we have $a(bv) = (ab)v$.
- (VII. Multiplication is linear) For any scalar $a \in \mathbf{R}$ and any vectors $v, w \in V$, we have $a(v + w) = av + aw$.
- (VIII. Multiplication distributes over addition) For any scalars $a, b \in \mathbf{R}$ and any vector $v \in V$, we have $(a + b)v = av + bv$.

* * * * *

(Not very important) Remarks

- The number of properties listed is long, but they can be summarized briefly as: the laws of algebra work! They are all eminently reasonable; one would not want to work with vectors for which $v + w \neq w + v$, for instance. Verifying all the vector space axioms seems rather tedious, but later we will see that in most cases we don't need to verify all of them.
- Because addition is associative (axiom II), we will often write expressions such as $u + v + w$ without worrying about which order the vectors are added in. Similarly from axiom VI we can write things like abv. We also write $v - w$ as shorthand for $v + (-w)$.
- A philosophical point: we never say exactly what vectors *are*, only what vectors *do*. This is an example of *abstraction*, which appears everywhere in mathematics (but especially in algebra): the exact substance of an object is not important, only its properties and functions. (For instance, when using the number "three" in mathematics, it is unimportant whether we refer to three rocks, three sheep, or whatever; what is important is how to add, multiply, and otherwise manipulate these numbers, and what properties these operations have). This is tremendously powerful: it means that we can use a single theory (linear algebra) to deal with many very different subjects (physical vectors, population vectors in biology, portfolio vectors in finance, probability distributions in probability, functions in analysis, etc.). [A similar philosophy underlies "object-oriented programming" in computer science.] Of course, even though vector spaces can be abstract, it is often very

helpful to keep concrete examples of vector spaces such as $\mathbf{R}^2$ and $\mathbf{R}^3$ handy, as they are of course much easier to visualize. For instance, even when dealing with an abstract vector space we shall often still just draw arrows in $\mathbf{R}^2$ or $\mathbf{R}^3$, mainly because our blackboards don't have all that many dimensions.

- Because we chose our field of scalars to be the field of real numbers $\mathbf{R}$, these vector fields are known as *real vector fields*, or *vector fields over* $\mathbf{R}$. Occasionally people use other fields, such as complex numbers $\mathbf{C}$, to define the scalars, thus creating complex vector fields (or vector fields over $\mathbf{C}$), etc. Another interesting choice is to use functions instead of numbers as scalars (for instance, one could have an indeterminate x, and let things like $4x^3 + 2x^2 + 5$ be scalars, and $(4x^3 + 2x^2 + 5, x^4 - 4)$ be vectors). We will stick almost exclusively with the real scalar field in this course, but because of the abstract nature of this theory, almost everything we say in this course works equally well for other scalar fields.

- A pedantic point: The zero vector is often denoted 0, but technically it is not the same as the zero scalar 0. But in practice there is no harm in confusing the two objects: zero of one thing is pretty much the same as zero of any other thing.

* * * * *

Examples of vector spaces

- **n-tuples as vectors.** For any integer $n \geq 1$, the vector space $\mathbf{R}^n$ is defined to be the space of all n-tuples of reals $(x_1, x_2, \ldots, x_n)$. These are ordered n-tuples, so for instance $(3, 4)$ is not the same as $(4, 3)$; two vectors are equal $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ are only equal if $x_1 = y_1$, $x_2 = y_2$, $\ldots$, and $x_n = y_n$. Addition of vectors is defined by

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

and scalar multiplication by

$$c(x_1, x_2, \ldots, x_n) := (cx_1, cx_2, \ldots, cx_n).$$

The zero vector is

$$0 := (0, 0, \ldots, 0)$$

and additive inverse is given by

$$-(x_1, x_2, \ldots, x_n) := (-x_1, -x_2, \ldots, -x_n).$$

- A typical use of such a vector is to count several types of objects. For instance, a simple ecosystem consisting of X units of plankton, Y units of fish, and Z whales might be represented by the vector (X, Y, Z). Combining two ecosystems together would then correspond to adding the two vectors; natural population growth might correspond to multiplying the vector by some scalar corresponding to the growth rate. (More complicated operations, dealing with how one species impacts another, would probably be dealt with via matrix operations, which we will come to later). As one can see, there is no reason for n to be restricted to two or three dimensions.

- The vector space axioms can be verified for $\mathbf{R}^n$, but it is tedious to do so. We shall just verify one axiom here, axiom VIII: $(a+b)v = av + bv$. We can write the vector v in the form $v := (x_1, x_2, \ldots, x_n)$. The left-hand side is then

  $$(a+b)v = (a+b)(x_1, x_2, \ldots, x_n) = ((a+b)x_1, (a+b)x_2, \ldots, (a+b)x_n)$$

  while the right-hand side is

  $$av + bv = a(x_1, x_2, \ldots, x_n) + b(x_1, x_2, \ldots, x_n)$$

  $$= (ax_1, ax_2, \ldots, ax_n) + (bx_1, bx_2, \ldots, bx_n)$$

  $$= (ax_1 + bx_1, ax_2 + bx_2, \ldots, ax_n + bx_n)$$

  and the two sides match since $(a+b)x_j = ax_j + bx_j$ for each $j = 1, 2, \ldots, n$.

- There are of course other things we can do with $\mathbf{R}^n$, such as taking dot products, lengths, angles, etc., but those operations are not common to all vector spaces and so we do not discuss them here.

- **Scalars as vectors**. The scalar field $\mathbf{R}$ can itself be thought of as a vector space - after all, it has addition and scalar multiplication. It is essentially the same space as $\mathbf{R}^1$. However, this is a rather boring vector space and it is often confusing (though technically correct) to refer to scalars as a type of vector. Just as $\mathbf{R}^2$ represents vectors in a plane and $\mathbf{R}^3$ represents vectors in space, $\mathbf{R}^1$ represents vectors in a line.

- **The zero vector space**. Actually, there is an even more boring vector space than $\mathbf{R}$ - the zero vector space $\mathbf{R}^0$ (also called $\{0\}$), consisting solely of a single vector 0, the zero vector, which is also sometimes denoted $()$ in this context. Addition and multiplication are trivial: $0 + 0 = 0$ and $c0 = 0$. The space $\mathbf{R}^0$ represents vectors in a point. Although this space is utterly uninteresting, it is necessary to include it in the pantheon of vector spaces, just as the number zero is required to complete the set of integers.

- **Complex numbers as vectors**. The space $\mathbf{C}$ of complex numbers can be viewed as a vector space over the reals; one can certainly add two complex numbers together, or multiply a complex number by a (real) scalar, with all the laws of arithmetic holding. Thus, for instance, $3+2i$ would be a vector, and an example of scalar multiplication would be $5(3+2i) = 15+10i$. This space is very similar to $\mathbf{R}^2$, although complex numbers enjoy certain operations, such as complex multiplication and complex conjugate, which are not available to vectors in $\mathbf{R}^2$.

- **Polynomials as vectors I.** For any $n \geq 0$, let $P_n(\mathbf{R})$ denote the vector space of all polynomials of one indeterminate variable x whose degree is at most n. Thus for instance $P_3(\mathbf{R})$ contains the "vectors"

  $$x^3 + 2x^2 + 4; \quad x^2 - 4; \quad -1.5x^3 + 2.5x + \pi; \quad 0$$

  but not

  $$x^4 + x + 1; \quad \sqrt{x}; \quad \sin(x) + e^x; \quad x^3 + x^{-3}.$$

  Addition, scalar multiplication, and additive inverse are defined in the standard manner, thus for instance

  $$(x^3 + 2x^2 + 4) + (-x^3 + x^2 + 4) = 3x^2 + 8 \tag{0.1}$$

  and

  $$3(x^3 + 2x^2 + 4) = 3x^3 + 6x^2 + 12.$$

  The zero vector is just 0.

