Lecture notes for Math 115A (linear algebra)
Fall of 2002
Terence Tao, UCLA
http://www.math.ucla.edu/~tao/resource/general/115a.3.02f/
The textbook used was Linear Algebra, S.H. Friedberg, A.J. Insel, L.E. Spence, Third Edition. Prentice Hall, 1999.
Thanks to Radhakrishna Bettadapura, Yu Cao, Cristian Gonzales, Hannah Kim, Michael Smith, Wilson Sov, Luqing Ye, and Shijia Yu for corrections.

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

$$\begin{aligned} F_x &= ma_x + 0a_y + 0a_z \\ F_y &= 0a_x + ma_y + 0a_z \\ F_z &= 0a_x + 0a_y + ma_z. \end{aligned}$$

  This linear transformation is associated to the *matrix*

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

- Here is another example of a linear transformation with vector inputs and vector outputs:

$$\begin{aligned} y_1 &= 3x_1 + 5x_2 + 7x_3 \\ y_2 &= 2x_1 + 4x_2 + 6x_3; \end{aligned}$$

  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 $\mathbf{R}$, the field of real numbers, but occasionally one likes to work with other fields such as $\mathbf{C}$, the field of complex numbers, or $\mathbf{Q}$, the field of rational numbers. However in this course the field of scalars will almost always be $\mathbf{R}$. (In the textbook the scalar field is often denoted $\mathbf{F}$, just to keep aside the possibility that it might not be the reals $\mathbf{R}$; 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 $x_1 = y_1$, $x_2 = y_2$, ..., 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 $\mathbf{R}^2$, $\mathbf{R}^3$, $\ldots$, 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:

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

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

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 W is also a vector in V), 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:

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

- 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

$$\mathrm{span}(\{v_1, \ldots, v_n\}) := \{a_1 v_1 + a_2 v_2 + \ldots + a_n v_n : a_1, \ldots, a_n \in \mathbf{R}\}.$$

(Why?)

- Occasionally we will need to deal when S is empty. In this case we set the span $\mathrm{span}(\emptyset)$ 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 $a_1 v_1 + \ldots + a_n v_n$, 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 $a_1 v_1 + \ldots + a_n v_n$ and $b_1 v_1 + \ldots + b_n v_n$, 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

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

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 vector addition, we take a typical element of this space, say $a_1 v_1 + \ldots a_n v_n$, and a typical scalar c. We want to verify that the scalar product

$$c(a_1 v_1 + \ldots + a_n v_n)$$

is also in $\mathrm{span}(\{v_1, \ldots, v_n\})$. 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 $v = a_1 v_1 + \ldots + a_n v_n$ 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 $a_1 v_1$ lies in W. Similarly $a_2 v_2, \ldots, a_n v_n$ lie in W. But W is closed under vector addition, thus $a_1 v_1 + \ldots + a_n v_n$ lies in W, as desired. This concludes the proof of the Theorem. □

- We remark that the span of a set of vectors does not depend on what order we list the set S: for instance, $\mathrm{span}(\{\mathrm{u}, \mathrm{v}, \mathrm{w}\})$ is the same as $\mathrm{span}(\{\mathrm{w}, \mathrm{v}, \mathrm{u}\})$. (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 of 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_2a_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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 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, 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:

$$\mathrm{span}(\{\mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3\}) := \{\mathrm{a}_1\mathrm{v}_1 + \mathrm{a}_2\mathrm{v}_2 + \mathrm{a}_3\mathrm{v}_3 : \mathrm{a}_1, \mathrm{a}_2, \mathrm{a}_3 \in \mathbf{R}\}.$$

  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:

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

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

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

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

- 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

$$v = a_1v_1 + a_2v_2 + \ldots + a_nv_n,$$

  or in other words

$$(-1)v + a_1v_1 + a_2v_2 + \ldots + a_nv_n = 0.$$

  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. □

* * * * *

Bases

- A *basis* of a vector space V is a set S which spans V, while also being 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 $a_1 v_1 + \ldots + a_n v_n$, $b_1 v_1 + \ldots + b_n v_n$ 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.

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 span(S) and is a subspace of V. If 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 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

$$\begin{aligned} a_1 \quad +a_2 &= x \\ -a_2 &= y \\ -a_1 \quad\quad &= z \end{aligned}$$

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 span(S), then $S \cup \{v\}$ is linearly dependent, and span(S$\cup$ {v}) = span(S).
- (b) If v does not lie in span(S), then $S \cup \{v\}$ is linearly independent, and span(S $\cup$ {v}) $\supsetneq$ 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 span(S); this will also keep S linearly independent.
- **Proof** We first prove (a). If v lies in 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

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

and thus

$$0 = (-1)v + a_1v_1 + \ldots + a_nv_n.$$

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 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 $v \not\in$ 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. $\square$

* * * * *

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. 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

$$v_m = a_1 v_1 + \ldots + a_{m-1} v_{m-1} + b_1 w_1 + \ldots + b_{n-m+1} w_{n-m+1} \tag{0.2}$$

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

$$v_m = a_1 v_1 + \ldots + a_{m-1} v_{m-1} \tag{0.3}$$

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 (0.3) 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 (0.2) 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,

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

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

- **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 $\mathrm{span}(\{\mathrm{v}_1\})$; 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 $\mathrm{span}(\{\mathrm{v}_1, \mathrm{v}_2\})$. 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 span(B) = V and span(B) = W, which implies that $W = V$ as desired. $\square$

* * * * *

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 $f \in P_Pn(\mathbf{R})$ 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. $\square$

- 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

$$f = a_1 f_1 + \ldots + a_n f_n \tag{0.4}$$

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 (0.4), 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

$$f := y_1 f_1 + \ldots + y_n f_n = \sum_{j=1}^{n} y_j f_j \tag{0.5}$$

which is the Lagrange interpolation formula. Conversely, it is easy to check that if we define f by the formula (0.5), then $f(x_1) = y_1$, $f(x_2) = y_2$, etc. so f is indeed the unique interpolating polynomial between the data points $(x_1, y_1), \ldots, (x_3, y_3)$. $\square$

- 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.)

Math 115A - Week 3
Textbook sections: 2.1-2.3
Topics covered:

- Null spaces and nullity of linear transformations
- Range and rank of linear transformations
- The Dimension Theorem
- Linear transformations and bases
- Co-ordinate bases
- Matrix representation of linear transformations
- Sum, scalar multiplication, and composition of linear transformations

* * * * *

Review of linear transformations

- A linear transformation is any map $T : V \to W$ from one vector space V to another W such that T preserves vector addition (i.e. $T(v+v') = Tv + Tv'$ for all $v, v' \in V$) and T preserves scalar multiplication (i.e. $T(cv) = cTv$ for all scalars c and all $v \in V$).
- A map which preserves vector addition is sometimes called *additive*; a map which preserves scalar multiplication is sometimes called *homogeneous.*
- We gave several examples of linear transformations in the previous notes; here are a couple more.
- **Sampling as a linear transformation** Recall that $\mathcal{F}(\mathbf{R}, \mathbf{R})$ is the space of all functions from $\mathbf{R}$ to $\mathbf{R}$. This vector space might be used to represent, for instance, sound signals $f(t)$. In practice, a measuring device cannot capture all the information in a signal (which contains an infinite amount of data); instead it only samples a finite amount, at some fixed times. For instance, a measuring device might only sample

$f(t)$ for $t = 1, 2, 3, 4, 5$ (this would correspond to sampling at 1Hz for five seconds). This operation can be described by a linear transformation $S : \mathcal{F}(\mathbf{R}, \mathbf{R}) \to \mathbf{R}^5$, defined by

$$Sf := (f(1), f(2), f(3), f(4), f(5));$$

i.e. S transforms a signal $f(t)$ into a five-dimensional vector, consisting of f sampled at five times. For instance,

$$S(x^2) = (1, 4, 9, 16, 25)$$

$$S(\sqrt{x}) = (\sqrt{1}, \sqrt{2}, \sqrt{3}, \sqrt{4}, \sqrt{5})$$

etc. (Why is this map linear?)

- One can similarly sample polynomial spaces. For instance, the map $S : P_2(\mathbf{R}) \to \mathbf{R}^3$ defined by

  $$Sf := (f(0), f(1), f(2))$$

  is linear.

- **Interpolation as a linear transformation** Interpolation can be viewed as the reverse of sampling. For instance, given three numbers y_1, y_2, y_3, the Lagrange interpolation formula gives us a polynomial $f \in P_2(\mathbf{R})$ such that $f(0) = y_1$, $f(1) = y_2$, and $f(2) = y_3$:

  $$f(x) = y_1 \frac{(x-1)(x-2)}{(0-1)(0-2)} + y_2 \frac{(x-0)(x-2)}{(1-0)(1-2)} + y_3 \frac{(x-0)(x-1)}{(2-0)(2-1)}.$$

  One can view this as a linear transformation $S : \mathbf{R}^3 \to P_2(\mathbf{R})$ defined by $S(y_1, y_2, y_3) := f$, e.g.

  $$S(3, 4, 7) = y_1 \frac{(x-1)(x-2)}{(0-1)(0-2)} + y_2 \frac{(x-0)(x-2)}{(1-0)(1-2)} + y_3 \frac{(x-0)(x-1)}{(2-0)(2-1)}.$$

  (Why is this linear?). This is the *inverse* of the transformation S defined in the previous paragraph - but more on that later.

- **Linear combinations as a linear transformation** Let V be a vector space, and let $v_1, \ldots, v_n$ be a set of vectors in V. Then the transformation $T : \mathbf{R}^n \to V$ defined by

$$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

is a linear transformation (why?). Also, one can express many of the statements from previous notes in terms of this transformation T. For instance, $\mathrm{span}(\{\mathrm{v}_1, \ldots, \mathrm{v}_\mathrm{n}\})$ is the same thing as the image $T(\mathbf{R}^n)$ of T; thus $\{v_1, \ldots, v_n\}$ spans V if and only if T is onto. On the other hand, T is one-to-one if and only if $\{v_1, \ldots, v_n\}$ is linearly independent (more on this later). Thus T is a bijection if and only if $\{v_1, \ldots, v_n\}$ is a basis.

* * * * *

Null spaces and nullity

- A note on notation: in this week's notes, we shall often be dealing with two different vector spaces V and W, so we have two different types of vectors. We will try to reserve the letter v to denote vectors in V, and w to denote vectors in W, in what follows.

- Not all linear transformations are alike; for instance, the zero transformation $T : V \to W$ defined by $Tv := 0$ behaves rather differently from, say, the identity transformation $T : V \to V$ defined by $Tv := v$. Now we introduce some characteristics of linear transformations to start telling them apart.

- **Definition** Let $T : V \to W$ be a linear transformation. The *null space* of T, called $N(T)$, is defined to be the set

$$N(T) := \{v \in V : Tv = 0\}.$$

- In other words, the null space consists of all the stuff that T sends to zero (this is the zero vector 0_W of W, not the zero vector 0_V of V): $N(T) = T^{-1}(\{0\})$. Some examples: if $T : V \to W$ is the zero transformation $Tv := 0$, then the null space $N(T) = V$. If instead $T : V \to V$ is the identity transformation $Tv := v$, then $N(T) = \{0\}$. If $T : \mathbf{R}^3 \to \mathbf{R}$ is the linear transformation $T(x, y, z) = x + y + z$, then $N(T)$ is the plane $\{(x, y, z) \in \mathbf{R}^3 : x + y + z = 0\}$.

- The null space of T is sometimes also called the *kernel* of T, and is sometimes denoted $\ker(T)$; but we will use the notation $N(T)$ throughout this course.

- The null space $N(T)$ is always a subspace of V; this is an exercise. Intuitively, the larger the null space, the more T resembles the 0 transformation. The null space also measures the extent to which T fails to be one-to-one:

- **Lemma 1.** Let $T : V \to W$ be a linear transformation. Then T is one-to-one if and only if $N(T) = \{0\}$.

- **Proof.** First suppose that T is one-to-one; we have to show that $N(T) = \{0\}$. First of all, it is clear that $0 \in N(T)$, because $T0 = 0$. Now we show that no other element is in $N(T)$. Suppose for contradiction that there was a non-zero vector $v \in V$ such that $v \in N(T)$, i.e. that $Tv = 0$. Then $Tv = T0$. But T is one-to-one, so this forces $v = 0$, contradiction.

- Now suppose that $N(T) = \{0\}$; we have to show that T is one-to-one. In other words, we need to show that whenever $Tv = Tv'$, then we must have $v = v'$. So suppose that $Tv = Tv'$. Then $Tv - Tv' = 0$, so that $T(v - v') = 0$. Thus $v - v' \in N(T)$, which means by hypothesis that $v - v' = 0$, so $v = v'$, as desired. $\square$

- Example: Take the transformation $T : \mathbf{R}^n \to V$ defined by

  $$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

  which we discussed earlier. If $\{v_1, \ldots, v_n\}$ is linearly dependent, then there is a non-zero n-tuple $(a_1, \ldots, a_n)$ such that $0 = a_1 v_1 + \ldots + a_n v_n$; i.e. $N(T)$ will consist of more than just the 0 vector. Conversely, if $N(T) \neq \{0\}$, then $\{v_1, \ldots, v_n\}$ is linearly dependent. Thus by Lemma 1, T is injective if and only if $\{v_1, \ldots, v_n\}$ is linearly independent.

- Since $N(T)$ is a vector space, it has a dimension. We define the *nullity* of T to be the dimension of $N(T)$; this may be infinite, if $N(T)$ is infinite dimensional. The nullity of T will be denoted $\mathrm{nullity}(T)$, thus $\mathrm{nullity}(T) = \dim(N(T))$.

- Example: let $\pi : \mathbf{R}^5 \to \mathbf{R}^5$ be the operator

$$\pi(x_1, x_2, x_3, x_4, x_5) := (x_1, x_2, x_3, 0, 0)$$

(Why is this linear?). Then

$$N(\pi) = \{(0, 0, 0, x_4, x_5) : x_4, x_5 \in \mathbf{R}\}$$

(why?); this is a two-dimensional space (it has a basis consisting of $(0, 0, 0, 1, 0)$ and $(0, 0, 0, 0, 1)$ and so $\text{nullity}(\pi) = 2$.

- Example: By Lemma 1, a transformation is injective if and only if it has a nullity of 0.

- The nullity of T measures how much information (or degrees of freedom) is lost when applying T. For instance, in the above projection, two degrees of freedom are lost: the freedom to vary the x_4 and x_5 co-ordinates are lost after applying π. An injective transformation does not lose any information (if you know Tv, then you can reconstruct v).

* * * * *

Range and rank

- You may have noticed that many concepts in this field seem to come in complementary pairs: spanning set versus linearly independent set, one-to-one versus onto, etc. Another such pair is null space and range, or nullity and rank.

- **Definition** The *range* $R(T)$ of a linear transformation $T : V \to W$ is defined to be the set

$$R(T) := \{Tv : v \in V\}.$$

- In other words, $R(T)$ is all the stuff that T maps into: $R(T) = T(V)$. (Unfortunately, the space W is also sometimes called the range of T; to avoid confusion we will try to refer to W instead as the *target space* for T; V is the *initial space* or *domain* of T.)

- Just as the null space $N(T)$ is always a subspace of V, it can be shown that $R(T)$ is a subspace of W (this is part of an exercise).

- Examples: If $T : V \to W$ is the zero transformation $Tv := 0$, then $R(T) = \{0\}$. If $T : V \to V$ is the identity transformation $Tv := v$, then $R(T) = V$. If $T : \mathbf{R}^n \to V$ is the transformation

$$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

discussed earlier, then $R(T) = \mathrm{span}(\{\mathrm{v}_1, \ldots, \mathrm{v}_\mathrm{n}\})$.

- Example: A map $T : V \to W$ is onto if and only if $R(T) = W$.

- **Definition** The *rank* $\mathrm{rank}(T)$ of a linear transformation $T : V \to W$ is defined to be the dimension of $R(T)$, thus $\mathrm{rank}(T) = \dim(R(T))$.

- Examples: The zero transformation has rank 0 (and indeed these are the only transformations with rank 0). The transformation

$$\pi(x_1, x_2, x_3, x_4, x_5) := (x_1, x_2, x_3, 0, 0)$$

defined earlier has range

$$R(\pi) = \{(x_1, x_2, x_3, 0, 0) : x_1, x_2, x_3 \in \mathbf{R}\}$$

(why?), and so has rank 3.

- The rank measures how much information (or degrees of freedom) is *retained* by the transformation T. For instance, with the example of π above, even though two degrees of freedom have been lost, three degrees of freedom remain.

* * * * *

The dimension theorem

- Let $T : V \to W$ be a linear transformation. Intuitively, $\mathrm{nullity}(T)$ measures how many degrees of freedom are lost when applying T; $\mathrm{rank}(T)$ measures how many degrees of freedom are retained. Since the initial space V originally has $\dim(V)$ degrees of freedom, the following theorem should not be too surprising.

- **Dimension Theorem** Let V be a finite-dimensional space, and let $T : V \to W$ be a linear transformation. Then

$$\mathrm{nullity}(T) + \mathrm{rank}(T) = \dim(V).$$

- The proof here will involve a lot of shuttling back and forth between V and W using T; and is an instructive example as to how to analyze linear transformations.

- **Proof.** By hypothesis, $\dim(V)$ is finite; let's define $n := \dim(V)$. Since $N(T)$ is a subspace of V, it must also be finite-dimensional; let's call $k := \dim(N(T)) = \mathrm{nullity}(T)$. Then we have $0 \leq k \leq n$. Our task is to show that $k + \mathrm{rank}(T) = n$, or in other words that $\dim(R(T)) = n - k$.

- By definition of dimension, the space $N(T)$ must have a basis $\{v_1, \ldots, v_k\}$ of k elements. (Probably it has many such bases, but we just need one such for this argument). This set of k elements lies in $N(T)$, and thus in V, and is linearly independent; thus by part (f) of Corollary 1 of last week's notes, it must be part of a basis of V, which must then have $n = \dim(V)$ elements (by part (c) of Corollary 1). Thus we may add $n - k$ extra elements $v_{k+1}, \ldots, v_n$ to our $N(T)$-basis to form an V-basis $\{v_1, \ldots, v_n\}$.

- Since $v_{k+1}, \ldots, v_n$ lie in V, the elements $Tv_{k+1}, \ldots, Tv_n$ lie in $R(T)$. We now claim that $\{Tv_{k+1}, \ldots, Tv_n\}$ are a basis for $R(T)$; this will imply that $R(T)$ has dimension $n - k$, as desired.

- To verify that $\{Tv_{k+1}, \ldots, Tv_n\}$ form a basis, we must show that they span $R(T)$ and that they are linearly independent. First let's show they span $R(T)$. This means that every vector in $R(T)$ is a linear combination of $Tv_{k+1}, \ldots, Tv_n$. So let's pick a typical vector w in $R(T)$; our job is to show that w is a linear combination of $Tv_{k+1}, \ldots, Tv_n$. By definition of $R(T)$, w must equal Tv for some v in V.

- On the other hand, we know that $\{v_1, \ldots, v_n\}$ spans V, thus we must have
$$v = a_1v_1 + \ldots + a_nv_n$$
for some scalars $a_1, \ldots, a_n$. Applying T to both sides and using the fact that T is linear, we obtain
$$Tv = a_1Tv_1 + \ldots + a_nTv_n.$$

- Now we use the fact that $v_1, \ldots, v_k$ lie in $N(T)$, so $Tv_1 = \ldots = Tv_k = 0$. Thus
$$Tv = a_{k+1}Tv_{k+1} + \ldots + a_nTv_n.$$
Thus $w = Tv$ is a linear combination of $Tv_{k+1}, \ldots, Tv_n$, as dsired.

- Now we show that $\{Tv_{k+1}, \ldots, Tv_n\}$ is linearly independent. Suppose for contradiction that this set was linearly dependent, thus
$$a_{k+1}Tv_{k+1} + \ldots + a_nTv_n = 0$$
for some scalars $a_{k+1}, \ldots, a_n$ which were not all zero. Then by the linearity of T again, we have
$$T(a_{k+1}v_{k+1} + \ldots + a_nv_n) = 0$$
and thus by definition of null space
$$a_{k+1}v_{k+1} + \ldots + a_nv_n \in N(T).$$
Since $N(T)$ is spanned by $\{v_1, \ldots, v_k\}$, we thus have
$$a_{k+1}v_{k+1} + \ldots + a_nv_n = a_1v_1 + \ldots a_kv_k$$
for some scalars $a_1, \ldots, a_k$. We can rearrange this as
$$-a_1v_1 - \ldots - a_kv_k + a_{k+1}v_{k+1} + \ldots + a_nv_n = 0.$$
But the set $\{v_1, \ldots, v_n\}$ is linearly independent, which means that all the a's must then be zero. But that contradicts our hypothesis that not all of the $a_{k+1}, \ldots, a_n$ were zero. Thus $\{Tv_{k+1}, \ldots, Tv_n\}$ must have been linearly independent, and we are done. $\square$.

- **Example** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ denote the linear transformation
$$T(x, y) := (x + y, 2x + 2y).$$
The null space of this transformation is
$$N(T) = \{(x, y) \in \mathbf{R}^2 : x + y = 0\}$$

(why?); this is a line in $\mathbf{R}^2$, and thus has dimension 1 (for instance, it has $\{(1,-1)\}$ as a basis). The range of this transformation is

$$R(T) = \{(t, 2t) : t \in \mathbf{R}\}$$

(why?); this is another line in $\mathbf{R}^2$ and has dimension 1. Since 1+1=2, the Dimension Theorem is verified in this case.

- **Example** For the zero transformation $Tx := 0$, we have $\text{nullity}(T) = \dim(X)$ and $\text{rank}(T) = 0$ (so all the degrees of freedom are lost); while for the identity transformation $Tx := x$ we have $\text{nullity}(T) = 0$ and $\text{rank}(T) = \dim(X)$ (so all the degrees of freedom are retained). In both cases we see that the Dimension Theorem is verified.

- One important use of the Dimension Theorem is that it allows us to discover facts about the range of T just from knowing the null space of T, and vice versa. For instance:

- **Example** Let $T : P_5(\mathbf{R}) \to P_4(\mathbf{R})$ denote the differentiation map

$$Tf := f';$$

thus for instance $T(x^3+2x) = 3x^2+2$. The null space of T consists of all polynomials f in $P_5(\mathbf{R})$ for which $f' = 0$; i.e. the constant polynomials

$$N(T) = \{c : c \in \mathbf{R}\} = P_0(\mathbf{R}).$$

Thus $N(T)$ has dimension 1 (it has $\{1\}$ as a basis). Since $P_5(\mathbf{R})$ has dimension 6, we thus see from the dimension theorem that $R(T)$ must have dimension 5. But $R(T)$ is a subspace of $P_4(\mathbf{R})$, and $P_4(\mathbf{R})$ has dimension 5. Thus $R(T)$ must equal all of $P_4(\mathbf{R})$. In other words, every polynomial of degree at most 4 is the derivative of some polynomial of degree at most 5. (This is of course easy to check by integration, but the amazing fact was that we could deduce this fact purely from linear algebra - using only a very small amount of calculus).

- Here is another example:

- **Lemma 2** Let V and W be finite-dimensional vector spaces of the same dimension ($\dim(V) = \dim(W)$), and let $T : V \to W$ be a linear transformation from V to W. Then T is one-to-one if and only if T is onto.

- **Proof** If T is one-to-one, then nullity$(T) = 0$, which by the dimension theorem implies that rank$(T) = \dim(V)$. Since $\dim(V) = \dim(W)$, we thus have $\dim R(T) = \dim(W)$. But $R(T)$ is a subspace of W, thus $R(T) = W$, i.e. T is onto. The reverse implication then follows by reversing the above steps (we leave as an exercise to verify that all the steps are indeed reversible). $\square$

- Exercise: re-interpret Corollary 1(de) from last week's notes using this Lemma, and the linear transformation

$$T(a_1, \ldots, a_n) := a_1 v_1 + \ldots + a_n v_n$$

discussed earlier.

Linear transformations and bases

- Let $T : V \to W$ be a linear transformation, and let $\{v_1, \ldots, v_n\}$ be a collection of vectors in V. Then $\{Tv_1, \ldots, Tv_n\}$ is a collection of vectors in W. We now study how similar these two collections are; for instance, if one is a basis, does this mean the other one is also a basis?

- **Theorem 3** If $T : V \to W$ is a linear transformation, and $\{v_1, \ldots, v_n\}$ spans V, then $\{Tv_1, \ldots, Tv_n\}$ spans $R(T)$.

- **Proof**. Let w be any vector in $R(T)$; our job is to show that w is a linear combination of $Tv_1, \ldots, Tv_n$. But by definition of $R(T)$, $w = Tv$ for some $v \in V$. Since $\{v_1, \ldots, v_n\}$ spans V, we thus have $v = a_1 v_1 + \ldots + a_n v_n$ for some scalars $a_1, \ldots, a_n$. Applying T to both sides, we obtain $Tv = a_1 Tv_1 + \ldots + a_n Tv_n$. Thus we can write $w = Tv$ as a linear combination of $Tv_1, \ldots, Tv_n$, as desired. $\square$

- **Theorem 4** If $T : V \to W$ is a linear transformation which is one-to-one, and $\{v_1, \ldots, v_n\}$ is linearly independent, then $\{Tv_1, \ldots, Tv_n\}$ is also linearly independent.

- **Proof** Suppose we can write 0 as a linear combination of $\{Tv_1, \ldots, Tv_n\}$:

$$0 = a_1 Tv_1 + \ldots + a_n Tv_n.$$

Our job is to show that the $a_1, \ldots, a_n$ must all be zero. Using the linearity of T, we obtain

$$0 = T(a_1 v_1 + \ldots + a_n v_n).$$

Since T is one-to-one, $N(T) = \{0\}$, and thus

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

But since $\{v_1, \ldots, v_n\}$ is linearly independent, this means that $a_1, \ldots, a_n$ are all zero, as desired. $\square$

- **Corollary 5** If $T : V \to W$ is both one-to-one and onto, and $\{v_1, \ldots, v_n\}$ is a basis for V, then $\{Tv_1, \ldots, Tv_n\}$ is a basis for W. (In particular, we see that $\dim(V) = \dim(W)$).

- **Proof** Since $\{v_1, \ldots, v_n\}$ is a basis for V, it spans V; and hence, by Theorem 3, $\{Tv_1, \ldots, Tv_n\}$ spans $R(T)$. But $R(T) = W$ since T is onto. Next, since $\{v_1, \ldots, v_n\}$ is linearly independent and T is one-to-one, we see from Theorem 4 that $\{Tv_1, \ldots, Tv_n\}$ is also linearly independent. Combining these facts we see that $\{Tv_1, \ldots, Tv_n\}$ is a basis for W. $\square$

- The converse is also true: if $T : V \to W$ is one-to-one, and $\{Tv_1, \ldots, Tv_n\}$ is a basis, then $\{v_1, \ldots, v_n\}$ is also a basis; we leave this as an exercise (it's very similar to the previous arguments).

- **Example** The map $T : P_3(\mathbf{R}) \to \mathbf{R}^4$ defined by

$$T(ax^3 + bx^2 + cx + d) := (a, b, c, d)$$

is both one-to-one and onto (why?), and is also linear (why?). Thus we can convert every basis of $P_3(\mathbf{R})$ to a basis of $\mathbf{R}^4$ and vice versa. For instance, the standard basis $\{1, x, x^2, x^3\}$ of $P_3(\mathbf{R})$ can be converted to the basis $\{(0,0,0,1), (0,0,1,0), (0,1,0,0), (1,0,0,0)\}$ of $\mathbf{R}^4$. In principle, this allows one to convert many problems about the vector space $P_3(\mathbf{R})$ into one about $\mathbf{R}^4$, or vice versa. (The formal way of saying this is that $P_3(\mathbf{R})$ and $\mathbf{R}^4$ are *isomorphic*; more about this later).

* * * * *

Using a basis to specify a linear transformation

- In this section we discuss one of the fundamental reasons why bases are important; one can use them to describe linear transformations in a compact way.

- In general, to specify a function $f : X \to Y$, one needs to describe the value $f(x)$ for every point x in X; for instance, if $f : \{1, 2, 3, 4, 5\} \to \mathbf{R}$, then one needs to specify $f(1)$, $f(2)$, $f(3)$, $f(4)$, $f(5)$ in order to completely describe the function. Thus, when X gets large, the amount of data needed to specify a function can get quite large; for instance, to specify a function $f : \mathbf{R}^2 \to \mathbf{R}^3$, one needs to specify a vector $f(x) \in \mathbf{R}^3$ for every single point x in $\mathbf{R}^2$ - and there are infinitely many such points! The remarkable thing, though, is that if f is *linear*, then one does not need to specify f at every single point - one just needs to specify f on a basis and this will determine the rest of the function.

- **Theorem 6** Let V be a finite-dimensional vector space, and let $\{v_1, \ldots, v_n\}$ be a basis for V. Let W be another vector space, and let $w_1, \ldots, w_n$ be some vectors in W. Then there exists exactly one linear transformation $T : V \to W$ such that $Tv_j = w_j$ for each $j = 1, 2, \ldots, n$.

- **Proof** We need to show two things: firstly, that there exists a linear transformation T with the desired properties, and secondly that there is at most one such transformation.

- Let's first show that there is at most one transformation. Suppose for contradiction that we had two different linear transformations $T : V \to W$ and $U : V \to W$ such that $Tv_j = w_j$ and $Uv_j = w_j$ for each $j = 1, \ldots, n$. Now take any vector $v \in V$, and consider Tv and Uv.

- Since $\{v_1, \ldots, v_n\}$ is a basis of V, we have a unique representation

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

where $a_1, \ldots, a_n$ are scalars. Thus, since T is linear

$$Tv = a_1 T v_1 + a_2 T v_2 + \ldots + a_n T v_n$$

but since $Tv_j = w_j$, we have

$$Tv = a_1 w_1 + a_2 w_2 + \ldots + a_n w_n.$$

Arguing similarly with U instead of T, we have

$$Uv = a_1 w_1 + a_2 w_2 + \ldots + a_n w_n.$$

so in particular $Tv = Uv$ for all vectors v. Thus T and U are exactly the same linear transformation, a contradiction. Thus there is at most one linear transformation.

- Now we need to show that there is at least one linear transformation T for which $Tv_j = w_j$. To do this, we need to specify Tv for every vector $v \in V$, and then verify that T is linear. Well, guided by our previous arguments, we know how to find Tv: we first decompose v as a linear combination of $v_1, \ldots, v_n$

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

and then *define* Tv by the formula above:

$$Tv := a_1 w_1 + \ldots + a_n w_n.$$

This is a well-defined construction, since the scalars $a_1, \ldots, a_n$ are unique (see the Lemma on page 36 of week 1 notes). To check that $Tv_j = w_j$, note that

$$v_j = 0v_1 + \ldots + 0v_{j-1} + 1v_j + 0v_{j+1} + \ldots + 0v_n$$

and thus by definition of T

$$Tv_j = 0w_1 + \ldots + 0w_{j-1} + 1w_j + 0v_{j+1} + \ldots + 0w_n = w_j$$

as desired.

- It remains to verify that T is linear; i.e. that $T(v + v') = Tv + Tv'$ and that $T(cv) = cTv$ for all vectors $v, v' \in V$ and scalars c.

- We'll just verify that $T(v + v') = Tv + Tv'$, and leave $T(cv) = cTv$ as an exercise. Fix any $v, v' \in V$. We can decompose

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

and

$$v' = b_1 v_1 + \ldots + b_n v_n$$

for some scalars $a_1, \ldots, a_n, b_1, \ldots, b_n$. Thus, by definition of T,

$$Tv = a_1 w_1 + \ldots + a_n w_n$$

and

$$Tv' = b_1 w_1 + \ldots + b_n w_n$$

and thus

$$Tv + Tv' = (a_1 + b_1) w_1 + \ldots + (a_n + b_n) w_n.$$

On the other hand, adding our representations of v and v' we have

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

and thus by the definition of T again

$$T(v + v') = (a_1 + b_1) w_1 + \ldots + (a_n + b_n) w_n$$

and so $T(v+v') = Tv + Tv'$ as desired. The derivation of $T(cv) = cTv$ is similar and is left as an exercise. This completes the construction of T and the verification of the desired properties. $\square$

- Example: We know that $\mathbf{R}^2$ has $\{(1,0),(0,1)\}$ as a basis. Thus, by Theorem 6, for any vector space W and any vectors w_1, w_2 in W, there is exactly one linear transform $T : \mathbf{R}^2 \to W$ such that $T(1,0) = w_1$ and $T(0,1) = w_2$. Indeed, this transformation is given by

$$T(x,y) := x w_1 + y w_2$$

(why is this transformation linear, and why does it have the desired properties?).

- Example: Let θ be an angle. Suppose we want to understand the operation $Rot_\theta : \mathbf{R}^2 \to \mathbf{R}^2$ of anti-clockwise rotation of $\mathbf{R}^2$ by θ. From elementary geometry one can see that this is a linear transformation. By some elementary trigonometry we see that $Rot_\theta(1,0) = (\cos\theta, \sin\theta)$ and $Rot_\theta(0,1) = (-\sin\theta, \cos\theta)$. Thus from the previous example, we see that

$$Rot_\theta(x,y) = x(\cos\theta, \sin\theta) + y(-\sin\theta, \cos\theta).$$

* * * * *

Co-ordinate bases

- Of all the vector spaces, $\mathbf{R}^n$ is the easiest to work with; every vector v consists of nothing more than n separate scalars - the n *co-ordinates* of the vector. Vectors from other vector spaces - polynomials, matrices, etc. - seem to be more complicated to work with. Fortunately, by using *co-ordinate bases*, one can convert every (finite-dimensional) vector space into a space just like $\mathbf{R}^n$.

- **Definition.** Let V be a finite dimensional vector space. An *ordered basis* of V is an *ordered* sequence $(v_1, \ldots, v_n)$ of vectors in V such that the set $\{v_1, \ldots, v_n\}$ is a basis.

- **Example** The sequence $((1, 0, 0), (0, 1, 0), (0, 0, 1))$ is an ordered basis of $\mathbf{R}^3$; the sequence $((0, 1, 0), (1, 0, 0), (0, 0, 1))$ is a different ordered basis of $\mathbf{R}^3$. (Thus sequences are different from sets; rearranging the elements of a set does not affect the set).

- More generally, if we work in $\mathbf{R}^n$, and we let e_j be the vector with j^{th} co-ordinate 1 and all other co-ordinates 0, then the sequence $(e_1, e_2, \ldots, e_n)$ is an ordered basis of $\mathbf{R}^n$, and is known as the *standard ordered basis* for $\mathbf{R}^n$. In a similar spirit, $(1, x, x^2, \ldots, x^n)$ is known as the *standard ordered basis* for $P_n(\mathbf{R})$.

- Ordered bases are also called *co-ordinate bases*; we shall often give bases names such as β. The reason why we need ordered bases is so that we can refer to the *first* basis vector, *second* basis vector, etc. (In a set, which is unordered, one cannot refer to the first element, second element, etc. - they are all jumbled together and are just plain elements).

- Let $\beta = (v_1, \ldots, v_n)$ be an ordered basis for V, and let v be a vector in V. From the Lemma on page 36 of Week 1 notes, we know that v has a unique representation of the form

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

The scalars $a_1, \ldots, a_n$ will be referred to as the *co-ordinates of v with respect to β*, and we define the *co-ordinate vector of v relative to β*, denoted $[v]^\beta$, by

$$[v]^\beta := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

(In the textbook, $[v]_\beta$ is used instead of $[v]^\beta$. I believe this is a mistake - there is a convention that superscripts should refer to column vectors and subscripts to row vectors - although this distinction is of course very minor. This convention becomes very useful in physics, especially when one begins to study *tensors* - a generalization of vectors and matrices - but for this course, please don't worry too much about whether an index should be a subscript or superscript.)

- **Example** Let's work in $\mathbf{R}^3$, and let $v := (3, 4, 5)$. If β is the standard ordered basis $\beta := ((1, 0, 0), (0, 1, 0), (0, 0, 1))$, then

$$[v]^\beta = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$$

since

$$(3, 4, 5) = 3(1, 0, 0) + 4(0, 1, 0) + 5(0, 0, 1).$$

On the other hand, if we use the ordered basis $\beta' := ((0, 1, 0), (1, 0, 0), (0, 0, 1))$, then

$$[v]^{\beta'} = \begin{pmatrix} 4 \\ 3 \\ 5 \end{pmatrix}$$

since

$$(3, 4, 5) = 4(0, 1, 0) + 3(1, 0, 0) + 5(0, 0, 1).$$

If instead we use the basis $\beta'' := ((3, 4, 5), (0, 1, 0), (0, 0, 1))$, then

$$[v]^{\beta'} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

since

$$(3, 4, 5) = 1(3, 4, 5) + 0(0, 1, 0) + 0(0, 0, 1).$$

For more general bases, one would probably have to do some Gaussian elimination to work out exactly what the co-ordinate vector is (similar to what we did in Week 1).

- **Example** Now let's work in $P_2(\mathbf{R})$, and let $f = 3x^2 + 4x + 6$. If β is the standard ordered basis $\beta := (1, x, x^2)$, then

$$[f]^\beta = \begin{pmatrix} 6 \\ 4 \\ 3 \end{pmatrix}$$

since

$$f = 6 \times 1 + 4 \times x + 3 \times x^2.$$

Or using the reverse standard ordered basis $\beta' := (x^2, x, 1)$, we have

$$[f]^{\beta'} = \begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix}$$

since

$$f = 3 \times x^2 + 4 \times x + 6 \times 1.$$

Note that while $\begin{pmatrix} 6 \\ 4 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix}$ are clearly different column vectors, they both came from the same object f. It's like how one person may perceive a pole as being 12 feet long and another may perceive it as being 4 yards long; both are correct, even though 12 is not equal to 4. It's just that one person is using feet as a basis for length and the other is using yards as a basis for length. (Units of measurement are to scalars as bases are to vectors. To be pedantic, the space V of all possible lengths is a one-dimensional vector space, and both $(yard)$ and $(foot)$ are bases. A length v might be equal to 4 yards, so that $[v]^{(yard)} = (4)$, while also being equal to 12 feet, so $[v]^{(foot)} = 12$).

- Given any vector v and any ordered basis β, we can construct the co-ordinate vector $[v]^\beta$. Conversely, given the co-ordinate vector

$$[v]^\beta = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

and the ordered basis $\beta = (v_1, \ldots, v_n)$, one can reconstruct v by the formula

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

Thus, for any fixed basis, one can go back and forth between vectors v and column vectors $[v]^\beta$ without any difficulty.

- Thus, the use of co-ordinate vectors gives us a way to represent any vector as a familiar column vector, provided that we supply a basis β. The above examples show that the choice of basis β is important; different bases give different co-ordinate vectors.

- A philosophical point: This flexibility in choosing bases underlies a basic fact about the standard Cartesian grid structure, with its x and y axes, etc: it is artificial! (though of course very convenient for computations). The plane $\mathbf{R}^2$ is a very natural object, but our Cartesian grid is not (the ancient Greeks were working with the plane back in 300 BC, but Descartes only introduced the grid in the 1700s). Why couldn't we make, for instance, the x-axis point northwest and the y-axis point northeast? This would correspond to a different basis (for instance, using $((1,1),(1,-1))$ instead of $((1,0),(0,1))$ but one could still do all of geometry, calculus, etc. perfectly well with this grid.

- (The way mathematicians describe this is: the plane is *canonical*, but the Cartesian co-ordinate system is *non-canonical*. Canonical means that there is a natural way to define this object uniquely, without recourse to any artificial convention.)

- As we will see later, it does make sense every now and then to shift one's co-ordinate system to suit the situation - for instance, the above basis $((1,1),(1,-1))$ might be useful in dealing with shapes which were always at 45 degree angles to the horizontal (i.e. diamond-shaped objects). But in the majority of cases, the standard basis suffices, if for no reason other than tradition.

- The very operation of sending a vector v to its co-ordinate vector $[v]^\beta$ is itself a linear transformation, from V to $\mathbf{R}^n$: see this week's homework.

* * * * *

The matrix representation of linear transformations

- We have just seen that by using an ordered basis of V, we can represent vectors in V as column vectors. Now we show that by using an ordered basis of V and another ordered basis of W, we can represent *linear transformations* from V to W as *matrices*. This is a very fundamental observation in this course; it means that from now on, we can study linear transformations by focusing on matrices, which is exactly what we will be doing for the rest of this course.

- Specifically, let V and W be finite-dimensional vector spaces, and let $\beta := (v_1, \ldots, v_n)$ and $\gamma = (w_1, \ldots, w_m)$ be ordered bases for V and W respectively; thus $\{v_1, \ldots, v_n\}$ is a basis for V and $\{w_1, \ldots, w_m\}$ is a basis for W, so that V is n-dimensional and W is m-dimensional. Let T be a linear transformation from V to W.

- **Example** Let $V = P_3(\mathbf{R})$, $W = P_2(\mathbf{R})$, and $T : V \to W$ be the differentiation map $Tf := f'$. We use the standard ordered basis $\beta := (1, x, x^2, x^3)$ for V, and the standard ordered basis $\gamma := (1, x, x^2)$ for W. We shall continue with this example later.

- Returning now to the general situation, let us take a vector v in V and try to compute Tv using our bases. Since v is in V, it has a co-ordinate representation
$$[v]^\beta = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

with respect to β. Similarly, since Tv is in W, it has a co-ordinate representation

$$[Tv]^\gamma = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$$

with respect to γ. Our question is now: how are the column vectors $[v]^\beta$ and $[Tv]^\gamma$ related? More precisely, if we know the column vector $[v]^\beta$, can we work out what $[Tv]^\gamma$ will be? Of course, the answer will depend on T; but as we shall see, we can quantify this more precisely, by saying that the answer will depend on a certain *matrix representation* of T with respect to β and γ.

- **Example** Continuing our previous example, let's pick a $v \in P_3(\mathbf{R})$ at random, say $v := 3x^2 + 7x + 5$, so that

$$[v]^\beta = \begin{pmatrix} 5 \\ 7 \\ 3 \\ 0 \end{pmatrix}.$$

Then we have $Tv = 6x + 7$, so that

$$[Tv]^\gamma = \begin{pmatrix} 7 \\ 6 \\ 0 \end{pmatrix}.$$

The question here is this: starting from the column vector $\begin{pmatrix} 5 \\ 7 \\ 3 \\ 0 \end{pmatrix}$ for v, how does one work out the column vector $\begin{pmatrix} 7 \\ 6 \\ 0 \end{pmatrix}$ for Tv?

- Return now to the general case. From our formula for $[v]_\beta$, we have

$$v = x_1 v_1 + x_2 v_2 + \ldots + x_n v_n,$$

so if we apply T to both sides we obtain

$$Tv = x_1 T v_1 + x_2 T v_2 + \ldots + x_n T v_n. \tag{0.6}$$

while from our formula for $[Tv]^\gamma$ we have

$$Tv = y_1 w_1 + y_2 w_2 + \ldots + y_m w_m. \tag{0.7}$$

Now to connect the two formulae. The vectors $Tv_1, \ldots, Tv_n$ lie in W, and so they are linear combinations of $w_1, \ldots, w_m$:

$$\begin{aligned}
Tv_1 &= a_{11} w_1 + a_{21} w_2 + \ldots + a_{m1} w_m \\
Tv_2 &= a_{12} w_1 + a_{22} w_2 + \ldots + a_{m2} w_m \\
&\vdots \\
Tv_n &= a_{1n} w_1 + a_{2n} w_2 + \ldots + a_{mn} w_m;
\end{aligned}$$

note that the numbers $a_{11}, \ldots, a_{nm}$ are scalars that only depend on T, β, and γ (the vector v is only relevant for computing the x's and y's). Substituting the above formulae into (0.6) we obtain

$$\begin{aligned}
Tv = \ & x_1(a_{11} w_1 + \ldots + a_{m1} w_m) \\
& + x_2(a_{12} w_1 + \ldots + a_{m2} w_m) \\
& \vdots \\
& + x_n(a_{1n} w_1 + \ldots + a_{mn} w_m)
\end{aligned}$$

Collecting coefficients and comparing this with (0.7) (remembering that $\{w_1, \ldots, w_m\}$ is a basis, so there is only one way to write Tv as a linear combination of $w_1, \ldots, w_m$) - we obtain

$$\begin{aligned}
y_1 &= a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \\
y_2 &= a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n \\
&\vdots \\
y_m &= a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n.
\end{aligned}$$

This may look like a mess, but it becomes cleaner in matrix form:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

- Thus, if we define $[T]_\beta^\gamma$ to be the matrix

$$[T]_\beta^\gamma := \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

then we have answered our question of how to link $[v]^\beta$ with $[Tv]^\gamma$:

$$[Tv]^\gamma = [T]_\beta^\gamma [v]^\beta.$$

(If you like, the β subscript on the T has "cancelled" the β superscript on the v. This is part of a more general rule, known as the *Einstein summation convention*, which you might encounter in advanced physics courses when dealing with things called *tensors*).

