Linear Algebra

Pete L. Clark

Contents

CHAPTER 1

Basics

1. Some Motivating Examples

Linear algebra is the study of linear systems, matrices, vector spaces and linear transformations. As with most higher mathematics courses, it will take time to present and appreciate these new concepts and objects. Rather than proceeding in a strict logical (linear?) order, I want to begin with some motivational examples: some problems we can solve using linear algebra.

1.1. Example 1: Stoichiometry.

In chemical reactions, adding one compound to another may yield one or more new compounds using the component atoms. We write equations to describe these reactions, but they need to be *balanced*. For example, consider

$$N_2 + H_2 \mapsto NH_3.$$

This reaction cannot happen as written: on the left hand side there are 2 nitrogen atoms and 2 hydrogen atoms, whereas on the right hand side there is only 1 nitrogen atom and there are 3 hydrogen atoms. We need to balance the equation by supplying positive whole number coefficients to make the number of atoms of each element on the left hand side equal the number of atoms of each element on the right hand side: say

$$?N_2 + ?H_2 \mapsto ?NH_3.$$

Notice that we have twice as many nitrogen atoms on the left hand side as the right hand side, so why not multiply the N_2 by 1 and the NH_3 by 2:

$$1N_2 + ?H_2 \mapsto 2NH_3.$$

This balances the nitrogen. What about the hydrogen? Well, no problem: we have 6 on the right so we need six on the left, so the last "?" should be 3:

$$N_2 + 3H_2 \implies 2NH_3.$$

Is it always so easy? No, it isn't, as you know if you've taken chemistry. Here is another example, taken from an online chemistry guide to balancing equations.

$$?\,\mathrm{Zn_3\,Sb_2} + ?H_2O \mapsto ?\,\mathrm{Zn}(OH)_2 + ?\,\mathrm{Sb}\,H_3.$$

Suppose we start by balancing the O:

$$?\,\mathrm{Zn_3}\,Sb_2 + 2H_2O \mapsto 1\,\mathrm{Zn}(OH)_2 + ?\,\mathrm{Sb}\,H_3.$$

And suppose we continue by balancing the Zn (zinc): we get stuck, because whatever positive integer we take for the coefficient of $\mathrm{Zn_3\,Sb_2}$ we'll get at least 3 zinc atoms on the left and we only have one on the right. What do we do??

The handout instructs us to start again with an atom that only appears once on each side of the equation, say Sb.[1] This would lead us to

$$1\,\mathrm{Zn_3\,Sb_2} + ?H_2O \mapsto ?\,\mathrm{Zn}(OH)_2 + 2\,\mathrm{Sb}\,H_3.$$

[1] Chemical trivia question: what element is denoted Sb?!?

Maybe we try the Zn next? We have to put a coefficient of 3 on the right, getting

$$1\,\mathrm{Zn_3\,Sb_2} + ?H_2O \mapsto 3\,\mathrm{Zn}(OH)_2 + 2\,\mathrm{Sb}\,H_3.$$

Finally, look at the H and the O. If we put a 6 on the left, it works out:

$$1Zn_3Sb_2 + 6H_2O \mapsto 3Zn(OH)_2 + 2SbH_3.$$

Well, that was fortunate. The same handout includes the following hints:

1. **Start with** an atom that appears only once on each side of the equation.
2. **Continue choosing** atoms needing only one new coefficient to become balanced.
3. Balance polyatomic ions as a group.[2]
4. Save hydrogen and oxygen atoms for last.
5. **If you get totally stuck**, try doubling (or tripling, etc.) all the coefficients already entered.

Oh, dear. The last bit suggests that we are being given less than a complete recipe. This stoichiometry business looks like more of an art than a science.