- Notice in this example it does not really matter what x is. The space $P_n(\mathbf{R})$ is very similar to the vector space $\mathbf{R}^{n+1}$; indeed one can match one to the other by the pairing

$$a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \iff (a_n, a_{n-1}, \ldots, a_1, a_0),$$

  thus for instance in $P_3(\mathbf{R})$, the polynomial $x^3 + 2x^2 + 4$ would be associated with the 4-tuple $(1, 2, 0, 4)$. The more precise statement here is that $P_n(\mathbf{R})$ and $\mathbf{R}^{n+1}$ are *isomorphic* vector spaces; more on this later. However, the two spaces are still different; for instance we can do certain operations in $P_n(\mathbf{R})$, such as differentiate with respect to x, which do not make much sense for $\mathbf{R}^{n+1}$.

- Notice that we allow the polynomials to have degree less than n; if we only allowed polynomials of degree *exactly* n, then we would not have a vector space because the sum of two vectors would not necessarily be a vector (see (0.1)). (In other words, such a space would not be closed under addition).

- **Polynomials as vectors II.** Let $P(\mathbf{R})$ denote the vector space of *all* polynomials of one indeterminate variable x - regardless of degree. (In other words, $P(\mathbf{R}) := \bigcup_{n=0}^{\infty} P_n(\mathbf{R})$, the union of all the $P_n(\mathbf{R})$). Thus this space in particular contains the *monomials*

$$1, x, x^2, x^3, x^4, \ldots$$

  though of course it contains many other vectors as well.

- This space is much larger than any of the $P_n(\mathbf{R})$, and is not isomorphic to any of the standard vector spaces $\mathbf{R}^n$. Indeed, it is an *infinite dimensional space* - there are infinitely many "independent" vectors in this space. (More on this later).

- **Functions as vectors I.** Why stick to polynomials? Let $C(\mathbf{R})$ denote the vector space of all continuous functions of one real variable x - thus this space includes as vectors such objects as

$$x^4 + x + 1; \quad \sin(x) + e^x; \quad x^3 + \pi - \sin(x); \quad |x|.$$

One still has addition and scalar multiplication:

$$(\sin(x) + e^x) + (x^3 + \pi - \sin(x)) = x^3 + e^x + \pi$$

$$5(\sin(x) + e^x) = 5\sin(x) + 5e^x,$$

and all the laws of vector spaces still hold. This space is substantially larger than $P(\mathbf{R})$, and is another example of an infinite dimensional vector space.

- **Functions as vectors II.** In the previous example the real variable x could range over all the real line $\mathbf{R}$. However, we could instead restrict the real variable to some smaller set, such as the interval $[0, 1]$, and just consider the vector space $C([0, 1])$ of continuous functions on $[0, 1]$. This would include such vectors such as

$$x^4 + x + 1; \quad \sin(x) + e^x; \quad x^3 + \pi - \sin(x); \quad |x|.$$

  This looks very similar to $C(\mathbf{R})$, but this space is a bit smaller because more functions are equal. For instance, the functions x and $|x|$ are the same vector in $C([0, 1])$, even though they are different vectors in $C(\mathbf{R})$.

- **Functions as vectors III.** Why stick to continuous functions? Let $\mathcal{F}(\mathbf{R}, \mathbf{R})$ denote the space of *all* functions of one real variable $\mathbf{R}$, regardless of whether they are continuous or not. In addition to all the vectors in $C(\mathbf{R})$ the space $\mathcal{F}(\mathbf{R}, \mathbf{R})$ contains many strange objects, such as the function

$$f(x) := \begin{cases} 1 & \text{if } x \in \mathbf{Q} \\ 0 & \text{if } x \notin \mathbf{Q} \end{cases}$$

  This space is much, much, larger than $C(\mathbf{R})$; it is also infinite dimensional, but it is in some sense "more infinite" than $C(\mathbf{R})$. (More precisely, the dimension of $C(\mathbf{R})$ is countably infinite, but the dimension of $\mathcal{F}(\mathbf{R}, \mathbf{R})$ is uncountably infinite. Further discussion is beyond the scope of this course, but see Math 112).

- **Functions as vectors IV.** Just as the vector space $C(\mathbf{R})$ of continuous functions can be restricted to smaller sets, the space $\mathcal{F}(\mathbf{R}, \mathbf{R})$ can also be restricted. For any subset S of the real line, let $\mathcal{F}(S, \mathbf{R})$ denote the vector space of all functions from S to $\mathbf{R}$, thus a vector in this space is a function f which assigns a real number $f(x)$ to each x in S. Two vectors f, g would be considered equal if $f(x) = g(x)$ for each x in S. For instance, if S is the two element set $S := \{0, 1\}$, then the two functions $f(x) := x^2$ and $g(x) := x$ would be considered the same vector in $\mathcal{F}(\{0,1\}, \mathbf{R})$, because they equal the same value at 0 and 1. Indeed, to specify any vector f in $\{0, 1\}$, one just needs to specify $f(0)$ and $f(1)$. As such, this space is very similar to $\mathbf{R}^2$.

- **Sequences as vectors.** An *infinite sequence* is a sequence of real numbers
$$(a_1, a_2, a_3, a_4, \ldots);$$
for instance, a typical sequence is
$$(2, 4, 6, 8, 10, 12, \ldots).$$
Let $\mathbf{R}^\infty$ denote the vector space of all infinite sequences. These sequences are added together by the rule
$$(a_1, a_2, \ldots) + (b_1, b_2, \ldots) := (a_1 + b_1, a_2 + b_2, \ldots)$$
and scalar multiplied by the rule
$$c(a_1, a_2, \ldots) := (ca_1, ca_2, \ldots).$$
This vector space is very much like the finite-dimensional vector spaces $\mathbf{R}^2$, $\mathbf{R}^3$, ..., except that these sequences do not terminate.

- **Matrices as vectors.** Given any integers $m, n \geq 1$, we let $M_{m\times n}(\mathbf{R})$ be the space of all $m \times n$ matrices (i.e. m rows and n columns) with real entries, thus for instance $M_{2\times 3}$ contains such "vectors" as
$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 & -2 \\ -3 & -4 & -5 \end{pmatrix}.$$
Two matrices are equal if and only if all of their individual components match up; rearranging the entries of a matrix will produce a different

matrix. Matrix addition and scalar multiplication is defined similarly to vectors:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 0 & -1 & -2 \\ -3 & -4 & -5 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$10 \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix}.$$

Matrices are useful for many things, notably for solving linear equations and for encoding linear transformations; more on these later in the course.

- As you can see, there are (infinitely!) many examples of vector spaces, some of which look very different from the familiar examples of $\mathbf{R}^2$ and $\mathbf{R}^3$. Nevertheless, much of the theory we do here will cover all of these examples simultaneously. When we depict these vector spaces on the blackboard, we will draw them as if they were $\mathbf{R}^2$ or $\mathbf{R}^3$, but they are often much larger, and each point we draw in the vector space, which represents a vector, could in reality stand for a very complicated object such as a polynomial, matrix, or function. So some of the pictures we draw should be interpreted more as analogies or metaphors than as a literal depiction of the situation.