- It is no co-incidence that matrices were so conveniently suitable for this problem; in fact matrices were initially invented for the express purpose of understanding linear transformations in co-ordinates.

- **Example.** We return to our previous example. Note

$$Tv_1 = T1 = 0 = 0w_1 + 0w_2 + 0w_3$$

$$Tv_2 = Tx = 1 = 1w_1 + 0w_2 + 0w_3$$

$$Tv_3 = Tx^2 = 2x = 0w_1 + 2w_2 + 0w_3$$

$$Tv_4 = Tx^3 = 3x^2 = 0w_1 + 0w_2 + 3w_3$$

and hence

$$[T]_\beta^\gamma := \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Thus $[v]^\beta$ and $[Tv]^\gamma$ are linked by the equation

$$[Tv]^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} [v]^\beta,$$

thus for instance returning to our previous example

$$\begin{pmatrix} 7 \\ 6 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 5 \\ 7 \\ 3 \\ 0 \end{pmatrix}.$$

- The matrix $[T]_\beta^\gamma$ is called the *matrix representation of T with respect to the bases β and γ*. Notice that the j^{th} column of $[T]_\beta^\gamma$ is just the co-ordinate vector of Tv_j with respect to γ:

$$[T]_\beta^\gamma = ([Tv_1]^\gamma \quad [Tv_2]^\gamma \quad \dots \quad [Tv_n]^\gamma)$$

(see for instance the previous Example).

- In many cases, V will be equal to W, and β equal to γ; in that case we may abbreviate $[T]_\beta^\gamma$ as $[T]_\beta$.

- Just like a vector v can be reconstructed from its co-ordinate vector $[v]^\beta$ and vice versa (provided one knows what β is, of course), a linear transformation T can be reconstructed from its co-ordinate matrix $[T]_\beta^\gamma$ and vice versa (provided β and γ are given). Indeed, if one knows $[T]_\beta^\gamma$, then one can work out the rule to get from v to Tv as follows: first write v in terms of β, obtaining the co-ordinate vector $[v]^\beta$; multiply this column vector by $[T]_\beta^\gamma$ to obtain $[Tv]^\gamma$, and then use γ to convert this back into the vector Tv.

- **The scalar case** All this stuff may seem very abstract and foreign, but it is just the vector equivalent of something you are already familiar with in the scalar case: conversion of units. Let's give an example. Suppose a car is travelling in a straight line at a steady speed T for a period v of time (yes, the letters are strange, but this is deliberate). Then the distance that this car traverses is of course Tv. Easy enough, but now let's do everything with units.

- Let's say that the period of time v was half an hour, or thirty minutes. It is not quite accurate to say that $v = 1/2$ or $v = 30$; the precise statement (in our notation) is that $[v]^{(hour)} = (1/2)$, or $[v]^{(minute)} = (30)$. (Note that (*hour*) and (*minute*) are both ordered bases for time, which is a one-dimensional vector space). Since our bases just have one element, our column vector has only one row, which makes it a rather silly vector in our case.

- Now suppose that the speed T was twenty miles an hour. Again, it is not quite accurate to say that $T = 20$; the correct statement is that

  $$[T]^{(mile)}_{(hour)} = (20)$$

  since we clearly have

  $$T(1 \times hour) = 20 \times mile.$$

  We can also represent T in other units:

  $$[T]^{(mile)}_{(minute)} = (1/3)$$

  $$[T]^{(kilometer)}_{(hour)} = (32)$$

  etc. In this case our "matrices" are simply 1×1 matrices - pretty boring!

  Now we can work out Tv in miles or kilometers:

  $$[Tv]^{(mile)} = [T]^{(mile)}_{(hour)}[v]^{(hour)} = (20)(1/2) = (10)$$

  or to do things another way

  $$[Tv]^{(mile)} = [T]^{(mile)}_{(minute)}[v]^{(minute)} = (1/3)(30) = (10).$$

  Thus the car travels for 10 miles - which was of course obvious from the problem. The point here is that these strange matrices and bases are not alien objects - they are simply the vector versions of things that you have seen even back in elementary school mathematics.

- **A matrix example** Remember the car company example from Week 1? Let's run an example similar to that. Suppose the car company needs money and labor to make cars. To keep things very simple, let's suppose that the car company only makes exteriors - doors and wheels. Let's say that there are two types of cars: *coupes*, which have two doors and four wheels, and *sedans*, which have four doors and four wheels. Let's say that a wheel requires 2 units of money and 3 units of labor, while a door requires 4 units of money and 5 units of labor.

- We're going to have two vector spaces. The first vector space, V, is the space of *orders* - the car company may have an order v of 2 coupes and 3 sedans, which translates to 16 doors and 20 wheels. Thus

$$[v]^{(coupe,sedan)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and

$$[v]^{(door,wheel)} = \begin{pmatrix} 16 \\ 20 \end{pmatrix};$$

both $(coupe, sedan)$ and $(door, wheel)$ are ordered bases for V. (One could also make other bases, such as $(coupe, wheel)$, although those are rather strange).

- The second vector space, W, is the space of *resources* - in this case, just money and labor. We're only going to use one ordered basis here: $(money, labor)$.

- There is an obvious linear transformation T from V to W - the *cost* (actually, *price* is a more accurate name for T; *cost* should really refer to Tv). Thus, for any order v in V, Tv is the amount of resources required to create v. By our hypotheses,

$$T(door) = 4 \times money + 5 \times labor$$

and

$$T(wheel) = 2 \times money + 3 \times labor$$

so

$$[T]^{(money,labor)}_{(door,wheel)} = \begin{pmatrix} 4 & 2 \\ 5 & 3 \end{pmatrix}.$$

You may also check that

$$[T]^{(money,labor)}_{(coupe,sedan)} = \begin{pmatrix} 16 & 24 \\ 22 & 32 \end{pmatrix}.$$

Thus, for our order v, the cost to make v can be computed as

$$[Tv]^{(money,labor)} = [T]^{(money,labor)}_{(door,wheel)}[v]^{(door,wheel)} = \begin{pmatrix} 4 & 2 \\ 5 & 3 \end{pmatrix}\begin{pmatrix} 16 \\ 20 \end{pmatrix} = \begin{pmatrix} 104 \\ 140 \end{pmatrix}$$

or equivalently as

$$[Tv]^{(money,labor)} = [T]^{(money,labor)}_{(coupe,sedan)}[v]^{(coupe,sedan)} = \begin{pmatrix} 16 & 24 \\ 22 & 32 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 104 \\ 140 \end{pmatrix} -$$

i.e. one needs 104 units of money and 140 units of labor to complete the order. (Can you explain why these two apparently distinct computations gave exactly the same answer, and why this answer is actually the correct cost of this order?). Note how the different bases $(coupe, sedan)$ and $(door, wheel)$ have different advantages and disadvantages; the $(coupe, sedan)$ basis makes the co-ordinate vector for v nice and simple, while the $(door, wheel)$ basis makes the co-ordinate matrix for T nice and simple.

* * * * *

Things to do with linear transformations

- We know that certain operations can be performed on vectors; they can be added together, or multiplied with a scalar. Now we will observe that there are similar operations on *linear transformations*; they can also be added together and multiplied by a scalar, but also (under certain conditions) can also be multiplied with each other.

- **Definition.** Let V and W be vector spaces, and let $S : V \to W$ and $T : V \to W$ be two linear transformations from V to W. We define the *sum* $S + T$ of these transformations to be a third transformation $S + T : V \to W$, defined by

$$(S + T)(v) := Sv + Tv.$$

- **Example.** Let $S : \mathbf{R}^2 \to R^2$ be the doubling transformation, defined by $Sv := 2v$. Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the identity transformation, defined by $Tv := v$. Then $S + T$ is the tripling transformation

$$(S + T)v = Sv + Tv = 2v + v = 3v.$$

- **Lemma 7** The sum of two linear transformations is again a linear transformation.

- **Proof** Let $S : V \to W$, $T : V \to W$ be linear transformations. We need to show that $S + T : V \to W$ is also linear; i.e. it preserves addition and preserves scalar multiplication. Let's just show that it preserves scalar multiplication, i.e. for any $v \in V$ and scalar c, we have to show that
$$(S + T)(cv) = c(S + T)v.$$
But the left-hand side, by definition, is
$$S(cv) + T(cv) = cSv + cTv$$
since S, T are linear. Similarly, the right-hand side is
$$c(Sv + Tv) = cSv + cTv$$
by the axioms of vector spaces. Thus the two are equal. The proof that $S + T$ preserves addition is similar and is left as an exercise. $\square$

- Note that we can only add two linear transformations S, T if they have the same domain and target space; for instance it is not permitted to add the identity transformation on $\mathbf{R}^2$ to the identity transformation on $\mathbf{R}^3$. This is similar to how vectors can only be added if they belong to the same space; a vector in $\mathbf{R}^2$ cannot be added to a vector in $\mathbf{R}^3$.

- **Definition.** Let $T : V \to W$ be a linear transformation, and let c be a scalar. We define the *scalar multiplication* cT of c and T to be the transformation $cT : V \to W$, defined by
$$(cT)(v) = c(Tv).$$

- It is easy to verify that cT is also a linear transformation; we leave this as an exercise.

- **Example** Let $S : \mathbf{R}^2 \to \mathbf{R}^2$ be the doubling transformation, defined by $Sv := 2v$. Then $2S : \mathbf{R}^2 \to \mathbf{R}^2$ is the quadrupling transformation, defined by $2Sv := 4v$.

- **Definition** Let $\mathcal{L}(V, W)$ be the space of linear transformations from V to W.

- **Example** In the examples above, the transformations S, T, $S+T$, and $2S$ all belonged to $\mathcal{L}(\mathbf{R}^2, \mathbf{R}^2)$.

- **Lemma 8** The space $\mathcal{L}(V, W)$ is a subspace of $\mathcal{F}(V, W)$, the space of all functions from V to W. In particular, $\mathcal{L}(V, W)$ is a vector space.

- **Proof** Clearly $\mathcal{L}(V, W)$ is a subset of $\mathcal{F}(V, W)$, since every linear transformation is a transformation. Also, we have seen that the space $\mathcal{L}(V, W)$ of linear transformations from V to W is closed under addition and scalar multiplication. Hence, it is a subspace of the vector space $\mathcal{F}(V, W)$, and is hence itself a vector space. (Alternatively, one could verify each of the vector space axioms (I-VIII) in turn for $\mathcal{L}(V, W)$; this is a tedious but not very difficult exercise). $\square$

- The next basic operation is that of *multiplying* or *composing* two linear transformations.

- **Definition** Let U, V, W be vector spaces. Let $S : V \to W$ be a linear transformation from V to W, and let $T : U \to V$ be a linear transformation from U to V. Then we define the *product* or *composition* $ST : U \to W$ to be the transformation

$$ST(u) := S(T(u)).$$

- **Example** Let $U : \mathbf{R}^\infty \to \mathbf{R}^\infty$ be the right shift operator

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

Then the operator $UU = U^2$ is given by

$$U^2(x_1, x_2, \ldots) := U(U(x_1, x_2, \ldots)) = U(0, x_1, x_2, \ldots) = (0, 0, x_1, x_2, \ldots),$$

i.e. the double right shift.

- **Example** Let $U^* : \mathbf{R}^\infty \to \mathbf{R}^\infty$ be the left-shift operator

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

Then U^*U is the identity map:

$$U^*U(x_1, x_2, \ldots) = U^*(0, x_1, x_2, \ldots) = (x_1, x_2, \ldots)$$

but UU^* is not:

$$UU^*(x_1, x_2, \ldots) = U(x_2, \ldots) = (0, x_2, \ldots).$$

Thus multiplication of operators is *not* commutative.

- Note that in order for ST to be defined, the target space of T has to match the initial space of S. (This is very closely related to the fact that in order for matrix multiplication AB to be well defined, the number of columns of A must equal the number of rows of B).

* * * * *

Addition and multiplication of matrices

- We have just defined addition, scalar multiplication, and composition of linear transformations. On the other hand, we also know how to add, scalar multiply, and multiply matrices. Since linear transformations can be represented (via bases) as matrices, it is thus a natural question as to whether the linear transform notions of addition, scalar multiplication, and composition are in fact compatible with the matrix notions of addition, scalar multiplication, and multiplication. This is indeed the case; we will now show this.

- **Lemma 9** Let V, W be finite-dimensional spaces with ordered bases β, γ respectively. Let $S : V \to W$ and $T : V \to W$ be linear transformations from V to W, and let c be a scalar. Then

  $$[S + T]_\beta^\gamma = [S]_\beta^\gamma + [T]_\beta^\gamma$$

  and

  $$[cT]_\beta^\gamma = c[T]_\beta^\gamma.$$

- **Proof.** We'll just prove the second statement, and leave the first as an exercise. Let's write $\beta = (v_1, \ldots, v_n)$ and $\gamma = (w_1, \ldots, w_n)$, and denote the matrix $[T]_\beta^\gamma$ by

  $$[T]_\beta^\gamma := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Thus

$$\begin{aligned}
Tv_1 &= a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \\
Tv_2 &= a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m \\
&\vdots \\
Tv_n &= a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m;
\end{aligned}$$

Multiplying by c, we obtain

$$\begin{aligned}
(cT)v_1 &= ca_{11}w_1 + ca_{21}w_2 + \ldots + ca_{m1}w_m \\
(cT)v_2 &= ca_{12}w_1 + ca_{22}w_2 + \ldots + ca_{m2}w_m \\
&\vdots \\
(cT)v_n &= ca_{1n}w_1 + ca_{2n}w_2 + \ldots + ca_{mn}w_m
\end{aligned}$$

and thus

$$[cT]_\beta^\gamma := \begin{matrix} ca_{11} & ca_{12} & \ldots & ca_{1n} \\ ca_{21} & ca_{22} & \ldots & ca_{2n} \\ & & \vdots & \\ ca_{m1} & ca_{m2} & \ldots & ca_{mn} \end{matrix},$$

i.e. $[cT]_\beta^\gamma = c[T]_\beta^\gamma$ as desired. □

- We'll leave the corresponding statement connecting composition of linear transformations with matrix multiplication for next week's notes.

Math 115A - Week 4
Textbook sections: 2.3-2.4
Topics covered:

- A quick review of matrices
- Co-ordinate matrices and composition
- Matrices as linear transformations
- Invertible linear transformations (isomorphisms)
- Isomorphic vector spaces

* * * * *

A quick review of matrices

- An $m \times n$ *matrix* is a collection of mn scalars, organized into m rows and n columns:

$$A = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ & & \vdots & \\ A_{m1} & A_{m2} & \ldots & A_{mn} \end{pmatrix}.$$

  If A is a matrix, then A_{jk} refers to the scalar entry in the j^{th} row and k^{th} column. Thus if
  $$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$
  then $A_{11} = 1$, $A_{12} = 2$, $A_{21} = 3$, and $A_{22} = 4$.

- (The word "matrix" is late Latin for "womb"; it is the same root as maternal or matrimony. The idea being that a matrix is a receptacle for holding numbers. Thus the title of the recent Hollywood movie "the Matrix" is a play on words).

- A special example of a matrix is the $n \times n$ *identity matrix* I_n, defined by

$$I_n := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

  or equivalently that $(I_n)_{jk} := 1$ when $j = k$ and $(I_n)_{jk} := 0$ when $j \neq k$.

- If A and B are two $m \times n$ matrices, the sum $A + B$ is another $m \times n$ matrix, defined by adding each component separately, for instance

$$(A + B)_{11} := A_{11} + B_{11}$$

  and more generally

$$(A + B)_{jk} := A_{jk} + B_{jk}.$$

  If A and B have different shapes, then $A + B$ is left undefined.

- The scalar product cA of a scalar c and a matrix A is defined by multiplying each component of the matrix by c:

$$(cA)_{jk} := cA_{jk}.$$

- If A is an $m \times n$ matrix, and B is an $l \times m$ matrix, then the matrix product BA is an $l \times n$ matrix, whose co-ordinates are given by the formula

$$(BA)_{jk} = B_{j1}A_{1k} + B_{j2}A_{2k} + \ldots + B_{jm}A_{mk} = \sum_{i=1}^{m} B_{ji}A_{ik}.$$

  Thus for instance if

$$A := \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

  and

$$B := \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

then

$$(BA)_{11} = B_{11}A_{11} + B_{12}A_{21}; \quad (BA)_{12} = B_{11}A_{12} + B_{12}A_{22}$$

$$(BA)_{21} = B_{21}A_{11} + B_{22}A_{21}; \quad (BA)_{22} = B_{21}A_{12} + B_{22}A_{22}$$

and so

$$BA = \begin{pmatrix} B_{11}A_{11} + B_{12}A_{21} & B_{11}A_{12} + B_{12}A_{22} \\ B_{21}A_{11} + B_{22}A_{21} & B_{21}A_{12} + B_{22}A_{22} \end{pmatrix}$$

or in other words

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} B_{11}A_{11} + B_{12}A_{21} & B_{11}A_{12} + B_{12}A_{22} \\ B_{21}A_{11} + B_{22}A_{21} & B_{21}A_{12} + B_{22}A_{22} \end{pmatrix}.$$

If the number of columns of B does not equal the number of rows of A, then BA is left undefined. Thus for instance it is possible for BA to be defined while AB remains undefined.

- This matrix multiplication rule may seem strange, but we will explain why it is natural below.

- It is an easy exercise to show that if A is an $m \times n$ matrix, then $I_m A = A$ and $AI_n = A$. Thus the matrices I_m and I_n are multiplicative identities, assuming that the shapes of all the matrices are such that matrix multiplication is defined.

* * * * *

Co-ordinate matrices and composition

- Last week, we introduced the notion of a linear transformation $T : X \to Y$. Given two linear transformations $T : X \to Y$ and $S : Y \to Z$, where the target space of T matches up with the initial space of S, their *composition* $ST : X \to Z$, defined by

$$ST(v) = S(Tv)$$

is also a linear transformation; this is easy to check and I'll leave it as an exercise. Also, if $I_X : X \to X$ is the identity on X and $I_Y : Y \to Y$ is the identity on Y, it is easy to check that $TI_X = T$ and $I_Y T = T$.

- **Example** Suppose we are considering combinations of two molecules: methane CH_4 and water H_2O. Let X be the space of all linear combinations of such molecules, thus X is a two-dimensional space with $\alpha := (methane, water)$ as an ordered basis. (A typical element of X might be $3 \times methane + 2 \times water$). Let Y be the space of all linear combinations of Hydrogen, Carbon, and Oxygen atoms; this is a three-dimensional space with $\beta := (hydrogen, carbon, oxygen)$ as an ordered basis. Let Z be the space of all linear combinations of electrons, protons, and neutrons, thus it is a three-dimensional space with $\gamma := (electron, proton, neutron)$ as a basis. There is an obvious linear transformation $T : X \to Y$, defined by starting with a collection of molecules and breaking them up into component atoms. Thus

$$T(methane) = 4 \times hydrogen + 1 \times carbon$$

$$T(water) = 2 \times hydrogen + 1 \times oxygen$$

and so T has the matrix

$$[T]_\alpha^\beta = [T]_{(methane,water)}^{(hydrogen,carbon,oxygen)} = \begin{pmatrix} 4 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Similarly, there is an obvious linear transformation $S : Y \to Z$, defined by starting with a collection of atoms and breaking them up into component particles. Thus

$$S(hydrogen) = 1 \times electron + 1 \times proton$$

$$S(carbon) = 6 \times electron + 6 \times proton + 6 \times neutron$$

$$S(oxygen) = 8 \times electron + 8 \times proton + 8 \times neutron.$$

Thus

$$[S]_\beta^\gamma = [S]_{(hydrogen,carbon,oxygen)}^{(electron,proton,neutron)} = \begin{pmatrix} 1 & 6 & 8 \\ 1 & 6 & 8 \\ 0 & 6 & 8 \end{pmatrix}.$$

The composition $ST : X \to Z$ of S and T is thus the transformation which sends molecules to their component particles. (Note that even though S is to the left of T, the operation T is applied *first*. This rather unfortunate fact occurs because the conventions of mathematics place the operator T before the operand x, thus we have $T(x)$ instead of $(x)T$. Since all the conventions are pretty much entrenched, there's not much we can do about it). A brief calculation shows that

$$ST(methane) = 10 \times electron + 10 \times proton + 6 \times neutron$$

$$ST(water) = 10 \times electron + 10 \times proton + 8 \times neutron$$

and hence

$$[ST]_\alpha^\gamma = [ST]_{(methane,water)}^{(electron,proton,neutron)} = \begin{pmatrix} 10 & 10 \\ 10 & 10 \\ 6 & 8 \end{pmatrix}.$$

Now we ask the following question: how are these matrices $[T]_\alpha^\beta$, $[S]_\beta^\gamma$, and $[ST]_\alpha^\gamma$ related?

- Let's consider the 10 entry on the top left of $[ST]_\alpha^\gamma$. This number measures how many electrons there are in a methane molecule. From the matrix of $[T]_\alpha^\beta$ we see that each methane molecule has 4 hydrogen, 1 carbon, and 0 oxygen atoms. Since hydrogen has 1 electron, carbon has 6, and oxygen has 8, we see that the number of electrons in methane is

$$4 \times 1 + 1 \times 6 + 0 \times 8 = 10.$$

Arguing similarly for the other entries of $[ST]_\alpha^\gamma$, we see that

$$[ST]_\alpha^\gamma = \begin{pmatrix} 4 \times 1 + 1 \times 6 + 0 \times 8 & 2 \times 1 + 0 \times 6 + 1 \times 8 \\ 4 \times 1 + 1 \times 6 + 0 \times 8 & 2 \times 1 + 0 \times 6 + 1 \times 8 \\ 4 \times 0 + 1 \times 6 + 0 \times 8 & 2 \times 0 + 0 \times 6 + 1 \times 8 \end{pmatrix}.$$

But this is just the matrix product of $[S]_\beta^\gamma$ and $[T]_\alpha^\beta$:

$$[ST]_\alpha^\gamma = \begin{pmatrix} 1 & 6 & 8 \\ 1 & 6 & 8 \\ 0 & 6 & 8 \end{pmatrix} \begin{pmatrix} 4 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = [S]_\beta^\gamma [T]_\alpha^\beta.$$

- More generally, we have

- **Theorem 1.** Suppose that X is l-dimensional and has an ordered basis $\alpha = (u_1, \ldots, u_l)$, Y is m-dimensional and has an ordered basis $\beta = (v_1, \ldots, v_m)$, and Z is n-dimensional and has a basis γ of n elements. Let $T : X \to Y$ and $S : Y \to Z$ be linear transformations. Then

$$[ST]^\gamma_\alpha = [S]^\gamma_\beta [T]^\beta_\alpha.$$

- **Proof.** The transformation T has a co-ordinate matrix $[T]^\beta_\alpha$, which is an $m \times l$ matrix. If we write

$$[T]^\beta_\alpha =:= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1l} \\ a_{21} & a_{22} & \ldots & a_{2l} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{ml} \end{pmatrix}$$

then we have

$$\begin{aligned} Tu_1 &= a_{11}v_1 + a_{21}v_2 + \ldots + a_{m1}v_m \\ Tu_2 &= a_{12}v_1 + a_{22}v_2 + \ldots + a_{m2}v_m \\ &\vdots \\ Tu_l &= a_{1l}v_1 + a_{2l}v_2 + \ldots + a_{ml}v_m \end{aligned}$$

We write this more compactly as

$$Tu_i = \sum_{j=1}^m a_{ji} v_j \text{ for } i = 1, \ldots, l.$$

- Similarly, S has a co-ordinate matrix $[S]^\gamma_\beta$, which is an $n \times m$ matrix. If

$$[S]^\gamma_\beta := \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1m} \\ b_{21} & b_{22} & \ldots & b_{2m} \\ & & \vdots & \\ b_{n1} & b_{m2} & \ldots & b_{nm} \end{pmatrix}.$$

then

$$Sv_j = \sum_{k=1}^n b_{kj} w_k \text{ for } j = 1, \ldots, m.$$

Now we try to understand how ST acts on the basis $u_1, \ldots, u_l$. Applying S to both sides of the T equations, and using the fact that S is linear, we obtain

$$STu_i = \sum_{j=1}^{m} a_{ji} S v_j.$$

Applying our formula for Sv_j, we obtain

$$STu_i = \sum_{j=1}^{m} a_{ji} \sum_{k=1}^{n} b_{kj} w_k$$

which we can rearrange as

$$STu_i = \sum_{k=1}^{n} (\sum_{j=1}^{m} b_{kj} a_{ji}) w_k.$$

Thus if we define

$$c_{ki} := \sum_{j=1}^{m} b_{kj} a_{ji} = b_{k1} a_{1i} + b_{k2} a_{2i} + \ldots + b_{km} a_{mi}$$

then we have

$$STu_i = \sum_{k=1}^{n} c_{ki} w_k$$

and hence

$$[ST]_\alpha^\gamma = \begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1l} \\ c_{21} & c_{22} & \ldots & c_{2l} \\ & & \vdots & \\ c_{n1} & c_{m2} & \ldots & c_{nl} \end{pmatrix}.$$

However, if we perform the matrix multiplication

$$\begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1m} \\ b_{21} & b_{22} & \ldots & b_{2m} \\ & & \vdots & \\ b_{n1} & b_{m2} & \ldots & b_{nm} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1l} \\ a_{21} & a_{22} & \ldots & a_{2l} \\ & & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{ml} \end{pmatrix}$$

we get exactly the same matrix (this is because of our formula for c_{ki} in terms of the b and a co-efficients). This proves the theorem. $\square$

- This theorem illustrates *why* matrix multiplication is defined in that strange way - multiplying rows against columns, etc. It also explains why we need the number of columns of the left matrix to equal the number of rows of the right matrix; this is like how to compose two transformations $T : X \to Y$ and $S : Y \to Z$ to form a transformation $ST : X \to Z$, we need the target space of T to equal to the initial space of S.

* * * * *

Comparison between linear transformations and matrices

- To summarize what we have done so far:

- Given a vector space X and an ordered basis α for X, one can write vectors v in V as column vectors $[v]^\alpha$. Given two vector spaces X, Y, and ordered bases α, β for X and Y respectively, we can write linear transformations $T : X \to Y$ as matrices $[T]^\beta_\alpha$. The action of T then corresponds to matrix multiplication by $[T]^\gamma_\beta$:

$$[Tv]^\beta = [T]^\beta_\alpha [v]^\alpha;$$

  i.e. we can "cancel" the basis α. Similarly, composition of two linear transformations corresponds to matrix multiplication: if $S : Y \to Z$ and γ is an ordered basis for Z, then

$$[ST]^\gamma_\alpha = [S]^\gamma_\beta [T]^\beta_\alpha$$

  i.e. we can "cancel" the basis β.

- Thus, by using bases, one can understand the behavior of linear transformations in terms of matrix multiplication. This is not quite saying that linear transformations are the same as matrices, for two reasons: firstly, this correspondence only works for finite dimensional spaces X, Y, Z; and secondly, the matrix you get depends on the basis you choose - a single linear transformation can correspond to many different matrices, depending on what bases one picks.

- To clarify the relationship between linear transformations and matrices let us once again turn to the scalar case, and now consider currency conversions. Let X be the space of US currency - this is the one-dimensional space which has $(dollar)$ as an (ordered) basis; $(cent)$ is also a basis. Let Y be the space of British currency (with $(pound)$ or $(penny)$ as a basis; $pound = 100 \times penny$), and let Z be the space of Japanese currency (with (yen) as a basis). Let $T : X \to Y$ be the operation of converting US currency to British, and $S : Y \to Z$ the operation of converting British currency to Japanese, thus $ST : X \to Z$ is the operation of converting US currency to Japanese (via British).

- Suppose that one dollar converted to half a pound, then we would have

$$[T]^{(pound)}_{(dollar)} = (0.5),$$

or in different bases

$$[T]^{(pound)}_{(cent)} = (0.005); \quad [T]^{(penny)}_{(cent)} = (0.5); \quad [T]^{(penny)}_{(dollar)} = (50).$$

Thus the same linear transformation T corresponds to many different 1×1 matrices, depending on the choice of bases both for the domain X and the range Y. However, conversion works properly no matter what basis you pick (as long as you are consistent), e.g.

$$[v]^{(dollar)} = (6) \Rightarrow [Tv]^{(pound)} = [T]^{(pound)}_{(dollar)}[v]^{(dollar)} = (0.5)(6) = (3).$$

Furthermore, if each pound converted to 200 yen, so that

$$[S]^{(yen)}_{(pound)} = (200)$$

then we can work out the various matrices for ST by matrix multiplication (which in the 1×1 case is just scalar multiplication):

$$[ST]^{(yen)}_{(dollar)} = [S]^{(yen)}_{(pound)}[T]^{(pound)}_{(dollar)} = (200)(0.5) = (100).$$

One can of course do this computation in different bases, but still get the same result, since the intermediate basis just cancels itself out at the end:

$$[ST]^{(yen)}_{(dollar)} = [S]^{(yen)}_{(penny)}[T]^{(penny)}_{(dollar)} = (2)(50) = (100)$$

etc.

- You might amuse yourself concocting a vector example of currency conversion - for instance, suppose that in some country there was more than one type of currency, and they were not freely interconvertible. A US dollar might then convert to x amounts of one currency plus y amounts of another, and so forth. Then you could repeat the above computations except that the scalars would have to be replaced by various vectors and matrices.

- One basic example of a linear transformation is the *identity transformation* $I_V : V \to V$ on a vector space V, defined by $I_V v = v$. If we pick any basis $\beta = (v_1, \ldots, v_n)$ of V, then of course we have

$$I_V v_1 = 1 \times v_1 + 0 \times v_2 + \ldots + 0 \times v_n$$

$$I_V v_2 = 0 \times v_1 + 1 \times v_2 + \ldots + 0 \times v_n$$

$$\ldots$$

$$I_V v_n = 0 \times v_1 + 0 \times v_2 + \ldots + 1 \times v_n$$

and thus

$$[I_V]_\beta^\beta = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & & \vdots & \\ 0 & 0 & \ldots & 1 \end{pmatrix} = I_n.$$

Thus the identity transformation is connected to the identity matrix.

* * * * *

Matrices as linear transformations.

- We have now seen how linear transformations can be viewed as matrices (after selecting bases, etc.). Conversely, every matrix can be viewed as a linear transformation.

- **Definition** Let A be an $m \times n$ matrix. Then we define the linear transformation $L_A : \mathbf{R}^n \to \mathbf{R}^m$ by the rule

$$L_A x := Ax \text{ for all } x \in \mathbf{R}^n,$$

where we think of the vectors in $\mathbf{R}^n$ and $\mathbf{R}^m$ as column vectors.

- **Example** Let A be the matrix

$$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}.$$

Then $L_A : \mathbf{R}^2 \to \mathbf{R}^3$ is the linear transformation

$$L_A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{pmatrix}.$$

- It is easily checked that L_A is indeed linear. Thus for every $m \times n$ matrix A we can associate a linear transformation $L_A : \mathbf{R}^n \to \mathbf{R}^m$. Conversely, if we let α be the standard basis for $\mathbf{R}^n$ and β be the standard basis for $\mathbf{R}^m$, then for every linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$ we can associate an $m \times n$ matrix $[T]_\alpha^\beta$. The following simple lemma shows that these two operations invert each other:

- **Lemma 2.** Let the notation be as above. If A is an $m \times n$ matrix, then $[L_A]_\alpha^\beta = A$. If $T : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, then $L_{[T]_\alpha^\beta} = T$.

- **Proof** Let $\alpha = (e_1, e_2, \ldots, e_n)$ be the standard basis of $\mathbf{R}^n$. For any column vector

$$x = \begin{pmatrix} x_1 \\ \ldots \\ x_n \end{pmatrix}$$

in $\mathbf{R}^n$, we have

$$x = x_1 e_1 + \ldots x_n e_n$$

and thus

$$[x]^\alpha = \begin{pmatrix} x_1 \\ \ldots \\ x_n \end{pmatrix} = x.$$

Thus $[x]^\alpha = x$ for all $x \in \mathbf{R}^n$. Similarly we have $[y]^\beta = y$ for all $y \in \mathbf{R}^m$.

- Now let A be an $m \times n$ matrix, and let $x \in \mathbf{R}^n$. By definition

$$L_A x = Ax$$

On the other hand, we have

$$[L_A x]^\beta = [L_A]^\beta_\alpha [x]^\alpha$$

and hence (by the previous discussion)

$$L_A x = [L_A]^\beta_\alpha x.$$

Thus

$$[L_A]^\beta_\alpha x = Ax \text{ for all } x \in \mathbf{R}^n.$$

If we apply this with x equal to the first basis vector $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, we see that the first column of the matrices $[L_A]^\beta_\alpha$ and A are equal. Similarly we see that all the other columns of $[L_A]^\beta_\alpha$ and A match, so that $[L_A]^\beta_\alpha = A$ as desired.

- Now let $T : \mathbf{R}^n \to \mathbf{R}^m$ be a linear transformation. Then for any $x \in \mathbf{R}^n$

$$[Tx]^\beta = [T]^\beta_\alpha [x]^\alpha$$

which by previous discussion implies that

$$Tx = [T]^\beta_\alpha x = L_{[T]^\beta_\alpha} x.$$

Thus T and $L_{[T]^\beta_\alpha}$ are the same linear transformation, and the lemma is proved. □

- Because of the above lemma, any result we can say about linear transformations, one can also say about matrices. For instance, the following result is trivial for linear transformations:

- **Lemma 3. (Composition is associative)** Let $T : X \to Y$, $S : Y \to Z$, and $R : Z \to W$ be linear transformations. Then we have $R(ST) = (RS)T$.

- **Proof.** We have to show that $R(ST)(x) = (RS)T(x)$ for all $x \in X$. But by definition

$$R(ST)(x) = R((ST)(x)) = R(S(T(x)) = (RS)(T(x)) = (RS)T(x)$$

as desired. □

- **Corollary 4. (Matrix multiplication is associative)** Let A be an $m \times n$ matrix, B be a $l \times m$ matrix, and C be a $k \times l$ matrix. Then $C(BA) = (CB)A$.

- **Proof** Since $L_A : \mathbf{R}^n \to \mathbf{R}^m$, $L_B : \mathbf{R}^m \to \mathbf{R}^l$, and $L_C : \mathbf{R}^l \to \mathbf{R}^k$ are linear transformations, we have from the previous Lemma that

$$L_C(L_B L_A) = (L_C L_B) L_A.$$

Let $\alpha, \beta, \gamma, \delta$ be the standard bases of $\mathbf{R}^n$, $\mathbf{R}^m$, $\mathbf{R}^l$, and $\mathbf{R}^k$ respectively. Then we have

$$[L_C(L_B L_A)]_\alpha^\delta = [L_C]_\gamma^\delta [L_B L_A]_\alpha^\gamma = [L_C]_\gamma^\delta ([L_B]_\beta^\gamma [L_A]_\alpha^\beta) = C(BA)$$

while

$$[(L_C L_B) L_A]_\alpha^\delta = [L_C L_B]_\beta^\delta [L_A]_\alpha^\beta = ([L_C]_\gamma^\delta [L_B]_\beta^\gamma)[L_A]_\alpha^\beta = (CB)A$$

using Lemma 2. Combining these three identities we see that $C(BA) = (CB)A$. $\square$

- The above proof may seem rather weird, but it managed to prove the matrix identity $C(BA) = (CB)A$ without having to do lots and lots of matrix multiplication. Exercise: try proving $C(BA) = (CB)A$ directly by writing out C, B, A in co-ordinates and expanding both sides!

- We have just shown that matrix multiplication is associative. In fact, all the familiar rules of algebra apply to matrices (e.g. $A(B + C) = AB + AC$, and A times the identity is equal to A) provided that all the matrix operations make sense, of course. (The shapes of the matrices have to be compatible before one can even begin to add or multiply them together). The one important caveat is that matrix multiplication is not commutative: AB is usually not the same as BA! Indeed there is no guarantee that these two matrices are the same shape (or even that they are both defined at all).

- Some other properties of A and L_A are stated below. As you can see, the proofs are similar to the ones above.

- If A is an $m \times n$ matrix and B is an $l \times m$ matrix, then $L_{BA} = L_B L_A$. Proof: Let α, β, γ be the standard bases of $\mathbf{R}^n$, $\mathbf{R}^m$, $\mathbf{R}^l$ respectively. Then $L_B L_A$ is a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^l$, and so

$$[L_B L_A]_\alpha^\gamma = [L_B]_\beta^\gamma [L_A]_\alpha^\beta = BA,$$

  and so by taking L of both sides and using Lemma 2, we obtain $L_B L_A = L_{BA}$ as desired.

- If A is an $m \times n$ matrix, and B is another $m \times n$ matrix, then $L_{A+B} = L_A + L_B$. Proof: $L_A + L_B$ is a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$. Let α, β be the standard bases of $\mathbf{R}^n$ and $\mathbf{R}^m$ respectiely. Then

$$[L_A + L_B]_\alpha^\beta = [L_A]_\alpha^\beta + [L_B]_\alpha^\beta = A + B$$

  and so by taking L of both sides and using Lemma 2, we obtain $L_{A+B} = L_A + L_B$ as desired.

* * * * *

Invertible linear transformations

- We have already dealt with the concepts of a linear transformation being one-to-one, and of being onto. We now combine these two concepts to that of a transformation being *invertible*.

- **Definition.** Let $T : V \to W$ be a linear transformation. We say that a linear transformation $S : W \to V$ is the *inverse* of T if $TS = I_W$ and $ST = I_V$. We say that T is *invertible* if it has an inverse, and call the inverse T^{-1}; thus $TT^{-1} = I_W$ and $T^{-1}T = I_V$.

- **Example** Let $T : \mathbf{R}^3 \to \mathbf{R}^3$ be the doubling transformation $Tv := 2v$. Let $S : \mathbf{R}^3 \to R^3$ be the halving transformation $Sv := v/2$. Then S is the inverse of T: $ST(v) = S(2v) = (2v)/2 = v$, while $TS(v) = T(v/2) = 2(v/2) = v$, thus both ST and TS are the identity on $\mathbf{R}^3$.

- Note that this definition is symmetric: if S is the inverse of T, then T is the inverse of S.

- Why do we call S *the* inverse of T instead of just *an* inverse? This is because every transformation can have at most one inverse:

- **Lemma 6**. Let $T : V \to W$ be a linear transformation, and let $S : W \to V$ and $S' : W \to V$ both be inverses of T. Then $S = S'$.

- **Proof**
$$S = SI_W = S(TS') = (ST)S' = I_V S' = S'.$$
$\square$

- Not every linear transformation has an inverse:

- **Lemma 7**. If $T : V \to W$ has an inverse $S : W \to V$, then T must be one-to-one and onto.

- **Proof** Let's show that T is one-to-one. Suppose that $Tv = Tv'$; we have to show that $v = v'$. But by applying S to both sides we get $STv = STv'$, thus $I_V v = I_V v'$, thus $v = v'$ as desired. Now let's show that T is onto. Let $w \in W$; we have to find v such that $Tv = w$. But $w = I_W w = TSw = T(Sw)$, so if we let $v := Sw$ then we have $Tv = w$ as desired. $\square$.

- Thus, for instance, the zero transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ defined by $Tv = 0$ is not invertible.

- The converse of Lemma 7 is also true:

- **Lemma 8**. If $T : V \to W$ is a one-to-one and onto linear transformation , then it has an inverse $S : W \to V$, which is also a linear transformation.

- **Proof** Let $T : V \to W$ be one-to-one and onto. Let w be any element of W. Since T is onto, we have $w = Tv$ for some v in V; since T is one-to-one; this v is unique (we can't have two different elements v, v' of V such that Tv and Tv' are both equal to w). Let us define Sw as equal to this v, thus S is a transformation from W to V. For any $w \in W$, we have $w = Tv$ and $Sw = v$ for some $v \in V$, and hence $TSw = w$; thus TS is the identity I_W.

- Now we show that $ST = I_V$, i.e. that for every $v \in V$, we have $STv = v$. Since we already know that $TS = I_W$, we have that $TSw = w$ for all $w \in W$. In particular we have $TSTv = Tv$, since $Tv \in W$. But since T is injective, this implies that $STv = v$ as desired.

- Finally, we show that S is linear, i.e. that it preserves addition and scalar multiplication. We'll just show that it preserves addition, and leave scalar multiplication as an exercise. Let $w, w' \in W$; we need to show that $S(w + w') = Sw + Sw'$. But we have

  $$T(S(w+w')) = (TS)(w+w') = I_W(w+w') = I_W w + I_W w' = TSw + TSw' = T(Sw+Sw');$$

  since T is one-to-one, this implies that $S(w+w') = Sw+Sw'$ as desired. The preservation of scalar multiplication is proven similarly. $\square$

- Thus a linear transformation is invertible if and only if it is one-to-one and onto. Invertible linear transformations are also known as *isomorphisms*.

- **Definition** Two vector spaces V and W are said to be *isomorphic* if there is an invertible linear transformation $T : V \to W$ from one space to another.

- **Example** The map $T : \mathbf{R}^3 \to P_2(\mathbf{R})$ defined by

  $$T(a, b, c) := ax^2 + bx + c$$

  is easily seen to be linear, one-to-one, and onto, and hence an isomorphism. Thus $\mathbf{R}^3$ and $P_2(\mathbf{R})$ are isomorphic.

- Isomorphic spaces tend to have almost identical properties. Here is an example:

- **Lemma 9**. Two finite-dimensional spaces V and W are isomorphic if and only if $\dim(V) = \dim(W)$.

- **Proof** If V and W are isomorphic, then there is an invertible linear transformation $T : V \to W$ from V to W, which by Lemma 7 is one-to-one and onto. Since T is one-to-one, $\text{nullity}(T) = 0$. Since T is onto, $\text{rank}(T) = \dim(W)$. By the dimension theorem we thus have $\dim(V) = \dim(W)$.