Instead, let's be scientific about it. We only need a small change: instead of repeated question marks, let's write variables for the coefficients, say

$$aZn_3Sb_2 + bH_2O \mapsto cZn(OH)_2 + dSbH_3.$$

Now equating the total number of instances of Zinc on both sides gives

$$3a = c.$$

Equating the total number of instances of Sb on both sides gives

$$2a = d.$$

Equating the total number of instances of H on both sides gives

$$2b = 2c + 3d.$$

Equating the total number of instances of O on both sides gives

$$b = 2c.$$

So we get a system of four equations in four unknowns. If only we knew how to solve such things. Well, we'll learn! Note though that this is a pretty *sparse* system of equations. Many of the coefficients are 0:

$$\begin{aligned} 3a + 0b - c + 0d &= 0. \\ 2a + 0b + 0c - d &= 0. \\ 0a + 2b - 2c - 3d &= 0. \\ 0a + b - 2c + 0d &= 0. \end{aligned}$$

Maybe including the zero coefficients looks fastidious. But when we get serious about solving linear systems, we'll see that recopying the variables over and over again is unnecessary and even slightly distracting. So long as we keep the zeros as placeholders, we can just take the coefficients and put them in a rectangular array – a **matrix**:

$$\begin{bmatrix} 3 & 0 & -1 & 0 & 0 \\ 2 & 0 & 0 & -1 & 0 \\ 0 & 2 & -2 & -3 & 0 \\ 0 & 1 & -2 & 0 & 0 \end{bmatrix}.$$

Note that the final column consists entirely of zeros: this is characteristic of **homogeneous linear systems**. Such systems always have a solution: take all variables equal to zero! Here, as usual, we are looking for solutions *other than* the all zero solution.

[2] I'm not sure exactly what that means, but I wanted to give you all the advice.

In fact the zeros in the matrix are not just placeholders but welcome guests. The more zero coefficients in the corresponding matrix, the easier it is to solve the linear system. In this case solving the system certainly doesn't require any special knowledge or ingenuity: two of the variables are simply being given to us in terms of a. Suppose for the sake of argument that $a = 1$. Then we get $c = 3$ and $d = 2$, and using this information we get

$$2b = 2c + 3d = 2 \cdot 3 + 3 \cdot 2 = 12,$$

so

$$b = 6.$$

And now we have one more equation involving b. Luckily it is *consistent* with what we already know:

$$b = 2c = 2 \cdot 3 = 6.$$

Thus

$$(a, b, c, d) = (1, 6, 3, 2)$$

is a solution to the system...exactly the solution we found by hand above. It is not the only solution: no matter what a is we can solve uniquely for b, c and d. In fact we can do this simply by leaving a as is: we get

$$(a, b, c, d) = (a, 6a, 3a, 2a).$$

Notice that this amounts to taking our previous solution and just multiplying it through by a. However the solution with $a = 1$ is the one that the chemists want: the entries need to be positive integers, and we don't want redundancy: mathematically speaking, we don't want all of a, b, c, d to be divisible by any common factor greater than 1.

This simple mathematical analysis is very illuminating. Here are some key points:

I. The entire task is being reduced to solving a system of linear equations. If we know how to do that systematically, balancing equations has no fear for us.

II. We have in fact been given some good advice about how to solve linear systems. In particular, whenever a certain atom appears exactly once on each side, we'll get an equation of the form $\alpha a = \beta b$, where a and b are the variables we're trying to solve for any α and β are positive integers. This tells us that $b = \frac{\beta}{\alpha} a$, i.e., we've eliminated one of the variables from the system of equations, making it that much easier to solve.

III. It seems to be an implicit assumption that the system is close to having a unique solution: namely it has a unique solution if we require the variables to be positive integers without a common factor. This is much less clear, even if for those who have some knowledge in the solution of linear systems. Note for instance that we have four equations in four unknowns. As we will see later, "most of the time" this type of homogeneous system has only the all zero solution, so our stoichiometric system is somewhat atypical. Neither is it clear that we will always have the same number of variables as equations. In fact, the inspiration to motivate linear systems through stoichiometry came from the course text [**SA**], which does so on p. 66. However, their example leads to a system of three equations in four unknowns, which as we will learn later, *always* has a solution apart from the all zero solution.