* * * * *

Non-vector spaces

- Now for some examples of things which are *not* vector spaces.
- **Latitude and longitude.** The location of any point on the earth can be described by two numbers, e.g. Los Angeles is 34 N, 118 W. This may look a lot like a two-dimensional vector in $\mathbf{R}^2$, but the space of all latitude-longitude pairs is not a vector space, because there is no reasonable way of adding or scalar multiplying such pairs. For instance, how could you multiply Los Angeles by 10? 340 N, 1180 W does not make sense.
- **Unit vectors.** In $\mathbf{R}^3$, a *unit vector* is any vector with unit length, for instance $(0, 0, 1)$, $(0, -1, 0)$, and $(\frac{3}{5}, 0, \frac{4}{5})$ are all unit vectors. However the space of all unit vectors (sometimes denoted S^2, for two-dimensional sphere) is not a vector space as it is not closed under addition (or under scalar multiplication).

- **The positive real axis.** The space $\mathbf{R}^+$ of positive real numbers is closed under addition, and obeys most of the rules of vector spaces, but is not a vector space, because one cannot multiply by negative scalars. (Also, it does not contain a zero vector).

- **Monomials.** The space of monomials $1, x, x^2, x^3, \ldots$ does not form a vector space - it is not closed under addition or scalar multiplication.

$$* * * * *$$

Vector arithmetic

- The vector space axioms I-VIII can be used to deduce all the other familiar laws of vector arithmetic. For instance, we have

- **Vector cancellation law** If u, v, w are vectors such that $u+v = u+w$, then $v = w$.

- **Proof:** Since u is a vector, we have an additive inverse $-u$ such that $-u+u = 0$, by axiom IV. Now we add $-u$ to both sides of $u+v = u+w$:

$$-u + (u + v) = -u + (u + w).$$

  Now use axiom II:

$$(-u + u) + v = (-u + u) + w$$

  then axiom IV:

$$0 + v = 0 + w$$

  then axiom III:

$$v = w.$$

- As you can see, these algebraic manipulations are rather trivial. After the first week we usually won't do these computations in such painful detail.

- Some other simple algebraic facts, which you can amuse yourself with by deriving them from the axioms:

$$0v = 0; \quad (-1)v = -v; \quad -(v+w) = (-v)+(-w); \quad a0 = 0; \quad a(-x) = (-a)x = -ax$$

* * * * *

Vector subspaces

- Many vector spaces are *subspaces* of another. A vector space W is a *subspace* of a vector space V if $W \subseteq V$ (i.e. every vector in V is also a vector in W), and the laws of vector addition and scalar multiplication are consistent (i.e. if v_1 and v_2 are in W, and hence in V, the rule that W gives for adding v_1 and v_2 gives the same answer as the rule that V gives for adding v_1 and v_2.

- For instance, the space $P_2(\mathbf{R})$ - the vector space of polynomials of degree at most 2 is a subspace of $P_3(\mathbf{R})$. Both are subspaces of $P(\mathbf{R})$, the vector space of polynomials of arbitrary degree. $C([0,1])$, the space of continuous functions on $[0,1]$, is a subspace of $\mathcal{F}([0,1],\mathbf{R})$. And so forth. (Technically, $\mathbf{R}^2$ is not a subspace of $\mathbf{R}^3$, because a two-dimensional vector is not a three-dimensional vector. However, $\mathbf{R}^3$ does contain subspaces which are almost identical to $\mathbf{R}^2$. More on this later).

- If V is a vector space, and W is a subset of V (i.e. $W \subseteq V$), then of course we can add and scalar multiply vectors in W, since they are automatically vectors in V. On the other hand, W is not necessarily a subspace, because it may not be a vector space. (For instance, the set S^2 of unit vectors in $\mathbf{R}^3$ is a sub*set* of $\mathbf{R}^3$, but is not a sub*space*). However, it is easy to check when a subset is a subspace:

- **Lemma.** Let V be a vector space, and let W be a subset of V. Then W is a subspace of V if and only if the following two properties are satisfied:

- (W is closed under addition) If w_1 and w_2 are in W, then $w_1 + w_2$ is also in W.

- (W is closed under scalar multiplication) If w is in W and c is a scalar, then cw is also in W.

- **Proof.** First suppose that W is a subspace of V. Then W will be closed under addition and multiplication directly from the definition of vector space. This proves the "only if" part.

- Now we prove the harder "if part". In other words, we assume that W is a subset of V which is closed under addition and scalar multiplication, and we have to prove that W is a vector space. In other words, we have to verify the axioms I-VIII.

- Most of these axioms follow immediately because W is a subset of V, and V already obeys the axioms I-VIII. For instance, since vectors v_1, v_2 in V obey the commutativity property $v_1+v_2 = v_2+v_1$, it automatically follows that vectors in W also obey the property $w_1 + w_2 = w_2 + w_1$, since all vectors in W are also vectors in V. This reasoning easily gives us axioms I, II, V, VI, VII, VIII.

- There is a potential problem with III though, because the zero vector 0 of V might not lie in W. Similarly with IV, there is a potential problem that if w lies in W, then $-w$ might not lie in W. But both problems cannot occur, because $0 = 0w$ and $-w = (-1)w$ (Exercise: prove this from the axioms!), and W is closed under scalar multiplication. $\square$

- This Lemma makes it quite easy to generate a large number of vector spaces, simply by taking a big vector space and passing to a subset which is closed under addition and scalar multiplication. Some examples:

- (**Horizontal vectors**) Recall that $\mathbf{R}^3$ is the vector space of all vectors (x, y, z) with x, y, z real. Let V be the subset of $\mathbf{R}^3$ consisting of all vectors with zero z co-ordinate, i.e. $V := \{(x, y, 0) : x, y \in \mathbf{R}\}$. This is a subset of $\mathbf{R}^3$, but moreover it is also a subspace of $\mathbf{R}^3$. To see this, we use the Lemma. It suffices to show that V is closed under vector addition and scalar multiplication. Let's check the vector addition. If we have two vectors in V, say $(x_1, y_1, 0)$ and $(x_2, y_2, 0)$, we need to verify that the sum of these two vectors is still in V. But the sum is just $(x_1 + x_2, y_1 + y_2, 0)$, and this is in V because the z co-ordinate is zero. Thus V is closed under vector addition. A similar argument shows that V is closed under scalar multiplication, and so V is indeed a subspace of $\mathbf{R}^3$. (Indeed, V is very similar to - though technically not the same thing as - $\mathbf{R}^2$). Note that if we considered instead the space of all vectors with z co-ordinate 1, i.e. $\{(x, y, 1) : x, y \in \mathbf{R}\}$, then this would be a subset but not a subspace, because it is not closed under vector addition (or under scalar multiplication, for that matter).

- Another example of a subspace of $\mathbf{R}^3$ is the plane $\{(x, y, z) \in R^3 : x + 2y + 3z = 0\}$. A third example of a subspace of $\mathbf{R}^3$ is the line $\{(t, 2t, 3t) : t \in \mathbf{R}\}$. (Exercise: verify that these are indeed subspaces). Notice how subspaces tend to be very flat objects which go through the origin; this is consistent with them being closed under vector addition and scalar multiplication.