- Now suppose that $\dim(V)$ and $\dim(W)$ are equal; let's say that $\dim(V) = \dim(W) = n$. Then V has a basis $\{v_1, \ldots, v_n\}$, and W has a basis

$\{w_1, \ldots, w_n\}$. By Theorem 6 of last week's notes, we can find a linear transformation $T : V \to W$ such that $Tv_1 = w_1, \ldots, Tv_n = w_n$. By Theorem 3 of last week's notes, $w_1, \ldots, w_n$ must then span $R(T)$. But since $w_1, \ldots, w_n$ span W, we have $R(T) = W$, i.e. T is onto. By Lemma 2 of last week's notes, T is therefore one-to-one, and hence is an isomorphism. Thus V and W are isomorphic. $\square$

- Every basis leads to an isomorphism. If V has a finite basis $\beta = (v_1, \ldots, v_n)$, then the co-ordinate map $\phi_\beta : V \to \mathbf{R}^n$ defined by

$$\phi_\beta(x) := [x]^\beta$$

is a linear transformation (see last week's homework), and is invertible (this was discussed in last week's notes, where we noted that we can reconstruct x from $[x]^\beta$ and vice versa). Thus ϕ_β is an isomorphism between V to $\mathbf{R}^n$. In the textbook ϕ_β is called the *standard representation of V with respect to β.*

- Because of all this theory, we are able to essentially equate finite-dimensional vector spaces V with the standard vector spaces $\mathbf{R}^n$, to equate vectors $v \in V$ with their co-ordinate vectors $[v]^\alpha \in \mathbf{R}^n$ (provided we choose a basis α for V) and linear transformations $T : V \to W$ from one finite-dimensional space to another, with $n \times m$ matrices $[T]_\alpha^\beta$. This means that, for finite-dimensional linear algebra at least, we can reduce everything to the study of column vectors and matrices. This is what we will be doing for the rest of this course.

* * * * *

Invertible linear transformations and invertible matrices

- An $m \times n$ matrix A has an *inverse* B, if B is an $n \times m$ matrix such that $BA = I_n$ and $AB = I_m$. In this case we call A an *invertible matrix*, and denote B by A^{-1}.

- **Example.** If

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

then

$$A^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/4 \end{pmatrix}$$

is the inverse of A, as can be easily checked.

- The relationship between invertible linear transformations and invertible matrices is the following:

- **Theorem 10.** Let V be a vector space with finite ordered basis α, and let W be a vector space with finite ordered basis β. Then a linear transformation $T : V \to W$ is invertible if and only if the matrix $[T]_\alpha^\beta$ is invertible. Furthermore, $([T]_\alpha^\beta)^{-1} = [T^{-1}]_\beta^\alpha$

- **Proof.** Suppose that V is n-dimensional and W is m-dimensional; this makes $[T]_\alpha^\beta$ an $m \times n$ matrix.

- First suppose that $T : V \to W$ has an inverse $T^{-1} : W \to V$. Then

$$[T]_\alpha^\beta [T^{-1}]_\beta^\alpha = [TT^{-1}]_\beta^\beta = [I_W]_\beta^\beta = I_m$$

while

$$[T^{-1}]_\beta^\alpha [T]_\alpha^\beta = [T^{-1}T]_\alpha^\alpha = [I_V]_\alpha^\alpha = I_n,$$

thus $[T^{-1}]_\beta^\alpha$ is the inverse of $[T]_\alpha^\beta$ and so $[T]_\alpha^\beta$ is invertible.

- Now suppose that $[T]_\alpha^\beta$ is invertible, with inverse B. We'll prove shortly that there exists a linear transformation $S : W \to V$ with $[S]_\beta^\alpha = B$. Assuming this for the moment, we have

$$[ST]_\alpha^\alpha = [S]_\beta^\alpha [T]_\alpha^\beta = B[T]_\alpha^\beta = I_n = [I_V]_\alpha^\alpha$$

and hence $ST = I_V$. A similar argument gives $TS = I_W$, and so S is the inverse of T and so T is invertible.

- It remains to show that we can in fact find a transformation $S : W \to V$ with $[S]_\beta^\alpha = B$. Write $\alpha = (v_1, \ldots, v_n)$ and $\beta = (w_1, \ldots, w_m)$. Then we want a linear transformation $S : W \to V$ such that

$$Sw_1 = B_{11}v_1 + \ldots + B_{1n}v_n$$

$$Sw_2 = B_{21}v_1 + \ldots + B_{2n}v_n$$
$$\vdots$$
$$Sw_m = B_{m1}v_1 + \ldots + B_{mn}v_n.$$

But we can do this thanks to Theorem 6 of last week's notes. $\square$

- **Corollary 11**. An $m \times n$ matrix A is invertible if and only if the linear transformation $L_A : \mathbf{R}^n \to \mathbf{R}^m$ is invertible. Furthermore, the inverse of L_A is $L_{A^{-1}}$.

- **Proof.** If α is the standard basis for $\mathbf{R}^n$ and β is the standard basis for $\mathbf{R}^m$, then
$$[L_A]_\alpha^\beta = A.$$
Thus by Theorem 10, A is invertible if and only if L_A is. Also, from Theorem 10 we have
$$[L_A^{-1}]_\beta^\alpha = ([L_A]_\beta^\alpha)^{-1} = A^{-1} = [L_{A^{-1}}]_\beta^\alpha$$
and hence
$$L_A^{-1} = L_{A^{-1}}$$
as desired. $\square$

- **Corollary 12.** In order for a matrix A to be invertible, it must be square (i.e. $m = n$).

- **Proof**. This follows immediately from Corollary 11 and Lemma 9. $\square$

- On the other hand, not all square matrices are invertible; for instance the zero matrix clearly does not have an inverse. More on this in a later week.

Math 115A - Week 5
Textbook sections: 1.1-2.5
Topics covered:

- Co-ordinate changes
- Stuff about the midterm

* * * * *

Changing the basis

- In last week's notes, we used bases to convert vectors to co-ordinate vectors, and linear transformations to matrices. We have already mentioned that if one changes the basis, then the co-ordinate vectors and matrices also change. Now we study this phenomenon more carefully, and quantify exactly how changing the basis changes these co-ordinate vectors and matrices.

- Let's begin with co-ordinate vectors. Suppose we have a vector space V and two ordered bases β, β' of that vector space. Suppose we also have a vector v in V. Then one can write v as a co-ordinate vector either with respect to β - thus obtaining $[v]^\beta$ - or with respect to $[v]^{\beta'}$. The question is now: how are $[v]^\beta$ and $[v]^{\beta'}$ related?

- Fortunately, this question can easily be resolved with the help of the identity operator $I_V : V \to V$ on V. By definition, we have
$$I_V v = v.$$
We now convert this equation to matrices, but with a twist: we measure the domain V using the basis β, but the range V using the basis β'! This gives
$$[I_V]_\beta^{\beta'}[v]^\beta = [v]^{\beta'}.$$
Thus we now know how to convert from basis β to basis β':
$$[v]^{\beta'} = [I_V]_\beta^{\beta'}[v]^\beta. \qquad (0.8)$$
(If you like, the two bases β have "cancelled each other" on the right-hand side).

- **Example.** Let $V = \mathbf{R}^2$, and consider both the standard ordered basis $\beta := ((1,0),(0,1))$ and a non-standard ordered basis $\beta' := ((1,1),(1,-1))$. Let's pick a vector $v \in \mathbf{R}^2$ at random; say $v := (5,3)$. Then one can easily check that $[v]^\beta = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$ and $[v]^{\beta'} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ (why?).

- Now let's work out $[I_V]_\beta^{\beta'}$. Thus we are applying I_V to elements of β and writing them in terms of β'. Since

$$I_V(1,0) = (1,0) = \frac{1}{2}(1,1) + \frac{1}{2}(1,-1)$$

and

$$I_V(0,1) = (0,1) = \frac{1}{2}(1,1) - \frac{1}{2}(1,-1)$$

we thus see that

$$[I_V]_\beta^{\beta'} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

We can indeed verify the formula (0.8), which in this case becomes

$$\begin{pmatrix} 4 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

Note that $[I_V]_\beta^{\beta'}$ is different from $[I_V]_\beta^\beta$, $[I_V]_{\beta'}^{\beta'}$, or $[I_V]_{\beta'}^\beta$. For instance, we have

$$[I_V]_\beta^\beta = [I_V]_{\beta'}^{\beta'} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(why?), while

$$[I_V]_{\beta'}^\beta = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

(why?). Note also that $[I_V]_{\beta'}^\beta$ is the inverse of $[I_V]_\beta^{\beta'}$ (can you see why this should be the case, without doing any matrix multiplication?)

- **A scalar example**. Let V be the space of all lengths; this has a basis $\beta := (yard)$, and a basis $\beta' := (foot)$. Since the identity I_V applied to a yard yields three feet, we have

$$[I_V]_{(yard)}^{(foot)} = (3)$$

(i.e. the identity on lengths is three feet per yard). Thus for any length v,

$$[v]^{(foot)} = [I_V]^{(foot)}_{(yard)}[v]^{(yard)} = (3)[v]^{(yard)}.$$

Thus for instance, if $[v]^{(yard)} = (4)$ (so v is four yards), then $[v]^{(foot)}$ must equal $(3)(4) = (12)$ (i.e. v is also twelve feet).

- Conversely, we have $[I_V]^{(yard)}_{(foot)} = (1/3)$ (i.e. we have 1/3 yards per foot).

- The matrix $[I_V]^{\beta'}_{\beta}$ is called the *change of co-ordinate matrix from β to β'*; it is the matrix we use to multiply by when we want to convert β co-ordinates to β'. Very loosely speaking, $[I_V]^{\beta'}_{\beta}$ measures how much of β' lies in β (just as $[I_V]^{(foot)}_{(yard)}$ measures how many feet lie in a yard).

- Change of co-ordinate matrices are always square (why?) and always invertible (why?).

- **Example** Let $V := P_2(\mathbf{R})$, and consider the two bases $\beta := (1, x, x^2)$ and $\beta' := (x^2, x, 1)$ of V. Then

$$[I_V]^{\beta'}_{\beta} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

(why?).

- **Example** Suppose we have a mixture of carbon dioxide CO_2 and carbon monoxide CO molecule. Let V be the vector space of all such mixtures, so it has an ordered basis $\beta := (CO_2, CO)$. One can also use just the basis $\beta' = (C, O)$ of carbon and oxygen atoms (where we are ignoring the chemical bonds, etc., and treating each molecule as simply the sum of its components, thus $CO_2 = 1 \times C + 2 \times O$ and $CO = 1 \times C + 1 \times O$). Then we have

$$[I_V]^{\beta'}_{\beta} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix};$$

this can be interpreted as saying that CO_2 contains 1 atoms of carbon and 2 atoms of oxygen, while CO contains 1 atom of carbon and 1 atom of oxygen. The inverse change of co-ordinate matrix is

$$[I_V]^{\beta}_{\beta'} = ([I_V]^{\beta'}_{\beta})^{-1} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix};$$

this can be interpreted as saying that an atom of carbon is equivalent to -1 molecules of CO_2 and $+2$ molecules of CO, while an atom of oxygen is equivalent to 1 molecule of CO_2 and -1 molecules of CO. Note how the inverse matrix does not quite behave the way one might naively expect; for instance, given the factor of 2 in the conversion matrix $[I_V]^{\beta'}_{\beta}$, one might expect a factor of 1/2 in the inverse conversion matrix $[I_V]^{\beta}_{\beta'}$; instead we get strange negative numbers all over the place. (To put it another way, since CO_2 contains two atoms of oxygen, why shouldn't oxygen consist of $\frac{1}{2}$ of a molecule of CO_2? Think about it).

- **Example** (This rather lengthy example is only for physics-oriented students who have some exposure to special relativity; everyone else can safely skip this example). One of the fundamental concepts of special relativity is that space and time should be treated together as a single vector space, and that different observers use different bases to measure space and time.

- For simplicity, let us assume that space is one-dimensional; people can move in only two directions, which we will call right and left. Let's say that observers measure time in years, and space in light-years.

- Let's say there are two observers, Alice and Bob. Alice is an *inertial observer*, which means that she is not accelerating. Bob is another inertial observer, but travelling at a fixed speed $\frac{3}{5}c$ to the right, as measured by Alice; here c is of course the speed of light.

- An *event* is something which happens at a specific point in space and time. Any two events are separated by some amount of time and some amount of distance; however the amount of time and distance that separates them depends on which observer is perceiving. For instance, Alice might measure that event Y occurred 8 years later and 4 light-years to the right of event X, while Bob might measure the distance and duration between the two events differently. (In this case, it turns out that B measures Y as occurring 7 years later and 1 light-year to the *left* of event X).

- Let V denote the vector space of all possible displacements in both space and time; this is a two-dimensional vector space, because we are assuming space to be one-dimensional. (In real life, space is three dimensional, and so spacetime is four dimensional). To measure things in this vector space, Alice has a unit of length - let's call it the *Alice-light-year*, and a unit of time - the *Alice-year*. Thus in the above example, if we call v the vector from event X to event Y, then $v = 8 \times Alice - year + 4 \times Alice - light - year$. (We'll adopt the convention that a displacement of length in the right direction is positive, while a displacement in the left direction is negative). Thus Alice uses the ordered basis $(Alice - light - year, Alice - year)$ to span the space V.

- Similarly, Bob has the ordered basis $(Bob - light - year, Bob - year)$. These bases are related by the *Lorentz transformations*

$$Alice - light - year = \frac{5}{4} Bob - light - year - \frac{3}{4} Bob - year$$

$$Alice - year = -\frac{3}{4} Bob - light - year + \frac{5}{4} Bob - year$$

(this is because of Bob's velocity $\frac{3}{5}c$; different velocities give different transformations, of course. A derivation of the Lorentz transformations from Einstein's postulates of relativity is not too difficult, but is beyond the scope of this course). In other words, we have

$$[I_V]^{(Bob-light-year, Bob-year)}_{(Alice-light-year, Alice-year)} = \begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix}.$$

- Some examples. Suppose Alice emits a flash of light (event X), waits for one year without moving, and emits another flash of light (event Y). Let v denote the vector from X to Y. Then from Alice's point of view, v consists of one year and 0 light-years (because she didn't move between events):

$$[v]^{(Alice-light-year, Alice-year)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and thus

$$[v]^{(Bob-light-year, Bob-year)} = \begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 5/4 \\ -3/4 \end{pmatrix};$$

in other words, Bob perceives event Y occurring 5/4 years afterward and 3/4 light-years to the left of event X. This is consistent with the assumption that Bob was moving to the right at $\frac{3}{5}c$ with respect to Alice, so that from Bob's point of view Alice is receding to the left at $-\frac{3}{5}c$. This also illustrates the phenomenon of *time dilation* - what appears to be a single year from Alice's point of view becomes 5/4 years when measured by Bob.

- Another example. Suppose a beam of light was emitted by some source (event A) in a left-ward direction and absorbed by some receiver (event B) some time later. Suppose that Alice perceives event B as occurring one year after, and one light-year to the left of, event A; this is of course consistent with light traveling at $1c$. Thus if w is the vector from A to B, then

$$[v]^{(Alice-light-year, Alice-year)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and thus

$$[v]^{(Bob-light-year, Bob-year)} = \begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

Thus Bob views event B as occurring *two* years after and *two* light-years to the left of event A. Thus Bob still measures the speed of light as $1c$ (indeed, one of the postulates of relativity is that the speed of light is always a constant c to all inertial observers), but the light is "stretched out" over two years instead of one, resulting in Bob seeing the light at half the frequency that Alice would. This is the famous *Doppler red shift* effect in relativity (receding light has lower frequency and is red-shifted; approaching light has higher frequency and is blue-shifted).

- It may not seem like it, but this situation is symmetric with respect to Alice and Bob. We have

$$[I_V]^{(Bob-light-year, Bob-year)}_{(Alice-light-year, Alice-year)} = ([I_V]^{(Bob-light-year, Bob-year)}_{(Alice-light-year, Alice-year)})^{-1}$$

$$= \begin{pmatrix} 5/4 & -3/4 \\ -3/4 & 5/4 \end{pmatrix}^{-1} = \begin{pmatrix} 5/4 & 3/4 \\ 3/4 & 5/4 \end{pmatrix},$$

as one can verify by multiplying the above two matrices together (we will discuss matrix inversion in more detail next week). In other words,

$$Bob - light - year = \frac{5}{4} Alice - light - year + \frac{3}{4} Alice - year$$

$$Bob - year = \frac{3}{4} Alice - light - year + \frac{5}{4} Alice - year.$$

Thus for instance, just as Alice's years are time-dilated when measured by Bob, Bob's years are time-dilated when measured by Bob: if Bob emits light (event X'), waits for one year without moving (though of course still drifting at $\frac{3}{5}c$ as measured by Alice), and emits more light (event Y'), then Alice will perceive event Y' as occurring 5/4 years after and 3/4 light-years to the right of event X'; this is consistent with Bob travelling at $\frac{3}{5}c$, but Bob's year has been time dilated to $\frac{5}{4}$ years. (Why is it not contradictory for Alice's years to be time dilated when measured by Bob, and for Bob's years to be time dilated when measured by Alice? This is similar to the (CO_2, CO) versus (C, O) example: one molecule CO_2 contains two atoms of oxygen (plus some carbon), while an atom of oxygen consists of one molecule of CO_2 (minus some CO), and this is not contradictory. Vectors behave slightly different from scalars sometimes).

* * * * *

Co-ordinate change and matrices

- Let $T : V \to V$ be a linear transformation from a vector space V to itself. (Such transformations are sometimes called *automorphisms*, because they map onto themselves). Given any basis β of V, we can form a matrix $[T]_\beta^\beta$ representing T in the basis β. Of course, if we change the basis, from β to a different basis, say β', then the matrix changes also, to $[T]_{\beta'}^{\beta'}$. However, the two are related.

- **Lemma 1.** Let V be a vector space with two bases β' and β, and let $Q := [I_V]_\beta^{\beta'}$ be the change of co-ordinates matrix from β to β'. Let

$T : V \to V$ be a linear transformation. Then $[T]^\beta_\beta$ and $[T]^{\beta'}_{\beta'}$ are related by the formula

$$[T]^{\beta'}_{\beta'} = Q[T]^\beta_\beta Q^{-1}.$$

- **Proof** We begin with the obvious identity

$$T = I_V T I_V$$

and take bases (using Theorem 1 from last week's notes) to obtain

$$[T]^{\beta'}_{\beta'} = [I_V]^{\beta'}_\beta [T]^\beta_\beta [I_V]^\beta_{\beta'}.$$

Substituting $[I_V]^{\beta'}_\beta = Q$ and $[I_V]^\beta_{\beta'} = ([I_V]^{\beta'}_\beta)^{-1} = Q^{-1}$ we obtain the Lemma.

- **Example** Let's take a very simple portfolio model, where the portfolio consists of one type of stock (let's say GM stock), and one type of cash (let's say US dollars, invested in a money market fund). Thus a portfolio lives in a two-dimensional space V, with a basis $\beta := (Stock, Dollar)$. Let's say that over the course of a year, a unit of GM stock issues a dividend of two dollars, while a dollar invested in the money market fund would earn 2 percent, so that 1 dollar becomes 1.02 dollars. We can then define the linear transformation $T : V \to V$, which denotes how much a portfolio will appreciate within one year. (If one wants to do other operations on the portfolio, such as buy and sell stock, etc., this would require other linear transformations; but in this example we will just analyze plain old portfolio appreciation). Since

$$T(1 \times Stock) = 1 \times Stock + 2 \times Dollar$$

$$T(1 \times Dollar) = 1.02 \times Dollar$$

we thus see that

$$[T]^\beta_\beta = \begin{pmatrix} 1 & 0 \\ 2 & 1.02 \end{pmatrix}.$$

Now let's measure T in a different basis. Suppose that GM's stock has split, so that each old unit of Stock becomes two units of Newstock (so $Newstock = 0.5Stock$). Also, suppose for some reason (decimalization?) we wish to measure money in cents instead of dollars. So we now have a new basis $\beta' = (Newstock, Cent)$. Then we have

$$Newstock = 0.5Stock + 0Dollar; \quad Cent = 0Stock + 0.01Dollar,$$

so

$$Q = [I_V]_\beta^{\beta'} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.01 \end{pmatrix}$$

while similar reasoning gives

$$Q^{-1} = [I_V]_{\beta'}^{\beta} = \begin{pmatrix} 2 & 0 \\ 0 & 100 \end{pmatrix}.$$

Thus

$$[T]_{\beta'}^{\beta'} = Q[T]_\beta^\beta Q^{-1} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.01 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1.02 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 100 \end{pmatrix}$$

which simplifies to

$$[T]_{\beta'}^{\beta'} = \begin{pmatrix} 1 & 0 \\ 100 & 1.02 \end{pmatrix}.$$

Thus

$$T(Newstock) = 1 \times Newstock + 100 \times Cent$$

$$T(Cent) = 0 \times Newstock + 1.02 \times Cent.$$

This can of course be deduced directly from our hypotheses; it is instructive to do so and to compare that with the matrix computation.

- **Definition** Two $n \times n$ matrices A, B are said to be *similar* if one has $B = QAQ^{-1}$ for some invertible $n \times n$ matrix Q.

- Thus the two matrices $[T]_\beta^\beta$ and $[T]_{\beta'}^{\beta'}$ are similar. Similarity is an important notion in linear algebra and we will return to this property later.

* * * * *

Common sources of confusion

- This course is very much about concepts, and on thinking clearly and precisely about these concepts. It is particularly important not to confuse two concepts which are similar but not identical, otherwise this can lead to one getting hopelessly lost when trying to work one's way through a problem. (This is true not just of mathematics, but also of other languages, such as English. If one confuses similar words - e.g. the adjective "happy" with the noun "happiness" - then one still might be able to read and write simple sentences and still be able to communicate (although you may sound unprofessional while doing so), but complex sentences will become very difficult to comprehend). Here I will list some examples of similar concepts that should be distinguished. These points may appear pedantic, but an inability to separate these concepts is usually a sign of some more fundamental problem in comprehending the material, and should be addressed as quickly as possible.

- **"Vector" versus "Vector space"**. A vector space consists of vectors, but is not actually a vector itself. Thus questions like "What is the dimension of $(1, 2, 3, 4)$?" are meaningless; $(1, 2, 3, 4)$ is a vector, not a vector space, and only vector spaces have a concept of dimension. A question such as "What is the dimension of (x_1, x_1, x_1)?" or "What is the dimension of $x_1 + x_2 + x_3 = 0$?" is also meaningless for the same reason, although "What is the dimension of $\{(x_1, x_1, x_1) : x_1 \in \mathbf{R}\}$?" or "What is the dimension of $\{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 0\}$?" are not.

- In a similar spirit, the zero vector 0 is distinct from the zero vector space $\{0\}$, and is in turn distinct from the zero linear transformation T_0. (And then there is also the zero scalar 0).

- **A set S of vectors, versus the span** $\mathrm{span}(S)$ **of that set**. This is a similar problem. A statement such as "What is the dimension of $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$?" is meaningless, because the set $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is not a vector space. It is true that this set *spans* a vector space - $\mathbf{R}^3$ in this example - but that is a different object. Similarly, it is not correct to say that the sets $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ and $\{(1, 1, 0), (1, 0, 1), (0, 0, 1)\}$ are equal - they may span the same space, but they are certainly not the same set.

- For a similar reason, it is not true that if you add $(0,0,1)$ to $\mathbf{R}^2$, that you "get" $\mathbf{R}^3$. First of all, $\mathbf{R}^2$ is not even contained in $\mathbf{R}^3$, although the xy-plane $V := \{(x,y,0) : x,y, \in \mathbf{R}\}$, which is isomorphic to $\mathbf{R}^2$, *is* contained in $\mathbf{R}^3$. But also, if you take the union of V with $\{(0,0,1)\}$, one just gets the set $V \cup \{(0,0,1)\}$, which is a plane with an additional point added to it. The correct statement is that V and $(0,0,1)$ together *span* or *generate* $\mathbf{R}^3$.

- **"Finite" versus "finite-dimensional".** A set is *finite* if it has finitely many elements. A vector space is *finite-dimensional* if it has a finite basis. The two notions are distinct. For instance, $\mathbf{R}^2$ is infinite (there are infinitely many points in the plane), but is finite-dimensional because it has a finite basis $\{(1,0),(0,1)\}$. The zero vector space $\{0\}$ contains just one element, but is zero-dimensional. The set $\{(1,0,0),(0,1,0),(0,0,1)\}$ is finite, but it does not have a dimension because it is not a vector space.

- **"Subset" versus "subspace"** Let V be a vector space. A *subset* U of V is any collection of vectors in V. If this subset is also closed under addition and scalar multiplication, we call U a *subspace*. Thus every subspace is a subset, but not conversely. For instance, $\{(1,0,0),(0,1,0),(0,0,1)\}$ is a subset of $\mathbf{R}^3$, but is not a subspace (so it does not have a dimension, for instance).

- **"Nullity" versus "null space"** The *null space* $N(T)$ of a linear transformation $T : V \to W$ is defined as the space $N(T) = \{v \in V : Tv = 0\}$; it is a subspace of V. The *nullity* nullity(T) is the dimension of $N(T)$; it is a *number*. So statements such as "the null space of T is three" or "the nullity of T is $\mathbf{R}^3$" are meaningless. Similarly one should distinguish "range" and "rank"

- **"Range $R(T)$" versus "final space"** This is an unfortunate notation problem. If one has a transformation $T : V \to W$, which maps elements in V to elements in W, V is sometimes referred to as the *domain* and W is referred to as the *range*. However, this notation is in conflict with the range $R(T)$, defined by $R(T) = \{Tv : v \in V\}$. If the transformation T is *onto*, then $R(T)$ and W are the same, but otherwise they are not. To avoid this confusion, I will try to refer to V as the *initial space* of

T, and W as the *final space*. Thus in the map $T : \mathbf{R}^2 \to \mathbf{R}^3$ defined by $T(x_1, x_2) := (x_1, x_2, 0)$, the final space is $\mathbf{R}^3$, but the range is only the xy-plane, which is a subspace of $\mathbf{R}^3$.

- To re-iterate, in order to be able to say that a transformation T maps V to W, it is not required that every element of W is actually covered by T; if that is the case, we say that T maps V *onto* W, and not just *to* W. It is also not required that different elements of V have to map to different elements of W; that is true for *one-to-one* transformations, but not for general transformations. So in particular, if you are given that $T : V \to W$ is a linear transformation, you cannot necessarily assume that T is one-to-one or onto unless that is explicitly indicated in the question.

- **"Vector" versus "Co-ordinate vector"** A vector v in a vector space V is not the same thing as its co-ordinate vector $[v]^\beta$ with respect to some basis β of V. For one thing, v needn't be a traditional vector (a row or column of numbers) at all; it may be a polynomial, or a matrix, or a function - these are all valid examples of vectors in a vector space. Secondly, the co-ordinate vector $[v]^\beta$ depends on the basis β as well as on the vector v itself; change β and you change the co-ordinate vector $[v]^\beta$, even if v itself is kept fixed. For instance, if $V = P_2(\mathbf{R})$ and $v = 3x^2+4x+5$, then v is a vector (even though it's also a polynomial). If $\beta := (1, x, x^2)$, then $[v]^\beta = \begin{pmatrix} 5 \\ 4 \\ 3 \end{pmatrix}$, but this is not the same object as $3x^2 + 4x + 5$; one is a column vector and the other is a polynomial. Calling them "the same" will lead to trouble if one then switches to another basis, such as $\beta' := (x^2, x, 1)$, since the co-ordinate vector is now $[v]^{\beta'} = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$, which is clearly different from $\begin{pmatrix} 5 \\ 4 \\ 3 \end{pmatrix}$.

- There is of course a similar distinction between a linear transformation T, and the matrix representation $[T]^\gamma_\beta$ of that transformation T.

- **"Closed under addition" versus "preserves addition"**. A set U is *closed under addition* if, whenever x and y are elements of U, the sum

$x + y$ is also in U. A transformation $T : V \to W$ *preserves addition* if, whenever x and y are elements in V, that $T(x + y) = Tx + Ty$. The two concepts refer to two different things - one has to do with a *set*, and the other with a *transformation* - but surprisingly it is still possible to confuse the two. For instance, in one of the homework questions, one has to show that the *set* $T(U)$ is closed under addition, and someone thought this meant that one had to show that $T(x+y) = Tx+Ty$ for all x in U; probably what happened was that the presence of the letter T in the set $T(U)$ caused one to automatically think of the preserving addition property rather the closed-under-addition property. Of course, there is a similar distinction between "closed under scalar multiplication" and "preserves scalar multiplication".

- **A vector v, versus the image Tv of that vector**. A linear transformation $T : V \to W$ can take any vector v in V, and transform it to a new vector Tv in W. However, it is dangerous to say things like v "becomes" Tv, or v "is now" Tv. If one is not careful, one may soon write that v "is" Tv, or think that every property that v has, automatically also holds for Tv. The correct thing to say is that Tv is the *image* of v under T; this image may preserve some properties of the original vector v, but it may distort or destroy others. In general, Tv is not equal to v (except in special cases, such as when T is the identity transformation).

- **Hypothesis versus Conclusion**. This is not a confusion specific to linear algebra, but nevertheless is an important distinction to keep in mind when doing any sort of "proof" type question. You should always know what you are currently *assuming* (the hypotheses), and what you are trying to prove (the conclusion). For instance, if you are trying to prove

  Show that if $T : V \to W$ is linear, then
  $T(cx + y) = cTx + Ty$ for all $x, y \in V$ and all scalars c

  then your hypothesis is that $T : V \to W$ is a linear transformation (so that $T(x + y) = Tx + Ty$ and $T(cx) = cTx$ for all $x, y \in V$ and all scalars c, and your objective is to prove that for any vectors $x, y \in V$ and any scalar c, we have $T(cx + y) = cTx + Ty$.

- On the other hand, if you are trying to prove that

Show that if $T : V \to W$ is such that $T(cx + y) = cTx + Ty$ for all $x, y \in V$ and all scalars c, then T is linear.

then your hypotheses are that T maps V to W, and that $T(cx + y) = cTx + Ty$ for all $x, y \in V$ and all scalars c, and your objective is to prove that T is linear, i.e. that $T(x + y) = Tx + Ty$ for all vectors $x, y \in V$, and that $T(cx) = cTx$ for all vectors $x \in V$ and scalars c. This is a completely different problem from the previous one. (Part 2. of Question 7 of 2.2 requires you to prove *both* of these implications, because it is an "if and only if" question).

- In your proofs, it may be a good idea to identify which of the statements that you write down are things that you *know* from the hypotheses, and which ones are those that you *want*. Little phrases like "We are given that", "It follows that", or "We need to show that" in the proof are very helpful, and will help convince the grader that you actually know what you are doing! (provided that those phrases are being used correctly, of course).

- **"For all" versus "For some"**. Sometimes it is really important to read the "fine print" of a question - it is all to easy to jump to the equations without reading all the English words which surround those equations. For instance, the statements

  "Show that $Tv = 0$ for some non-zero $v \in V$"

  is completely different from

  "Show that $Tv = 0$ for all non-zero $v \in V$"

  In the first case, one only needs to exhibit a *single* non-zero vector v in V for which $Tv = 0$; this is a statement which could be proven by an example. But in the second case, no amount of examples will help; one has to show that $Tv = 0$ for *every* non-zero vector v. In the second case, probably what one would have to do is start with the hypotheses that $v \in V$ and that $v \neq 0$, and somehow work one's way to proving that $Tv = 0$.

- Because of this, it is very important that you read the question carefully before answering. If you don't understand exactly what the question is asking, you are unlikely to write anything for the question that the grader will find meaningful.

Math 115A - Week 6
Textbook sections: 3.1-5.1
Topics covered:

- Review: Row and column operations on matrices
- Review: Rank of a matrix
- Review: Inverting a matrix via Gaussian elimination
- Review: Determinants

* * * * *

Review: Row and column operations on matrices

- We now quickly review some material from Math 33A which we will need later in this course. The first concept we will need is that of an *elementary row operation.*

- An *elementary row operation* takes an $m \times n$ matrix as input and returns a different $m \times n$ matrix as output. (In other words, each elementary row operation is a map from $M_{m\times n}(\mathbf{R})$ to $M_{m\times n}(\mathbf{R})$. There are three types of elementary row operations:

- **Type 1 (row interchange)**. This type of row operation interchanges row i with row j for some $i, j \in \{1, 2, \ldots, m\}$. For instance, the operation of interchanging rows 2 and 4 in a 4×3 matrix would change

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$$

to

$$\begin{pmatrix} 1 & 2 & 3 \\ 10 & 11 & 12 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix}.$$

Observe that the final matrix can be obtained from the initial matrix by multiplying on the left by the 4×4 matrix

$$E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

which is the identity matrix with rows 2 and 4 switched. (Why?) Thus for instance

$$\begin{pmatrix} 1 & 2 & 3 \\ 10 & 11 & 12 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$$

and more generally, if A is a 4×3 matrix, then the interchange of rows 2 and 4 replaces A with EA. We refer to E as an 4×4 *elementary matrix of type 1.*

- Also observe that row interchange is its own inverse; if one replaces A with EA, and then replaces EA with EEA (i.e. we interchange rows 2 and 4 twice), we get back to where we started, because $EE = I_4$.

- **Type 2 (row multiplication)** This type of elementary row operation takes a row i and multiplies it by a *non-zero* scalar c. For instance, the elementary row operation that multiplies row 2 by 10 would map

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$$

to

$$\begin{pmatrix} 1 & 2 & 3 \\ 40 & 50 & 60 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}.$$

This operation is the same as multiplying on the left by the matrix

$$E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which is what one gets by starting with the identity matrix and multiplying row 2 by 10. (Why?) We call E an example of a 4×4 *elementary matrix of type 2.*

- Row multiplication is invertible; the operation of multiplying a row i by a non-zero scalar c is inverted by multiplying a row i by the non-zero scalar $1/c$. In the above example, the inverse operation is given by

$$E^{-1} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/10 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

i.e. to invert the operation of multiplying row 2 by 10, we then multiply row 2 by 1/10.

- **Type 3 (row addition)** For this row operation we need two rows i, j, and a scalar c. The row operation adds c multiples of row i to row j. For instance, if one were to add 10 multiples of row 2 to row 3, then

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$$

would become

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 47 & 58 & 69 \\ 10 & 11 & 12 \end{pmatrix}.$$

Equivalently, this row operation amounts to multiplying the original matrix on the left by the matrix

$$E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 10 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which is what one gets by starting with the identity matrix and adding 10 copies of row 2 to row 3. (Why?) We call E an example of a 4×4 *elementary matrix of type* 3. It has an inverse

$$E^{-1} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -10 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

thus to invert the operation of adding 10 copies of row 2 to row 3, we subtract 10 copies of row 2 to row 3 instead.

- Thus to summarize: there are special matrices, known as *elementary row matrices*, and an elementary row operation amounts to multiplying a given matrix on the left by one of these elementary row matrices. Each of the elementary row matrices is invertible, and the inverse of an elementary matrix is another elementary matrix. (In the above discussion, we only verified this for specific examples of elementary matrices, but it is easy to see that the same is true for general elementary matrices. See the textbook).

- There are also *elementary column operations*, which are very similar to elementary row operations, but arise from multiplying a matrix on the *right* by an elementary matrix, instead of the left. For instance, if one multiplies a matrix A with 4 columns on the right by the elementary matrix

$$E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

then this amounts to switching column 2 and column 4 of A (why?). This is a type 1 (column interchange) elementary column move. Similarly, if one multiplies A on the right by

$$E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

then this amounts to multiplying column 2 of A by 10 (why?). This is a type 2 (column multiplication) elementary column move. Finally, if one multiplies A on the right by

$$E := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 10 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

then this amounts to adding 10 copies of column 3 to column 2 (why?). This is a type 3 (column addition) elementary column move.

- Elementary row (or column) operations have several uses. One important use is to simplify systems of linear equations of the form $Ax = b$, where A is some matrix and x, b are vectors. If E is an elementary matrix, then the equation $Ax = b$ is equivalent to $EAx = Eb$ (why are these two equations equivalent? Hint: E is invertible). Thus one can simplify the equation $Ax = b$ by performing the same row operation to both A and b simultaneously (one can concatenate A and b into a single (artificial) matrix in order to do this). Eventually one can use row operations to reduce A to row-echelon form (in which each row is either zero, or begins with a 1, and below each such 1 there are only zeroes), at which point it becomes straightforward to solve for $Ax = b$ (or to determine that there are no solutions). We will not review this procedure here, because it will not be necessary for this course; see however the textbook (or your Math 33A notes) for more information. However, we will remark that every matrix can be row-reduced into row-echelon form (though there is usually more than one way to do so).

- Another purpose of elementary row or column operations is to determine the *rank* of a matrix, which is a more precise measurement of its invertibility. This will be the purpose of the next section.

* * * * *

Rank of a matrix

- Recall that the *rank* of a linear transformation $T : V \to W$ is the dimension of its range $R(T)$. Rank has a number of uses, for instance it can be used to tell whether a linear transformation is invertible:

- **Lemma.** Let V and W be n-dimensional spaces, and let $\dim(V) = \dim(W) = n$. Let $T : V \to W$ be a linear transformation. Then T is invertible if and only if $\text{rank}(T) = n$ (i.e. T has the maximum rank possible).

- **Proof.** If $\text{rank}(T) = n$, then $R(T)$ has the same dimension n as W. But $R(T)$ is a subspace of W, so this forces $R(T)$ to actually equal W (see Theorem 2 of Week 2 notes). Thus T is onto. Also, from the dimension theorem we see that $\text{nullity}(T) = 0$, and so T is one-to-one. Thus T is invertible.

- Conversely, if T is invertible, then it is one-to-one, hence $\text{nullity}(T) = 0$, and hence by the dimension theorem $\text{rank}(T) = n$. $\square$

- One interesting thing about rank is that if you multiply a linear transformation on the left or right by an invertible transformation, then the rank doesn't change:

- **Lemma 1.** Let $T : V \to W$ be a linear transformation from one finite-dimensional space to another, let $S : U \to V$ be an invertible transformation, and let $Q : W \to Z$ be another invertible transformation. Then

$$\text{rank}(T) = \text{rank}(QT) = \text{rank}(TS) = \text{rank}(QTS).$$

- **Proof.** First let us show that $\text{rank}(T) = \text{rank}(TS)$. To show this, we first compute the ranges of $T : V \to W$ and $TS : U \to W$. By definition of range, $R(T) = T(V)$, the image of V under T. Similarly,

$R(TS) = TS(U)$. But since $S : U \to V$ is invertible, it is onto, and so $S(U) = V$. Thus

$$R(TS) = TS(U) = T(S(U)) = T(V) = R(T)$$

and so

$$\mathrm{rank}(TS) = \mathrm{rank}(T).$$

A similar argument gives that $\mathrm{rank}(QTS) = \mathrm{rank}(QT)$ (just replace T by RT in the above. To finish the argument we need to show that $\mathrm{rank}(QT) = \mathrm{rank}(T)$. We compute ranges again:

$$R(QT) = QT(V) = Q(T(V)) = Q(R(T)),$$

so that

$$\mathrm{rank}(QT) = \dim(Q(R(T))).$$

But since Q is invertible, we have

$$\dim(Q(R(T))) = \dim(R(T)) = \mathrm{rank}(T)$$

(see Q3 of the midterm!). Thus $\mathrm{rank}(QT) = \mathrm{rank}(T)$ as desired. $\square$.

- Now to compute the rank of an arbitrary linear transformation can get messy. The best way to do this is to convert the linear transform to a matrix, and compute the rank of that.

- **Definition** If A is a matrix, then the *rank* of A, denoted $\mathrm{rank}(A)$, is defined by $\mathrm{rank}(A) = \mathrm{rank}(L_A)$.

- **Example** Consider the matrix

$$A := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then L_A is the transformation

$$L_A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \end{pmatrix}.$$

The range of this operator is thus three-dimensional (why?) and so the rank of A is 3.

- Let A be an $m \times n$ matrix, so that L_A maps $\mathbf{R}^n$ to $\mathbf{R}^m$. Let $(e_1, e_2, \ldots, e_n)$ be the standard basis for $\mathbf{R}^n$. Since $e_1, \ldots, e_n$ span $\mathbf{R}^n$, we see that $L_A(e_1), L_A(e_2), \ldots, L_A(e_n)$ span $R(L_A)$ (see Theorem 3 of Week 3 notes). Thus the rank of A is the dimension of the space spanned by $L_A(e_1), L_A(e_2), \ldots, L_A(e_n)$. But $L_A(e_1)$ is simply the first column of A (why?), $L_A(e_2)$ is the second column of A, etc. Thus we have shown

- **Lemma 2.** The rank of a matrix A is equal to the dimension of the space spanned by its columns.

- **Example** If A is the matrix used in the previous example, then the rank of A is the dimension of the span of the columns $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$; this span is easily seen to be three-dimensional.

- As one corollary of this Lemma, if only k of the columns of a matrix are non-zero, then the rank of the matrix can be at most k (though it could be smaller than k; can you think of an example?).

- This Lemma does not necessarily make computing the rank easy, because finding the dimension of the space spanned by the columns could be difficult. However, one can use elementary row or column operations to simplify things. From Lemma 1 we see that if E is an elementary matrix and A is a matrix, then EA or AE has the same rank as A (provided that the matrix multiplication makes sense, of course). Thus elementary row or column operations do not change the rank. Thus, we can use these operations to simplify the matrix into row-echelon form.

- **Lemma 3.** Let A be a matrix in row-echelon form (thus every row is either zero, or begins with a 1, and each of those 1s has nothing but 0s below it). Then the rank of A is equal to the number of non-zero rows.

- **Proof.** Let's say that A has k non-zero rows, so that we have to show that A has rank k. Each column of A can only has the top k entries non-zero; all entries below that must be zero. Thus the span of the columns of A is contained in the k-dimensional space

$$V = \{ \begin{pmatrix} x_1 \\ \dots \\ x_k \\ 0 \\ \dots \\ 0 \end{pmatrix} : x_1, \dots, x_k \in \mathbf{R}\},$$

  and so the rank is at most k.

- Now we have to show that the rank is at least k. To do this it will suffice to show that every vector in V is in the span of the columns of A, since this will mean that the span of the columns of A is exactly the k-dimensional space V. So, let us pick a vector $v := \begin{pmatrix} x_1 \\ \dots \\ x_k \\ 0 \\ \dots \\ 0 \end{pmatrix}$ in V. Since A is in row-echelon form, the k^{th} row of A must contain a 1 somewhere, which means that there is a column whose k^{th} entry is 1 (with all entries below that equal to 0). If we subtract x_k multiples of this column from v, then we get a new vector whose k^{th} entry (and all the ones below it) are zero.

- Now we look at the $(k-1)^{th}$ row. Again, since we are in row-echelon form, there is a 1 somewhere, with 0s below it; this implies that there is a column whose $(k-1)^{th}$ entry is 1 (with all entries below that equal to 0). Thus we can subtract a multiple of this vector to get a new vector whose $(k-1)^{th}$ entry (and all the ones below it) are zero.

- Continuing in this fashion we can subtract multiples of various columns of A from v until we manage to zero out all the entries. In other words, we have expressed v as a linear combination of columns in A, and hence

v is in the span of the columns. Thus the span of the columns is exactly V, and we are done. $\square$.

- Thus we now have a procedure to compute the rank of a matrix: we row reduce (or column reduce) until we reach row-echelon form, and then just count the number of non-zero rows. (Actually, one doesn't necessarily have to reduce all the way to row-echelon form; it may be that the rank becomes obvious some time before then, because the span of the columns can be determined by inspection).

- If one only uses elementary row operations, then usually one cannot hope to make the matrix much simpler than row-echelon form. But if one is allowed to also use elementary column operations, then one can get the matrix into a particularly simple form:

- **Theorem 4.** Let A be an $m \times n$ matrix with rank r. Then one can use elementary row and column matrices to place A in the form

$$\begin{pmatrix} I_r & 0_{r \times n-r} \\ 0_{m-r \times r} & 0_{m-r \times n-r} \end{pmatrix},$$

where I_r is the $r \times r$ identity matrix, and $0_{m \times n}$ is the $m \times n$ zero matrix. (Thus, we have reduced the matrix to nothing but a string of r ones down the diagonal, with everything else being zero.

- **Proof** To begin with, we know that we can use elementary row operations to place A in row-echelon form. Thus the first r rows begin with a 1, with 0s below the 1, while the remaining $m - r$ rows are entirely zero.