Could we in fact *prove* that the solutions to these stoichiometric systems always have a unique solution in positive integers with no common factor? Or are there chemical reactions that are "stoichiometrically impossible"? This is an interesting question which we'll come back to later.

1.2. Partial Fractions Decomposition.

Suppose I want to find an antiderivative of the function $\frac{2x+3}{x^3+x}$. In second semester calculus we learn to do this via the method of **partial fractions**, namely we posit an algebraic identity of the form

$$\frac{x^2+2x+3}{x^3+x} = \frac{A}{x} + \frac{Bx+C}{x^2+1}$$

and try to solve it for real numbers A, B, C. How is this done? Well, if we multiply both sides by x^3+x to clear denominators we get

$$x^2+2x+3 = A(x^2+1) + (Bx+C)x = (A+B)x^2 + Cx + A.$$

Now the polynomial on the left will certainly be equal to the polynomial on the right if they are equal coefficient by coefficient (in fact this is the only way for two polynomials with real numbers as coefficients to be equal, as we will probably have occasion to recall later on), so it is enough to enforce

$$A = 3,$$
$$C = 2,$$
$$A + B = 1.$$

Again we get a linear system to solve! (This time the system is **inhomogeneous**: the right hand sides of the equations are not all zero.) And again it's an easier system than the general case, in this case very easy: clearly $A = 3$ and $C = 2$, which tells us that $B = 1 - A = -2$, thus the desired identity is

$$\frac{x^2+2x+3}{x^3+x} = \frac{3}{x} + \frac{-2x+2}{x^2+1},$$

so

$$\begin{aligned}\int \frac{x^2+2x+3}{x^3+x} &= \int \frac{3}{x} - \int \frac{2x}{x^2+1} + 2\int \frac{1}{x^2+1} \\ &= 3\log x - \log(x^2+1) + 2\arctan x + c.\end{aligned}$$

Of course this was a relatively benign example: in general, to integrate a proper rational function $\frac{P(x)}{Q(x)}$ when the denominator Q is a polynomial of degree n (i.e., the highest power of x which appears is x^n), then this method gives us an $n \times n$ system of linear equations to solve. It is not always the case that one can solve the system so easily: sometimes there is still nontrivial work to do. In fact, the class of rational functions you are asked to integrate in second semester calculus is limited to those for which solving the corresponding linear systems is sufficiently easy to do without knowledge of the methods of linear algebra.

Again though there is also a theoretical question here: how do we know that the linear system we set up to do a partial fractions decomposition will always have a unique solution? This is the type of question that linear algebra can answer.

1.3. Polynomial Interpolation.

It is a well-known adage that "two points determine a line". What this means is that given any two distinct points P_1 and P_2 in the plane, there is exactly one line passing through both of them. If we dismiss vertical lines as being a not especially fascinating degenerate case, then the line is the graph of a function $f : \mathbb{R} \to \mathbb{R}$ and the problem is one of **interpolation**: suppose we have x-coordinates $x_1 < x_2$ and numbers y_1, y_2, and we want to find the unique linear function $\ell = mx + b$ which passes through the points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$. What do we do?

Really, we just do it: brute force algebra will work nicely. Namely, writing out the equations $\ell(x_1) = y_1$ and $\ell(x_2) = y_2$ we get

$$y_1 = mx_1 + b,$$
$$y_2 = mx_2 + b.$$

Again this gives us a linear system: this time the two unknowns are m and b. We can solve it, for instance, by subtracting the equatations, giving