- In $\mathbf{R}^3$, the only subspaces are lines through the origin, planes through the origin, the whole space $\mathbf{R}^3$, and the zero vector space $\{0\}$. In $\mathbf{R}^2$, the only subspaces are lines through the origin, the whole space $\mathbf{R}^2$, and the zero vector space $\{0\}$. (This is another clue as to why this subject is called *linear* algebra).

- (**Even polynomials**) Recall that $P(\mathbf{R})$ is the vector space of all polynomials $f(x)$. Call a polynomial *even* if $f(x) = f(-x)$; for instance, $f(x) = x^4 + 2x^2 + 3$ is even, but $f(x) = x^3 + 1$ is not. Let $P_{even}(\mathbf{R})$ denote the set of all even polynomials, thus $P_{even}(\mathbf{R})$ is a subset of $P(\mathbf{R})$. Now we show that $P_{even}(\mathbf{R})$ is not just a subset, it is a subspace of $P(\mathbf{R})$. Again, it suffices to show that $P_{even}(\mathbf{R})$ is closed under vector addition and scalar multiplication. Let's show it's closed under vector addition - i.e. if f and g are even polynomials, we have to show that $f + g$ is also even. In other words, we have to show that $f(-x) + g(-x) = f(x) + g(x)$. But this is clear since $f(-x) = f(x)$ and $g(-x) = g(x)$. A similar argument shows why even polynomials are closed under scalar multiplication.

- (**Diagonal matrices**) Let $n \geq 1$ be an integer. Recall that $M_{n\times n}(\mathbf{R})$ is the vector space of $n \times n$ real matrices. Call a matrix *diagonal* if all the entries away from the main diagonal (from top left to bottom right) are zero, thus for instance

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

is a diagonal matrix. Let $D_n(\mathbf{R})$ denote the space of all diagonal $n \times n$ matrices. This is a subset of $M_{n\times n}(\mathbf{R})$, and is also a subspace, because the sum of any two diagonal matrices is again a diagonal matrix, and the scalar product of a diagonal matrix and a scalar is still a diagonal matrix. The notation of a diagonal matrix will become very useful much later in the course.

- (**Trace zero matrices**) Let $n \geq 1$ be an integer. If A is an $n \times n$ matrix, we define the *trace* of that matrix, denoted $tr(A)$, to be the sum of all the entries on the diagonal. For instance, if

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

then

$$tr(A) = 1 + 5 + 9 = 15.$$

Let $M^0_{n\times n}(\mathbf{R})$ denote the set of all $n \times n$ matrices whose trace is zero:

$$M^0_{n\times n}(\mathbf{R}) := \{A \in M_{n\times n} : tr(A) = 0\}.$$

One can easily check that this space is a subspace of $M_{n\times n}$. We will return to traces much later in this course.

- Technically speaking, every vector space V is considered a subspace of itself (since V is already closed under addition and scalar multiplication). Also the zero vector space $\{0\}$ is a subspace of every vector space (for a similar reason). But these are rather uninteresting examples of subspaces. We sometimes use the term *proper subspace* of V to denote a subspace W of V which is not the whole space V or the zero vector space $\{0\}$, but instead is something in between.

- The intersection of two subspaces is again a subspace (why?). For instance, since the diagonal matrices $D_n(\mathbf{R})$ and the trace zero matrices $M^0_{n\times n}(\mathbf{R})$ are both subspaces of $M_{n\times n}(\mathbf{R})$, their intersection $D_n(\mathbf{R}) \cap M^0_{n\times n}(\mathbf{R})$ is also a subspace of $M_{n\times n}(\mathbf{R})$. On the other hand, the union of two subspaces is usually not a subspace. For instance, the x-axis $\{(x, 0) : x \in \mathbf{R}\}$ and y-axis $\{(0, y) : y \in \mathbf{R}\}$, but their union $\{(x, 0) : x \in \mathbf{R}\} \cup \{(0, y) : y \in \mathbf{R}\}$ is not (why?). See Assignment 1 for more details.

- In some texts one uses the notation $W \leq V$ to denote the statement "W is a subspace of V". I'll avoid this as it may be a little confusing at first. However, the notation is suggestive. For instance it is true that if $U \leq W$ and $W \leq V$, then $U \leq V$; i.e. if U is a subspace of W, and W is a subspace of V, then U is a subspace of V. (Why?)

* * * * *

Linear combinations

- Let's look at the standard vector space $\mathbf{R}^3$, and try to build some subspaces of this space. To get started, let's pick a random vector in $\mathbf{R}^3$, say $v := (1, 2, 3)$, and ask how to make a subspace V of $\mathbf{R}^3$ which would contain this vector $(1, 2, 3)$. Of course, this is easy to accomplish by setting V equal to all of $\mathbf{R}^3$; this would certainly contain our single vector v, but that is overkill. Let's try to find a smaller subspace of $\mathbf{R}^3$ which contains v.

- We could start by trying to make V just consist of the single point $(1, 2, 3)$: $V := \{(1, 2, 3)\}$. But this doesn't work, because this space is not a vector space; it is not closed under scalar multiplication. For instance, $10(1, 2, 3) = (10, 20, 30)$ is not in the space. To make V a vector space, we cannot just put $(1, 2, 3)$ into V, we must also put in all the scalar multiples of $(1, 2, 3)$: $(2, 4, 6)$, $(3, 6, 9)$, $(-1, -2, -3)$, $(0, 0, 0)$, etc. In other words,

$$V \supseteq \{a(1, 2, 3) : a \in \mathbf{R}\}.$$

  Conversely, the space $\{a(1, 2, 3) : a \in \mathbf{R}\}$ is indeed a subspace of $\mathbf{R}^3$ which contains $(1, 2, 3)$. (Exercise!). This space is the one-dimensional space which consists of the line going through the origin and $(1, 2, 3)$.

- To summarize what we've seen so far, if one wants to find a subspace V which contains a specified vector v, then it is not enough to contain v; one must also contain the vectors av for all scalars a. As we shall see later, the set $\{av : a \in \mathbf{R}\}$ will be called the *span* of v, and is denoted $\mathrm{span}(\{v\})$.

- Now let's suppose we have two vectors, $v := (1, 2, 3)$ and $w := (0, 0, 1)$, and we want to construct a vector space V in $\mathbf{R}^3$ which contains both

v and w. Again, setting V equal to all of $\mathbf{R}^3$ will work, but let's try to get away with as small a space V as we can.

- We know that at a bare minimum, V has to contain not just v and w, but also the scalar multiples av and bw of v and w, where a and b are scalars. But V must also be closed under vector addition, so it must also contain vectors such as $av+bw$. For instance, V must contain such vectors as

  $$3v + 5w = 3(1, 2, 3) + 5(0, 0, 1) = (3, 6, 9) + (0, 0, 5) = (3, 6, 14).$$

  We call a vector of the form $av + bw$ a *linear combination* of v and w, thus $(3, 6, 14)$ is a linear combination of $(1, 2, 3)$ and $(0, 0, 1)$. The space $\{av + bw : a, b \in \mathbf{R}\}$ of all linear combinations of v and w is called the *span* of v and w, and is denoted $\mathrm{span}(\{v, w\})$. It is also a subspace of $\mathbf{R}^3$; it turns out to be the plane through the origin that contains both v and w.

- More generally, we define the notions of linear combination and span as follows.