- Now consider the first row of this reduced matrix; let's suppose that it is not identically zero. After some zeroes, it has a 1, and then some other entries which may or may not be zero. But by subtracting multiples of the column with 1 in it from the columns with other non-zero entries (i.e. type 3 column operations), we can make all the other entries in this row equal to zero. Note that these elementary column operations only affect the top row, leaving the other rows unchanged, because the column with 1 in it has 0s everywhere else.

- A similar argument then allows one to take the second row of the reduced matrix and make all the entries (apart from the leading 1) equal to 0. And so on and so forth. At the end of this procedure, we get that the first r rows each contain one 1 and everything else being zero. Furthermore, these 1s have 0s both above and below, so they lie in different columns. Thus by switching columns appropriately (using type 1 column operations) we can get into the form required for the Theorem. $\square$

- Let A be an $m \times n$ matrix with rank r. Every elementary row operation corresponds to multiplying A on the left by an $m \times m$ elementary matrix, while every elementary column operation corresponds to multiplying A on the right by an $n \times n$ elementary matrix. Thus by Theorem 4, we have

$$E_1 E_2 \ldots E_a A F_1 F_2 \ldots F_b = \begin{pmatrix} I_r & 0_{r \times n-r} \\ 0_{m-r \times r} & 0_{m-r \times n-r} \end{pmatrix},$$

where $E_1, \ldots, E_a$ are elementary $m \times m$ matrices, and $F_1 F_2 \ldots F_b$ are elementary $n \times n$ matrices. All the elementary matrices are invertible. After some matrix algebra, this becomes

$$A = E_a^{-1} \ldots E_2^{-1} E_1^{-1} \begin{pmatrix} I_r & 0_{r \times n-r} \\ 0_{m-r \times r} & 0_{m-r \times n-r} \end{pmatrix} F_b^{-1} \ldots F_2^{-1} F_1^{-1}$$

(why did the order of the matrices get reversed?). We have thus proven

- **Proposition 5.** Let A be an $m \times n$ matrix with rank r. Then we have an $m \times m$ matrix B which is a product of elementary matrices and an $n \times n$ matrix C, also a product of elementary matrices such that

$$A = B \begin{pmatrix} I_r & 0_{r \times n-r} \\ 0_{m-r \times r} & 0_{m-r \times n-r} \end{pmatrix} C.$$

- Note (from Q6 of Assignment 4) that the product of invertible matrices is always invertible. Thus the matrices B and C above are invertible.

- Proposition 5 is an example of a *factorization theorem*, which takes a general matrix and factors it into simpler pieces. There are many other examples of factorization theorems which you will encounter later in the 115 sequence, and they have many applications.

- Some more properties of rank. We know from Lemma 1 that rank of a *linear transformation* is unchanged by multiplying on the left or right by invertible transformations. Given the close relationship between linear transformations and matrices, it is unsurprising that the same thing is true for matrices:

- **Lemma 6.** Let A be an $m \times n$ matrix, B be an $m \times m$ invertible matrix, and C be an $n \times n$ invertible matrix. Then

$$\operatorname{rank}(A) = \operatorname{rank}(BA) = \operatorname{rank}(AC) = \operatorname{rank}(BAC).$$

- **Proof.** Since B is invertible, so is L_B (see Theorem 10 from Week 4 notes). Similarly L_C is invertible. From Lemma 1 we have

$$\operatorname{rank}(L_A) = \operatorname{rank}(L_B L_A) = \operatorname{rank}(L_A L_C) = \operatorname{rank}(L_B L_A L_C).$$

  Since $L_B L_A = L_{BA}$, etc. we thus have

$$\operatorname{rank}(L_A) = \operatorname{rank}(L_{BA}) = \operatorname{rank}(L_{AC}) = \operatorname{rank}(L_{BAC}).$$

  The claim then follows from the definition of rank for matrices. $\square$

- Note that Lemma 6 is consistent with Proposition 5, since the matrix $\begin{pmatrix} I_r & 0_{r \times n-r} \\ 0_{m-r \times r} & 0_{m-r \times n-r} \end{pmatrix}$ has rank r.

- One important consequence of the above theory concerns the rank of a *transpose* A^t of a matrix A. Recall that the transpose of an $m \times n$ matrix is the $n \times m$ matrix obtained by reflecting around the diagonal, so for instance

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{pmatrix}^t = \begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{pmatrix}.$$

  Thus transposes swap rows and columns. From the definition of matrix multiplication it is easy to verify the identity $(AB)^t = B^t A^t$ (why?). In particular, if A is invertible, then $I = I^t = (AA^{-1})^t = A^t (A^{-1})^t$, which implies that A^t is also invertible.

- **Lemma 7** Let A be an $m \times n$ matrix with rank r. Then A^t has the same rank as A.
- From Proposition 5 we have

$$A = B \begin{pmatrix} I_r & 0_{r \times n-r} \\ 0_{m-r \times r} & 0_{m-r \times n-r} \end{pmatrix} C.$$

Taking transposes of both sides we obtain

$$A^t = C^t \begin{pmatrix} I_r & 0_{r \times m-r} \\ 0_{n-r \times r} & 0_{n-r \times m-r} \end{pmatrix} B^t.$$

The inner matrix on the right-hand side has rank r. Since B and C are invertible, so are B^t and C^t, and so by Lemma 6 A^t has rank r, and we are done. $\square$

- From Lemma 7 and Lemma 2 we thus have
- **Corollary 8.** The rank of a matrix is equal to the dimension of the space spanned by its *rows*.
- As one corollary of this Lemma, if only k of the rows of a matrix are non-zero, then the rank of the matrix can be at most k.
- Finally, we give a way to compute the rank of any linear transformation from one finite-dimensional space to another.
- **Lemma 9.** Let $T : V \to W$ be a linear transformation, and let β and γ be finite bases for V and W respectively. Then $\text{rank}(T) = \text{rank}([T]_\beta^\gamma)$.
- **Proof.** Suppose V is n-dimensional and W is m-dimensional. Then the co-ordinate map $\phi_\beta : V \to \mathbf{R}^n$ defined by $\phi_\beta(v) := [v]^\beta$ is an isomorphism, as is the co-ordinate map $\phi_\gamma : W \to \mathbf{R}^m$ defined by $\phi_\gamma(w) := [w]^\gamma$. Meanwhile, the map $L_{[T]_\beta^\gamma}$ is a linear transformation from $\mathbf{R}^n$ to $\mathbf{R}^m$. The identity

$$[Tv]^\gamma = [T]_\beta^\gamma [v]^\beta$$

can thus be rewritten as

$$\phi_\gamma(Tv) = L_{[T]_\beta^\gamma} \phi_\beta(v)$$

and thus

$$\phi_\gamma T = L_{[T]_\beta^\gamma} \phi_\beta$$

and hence (since ϕ_β is invertible)

$$T = \phi_\gamma^{-1} L_{[T]_\beta^\gamma} \phi_\beta.$$

Taking rank of both sides and using Lemma 6, we obtain

$$\mathrm{rank}(T) = \mathrm{rank}(L_{[T]_\beta^\gamma}) = \mathrm{rank}([T]_\beta^\gamma)$$

as desired. $\square$

- **Example** Let $T : P_3(\mathbf{R}) \to P_3(\mathbf{R})$ be the linear transformation

$$Tf := f - xf',$$

thus for instance

$$Tx^2 = x^2 - x(2x) = -x^2.$$

To find the rank of this operator, we let $\beta := (1, x, x^2, x^3)$ be the standard basis for $P_3(\mathbf{R})$. A simple calculation shows that

$$T1 = 1; \quad Tx = 0; \quad Tx^2 = -x^2; \quad Tx^3 = -2x^3,$$

so

$$[T]_\beta^\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

This matrix clearly has rank 3 (row operations can convert it to row-echelon form with three non-zero rows), so T has rank 3.

- The rank of a matrix measures, in some sense, how close to zero (or how "degenerate") a matrix is; the only matrix with rank 0 is the 0 matrix (why). The largest rank an $m \times n$ matrix can have is $\min(m, n)$ (why? See Lemma 2 and Corollary 8). For instance, a 3×5 matrix can have rank at most 3.

* * * * *

Inverting matrices via row operations

- Proposition 5 has the following consequence.

- **Lemma 10.** Let A be an $n \times n$ matrix. Then A is invertible if and only if it is the product of elementary $n \times n$ matrices.

- **Proof** First suppose that A is the product of elementary matrices. We already know that every elementary matrix is invertible; also from Q6 from the Week 4 homework, we know that the product of two invertible matrices is also invertible. Applying this fact repeatedly we see that A is also invertible.

- Now suppose that A is invertible, thus $L_A : \mathbf{R}^n \to \mathbf{R}^n$ is invertible, and in particular is onto. Thus $R(L_A) = \mathbf{R}^n$, and so $\mathrm{rank}(L_A) = n$, and so A itself must have rank n. By Proposition 5 we thus have

$$A = E_1 E_2 \ldots E_a I_n F_1 \ldots F_b$$

where $E_1, \ldots, E_a, F_1, \ldots, F_b$ are elementary $n \times n$ matrices. Since the identity matrix I_n cancels out, we are done. $\square$

- This gives us a way to use row operations to invert a matrix A. Suppose we manage to use a sequence of row operations E_1, $E_2, \ldots, E_a$ in turn a matrix A into the identity, thus

$$E_a \ldots E_2 E_1 A = I.$$

Then by multiplying both sides on the right by A^{-1} we get

$$E_a \ldots E_2 E_1 I = A^{-1}.$$

Thus, if we concatenate A and I together, and apply row operations on the concatenated matrix to turn the A component into I, then the I component will automatically turn to A^{-1}. This is a way of computing the inverse of A.

- **Example**. Suppose we want to invert the matrix

$$A := \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

We combine A and the identity I_2 into a single matrix:

$$(A|I_2) = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array} \right).$$

Then we row reduce to turn the left matrix into the identity. For instance, by subtracting three copies of row 1 from row 2 we obtain

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array} \right)$$

and then by adding row 2 to row 1 we obtain

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

Dividing the second row by -2 we obtain

$$\left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{array} \right).$$

This the inverse of A is

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ 3/2 & -1/2 \end{pmatrix},$$

since the elementary transformations which convert A to I_2, also convert I_2 to A^{-1}.

* * * * *

Determinants

- We now review a very useful characteristic of matrices - the *determinant* of a matrix. The determinant of a square ($n \times n$) matrix is a number which depends in a complicated way on the entries of that matrix. Despite the complicated definition, it has some very remarkable properties, especially with regard to matrix multiplication, and row and column operations. Unfortunately we will not be able to give the proofs of many of these remarkable properties here; the best way to understand determinants is by means of something called *exterior* *algebra*, which is beyond the scope of this course. Without the tools of exterior algebra (in particular, something called a *wedge product*), proving any of the fundamental properties of determinants becomes very messy. So we will settle just for describing the determinant and stating its basic properties.

- The determinant of a 1×1 matrix is just its entry:

$$\det(a) = a.$$

- The 2×2 determinant is given by the formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc.$$

- The $n \times n$ determinant is messier, and is defined in the following strange way. For any row i and column j, define the *minor* $\tilde{A}_{ij}$ of an $n \times n$ matrix A to be the $n-1 \times n-1$ matrix which is A with the i^{th} row and j^{th} column removed. For instance, if

$$A := \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

then

$$\tilde{A}_{11} = \begin{pmatrix} e & f \\ h & i \end{pmatrix}, \quad \tilde{A}_{12} = \begin{pmatrix} d & f \\ g & i \end{pmatrix}, \quad \tilde{A}_{13} = \begin{pmatrix} d & e \\ g & h \end{pmatrix},$$

etc.

This should not be confused with A_{ij}, which is the entry of A in the i^{th} row and j^{th} column. FOr instance, in the above example $A_{11} = a$ and $A_{12} = b$.

- We can now define the $n \times n$ determine recursively in terms of the $n-1$ determinant by what is called the *cofactor expansion.* To find the determinant of an $n \times n$ matrix A, we pick a row i (any row will do) and set

$$\det(A) := \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(\tilde{A}_{ij}).$$

For instance, we have

$$\det(\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a\det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - b\det\begin{pmatrix} d & f \\ g & i \end{pmatrix} + c\det\begin{pmatrix} d & e \\ g & h \end{pmatrix},$$

or

$$\det(\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = -d\det\begin{pmatrix} b & c \\ h & i \end{pmatrix} + e\det\begin{pmatrix} a & c \\ g & i \end{pmatrix} - f\det\begin{pmatrix} a & b \\ g & h \end{pmatrix},$$

or

$$\det(\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = g\det\begin{pmatrix} b & c \\ e & f \end{pmatrix} - h\det\begin{pmatrix} a & c \\ d & f \end{pmatrix} + i\det\begin{pmatrix} a & b \\ d & e \end{pmatrix}.$$

It seems that this definition depends on which row you use to perform the cofactor expansion, but the amazing thing is that it doesn't! For instance, in the above example, any three of the computations will lead to the same answer, namely

$$aei - ahf - bdi + bgf + cdh - cge.$$

We would like to explain why it doesn't matter which row (or column; see below) to perform cofactor expansion, but it would require one to develop some new material (on the signature of permutations) which would take us far afield, so we will regretfully skip the derivation.

- The quantities $\det(\tilde{A}^{ij})$ are sometimes known as *cofactors*. As one can imagine, this cofactor formula becomes extremely messy for large matrices (to compute the determinant of an $n \times n$ matrix, the above formula will eventually require us to add or subtract $n!$ terms together!); there are easier ways to compute using row and column operations which we will describe below.

- One can also perform cofactor expansion along a column j instead of a row i:

$$\det(A) := \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det(\tilde{A}_{ij}).$$

This ultimately has to do with a symmetry property $\det(A^t) = \det(A)$ for the determinant, although this symmetry is far from obvious given the definition.

- A special case of cofactor expansion: if A has the form

$$A = \begin{pmatrix} c & 0\ldots 0 \\ \vdots & B \end{pmatrix},$$

where the $0\ldots 0$ are a string of $n-1$ zeroes, the $\vdots$ represent a column vector of length $n-1$, and B is an $n-1 \times n-1$ matrix, then $\det(A) = c\det(B)$. In particular, from this and induction we see that the identity matrix always has determinant 1: $\det(I_n) = 1$. Also, we see that the determinant of a lower-diagonal matrix is just the product of the diagonal entries; for instance

$$\det(\begin{pmatrix} a & 0 & 0 \\ d & e & 0 \\ g & h & i \end{pmatrix} = aei.$$

Because of the symmetry property we also know that upper-diagonal matrices work the same way:

$$\det(\begin{pmatrix} a & b & c \\ 0 & e & f \\ 0 & 0 & i \end{pmatrix} = aei.$$

So in particular, diagonal matrices have a determinant which is just multiplication along the diagonal:

$$\det(\begin{pmatrix} a & 0 & 0 \\ 0 & e & 0 \\ 0 & 0 & i \end{pmatrix} = aei.$$

- An $n \times n$ matrix can be thought of as a collection of n row vectors in $\mathbf{R}^n$, or as a collection of n column vectors in $\mathbf{R}^n$. Thus one can talk about the determinant $\det(v_1, \ldots, v_n)$ of n column vectors in $\mathbf{R}^n$, or the determinant of n row vectors in $\mathbf{R}^n$, simply by arranging those n row or column vectors into a matrix. Note that the order in which we arrange these vectors will be somewhat important.

- Example: the determinant of the two vectors $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$ is $ad - bc$.

- The determinant of n vectors has two basic properties. One is that it is linear in each column separately. What we mean by this is that

$$\det(v_1, \ldots, v_{j-1}, v_j + w_j, v_{j+1}, \ldots, v_n) =$$

$$\det(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n) + \det(v_1, \ldots, v_{j-1}, w_j, v_{j+1}, \ldots, v_n)$$

  and

$$\det(v_1, \ldots, v_{j-1}, cv_j, v_{j+1}, \ldots, v_n) = c \det(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n).$$

  This linearity can be seen most easily by cofactor expansion in the column j.

- The other basic property it has is anti-symmetry: if one switches two column vectors (not necessarily adjacent), then the determinant changes sign. For instance, when $n = 5$,

$$\det(v_1, v_5, v_3, v_4, v_2) = -\det(v_1, v_2, v_3, v_4, v_5).$$

  This is not completely obvious from the cofactor expansion definition, although the presence of the factor $(-1)^{i+j}$ does suggest that some sort of sign change might occur when one switches rows or columns. We will not prove this anti-symmetry property here.

- (It turns out that the determinant is in fact the *only* expression which obeys these two properties, and which also has the property that the identity matrix has determinant one. But we will not prove that here).

- The same facts hold if we replace columns by rows; i.e. the determinant is linear in each of the rows separately, and if one switches two rows then the determinant changes sign.

- We now write down some properties relating to how determinants behave under elementary row operations.

- **Property 1** If A is an $n \times n$ matrix, and B is the matrix A with two distinct rows i and j interchanged, then $\det(B) = -\det(A)$. (I.e. row operations of the first type flip the sign of the determinant). This is just a restatement of the antisymmetry property for rows.

- Example:
$$\det \begin{pmatrix} c & d \\ a & b \end{pmatrix} = -\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

- Corollary of Property 1: If two rows of a matrix A are the same, then the determinant must be zero.

- **Property 2** If A is an $n \times n$ matrix, and B is the matrix B but with one row i multiplied by a scalar c, then $\det(B) = c\det(A)$. (I.e. row operations of the second type multiply the determinant by whatever scalar was used in the row operation). This is a special case of the linearity property for the i^{th} row.

- Example:
$$\det \begin{pmatrix} ka & kb \\ c & d \end{pmatrix} = k\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

- Corollary of Property 2: if a matrix A has one of its rows equal to zero, then $\det(A)$ is zero (just apply this Property with $c = 0$).

- **Property 3** If A is an $n \times n$ matrix, and B is the matrix B but with c copies of one row i added to another row j, then $\det(B) = \det(A)$. (I.e. row operations of the third type do not affect the determinant). This is a consequence of the linearity property for the j^{th} row, combined with the Corollary to Property 1 (why?).

- Example:
$$\det \begin{pmatrix} a + kc & b + kd \\ c & d \end{pmatrix} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

- Similar properties hold for elementary column operations (just replace "row" by "column" throughout in the above three properties).

- We can summarize the above three properties in the following lemma (which will soon be superceded by a more general statement):

- **Lemma 11.** If E is an elementary matrix, then $\det(EA) = \det(E)\det(A)$ and $\det(AE) = \det(A)\det(E)$.

- This is because the determinant of a type 1 elementary matrix is easily seen to be -1 (from Property 1 applied to the identity matrix), the determinant of a type 2 elementary matrix (multiplying a row by c) is easily seen to be c (from Property 2 applied to the identity matrix), and the determinant of a type 3 elementary matrix is easily seen to be 1 (from Property 3 applied to the identity matrix). In particular, elementary matrices always have non-zero determinant (recall that in the type 2 case, c must be non-zero).

- We are now ready to state one of the most important properties of a determinant: it measures how invertible a matrix is.

- **Theorem 12.** An $n \times n$ matrix is invertible if and only if its determinant is non-zero.

- **Proof** Suppose A is an invertible $n \times n$ matrix. Then by Lemma 10, it is the product of elementary matrices, times the identity I_n. The identity I_n has a non-zero determinant: $\det(I_n) = 1$. Each elementary matrix has non-zero determinant (see above), so by Lemma 11 if a matrix has non-zero determinant, then after multiplying on the left or right by an elementary matrix it still has non-zero determinant. Applying this repeatedly we see that A must have non-zero determinant.

  Now conversely suppose that A had non-zero determinant. By Lemma 11, we thus see that even after applying elementary row and column operations to A, one must still obtain a matrix with non-zero determinant. In particular, in row-echelon form A must still have non-zero determinant, which means in particular that it cannot contain any rows which are entirely zero. Thus A has full rank n, which means that $L_A : \mathbf{R}^n \to \mathbf{R}^n$ is onto. But then L_A would also be one-to-one by the dimension theorem - see Lemma 2 of Week 3 notes, hence L_A would be invertible and hence A is invertible. $\square$

- Not only does the determinant measure invertibility, it also measures linear independence.

- **Corollary 13.** Let $v_1, \ldots, v_n$ be n column vectors in $\mathbf{R}^n$. Then $v_1, \ldots, v_n$ are linearly dependent if and only if $\det(v_1, \ldots, v_n) = 0$.

- **Proof** Suppose that $\det(v_1, \ldots, v_n) = 0$, so that by Theorem 12 the $n \times n$ matrix $(v_1, \ldots, v_n)$ is not invertible. Then the linear transformation $L_{(v_1,\ldots,v_n)}$ cannot be one-to-one, and so there is a non-zero vector $\begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{pmatrix}$ in the null space, i.e.

$$(v_1, \ldots, v_n) \begin{pmatrix} a_1 \\ a_2 \\ \ldots \\ a_n \end{pmatrix} = 0$$

or in other words

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

i.e. $v_1, \ldots, v_n$ are not linearly independent. The converse statement follows by reversing all the above steps and is left to the reader. $\square$

- Note that if one writes down a typical $n \times n$ matrix, then the determinant will in general just be some random number and will usually not be zero. So "most" matrices are invertible, and "most" collections of n vectors in $\mathbf{R}^n$ are linearly independent (and hence form a basis for $\mathbf{R}^n$, since $\mathbf{R}^n$ is n-dimensional).

- Properties 1,2,3 also give a way to compute the determinant of a matrix - use row and column operations to convert it into some sort of triangular or diagonal form, for which the determinant is easy to compute, and then work backwards to recover the original determinant of the matrix.

- **Example**. Suppose we wish to compute the determinant of

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 4 \\ 3 & 2 & 1 \end{pmatrix}.$$

We perform row operations. Subtracting two copies of row 1 from row 2 and using Property 3, we obtain

$$\det(A) = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Similarly subtracting three copies of row 1 from row 2, we obtain

$$\det(A) = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -2 \\ 0 & -4 & -8 \end{pmatrix}.$$

Dividing the third row by $-1/4$ using Property 2, we obtain

$$\frac{-1}{4}\det(A) = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & -2 \\ 0 & 1 & 2 \end{pmatrix}$$

which after swapping two rows using Property 1, becomes

$$\frac{1}{4}\det(A) = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{pmatrix}.$$

But the right-hand side is triangular and has a determinant of -2. Hence $\frac{1}{4}\det(A) = -2$, so that $\det(A) = -8$. (One can check this using the original formula for determinant. Which approach is less work? Which approach is less prone to arithmetical error?)

- We now give another important property of a determinant, namely its multiplicative properties.

- **Theorem 14.** If A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.

- **Proof** First suppose that A is not invertible. Then L_A is not onto (cf. Lemma 2 of Week 3 notes), which implies that $L_A L_B$ is not onto (why? Note that the range of $L_A L_B$ must be contained in the range of L_A), so that AB is not invertible. Then by Theorem 12, both sides of

$\det(AB) = \det(A)\det(B)$ are zero, and we are done. Similarly, suppose that B is not invertible. Then L_B is not one-to-one, and so $L_A L_B$ is not one-to-one (why? Note that the null space of L_B must be contained in the null space of $L_A L_B$). So AB is not invertible. Thus both sides are again zero.

- The only remaining case is when A and B are both invertible. By Lemma 10 we may thus write

$$A = E_1 E_2 \ldots E_a; \quad B = F_1 F_2 \ldots F_b$$

where $E_1, \ldots, E_a, F_1, \ldots, F_b$ are elementary matrices. By many applications of Lemma 11 we thus have

$$\det(A) = \det(E_1)\det(E_2)\ldots\det(E_a)$$

and

$$\det(B) = \det(F_1)\det(F_2)\ldots\det(F_b).$$

But also

$$AB = E_1 \ldots E_A F_1 \ldots F_b$$

and so by taking det of both sides and using Lemma 11 many times again we obtain

$$\det(AB) = \det(E_1)\ldots\det(E_A)\det(F_1)\ldots\det(F_b)$$

and by combining all these equations we obtain $\det(AB) = \det(A)\det(B)$ as desired. $\square$

- **Warning:** The corresponding statement for addition is not true in general, i.e. $\det(A + B) \neq \det(A) + \det(B)$ in general. (Can you think up a counterexample? Even for diagonal matrices one can see this will not work. On the other hand, we still have linearity in each row and column).

- Note that Theorem 14 supercedes Lemma 11, although we needed Lemma 11 as an intermediate step to prove Theorems 12 and 14.

- (Optional) Remember the symmetry property $\det(A^t) = \det(A)$ we stated earlier? It can now be proved using the above machinery. We sketch a proof as follows. First of all, if A is non-invertible, then A^t is also non-invertible (why?), and so both sides are zero. Now if A is invertible, then by Lemma 10 it is the product of elementary matrices:

  $$A = E_1 E_2 \dots E_a$$

  and so

  $$\det(A) = \det(E_1) \dots \det(E_a).$$

  On the other hand, taking transposes (and recalling that transpose reverses multiplication order) we obtain

  $$A^t = E_a^t \dots E_2^t E_1^t$$

  and so

  $$\det(A^t) = \det(E_a^t) \dots \det(E_1^t).$$

  But a direct computation (checking the three types of elementary matrix separately) shows that $\det(E^t) = \det(E)$ for every elementary matrix, so

  $$\det(A^t) = \det(E_a) \dots \det(E_1).$$

  Thus $\det(A^t) = \det(A)$ as desired.

* * * * *

Geometric interpretation of determinants (optional)

- This material is optional, and is also not covered in full detail. It is intended only for those of you who are interested in the geometric provenance of determinants.

- Up until now we've treated the determinant as a mysterious algebraic expression which has a lot of remarkable properties. But we haven't delved much into what the determinant actually means, and why we have any right to have such a remarkable characteristic of matrices. It turns out that the determinant measures something very fundamental to the geometry of $\mathbf{R}^n$, namely n-dimensional volume. The one caveat is that determinants can be either positive or negative, while volume can only be positive, so determinants are in fact measuring *signed volume* - volume with a sign. (This is similar to how a definite integral $\int_a^b f(x)\,dx$ can be negative if f dips below the x axis, even though the "area under the curve" interpretation of $f(x)$ seems to suggest that integrals must always be positive).

- Let's begin with $\mathbf{R}^1$. The determinant $\det(v_1)$ of a single vector $v_1 = (a)$ in $\mathbf{R}^1$ is of course a, which is plus or minus the length $|a|$ of that vector; plus if the vector is pointing right, and minus if the vector is pointing left. In the degenerate case $v_1 = 0$, the determinant is of course zero.

- Now let's look at $\mathbf{R}^2$, and think about the determinant $\det(v_1, v_2)$ of two vectors v_1, v_2 in $\mathbf{R}^2$. This turns out to be (plus or minus) the area of the parallelogram with sides v_1 and v_2; plus if v_2 is anticlockwise of v_1, and minus if v_2 is clockwise of v_1. For instance, $\det((1,0),(0,1))$ is the area of the square with sides $(1,0),(0,1)$, i.e. 1. On the other hand, $\det((0,1),(1,0))$ is -1 because $(0,1)$ is clockwise of $(1,0)$. Similarly, $\det((3,0),(0,1))$ is 3, because the rectangle with sides $(3,0)$, $(0,1)$ has area 3, and $\det((3,1),(0,1))$ is also 3, because the parallelogram with sides $(3,1),(0,1)$ has the same area as the previous rectangle.

- This parallelogram property can be proven using cross products (recall that the cross product can be used to measure the area of a parallelogram). It is also interesting to interpret Properties 1, 2, 3 using this area interpretation. Property 1 says that if you swap the two vectors v_1 and v_2, then the sign of the determinant changes. Property 2 says that if you dilate one of the vectors by c, then the area of the parallelogram also dilates by c (note that if c is negative, then the determinant changes sign, even though the area is of course always positive, because you flip the clockwiseness of v_1 and v_2). Property 3 says that if you slide v_2 (say) by a constant multiple of v_1, then the area of the parallelogram doesn't change. (This is consistent with the familiar "base $\times$ height" formula for parallelograms - sliding v_2 by v_1 does not affect either the base or the height).

- Note also that if v_1 and v_2 are linearly dependent, then their parallelogram has area 0; this is consistent with Corollary 13.

- Now let's look at $\mathbf{R}^3$, and think about the determinant $\det(v_1, v_2, v_3)$ of three vectors in $\mathbf{R}^3$. Thus turns out to be (plus or minus) the volume of the *parallelopiped* with sides v_1, v_2, v_3 (you may remember this from Math 32A). To determine plus or minus, one uses the right-hand rule: if the thumb is at v_1 and the second finger is at v_2, and the middle finger is at v_3, then we have a plus sign if this can be achieved using the right hand, and a minus sign if it can be achieved using the left-hand. For instance, $\det((1,0,0),(0,1,0),(0,0,1)) = 1$, but $\det((0,1,0),(1,0,0),(0,0,1)) = -1$. It is an instructive exercise to interpret Properties 1,2,3 using this geometric picture, as well as Corollary 13.

- The two-dimensional rule of "determinant is positive if v_2 clockwise of v_1" can be interpreted as a right-hand rule using a two-dimensional hand, while the one-dimensional rule of "determinant is positive if v_1 is on the right" can be interpreted as a right-hand rule using a one-dimensional hand.

- There is a similar link between determinant and n-dimensional volume in higher dimensions $n \geq 3$, but it is of course much more difficult to visualize, and beyond the scope of this course (one needs some grounding in measure theory, anyway, in order to understand what "n-dimensional volume" means. Also, one needs n-dimensional hands.). But in particular, we see that the volume of a parallelopiped with edges $v_1, \ldots, v_n$ is the *absolute value* of the determinant $\det(v_1, \ldots, v_n)$. (Note that it doesn't particularly matter whether we use row or column vectors here since $\det(A^t) = \det(A)$).

- Let $v_1, \ldots, v_n$ be n column vectors in $\mathbf{R}^n$, so that $(v_1, \ldots, v_n)$ is an $n \times n$ matrix, and consider the linear transformation

$$T := L_{(v_1, \ldots, v_n)}.$$

Observe that

$$T(1, 0, \ldots, 0) = v_1$$
$$T(0, 1, \ldots, 0) = v_2$$
$$\ldots$$

$$T(0, 0, \ldots, 1) = v_n$$

(why is this the case?) So if we let Q be the unit cube with edges $(1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)$, then T will map Q to the n-dimensional parallelopiped with vectors $v_1, \ldots, v_n$. (If you are having difficulty imagining n-dimensional parallelopipeds, you may just want to think about the $n = 3$ case). Thus $T(Q)$ has volume $|\det(v_1, \ldots, v_n)|$, while Q of course had volume 1. Thus T *expands volume* by a factor of $|\det(v_1, \ldots, v_n)|$. Thus the *magnitude* $|\det(A)|$ of a determinant measures how much the linear transformation L_A expands (n-dimensional) volume.

- **Example** Consider the matrix

$$A := \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$$

which as we know has the corresponding linear transformation

$$L_A(x_1, x_2) = (5x_1, 3x_2).$$

This dilates the x_1 co-ordinate by 5 and the x_2 co-ordinate by 3, so area (which is 2-dimensional volume) is expanded by 15. This is consistent with $\det(A) = 15$. Note that if we replace 3 with -3, then the determinant becomes -15 but area still expands by a factor of 15 (why?). Also, if we replace 3 instead with 0, then the determinant becomes 0. What happens to the area in this case?

- **Example** Now consider the matrix

$$A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

which as we know has the corresponding linear transformation

$$L_A(x_1, x_2) = (x_1 + x_2, x_2).$$

This matrix shears the x_1 co-ordinate horizontally by an amount depending on the x_2 co-ordinate, but area is unchanged (why? It has to do with the base $\times$ height formula for area). This is consistent with the determinant being 1.

- So the magnitude of the determinant measures the volume-expanding properties of a linear transformation. The sign of a determinant will measure the *orientation-preserving* properties of a transformation: will a "right-handed" object remain right-handed when one applies the transformation? If so, the determinant is positive; if however right-handed objects become left-handed, then the determinant is negative.

- **Example** The reflection matrix

$$A := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

corresponds to reflection through the x_1-axis:

$$L_A(x_1, x_2) = (x_1, -x_2).$$

It is clear that a "right-handed" object (which in two-dimensions, means an arrow pointing anti-clockwise) will reflect to a "left-handed" object (an arrow pointing clockwise). This is why reflections have negative determinant.

- This interpretation of determinant, as measuring both the volume expanding and the orientation preserving properties of a transformation, also allow us to interpret Theorem 14 geometrically. For instance, if $T : \mathbf{R}^n \to \mathbf{R}^n$ expands volume by a factor of 4 and flips the orientation (so $\det[T]_\beta^\beta = -4$, where β is the standard ordered basis), and $S : \mathbf{R}^n \to \mathbf{R}^n$ expands volume by a factor of 3 and also flips the orientation (so $\det[S]_\beta^\beta = -3$), then one can now see why ST should expand volume by 12 and preserve orientation (so $\det[ST]_\beta^\beta = +12$).

- We now close with a little lemma that says that to take the determinant of a matrix, it doesn't matter what basis you use.

- **Lemma 15.** If two matrices are similar, then they have the same determinant.

- **Proof** If A is similar to B, then $B = Q^{-1}AQ$ for some invertible matrix Q. Thus by Theorem 13

$$\det(B) = \det(Q^{-1}) \det(A) \det(Q) = \det(A) \det(Q^{-1}) \det(Q)$$

$$= \det(A)\det(Q^{-1}Q) = \det(A)\det(I_n) = \det(A)$$

as desired. □

- **Corollary 16.** Let $T : \mathbf{R}^n \to \mathbf{R}^n$ be a linear transformation, and let β, β' be two ordered bases for $\mathbf{R}^n$. Then the matrices $[T]_\beta^\beta$ and $[T]_{\beta'}^{\beta'}$ have the same determinant.

- **Proof.** From last week's notes we know that $[T]_\beta^\beta$ and $[T]_{\beta'}^{\beta'}$ are similar, and the result follows from Lemma 15. □

Math 115A - Week 7
Textbook sections: 4.5, 5.1-5.2
Topics covered:

- Cramer's rule
- Diagonal matrices
- Eigenvalues and eigenvectors
- Diagonalization

* * * * *

Cramer's rule

- Let A be an $n \times n$ matrix. Last week we introduced the notion of the determinant $\det(A)$ of A, and also that of a *cofactor* $\tilde{A}_{ij}$ associated to each row i and column j. Given any row i, we then have the cofactor expansion formula

  $$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \tilde{A}_{ij}.$$

  For instance, if

  $$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

  then

  $$\det(A) = a\tilde{A}_{11} - b\tilde{A}_{12} + c\tilde{A}_{13}$$

  or in other words

  $$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}.$$

- Suppose we replace the row (a, b, c) by (d, e, f) in the above example. Then we have

$$\det \begin{pmatrix} d & e & f \\ d & e & f \\ g & h & i \end{pmatrix} = d \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - e \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + f \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}.$$

But the left-hand side is zero because two of the rows are the same (see Property 1 of determinants on the previous week's notes). Thus we have

$$0 = d\tilde{A}_{11} - e\tilde{A}_{12} + f\tilde{A}_{13}.$$

Similarly we have

$$0 = g\tilde{A}_{11} - h\tilde{A}_{12} + i\tilde{A}_{13}.$$

We can also do the same analysis with the cofactor expansion along the second row

$$\det(A) = -d\tilde{A}_{21} + e\tilde{A}_{22} - f\tilde{A}_{23}$$

yielding

$$0 = -a\tilde{A}_{21} + b\tilde{A}_{22} - c\tilde{A}_{23}$$
$$0 = -g\tilde{A}_{21} + h\tilde{A}_{22} - i\tilde{A}_{23}.$$

And similarly for the third row:

$$\det(A) = g\tilde{A}_{31} - h\tilde{A}_{32} + i\tilde{A}_{33}$$

$$0 = a\tilde{A}_{31} - b\tilde{A}_{33} + c\tilde{A}_{33}$$
$$0 = d\tilde{A}_{31} - e\tilde{A}_{32} + f\tilde{A}_{33}.$$

We can put all these nine identies together in a compact matrix form as

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} +\tilde{A}_{11} & -\tilde{A}_{21} & +\tilde{A}_{31} \\ -\tilde{A}_{12} & +\tilde{A}_{22} & -\tilde{A}_{32} \\ +\tilde{A}_{13} & -\tilde{A}_{23} & +\tilde{A}_{33} \end{pmatrix} = \begin{pmatrix} \det(A) & 0 & 0 \\ 0 & \det(A) & 0 \\ 0 & 0 & \det(A) \end{pmatrix}.$$

The second matrix on the left-hand side is known as the *adjugate* of A, and is denoted $adj(A)$:

$$adj(A) := \begin{pmatrix} \tilde{A}_{11} & -\tilde{A}_{21} & \tilde{A}_{31} \\ -\tilde{A}_{12} & \tilde{A}_{22} & -\tilde{A}_{32} \\ \tilde{A}_{13} & -\tilde{A}_{23} & \tilde{A}_{33} \end{pmatrix}.$$

Thus we have the identity for 3×3 matrices

$$Aadj(A) = \det(A) I_3.$$

The adjugate matrix is the transpose of the *cofactor matrix* $cof(A)$:

$$cof(A) := \begin{pmatrix} \tilde{A}_{11} & -\tilde{A}_{12} & \tilde{A}_{13} \\ -\tilde{A}_{21} & \tilde{A}_{22} & -\tilde{A}_{23} \\ \tilde{A}_{31} & -\tilde{A}_{32} & \tilde{A}_{33} \end{pmatrix}.$$

Thus $adj(A) = cof(A)^t$. To compute the cofactor matrix, at every row i and column j we extract the minor corresponding to that row and column, take the $n-1 \times n-1$ determinant of the minor, and then place that number in the ij entry of the cofactor matrix. Then we alternate the signs by $(-1)^{i+j}$.

- More generally, for $n \times n$ matrices, we can define the cofactor matrix by

  $$cof(A)_{ij} = (-1)^{i+j} \tilde{A}_{ij}$$

  and the adjugate matrix by $adj(A) = cof(A)^t$, so that

  $$adj(A)_{ij} = (-1)^{j+i} \tilde{A}_{ji},$$

  and then we have the identity

  $$Aadj(A) = \det(A) I_n.$$

  If $\det(A)$ is non-zero, then A is invertible and we thus have

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

  This is known as *Cramer's rule* - it allows us to compute the inverse of a matrix using determinants. (There is also a closely related rule, also known as Cramer's rule, which allows one to solve equations $Ax = b$ when A is invertible; see the textbook).

- For example, in the 2×2 case

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then the cofactor matrix is

$$cof(A) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

and so the adjugate matrix is

$$adj(A) = cof(A)^t = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

and so, if $\det(A) \neq 0$, the inverse of A is

$$A^{-1} = \frac{1}{\det(A)} adj(A) = \frac{1}{ad - bc} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

* * * * *

Diagonal matrices

- Matrices in general are complicated objects to manipulate, and we are always looking for ways to simplify them into something better. Last week we explored one such way to do so: using elementary row (or column operations) to reduce a matrix into row-echelon form, or to even simpler forms. This type of simplification is good for certain purposes (computing rank, determinant, inverse), but is not good for other purposes. For instance, suppose you want to raise a matrix A to a large power, say A^{100}. Using elementary matrices to reduce A to, say, row-echelon form will not be very helpful, because (a) it is still not very easy to raise row-echelon form matrices to very large powers, and (b) one has to somehow deal with all the elementary matrices you used to convert A into row echelon form. However, to perform tasks like this there is a better factorization available, known as *diagonalization*. But before we do this, we first digress on diagonal matrices.

- **Definition** An $n \times n$ matrix A is said to be *diagonal* if all the off-diagonal entries are zero, i.e. $A_{ij} = 0$ whenever $i \neq j$. Equivalently, a diagonal matrix is of the form

$$A = \begin{pmatrix} A_{11} & 0 & \ldots & 0 \\ 0 & A_{22} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & A_{nn} \end{pmatrix}.$$

We write this matrix as $\mathrm{diag}(A_{11}, A_{22}, \ldots, A_{nn})$. Thus for instance

$$\mathrm{diag}(1, 3, 5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

- Diagonal matrices are very easy to add, scalar multiply, and multiply. One can easily verify that

$$\mathrm{diag}(a_1, a_2, \ldots, a_n) + \mathrm{diag}(b_1, b_2, \ldots, b_n) = \mathrm{diag}(a_1+b_1, a_2+b_2, \ldots, a_n+b_n),$$

$$c\,\mathrm{diag}(a_1, a_2, \ldots, a_n) = \mathrm{diag}(ca_1, ca_2, \ldots, ca_n)$$

and

$$\mathrm{diag}(a_1, a_2, \ldots, a_n)\mathrm{diag}(b_1, b_2, \ldots, b_n) = \mathrm{diag}(a_1 b_1, a_2 b_2, \ldots, a_n b_n).$$

Thus for instance

$$\mathrm{diag}(1, 3, 5)\mathrm{diag}(1, 3, 5) = \mathrm{diag}(1^2, 3^2, 5^2)$$

and more generally

$$\mathrm{diag}(1, 3, 5)^n = \mathrm{diag}(1^n, 3^n, 5^n).$$

Thus raising diagonal matrices to high powers is very easy. More generally, polynomial expressions of a diagonal matrix are very easy to compute. For instance, consider the polynomial $f(x) = x^3 + 4x^2 + 2$. We can apply this polynomial to any $n \times n$ matrix A, creating the new matrix $f(A) = A^3 + 4A^2 + 2$. In general, such a matrix may be difficult to compute. But for a diagonal matrix $A = \mathrm{diag(a_1, a_2, \ldots, a_n)}$, we have

$$A^2 = \mathrm{diag(a_1^2, a_2^2, \ldots, a_n^2)}$$
$$A^3 = \mathrm{diag(a_1^3, a_2^3, \ldots, a_n^3)}$$

and thus

$$f(A) = \mathrm{diag(a_1^3{+}4a_1^2{+}2, a_2^3{+}4a_2^2{+}2, \ldots, a_n^3{+}4a_n^2{+}2) = diag(f(a_1), \ldots, f(a_n))}.$$

This is true for more general polynomials f:

$$f(\mathrm{diag(a_1, \ldots, a_n)}) = \mathrm{diag(f(a_1), \ldots, f(a_n))}.$$

Thus to do any sort of polynomial operation to a diagonal matrix, one just has to perform it on the diagonal entries separately.

- If A is an $n \times n$ diagonal matrix $A = \mathrm{diag(a_1, \ldots, a_n)}$, then the linear transformation $L_A : \mathbf{R}^n \to \mathbf{R}^n$ is very simple:

$$L_A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_1 x_1 \\ a_2 x_2 \\ \vdots \\ a_n x_n \end{pmatrix}.$$

Thus L_A dilates the first co-ordinate x_1 by a_1, the second co-ordinate x_2 by a_2, and so forth. In particular, if $\beta = (e_1, e_2, \ldots, e_n)$ is the standard ordered basis for $\mathbf{R}^n$, then

$$L_A e_1 = a_1 e_1; \quad L_A e_2 = a_2 e_2; \ldots; \quad L_A e_n = a_n e_n.$$

This leads naturally to the concept of *eigenvalues* and *eigenvalues*, which we now discuss.

- Remember from last week that the rank of a matrix is equal to the number of non-zero rows in row echelon form. Thus it is easy to see that

- **Lemma 1.** The rank of a diagonal matrix is equal to the number of its non-zero entries.

- Thus, for instance, $\mathrm{diag}(3, 4, 5, 0, 0, 0)$ has rank 3.

* * * * *

Eigenvalues and eigenvectors

- Let $T : V \to V$ be a linear transformation from a vector space V to itself. One of the simplest possible examples of such a transformation is the identity transformation $T = I_V$, so that $Tv = v$ for all $v \in V$. After the identity operation, the next simplest example of such a transformation is a dilation $T = \lambda I_V$ for some scalar λ, so that $Tv = \lambda v$ for all $v \in V$.

- In general, though, T does not look like a dilation. However, there are often some special vectors in V for which T is as simple as a dilation, and these are known as *eigenvectors*.