$$y_1 - y_2 = m(x_1 - x_2),$$

so – perhaps we could have gotten here faster! –

$$m = \frac{y_1 - y_2}{x_1 - x_2}.$$

(The denominator cannot be zero since we have assumed $x_1 < x_2$.) Then

$$b = y_1 - mx_1 = y_1 - \left(\frac{y_1 - y_2}{x_1 - x_2}\right) x_1.$$

Now suppose that we are given three points

$$P_1 = (x_1, y_1),\ P_2 = (x_2, y_2),\ P_3 = (x_3, y_3).$$

Of course there will usually be no line that passes through all three points: if at first there is a line passing through all three, then change the value of y_3! However this time there is a unique **quadratic function**

$$f(x) = ax^2 + bx + c$$

such that $f(x_1) = y_1$, $f(x_2) = y_2$, $f(x_3) = y_3$. By plugging everything in we get a linear system:

$$y_1 = f(x_1) = ax_1^2 + bx_1 + c$$
$$y_2 = f(x_2) = ax_2^2 + bx_2 + c$$
$$y_3 = f(x_3) = ax_3^2 + bx_3 + c.$$

(Note that these are not purely linear equations, but they are linear in the unknown variables a, b and c.) For particular $x_1 < x_2 < x_3$ and y_1, y_2, y_3, we can try to solve the linear system: if there is indeed a unique solution, we will find it. However this time we do not have a *sparse* system of three equations and three unknowns: we really have to do some work to solve it. It is another matter entirely to explain why there is always exactly one solution (a, b, c). In fact there are some general theorems along these lines, for instance.

THEOREM 1.1. *(Lagrange Interpolation) Let n be a positive integer, let $x_1 < \ldots < x_n < x_{n+1}$ be real numbers, and let $y_1, \ldots, y_n, y_{n+1}$ be real numbers. Then there is exactly one polynomial $P(x)$ of the form $P(x) = a_n x^n + \ldots + a_1 x + a_0$ – i.e., of degree at most n – such that $P(x_1) = y_1$, $P(x_2) = y_2$, $\ldots$, $P(x_{n+1}) = y_{n+1}$.*

We will explain how to prove this theorem later on in the course. For now let me notice that there is another polynomial interpolation theorem which is even more familiar. Namely, given an n times differentiable function f defined on an interval containing $c \in \mathbb{R}$, there is a unique polynomial function $T_n(x)$ of degree at most n such that: for all $0 \leq i \leq n$, the ith derivative of T_n at c is equal to the ith derivative of f at c:

$$T_n^{(i)}(c) = f^{(i)}(c).$$

Namely, T_n must be the **degree n Taylor polynomial of f**,

$$T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(c)}{i!}(x - c)^i.$$

This **Taylor Interpolation Theorem** can be (and is, say in Math 3100) proved without using linear algebra. In fact one can give a linear-algebra free proof of Lagrange Interpolation: see e.g. [**HC**, § 12.5].

But what if we want to interpolate *between* Lagrange Interpolation and Taylor Interpretation? For instance, suppose I have a function f, and I want a polynomial $P(x)$ which matches the value of the function and the first two derivatives at 1, the value of the function at 3 and the value of the function

and the first three derivatives at 7. If you are patient enough to write all this out you will see that this amounts to $3 + 1 + 4 = 8$ different linear equations on the coefficients of an unknown polynomial. Since a degree n polynomial has $n + 1$ different coefficients, it is plausible that to do this we should look for a polynomial of degree (at most) 8. The **Hermite Interpolation Theorem** says that one can always interpolate in this way by a polynomial of degree at most n, as long as $n + 1$ is at least as large as "the number of conditions" we are imposing. It is a very satisfying generalization of both Lagrange and Taylor interpolation, and in contrast to the above I *only* know how to prove this result using linear algebra. We will do so later in the course.