- **Definition.** Let S be a collection of vectors in a vector space V (either finite or infinite). A *linear combination* of S is defined to be any vector in V of the form

  $$a_1v_1 + a_2v_2 + \ldots + a_nv_n$$

  where $a_1, \ldots, a_n$ are scalars (possibly zero or negative), and $v_1, \ldots, v_n$ are some elements in S. The *span* of S, denoted $\mathrm{span}(S)$, is defined to be the space of all linear combinations of S:

  $$\mathrm{span}(S) := \{a_1v_1 + a_2v_2 + \ldots + a_nv_n : a_1, \ldots, a_n \in \mathbf{R}; v_1, \ldots, v_n \in S\}.$$

- Usually we deal with the case when the set S is just a finite collection

  $$S = \{v_1, \ldots, v_n\}$$

  of vectors. In that case the span is just

  $$\mathrm{span}(\{v_1, \ldots, v_n\}) := \{a_1v_1 + a_2v_2 + \ldots + a_nv_n : a_1, \ldots, a_n \in \mathbf{R}\}.$$

  (Why?)

- Occasionally we will need to deal when S is empty. In this case we set the span span($\emptyset$) of the empty set to just be $\{0\}$, the zero vector space. (Thus 0 is the only vector which is a linear combination of an empty set of vectors. This is part of a larger mathematical convention, which states that any summation over an empty set should be zero, and every product over an empty set should be 1.)

- Here are some basic properties of span.

- **Theorem.** Let S be a subset of a vector space V. Then span(S) is a subspace of V which contains S as a subset. Moreover, any subspace of V which contains S as a subset must in fact contain all of span(S).

- We shall prove this particular theorem in detail to illustrate how to go about giving a proof of a theorem such as this. In later theorems we will skim over the proofs more quickly.

- **Proof.** If S is empty then this theorem is trivial (in fact, it is rather vacuous - it says that the space $\{0\}$ contains all the elements of an empty set of vectors, and that any subspace of V which contains the elements of an empty set of vectors, must also contain $\{0\}$), so we shall assume that $n \geq 1$. We now break up the theorem into its various components.

  (a) First we check that span(S) is a subspace of V. To do this we need to check three things: that span(S) is contained in V; that it is closed under addition; and that it is closed under scalar multiplication.

  (a.1) To check that span(S) is contained in V, we need to take a typical element of the span, say $a_1v_1 + \ldots + a_nv_n$, where $a_1, \ldots, a_n$ are scalars and $v_1, \ldots, v_n \in S$, and verify that it is in V. But this is clear since $v_1, \ldots, v_n$ were already in V and V is closed under addition and scalar multiplication.

  (a.2) To check that the space span(S) is closed under vector addition, we take two typical elements of this space, say $a_1v_1 + \ldots + a_nv_n$ and $b_1v_1 + \ldots + b_nv_n$, where the a_j and b_j are scalars and $v_j \in S$ for $j = 1, \ldots n$, and verify that their sum is also in span(S). But the sum is

$$(a_1v_1 + \ldots + a_nv_n) + (b_1v_1 + \ldots + b_nv_n)$$

which can be rearranged as

$$(a_1 + b_1)v_1 + \ldots + (a_n + b_n)v_n$$

[Exercise: which of the vector space axioms I-VIII were needed in order to do this?]. But since $a_1 + b_1, \ldots, a_n + b_n$ are all scalars, we see that this is indeed in span(S).

(a.3) To check that the space span(S) is closed under scalar multiplication, we take a typical element of this space, say $a_1v_1 + \ldots a_nv_n$, and a typical scalar c. We want to verify that the scalar product

$$c(a_1v_1 + \ldots + a_nv_n)$$

is also in span$(\{v_1, \ldots, v_n\})$. But this can be rearranged as

$$(ca_1)v_1 + \ldots + (ca_n)v_n$$

(which axioms were used here?). Since $ca_1, \ldots, ca_n$ were scalars, we see that we are in span(S) as desired.

(b) Now we check that span(S) contains S. It will suffice of course to show that span(S) contains v for each $v \in S$. But each v is clearly a linear combination of elements in S, in fact $v = 1.v$ and $v \in S$. Thus v lies in span(S) as desired.

(c) Now we check that every subspace of V which contains S, also contains span(S). In order to stop from always referring to "that subspace", let us use W to denote a typical subspace of V which contains S. Our goal is to show that W contains span(S).

This the same as saying that every element of span(S) lies in W. So, let $v = a_1v_1 + \ldots + a_nv_n$ be a typical element of span(S), where the a_j are scalars and $v_j \in S$ for $j = 1, \ldots, n$. Our goal is to show that v lies in W.

Since v_1 lies in W, and W is closed under scalar multiplication, we see that a_1v_1 lies in W. Similarly $a_2v_2, \ldots, a_nv_n$ lie in W. But W is closed under vector addition, thus $a_1v_1 + \ldots + a_nv_n$ lies in W, as desired. This concludes the proof of the Theorem. $\square$

- We remark that the span of a set of vectors does not depend on what order we list the set S: for instance, $\mathrm{span}(\{u, v, w\})$ is the same as $\mathrm{span}(\{w, v, u\})$. (Why is this?)

- The span of a set of vectors comes up often in applications, when one has a certain number of "moves" available in a system, and one wants to see what options are available by combining these moves. We give a example, from a simple economic model, as follows.

- Suppose you run a car company, which uses some basic raw materials - let's say money, labor, metal, for sake of argument - to produce some cars. At any given point in time, your resources might consist of x units of money, y units of labor (measured, say, in man-hours), z units of metal, and w units of cars, which we represent by a vector (x, y, z, w).

  Now you can make various decisions to alter your balance of resources. For instance, suppose you could purchase a unit of metal for two units of money - this amounts to adding $(-2, 0, 1, 0)$ to your resource vector. You could do this repeatedly, thus adding $a(-2, 0, 1, 0)$ to your resource vector for any positive a. (If you could also sell a unit of metal for two units of money, then a could also be negative. Of course, a can always be zero, simply by refusing to buy or sell any metal). Similarly, one might be able to purchase a unit of labor for three units of money, thus adding $(-3, 1, 0, 0)$ to your resource vector. Finally, to produce a car requires 4 units of labor and 5 units of metal, thus adding $(0, -4, -5, 1)$ to your resource vector. (This is course an extremely oversimplified model, but will serve to illustrate the point).

- Now we ask the question of how much money it will cost to create a car - in other words, for what price x can we add $(-x, 0, 0, 1)$ to our resource vector? The answer is 22, because

  $$(-22, 0, 0, 1) = 5(-2, 0, 1, 0) + 4(-3, 1, 0, 0) + 1(0, -4, -5, 1)$$

  and so one can convert 22 units of money to one car by buying 5 units of metal, 4 units of labor, and producing one car. On the other hand, it is not possible to obtain a car for a smaller amount of money using the moves available (why?). In other words, $(-22, 0, 0, 1)$ is the unique vector of the form $(-x, 0, 0, 1)$ which lies in the *span* of the vectors $(-2, 0, 1, 0)$, $(-3, 1, 0, 0)$, and $(0, -4, -5, 1)$.

- Of course, the above example was so simple that we could have worked out the price of a car directly. But in more complicated situations (where there aren't so many zeroes in the vector entries) one really has to start computing the span of various vectors. [Actually, things get more complicated than this because in real life there are often other constraints. For instance, one may be able to buy labor for money, but one cannot sell labor to get the money back - so the scalar in front of $(-3, 1, 0, 0)$ can be positive but not negative. Or storage constraints might limit how much metal can be purchased at a time, etc. This passes us from linear algebra to the more complicated theory of *linear programming*, which is beyond the scope of this course. Also, due to such things as the law of diminishing returns and the law of economies of scale, in real life situations are not quite as linear as presented in this simple model. This leads us eventually to *non-linear optimization and control theory*, which is again beyond the scope of this course.]