- **Definition** An *eigenvector* v of T is a non-zero vector $v \in V$ such that $Tv = \lambda v$ for some scalar λ. The scalar λ is known as the *eigenvalue* corresponding to v.

- **Example** Consider the linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) := (5x, 3y)$. Then the vector $v = (1, 0)$ is an eigenvector of T with eigenvalue 5, since $Tv = T(1, 0) = (5, 0) = 5v$. More generally, any non-zero vector of the form $(x, 0)$ is an eigenvector with eigenvalue 5. Similarly, $(0, y)$ is an eigenvector of T with eigenvalue 3, if y is non-zero. The vector $v = (1, 1)$ is *not* an eigenvector, because $Tv = (5, 3)$ is not a scalar multiple of v.

- **Example** More generally, if $A = \mathrm{diag}(\mathrm{a_1}, \ldots, \mathrm{a_n})$ is a diagonal matrix, then the basis vectors $e_1, \ldots, e_n$ are eigenvectors for L_A, with eigenvalues $a_1, \ldots, a_n$ respectively.

- **Example** If T is the identity operator, then every non-zero vector is an eigenvector, with eigenvalue 1 (why?). More generally, if $T = \lambda I_V$ is λ times the identity operator, then every non-zero vector is an eigenvector, with eigenvalue λ (Why?).

- **Example** If $T : V \to V$ is any linear transformation, and v is any non-zero vector in the null space $N(T)$, then v is an eigenvector with eigenvalue 0. (Why?)

- **Example** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the reflection through the line l connecting the origin to $(4,3)$. Then $(4,3)$ is an eigenvector with eigenvalue 1 (why?), and $(3,-4)$ is an eigenvector with eigenvalue -1 (why?).

- We do not consider the 0 vector as an eigenvector, even though $T0$ is always 0, because we cannot determine what eigenvalue 0 should have.

- If A is an $n \times n$ matrix, we say that v is an *eigenvector* for A with eigenvalue λ if it is already an eigenvector for L_A with eigenvalue λ, i.e. $Av = \lambda v$. In other words, for the purposes of computing eigenvalues and eigenvectors we do not distinguish between a matrix A and its linear transformation L_A.

- (Incidentally, the word "eigen" is German for "own". An eigenvector is a vector which keeps its *own* direction when acted on by T. The terminology is thus a hybrid of German and English, though some people prefer "principal value" and "principal vector" to avoid this (or "characteristic" or "proper" instead of "principal"). Then again, "vector" is pure Latin. English is very cosmopolitan).

- **Definition** Let $T : V \to V$ be a linear transformation, and let λ be a scalar. Then the *eigenspace* of T corresponding to λ is the set of all vectors (including 0) such that $Tv = \lambda v$.

- Thus an eigenvector with eigenvalue λ is the same thing as a non-zero element of the eigenspace with eigenvalue λ. Since $Tv = \lambda v$ is equivalent to $(T - \lambda I_V)v = 0$, we thus see that the eigenspace of T with eigenvalue λ is the same thign as the null space $N(T - \lambda I_V)$ of $T - \lambda I_V$. In particular, the eigenspace is always a subspace of V. From the above discussion we also see that λ is an eigenvalue of T if and only if $N(T - \lambda I_V)$ is non-zero, i.e. when $T - \lambda I_V$ is not one-to-one.

- **Example** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the transformation $T(x,y) := (5x, 3y)$. Then the x-axis is the eigenspace $N(T - 3I_{\mathbf{R}^2})$ with eigenvalue 5, while the y-axis is the eigenspace $N(T - 5I_{\mathbf{R}^2})$ with eigenvalue 5. For all other values $\lambda \neq 3, 5$, the eigenspace $N(T - \lambda I_{\mathbf{R}^2})$ is just the zero vector space $\{0\}$ (why?).

- The relationship between eigenvectors and diagonal matrices is the following.

- **Lemma 2.** Suppose that V is an n-dimensional vector space, and suppose that $T : V \to V$ is a linear transformation. Suppose that V has an ordered basis $\beta = (v_1, v_2, \ldots, v_n)$, such that each v_j is an eigenvector of T with eigenvalue λ_j. Then the matrix $[T]_\beta^\beta$ is a diagonal matrix; in fact $[T]_\beta^\beta = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n})$.

- Conversely, if $\beta = (v_1, v_2, \ldots, v_n)$ is a basis such that $[T]_\beta^\beta = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n})$, then each v_j is an eigenvector of T with eigenvalue λ.

- **Proof** Suppose that v_j is an eigenvector of T with eigenvalue λ_j. Then $Tv_j = \lambda_j v_j$, so $[Tv_j]^\beta$ is just the column vector with j^{th} entry equal to λ_j, and all other entries zero. Putting all these column vectors together we see that $[T]_\beta^\beta = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n})$. Conversely, if $[T]_\beta^\beta = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n})$, then by definition of $[T]_\beta^\beta$ we see that $Tv_j = \lambda_j v_j$, and so v_j is an eigenvector with eigenvalue λ_j. $\square$

- **Definition.** A linear transformation $T : V \to V$ is said to be *diagonalizable* if there is an ordered basis β of V for which the matrix $[T]_\beta^\beta$ is diagonal.

- Lemma 2 thus says that a transformation is diagonalizable if and only if it has a basis consisting entirely of eigenvectors.

- **Example** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the reflection through the line l connecting the origin to $(4, 3)$. Then $(4, 3)$ and $(3, -4)$ are both eigenvectors for T. Since these two vectors are linearly independent and $\mathbf{R}^2$ is two-dimensional, they form a basis for T. Thus T is diagonalizable; indeed, if $\beta := ((4, 3), (3, -4))$, then

$$[T]_\beta^\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \mathrm{diag}(1, -1).$$

- If one knows how to diagonalize a transformation, then it becomes very easy to manipulate. For example, in the above reflection example we see very quickly that T must have rank 2 (since $\mathrm{diag}(1,-1)$ has two non-zero entries). Also, we can square T easily:

$$[T^2]_\beta^\beta = \mathrm{diag}(1,-1)^2 = \mathrm{diag}(1,1) = \mathrm{I}_2 = [\mathrm{I}_{\mathbf{R}^2}]_\beta^\beta$$

and hence $T^2 = I_{\mathbf{R}^2}$, the identity transformation. (Geometrically, this amounts to the fact that if you reflect twice around the same line, you get the identity).

- **Definition.** An $n \times n$ matrix A is said to be *diagonalizable* if the corresponding linear transformation L_A is diagonalizable.

- **Example** The matrix $A = \mathrm{diag}(5,3)$ is diagonalizable, because the linear operator L_A in the standard basis $\beta = ((1,0),(0,1))$ is just A itself: $[L_A]_\beta^\beta = A$, which is diagonal. So all diagonal matrices are diagonalizable (no surprise there).

- **Lemma 3.** A matrix A is diagonalizable if and only if $A = QDQ^{-1}$ for some invertible matrix Q and some diagonal matrix D. In other words, a matrix is diagonalizable if and only if it is similar to a diagonal matrix.

- **Proof** Suppose A was diagonalizable. Then $[L_A]_{\beta'}^{\beta'}$ would be equal to some diagonal matrix D, for some choice of basis β' (which may be different from the standard ordered basis β. But by the change of variables formula,

$$A = [L_A]_\beta^\beta = Q[L_A]_{\beta'}^{\beta'}Q^{-1} = QDQ^{-1}$$

as desired, where $Q := [I_{\mathbf{R}^n}]_{\beta'}^\beta$ is the change of variables matrix.

Conversely, suppose that $A = QDQ^{-1}$ for some invertible matrix Q. Write $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n})$, so that $De_j = \lambda_j$. Then

$$A(Qe_j) = QDQ^{-1}Qe_j = QDe_j = Q\lambda_j e_j = \lambda_j(Qe_j)$$

and so Qe_j is an eigenvector for A with eigenvalue λ_j. Since Q is invertible and $e_1, \ldots, e_n$ is a basis, we see that $Qe_1, \ldots, Qe_n$ is also a basis (why?). Thus we have found a basis of $\mathbf{R}^n$ consisting entirely of eigenvectors of A, and so A is diagonalizable. $\square$

- From Lemma 2 and Lemma 3 we see that if we can find a basis $(v_1, \ldots, v_n)$ of $\mathbf{R}^n$ which consists entirely of eigenvectors of A, then A is diagonalizable and $A = QDQ^{-1}$ for some diagonal matrix D and some invertible matrix Q. We now make this statement more precise, specifying precisely what Q and D are.

- **Lemma 4.** Let A be an $n \times n$ matrix, and suppose that $(v_1, \ldots, v_n)$ is an ordered basis of $\mathbf{R}^n$ such that each v_j is an eigenvector of A with eigenvalue λ_j (i.e. $Av_j = \lambda v_j$ for $j = 1, \ldots, n$). Then we have

$$A = Q\mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n})\mathrm{Q}^{-1}$$

where Q is the $n \times n$ matrix with columns $v_1, v_2, \ldots, v_n$:

$$Q = (v_1, v_2, \ldots, v_n).$$

- **Proof.** Let β' be the ordered basis $\beta' := (v_1, v_2, \ldots, v_n)$ of $\mathbf{R}^n$, and let $\beta := (e_1, e_2, \ldots, e_n)$ be the standard ordered basis of $\mathbf{R}^n$. Then

$$[I_{\mathbf{R}^n}]_{\beta'}^{\beta} = Q$$

(why?). So by the change of variables formula

$$A = [L_A]_\beta^\beta = Q[L_A]_{\beta'}^{\beta'} Q^{-1}.$$

On the other hand, since $L_A v_j = \lambda_j v_j$, we see that

$$[L_A]_{\beta'}^{\beta'} = \mathrm{diag}(\lambda_1, \ldots, \lambda_\mathrm{n}).$$

Combining these two equations we obtain the lemma. $\square$.

* * * * *

Computing eigenvalues

- Now we compute the eigenvalues and eigenvectors of a general matrix. The key lemma here is

- **Lemma 5.** A scalar λ is an eigenvalue of an $n \times n$ square matrix A if and only $\det(A - \lambda I_n) = 0$.

- **Proof.** If λ is an eigenvalue of A, then $Av = \lambda v$ for some non-zero v, thus $(A - \lambda I_n)v = 0$. Thus $A - \lambda I_n$ is not invertible, and so $\det(A - \lambda I_n) = 0$ (Theorem 12 from last week's notes). Conversely, if $\det(A - \lambda I_n) = 0$, then $A - \lambda I_n$ is not invertible (again by Theorem 12), which means that the corresponding linear transformation is not one-to-one (recall that one-to-one and onto are equivalent when the domain and range have the same dimension; see Lemma 2 of Week 3 notes). So we have $(A - \lambda I_n)v = 0$ for some non-zero v, which means that $Av = \lambda v$ and hence λ is an eigenvalue. $\square$

- Because of this lemma, we call $\det(A - \lambda I_n)$ the *characteristic polynomial* of A, and sometimes call it $f(\lambda)$. Lemma 5 then says that the eigenvalues of A are precisely the zeroes of $f(\lambda)$.

- **Example** Let A be the matrix

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

Then the characteristic polynomial $f(\lambda)$ is given by

$$f(\lambda) = \det \begin{pmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix} = -\lambda(1-\lambda) - 1 \times 1 = \lambda^2 - \lambda - 1.$$

From the quadratic formula, this polynomial has zeroes when $\lambda = (1 \pm \sqrt{5})/2$, and so the eigenvalues are $\lambda_1 := (1+\sqrt{5})/2 = 1.618\ldots$ and $\lambda_2 = (1-\sqrt{5})/2 := -0.618\ldots$.

- Once we have the eigenvalues of A, we can compute eigenvectors, because the eigenvectors with eigenvalues λ are precisely those non-zero vectors in the null-space of $A - \lambda I_n$ (or equivalently, of the null space of $L_A - \lambda I_{\mathbf{R}^n}$).

- **Example:** Let A be the above matrix. Let us try to find the eigenvectors $\begin{pmatrix} x \\ y \end{pmatrix}$ with eigenvalue $\lambda_1 = (1+\sqrt{5})/2$. In other words, we want to solve the equation

$$(A - (1+\sqrt{5})/2) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or in other words

$$\begin{pmatrix} -(1+\sqrt{5})/2 & 1 \\ 1 & (1-\sqrt{5})/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or equivalently

$$\begin{aligned} -\frac{1+\sqrt{5}}{2}x + y &= 0 \\ x + \frac{1-\sqrt{5}}{2}y &= 0. \end{aligned}$$

This matrix does not have full rank (since its determinant is zero - why should this not be surprising?). Indeed, the second equation here is just $(1-\sqrt{5})/2$ times the first equation. So the general solution is y arbitrary, and x equal to $-\frac{1-\sqrt{5}}{2}y$. In particular, we have

$$v_1 := \begin{pmatrix} -\frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix}$$

as an eigenvector of A with eigenvalue $\lambda_1 = (1+\sqrt{5})/2$.

A similar argument gives

$$v_2 := \begin{pmatrix} -\frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix}$$

as an eigenvector of A with eigenvalue $\lambda_2 = (1-\sqrt{5})/2$. Thus we have $Av_1 = \lambda_1 v_1$ and $Av_2 = \lambda_2 v_2$. Thus, if we let β' be the ordered basis $\beta' := (v_1, v_2)$, then

$$[L_A]_{\beta'}^{\beta'} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Thus A is diagonalizable. Indeed, from Lemma 4 we have

$$A = QDQ^{-1}$$

where $D := \mathrm{diag}(\lambda_1, \lambda_2)$ and $Q := (v_1, v_2)$.

- As an application, we recall the example of Fibonacci's rabbits from Week 2. If at the beginning of a year there are x pairs of juvenile rabbits and y pairs of adult rabbits - which we represent by the vector $\begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbf{R}^2$ - then at the end of the next year there will be y juvenile pairs and $x+y$ adult pairs - so the new vector is

$$\begin{pmatrix} y \\ x+y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix}.$$

Thus, each passage of a year multiplies the population vector by A. So if we start with one juvenile pair and no adult pairs - so the population vector is initially $v_0 := \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ - then after n years, the population vector should become $A^n v_0$. To compute this, one would have to multiply A by itself n times, which appears to be difficult (try computing A^5 by hand, for instance!). However, this can be done efficiently using the diagonalization $A = QDQ^{-1}$ we have. Observe that

$$A^2 = QDQ^{-1}QDQ^{-1} = QD^2Q^{-1}$$

$$A^3 = A^2A = QD^2Q^{-1}QDQ^{-1} = QD^3Q^{-1}$$

and more generally (by induction on n)

$$A^n = QD^nQ^{-1}.$$

In particular, our population vector after n years is

$$A^n v_0 = QD^nQ^{-1}v_0.$$

But since D is the diagonal matrix $D = \operatorname{diag}(\lambda_1, \lambda_2)$, D^n is easy to compute:

$$D^n = \operatorname{diag}(\lambda_1^{\mathrm{n}}, \lambda_2^{\mathrm{n}}).$$

Now we can compute $QD^nQ^{-1}v_0$. Since

$$Q = (v_1, v_2) = \begin{pmatrix} -\frac{1-\sqrt{5}}{2} & -\frac{1+\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}$$

we have

$$\det(Q) = -\frac{1-\sqrt{5}}{2} + \frac{1-\sqrt{5}}{2} = \sqrt{5}$$

and so by Cramer's rule

$$Q^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & \frac{1+\sqrt{5}}{2} \\ -1 & -\frac{1-\sqrt{5}}{2} \end{pmatrix}$$

and so

$$Q^{-1} v_0 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

and hence

$$D^n Q^{-1} v_0 = \mathrm{diag}(\lambda_1^{\mathrm{n}}, \lambda_2^{\mathrm{n}}) \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1^n \\ -\lambda_2^n \end{pmatrix}.$$

Since $\frac{1-\sqrt{5}}{2} \frac{1+\sqrt{5}}{2} = \frac{1-5}{4} = -1$, we have

$$Q = \begin{pmatrix} -\frac{1-\sqrt{5}}{2} & -\frac{1+\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \lambda_1^{-1} & \lambda_2^{-1} \\ 1 & 1 \end{pmatrix}$$

and hence

$$A^n v_0 = Q D^n Q^{-1} v_0 = \begin{pmatrix} \lambda_1^{-1} & \lambda_2^{-1} \\ 1 & 1 \end{pmatrix} \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1^n \\ -\lambda_2^n \end{pmatrix} = \begin{pmatrix} (\lambda_1^{n-1} - \lambda_2^{n-1})/\sqrt{5} \\ (\lambda_1^n - \lambda_2^n)/\sqrt{5} \end{pmatrix}.$$

Thus, after n years, the number of pairs of juvenile rabbits is

$$F_{n-1} = (\lambda_1^{n-1} - \lambda_2^{n-1})/\sqrt{5} = ((1.618\ldots)^{n-1} - (-0.618\ldots)^{n-1})/2.236\ldots,$$

and the number of pairs of adult rabbits is

$$F_n = (\lambda_1^n - \lambda_2^n)/\sqrt{5} = ((1.618\ldots)^n - (-0.618\ldots)^n)/2.236\ldots.$$

This is a remarkable formula - it does not look like it at all, but the expressions F_{n-1}, F_n are always integers. For instance

$$F_3 = ((1.618\ldots)^3 - (-0.618\ldots)^3)/2.236\ldots = 2.$$

(Check this!). The numbers

$$F_0 = 0, F_1 = 1, F_2 = 1, F_3 = 2, F_4 = 3, F_5 = 5, F_6 = 8, F_7 = 13, F_8 = 25, \ldots$$

are known as *Fibonacci numbers* and come up in all sorts of places (including, oddly enough, the number of petals on flowers and pine cones). The above formula shows that these numbers grow exponentially, and are comparable to $(1.618)^n$ when n gets large. The number $1.618\ldots = (1+\sqrt{5})/2$ is known as the *golden ratio* and has several interesting properties, which we will not go into here.

- **A final note.** Up until now, we have always chosen the field of scalars to be real. However, it will now sometimes be convenient to change the field of scalars to be complex, because one gets more eigenvalues and eigenvectors this way. For instance, consider the matrix

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

The characteristic polynomial $f(\lambda)$ is given by

$$f(\lambda) = \det \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} = \lambda^2 + 1.$$

If one restricts the field of scalars to be real, then $f(\lambda)$ has no zeroes, and so there are no real eigenvalues (and thus no real eigenvectors). On the other hand, if one expands the field of scalars to be complex, then $f(\lambda)$ has zeroes at $\lambda = \pm i$, and one can easily show that vectors such as $\begin{pmatrix} 1 \\ i \end{pmatrix}$ are eigenvectors with eigenvalue i, while $\begin{pmatrix} 1 \\ i \end{pmatrix}$ is an eigenvector with eigenvalue $-i$. Thus it is sometimes advantageous to introduce complex numbers into a problem which seems purely concerned with real numbers, because it can introduce such useful concepts as eigenvectors and eigenvalues into the situation. (An example of this appears in Q10 of this week's assignment).

Math 115A - Week 8
Textbook sections: 5.2, 6.1
Topics covered:

- Characteristic polynomials
- Tests for diagonalizability
- Inner products
- Inner products and length

* * * * *

Characteristic polynomials

- Let A be an $n \times n$ matrix. Last week we introduced the *characteristic polynomial*
$$f(\lambda) = \det(A - \lambda I)$$
of that matrix; this is a polynomial in λ. For instance, if
$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$
then
$$\begin{aligned} f(\lambda) &= \det \begin{pmatrix} a-\lambda & b & c \\ d & e-\lambda & f \\ g & h & i-\lambda \end{pmatrix} \\ &= (a-\lambda)((e-\lambda)(i-\lambda) - fh)) - b(d(i-\lambda) - gf) + c(dh - (e-\lambda)g), \end{aligned}$$
which simplifies to some degree 3 polynomial in λ (we think of a, b, c, d, e, f, g as just constant scalars). Last week we saw that the zeroes of this polynomial give the eigenvalues of A.

- As you can see, the characteristic polynomial looks pretty messy. But in the special case of a diagonal matrix, e.g

$$A = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

the characteristic polynomial is quite simple, in fact

$$f(\lambda) = (a - \lambda)(b - \lambda)(c - \lambda)$$

(why?). This has zeroes when $\lambda = a, b, c$, and so the eigenvalues of this matrix are a, b, and c.

- **Lemma 1.** Let A and B be similar matrices. Then A and B have the same characteristic polynomial.

- An algebraist would phrase this as: "the characteristic polynomial is invariant under similarity".

- **Proof.** Since A and B are similar, we have $B = QAQ^{-1}$ for some invertible matrix Q. So the characteristic polynomial of B is

$$\begin{aligned} \det(B - \lambda I) &= \det(QAQ^{-1} - \lambda I) \\ &= \det(QAQ^{-1} - Q\lambda IQ^{-1}) \\ &= \det(Q(A - \lambda I)Q^{-1}) \\ &= \det(Q)\det(A - \lambda I)\det(Q^{-1}) \\ &= \det(Q)\det(Q^{-1})\det(A - \lambda I) \\ &= \det(QQ^{-1}(A - \lambda I)) \\ &= \det(A - \lambda I) \end{aligned}$$

and hence the characteristic polynomials are the same. □

- Now let's try to understand the characteristic polynomial for general matrices. Let $P_1(\mathbf{R})$ be all the polynomials $a\lambda + b$ of degree at most 1; we shall make the free variable λ instead of x. Note that all the entries in the matrix $A - \lambda I$ lie in $P_1(\mathbf{R})$.

- **Lemma 2.** Let B be an $n \times n$ matrix, all of whose entries lie in $P_1(\mathbf{R})$. Then $\det(B)$ lies in $P_n(\mathbf{R})$ (i.e. $\det(B)$ is a polynomial in λ of degree at most n).

- **Proof.** We prove this by induction on n. When $n = 1$ the claim is trivial, since a 1×1 matrix with an entry in $P_1(\mathbf{R})$ looks like $B = (a\lambda + b)$, and clearly $\det(B) = a\lambda + b \in P_n(\mathbf{R})$.

  Now let's suppose inductively that $n > 1$, and that we have already proved the lemma for $n - 1$. We expand $\det(B)$ using cofactor expansion along some row or column (it doesn't really matter which row or column we use). This expands $\det(B)$ as an (alternating-sign) sum of expressions, each of which is the product of an entry of B, and a cofactor of B. The entry of B is in $P_1(\mathbf{R})$, while the cofactor of B is in $P_{n-1}(\mathbf{R})$ by the induction hypothesis. So each term in $\det(B)$ is in $P_n(\mathbf{R})$, and so $\det(B)$ is also in $P_n(\mathbf{R})$. This finishes the induction. $\square$

- From this lemma we see that $f(\lambda)$ lies in $P_n(\mathbf{R})$, i.e it is a polynomial of degree at most n. But we can be more precise. In fact the characteristic polynomial in general looks a lot like the characteristic polynomial of a diagonal matrix, except for an error which is a polynomial of degree at most $n - 2$:

- **Lemma 3.** Let $n \geq 2$. Let A be the $n \times n$ matrix

$$A := \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix}.$$

  Then we have

$$f(\lambda) = (A_{11} - \lambda)(A_{22} - \lambda) \ldots (A_{nn} - \lambda) + g(\lambda)$$

  where $g(\lambda) \in P_{n-2}(\mathbf{R})$.

- **Proof.** Again we induct on n. If $n = 2$ then $f(\lambda) = (A_{11}-\lambda)(A_{22}-\lambda)-A_{12}A_{21}$ (why?) and so the claim is true with $g := -A_{12}A_{21} \in P_0(\mathbf{R})$. Now suppose inductively that $n > 2$, and the claim has already been proven for $n - 1$. We write out $f(\lambda)$ as

$$f(\lambda) = \begin{pmatrix} A_{11} - \lambda & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} - \lambda & \ldots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} - \lambda \end{pmatrix}.$$

- Now we do cofactor expansion along the first row. The first term in this expansion is

$$(A_{11} - \lambda) \det \begin{pmatrix} A_{22} - \lambda & \dots & A_{2n} \\ \vdots & \vdots & \vdots \\ A_{n2} & \dots & A_{nn} - \lambda \end{pmatrix}.$$

  But this determinant is just the characteristic polynomial of an $n-1 \times n-1$ matrix, and so by the induction hypothesis we have

$$\det \begin{pmatrix} A_{22} - \lambda & \dots & A_{2n} \\ \vdots & \vdots & \vdots \\ A_{n2} & \dots & A_{nn} - \lambda \end{pmatrix} = (A_{22}-\lambda) \dots (A_{nn}-\lambda) + \text{ something in } P_{n-3}(\mathbf{R}).$$

  Thus the first term in the cofactor expansion is

$$(A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda) + \text{ something in } P_{n-2}(\mathbf{R}).$$

  (Why did the $P_{n-3}(\mathbf{R})$ become a $P_{n-2}(\mathbf{R})$ when multiplying by $(A_{11} - \lambda)$?).

- Now let's look at the second term in the cofactor expansion; this is

$$-A_{12} \det \begin{pmatrix} A_{21} & \dots & A_{2n} \\ \vdots & \vdots & \vdots \\ A_{n1} & \dots & A_{nn} - \lambda \end{pmatrix}.$$

  We do cofactor expansion again on the second row of this $n-1 \times n-1$ determinant. We can expand this determinant as an alternating-sign sum of terms, which look like A_{2i} times some $n-2 \times n-2$ determinant. By Lemma 2, this $n-2 \times n-2$ determinant lies in $P_{n-2}(\mathbf{R})$, while A_{2i} is a scalar. Thus all the terms in this determinant lie in $P_{n-2}(\mathbf{R})$, and so the determinant itself must lie in $P_{n-2}(\mathbf{R})$ (recall that $P_{n-2}(\mathbf{R})$ is closed under addition and scalar multiplication). Thus this second term in the cofactor expansion lies in $P_{n-2}(\mathbf{R})$.

- A similar argument shows that the third, fourth, etc. terms in the cofactor expansion of $\det(A - \lambda I)$ all lie in $P_{n-2}(\mathbf{R})$. Adding up all these terms we obtain

$$\det(A - \lambda I) = (A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda) + \text{ something in } P_{n-2}(\mathbf{R})$$

as desired. $\square$

- If we multiply out

$$(A_{11} - \lambda)(A_{22} - \lambda) \dots (A_{nn} - \lambda)$$

we get

$$(-\lambda)^n + (-\lambda)^{n-1}(A_{11} + A_{22} + \dots + A_{nn}) + \text{ stuff of degree at most } n-2$$

(why?). Note that $(A_{11} + \dots + A_{nn})$ is just the trace tr(A) of A. Thus from Lemma 3 we have

$$f(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \mathrm{tr(A)} \lambda^{\mathrm{n}-1} + \mathrm{a}_{\mathrm{n}-2} \lambda^{\mathrm{n}-2} + \mathrm{a}_{\mathrm{n}-3} \lambda^{\mathrm{n}-3} + \dots + \mathrm{a}_1 \lambda + \mathrm{a}_0$$

for some scalars $a_{n-2}, \dots, a_0$. These coefficients $a_{n-2}, \dots, a_0$ are quite interesting, but hard to compute. However, a_0 can be obtained by a simple trick: if we evaluate the above expression at 0, we get

$$f(0) = a_0,$$

but $f(0) = \det(A - 0I) = \det(A)$. We have thus proved the following result.

- **Theorem 4.** The characteristic polynomial $f(\lambda)$ of an $n \times n$ matrix A has the form

$$f(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \mathrm{tr(A)} \lambda^{\mathrm{n}-1} + \mathrm{a}_{\mathrm{n}-2} \lambda^{\mathrm{n}-2} + \mathrm{a}_{\mathrm{n}-3} \lambda^{\mathrm{n}-3} + \dots + \mathrm{a}_1 \lambda + \mathrm{det(A)}.$$

- Thus the characteristic polynomial encodes the trace and the determinant, as well as some additional information which we will not study further in this course.

- **Example.** The characteristic polynomial of the 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is

$$(a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + (ad - bc)$$

(why?). Note that $a + d$ is the trace and $ad - bc$ is the determinant.

- **Example.** The characteristic polynomial of the 3×3 matrix

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

is

$$(a-\lambda)(b-\lambda)(c-\lambda) = -\lambda^3 + (a+b+c)\lambda^2 - (ab+bc+ca)\lambda + abc.$$

Note that $a+b+c$ is the trace and abc is the determinant.

- Since the characteristic polynomial is of degree n and has a leading coefficient of -1, it is possible that it factors into n linear factors, i.e.

$$f(\lambda) = -(\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_n)$$

for some scalars $\lambda_1, \ldots, \lambda_n$ in the field of scalars (which we will call F for a change... this F may be either $\mathbf{R}$ or $\mathbf{C}$). These scalars do not necessarily have to be distinct (i.e. we can have releated roots). If this is the case we say that f *splits over* F, or more simply that f *splits*.

- **Example.** The characteristic polynomial of the 2×2 matrix

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

is $\lambda^2 - 1$ (why?), which splits over the reals as $(\lambda - 1)(\lambda - (-1))$. It also splits over the complex numbers because $+1$ and -1 are real numbers, and hence also complex numbers. On the other hand, the characteristic polynomial of

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

is $\lambda^2 + 1$, which doesn't split over the reals, but does split over the complexes as $(\lambda - i)(\lambda + i)$. Finally, the characteristic polynomial of

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

is λ^2, which splits over both the reals and the complexes as $(\lambda - 0)(\lambda - 0)$.

- **Example** The characteristic polynomial of a diagonal matrix will always split. For instance the characteristic polynomial of

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

is

$$(a-\lambda)(b-\lambda)(c-\lambda) = -(\lambda - a)(\lambda - b)(\lambda - c).$$

- From the previous example, and Lemma 1, we see that the characteristic polynomial of any diagonalizable matrix will always split (since diagonalizable matrices are similar to diagonal matrices). In particular, if the characteristic polynomial of a matrix *doesn't* split, then it can't be diagonalizable.

- **Example.** The matrix
$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
from an earlier example cannot be diagonalizable over the reals, because its characteristic polynomial does not split over the reals. (However, it can be diagonalized over the complex numbers; we leave this as an exercise).

- It turns out that the complex numbers have a significant advantage over the reals, in that polynomials *always* split:

- **Fundamental Theorem of Algebra.** Every polynomial splits over the complex numbers.

- This theorem is a basic reason why the complex numbers are so useful; unfortunately, the proof of this theorem is far beyond the scope of this course. (You can see a proof in Math 132, however).

* * * * *

Tests for diagonalizability

- Recall that an $n \times n$ matrix A is *diagonalizable* if there is an invertible matrix Q and a diagonal matrix D such that $A = QDQ^{-1}$. It is often useful to know when a matrix can be diagonalized. We already know one such characterization: A is diagonalizable if and only if there is a basis of $\mathbf{R}^n$ which consists entirely of eigenvectors of A. Equivalently:

- **Lemma 5.** An $n \times n$ matrix A is diagonalizable if and only if one can find n linearly independent vectors $v_1, v_2, \ldots, v_n$ in $\mathbf{R}^n$, such that each vector v_j is an eigenvector of A.

- This is because n linearly independent vectors in $\mathbf{R}^n$ automatically form a basis of $\mathbf{R}^n$.

- It is thus important to know when the eigenvectors of A are linearly independent. Here is one useful test:

- **Proposition 6.** Let A be an $n \times n$ matrix. Let $v_1, v_2, \ldots, v_k$ be eigenvectors of A with eigenvalues $\lambda_1, \ldots, \lambda_k$ respectively. Suppose that the eigenvalues $\lambda_1, \ldots, \lambda_k$ are all *distinct*. Then the vectors $v_1, \ldots, v_k$ are linearly independent.

- **Proof.** Suppose for contradiction that $v_1, \ldots, v_k$ were not independent, i.e. there was some scalars $a_1, \ldots, a_k$, not all equal to zero, such that

$$a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = 0.$$

  At least one of the a_j is non-zero; without loss of generality we may assume that a_1 is non-zero.

- Now we use a trick to eliminate v_k: We apply $(A - \lambda_k I)$ to both sides of this equation. Using the fact that $A - \lambda_k I$ is linear, we obtain

$$a_1(A - \lambda_k I)v_1 + a_2(A - \lambda_k I)v_2 + \ldots + a_k(A - \lambda_k I)v_k = 0.$$

  But observe that

$$(A - \lambda_k I)v_1 = Av_1 - \lambda_k v_1 = \lambda_1 v_1 - \lambda_k v_1 = (\lambda_1 - \lambda_k)v_1$$

  and more generally

$$(A - \lambda_k I)v_j = (\lambda_j - \lambda_k)v_j.$$

In particular we have
$$(A - \lambda_k I)v_k = 0.$$
Putting this all together, we obtain

$$a_1(\lambda_1 - \lambda_k)v_1 + a_2(\lambda_2 - \lambda_k)v_2 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0.$$

Now we eliminate v_{k-1} by applying $A - \lambda_{k-1}I$ to both sides of the equation. Arguing as before, we obtain

$$a_1(\lambda_1 - \lambda_k)(\lambda_1 - \lambda_{k-1})v_1 + a_2(\lambda_2 - \lambda_k)(\lambda_2 - \lambda_{k-1})v_2 + \ldots$$

$$+a_{k-2}(\lambda_{k-2} - \lambda_k)(\lambda_{k-2} - \lambda_{k-1})v_{k-2} = 0.$$

We then eliminate v_{k-2}, then v_{k-3}, and so forth all the way down to eliminating v_2, until we obtain

$$a_1(\lambda_1 - \lambda_k)(\lambda_1 - \lambda_{k-1}) \ldots (\lambda_1 - \lambda_2)v_1 = 0.$$

But since the λ_i are all distinct, and a_1 is non-zero, this forces v_1 to equal zero. But this contradicts the definition of eigenvector (eigenvectors are not allowed to be zero). Thus the vectors $v_1, \ldots, v_k$ must have been linearly independent. $\square$

- Proposition 5 holds for linear transformations as well as matrices: see Theorem 5.10 of the textbook.

- **Corollary 6** Let A be an $n \times n$ matrix. If the characteristic polynomial of A splits into n *distinct* factors, then A is diagonalizable.

- **Proof.** By assumption, the characteristic polynomial $f(\lambda)$ splits as

$$f(\lambda) = -(\lambda - \lambda_1) \ldots (\lambda - \lambda_n)$$

for some distinct scalars $\lambda_1, \ldots, \lambda_n$. Thus we have n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. For each eigenvalue λ_j let v_j be an eigenvector with that eigenvalue, then by Proposition 5 $v_1, \ldots, v_n$ are linearly independent, and hence by Lemma 4 A is diagonalizable. $\square$

- **Example** Consider the matrix

$$A = \begin{pmatrix} 1 & -2 \\ 1 & 4 \end{pmatrix}.$$

The characteristic polynomial here is

$$f(\lambda) = (1-\lambda)(4-\lambda) + 2 = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3),$$

so the characteristic polynomial splits into n distinct factors (regardless of whether our scalar field is the reals or the complexes). So we know that A is diagonalizable. (If we actually wanted the explicit diagonalization, we would find the eigenvalues (which are 2,3) and then some eigenvectors, and use the previous week's notes).

- To summarize what we know so far: if the characteristic polynomial doesn't split, then we can't diagonalize the matrix; while if it does split into distinct factors, then we can diagonalize the matrix. There is still a remaining case in which the characteristic function splits, but into repeated factors. Unfortunately this case is much more complicated; the matrix may or may not be diagonalizable. For instance, the matrix

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

has a characteristic polynomial of $(\lambda - 2)^2$ (why?), so it splits but not into distinct linear factors. It is clearly diagonalizable (indeed, it is diagonal). On the other hand, the matrix

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

has the same characteristic polynomial of $(\lambda - 2)^2$ (why?), but it turns out not to be diagonalizable, for the following reason. If it were diagonalizable, then we could find a basis of $\mathbf{R}^n$ which consists entirely of eigenvectors. But since the only root of the characteristic polynomial is 2, the only eigenvalue is 2. Now let's work out what the eigenvectors are. Since the only eigenvalue is 2, we only need to look in the eigenspace with eigenvalue 2. We have to solve the equation

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2 \begin{pmatrix} x \\ y \end{pmatrix},$$

i.e. we have to solve the system of equations

$$2x + y = 2x; \quad 2y = 2y.$$

The general solution of this system occurs when $y = 0$ and x is arbitrary, so the eigenspace with eigenvalue 2 is just the x-axis. But the vectors from this eigenspace are not enough to span all of $\mathbf{R}^2$, so we cannot find a basis of eigenvectors. Thus this matrix is not diagonalizable.

- The moral of this story is that, while the characteristic polynomial does carry a large amount of information, it does not completely solve the problem of whether a matrix is diagonalizable or not. However, even when the characteristic polynomial is inconclusive, it is still possible to determine whether a matrix is diagonalizable or not by computing its eigenspaces and seeing if it is possible to make a basis consisting entirely of eigenvectors. We will not pursue the full solution of the diagonalization problem here, but defer it to 115B (where you will learn about two more tools to study diagonalization - the minimal polynomial and the Jordan normal form).

- **One last example.** Consider the matrix

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

this is the same matrix as the previous example but we attach another row and column, and add a 3. (This is not a diagonal matrix, but is an example of a *block-diagonal matrix*: see this week's homework for more information). The characteristic polynomial here is

$$f(\lambda) = -(\lambda - 2)^2(\lambda - 3)$$

(why?), so the eigenvalues are 2 and 3. To find the eigenspace with eigenvalue 2, we solve the equation

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 2 \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

and a little bit of work shows that the general solution to this equation occurs when $y = z = 0$ and x is arbitrary, thus the eigenspace is just the x-axis $\{\begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} : x \in \mathbf{R}\}$. Similarly the eigenspace with eigenvalue 3 is the z-axis. But this is not enough eigenvectors to span $\mathbf{R}^3$ (the 2-eigenspace only contributes one linearly independent eigenvector, and the 3-eigenspace contributes only one linearly independent eigenvector, whereas we need three linearly independent eigenvectors in order to span $\mathbf{R}^3$.

* * * * *

Inner product spaces

- We now leave matrices and eigenvalues and eigenvectors for the time being, and begin a very different topic - the concept of an inner product space.

- Up until now, we have been preoccupied with *vector spaces* and various things that we can do with these vector spaces. If you recall, a vector space comes equipped with only two basic operations: addition and scalar multiplication. These operations have already allowed us to introduce many more concepts (bases, linear transformations, etc.) but they cannot do everything that one would like to do in applications.

- For instance, how does one compute the length of a vector? In $\mathbf{R}^2$ or $\mathbf{R}^3$ one can use Pythagoras's theorem to work out the length, but what about, say, a vector in $P_3(\mathbf{R})$? What is the length of $x^3 + 3x^2 + 6$? It turns out that such spaces do not have an inherent notion of length: you can add and scalar multiply two polynomials, but we have not given any rule to determine the length of a polynomial. Thus, vector spaces are not equipped to handle certain geometric notions such as length (or angle, or orthogonality, etc.)

- To resolve this, mathematicians have introduced several "upgraded" versions of vector spaces, in which you can not only add and scalar multiply vectors, but can also compute lengths, angles, inner products, etc. One particularly common such "upgraded" vector space is something called an *inner product space*, which we will now discuss. (There are also *normed vector spaces*, which have a notion of length but not angle; *topological vector spaces*, which have a notion of convergence but not length; and if you wander into infinite dimensions then there are slightly fancier things such as *Hilbert spaces* and *Banach spaces*. Then there are *vector algebras*, where you can multiply vectors with other vectors to get more vectors. Then there are hybrids of these notions, such as *Banach algebras*, which are a certain type of infinite-dimensional vector algebra. None of these will be covered in this course; they are mostly graduate level topics).

- The problem with length is that it is not particularly linear: the length of a vector $v+w$ is not just the length of v plus the length of w. However, in $\mathbf{R}^2$ or $\mathbf{R}^3$ we can rewrite the length of a vector v as the square root of the *dot product* $v \cdot v$. Unlike length, the dot product is linear in the sense that $(v + v') \cdot w = v \cdot w + v' \cdot w$ and $v \cdot (w + w') = v \cdot w + v \cdot w'$, with a similar rule for scalar multiplication. (Actually, to be precise, the dot product is considered *bilinear* rather than linear, just as the determinant is considered multilinear, because it has two inputs v and w, instead of just one for linear transformations).

- Thus, the idea behind an inner product space is to introduce length indirectly, by means of something called an *inner product*, which is a generalization of the dot product. Depending on whether the field of scalars is real or complex, we have either a real inner product space or a complex inner product space. Complex inner product spaces are similar to real ones, except the complex conjugate operation $z \mapsto \overline{z}$ makes an appearance. Here's a clue why: the length $|z|$ of a complex number $z = a + bi$, is not the square root of $z \cdot z$, but is instead the square root of $z \cdot \overline{z}$.

- We will now use both real and complex vector spaces, and will try to take care to distinguish between the two. When we just say "vector space" without the modifier "real" or "complex", then the field of scalars might be either the reals or the complex numbers.

- **Definition** An *inner product space* is a vector space V equipped with an additional operation, called an *inner product*, which takes two vectors

$v, w \in V$ as input and returns a scalar $\langle v, w\rangle$ as output, which obeys the following three properties:

- (Linearity in the first variable) For any vectors $v, v', w \in V$ and any scalar c, we have $\langle v + v', w\rangle = \langle v, w\rangle + \langle v', w\rangle$ and $\langle cv, w\rangle = c\langle v, w\rangle$.
- (Conjugate symmetry) If v and w are vectors in V, then $\langle w, v\rangle$ is the complex conjugate of $\langle v, w\rangle$: $\langle w, v\rangle = \overline{\langle v, w\rangle}$.
- (Positivity) If v is a non-zero vector in V, then $\langle v, v\rangle$ is a positive real number: $\langle v, v\rangle > 0$.
- If the field of scalars is real, then every number is its own conjugate (e.g. $\overline{3} = 3$) and so the conjugate-symmetry property simplifies to just the symmetry property $\langle w, v\rangle = \langle v, w\rangle$.
- We now give some examples of inner product spaces.
- $\mathbf{R}^n$ **as an inner product space.** We already know that $\mathbf{R}^n$ is a real vector space. If we now equip $\mathbf{R}^n$ with the inner product equal to the *dot product*
$$\langle x, y\rangle := x \cdot y$$
i.e.
$$\langle (x_1, x_2, \ldots, x_n), (y_1, \ldots, y_n)\rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n = \sum_{j=1}^{n} x_jy_j$$
then we obtain an inner product space. For instance, we now have $\langle (1, 2), (3, 4)\rangle = 11$.
- To verify that we have an inner product space, we have to verify the linearity property, conjugate symmetry property, and the positivity property. To verify the linearity property, observe that
$$\langle x + x', y\rangle = (x + x') \cdot y = x \cdot y + x' \cdot y = \langle x, y\rangle + \langle x', y\rangle$$
and
$$\langle cx, y\rangle = (cx) \cdot y = c(x \cdot y) = c\langle x, y\rangle$$

while the conjugate symmetry follows since

$$\langle y, x \rangle = y \cdot x = x \cdot y = \langle x, y \rangle = \overline{\langle x, y \rangle}$$

(since the conjugate of a real number is itself. To verify the positivity property, observe that

$$\langle (x_1, x_2, \ldots, x_n), (x_1, \ldots, x_n) \rangle = x_1^2 + x_2^2 + \ldots + x_n^2$$

which is clearly positive if the $x_1, \ldots, x_n$ are not all zero.

- The difference between the dot product and the inner product is that the dot product is specific to $\mathbf{R}^n$, while the inner product is a more general concept and is applied to many other vector spaces.

- One can interpret the dot product $x \cdot y$ as measuring the amount of "correlation" or "interaction" between x and y; the longer that x and y are, and the more that they point in the same direction, the larger the dot product becomes. If x and y point in opposing directions then the dot product is negative, while if x and y point at right angles then the dot product is zero. Thus the dot product combines both the length of the vectors, and their angle (as can be seen by the famous formula $x \cdot y = |x||y| \cos \theta$ but easier to work with than either length or angle because it is (bi-)linear (while length and angle individually are not linear quantities).

- **$\mathbf{R}^n$ as an inner product space II.** One doesn't have to use the dot product as the inner product; other dot products are possible. For instance, one could endow $\mathbf{R}^n$ with the non-standard inner product

$$\langle x, y \rangle' := 10 x \cdot y,$$

so for instance $\langle (1, 2), (3, 4) \rangle' = 110$. While this is not the standard inner product, it still obeys the three properties of linearity, conjugate symmetry, and positivity (why?), so this is still an inner product (though to avoid confusion we have labeled it as $\langle , \rangle'$ instead of $\langle , \rangle$. The situation here is similar to bases of vector spaces; a vector space such as $\mathbf{R}^n$ can have a standard basis but also have several non-standard bases (for instance, we could multiply every vector in the standard basis by 10), and the same is often true of inner products. However in the vast majority of cases we will use a standard inner product.

- More generally, we can multiply any inner product by a positive constant, and still have an inner product.

- **R as an inner product space.** A special case of the previous example of $\mathbf{R}^n$ with the standard inner product occurs when $n = 1$. Then our inner product space is just the real numbers, and the inner product is given by the ordinary product: $\langle x, y\rangle := xy$. For instance $\langle 2, 3\rangle = 6$. Thus, plain old multiplication is itself an example of an inner product space.

- **C as an inner product space.** Now let's look at the complex numbers $\mathbf{C}$, which is a one-dimensional *complex* vector space (so the field of scalars is now $\mathbf{C}$). Here, we could reason by analogy with the previous example and guess that $\langle z, w\rangle := zw$ would be an inner product, but this does not obey either the conjugate-symmetry property or the positivity property: if z were a complex number, then $\langle z, z\rangle = z^2$ would not necessarily be a positive real number (or even a real number); for instance $\langle i, i\rangle = -1$.

- To fix this, the correct way to define an inner product on $\mathbf{C}$ is to set $\langle z, w\rangle := z\overline{w}$; in other words we have to conjugate the second factor. This inner product is now linear in the first variable (why?) and conjugate-symmetric (why?). To verify positivity, observe that $\langle a + bi, a + bi\rangle = (a + bi)(a - bi) = a^2 + b^2$ which will be a positive real number if $a + bi$ is non-zero.

- **$\mathbf{C}^n$ as an inner product space.** Now let's look at the complex vector space
$$\mathbf{C}^n := \{(z_1, z_2, \ldots, z_n) : z_1, z_2, \ldots, z_n \in \mathbf{C}\}.$$
This is just like $\mathbf{R}^n$ but with the scalars being complex instead of real; for instance $(3, 1 + i, 3i)$ would lie in $\mathbf{C}^3$ but it wouldn't be a vector in $\mathbf{R}^3$. We can define an inner product here by
$$\langle (z_1, z_2, \ldots, z_n), (w_1, w_2, \ldots, w_n)\rangle := z_1\overline{w_1} + z_2\overline{w_2} + \ldots + z_n\overline{w_n};$$
note how this definition is a hybrid of the $\mathbf{R}^n$ inner product and the $\mathbf{C}$ inner product. This is an inner product space (why?).

- **Functions as an inner product space.** Consider $C([0,1];\mathbf{R})$, the space of continuous real-valued functions from the interval $[0,1]$ to $\mathbf{R}$. This is an infinite-dimensional real vector space, containing such functions as $\sin(x)$, x^2+3, $1/(x+1)$, and so forth. We can define an inner product on this space by defining

$$\langle f, g \rangle := \int_0^1 f(x)g(x)\ dx;$$

for instance,

$$\langle x+1, x^2 \rangle = \int_0^1 (x+1)x^2\ dx = (\frac{x^4}{4} + \frac{x^3}{3})|_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}.$$

Note that we need the continuity property in order to make sure that this integral actually makes sense (as opposed to diverging to infinity or otherwise doing something peculiar). One can verify fairly easily that this is an inner product space; we just give parts of this verification. One of the things we have to show is that

$$\langle f_1 + f_2, g \rangle = \langle f_1, g \rangle + \langle f_2, g \rangle,$$

but this follows since the left-hand side is

$$\int_0^1 (f_1(x)+f_2(x))g(x)\ dx = \int_0^1 f_1(x)g(x)\ dx + \int_0^1 f_2(x)g(x)\ dx = \langle f_1, g \rangle + \langle f_2, g \rangle.$$

To verify positivity, observe that

$$\langle f, f \rangle = \int_0^1 f(x)^2\ dx.$$

The function $f(x)^2$ is always non-negative, and if f is not the zero function on $[0,1]$, then $f(x)^2$ must be strictly positive for some $x \in [0,1]$. Thus there is a strictly positive area under the graph of $f(x)^2$, and so $\int_0^1 f(x)^2\ dx > 0$.

- One can view this example as an infinite-dimensional version of the finite-dimensional inner product space example of $\mathbf{R}^n$. To see this, let N be a very large number. Remembering that integrals can be approximated by Riemann sums, we have (non-rigorously) that

$$\int_0^1 f(x)g(x)\, dx \approx \sum_{j=1}^N f(\frac{j}{N})g(\frac{j}{N})\frac{1}{N},$$

or in other words

$$\langle f, g\rangle \approx \frac{1}{N}(f(\frac{1}{N}), f(\frac{2}{N}), \ldots, f(\frac{N}{N})) \cdot (g(\frac{1}{N}), g(\frac{2}{N}), \ldots, g(\frac{N}{N})),$$

and the right-hand side resembles an example of the inner product on $\mathbf{R}^N$ (admittedly there is an additional factor of $\frac{1}{N}$, but as observed before, putting a constant factor in the definition of an inner product just gives you another inner product.

- **Functions as an inner product space II.** Consider $C([-1,1];\mathbf{R})$, the space of continuous real-valued functions on $[-1,1]$. Here we can define an inner product as

  $$\langle f, g\rangle := \int_{-1}^1 f(x)g(x)\, dx.$$

  Thus for instance

  $$\langle x+1, x^2\rangle = \int_{-1}^1 (x+1)x^2\, dx = (\frac{x^4}{4} + \frac{x^3}{3})|_{-1}^1 = \frac{2}{3}.$$

  Note that this inner product of $x+1$ and x^2 was different from the inner product of $x+1$ and x^2 given in the previous example! Thus it is important, when dealing with functions, to know exactly what the domain of the functions is, and when dealing with inner products, to know exactly which inner product one is using - confusing one inner product for another can lead to the wrong answer! To avoid confusion, one sometimes labels the inner product with some appropriate subscript, for instance the inner product here might be labeled $\langle, \rangle_{C([-1,1];\mathbf{R})}$ and the previous one labeled $\langle, \rangle_{C([0,1];\mathbf{R})}$.

- **Functions as an inner product space III.** Now consider $C([0,1];\mathbf{C})$, the space of continuous complex-valued functions from the interval $[0,1]$

to $\mathbf{C}$; this includes such functions as $\sin(x)$, $x^2 + ix - 3 + i$, $i/(x - i)$, and so forth. Note that while the range of this function is complex, the domain is still real, so x is still a real number. This is a infinite-dimensional complex vector space (why?). We can define an inner product on this space as

$$\langle f, g \rangle := \int_0^1 f(x)\overline{g(x)}\ dx.$$

Thus, for instance

$$\langle x^2, x + i \rangle = \int_0^1 x^2(x - i)\ dx = (\frac{x^4}{4} - \frac{ix^3}{3})|_0^1 = \frac{1}{4} - \frac{i}{3}.$$

This can be easily verified to be an inner product space. For the positivity, observe that

$$\langle f, f \rangle = \int_0^1 f(x)\overline{f(x)}\ dx = \int_0^1 |f(x)|^2\ dx.$$

Even though $f(x)$ can be any complex number, $|f(x)|^2$ must be a non-negative real number, and an argument similar to that for real functions shows that $\int_0^1 |f(x)|^2\ dx$ is a positive real number when f is not the zero function on $[0, 1]$.

- **Polynomials as an inner product space.** The inner products in the above three examples work on functions. Since polynomials are a special instance of functions, the above inner products are also inner products on polynomials. Thus for instance we can give $P_3(\mathbf{R})$ the inner product

$$\langle f, g \rangle := \int_0^1 f(x)g(x)\ dx,$$

so that for instance

$$\langle x, x^2 \rangle = \int_0^1 x^3\ dx = \frac{x^4}{4}|_0^1 = \frac{1}{4}.$$

Or we could instead give $P_3(\mathbf{R})$ a different inner product

$$\langle f, g \rangle := \int_{-1}^1 f(x)g(x)\ dx,$$

so that for instance

$$\langle x, x^2 \rangle = \int_{-1}^{1} x^3 \, dx = \frac{x^4}{4}|_{-1}^{1} = 0.$$

Unfortunately, we have here a large variety of inner products and it is not obvious what the "standard" inner product should be. Thus whenever we deal with polynomials as an inner product space we shall be careful to specify exactly which inner product we will use. However, we can draw one lesson from this example, which is that if V is an inner product space, then any subspace W of V is also an inner product space. (Note that if the properties of linearity, conjugate symmetry, and positivity hold for the larger space V, then they will automatically hold for the smaller space W (why?)).

- **Matrices as an inner product space.** Let $M_{m\times n}(\mathbf{R})$ be the space of real matrices with m rows and n columns; a typical element is

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix}.$$

- This is a mn-dimensional real vector space. We can turn this into an inner product space by defining the inner product

$$\langle A, B \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij};$$

i.e. for every row and column we multiply the corresponding entries of A and B together, and then sum. For instance,

$$\langle \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \rangle = 1 \times 5 + 2 \times 6 + 3 \times 7 + 4 \times 8 = 70.$$

It is easy to verify that this is also an inner product; note how similar this is to the standard $\mathbf{R}^n$ inner product. This inner product can also be written using transposes and traces:

$$\langle A, B \rangle = \mathrm{tr(AB^t)}.$$

To see this, note that AB^t is an $m \times m$ matrix, with the top left entry being $A_{11}B_{11} + A_{12}B_{12} + \ldots + A_{1n}B_{1n}$, the second entry on the diagonal being $A_{21}B_{21} + A_{22}B_{22} + \ldots + A_{2n}B_{2n}$, and so forth down the diagonal (why?). Adding these together we obtain $\langle A, B \rangle$. For instance,

$$\mathrm{tr}(\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}^{\mathrm{t}}) = \mathrm{tr} \begin{pmatrix} 1 \times 5 + 2 \times 6 & 1 \times 7 + 2 \times 8 \\ 3 \times 5 + 4 \times 5 & 3 \times 7 + 4 \times 8 \end{pmatrix}$$

$$= 1 \times 5 + 2 \times 6 + 3 \times 7 + 4 \times 8 = \langle \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \rangle.$$

- **Matrices as an inner product space II.** Let $M_{m \times n}(\mathbf{C})$ be the space of complex matrices with m rows and n columns; this is an mn-dimensional complex vector space. This is an inner product space with inner product
$$\langle A, B \rangle := \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} \overline{B_{ij}}.$$

- It is not hard to verify that this is indeed an inner product space. Unlike the previous example, the inner product is not given by the formula $\langle A, B \rangle = \mathrm{tr}(\mathrm{AB}^{\mathrm{t}})$, however there is a very similar formula. Define the *adjoint* $B^\dagger$ of a matrix to be the complex conjugate of the transpose B^t; i.e. $B^\dagger$ is the same matrix as B^t but with every entry replaced by its complex conjugate. For instance,

$$\begin{pmatrix} 1 + 2i & 3 + 4i \\ 5 + 6i & 7 + 8i \\ 9 + 10i & 11 + 12i \end{pmatrix}^\dagger = \begin{pmatrix} 1 - 2i & 5 - 6i & 9 - 10i \\ 3 - 4i & 7 - 8i & 11 - 12i \end{pmatrix}.$$

  The adjoint is the complex version of the transpose; it is completely unrelated to the adjugate matrix in the previous week's notes. It is easy to verify that $\langle A, B \rangle = \mathrm{tr}(\mathrm{AB}^\dagger)$.

- (Optional remarks) To summarize: many of the vector spaces we have encountered before, can be upgraded to inner product spaces. As we shall see, the additional capabilities of inner product spaces can be useful in many applications when the more basic capabilities of vector spaces are not enough. On the other hand, inner products add complexity (and possible confusion, if there is more than one choice of inner product available) to a problem, and in many situations (for instance, in the last six weeks of material) they are unnecessary. So it is sometimes better to just deal with bare vector spaces, with no inner product attached; it depends on the situation. (A more subtle reason why sometimes adding extra structure can be bad, is because it reduces the amount of symmetry in a situation; it is relatively easy for a transformation to be *linear* (i.e. it preserves the vector space structures of addition and scalar multiplication) but it is much harder to be *isometric* (which means that it not only preserves addition and scalar multiplication, but also inner products as well.). So if one insists on dealing with inner products all the time, then one loses a lot of symmetries, because there is more structure to preserve, and this can sometimes make a problem appear harder than it actually is. Some of the deepest advances in physics, for instance, particularly in relativity and quantum mechanics, were only possible because the physicists removed a lot of unnecessary structure from their models (e.g. in relativity they removed separate structures for space and time, keeping only something called the spacetime metric), and then gained so much additional symmetry that they could then use those symmetries to discover new laws of physics (e.g. Einstein's law of gravitation).)

- Some basic properties of inner products:

- From the linearity and conjugate symmetry properties it is easy to see that $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$ and $\langle v, cw \rangle = \overline{c} \langle v, w \rangle$ for all vectors v, w, w' and scalars c (why?) Note that when you pull a scalar c out of the second factor, it gets conjugated, so be careful about that. (Another way of saying this is that the inner product is *conjugate linear*, rather than linear, in the second variable. Because the inner product is linear in the first variable and only sort-of-linear in the second, it is sometimes said that the inner product is *sesquilinear* (sesqui is Latin for "one and a half").

- The inner product of 0 with anything is 0: $\langle 0, v \rangle = \langle v, 0 \rangle = 0$. (This is an easy consequence of the linearity (or conjugate linearity) - Why?). In particular, $\langle 0, 0 \rangle = 0$. Thus, by the positivity property, $\langle v, v \rangle$ is positive if v is non-zero and zero if v is zero. In particular, if we ever know that $\langle v, v\rangle = 0$, we can deduce that v itself is 0.

Inner products and length

- Once you have an inner product space, you can then define a notion of length:

- **Definition** Let V be an inner product space, and let v be a vector of V. Then the *length* of v, denoted $\|v\|$, is given by the formula $\|v\| := \sqrt{\langle v, v\rangle}$. (In particular, $\langle v, v\rangle = \|v\|^2$).

- **Example** In $\mathbf{R}^2$ with the standard inner product, the vector $(3, 4)$ has length
$$\|(3,4)\| = \sqrt{\langle (3,4), (3,4)\rangle} = \sqrt{3^2 + 4^2} = 5.$$

- If instead we use the non-standard inner product $\langle x, y\rangle = 10x \cdot y$, then the length is now
$$\|(3,4)\| = \sqrt{\langle (3,4), (3,4)\rangle} = \sqrt{10(3^2 + 4^2)} = 5\sqrt{10}.$$

  Thus the notion of length depends very much on what inner product you choose (although in most cases this will not be an issue since we will use the standard inner product).

- From the positivity property we see that every non-zero vector has a positive length, while the zero vector has zero length. Thus $\|v\| = 0$ if and only if $v = 0$.

- If c is a scalar, then
$$\|cv\| = \sqrt{\langle cv, cv\rangle} = \sqrt{c\bar{c}\langle v, v\rangle} = \sqrt{|c|^2\|v\|^2} = |c|\|v\|.$$

- **Example** In a complex vector space, the vector $(3 + 4i)v$ is five times as long as v. The vector $-v$ has exactly the same length as v.

- The inner product in some special cases can be expressed in terms of length. We already know that $\langle v, v\rangle = \|v\|^2$. More generally, if w is a positive scalar multiple of v (so that v and w are parallel and in the same direction), then $w = cv$ for some positive real number c, and

$$\langle v, w\rangle = \langle v, cv\rangle = c\langle v, v\rangle = c\|v\|^2 = \|v\|\|cv\| = \|v\|\|w\|,$$

i.e. when v and w are pointing in the same direction then the inner product is just the product of the norms. In the other extreme, if w is a negative scalar multiple of v, then $w = -cv$ for some positive c, and

$$\langle v, w\rangle = \langle v, -cv\rangle = -c\langle v, v\rangle = -c\|v\|^2$$

$$= -\|v\|| - c|\|v\| = -\|v\|\| - cv\| = -\|v\|\|w\|,$$

and so the inner product is negative the product of the norms. In general the inner product lies in between these two extremes:

- **Cauchy-Schwarz inequality** Let V be an inner product space. For any $v, w \in V$, we have

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

- **Proof.** If $w = 0$ then both sides are zero, so we can assume that $w \neq 0$. From the positivity property we know that $\langle v, v\rangle \geq 0$. More generally, for any scalars a, b we know that $\langle av + bw, av + bw\rangle \geq 0$. But

$$\begin{aligned}\langle av + bw, av + bw\rangle &= a\langle v, av + bw\rangle + b\langle w, av + bw\rangle \\ &= a\overline{a}\langle v, v\rangle + a\overline{b}\langle v, w\rangle + b\overline{a}\langle w, v\rangle + b\overline{b}\langle w, w\rangle \\ &= |a|^2\|v\|^2 + a\overline{b}\langle v, w\rangle + b\overline{a}\overline{\langle v, w\rangle} + |b|^2\|w\|^2.\end{aligned}$$

Since $\langle av + bw, av + bw\rangle \geq 0$ for any choice of scalars a, b, we thus have

$$|a|^2\|v\|^2 + a\overline{b}\langle v, w\rangle + b\overline{a}\overline{\langle v, w\rangle} + |b|^2\|w\|^2 \geq 0$$

for any choice of scalars a, b. We now select a and b in order to obtain some cancellation. Specifically, we set

$$a := \|w\|^2; b := -\langle v, w\rangle.$$

Then we see that

$$\|w\|^4\|v\|^2 - \|w\|^2\overline{\langle v, w\rangle}\langle v, w\rangle - \langle v, w\rangle\|w\|^2\overline{\langle v, w\rangle} + |\langle v, w\rangle|^2\|w\|^2 \geq 0;$$

this simplifies to

$$\|w\|^4\|v\|^2 \geq \|w\|^2|\langle v, w\rangle|^2.$$

Dividing by $\|w\|^2$ (recall that w is non-zero, so that $\|w\|$ is non-zero) and taking square roots we obtain the desired inequality. $\square$

- Thus for real vector spaces, the inner product $\langle v, w\rangle$ always lies somewhere between $+\|v\|\|w\|$ and $-\|v\|\|w\|$. For complex vector spaces, the inner product $\langle v, w\rangle$ can lie anywhere in the disk centered at the origin with radius $\|v\|\|w\|$. For instance, in the complex vector space $\mathbf{C}$, if $v = 3 + 4i$ and $w = 4 - 3i$ then $\langle v, w\rangle = v\overline{w} = 25i$, while $\|v\| = 5$ and $\|w\| = 5$.

- The Cauchy-Schwarz inequality is extremely useful, especially in analysis; it tells us that if one of the vectors v, w have small length then their inner product will also be small (unless of course the other vector has very large length).

- Another fundamental inequality concerns the relationship between length and vector addition. It is clear that length is not linear: the length of $v + w$ is not just the sum of the length of v and the length of w. For instance, in $\mathbf{R}^2$, if $v := (1, 0)$ and $w = (0, 1)$ then $\|v + w\| = \|(1, 1)\| = \sqrt{2} \neq 1 + 1 = \|v\| + \|w\|$. However, we do have

- **Triangle inequality** Let V be an inner product space. For any $v, w \in V$, we have

$$\|v + w\| \leq \|v\| + \|w\|.$$

- **Proof.** To prove this inequality, we can square both sides (note that this is OK since both sides are non-negative):

$$\|v + w\|^2 \leq (\|v\| + \|w\|)^2.$$

The left-hand side is

$$\langle v + w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle.$$

The quantities $\langle v, v\rangle$ and $\langle w, w\rangle$ are just $\|v\|^2$ and $\|w\|^2$ respectively. From the Cauchy-Schwarz inequality, the two quantities $\langle v, w\rangle$ and $\langle w, v\rangle$ have absolute value at most $\|v\|\|w\|$. Thus

$$\langle v + w, v + w\rangle \leq \|v\|^2 + \|v\|\|w\| + \|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2$$

as desired. $\square$

- The reason this is called the triangle inequality is because it has a natural geometric interpretation: if one calls two sides of a triangle v and w, so that the third side is $v + w$, then the triangle inequality says that the length of the third side is less than or equal to the sum of the lengths of the other two sides. In other words, a straight line has the shortest distance between two points (at least when compared to triangular alternatives).

- The triangle inequality has a couple of variants. Here are a few:

$$\|v - w\| \leq \|v\| + \|w\|$$

$$\|v + w\| \geq \|v\| - \|w\|$$

$$\|v + w\| \geq \|w\| - \|v\|$$

$$\|v - w\| \geq \|v\| - \|w\|$$

$$\|v - w\| \geq \|w\| - \|v\|$$

Thus for instance, if v has length 10, and w has length 3, then both $v+w$ and $v-w$ have length somewhere between 7 and 13. (Can you see this geometrically?). These inequalities can be proven in a similar manner to the original triangle inequality, or alternatively one can start with the original triangle inequality and do some substitutions (e.g. replace w by $-w$, or replace v by $v - w$, on both sides of the inequality; try this!).

Math 115A - Week 9
Textbook sections: 6.1-6.2
Topics covered:

- Orthogonality
- Orthonormal bases
- Gram-Schmidt orthogonalization
- Orthogonal complements

* * * * *

Orthogonality

- From your lower-division vector calculus you know that two vectors v, w in $\mathbf{R}^2$ or $\mathbf{R}^3$ are perpendicular if and only if $v \cdot w = 0$; for instance, $(3, 4)$ and $(-4, 3)$ are perpendicular.
- Now that we have inner products - a generalization of dot products - we can now give a similar notion for all inner product spaces.
- **Definition.** Let V be an inner product space. If v, w are vectors in V, we say that v and w are *orthogonal* if $\langle v, w\rangle = 0$.
- **Example.** In $\mathbf{R}^4$ (with the standard inner product), the vectors $(1, 1, 0, 0)$ and $(0, 0, 1, 1)$ are orthogonal, as are $(1, 1, 1, 1)$ and $(1, -1, 1, -1)$, but the vectors $(1, 1, 0, 0)$ and $(1, 0, 1, 0)$ are not orthogonal. In $\mathbf{C}^2$, the vectors $(1, i)$ and $(1, -i)$ are orthogonal, but $(1, 0)$ and $(i, 0)$ are not.
- **Example.** In any inner product space, the 0 vector is orthogonal to everything (why?). On the other hand, a non-zero vector cannot be orthogonal to itself (why? Recall that $\langle v, v\rangle = \|v\|^2$).
- **Example.** In $C([0, 1]; \mathbf{C})$ with the inner product

$$\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}\, dx,$$

the functions 1 and $x - \frac{1}{2}$ are orthogonal (why?), but 1 and x are not. However, in $C([-1,1];\mathbf{C})$ with the inner product

$$\langle f, g \rangle := \int_{-1}^{1} f(x)\overline{g(x)}\, dx,$$

the functions 1 and $x - \frac{1}{2}$ are no longer orthogonal, however the functions 1 and x now are. Thus the question of whether two vectors are orthogonal depends on which inner product you use.

- Sometimes we say that v and w are *perpendicular* instead of *orthogonal*. This makes the most sense for $\mathbf{R}^n$, but can be a bit confusing when dealing with other inner product spaces such as $C([-1,1],\mathbf{C})$ - how would one visualize the functions 1 and x being "perpendicular", for instance (or i and x, for that matter)? So I prefer to use the word orthogonal when dealing with general inner product spaces.

- Sometimes we write $v \perp w$ to denote the fact that v is orthogonal to w.

- Being orthogonal is at the opposite extreme of being parallel; recall from the Cauchy-Schwarz inequality that $|\langle v, w\rangle|$ must lie between 0 and $\|v\|\|w\|$. When v and w are parallel then $|\langle v, w\rangle|$ attains its maximum possible value of $\|v\|\|w\|$, while when v and w are orthogonal then $|\langle v, w\rangle|$ attains its minimum value of 0.

- Orthogonality is symmetric: if v is orthgonal to w then w is orthogonal to v. (Why? Use the conjugate symmetry property and the fact that the conjugate of 0 is 0).

- Orthogonality is preserved under linear combinations:

- **Lemma 1.** Suppose that $v_1, \ldots, v_n$ are vectors in an inner product space V, and suppose that w is a vector in V which is orthogonal to all of $v_1, v_2, \ldots, v_n$. Then w is also orthogonal to any linear combination of $v_1, \ldots, v_n$.

- **Proof** Let $a_1v_1 + \ldots + a_nv_n$ be a linear combination of $v_1, \ldots, v_n$. Then by linearity

$$\langle a_1v_1 + \ldots + a_nv_n, w\rangle = a_1\langle v_1, w\rangle + \ldots + a_n\langle v_n, w\rangle.$$

But since w is orthogonal to each of $v_1, \ldots, v_n$, all the terms on the right-hand side are zero. Thus w is orthogonal to $a_1 v_1 + \ldots + a_n v_n$ as desired. $\square$

- In particular, if v and w are orthogonal, then cv and w are also orthogonal for any scalar c (why is this a special case of Lemma 1?)
- You are all familiar with the following theorem about orthogonality.
- **Pythagoras's theorem** If v and w are orthogonal vectors, then $\|v + w\|^2 = \|v\|^2 + \|w\|^2$.
- **Proof.** We compute

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle.$$

But since v and w are orthogonal, $\langle v, w \rangle$ and $\langle w, v \rangle$ are zero. Since $\langle v, v \rangle = \|v\|^2$ and $\langle w, w \rangle = \|w\|^2$, we obtain $\|v + w\|^2 = \|v\|^2 + \|w\|^2$ as desired. $\square$

- This theorem can be generalized:
- **Generalized Pythagoras's theorem** If $v_1, v_2, \ldots, v_n$ are all orthogonal to each other (i.e. $v_i \perp v_j = 0$ for all $i \neq j$) then

$$\|v_1 + v_2 + \ldots + v_n\|^2 = \|v_1\|^2 + \|v_2\|^2 + \ldots + \|v_n\|^2.$$

- **Proof.** We prove this by induction. If $n = 1$ the claim is trivial, and for $n = 2$ this is just the ordinary Pythagoras theorem. Now suppose that $n > 2$, and the claim has already been proven for $n - 1$. From Lemma 1 we know that v_n is orthogonal to $v_1 + \ldots + v_{n-1}$, so

$$\|v_1 + v_2 + \ldots + v_n\|^2 = \|v_1 + \ldots + v_{n-1}\|^2 + \|v_n\|^2.$$

On the other hand, by the induction hypothesis we know that

$$\|v_1 + \ldots + v_{n-1}\|^2 = \|v_1\|^2 + \ldots + \|v_{n-1}\|^2.$$

Combining the two equations we obtain

$$\|v_1 + v_2 + \ldots + v_n\|^2 = \|v_1\|^2 + \|v_2\|^2 + \ldots + \|v_n\|^2$$

as desired. $\square$

- Recall that if two vectors are orthogonal, then they remain orthogonal even when you multiply one or both of them by a scalar. So we have

- **Corollary 2.** If $v_1, v_2, \ldots, v_n$ are all orthogonal to each other (i.e. $v_i \perp v_j$ for all $i \neq j$) and $a_1, \ldots, a_n$ are scalars, then

  $$\|a_1v_1 + a_2v_2 + \ldots + a_nv_n\|^2 = |a_1|^2\|v_1\|^2 + |a_2|^2\|v_2\|^2 + \ldots + |a_n|^2\|v_n\|^2.$$

- **Definition** A collection $(v_1, v_2, \ldots, v_n)$ of vectors is said to be *orthogonal* if every pair of vectors is orthogonal to each other (i.e. $\langle v_i, v_j\rangle = 0$ for all $i \neq j$). If a collection is orthogonal, and furthermore each vector has length 1 (i.e. $\|v_i\| = 1$ for all i) then we say that the collection is *orthonormal*.

- **Example** In $\mathbf{R}^4$, the collection $((3,0,0,0), (0,4,0,0), (0,0,5,0))$ is orthogonal but not orthonormal. But the collection $((1,0,0,0), (0,1,0,0), (0,0,1,0))$ is orthonormal (and therefore orthogonal). Note that any single vector v_1 is always considered an orthogonal collection (why?).

- **Corollary 3**. If $(v_1, v_2, \ldots, v_n)$ is an orthonormal collection of vectors, and $a_1, \ldots, a_n$ are scalars, then

  $$\|a_1v_1 + a_2v_2 + \ldots + a_nv_n\|^2 = |a_1|^2 + |a_2|^2 + \ldots + |a_n|^2.$$

  Note that the right-hand side $|a_1|^2 + |a_2|^2 + \ldots + |a_n|^2$ is always positive, unless $a_1, \ldots, a_n$ are all zero. Thus $a_1v_1 + \ldots + a_nv_n$ is always non-zero, unless $a_1, \ldots, a_n$ are all zero. Thus

- **Corollary 4**. Every orthonormal collection of vectors is linearly independent.

* * * * *

Orthonormal bases

- As we have seen, orthonormal collections of vectors have many nice properties. As we shall see, things are even better when this collection is also a basis:

- **Definition** An *orthonormal basis* of an inner product space V is a collection $(v_1, \ldots, v_n)$ of vectors which is orthonormal and is also an ordered basis.

- **Example.** In $\mathbf{R}^4$, the collection $((1,0,0,0),(0,1,0,0),(0,0,1,0))$ is orthonormal but is not a basis. However, the collection $((1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,$ is an orthonormal basis. More generally, the standard ordered basis of $\mathbf{R}^n$ is always an orthonormal basis, as is the standard ordered basis of $\mathbf{C}^n$. (Actually, the standard bases of $\mathbf{R}^n$ and of $\mathbf{C}^n$ are the same; only the field of scalars is different). The collection $((1,0,0,0),(1,1,0,0),(1,1,1,0),(1,1,1,1))$ is a basis of $\mathbf{R}^4$ but is not an orthonormal basis.

- From Corollary 4 we have

- **Corollary 5** Let $(v_1, \ldots, v_n)$ be an orthonormal collection of vectors in an n-dimensional inner product space. Then $(v_1, \ldots, v_n)$ is an orthonormal basis.

- **Proof.** This is just because any n linearly independent vectors in an n-dimensional space automatically form a basis. $\square$

- **Example** Consider the vectors $(3/5, 4/5)$ and $(-4/5, 3/5)$ in $\mathbf{R}^2$. It is easy to check that they have length 1 and are orthogonal. Since $\mathbf{R}^2$ is two-dimensional, they thus form an orthonormal basis.

- Let $(v_1, \ldots, v_n)$ be an ordered basis of an n-dimensional inner product space V. Since $(v_1, \ldots, v_n)$ is a basis, we know that every vector v in V can be written as a linear combination of $v_1, \ldots, v_n$:

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

  In general, finding these scalars $a_1, \ldots, a_n$ can be tedious, and often requires lots of Gaussian elimination. (Try writing $(1,0,0,0)$ as a linear combination of $(1,1,1,1)$, $(1,2,3,4)$, $(2,2,1,1)$ and $(1,2,1,2)$, for instance). However, if we know that the basis is an *orthonormal* basis, then finding these coefficients is much easier.

- **Theorem 6.** Let $(v_1, \ldots, v_n)$ be an orthonormal basis of an inner product space V. Then for any vector $v \in V$, we have

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

where the scalars $a_1, \ldots, a_n$ are given by the formula

$$a_j = \langle v, v_j \rangle \text{ for all } j = 1, \ldots, n$$

- **Proof.** Since $(v_1, \ldots, v_n)$ is a basis, we know that $v = a_1 v_1 + \ldots + a_n v_n$ for *some* scalars $a_1, \ldots, a_n$. To finish the proof we have to solve for $a_1, \ldots, a_n$ and verify that $a_j = \langle v, v_j \rangle$ for all $j = 1, \ldots, n$. To do this we take our equation for v and take inner products of both sides with v_j:
$$\langle v, v_j \rangle = \langle a_1 v_1 + a_2 v_2 + \ldots + a_n v_n, v_j \rangle.$$
We expand out the right-hand side as
$$a_1 \langle v_1, v_j \rangle + a_2 \langle v_2, v_j \rangle + \ldots + a_n \langle v_n, v_j \rangle.$$
Since $v_1, \ldots, v_n$ are orthogonal, all the inner products vanish except $\langle v_j, v_j \rangle = \|v_j\|^2$. But $\|v_j\| = 1$ since $v_1, \ldots, v_n$ is also orthonormal. So we get
$$\langle v, v_j \rangle = 0 + \ldots + 0 + a_j \times 1 + 0 + \ldots + 0$$
as desired. □

From the definition of co-ordinate vector $[v]^\beta$, we thus have a simple way to compute co-ordinate vectors:

- **Corollary 7** Let $\beta = (v_1, \ldots, v_n)$ be an orthonormal basis of an inner product space V. Then the co-ordinate vector $[v]^\beta$ of any vector v is then given by
$$[v]^\beta = \begin{pmatrix} \langle v, v_1 \rangle \\ \langle v, v_2 \rangle \\ \vdots \\ \langle v, v_n \rangle \end{pmatrix}.$$

- Note that Corollary 7 also gives us a reltaively quick way to compute the co-ordinate matrix $[T]_\beta^\gamma$ of a linear operator $T : V \to W$ provided that γ is an orthonormal basis, since the columns of $[T]_\beta^\gamma$ are just $[Tv_j]^\gamma$, where v_j are the basis vectors of β.

- **Example.** Let $v_1 := (1,0,0)$, $v_2 := (0,1,0)$, $v_3 := (0,0,1)$, and $v := (3,4,5)$. Then (v_1, v_2, v_3) is an orthonormal basis of $\mathbf{R}^3$. Thus

$$v = a_1 v_1 + a_2 v_2 + a_3 v_3,$$

where $a_1 := \langle v, v_1 \rangle = 3$, $a_2 := \langle v, v_2 \rangle = 4$, and $a_3 := \langle v, v_3 \rangle = 5$. (Of course, in this case one could expand v as a linear combination of v_1, v_2, v_3 just by inspection.)

- **Example.** Let $v_1 := (3/5, 4/5)$, $v_2 := (-4/5, 3/5)$, and $v := (1,0)$. Then (v_1, v_2) is an orthonormal basis for $\mathbf{R}^2$. Now suppose we want to write v as a linear combination of v_1 and v_2. We could use Gaussian elimination, but because our basis is orthogonal we can use Theorem 6 instead to write

$$v = a_1 v_1 + a_2 v_2$$

where $a_1 := \langle v, v_1 \rangle = 3/5$ and $a_2 := \langle v, v_2 \rangle := -4/5$. Thus $v = \frac{3}{5}v_1 - \frac{4}{5}v_2$. (Try doing the same thing using Gaussian elimination, and see how much longer it would take!). Equivalently, we have

$$[v]^{(v_1, v_2)} = \begin{pmatrix} 3/5 \\ -4/5 \end{pmatrix}.$$

- **The example of Fourier series.** We now give an example which is important in many areas of mathematics (though we won't use it much in this particular course) - the example of *Fourier series*. Let $C([0,1]; \mathbf{C})$ be the inner product space of continuous complex-valued functions on the interval $[0,1]$, with the inner product

$$\langle f, g \rangle := \int_0^1 f(x)\overline{g(x)}\ dx.$$

Now consider the functions $\ldots, v_{-3}, v_{-2}, v_{-1}, v_0, v_1, v_2, v_3, \ldots$ in $C([0,1]; \mathbf{C})$ defined by

$$v_k(x) := e^{2\pi i k x};$$

these functions are sometimes known as *complex harmonics*.

- Observe that these functions all have length 1:

$$\|v_k\| = \langle v_k, v_k\rangle^{1/2} = (\int_0^1 v_k(x)\overline{v_k(x)}\ dx)^{1/2}$$

$$= (\int_0^1 e^{2\pi ikx}e^{-2\pi ikx}\ dx)^{1/2} = (\int_0^1 1\ dx)^{1/2} = 1.$$

Also, they are all orthogonal: if $j \neq k$ then

$$\langle v_j, v_k\rangle = \int_0^1 v_j(x)\overline{v_k(x)}\ dx = \int_0^1 e^{2\pi ijx}e^{-2\pi ikx}\ dx$$

$$= \int_0^1 e^{2\pi i(j-k)x}\ dx = \frac{e^{2\pi i(j-k)x}}{2\pi i(j-k)}|_0^1 = \frac{e^{2\pi i(j-k)}-1}{(2\pi i(j-k)}$$

$$= \frac{1-1}{2\pi i(j-k)} = 0.$$

Thus the collection $\ldots, v_{-3}, v_{-2}, v_{-1}, v_0, v_1, v_2, v_3, \ldots$ is an infinite orthonormal collection of vectors in $C([0,1];\mathbf{C})$.

- We have not really discussed infinite bases, but it does turn out that, in some sense, that the above collection is an orthonormal basis; thus every function in $C([0,1];\mathbf{C})$ is a linear combination of complex harmonics. (The catch is that this is an *infinite* linear combination, and one needs the theory of infinite series (as in Math 33B) to make this precise. This would take us too far afield from this course, unfortunately). This statement - which is not very intuitive at first glance - was first conjectured by Fourier, and forms the basis for something called *Fourier analysis* (which is an entire course in itself!). For now, let us work with a simpler situation.

- Define the space T_n of *trigonometric polynomials of degree at most n* to be the span of $v_0, v_1, v_2, \ldots, v_n$. In other words, T_n consists of all the functions $f \in C([0,1];\mathbf{C})$ of the form

$$f = a_0 + a_1e^{2\pi ix} + a_2e^{2\pi i2x} + \ldots + a_ne^{2\pi inx}.$$

Notice that this is very similar to the space $P_n(\mathbf{R})$ of polynomials of degree at most n, since an element $f \in P_n(\mathbf{R})$ has the form

$$f = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

and so the difference between polynomials and trigonometric polynomials is that x has been replaced by $e^{2\pi ix} = \cos(2\pi x) + i\sin(2\pi x)$ (hence the name, trigonometric polynomial).

- T_n is a subspace of the inner product space $C([0,1]; \mathbf{C})$, and is thus itself an inner product space. Since the vectors $v_0, v_1, \ldots, v_n$ are orthonormal, they are linearly independent, and thus $(v_0, v_1, \ldots, v_n)$ is an orthonormal basis for T_n. Thus by Theorem 6, every function f in T_n can be written as a series

$$f = a_0 + a_1 e^{2\pi ix} + a_2 e^{2\pi i2x} + \ldots + a_n e^{2\pi inx} = \sum_{j=0}^{n} a_j e^{2\pi ijx}.$$

where the (complex) scalars $a_0, a_1, \ldots, a_n$ are given by the formula

$$a_j := \langle f, v_j \rangle = \int_0^1 f(x) e^{-2\pi ijx}\, dx.$$

The coefficients a_j are known as the *Fourier coefficients* of f, and the above series is known as the *Fourier series* of f. From Corollary 3 we have the formula

$$\int_0^1 |f(x)|^2\, dx = \|f\|^2 = |a_0|^2 + |a_1|^2 + \ldots + |a_n|^2 = \sum_{j=0}^{n} |a_j|^2;$$

this is known as *Plancherel's formula*. These formulas form the foundation of Fourier analysis, and are useful in many other areas, such as signal processing, partial differential equations, and number theory. (Actually, to be truly useful, one needs to generalize these formulas to handle all kinds of functions, not just trigonometric polynomials, but to do so is beyond the scope of this course).

* * * * *

The Gram-Schmidt orthogonalization process.

- In this section all vectors are assumed to belong to a fixed inner product space V.

- In the last section we saw how many more useful properties orthonormal bases had, in comparison with ordinary bases. So it would be nice if we had some way of converting a non-orthonormal basis into an orthonormal one. Fortunately, there is such a process, and it is called *Gram-Schmidt orthogonalization*.

- To make a basis orthonormal there are really two steps; first one has to make a basis orthogonal, and then once it is orthogonal, one has to make it orthonormal. The second procedure is easier to describe than the first, so let us describe that first.

- **Definition.** A *unit vector* is any vector v of length 1 (i.e. $\|v\| = 1$, or equivalently $\langle v, v\rangle = 1$).

- **Example.** In $\mathbf{R}^2$, the vector $(3/5, 4/5)$ is a unit vector, but $(3, 4)$ is not. In $C([0, 1]; \mathbf{C})$, the function x is not a unit vector ($\|x\|^2 = \int_0^1 xx\, dx = 1/2$), but $\sqrt{2}x$ is (why?). The 0 vector is never a unit vector. In $\mathbf{R}^3$, the vectors $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ are all unit vectors.

- Unit vectors are sometimes known as *normalized* vectors. Note that an orthogonal basis will be orthonormal if it consists entirely of unit vectors.

- Most non-zero vectors are not unit vectors, e.g. $(3, 4)$ is not a unit vector. However, one can always turn a non-zero vector into a unit vector by dividing out by its length:

- **Lemma 8.** If v is a non-zero vector, then $v/\|v\|$ is a unit vector.

- **Proof.** Since v is non-zero, $\|v\|$ is non-zero, so $v/\|v\|$ is well defined. But then
$$\|v/\|v\|\| = \|\frac{1}{\|v\|}v\| = \frac{1}{\|v\|}\|v\| = 1$$
and so $v/\|v\|$ is a unit vector. $\square$

- We sometimes call $v/\|v\|$ the *normalization* of v. If $(v_1, v_2, \ldots, v_n)$ is a basis, then we can *normalize* this basis by replacing each vector v_j by its normalization $v_j/\|v_j\|$, obtaining a new basis $(v_1/\|v_1\|, v_2/\|v_2\|, \ldots, v_n/\|v_n\|)$ which now consists entirely of unit vectors. (Why is this still a basis?)

- **Lemma 9.** If $(v_1, v_2, \ldots, v_n)$ is an orthogonal basis of an inner product space V, then the normalization $(v_1/\|v_1\|, v_2/\|v_2\|, \ldots, v_n/\|v_n\|)$ is an orthonormal basis.

- **Proof.** Since the basis $(v_1, \ldots, v_n)$ has n elements, V must be n-dimensional. Since the vectors v_j are orthogonal to each other, the vectors $v_j/\|v_j\|$ must also be orthogonal to each other (multiplying vectors by a scalar does not affect orthogonality). By Lemma 8, these vectors are also unit vectors. So the claim follows from Corollary 5. $\square$

- **Example.** The basis $((3, 4), (-40, 30))$ is an orthogonal basis of $\mathbf{R}^2$ (why?). If we normalize this basis we obtain $((3/5, 4/5), (-4/5, 3/5))$, and this is now an orthonormal basis of $\mathbf{R}^2$.

- So we now know how to turn an orthogonal basis into an orthonormal basis - we normalize all the vectors by dividing out their length. Now we come to the tricker part of the procedure - how to turn a non-orthogonal basis into an orthogonal one. The idea is now to subtract scalar multiples of one vector from another to make them orthogonal (you might see some analogy here with row operations of the second and third kind).

- To illustrate the idea, we first consider the problem of how to make just two vectors v, w orthogonal to each other.

- **Lemma 10.** If v and w are vectors, and w is non-zero, then the vector $v - cw$ is orthogonal to w, where the scalar c is given by the formula $c := \frac{\langle v,w\rangle}{\|w\|^2}$.