- This leads us to ask the following question: How can we tell when one given vector v is in the span of some other vectors $v_1, v_2, \ldots v_n$? For instance, is the vector $(0, 1, 2)$ in the span of $(1, 1, 1)$, $(3, 2, 1)$, $(1, 0, 1)$? This is the same as asking for scalars a_1, a_2, a_3 such that

  $$(0, 1, 2) = a_1(1, 1, 1) + a_2(3, 2, 1) + a_3(1, 0, 1).$$

  We can multiply out the left-hand side as

  $$(a_1 + 3a_2 + a_3, a_1 + 2a_2, a_1 + a_2 + a_3)$$

  and so we are asking to find a_1, a_2, a_3 that solve the equations

  $$\begin{array}{llll} a_1 & +3a_2 & +a_3 & = 0 \\ a_1 & +2a_2 & & = 1 \\ a_1 & +a_2 & a_3 & = 2. \end{array}$$

  This is a *linear system of equations*; "system" because it consists of more than one equation, and "linear" because the variables a_1, a_2, a_3 only appear as linear factors (as opposed to quadratic factors such as a_1^2 or $a_2 a_3$, or more non-linear factors such as $\sin(a_1)$). Such a system can also be written in *matrix form*

  $$\begin{pmatrix} 1 & 3 & 1 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

or *schematically* as

$$\left(\begin{array}{ccc|c} 1 & 3 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{array}\right).$$

To actually solve this system of equations and find a_1, a_2, a_3, one of the best methods is to use *Gaussian elimination.* The idea of Gaussian elimination is to try to make as many as possible of the numbers in the matrix equal to zero, as this will make the linear system easier to solve. There are three basic moves:

- *Swap two rows*: Since it does not matter which order we display the equations of a system, we are free to swap any two rows of the system. This is mostly a cosmetic move, useful in making the system look prettier.

- *Multiply a row by a constant*: We can multiply (or divide) both sides of an equation by any constant (although we want to avoid multiplying a row by 0, as that reduces that equation to the trivial 0=0, and the operation cannot be reversed since division by 0 is illegal). This is again a mostly cosmetic move, useful for setting one of the co-efficients in the matrix to 1.

- *Subtract a multiple of one row from another*: This is the main move. One can take any row, multiply it by any scalar, and subtract (or add) the resulting object from a second row; the original row remains unchanged. The main purpose of this is to set one or more of the matrix entries of the second row to zero.

  We illustrate these moves with the above system. We could use the matrix form or the schematic form, but we shall stick with the linear system form for now:

$$\begin{array}{llll} a_1 & +3a_2 & +a_3 & = 0 \\ a_1 & +2a_2 & & = 1 \\ a_1 & +a_2 & +a_3 & = 2. \end{array}$$

We now start zeroing the a_1 entries by subtracting the first row from the second:

$$\begin{array}{llll} a_1 & +3a_2 & +a_3 & = 0 \\ & -a_2 & -a_3 & = 1 \\ a_1 & +a_2 & +a_3 & = 2 \end{array}$$

and also subtracting the first row from the third:

$$\begin{array}{llll} a_1 & +3a_2 & +a_3 & = 0 \\ & -a_2 & -a_3 & = 1 \\ & -2a_2 & & = 2. \end{array}$$

The third row looks simplifiable, so we swap it up

$$\begin{array}{llll} a_1 & +3a_2 & +a_3 & = 0 \\ & -2a_2 & & = 2 \\ & -a_2 & -a_3 & = 1 \end{array}$$

and then divide it by -2:

$$\begin{array}{llll} a_1+ & 3a_2 & +a_3 & = 0 \\ & a_2 & & = -1 \\ & -a_2 & -a_3 & = 1. \end{array}$$

Then we can zero the a_2 entries by subtracting 3 copies of the second row from the first, and adding one copy of the second row to the third:

$$\begin{array}{llll} a_1 & & +a_3 & = 3 \\ & a_2 & & = -1 \\ & & -a_3 & = 0. \end{array}$$

If we then multiply the third row by -1 and then subtract it from the first, we obtain

$$\begin{array}{llll} a_1 & & & = 3 \\ & a_2 & & = -1 \\ & & a_3 & = 0 \end{array}$$

and so we have found the solution, namely $a_1 = 3$, $a_2 = -1$, $a_3 = 0$. Getting back to our original problem, we have indeed found that $(0, 1, 2)$ is in the span of $(1, 1, 1)$, $(3, 2, 1)$, $(1, 0, 1)$:

$$(0, 1, 2) = 3(1, 1, 1) + (-1)(3, 2, 1) + 0(1, 0, 1).$$

In the above case we found that there was only one solution for a_1, a_2, a_3 - they were *exactly determined* by the linear system. Sometimes there can be more than one solution to a linear system, in which case we say that the system is *under-determined* - there are not enough equations to pin down all the variables exactly. This usually happens when the number of unknowns exceeds the number of equations. For instance, suppose we wanted to show that $(0,1,2)$ is in the span of the four vectors $(1,1,1)$, $(3,2,1)$, $(1,0,1)$, $(0,0,1)$:

$$(0,1,2) = a_1(1,1,1) + a_2(3,2,1) + a_3(1,0,1) + a_4(0,0,1).$$

This is the system

$$\begin{array}{lllll} a_1 & +3a_2 & +a_3 & & = 0 \\ a_1 & +2a_2 & & & = 1 \\ a_1 & +a_2 & +a_3 & +a_4 & = 2. \end{array}$$

Now we do Gaussian elimination again. Subtracting the first row from the second and third:

$$\begin{array}{lllll} a_1 & +3a_2 & +a_3 & & = 0 \\ & -a_2 & -a_3 & & = 1 \\ & -2a_2 & & +a_4 & = 2. \end{array}$$

Multiplying the second row by -1, then eliminating a_2 from the first and third rows:

$$\begin{array}{lllll} a_1 & & -2a_3 & & = 3 \\ & a_2 & +a_3 & & = -1 \\ & & 2a_3 & +a_4 & = 0. \end{array}$$

At this stage the system is in *reduced normal form*, which means that, starting from the bottom row and moving upwards, each equation introduces at least one new variable (ignoring any rows which have collapsed to something trivial like $0 = 0$). Once one is in reduced normal form, there isn't much more simplification one can do. In this case there is no unique solution; one can set a_4 to be arbitrary. The third equation then allows us to write a_3 in terms of a_4:

$$a_3 = -a_4/2$$

while the second equation then allows us to write a_2 in terms of a_3 (and thus of a_4:

$$a_2 = -1 - a_3 = -1 + a_4/2.$$

Similarly we can write a_1 in terms of a_4:

$$a_1 = 3 + 2a_3 = 3 - a_4.$$

Thus the general way to write $(0,1,2)$ as a linear combination of $(1,1,1)$, $(3,2,1)$, $(1,0,1)$, $(0,0,1)$ is

$$(0,1,2) = (3-a_4)(1,1,1)+(-1+a_4/2)(3,2,1)+(-a_4/2)(1,0,1)+a_4(0,0,1);$$

for instance, setting $a_4 = 4$, we have

$$(0,1,2) = -(1,1,1) + (3,2,1) - 2(1,0,1) + 4(0,0,1)$$

while if we set $a_4 = 0$, then we have

$$(0,1,2) = 3(1,1,1) - 1(3,2,1) + 0(1,0,1) + 0(0,0,1)$$

as before. Thus not only is $(0,1,2)$ in the span of $(1,1,1)$, $(3,2,1)$, $(1,0,1)$, and $(0,0,1)$, it can be written as a linear combination of such vectors in many ways. This is because some of the vectors in this set are redundant - as we already saw, we only needed the first three vectors $(1,1,1)$, $(3,2,1)$ and $(1,0,1)$ to generate $(0,1,2)$; the fourth vector $(0,0,1)$ was not necessary. As we shall see, this is because the four vectors $(1,1,1)$, $(3,2,1)$, $(1,0,1)$, and $(0,0,1)$ are *linearly dependent*. More on this later.