- **Proof.** We compute

$$\langle v - cw, w\rangle = \langle v, w\rangle - c\langle w, w\rangle = \langle v, w\rangle - c\|w\|^2.$$

But since $c := \frac{\langle v,w\rangle}{\|w\|^2}$, we have $\langle v - cw, w\rangle = 0$ and so $v - cw$ is orthogonal to w. $\square$

- **Example.** Let $v = (3, 4)$ and $w = (5, 0)$. Then v and w are not orthogonal; in fact, $\langle v, w\rangle = 15 \neq 0$. But if we replace v by the vector $v' := v - cw = (3, 4) - \frac{15}{5^2}(5, 0) = (3, 4) - (3, 0) = (0, 4)$, then v' is now orthogonal to w.

- Now we suppose that we have already made k vectors orthogonal to each other, and now we work out how to make a $(k+1)^{th}$ vector also orthogonal to the first k.

- **Lemma 11.** Let $w_1, w_2, \ldots, w_k$ be orthogonal non-zero vectors, and let v be another vector. Then the vector v', defined by

$$v' := v - c_1 w_1 - c_2 w_2 - \ldots - c_k w_k$$

is orthogonal to all of $w_1, w_2, \ldots, w_k$, where the scalars $c_1, c_2, \ldots, c_k$ are given by the formula

$$c_j := \frac{\langle v, w_j \rangle}{\|w_j\|^2}$$

for all $j = 1, \ldots, k$.

- Note that Lemma 10 is just Lemma 11 applied to the special case $k = 1$. We can write v' in series notation as

$$v' := v - \sum_{j=1}^{k} \frac{\langle v, w_j \rangle}{\|w_j\|^2} w_j.$$

- **Proof.** We have to show that v' is orthogonal to each w_j. We compute

$$\langle v', w_j \rangle = \langle v, w_j \rangle - c_1 \langle w_1, w_j \rangle - c_2 \langle w_2, w_j \rangle - \ldots - c_k \langle w_k, w_j \rangle.$$

But we are assuming that the $w_1, \ldots, w_k$ are orthogonal, so all the inner products $\langle w_i, w_j \rangle$ are zero, except for $\langle w_j, w_j \rangle$, which is equal to $\|w_j\|^2$. Thus

$$\langle v', w_j \rangle = \langle v, w_j \rangle - c_j \|w_j\|^2.$$

But since $c_j := \frac{\langle v, w_j \rangle}{\|w_j\|^2}$, we thus have $\langle v', w_j \rangle$ and so v' is orthogonal to w_j. $\square$

- We can now use Lemma 11 to turn any linearly independent set of vectors into an orthogonal set.

- **Gram-Schmidt orthogonalization process.** Let $v_1, v_2, \ldots, v_n$ be a linearly independent set of vectors. Suppose we construct the vectors $w_1, \ldots, w_n$ by the formulae

$$w_1 := v_1$$

$$w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1$$

$$w_3 := v_3 - \frac{\langle v_3, w_1 \rangle}{\|w_1\|^2} w_1 - \frac{\langle v_3, w_2 \rangle}{\|w_2\|^2} w_2$$

$$\ldots$$

$$w_n := v_n - \frac{\langle v_n, w_1 \rangle}{\|w_1\|^2} w_1 - \ldots - \frac{\langle v_n, w_{n-1} \rangle}{\|w_{n-1}\|^2} w_{n-1}.$$

Then the vectors $w_1, \ldots, w_n$ are orthogonal, non-zero, and the vector space $\operatorname{span}(w_1, \ldots, w_n)$ is the same vector space as $\operatorname{span}(v_1, \ldots, v_n)$ (i.e. the vectors $w_1, \ldots, w_n$ have the same span as $v_1, \ldots, v_n$). More generally, we have that $w_1, \ldots, w_k$ has the same span as $v_1, \ldots, v_k$ for all $1 \leq k \leq n$.

- **Proof.** We prove this by induction on n. In the base case $n = 1$ we just have $w_1 := v_1$, and so clearly v_1 and w_1 has the same span. Also v_1 is an orthonormal collection of vectors (by default, since there is nobody else to be orthonormal to).

  Now suppose inductively that $n > 1$, and that we have already proven the claim for $n - 1$. In particular, we already know that the vectors $w_1, \ldots, w_{n-1}$ are orthogonal, non-zero, and that $v_1, \ldots, v_k$ has the same span as $w_1, \ldots, w_k$ for any $1 \leq k \leq n - 1$. By Lemma 11, we thus see that the vector w_n is orthogonal to $w_1, \ldots, w_{n-1}$. Now we have to show that w_n is non-zero and that $w_1, \ldots, w_n$ has the same span as $v_1, \ldots, v_n$.

  Let V denote the span of $v_1, \ldots, v_n$, and W denote the span of $w_1, \ldots, w_n$. We have to show that $V = W$. Note that W contains the span of $w_1, \ldots, w_{n-1}$, and hence contains the span of $v_1, \ldots, v_{n-1}$. In particular it contains $v_1, \ldots, v_{n-1}$, and also contains w_n. But from the formula

$$v_n = w_n + \frac{\langle v_n, w_1 \rangle}{\|w_1\|^2} w_1 + \ldots + \frac{\langle n_3, w_{n-1} \rangle}{\|w_{n-1}\|^2} w_{n-1}$$

  we thus see that W contains v_n. Thus W contains the span of $v_1, \ldots, v_n$, i.e. W contains V. But V is n-dimensional (since it is the span of n linearly independent vectors), and W is at most n dimensional (since W is also the span of n vectors), and so V and W must actually be equal. Furthermore this shows that $w_1, \ldots, w_n$ are linearly independent (otherwise W would have dimension less than n). In particular w_n is non-zero. This completes everything we need to do to finish the induction. $\square$

- **Example** Let $v_1 := (1, 1, 1)$, $v_2 := (1, 1, 0)$, and $v_3 := (1, 0, 0)$. The vectors v_1, v_2, v_3 are independent (in fact, they form a basis for $\mathbf{R}^3$) but are not orthogonal. To make them orthogonal, we apply the Gram-Schmidt orthogonalization process, setting

$$w_1 := v_1 = (1, 1, 1)$$

$$w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1 = (1, 1, 0) - \frac{2}{3}(1, 1, 1) = (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3})$$

$$w_3 := v_3 - \frac{\langle v_3, w_1 \rangle}{\|w_1\|^2} w_1 - \frac{\langle v_3, w_2 \rangle}{\|w_2\|^2} w_2$$

$$= (1, 0, 0) - \frac{1}{3}(1, 1, 1) - \frac{1/3}{6/9}(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}) = (1/2, -1/2, 0).$$

Thus we have created an orthogonal set $(1, 1, 1), (\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}), (\frac{1}{2}, -\frac{1}{2}, 0)$, which has the same span as v_1, v_2, v_3, i.e. it is also a basis for $\mathbf{R}^3$. Note that we can then use Lemma 8 to normalize this basis and make it orthonormal, obtaining the orthonormal basis

$$\frac{1}{\sqrt{3}}(1, 1, 1), \frac{1}{\sqrt{6}}(1, 1, -2), \frac{1}{\sqrt{2}}(1, -1, 0).$$

- We shall call the *normalized Gram-Schmidt orthogonalization process* the procedure of first applying the ordinary Gram-Schmidt orthogonalization process, and then normalizing all the vectors one obtains as a result of that process in order for them to have unit length.

- One particular consequence of Gram-Schmidt is that we always have at least one orthonormal basis lying around, at least for finite-dimensional inner product spaces.

- **Corollary 12.** Every finite-dimensional inner product space V has an orthonormal basis.

- **Proof.** Let's say V is n-dimensional. Then V has some basis $(v_1, \ldots, v_n)$. By the Gram-Schmidt orthogonalization process, we can thus create a new collection $(w_1, \ldots, w_n)$ of non-zero orthogonal vectors. By Lemma 9, we can then create a collection $(y_1, \ldots, y_n)$ of orthonormal vectors. By Corollary 5, it is an orthonormal basis of V. $\square$

* * * * *

Orthogonal complements

- We know what it means for two vectors to be orthogonal to each other, $v \perp w$; it just means that $\langle v, w \rangle = 0$. We now state what it means for two *subspaces* to be orthogonal to each other.

- **Definition.** Two subspaces V_1, V_2 of an inner product space V are said to be *orthogonal* if we have $v_1 \perp v_2$ for all $v_1 \in V_1$ and $v_2 \in V_2$, and we denote this by $V_1 \perp V_2$.

- **Example.** The subspaces $V_1 := \{(x, y, 0, 0, 0) : x, y \in \mathbf{R}\}$ and $V_2 := \{(0, 0, z, w, 0) : z, w \in \mathbf{R}\}$ of $\mathbf{R}^5$ are orthogonal, because $(x, y, 0, 0, 0) \perp (0, 0, z, w, 0)$ for all $x, y, z, w \in \mathbf{R}$. The space V_1 is similarly orthogonal to the three-dimensional space $V_3 := \{(0, 0, z, w, u) : z, w, u \in \mathbf{R}\}$. However, V_1 is not orthogonal to the one-dimensional space $V_4 := \{(t, t, t, t, t) : t \in \mathbf{R}\}$, since the inner product of $(x, y, 0, 0, 0)$ and (t, t, t, t, t) can be non-zero (e.g. take $x = y = t = 1$).

- **Example.** The zero vector space $\{0\}$ is orthogonal to any other subspace of V (why?)

- Orthogonal spaces have to be disjoint:

- **Lemma 13.** If $V_1 \perp V_2$, then $V_1 \cap V_2 = \{0\}$.

- Clearly 0 lies in $V_1 \cap V_2$ since every vector space contains 0. Now suppose for contradiction that $V_1 \cap V_2$ contained at least one other vector v, which must of course be non-zero. Then $v \in V_1$ and $v \in V_2$; since $V_1 \perp V_2$, this implies that $v \perp v$, i.e. that $\langle v, v \rangle = 0$. But this implies that $\|v\|^2 = 0$, hence $v = 0$, contradiction. Thus $V_1 \cap V_2$ does not contain any vector other than zero. $\square$

- As we can see from the above example, a subspace V_1 can be orthogonal to many other subspaces V_2. However, there is a maximal orthogonal subspace to V_1 which contains all the others:

- **Definition** The *orthogonal complement* of a subspace V_1 of an inner product space V, denoted $V_1^\perp$, is defined to be the space of all vectors perpendicular to V_1:

$$V_1^\perp := \{v \in V : v \perp w \text{ for all } w \in V_1\}.$$

- **Example.** Let $V_1 := \{(x, y, 0, 0, 0) : x, y \in \mathbf{R}\}$. Then $V_1^\perp$ is the space of all vectors $(a, b, c, d, e) \in \mathbf{R}^5$ such that (a, b, c, d, e) is perpendicular to V_1, i.e.

$$\langle (a, b, c, d, e), (x, y, 0, 0, 0) \rangle = 0 \text{ for all } x, y \in \mathbf{R}.$$

In other words,

$$ax + by = 0 \text{ for all } x, y \in \mathbf{R}.$$

This can only happen when $a = b = 0$, although c, d, e can be arbitrary. Thus we have

$$V_1^\perp = \{(0, 0, c, d, e) : c, d, e \in \mathbf{R}\},$$

i.e. $V_1^\perp$ is the space V_3 from the previous example.

- **Example.** If $\{0\}$ is the zero vector space, then $\{0\}^\perp = V$ (why?). A little trickier is that $V^\perp = \{0\}$. (Exercise! Hint: if v is perpendicular to every vector in V, then in particular it must be perpendicular to itself).

- From Lemma 1 we can check that $V_1^\perp$ is a subspace of V (exercise!), and is hence an inner product space.

- **Lemma 14.** If $V_1 \perp V_2$, then V_2 is a subspace of $V_1^\perp$. Conversely, if V_2 is a subspace of $V_1^\perp$, then $V_1 \perp V_2$.

- **Proof.** First suppose that $V_1 \perp V_2$. Then every vector v in V_2 is orthogonal to all of V_1, and hence lies in $V_1^\perp$ by definition of $V_1^\perp$. Thus V_2 is a subspace of $V_1^\perp$. Conversely, if V_2 is a subspace of $V_1^\perp$, then every vector v in V_2 is in $V_1^\perp$ and is thus orthogonal to every vector in V_1. Thus V_1 and V_2 are orthogonal. $\square$

- Sometimes it is not so easy to compute the orthogonal complement of a vector space, but the following result gives one way to do so.

- **Theorem 15.** Let W be a k-dimensional subspace of an n-dimensional inner product space V. Let $(v_1, \ldots, v_k)$ be a basis of W, and let $(v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n)$ be an extension of that basis to be a basis of V. Let $(w_1, \ldots, w_n)$ be the normalized Gram-Schmidt orthogonalization of $(v_1, \ldots, v_n)$. Then $(w_1, \ldots, w_k)$ is an orthonormal basis of W, and $(w_{k+1}, \ldots, w_n)$ is an orthonormal basis of $W^\perp$.

- **Proof.** From the Gram-Schmidt orthogonalization process, we know that $(w_1, \ldots, w_k)$ spans the same space as $(v_1, \ldots, v_k)$ - i.e. it spans W. Since W is k-dimensional, this means that $(w_1, \ldots, w_k)$ is a basis for W, which is orthonormal by the normalized Gram-Schmidt process. Similarly $(w_1, \ldots, w_n)$ spans the n-dimensional space V, which implies that it is a basis for V.

- Thus the the vectors $w_{k+1}, \ldots, w_n$ are orthonormal and thus (by Corollary 4) linearly independent. It remains to show that they span $W^\perp$. First we show that they lie in $W^\perp$. Let w_j be one of these vectors. Then w_j is orthogonal to $w_1, \ldots, w_k$, and is thus (by Lemma 1) orthogonal to their span, which is W. Thus w_j lies in $W^\perp$. In particular, the span of $w_{k+1}, \ldots, w_n$ must lie inside $W^\perp$.

- Now we show that every vector $V^\perp$ lies in the span of $w_1, \ldots, w_k$. Let v be any vector in $W^\perp$. By Theorem 6 we have

  $$v = \langle v, w_1 \rangle w_1 + \ldots + \langle v, w_n \rangle w_n.$$

  But since $v \in W^\perp$, v is orthogonal to $w_1, \ldots, w_k$, and so the first k terms on the right-hand side vanish. Thus we have

  $$v = \langle v, w_{k+1} \rangle w_{k+1} + \ldots + \langle v, w_n \rangle w_n$$

  and in particular v is in the span of $w_{k+1}, \ldots, w_n$ as desired. $\square$

- **Corollary 16 (Dimension theorem for orthogonal complements)** If W is a subspace of a finite-dimensional inner product space V, then $\dim(W) + \dim(W^\perp) = \dim(V)$.

- **Example.** Suppose we want to find the orthogonal complement of the line $W := \{(x, y) \in \mathbf{R}^2 : 3x + 4y = 0\}$ in $\mathbf{R}^2$. This example is so simple that one could do this directly, but instead we shall choose do this via Theorem 15 for sake of illustration. We first need to find a basis for W; since W is one-dimensional, we just to find one non-zero vector in W, e.g. $v_1 := (-4, 3)$, and this will be our basis. Then we extend this basis to a basis of the two-dimensional space $\mathbf{R}^2$ by adding one more linearly independent vector, for instance we could take $v_2 := (1, 0)$. This basis is not orthogonal or orthonormal, but we can apply the Gram-Schmidt process to make it orthogonal:

$$w_1 := v_1 = (-4, 3); \quad w_2 := v_2 - \frac{\langle v_2, w_1 \rangle}{\|w_1\|^2} w_1 = (1, 0) - \frac{-4}{25}(-4, 3) = (\frac{9}{25}, \frac{12}{25}).$$

We can then normalize:

$$w_1' := w_1/\|w_1\| = (-\frac{4}{5}, \frac{3}{5}); \quad w_2' := w_2/\|w_2\| = (\frac{3}{5}, \frac{4}{5}).$$

Thus w_1' is an orthonormal basis for W, w_2' is an orthonormal basis for $W^\perp$, and (w_1, w_2) is an orthonormal basis for $\mathbf{R}^2$. (Note that we could skip the normalization step at the end if one only wanted an orthogonal basis for these spaces, as opposed to an orthonormal basis).

- **Example.** Let's give $P_2(\mathbf{R})$ the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\overline{g(x)}\, dx.$$

The space $P_2(\mathbf{R})$ contains $P_1(\mathbf{R})$ as a subspace. Suppose we wish to compute the orthogonal complement of $P_1(\mathbf{R})$. We begin by taking a basis of $P_1(\mathbf{R})$ - let's use the standard basis $(1, x)$, and then extend it to a basis of $P_2(\mathbf{R})$ - e.g. $(1, x, x^2)$. We then apply the Gram-Schmidt orthogonalization procedure:

$$w_1 := 1$$

$$w_2 := x - \frac{\langle x, w_1 \rangle}{\|w_1\|^2} 1 = x - \frac{0}{2} 1 = x$$

$$w_3 := x^2 - \frac{\langle x^2, w_1 \rangle}{\|w_1\|^2} 1 - \frac{\langle x^2, w_2 \rangle}{\|w_2\|^2} x$$
$$= x^2 - \frac{2/3}{2} 1 - \frac{0}{2/3} x = x^2 - 1/3.$$

We can then normalize:

$$w_1' := w_1/\|w_1\| = \frac{1}{\sqrt{2}}$$

$$w_2' := w_2/\|w_2\| = \frac{\sqrt{3}}{\sqrt{2}} x$$

$$w_3' := w_3/\|w_3\| = \frac{\sqrt{45}}{\sqrt{8}} (x^2 - \frac{1}{3}).$$

Thus $W^\perp$ has w_3' as an orthonormal basis. Or one can just use w_3 as a basis, so that

$$W^\perp = \{a(x^2 - \frac{1}{3}) : a \in \mathbf{R}\}.$$

- **Corollary 17** If W is a k-dimensional subspace of an n-dimensional inner product space V, then every vector $v \in V$ can be written in exactly one way as $w + u$, where $w \in W$ and $u \in W^\perp$.

- **Proof.** By Theorem 15, we can find a orthonormal basis $(w_1, w_2, \ldots, w_n)$ of V such that $(w_1, \ldots, w_k)$ is an orthonormal basis of W and $(w_{k+1}, \ldots, w_n)$ is an orthonormal basis of $W^\perp$. If v is a vector in V, then we can write

$$v = a_1 w_1 + \ldots + a_n w_n$$

for some scalars $a_1, \ldots, a_n$. If we write

$$w := a_1 w_1 + \ldots + a_k w_k; \quad u := a_{k+1} w_{k+1} + \ldots + a_n w_n$$

then we have $w \in W, u \in W^\perp$, and $v = w + u$. Now we show that this is the only way to decompose v in this manner. If $v = w' + u'$ for some $w' \in W$, $u' \in W^\perp$, then

$$w + u = w' + u'$$

and so

$$w - w' = u' - u.$$

But $w - w'$ lies in W, and $u' - u$ lies in $W^\perp$. By Lemma 13, this vector must be 0, so that $w = w'$ and $u = u'$. Thus there is no other way to write $v = w + u$. $\square$

- We call the vector w obtained in the above manner the *orthogonal projection* of v onto W; this terminology can be made clear by a picture (at least in $\mathbf{R}^3$ or $\mathbf{R}^2$), since w, u, and v form a right-angled triangle whose base w lies in W. This projection can be computed as

$$w = a_1 w_1 + \ldots + a_k w_k = \langle v, w_1 \rangle w_1 + \ldots + \langle v, w_k \rangle w_k$$

where $w_1, \ldots, w_k$ is any orthonormal basis of W.

- **Example** Let $W := \{(x, y) \in \mathbf{R}^2 : 3x + 4y = 0\}$ be as before. Suppose we wish to find the orthogonal projection of the vector $(1, 1)$ to W. Since we have an orthonormal basis given by $w_1' := (-\frac{4}{5}, \frac{3}{5})$, we can compute the orthogonal projection as

$$w = \langle (1,1), w_1' \rangle w_1' = -\frac{1}{5}(-\frac{4}{5}, \frac{3}{5}) = (\frac{4}{25}, -\frac{3}{25}).$$

- The orthogonal projection has the "nearest neighbour" property:

- **Theorem 18.** Let W be a subspace of a finite-dimensional inner product space V, let v be a vector in V, and let w be the orthogonal projection of v onto W. Then w is closer to v than any other element of W; more precisely, we have $\|v - w'\| > \|v - w\|$ for all vectors w' in W other than w.

- **Proof.** Write $v = w + u$, where $w \in W$ is the orthogonal projection of v onto W, and $u \in W^\perp$. Then we have $\|v - w\| = \|u\|$. On the other hand, to compute $v - w'$, we write $v - w' = (v - w) + (w - w') = u + (w - w')$. Since w, w' lie in W, $w - w'$ does also. But u lies in $W^\perp$, thus $u \perp w - w'$. By Pythagoras's theorem we thus have

$$\|v - w'\|^2 = \|u\|^2 + \|w - w'\|^2 > \|u\|^2$$

(since $w \neq w'$) and so $\|v - w'\| > \|u\| = \|v - w\|$ as desired. $\square$

- This Theorem makes the orthogonal projection useful for approximating a given vector v in V by a another vector in a given subspace W of V.

- **Example** Consider the vector x^2 in $P_2(\mathbf{R})$. Suppose we want to find the linear polynomial $ax + b \in P_1(\mathbf{R})$ which is closest to x^2 (using the inner product on $[-1, 1]$ from the previous example to define length). By Theorem 18, this linear polynomial will be the orthogonal projection of x^2 to $P_1(\mathbf{R})$. Using the orthonormal basis $w_1' = \frac{1}{\sqrt{2}}$, $w_2' = \frac{\sqrt{3}}{\sqrt{2}}x$ from the prior example, we thus see that this linear polynomial is

$$\langle x^2, w_1' \rangle w_1' + \langle x^2, w_2' \rangle w_2' = \frac{2/3}{\sqrt{2}} \frac{1}{\sqrt{2}} + 0 \frac{\sqrt{3}}{\sqrt{2}} x = \frac{1}{3} + 0x.$$

  Thus the function $1/3 + 0x$ is the closest linear polynomial to x^2 using the inner product on $[-1, 1]$. (If one uses a different inner product, one can get a different "closest approximation"; the notion of closeness depends very much on how one measures length).

Math 115A - Week 10
Textbook sections: 3.1-5.1
Topics covered:

- Linear functionals
- Adjoints of linear operators
- Self-adjoint operators
- Normal operators
- Stuff about the final

* * * * *

Linear functionals

- Let F be either the real or complex numbers, and let V, W be vector spaces over the field of scalars F. We know what a linear transformation T from V to W is; it is a transformation that takes as input a vector v in V and returns a vector Tv in W, which preserves addition $T(v+v') = Tv + Tv'$ and scalar multiplication $T(cv) = cTv$.

- We now look at some special types of linear transformation, where the input space V or the output space W is very small. We first look at what happens when the input space is just F, the field of scalars.

- **Example** The linear transformation $T : \mathbf{R} \to \mathbf{R}^3$ defined by $Tc := (3c, 4c, 5c)$ is a linear transformation from the field of scalars $\mathbf{R}$ to a vector space $\mathbf{R}^3$.

- Note that the above example can be written as $Tc := cw$, where w is the vector $(3, 4, 5)$ in $\mathbf{R}^3$. The following lemma says, in fact, that all linear transformations from the field of scalars to another vector space are of this form:

- **Lemma 1.** Let $T : F \to W$ be a linear transformation from F to W. Then there is a vector $w \in W$ such that $Tc = cw$ for all $c \in F$.

- **Proof** Since $c = c1$, we have $Tc = T(c1) = c(T1)$. So if we set $w := T1$, then we have $Tc = cw$ for all $c \in F$. $\square$

- Now we look at what happens when the *output* space is the field of scalars.

- **Definition** A *linear functional* on a vector space V is a linear transformation $T : V \to F$ from V to the field of scalars F.

- Thus linear functionals are in some sense the "opposite" of vectors: they "eat" a vector as input and spit out a scalar as output. (They are sometimes called *covectors* or *dual vectors* for this reason; sometimes physicists call them *axial vectors*. Another name used is *1-forms*. In quantum mechanics, one sometimes uses Dirac's "braket" notation, in which vectors are called "kets" and covectors are called "bras").

- **Example 1.** The linear transformation $T : \mathbf{R}^3 \to \mathbf{R}$ defined by $T(x, y, z) := 3x + 4y + 5z$ is a linear functional on $\mathbf{R}^3$. Another example is *altitude*: the linear transformation $A : \mathbf{R}^3 \to \mathbf{R}$ defined by $A(x, y, z) := z$; this takes a vector in three-dimensional space as input and returns its altitude (the z co-ordinate).

- **Example 2 (integration as a linear functional).** The linear transformation $I : C([0, 1]; \mathbf{R}) \to \mathbf{R}$ defined by $If := \int_0^1 f(x)\, dx$ is a linear functional, for instance $I(x^2) = 1/3$.

- **Example 3 (evaluation as a linear functional).** The linear transformation $E : C([0, 1]; \mathbf{R}) \to \mathbf{R}$ defined by $Ef = f(0)$ is a linear functional, thus for instance $E(x^2) = 0$, and $E(e^x) = 1$.

- **Example 4.** Let V be any inner product space, and let w be any vector in V. Then the linear transformation $T : V \to F$ defined by $Tv := \langle v, w\rangle$ is a linear functional on V (this is because inner product is linear in the first variable v). For instance, the linear functional $T(x, y, z) := 3x + 4y + 5z$ in Example 1 is of this type, since $T(x, y, z) = \langle (x, y, z), (3, 4, 5)\rangle$; similarly the altitude function can be written in this form, as $A(x, y, z) = \langle (x, y, z), (1, 0, 0)\rangle$. Also, the integration functional I in Example 2 is also of this form, since $If = \langle f, 1\rangle$.

(As it turns out, the evaluation function E from Example 3 is not of this form, at least on $C([0,1];\mathbf{R})$; but see below.)

- It turns out that on an *finite-dimensional* inner product space V, every linear functional is of the form given in the previous example:

- **Riesz representation theorem.** Let V be a finite-dimensional inner product space, and let $T : V \to F$ be a linear functional on V. Then there is a vector $w \in V$ such that $Tv = \langle v, w \rangle$ for all $v \in V$.

- **Proof.** Let's say that V is n-dimensional. By the Gram-Schmidt orthogonalization process we can find an orthonormal basis $v_1, v_2, \ldots, v_n$ of V. Let v be any vector in V. From the previous week's notes we have the formula

$$v = \langle v, v_1 \rangle v_1 + \ldots + \langle v, v_n \rangle v_n.$$

Applying T to both sides, we obtain

$$Tv = \langle v, v_1 \rangle Tv_1 + \ldots + \langle v, v_n \rangle Tv_n.$$

Since $Tv_1, \ldots, Tv_n$ are all scalars, and $\langle v, w \rangle c = \langle v, \bar{c} w \rangle$ for any scalar c, and we thus have

$$Tv = \langle v, \overline{Tv_1} v_1 + \ldots \overline{Tv_n} v_n \rangle.$$

Thus if we let $w \in V$ be the vector

$$w := \overline{Tv_1} v_1 + \ldots \overline{Tv_n} v_n$$

then we have $Tv = \langle v, w \rangle$ for all $v \in V$, as desired. $\square$

- (Actually, this is only the Riesz representation theorem for finite dimensional spaces. There are more general Riesz representation theorems for such infinite-dimensional spaces as $C([0,1];\mathbf{R})$, but they are beyond the scope of this course).

- **Example** Consider the linear functional $T : \mathbf{C}^3 \to \mathbf{C}$ defined by $T(x, y, z) := 3x + iy + 5z$. From the Riesz representation theorem we know that there must be some vector $w \in \mathbf{C}^3$ such that $Tv := \langle v, w \rangle$

for all $v \in \mathbf{C}^3$. In this case we can see what w is by inspection, but let us pretend that we are unable to see this, and instead use the formula in the proof of the Riesz representation theorem. Namely, we know that

$$w := \overline{Tv_1}v_1 + \ldots \overline{Tv_n}v_n$$

whenever $v_1, \ldots, v_n$ is an orthonormal basis for $\mathbf{C}^3$. Thus, using the standard basis $(1,0,0), (0,1,0), (0,0,1)$, we obtain

$$w := \overline{T(1,0,0)}(1,0,0) + \overline{T(0,1,0)}(0,1,0) + \overline{T(0,0,1)}(0,0,1)$$

$$= \overline{3}(1,0,0) + \overline{i}(0,1,0) + \overline{5}(0,0,1) = (3,-i,5).$$

Thus $Tv = \langle v, (3,-i,5)\rangle$, which one can easily check is consistent with our definition of T.

- More generally, we see that any linear functional $T : F^n \to F$ (where $F = \mathbf{R}$ or $\mathbf{C}$) can be written in the form $Tv := \langle v, w\rangle$, where w is the vector

$$w = (\overline{Te_1}, \overline{Te_2}, \ldots, \overline{Te_n}),$$

and $e_1, \ldots, e_n$ is the standard basis for F^n. (i.e. the first component of w is $\overline{Te_1}$, etc. For instance, in the previous example $Te_1 = T(1,0,0) = 3$, so the first component of w is $\overline{3} = 3$.

- **Example** Let $P_2(\mathbf{R})$ be the polynomials of degree at most 2, with the inner product

$$\langle f, g\rangle := \int_{-1}^{1} f(x)g(x)\ dx.$$

Let $E : P_2(\mathbf{R}) \to \mathbf{R}$ be the evaluation function $E(f) := f(0)$, for instance $E(x^2 + 2x + 3) = 3$. From the Riesz representation theorem we know that $E(f) = \langle f, w\rangle$ for some $w \in P_2(\mathbf{R})$; we now find what this w is. We first find an orthonormal basis for $P_2(\mathbf{R})$. From last week's notes, we know that

$$v_1 := \frac{1}{\sqrt{2}}; v_2 := \frac{\sqrt{3}}{\sqrt{2}}x; v_3 := \frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3})$$

is an orthonormal basis for $P_2(\mathbf{R})$. Thus we can compute w using the formula

$$w = \overline{Tv_1}v_1 + \overline{Tv_2}v_2 + \overline{Tv_3}v_3$$

from the proof of the Riesz representation theorem. Since $Tv_1 = \frac{1}{\sqrt{2}}$, $Tv_2 = 0$, and $Tv_3 = \frac{\sqrt{45}}{\sqrt{8}}(-\frac{1}{3})$, we thus have

$$w = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} + \frac{\sqrt{45}}{\sqrt{8}}(-\frac{1}{3})\frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3})$$

which simplifies to

$$w = \frac{1}{2} - \frac{5}{24}(3x^2 - 1) = \frac{17}{24} - \frac{5}{8}x^2.$$

- It may seem that the vector w that is obtained by the Riesz representation theorem would depend on which orthonormal basis $v_1, \ldots, v_n$ one chooses for V. But it turns out that this is not the case:

- **Lemma 2.** Let $T : V \to \mathbf{R}$ be a linear functional on an inner product space V. Then there can be at most one vector $w \in V$ with the property that $Tv = \langle v, w \rangle$ for all $v \in V$.

- **Proof.** Suppose for contradiction that there were at least two different vectors w, w' in V such that $Tv = \langle v, w \rangle$ and $Tv = \langle v, w' \rangle$ for all $v \in V$. Then we have

$$\langle v, w - w' \rangle = \langle v, w \rangle - \langle v, w' \rangle = Tv - Tv = 0$$

for all $v \in V$. In particular, if we apply this identity to the vector $v := w - w'$ we obtain

$$\|w - w'\|^2 = \langle w - w', w - w' \rangle = 0$$

which implies that $w - w' = 0$, so that w and w' are not different after all. This contradiction shows that there could only have been one such vector w to begin with, as desired. $\square$

- Another way to view Lemma 2 is the following: if $\langle v, w \rangle = \langle v, w' \rangle$ for all $v \in V$, then w and w' must be equal. (If you like, this is sort of like being able to "cancel" v from both sides of an identity involving an inner product, provided that you know the identity holds for *all* v).

* * * * *

Adjoints

- The Riesz representation theorem allows us to turn linear functionals $T : V \to \mathbf{R}$ into vectors $w \in V$, if V is a finite-dimensional inner product space. This leads us to a useful notion, that of the *adjoint* of a linear operator.

- Let $T : V \to W$ be a linear transformation from one inner product space to another. Then for every vector $w \in W$, we can define a linear functional $T_w : V \to \mathbf{R}$ by the formula

$$T_w v := \langle Tv, w \rangle.$$

- **Example.** If $T : \mathbf{R}^3 \to \mathbf{R}^2$ is the linear transformation

$$T(x, y, z) := (x + 2y + 3z, 4x + 5y + 6z)$$

and w was the vector $(10, 1) \in \mathbf{R}^2$, then $T_w : \mathbf{R}^3 \to \mathbf{R}$ would be the linear functional

$$T_w(x, y, z) = \langle (x + 2y + 3z, 4x + 5y + 6z), (10, 1) \rangle = 14x + 25y + 36z.$$

- One can easily check that T_w is indeed a linear functional on V:

$$T_w(v+v') = \langle T(v+v'), w \rangle = \langle Tv+Tv', w \rangle = \langle Tv, w \rangle + \langle Tv', w \rangle = T_w v + T_w v'$$

$$T_w(cv) = \langle T(cv), w \rangle = \langle cTv, w \rangle = c\langle Tv, w \rangle = cT_w v.$$

- By the Riesz representation theorem, there must be a vector, called $T^* w \in V$, such that $T_w v = \langle v, T^* w \rangle$ for all $v \in V$, or in other words that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for all $w \in W$ and $v \in V$; this is probably the most basic property of T^*. Note that by Lemma 2, there can only be one possible value for $T^* w$ for each w.

- **Example** Continuing the previous example, we see that

$$T_w(x, y, z) = \langle (x, y, z), (14, 25, 36) \rangle$$

and hence by Lemma 2, the only possible choice for $T^* w$ is

$$T^*(10, 1) = T^* w = (14, 25, 36).$$

- Note that while T is a transformation that turns a vector v in V to a vector Tv in W, T^* does the opposite, starting with a vector w in W as input and returning a vector $T^* w$ in V as output. This seems similar to how an inverse T^{-1} of T would work, but it is important to emphasize that T^* is *not* the inverse of T, and it makes sense even when T is not invertible.

- We refer to $T^* : W \to V$ as the *adjoint* of T. Thus when we move an operator T from one side of an inner product to another, we have to replace it with its adjoint. This is similar to how when one moves a scalar from one side of an inner product to another, you have to replace it by its complex conjugate: $\langle cv, w \rangle = \langle v, \overline{c} w \rangle$. Thus the adjoint is like the complex conjugate, but for linear transformations rather than for scalars.

- **Lemma 3.** If $T : V \to W$ is a linear transformation, then its adjoint $T^* : W \to V$ is also a linear transformation.

- **Proof.** We have to prove that $T^*(w + w') = T^* w + T^* w'$ and $T^*(cw) = cT^* w$ for all $w, w' \in W$ and scalars c.

- First we prove that $T^*(w + w') = T^* w + T^* w'$. By definition of T^*, we have

$$\langle v, T^*(w + w') \rangle = \langle Tv, w + w' \rangle$$

for all $v \in V$. But

$$\langle Tv, w+w' \rangle = \langle Tv, w \rangle + \langle Tv, w' \rangle = \langle v, T^* w \rangle + \langle v, T^* w' \rangle = \langle v, T^* w + T^* w' \rangle.$$

Thus we have

$$\langle v, T^*(w + w') \rangle = \langle v, T^* w + T^* w' \rangle$$

for all $v \in V$. By Lemma 2, we must therefore have $T^* w + T^* w' = T^*(w + w')$ as desired.

- Now we show that $T^*(cw) = cT^*w$. We have

$$\langle v, T^*(cw)\rangle = \langle Tv, cw\rangle = \overline{c}\langle Tv, w\rangle = \overline{c}\langle v, T^*w\rangle = \langle v, cT^*w\rangle$$

for all $v \in V$. By Lemma 2, we thus have $T^*(cw) = cT^*w$ as desired. $\square$

- **Example** Let us continue our previous example of the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

$$T(x, y, z) := (x + 2y + 3z, 4x + 5y + 6z).$$

Let us work out what $T^* : \mathbf{R}^2 \to \mathbf{R}^3$ is. Let (a, b) be any vector in $\mathbf{R}^2$. Then we have

$$\langle T(x, y, z), (a, b)\rangle = \langle (x, y, z), T^*(a, b)\rangle$$

for all $(x, y, z) \in \mathbf{R}^3$. The left-hand side is

$$\langle (x + 2y + 3z, 4x + 5y + 6z), (a, b)\rangle = a(x + 2y + 3z) + b(4x + 5y + 6z)$$

$$= (a+4b)x+(2a+5b)y+(3a+6b)z = \langle (x, y, z), (a+4b, 2a+5b, 3a+6b)\rangle.$$

Thus we have

$$\langle (x, y, z), (a + 4b, 2a + 5b, 3a + 6b)\rangle = \langle (x, y, z), T^*(a, b)\rangle$$

for all x, y, z; by Lemma 2, this implies that

$$T^*(a, b) = (a + 4b, 2a + 5b, 3a + 6b).$$

- This example was rather tedious to compute. However, things become easier with the aid of orthonormal bases. Recall (from Corollary 7 of last week's notes) that if v is a vector in V and $\beta := (v_1, \ldots, v_n)$ is an orthonormal basis of V, then the column vector $[v]^\beta$ is given by

$$[v]^\beta = \begin{pmatrix} \langle v, v_1\rangle \\ \vdots \\ \langle v, v_n\rangle \end{pmatrix}.$$

Thus the i^{th} row entry of $[v]^\beta$ is just $\langle v, v_i\rangle$.

- Now suppose that $T : V \to W$ is a linear transformation, and $\beta := (v_1, \ldots, v_n)$ is an orthonormal basis of V and $\gamma := (w_1, \ldots, w_m)$ is an orthonormal basis of W. Then $[T]^\gamma_\beta$ is a matrix with m rows and n columns, whose j^{th} column is given by $[Tv_j]^\beta$. In other words, we have

$$[T]^\gamma_\beta = \begin{pmatrix} \langle Tv_1, w_1\rangle & \langle Tv_2, w_1\rangle & \ldots & \langle Tv_n, w_1\rangle \\ \langle Tv_1, w_2\rangle & \langle Tv_2, w_2\rangle & \ldots & \langle Tv_n, w_2\rangle \\ & & \vdots & \\ \langle Tv_1, w_m\rangle & \langle Tv_2, w_m\rangle & \ldots & \langle Tv_n, w_m\rangle \end{pmatrix}.$$

In other words, the entry in the i^{th} row and j^{th} column is $\langle Tv_j, w_i\rangle$.

- We can apply similar reasoning to the linear transformation $T^* : W \to V$. Then $[T^*]^\beta_\gamma$ is a matrix with n rows and m columns, and the entry in the i^{th} row and j^{th} column is $\langle T^* w_j, v_i\rangle$. But

$$\langle T^* w_j, v_i\rangle = \overline{\langle v_i, T^* w_j\rangle} = \overline{\langle Tv_i, w_j\rangle}.$$

Thus, the matrix $[T^*]^\beta_\gamma$ is given by

$$[T^*]^\beta_\gamma = \begin{pmatrix} \overline{\langle Tv_1, w_1\rangle} & \overline{\langle Tv_1, w_2\rangle} & \ldots & \overline{\langle Tv_1, w_m\rangle} \\ \overline{\langle Tv_2, w_1\rangle} & \overline{\langle Tv_2, w_2\rangle} & \ldots & \overline{\langle Tv_2, w_m\rangle} \\ & & \vdots & \\ \overline{\langle Tv_n, w_1\rangle} & \overline{\langle Tv_n, w_2\rangle} & \ldots & \overline{\langle Tv_n, w_m\rangle} \end{pmatrix}.$$

Comparing this with our formula for $[T]^\gamma_\beta$ we see that $[T^*]^\beta_\gamma$ is the adjoint of $[T]^\gamma_\beta$:

- **Theorem 3.** If $T : V \to W$ is a linear transformation, β is an orthonormal basis of V, and γ is an orthonormal basis of W, then

$$[T^*]^\beta_\gamma = ([T]^\gamma_\beta)^\dagger.$$

- **Example** Let us once again take the example of the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

$$T(x, y, z) := (x + 2y + 3z, 4x + 5y + 6z).$$

Let $\beta := ((1,0,0),(0,1,0),(0,0,1))$ be the standard basis of $\mathbf{R}^3$, and let $\gamma := ((1,0),(0,1))$ be the standard basis of $\mathbf{R}^2$. Then we have

$$[T]_\beta^\gamma := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

(why?). On the other hand, if we write the linear transformation

$$T^*(a,b) = (a+4b, 2a+5b, 3a+6b)$$

in matrix form, we see that

$$[T^*]_\gamma^\beta := \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix},$$

which is the adjoint of $[T]_\beta^\gamma$. (In this example, the field of scalars is real, and so the complex conjugation aspect of the adjoint does not make an appearance.

- The following corollary connects the notion of adjoint of a *linear transformation* with that of adjoint of a *matrix*.

- **Corollary 4.** Let A be an $m \times n$ matrix with either real or complex entries. Then the adjoint of L_A is $L_{A^\dagger}$.

- **Proof.** Let F be the field of scalars that the entries of A lie in. Then L_A is a linear transformation from F^m to F^n, and $L_{A^\dagger}$ is a linear transformation from F^n to F^m. If we let β be the standard basis of F^m and γ be the standard basis of F^n, then by Theorem 3

$$[L_A^*]_\gamma^\beta = ([L_A]_\beta^\gamma)^\dagger = A^\dagger = [L_{A^\dagger}]_\gamma^\beta$$

and hence $L_A^* = L_{A^\dagger}$ as desired. □

- In particular, we see that

$$\langle Av, w\rangle = \langle v, A^\dagger w\rangle$$

for any $m \times n$ matrix A, any column vector v of length n, and any column vector w of length m.

- **Example** Let A be the matrix

$$A = \begin{pmatrix} 1 & i & 0 \\ 0 & 1+i & 3 \end{pmatrix},$$

so that $L_A : \mathbf{C}^3 \to \mathbf{C}^2$ is the linear transformation defined by

$$L_A \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = A \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_1 + iz_2 \\ (1+i)z_2 + 3z_3 \end{pmatrix}.$$

Then the adjoint of this transformation is given by $L_{A^\dagger}$, where $A^\dagger$ is the adjoint of A:

$$A^\dagger = \begin{pmatrix} 1 & 0 \\ -i & 1-i \\ 0 & 3 \end{pmatrix},$$

so

$$L_{A^\dagger} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = A \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} w_1 \\ -iw_1 + (1-i)w_2 \\ 3w_3 \end{pmatrix}.$$

- Some basic properties of adjoints. Firstly, the process of taking adjoints is conjugate linear: if $T : V \to W$ and $U : V \to W$ are linear transformations, and c is a scalar, then $(T+U)^* = T^* + U^*$ and $(cT)^* = \overline{c}T^*$. Let's just prove the second claim, as the first is similar (or can be found in the textbook). We look at the expression $\langle v, (cT)^* w\rangle$ for any $v \in V$ and $w \in W$, and compute:

$$\langle v, (cT)^* w\rangle = \langle cTv, w\rangle = c\langle Tv, w\rangle = c\langle v, T^* w\rangle = \langle v, \overline{c}T^* w\rangle.$$

Since this identity is true for all $v \in V$, we thus have (by Lemma 2) that $(cT)^* w = \overline{c}T^* w$ for all $w \in W$, and so $(cT)^* = \overline{c}T^*$ as desired.

- This argument shows a key trick in understanding adjoints: in order to understand a transformation T or its adjoint, it is often a good idea to start by looking at the expression $\langle Tv, w\rangle = \langle v, T^* w\rangle$ and rewrite it in some other way.