- Of course, sometimes a vector will not be in the span of other vectors at all. For instance, $(0,1,2)$ is not in the span of $(3,2,1)$ and $(1,0,1)$. If one were to try to solve the system

$$(0,1,2) = a_1(3,2,1) + a_2(1,0,1)$$

one would be solving the system

$$\begin{aligned} 3a_1 + \quad a_2 &= 0 \\ 2a_1 \qquad\quad &= 1 \\ a_1 + \quad a_2 &= 2. \end{aligned}$$

If one swapped the first and second rows, then divided the first by two, one obtains

$$\begin{array}{lll} a_1 & & = 1/2 \\ 3a_1 + & a_2 & = 0 \\ a_1 & +a_2 & = 2. \end{array}$$

Now zeroing the a_1 coefficient in the second and third rows gives

$$\begin{array}{lll} a_1 & & = 1/2 \\ & a_2 & = -3/2 \\ & a_2 & = 3/2. \end{array} .$$

Subtracting the second from the third, we get an absurd result:

$$\begin{array}{lll} a_1 & & = 1/2 \\ & a_2 & = -3/2 \\ & 0 & = 3. \end{array}$$

Thus there is no solution, and $(0, 1, 2)$ is not in the span.

* * * * *

Spanning sets

- **Definition.** A set S is said to *span* a vector space V if $\mathrm{span}(S) = V$; i.e. every vector in V is generated as a linear combination of elements of S. We call S a *spanning set* for V. (Sometimes one uses the verb "generated" instead of "spanned", thus V is *generated by* S and S is a *generating set* for V.)

- A model example of a spanning set is the set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ in $\mathbf{R}^3$; every vector in $\mathbf{R}^3$ can clearly be written as a linear combination of these three vectors, e.g.

$$(3, 7, 13) = 3(1, 0, 0) + 7(0, 1, 0) + 13(0, 0, 1).$$

There are of course similar examples for other vector spaces. For instance, the set $\{1, x, x^2, x^3\}$ spans $P_3(\mathbf{R})$ (why?).

- One can always add additional vectors to a spanning set and still get a spanning set. For instance, the set $\{(1,0,0),(0,1,0),(0,0,1),(9,14,23),(15,24,99)\}$ is also a spanning set for $\mathbf{R}^3$, for instance

  $$(3,7,13) = 3(1,0,0)+7(0,1,0)+13(0,0,1)+0(9,14,23)+0(15,24,99).$$

  Of course the last two vectors are not playing any significant role here, and are just along for the ride. A more extreme example: every vector space V is a spanning set for itself, $\mathrm{span}(V) = V$.

- On the other hand, removing elements from a spanning set can cause it to stop spanning. For instance, the two-element set $\{(1,0,0),(0,1,0)\}$ does not span, because there is no way to write $(3,7,13)$ (for instance) as a linear combination of $(1,0,0)$, $(0,1,0)$.

- Spanning sets are useful because they allow one to describe all the vectors in a space V in terms of a much smaller space S. For instance, the set $S := \{(1,0,0),(0,1,0),(0,0,1)\}$ only consists of three vectors, whereas the space $\mathbf{R}^3$ which S spans consists of infinitely many vectors. Thus, in principle, in order to understand the infinitely many vectors $\mathbf{R}^3$, one only needs to understand the three vectors in S (and to understand what linear combinations are).

- However, as we see from the above examples, spanning sets can contain "junk" vectors which are not actually needed to span the set. Such junk occurs when the set is *linearly dependent*. We would like to now remove such junk from the spanning sets and create a "minimal" spanning set - a set whose elements are all *linearly independent*. Such a set is known as a *basis*. In the rest of this series of lecture notes we discuss these related concepts of linear dependence, linear independence, and being a basis.

* * * * *

Linear dependence and independence

- Consider the following three vectors in $\mathbf{R}^3$: $v_1 := (1,2,3)$, $v_2 := (1,1,1)$, $v_3 := (3,5,7)$. As we now know, the span $\mathrm{span}(\{v_1,v_2,v_3\})$ of this set is just the set of all linear combinations of v_1, v_2, v_3:

  $$\mathrm{span}(\{v_1,v_2,v_3\}) := \{a_1v_1 + a_2v_2 + a_3v_3 : a_1,a_2,a_3 \in \mathbf{R}\}.$$

Thus, for instance $3(1,2,3)+4(1,1,1)+1(3,5,7)=(10,15,20)$ lies in the span. However, the $(3,5,7)$ vector is redundant because it can be written in terms of the other two:

$$v_3=(3,5,7)=2(1,2,3)+(1,1,1)=2v_1+v_2$$

or more symmetrically

$$2v_1+v_2-v_3=0.$$

Thus any linear combination of v_1, v_2, v_3 is in fact just a linear combination of v_1 and v_2:

$$a_1v_1+a_2v_2+a_3v_3=a_1v_1+a_2v_2+a_3(2v_1+v_2)=(a_1+2a_3)v_1+(a_2+a_3)v_2.$$

- Because of this redundancy, we say that the vectors v_1, v_2, v_3 are *linearly dependent*. More generally, we say that any collection S of vectors in a vector space V are *linearly dependent* if we can find distinct elements $v_1,\ldots,v_n \in S$, and scalars $a_1,\ldots,a_n$, not all equal to zero, such that

$$a_1v_1+a_2v_2+\ldots+a_nv_n=0.$$

- (Of course, 0 can always be written as a linear combination of $v_1,\ldots,v_n$ in a trivial way: $0=0v_1+\ldots+0v_n$. Linear dependence means that this is not the only way to write 0 as a linear combination, that there exists at least one non-trivial way to do so). We need the condition that the $v_1,\ldots,v_n$ are distinct to avoid silly things such as $2v_1+(-2)v_1=0$.

- In the case where S is a finite set $S=\{v_1,\ldots,v_n\}$, then S is linearly dependent if and only if we can find scalars $a_1,\ldots,a_n$ not all zero such that

$$a_1v_1+\ldots+a_nv_n=0.$$

  (Why is the same as the previous definition? It's a little subtle).

- If a collection of vectors S is not linearly dependent, then they are said to be *linearly independent*. An example is the set $\{(1,2,3),(0,1,2)\}$; it is not possible to find a_1, a_2, not both zero for which

$$a_1(1,2,3)+a_2(0,1,2)=0,$$

because this would imply

$$\begin{array}{lll} a_1 & & = 0 \\ 2a_1 & +a_2 & = 0 \\ 3a_1 & +2a_2 & = 0 \end{array},$$

which can easily be seen to only be true if a_1 and a_2 are both 0. Thus there is no non-trivial way to write the zero vector $0 = (0, 0, 0)$ as a linear combination of $(1, 2, 3)$ and $(0, 1, 2)$.

- By convention, an empty set of vectors (with $n = 0$) is always linearly independent (why is this consistent with the definition?)
- As indicated above, if a set is linearly dependent, then we can remove one of the elements from it without affecting the span.
- **Theorem.** Let S be a subset of a vector space V. If S is linearly dependent, then there exists an element v of S such that the smaller set $S - \{v\}$ has the same span as S:

  $$\mathrm{span}(S - \{v\}) = \mathrm{span}(S).$$

  Conversely, if S is linearly independent, then every proper subset $S' \subsetneq S$ of S will span a strictly smaller set than S:

  $$\mathrm{span}(S') \subsetneq \mathrm{span}(S).$$

- **Proof.** Let's prove the first claim: if S is a linearly dependent subset of V, then we can find $v \in S$ such that $\mathrm{span}(S - \{v\}) = \mathrm{span}(S)$.
- Since S is linearly dependent, then by definition there exists distinct $v_1, \ldots, v_n$ and scalars $a_1, \ldots, a_n$, not all zero, such that

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

  We know that at least one of the a_j are non-zero; without loss of generality we may assume that a_1 is non-zero (since otherwise we can just shuffle the v_j to bring the non-zero coefficient out to the front). We can then solve for v_1 by dividing by a_1:

  $$v_1 = -\frac{a_2}{a_1} v_2 - \ldots - \frac{a_n}{a_1} v_n.$$

Thus any expression involving v_1 can instead be written to involve $v_2, \ldots, v_n$ instead. Thus any linear combination of v_1 and other vectors in S not equal to v_1 can be rewritten instead as a linear combination of $v_2, \ldots, v_n$ and other vectors in S not equal to v_1. Thus every linear combination of vectors in S can in fact be written as a linear combination of vectors in $S - \{v_1\}$. On the other hand, every linear combination of $S - \{v_1\}$ is trivially also a linear combination of S. Thus we have $\mathrm{span}(S) = \mathrm{span}(S - \{v_1\})$ as desired.

- Now we prove the other direction. Suppose that $S \subseteq V$ is linearly independent. And le $S' \subsetneq S$ be a proper subset of S. Since every linear combination of S' is trivially a linear combination of S, we have that $\mathrm{span}(S') \subseteq \mathrm{span}(S)$. So now we just need argue why $\mathrm{span}(S') \neq \mathrm{span}(S)$.

  Let v be an element of S which is not contained in S'; such an element must exist because S' is a proper subset of S. Since $v \in S$, we have $v \in \mathrm{span}(S)$. Now suppose that v were also in $\mathrm{span}(S')$. This would mean that there existed vectors $v_1, \ldots, v_n \in S'$ (which in particular were distinct from v) such that

$$v = a_1 v_1 + a_2 v_2 + \ldots + a_n v_n,$$

  or in other words

$$(-1)v + a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0.$$

  But this is a non-trivial linear combination of vectors in S which sum to zero (it's nontrivial because of the -1 coefficient of v). This contradicts the assumption that S is linearly independent. Thus v cannot possibly be in $\mathrm{span}(S')$. But this means that $\mathrm{span}(S')$ and $\mathrm{span}(S)$ are different, and we are done. $\square$

* * * * *

Bases

- A *basis* of a vector space V is a set S which both spans S, but is also linearly independent. In other words, a basis consists of a bare minimum number of vectors needed to span all of V; remove one of them, and you fail to span V.

- Thus the set $\{(1,0,0),(0,1,0),(0,0,1)\}$ is a basis for $\mathbf{R}^3$, because it both spans and is linearly independent. The set $\{(1,0,0),(0,1,0),(0,0,1),(9,14,23)\}$ still spans $\mathbf{R}^3$, but is not linearly independent and so is not a basis. The set $\{(1,0,0),(0,1,0)\}$ is linearly independent, but does not span all of $\mathbf{R}^3$ so is not a basis. Finally, the set $\{(1,0,0),(2,0,0)\}$ is neither linearly independent nor spanning, so is definitely not a basis.

- Similarly, the set $\{1, x, x^2, x^3\}$ is a basis for $P_3(\mathbf{R})$, while the set $\{1, x, 1+x, x^2, x^2+x^3, x^3\}$ is not (it still spans, but is linearly dependent). The set $\{1, x+x^2, x^3\}$ is linearly independent, but doesn't span.

- One can use a basis to represent a vector in a unique way as a collection of numbers:

- **Lemma.** Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for a vector space V. Then every vector in v can be written *uniquely* in the form

$$v = a_1v_1 + \ldots + a_nv_n$$

for some scalars $a_1, \ldots, a_n$.

- **Proof.** Because $\{v_1, \ldots, v_n\}$ is a basis, it must span V, and so every vector v in V can be written in the form $a_1v_1 + \ldots + a_nv_n$. It only remains to show why this representation is unique. Suppose for contradiction that a vector v had two different representations

$$v = a_1v_1 + \ldots + a_nv_n$$

$$v = b_1v_1 + \ldots + b_nv_n$$

where $a_1, \ldots, a_n$ are one set of scalars, and $b_1, \ldots, b_n$ are a different set of scalars. Subtracting the two equations we get

$$(a_1 - b_1)v_1 + \ldots + (a_n - b_n)v_n = 0.$$

But the $v_1, \ldots, v_n$ are linearly independent, since they are a basis. Thus the only representation of 0 as a linear combination of $v_1, \ldots, v_n$ is the trivial representation, which means that the scalars $a_1 - b_1, \ldots, a_n - b_n$ must be equal. That means that the two representations $a_1v_1 + \ldots + a_nv_n$, $b_1v_1 + \ldots + b_nv_n$ must in fact be the same representation. Thus v cannot have two distinct representations, and so we have a unique representation as desired. $\square$

- As an example, let v_1 and v_2 denote the vectors $v_1 := (1,1)$ and $v_2 := (1,-1)$ in $\mathbf{R}^2$. One can check that these two vectors span $\mathbf{R}^2$ and are linearly independent, and so they form a basis. Any typical element, e.g. $(3,5)$, can be written uniquely in terms of v_1 and v_2:

$$(3,5) = 4(1,1) - (1,-1) = 4v_1 - v_2.$$

  In principle, we could write all vectors in $\mathbf{R}^2$ this way, but it would be a rather non-standard way to do so, because this basis is rather non-standard. Fortunately, most vector spaces have "standard" bases which we use to represent them:

- The standard basis of $\mathbf{R}^n$ is $\{e_1, e_2, \ldots, e_n\}$, where e_j is the vector whose j^{th} entry is 1 and all the others are 0. Thus for instance, the standard basis of $\mathbf{R}^3$ consists of $e_1 := (1,0,0)$, $e_2 := (0,1,0)$, and $e_2 := (0,0,1)$.

- The standard basis of the space $P_n(\mathbf{R})$ is $\{1, x, x^2, \ldots, x^n\}$. The standard basis of $P(\mathbf{R})$ is the infinite set $\{1, x, x^2, \ldots\}$.

- One can concoct similar standard bases for matrix spaces $M_{m\times n}(\mathbf{R})$ (just take those matrices with a single coefficient 1 and all the others zero). However, there are other spaces (such as $C(\mathbf{R})$) which do not have a reasonable standard basis.