- Some other properties, which we leave as exercises: $(T^*)^* = T$ (i.e. if T^* is the adjoint of T, then T is the adjoint of T^*); the adjoint of the identity operator is again the identity; and if $T : V \to W$ and $S : U \to V$ are linear transformations, then $(TS)^* = S^*T^*$. (This last identity can be verified by playing around with $\langle u, S^*T^*w\rangle$ for $u \in U$ and $w \in W$). If T is invertible, we also have $(T^{-1})^* = (T^*)^{-1}$ (i.e. the inverse of the adjoint is the adjoint of the inverse). This can be seen by starting with the identity $TT^{-1} = T^{-1}T = I$ and taking adjoints of all sides.

- Another useful property is that a matrix has the same rank as its adjoint. To see this, recall that the adjoint of a matrix is the conjugate of its transpose. From Lemma 7 of week 6 notes, we know that a matrix has the same rank as its transpose. It is also easy to see that a matrix has the same rank as its conjugate (this is basically because the conjugate of an elementary matrix is again an elementary matrix, and the conjugate of a matrix in row-echelon form is again a matrix in row echelon form.) Combining these two observations we see that the adjoint of a matrix must also have the same rank. From Theorem 3 (and Lemma 9 of week 6 notes) we see therefore that a linear operator from one finite-dimensional inner product space to another has the same rank as its adjoint.

- In a similar vein, if A is a square matrix with determinant d, then A^* will have determinant $\overline{d}$. (We will only sketch a proof of this fact here: first prove it for elementary matrices, and for diagonal matrices. Then to handle the general case, use Proposition 5 from week 6 notes, as well as the identity $(BA)^\dagger = A^\dagger B^\dagger$).

* * * * *

Normal operators

- Recall that in the Week 7 notes we discussed the problem of whether a linear transformation was diagonalizable, i.e. whether it had a basis of eigenvectors. We did not fully resolve this question, and in fact we will not be able to give a truly satisfactory answer to this question until Math 115B. However, there is a special class of linear transformations (aka operators) for which we can give a good answer - *normal* operators.

- **Definition** Let $T : V \to V$ be a linear transformation on V, so that the adjoint $T^* : V \to V$ is another linear transformation on V. We say that T is *normal* if $TT^* = T^*T$.

- **Example 1** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation $T(x, y) := (y, -x)$. Then $T^* : \mathbf{R}^2 \to \mathbf{R}^2$ can be computed to be the linear transformation $T^*(x, y) = (-y, x)$ (why?), and so

$$TT^*(x, y) = T(-y, x) = (x, y)$$

and

$$T^*T(x, y) = T^*(y, -x) = (x, y).$$

Thus $TT^*(x, y)$ and $T^*T(x, y)$ agree for all $(x, y) \in \mathbf{R}^2$, which implies that $TT^* = T^*T$. Thus this transformation is normal.

- **Example 2** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation $T(x, y) := (0, x)$. Then $T^*(x, y) = (y, 0)$ (why?). So

$$TT^*(x, y) = T(y, 0) = (0, y)$$

and

$$T^*T(x, y) = T^*(0, x) = (x, 0).$$

So in general $TT^*(x, y)$ and $T^*T(x, y)$ are not equal, and so $TT^* \neq T^*T$. Thus this transformation is not normal.

- In analogy to the above definition, we define a square matrix A to be *normal* if $AA^\dagger = A^\dagger A$. For instance, the matrix

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

can easily be checked to be normal, while the matrix

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

is not. (Why do these two examples correspond to Examples 1 and 2 above?)

- Another example, easily checked: every diagonal matrix is normal.

  From Theorem 3 we have

- **Proposition 5.** Let $T : V \to V$ be a linear transformation on a finite-dimensional inner product space, and let β be an orthonormal basis. Then $T : V \to V$ is normal if and only if the matrix $[T]_\beta^\beta$ is.

- **Proof.** If T is normal, then $TT^* = T^*T$. Now taking matrices with respect to β, we obtain

$$[T]_\beta^\beta [T^*]_\beta^\beta = [T^*]_\beta^\beta [T]_\beta^\beta.$$

  But by Theorem 3, $[T^*]_\beta^\beta$ is the adjoint of $[T]_\beta^\beta$. Thus $[T]_\beta^\beta$ is normal. This proves the "only if" portion of the Proposition; the "if" part follows by reversing the above steps. $\square$

- Normal transformations have several nice properties. First of all, when T is normal then T and T^* will have the same eigenvectors (but slightly different eigenvalues):

- **Lemma 6.** Let $T : V \to V$ be normal, and suppose that $Tv = \lambda v$ for some vector $v \in V$ and some scalar λ. Then $T^*v = \overline{\lambda} v$.

- Warning: the above lemma is only true for normal operators! For other linear transformations, it is quite possible that T and T^* have totally different eigenvectors and eigenvalues.

- **Proof** To show $T^*v = \overline{\lambda} v$, it suffices to show that $\|T^*v - \overline{\lambda} v\| = 0$, which in turn will follow if we can show that

$$\langle T^*v - \overline{\lambda} v, T^*v - \overline{\lambda} v \rangle = 0.$$

  We expand out the left-hand side as

$$\langle T^*v, T^*v \rangle - \langle \overline{\lambda} v, T^*v \rangle - \langle T^*v, \overline{\lambda} v \rangle + \langle \overline{\lambda} v, \overline{\lambda} v \rangle.$$

  Pulling the λs out and swapping the Ts over, this becomes

$$\langle v, TT^*v \rangle - \overline{\lambda} \langle Tv, v \rangle - \lambda \langle v, Tv \rangle + \lambda \overline{\lambda} \langle v, v \rangle.$$

Since T is normal and $Tv = \lambda v$, we have $T^*v = T^*Tv = \lambda T^*v$. Thus we can rewrite this expression as

$$\overline{\lambda}\langle v, T^*v\rangle - \lambda\overline{\lambda}\langle v, v\rangle - \lambda\overline{\lambda}\langle v, v\rangle + \lambda\overline{\lambda}\langle v, v\rangle.$$

But $\langle v, T^*v\rangle = \langle Tv, v\rangle = \lambda\langle v, v\rangle$. If we insert this in the above expression we then see that everything cancels to zero, as desired. □

- **Lemma 7.** Let $T : V \to V$ be normal, and let v_1, v_2 be two eigenvectors of T with *distinct* eigenvalues λ_1, λ_2. Then v_1 and v_2 must be orthogonal.

- (Compare this with Proposition 6 of the Week 8 notes, which merely asserts that these vectors v_1 and v_2 are linearly independent. Again, we caution that this orthogonality of eigenvectors is only true for normal operators.)

- **Proof.** We have $Tv_1 = \lambda_1 v_1$ and $Tv_2 = \lambda_2 v_2$. By Lemma 6 we thus have $T^*v_1 = \overline{\lambda_1}v_1$ and $T^*v_2 = \overline{\lambda_2}v_2$. Thus

$$\lambda_1\langle v_1, v_2\rangle = \langle Tv_1, v_2\rangle = \langle v_1, T^*v_2\rangle = \langle v_1, \overline{\lambda_2}v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

Since $\lambda_1 \neq \lambda_2$, this means that $\langle v_1, v_2\rangle = 0$, and so v_1 and v_2 are orthogonal as desired. □

- This lemma tells us that most linear transformations will not be normal, because in general the eigenvectors corresponding to different eigenvalues will not be orthogonal. (Take for instance the matrix involved in the Fibonacci rabbit example).

- In the other direction, if we have an orthonormal basis of eigenvectors, then the transformation must be normal:

- **Lemma 8.** Let $T : V \to V$ be a linear transformation o an inner product space V, and let β be an orthonormal basis which consists entirely of eigenvectors of T. Then T is normal.

- Compare this lemma to Lemma 2 of Week 7 notes, which sais that if you have a basis of eigenvectors (not necessarily orthonormal), then T is diagonalizable.

- **Proof.** From Lemma 2 of Week 7 notes, we know that the matrix $[T]_\beta^\beta$ is diagonal. But all diagonal matrices are normal (why?), and so $[T]_\beta^\beta$ is normal. By Proposition 5 we thus see that T is normal. $\square$

- We now come to an important theorem, that the converse of Lemma 8 is also true:

- **Spectral theorem for normal operators** Let $T : V \to V$ be a normal linear transformation on a complex finite dimensional inner product space V. Then there is an orthonormal basis β consisting entirely of eigenvectors of T. In particular, T is diagonalizable.

- Thus normal linear transformations are precisely those diagonalizable linear transformations which can be diagonalized using orthonormal bases (as opposed to just being plain diagonalizable, using bases which might not be orthonormal).

- There is also a spectral theorem for normal operators on infinite dimensional inner product spaces, but it is beyond the scope of this course.

- **Proof** Let the dimension of V be n. We shall prove this theorem by induction on n.

- First consider the base case $n = 1$. Then one can pick any orthonormal basis β of V (which in this case will just be a single unit vector), and the vector v in this basis will automatically be an eigenvector of T (because in a one-dimensional space every vector will be a scalar multiple of v). So the spectral theorem is trivially true when $n = 1$.

- Now suppose inductively that $n > 1$, and that the theorem has already been proven for dimension $n - 1$. Let $f(\lambda)$ be the characteristic polynomial of T (or of any matrix representation $[T]_\beta^\beta$ of T; recall that any two such matrix representations are similar and thus have the same characteristic polynomial). From the fundamental theorem of algebra, we know that this characteristic polynomial splits over the complex numbers. Hence there must be at least one root of this polynomial, and hence T has at least one (complex) eigenvalue, and hence at least one eigenvector.

- So now let us pick an eigenvector v_1 of T with eigenvalue λ_1, thus $Tv_1 = \lambda_1 v_1$ and $T^* v_1 = \overline{\lambda_1} v_1$ by Lemma 6. We can normalize v_1 to have length 1, so $\|v_1\| = 1$ (remember that if you multiply an eigenvector by a non-zero scalar you still get an eigenvector, so it's safe to normalize eigenvectors). Let $W := \{cv_1 : c \in \mathbf{C}\}$ denote the span of this eigenvector, thus W is a one-dimensional space. Let $W^\perp := \{v \in V : v \perp v_1\}$ denote the orthogonal complement of W; this is thus an $n-1$ dimensional space.

- Now we see what T and T^* do to $W^\perp$. Let w be any vector in $W^\perp$, thus $w \perp v_1$, i.e. $\langle w, v_1 \rangle = 0$. Then

$$\langle Tw, v_1 \rangle = \langle w, T^* v_1 \rangle = \langle w, \overline{\lambda_1} v_1 \rangle = \lambda_1 \langle w, v_1 \rangle = 0$$

and similarly

$$\langle T^* w, v_1 \rangle = \langle w, Tv_1 \rangle = \langle w, \lambda_1 v_1 \rangle = \overline{\lambda_1} \langle w, v_1 \rangle = 0.$$

Thus if $w \in W^\perp$, then Tw and $T^* w$ are also in $W^\perp$. Thus T and T^* are not only linear transformations from V to V, they are also linear transformations from $W^\perp$ to $W^\perp$. Also, we have

$$\langle Tw, w' \rangle = \langle w, T^* w' \rangle$$

for all $w, w' \in W^\perp$, because every vector in $W^\perp$ is a vector in V, and we already have this property for vectors in V. Thus T and T^* are still adjoints of each other even after we restrict the vector space from the n-dimensional space V to the $n-1$-dimensional space $W^\perp$.

- We now apply the induction hypothesis, and find that $W^\perp$ enjoys an orthonormal basis of eigenvectors of T. There are $n-1$ such eigenvectors, since $W^\perp$ is $n-1$ dimensional. Now v_1 is normalized and is orthogonal to all the vectors in this basis, since v_1 lies in W and all the other vectors lie in $W^\perp$. Thus if we add v_1 to this basis we get a new collection of n orthonormal vectors, which automatically form a basis by Corollary 5 of Week 9 notes. Each of these vectors is an eigenvector of T, and so we are done. $\square$

- **Example** The linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) := (y, -x)$ that we discussed earlier is normal, but not diagonalizable (its characteristic polynomial is $\lambda^2 + 1$, which doesn't split over the reals). This does not contradict the spectral theorem because that only concerns complex inner product spaces. If however we consider the complex linear transformation $T : \mathbf{C}^2 \to \mathbf{C}^2$ defined by $T(z, w) := (w, -z)$, then we can find an orthonormal basis of eigenvectors, namely

$$v_1 := \frac{1}{\sqrt{2}}(1, i); \quad v_2 := \frac{1}{\sqrt{2}}(1, -i)$$

(Exercise: cover up the above line and see if you can find these eigenvectors on your own). Indeed, you can check that v_1 and v_2 are orthonormal, and that $Tv_1 = iv_1$ and $Tv_2 = -iv_2$. Thus we can diagonalize T using an orthonormal basis, to become the diagonal matrix $\mathrm{diag}(i, -i)$.

* * * * *

Self-adjoint operators

- To summarize the previous section: in the world of complex inner product spaces, normal linear transformations (aka normal operators) are the best kind of linear transformations: they are not only diagonalizable, but they are diagonalizable using the best kind of basis, namely an orthonormal basis. However, there is a subclass of normal transformations which are even better: the *self-adjoint* transformations.

- **Definition** A linear transformation $T : V \to V$ on a finite-dimensional inner product space V is said to be *self-adjoint* if $T^* = T$, i.e. T is its own adjoint. A square matrix A is said to be *self-adjoint* if $A^\dagger = A$, i.e. A is its own adjoint.

- **Example.** The linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) := (y, -x)$ is normal, but not self-adjoint, because its adjoint $T^*(x, y) = (-y, x)$ is not the same as T. However, the linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) = (y, x)$ is self-adjoint, because its adjoint is given by $T^*(x, y) = (y, x)$ (why?), and this is the same as T.

- **Example.** The matrix

$$A := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

is normal, but not self-adjoint, because its adjoint

$$A^\dagger = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is not the same as A. However, the matrix

$$A := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is normal, but not self-adjoint, because its adjoint

$$A^\dagger = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is the same as A. (Why does this example correspond to the preceding one? It is easy to check, using Proposition 5, that a linear transformation is self-adjoint if and only if its matrix in some orthonormal basis is self-adjoint).

- **Example.** Every real diagonal matrix is self-adjoint, but any other type of diagonal matrix is not (e.g. $\text{diag}(2+i, 4+3i)$ has an adjoint of $\text{diag}(2-i, 4-3i)$ and is hence not self-adjoint, though it is still normal).

- It is clear that all self-adjoint linear transformations are normal, since if $T^* = T$ then T^*T and TT^* are both equal to T^2 and are hence equal to each other. Similarly, every self-adjoint matrix is normal. However, not every normal matrix is self-adjoint, and not every normal linear transformation is self-adjoint; see the above examples.

- A self-adjoint transformation over a complex inner product space is sometimes known as a *Hermitian* transformation. A self-adjoint transformation over a real inner product space is known as a *symmetric* transformation. Similarly, a complex self-adjoint matrix is known as a *Hermitian* matrix, while a real self-adjoint matrix is known as a *symmetric* matrix. (A matrix is *symmetric* if $A^t = A$. When the matrix is real, the transpose A^t is the same as the adjoint, thus self-adjoint and symmetric have the same meaning for real matrices, but not for complex matrices).

- **Example** The matrix
$$A := \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$
is its own adjoint (why?), and is hence Hermitian, but it is not symmetric, since it is not its own transpose. Note that every real symmetric matrix is automatically Hermitian, because every real matrix is also a complex matrix (with all the imaginary parts equal to 0).

- From the spectral theorem for normal matrices, we know that any Hermitian operator on a complex inner product space has an orthonormal basis of eigenvectors. But we can say a little bit more:

- **Theorem 9** All the eigenvalues of a Hermitian operator are real.

- **Proof.** Let λ be an eigenvalue of a Hermitian operator T, thus $Tv = \lambda v$ for some non-zero eigenvector v. But then by Lemma 6, $T^*v = \overline{\lambda}v$. But since T is Hermitian, $T = T^*$, and hence $\lambda v = \overline{\lambda}v$. Since v is non-zero, this means that $\lambda = \overline{\lambda}$, i.e. λ is real. Thus all the eigenvalues of T are real. $\square$

- A similar line of reasoning shows that all the eigenvalues of a Hermitian matrix are real.

- **Corollary 10.** The characteristic polynomial of a Hermitian matrix splits over the reals.

- **Proof.** We know already from the Fundamental Theorem of Algebra that the characteristic polynomial splits over the complex numbers. But since the matrix is Hermitian, every root of the characteristic polynomial must be real. Thus the polynomial must split over the reals. $\square$

- We can now prove

- **Spectral theorem for self-adjoint operators** Let T be a self-adjoint linear transformation on an inner product space V (which can be either real or complex). Then there is an orthonormal basis of V which consists entirely of eigenvectors of V, with *real* eigenvalues.

- **Proof.** We repeat the proof of the Spectral theorem for normal operators, i.e. we do an induction on the dimension n of the space V. When $n = 1$ the claim is again trivial (and we use the fact that every Lemma 9 to make sure the eigenvalue is real). Now suppose inductively that $n > 1$ and the claim has already been proven for $n - 1$.

  From Corollary 10 we know that T has at least one *real* eigenvalue. Thus we can find a real λ_1 and a non-zero vector v_1 such that $Tv_1 = \lambda_1 v_1$. We can then normalize v_1 to have unit length. We now repeat the rest of the proof of the spectral theorem for normal operators, to obtain the same conclusion except that the eigenvalues are now real. $\square$

- Notice one subtle difference between the spectral theorem for self-adjoint operators and the spectral theorem for normal operators: the spectral theorem for normal operators requires the inner product space to be complex, but the one for self-adjoint operators does not. In particular, every symmetric operator on a real vector space is diagonalizable.

- **Example** The matrix
$$A := \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$
  is Hermitian, and thus so is the linear transformation $L_A : \mathbf{C}^2 \to \mathbf{C}^2$, which is given by
$$L_A \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} iw \\ -iz \end{pmatrix}.$$
  By the spectral theorem, $\mathbf{C}^2$ must have an orthonormal basis of eigenvectors with real eigenvalues. One such basis is
$$v_1 := \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix}; \quad v_2 := \begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix};$$
  one can verify that v_1 and v_2 are an orthonormal basis for the complex two-dimensional inner product space $\mathbf{C}^2$, and that $L_A v_1 = -v_1$ and $L_A v_2 = +v_2$. Thus L_A can be diagonalized using an orthonormal basis to give the matrix $\mathrm{diag}(+1, -1)$. Note that while the eigen*values* of L_A are real, the eigen*vectors* are still complex. The spectral theorem says nothing as to how real or complex the eigenvectors are (indeed,

in many inner product spaces, such a question does not really make sense).

- Self-adjoint operators are thus the very best of all operators: not only are they diagonalizable, with an orthonormal basis of eigenvectors, the eigenvalues are also real. (Conversely, it is easy to modify Lemma 8 to show that any operator with these properties is necessarily self-adjoint). Fortunately, self-adjoint operators come up all over the place in real life. For instance, in quantum mechanics, almost all the linear transformations one sees there are Hermitian (this is basically because while quantum mechanics uses complex inner product spaces, the quantities we can actually observe in physical reality must be real-valued).

Assignment 1 Due October 10 Covers: Sections 1.1-1.6

- Q1. Let V be the space of real 3-tuples

$$V = \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \mathbf{R}\}$$

with the standard addition rule

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3)$$

but with the non-standard scalar multiplication rule

$$c(x_1, x_2, x_3) = (cx_1, x_2, x_3).$$

(In other words, V is the same thing as the vector space $\mathbf{R}^3$, but with the scalar multiplication law changed so that the scalar only multiplies the first co-ordinate of the vector.)

Show that V is *not* a vector space.

- Q2. (a) Find a subset of $\mathbf{R}^3$ which is closed under scalar multiplication, but is not closed under vector addition.

(b) Find a subset of $\mathbf{R}^3$ which is closed under vector addition, but not under scalar multiplication.

- Q3. Find three distinct non-zero vectors u, v, w in $\mathbf{R}^3$ such that $\text{span}(\{\text{u}, \text{v}\}) = \text{span}(\{\text{v}, \text{w}\}) = \text{span}(\{\text{u}, \text{v}, \text{w}\})$, but such that $\text{span}(\{\text{u}, \text{w}\}) \neq \text{span}(\{\text{u}, \text{v}, \text{w}\})$.

- Q4. Find a basis for $M^0_{2\times 2}(\mathbf{R})$, the vector space of 2×2 matrices with trace zero. Explain why the set you chose is indeed a basis.

- Q5. Do Exercise 1(abghk) of Section 1.2 in the textbook.

- Q6. Do Exercise 8(aef) of Section 1.2 in the textbook.

- Q7. Do Exercise 23 of Section 1.3 in the textbook.

- Q8*. Do Exercise 19 of Section 1.3 in the textbook. [**Hint**: Prove by contradiction. If $W_1 \not\subseteq W_2$, then there must be a vector w_1 which lies in W_1 but not in W_2. Similarly, if $W_2 \not\subseteq W_1$, then there must be a vector w_2 which lies in W_2 but not in W_1. Now suppose that both $W_1 \not\subseteq W_2$ and $W_2 \not\subseteq W_1$, and consider what one can say about the vector $w_1 + w_2$.]

- Q9. Do Exercise 4(a) of Section 1.4 in the textbook.
- Q10. Do Exercise 1(abdef) of Section 1.5 in the textbook.

Assignment 2 Due October 17 Covers: Sections 1.6-2.1

- Q1. Do Exercise 1(acdejk) of Section 1.6 in the textbook.
- Q2. Do Exercise 3(b) of Section 1.6 in the textbook.
- Q3. Do Exercise 7 of Section 2.1 in the textbook.
- Q4. Do Exercise 9 of Section 2.1 in the textbook.
- Q5. Find a polynomial $f(x)$ of degree at most three, such that $f(n) = 2^n$ for all $n = 0, 1, 2, 3$.
- Q6. Let V be a vector space, and let A, B be two subsets of V. Suppose that B spans V, and that span(A) contains B. Show that A spans V.
- Q7*. Let V be a vector space which is spanned by a finite set S of n elements. Show that V is finite dimensional, with dimension less than or equal to n. [Note: You cannot apply the Dimension Theorem directly, because we have not assumed that V is finite dimensional. To do that, we must first construct a finite basis for V; this can be done by modifying the proof of part (g) of the Dimension theorem, or the proof of Theorem 2.]
- Q8*. Show that $\mathcal{F}(\mathbf{R}, \mathbf{R})$, the space of functions from $\mathbf{R}$ to $\mathbf{R}$, is infinite-dimensional.
- Q9. Let V be a vector space of dimension 5, and let W be a subspace of V of dimension 3. Show that there exists a vector space U of dimension 4 such that $W \subset U \subset V$.
- Q10. Let $v_1 := (1, 0, 0)$, $v_2 := (0, 1, 0)$, $v_3 := (0, 0, 1)$, $v_4 := (1, 1, 0)$ be four vectors in $\mathbf{R}^3$, and let S denote the set $S := \{v_1, v_2, v_3, v_4\}$. The set S has 16 subsets, which are depicted on the reverse of this assignment. (This graph, incidentally, depicts (the shadow of) a *tesseract*, or 4-dimensional cube).
- Of these subsets, which ones span $\mathbf{R}^3$? which ones are linearly independent? Which ones are bases? (Feel free to color in the graph and turn it in with your assignment. You may find Corollary 1 in the Week 2 notes handy).

Assignment 3 Due October 24 Covers: Sections 2.1-2.3

- Q1. Do Exercise 1(cdfh) of Section 2.1 of the textbook.
- Q2. Do Exercise 7 of Section 2.1 of the textbook.
- Q3. Do Exercise 10 of Section 2.1 of the textbook.
- Q4*. Do Exercise 17 of Section 2.1 of the textbook.
- Q5. Do Exercise 1(bcdf) of Section 2.2 of the textbook.
- Q6. Do Exercise 2(aceg) of Section 2.2 of the textbook.
- Q7. Do Exercise 7 of Section 2.2 of the textbook.
- Q8. (a) Let V, W be vector spaces, and let $T : V \to W$ be a linear transformation. Let U be a subspace of W. Show that the set

  $$T^{-1}(U) := \{v \in V : T(v) \in U\}$$

  is a subspace of V. Explain why this shows that the null space $N(T)$ is also a subspace.
- (b) Let V, W be vector spaces, and let $T : V \to W$ be a linear transformation. Let X be a subspace of V. Show that the set

  $$T(X) := \{Tv : v \in X\}$$

  is a subspace of W. Explain why this shows that the range $R(T)$ is also a subspace.
- Q9*. Show, *without doing Gaussian elimination or any other computation*, that there must be a solution to the system

  $$\begin{array}{lllll} 12x_1 & +34x_2 & +56x_3 & +78x_4 & = 0 \\ 3x_1 & +6x_2 & +2x_3 & +10x_4 & = 0 \\ 43x_1 & +21x_2 & +98x_3 & +76x_4 & = 0 \end{array}$$

  such that the x_1, x_2, x_3, x_4 are not all equal to zero. [**Hint:** consider the linear transformation $T : \mathbf{R}^4 \to \mathbf{R}^3$ defined by

$$T(x_1, x_2, x_3, x_4) := (12x_1 + 34x_2 + 56x_3 + 78x_4,$$

$$3x_1 + 6x_2 + 2x_2 + 10x_4, 43x_1 + 21x_2 + 98x_3 + 76x_4).$$

What can you say about the rank and nullity of T?

- Q10. Find a *non-zero* vector $v \in \mathbf{R}^2$, and two ordered bases β, β' of $\mathbf{R}^2$, such that $[v]_\beta = [v]_{\beta'}$ but that $\beta \neq \beta'$.

Assignment 4 Due October 31 Covers: Sections 2.3-2.4

- Q1. Do Exercise 5(cdefg) of Section 2.2 of the textbook.
- Q2. Do Exercise 1(aegij) of Section 2.3 of the textbook.
- Q3. Do Exercise 4(c) of Section 2.3 of the textbook.
- Q4. Do Exercise 10 of Section 2.3 at the textbook. (T_0 is the zero transformation, so that $T_0 v = 0$ for all $v \in V$.
- Q5. Do Exercise 1(bcdefhi) of Section 2.4 of the textbook.
- Q6. Do Exercise 2 of Section 2.4 of the textbook.
- Q7. Do Exercise 4 of Section 2.4 of the textbook.
- Q8*. Do Exercise 9 of Section 2.4 of the textbook.
- Q9. Let U, V, W be vector spaces.
- (a) Show that U is isomorphic to U.
- (b) Show that if U is isomorphic to V, then V is isomorphic to U.
- (c) Show that if U is isomorphic to V, and V is isomorphic to W, then U is isomorphic to W.
- (Incidentally, the above three properties (a)-(c) together mean that isomorphism is an *equivalence relation*).
- Q10. From our notes on Lagrange interpolation, we know that given any three numbers y_1, y_2, y_3, there exists an interpolating polynomial $f \in P_2(\mathbf{R})$ such that $f(0) = y_1$, $f(1) = y_2$, and $f(2) = y_3$. Define the map $T : \mathbf{R}^3 \to P_2(\mathbf{R})$ by setting $T(y_1, y_2, y_3) := f$. (Thus for instance $T(0, 1, 4) = x^2$). Let $\alpha := ((1, 0, 0), (0, 1, 0), (0, 0, 1))$ be the standard basis for $\mathbf{R}^3$, and let $\beta := (1, x, x^2)$ be the standard basis for $P_2(\mathbf{R})$.
- (a) Compute the matrix $[T]_\alpha^\beta$. (You may assume without proof that T is linear.)

- (b) Let $S : P_2(\mathbf{R}) \to \mathbf{R}^3$ be the map

$$Sf := (f(0), f(1), f(2)).$$

Compute the matrix $[S]_\beta^\alpha$. (Again, you may assume without proof that S is linear).

- (c) Use matrix multiplication to verify the identities

$$[S]_\beta^\alpha [T]_\alpha^\beta = [T]_\alpha^\beta [S]_\beta^\alpha = I_3,$$

where I_3 is the 3×3 identity matrix. Can you explain why these identities should be true?

Midterm

- Q1. Let $T : \mathbf{R}^4 \to \mathbf{R}^4$ be the transformation

$$T(x_1, x_2, x_3, x_4) := (0, x_1, x_2, x_3).$$

- (a) What is the rank and nullity of T?
- (b) Let $\beta := ((1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1))$ be the standard ordered basis for T. Compute $[T]_\beta^\beta$, $[T^2]_\beta^\beta$, $[T^3]_\beta^\beta$, and $[T^4]_\beta^\beta$. (Here $T^2 = T \circ T$, $T^3 = T \circ T \circ T$, etc.)
- Q2. Let V denote the space

$$V := \{f \in P_3(\mathbf{R}) : f(0) = f(1) = 0\}.$$

  (a) Show that V is a vector space.

  (b) Find a basis for V. (Hint: if $f(0) = f(1) = 0$, what can one say about the factors of f?)

- Q3. Let V and W be vector spaces, and let $T : V \to W$ be a *one-to-one* linear transformation. Let U be a finite-dimensional subspace of V. Show that the vector space

$$T(U) := \{Tv : v \in U\}$$

  has the same dimension as U. (You may assume without proof that $T(U)$ is a vector space).

- Q4. Let V be a three-dimensional vector space with an ordered basis $\beta := (v_1, v_2, v_3)$. Let γ be the ordered basis $\gamma := ((1,1,0), (1,0,0), (0,0,1))$ of $\mathbf{R}^3$. (You may assume without proof that γ is indeed an ordered basis).
- Let $T : V \to \mathbf{R}^3$ be a linear transformation whose matrix representation $[T]_\beta^\gamma$ is given by

$$[T]_\beta^\gamma = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

  Compute $T(v_1 + 2v_2 + 3v_3)$.

- Q5. Find a linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ whose null space $N(T)$ is equal to the z-axis

$$N(T) = \{(0, 0, z) : z \in \mathbf{R}\}$$

and whose range $R(T)$ is equal to the plane

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

Assignment 5 Due November 7 Covers: Sections 2.4-2.5

- Q1. Do exercise 15(b) of Section 2.4 of the textbook. (Note that part (a) of this exercise was already done in Q8(b) of Assignment 3).
- Q2. Do exercise 1(abde) of Section 2.5 in the textbook.
- Q3. Do exercise 2(b) of Section 2.5 in the textbook.
- Q4. Do exercise 4 of Section 2.5 in the textbook.
- Q5. Do exercise 10 of Section 2.5 in the textbook.
- Q6. Let $\beta := ((1,0),(0,1))$ be the standard basis of $\mathbf{R}^2$, and let $\beta' := ((3,-4),(4,3))$ be another basis of $\mathbf{R}^2$. Let l be the line connecting the origin to $(4,3)$, and let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the operation of reflection through l (so if $v \in \mathbf{R}^2$, then Tv is the reflected image of v through the line l.
- (a) What is $[T]_{\beta'}^{\beta'}$? (You should do this entirely by drawing pictures).
- (b) Use the change of variables formula to determine $[T]_{\beta}^{\beta}$.
- (c) If $(x,y) \in \mathbf{R}^2$, give a formula for $T(x,y)$.
- Q7. Let $T : P_n(\mathbf{R}) \to \mathbf{R}^{n+1}$ be the map

  $$T(f) := (f(0), f(1), f(2), \ldots, f(n)).$$

  Thus, for instance if $n = 3$, then $T(x^2) = (0,1,4,9)$.
- (a) Prove that T is linear.
- (b) Prove that T is an isomorphism.
- Q8*. Let A, B be $n \times n$ matrices such that $AB = I_n$, where I_n is the $n \times n$ identity matrix.
- (a) Show that $L_A L_B = I_{\mathbf{R}^n}$, where $I_{\mathbf{R}^n}$ the identity on $\mathbf{R}^n$.
- (b) Show that L_B is one-to-one and onto. (**Hint:** Use (a) to obtain the one-to-one property. Then use the Dimension theorem to deduce the onto property).

- (c) Show that $L_B L_A = I_{\mathbf{R}^n}$. (**Hint:** First use (a) to show that $L_B L_A L_B = L_B$, and then use the fact that L_B is onto).
- (d) Show that $BA = I_n$.
- (To summarize the result of this problem: if one wants to show that two $n \times n$ matrices A, B are inverses, one only needs to show $AB = I_n$; the other condition $BA = I_n$ comes for free).
- Q9. Let A, B, C be $n \times n$ matrices.
- (a) Show that A is similar to A.
- (b) Show that if A is similar to B, then B is similar to A.
- (c) Show that if A is similar to B, and B is similar to C, then A is isomorphic to C.
- (Incidentally, the above three properties (a)-(c) together mean that similarity is an *equivalence relation*).
- Q10*. Let V be a finite-dimensional vector space, let $T : V \to V$ be a linear transformation, and let $S : V \to V$ be an invertible linear transformation.
- (a) Prove that $\mathbf{R}(STS^{-1}) = S(\mathbf{R}(T))$ and $\mathbf{N}(STS^{-1}) = S(\mathbf{N}(\mathbf{T}))$. (Recall that $\mathbf{R}(T) := \{Tv : v \in V\}$ is the range of T, while $\mathbf{N}(T) := \{v \in V : Tv = 0\}$ is the null space of T. Also, for any subspace W of V, recall that $S(W) := \{Sv : v \in W\}$ is the image of W under S.)
- (b) Prove that $\text{rank}(\mathbf{R}(T)) = \text{rank}(\mathbf{R}(STS^{-1}))$ and $\text{nullity}(\mathbf{R}(T)) = \text{nullity}(\mathbf{R}(STS^{-1}))$. (**Hint:** use part (a) as well as Q1).

Assignment 6 Due November 14 Covers: Sections 3.1-3.2; 4.1-4.4

- Q1. Do Question 1(abcdefgi) of Section 3.1 of the textbook.
- Q2. Do Question 1(acdefhi) of Section 3.2 of the textbook.
- Q3. Do Question 5(e) of Section 3.2 of the textbook.
- Q4. Do Question 6(ade) of Section 3.2 of the textbook.
- Q5. Do Question 1(abcdefgh) of Section 4.2 of the textbook.
- Q6. Do Question 25 of Section 4.2 of the textbook.
- Q7. Let U, V, W be finite-dimensional vector spaces, and let $S : V \to W$ and $T : U \to V$ be linear transformations.
- (a) Show that rank(ST) $\leq$ rank(S).
- (b) Show that rank(ST) $\leq$ rank(T).
- (c) Show that nullity(ST) $\geq$ nullity(T).
- (d) Give an example where nullity(ST) $>$ nullity(S).
- (e) Give an example where nullity(ST) $<$ nullity(S).
- Q8. Let A and B be $n \times n$ matrices. Prove (from the definition of transpose and matrix multiplication) that $(AB)^t = B^t A^t$.
- Q9. Let A be an invertible $n \times n$ matrix. Prove that $\det(A^{-1}) = 1/\det(A)$.
- Q10. Let A and B be invertible $n \times n$ matrices. Show that one there is a sequence of elementary row operations which transforms A to B. (Hint: first show that there is a sequence of row operations which transforms A to the identity matrix).

Assignment 7 Due November 21 Covers: Sections 4.4; 5.1-5.2

- Q1. Do Question 1(acdefgijk) of Question 1 of Section 4.4 of the textbook.
- Q2. Do Question 4(g) of Section 4.4 of the textbook.
- Q3. Do Question 2(a) of Section 5.1 of the textbook.
- Q4. Do Question 3(a) of Section 5.1 of the textbook. (Treat any occurrence of F as if it were $\mathbf{R}$ instead).
- Q5. Do Question 8 of Section 5.1 of the textbook. (You may assume that $T : V \to V$ is a linear transformation from some finite-dimensional vector space V to itself; this is what it means for T to be a linear transformation "on V).
- Q6*. Do Question 11 of Section 5.1 of the textbook.
- Q7. Do Question 15 of Section 5.1 of the textbook.
- Q8. Do Question 3(bf) of Section 5.2 of the textbook.
- Q9. Let A and B be similar $n \times n$ matrices. Show that A and B have the same set of eigenvalues (i.e. every eigenvalue of A is also an eigenvalue of B and vice versa).
- Q10*. For this question, the field of scalars will be the complex numbers $\mathbf{C} := \{x + yi : x, y \in \mathbf{R}\}$ instead of the reals (i.e. all matrices, etc. are allowed to have complex entries). Let θ be a real number, and let A be the 2×2 rotation matrix

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

- (a) Show that A has eigenvalues $e^{i\theta}$ and $e^{-i\theta}$. (You may use Euler formula $e^{i\theta} = \cos\theta + i\sin\theta$). What are the eigenvectors corresponding to $e^{i\theta}$ and $e^{-i\theta}$?

- (b) Write $A = QDQ^{-1}$ for some invertible matrix Q and diagonal matrix D (note that Q and D may have complex entries. Also, there are several possible answers to this question; you only need to give one of them).

- (c) Let $n \geq 1$ be an integer. Prove that

$$A^n = \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix}.$$

  (You may find the formulae $(e^{i\theta})^n = e^{in\theta} = \cos n\theta + i\sin n\theta$ and $(e^{-i\theta})^n = e^{-in\theta} = \cos n\theta - i\sin n\theta$ to be useful).

- (d) Can you explain why the operator $L_A : \mathbf{R}^2 \to \mathbf{R}^2$ corresponds to an anti-clockwise rotation of the plane $\mathbf{R}^2$ by angle θ?

- (e) Based on (d), can you think of a geometrical interpretation of the result proven in (c)?

Assignment 8 Due December 5 Covers: Sections 5.2,6.2-6.3

- Q1. Do Question 8 of Section 6.1 in the textbook.
- Q2. Do Question 11 of Section 6.1 in the textbook.
- Q3. Do Question 4 of Section 6.2 in the textbook.
- Q4. Do Question 13(a) of Section 6.2 in the textbook. (**Hint:** Use Theorem 6 from the Week 9 notes).
- Q5. Do Question 17(bc) of Section 6.2 of the textbook.
- Q6. Do Question 18(b) of Section 6.2 of the textbook.
- Q7. Do Question 2 of Section 6.3 of the textbook.
- Q8. Let A be an $n \times n$ matrix with n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that $\det(A) = \lambda_1\lambda_2 \ldots \lambda_n$ and $\mathrm{tr(A)} = \lambda_1 + \lambda_2 + \ldots + \lambda_\mathrm{n}$.
- Q9. Let V be a finite-dimensional inner product space, and let W be a subspace of V. Show that $(W^\perp)^\perp = W$; i.e. the orthogonal complement of the orthogonal complement of W is again W.
- Q10*. Find a 2×2 matrix A which has $(1, 1)$ and $(1, 0)$ as eigenvectors, is not equal to the identity matrix, and is such that $A^2 = I_2$, where I_2 is the 2×2 identity matrix. (Hint: you might want to use Q7 from last week's homework to work out what the eigenvalues of A must be).

Mathematics 115A/3
Terence Tao
Final Examination, Dec 10, 2002

Problem 1. (15 points) Let W be a finite-dimensional real vector space, and let U and V be two subspaces of W. Let $U+V$ be the space

$$U+V := \{u+v : u \in U \text{ and } v \in V\}.$$

You may use without proof the fact that $U+V$ is a subspace of W.

(a) (5 points) Show that $\dim(U+V) \leq \dim(U) + \dim(V)$.

(b) (5 points) Suppose we make the additional assumption that $U \cap V = \{0\}$. Now prove that $\dim(U+V) = \dim(U) + \dim(V)$.

Problem 1 continues on the next page.

Problem 1 continued.

(c) (5 points) Let U and V be two three-dimensional subspaces of $\mathbf{R}^5$. Show that there exists a non-zero vector $v \in \mathbf{R}^5$ which lies in both U and V. (Hint: Use (b) and argue by contradiction).

Problem 2. (10 points) Let $P_2(\mathbf{R})$ be the space of polynomials of degree at most 2, with real coefficients. We give $P_2(\mathbf{R})$ the inner product

$$\langle f, g\rangle := \int_0^1 f(x)g(x)\ dx.$$

You may use without proof the fact that this is indeed an inner product for $P_2(\mathbf{R})$.

(a) (5 points) Find an orthonormal basis for $P_2(\mathbf{R})$.

Ans.

(b) (5 points) Find a basis for $\mathrm{span}(1, \mathrm{x})^{\perp}$.

Ans.

Problem 3. (15 points) Let $P_3(\mathbf{R})$ be the space of polynomials of degree at most 3, with real coefficients. Let $T : P_3(\mathbf{R}) \to P_3(\mathbf{R})$ be the linear transformation

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

thus for instance $T(x^3 + 2x) = 3x^2 + 2$. You may use without proof the fact that T is indeed a linear transformation. Let $\beta := (1, x, x^2, x^3)$ be the standard basis for $P_3(\mathbf{R})$.

(a) (5 points) Compute the matrix $[T]_\beta^\beta$.

Ans.

(b) (3 points) Compute the characteristic polynomial of $[T]_\beta^\beta$.

Ans.

(c) (5 points) What are the eigenvalues and eigenvectors of T? (Warning: the eigenvectors of T are related to, but not quite the same as, the eigenvectors of $[T]_\beta^\beta$.

Ans.

Problem 3 continues on the next page.

Problem 3 continued.
(d) (2 points) Is T diagonalizable? Explain your reasoning.

Problem 4. (15 points) This question is concerned with the linear transformation $T : \mathbf{R}^4 \to \mathbf{R}^3$ defined by

$$T(x, y, z, w) := (x + y + z, y + 2z + 3w, x - z - 2w).$$

You may use without proof the fact that T is a linear transformation.

(a) (5 points) What is the nullity of T?

Ans.

(b) (5 points) Find a basis for the null space. (This basis does *not* need to be orthogonal or orthonormal).

Ans.

(c) (5 points) Find a basis for the range. (This basis does *not* need to be orthogonal or orthonormal).

Ans.

Problem 5. (10 points) Let V be a real vector space, and let $T : V \to V$ be a linear transformation such that $T^2 = T$. Let $R(T)$ be the range of T and let $N(T)$ be the null space of T.

(a) (5 points) Prove that $R(T) \cap N(T) = \{0\}$.

(b) (5 points) Let $R(T) + N(T)$ denote the space

$$R(T) + N(T) := \{x + y : x \in R(T) \text{ and } y \in N(T)\}.$$

Show that $R(T) + N(T) = V$. (**Hint:** First show that for any vector $v \in V$, the vector $v - Tv$ lies in the null space $N(T)$).

Problem 6. (15 points) Let A be the matrix

$$A := \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

(a) (5 points) Find a complex invertible matrix Q and a complex diagonal matrix D such that $A = QDQ^{-1}$. (Hint: A has -1 as one of its eigenvalues).

Ans.

Problem 6 continues on the next page.

Problem 6 continued.

(b) (5 points) Find three elementary matrices E_1, E_2, E_3 such that $A = E_1E_2E_3$. (Note: this problem is not directly related to (a)).

Ans.

(c) (5 points) Compute A^{-1}, by any means you wish.

Ans.

Problem 7. (10 points) Let f, g be continuous, complex-valued functions on $[-1, 1]$ such that $\int_{-1}^{1} |f(x)|^2 \, dx = 9$ and $\int_{-1}^{1} |g(x)|^2 \, dx = 16$.

(a) (5 points) What possible values can $\int_{-1}^{1} f(x)\overline{g(x)} \, dx$ take? Explain your reasoning.

Ans.

(b) (5 points) What possible values can $\int_{-1}^{1} |f(x)+g(x)|^2 \, dx$ take? Explain your reasoning.

Ans.

Problem 8. (10 points) Find a 2×2 matrix A with real entries which has trace 5, determinant 6, and has $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as one of its eigenvectors. (Hint: First work out what the characteristic polynomial of A must be. There are several possible answers to this question; you only have to supply one of them.)

Ans. $\boxed{}$