1.4. Fibonacci Numbers.

There is a very famous and ubiquitous sequence of positive integers defined by $F_1 = F_2 = 1$ and for all $n \geq 3$, $F_n = F_{n-1} = F_{n-2}$. In other words,

$$F_3 = F_1 + F_2 = 1 + 1 = 2,$$

$$F_4 = F_2 + F_3 = 1 + 2 = 3,$$

$$F_5 = F_3 + F_4 = 2 + 3 = 5,$$

$$F_6 = F_4 + F_5 = 3 + 5 = 8,$$

and so forth. There are all kinds of amazing identities surrounding the Fibonacci numbers. Here are three:

THEOREM 1.2. *(Cassini Identity) For all positive integers n,*

$$F_{n+1}F_{n-1} - F_n^2 = (-1)^n.$$

THEOREM 1.3. *(Addition Formula) For all positive integers m and n,*

$$F_{m+n} = F_m F_{n+1} + F_{m-1}F_n.$$

THEOREM 1.4. *(Binet's Formula) Let $\varphi = \frac{1+\sqrt{5}}{2}$ and $\overline{\varphi} = \frac{1-\sqrt{5}}{2}$. Then for all positive integers n,*

$$F_n = \frac{\varphi^n - \overline{\varphi}^n}{\sqrt{5}}.$$

It is in fact possible to prove all of these identities by induction on n. I have done so when teaching induction in Math 3200. But the kryptonite of mathematical induction is that it does not give you any help with the (often much more difficult) task of *coming up with* the statements you are trying to prove. It turns out that one can not only prove but also *discover* these identities by using the algebraic properties of a certain matrix

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

In particular, as we will see, one can *multiply* any two square matrices of the same size to get another square matrix of the same size. In particular, one can take powers M^k of any square matrix. Then the key to all three of the above identities is the following matrix identity.

THEOREM 1.5. *For all positive integers n, we have*

$$M^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix}.$$

1.5. Some Differential Equations.

One of the most important areas of both pure and applied mathematics is differential equations. For instance, Newton's Second Law is $F = ma = mx''$. If the force is given as a function of the position, then we get a second order differential equation whose solutions tell us how the body moves when subjected to the given force: this is mathematical magic. For instance, in the case of a mass suspended on a spring, Hooke's Law says that upon distending the mass a distance of x units from the equilibrium, the spring pulls back in the opposite direction and with a force which is simply proportional to x: $F = -kx$. This leads to

$$-kx(t) = mx''(t),$$

or

$$x'' = \frac{-m}{k}x.$$

Let us suppose for the sake of simplicity that $m = k$, so we get the equation

$$x'' = -x.$$

This has two fundamental solutions $x_1(t) = \cos t$, $x_2(t) = \sin t$, and the general solution is obtained by linearly combinbing them:

$$x(t) = C_1 \cos t + C_2 \sin t.$$

Note that the differential equaiton $x'' = -x$ implies that

$$x'''' = (x'')'' = (-x)'' = -x'' = -(-x) = x.$$

Thus we are looking for functions which are equal to their own fourth derivative. This larger space has four fundamental solutions

$$e^t, e^{-t}, \cos t, \sin t$$

and the general solution is a linear combination of them. However e^t and e^{-t} are "degenerate" solutions in that they satisfy $x'' = x$ rather than the desired $x'' = -x$.

It turns out that we are very lucky to be able to write down nondegenerate solutions to $x'''' = x$ by pure thought as we did above. Suppose for instance we want a function $x = x(t)$ which is equal to its own third derivative: $x''' = x$. Again there is the "degenerate solution" $x(t) = e^t$, which satisfies $x' = x$. What about a nondegenerate solution? The methods of linear algebra will help with this.

2. Row Operations and Row Echelon Form

We work over the real numbers (but everything we say would be valid for matrices with coefficients in any **field** of scalars). For positive integers m, n we denote by $M_{m,n}$ the set of $m \times n$ matrices.

Two matrices $A, B \in M_{m,n}$ are **row equivalent**, and write $A \sim B$, if we can get from A to B by performing a finite sequence of elementary row operations. If $A \sim B$, then for any column vectors $x = (x_1, \ldots, x_n)$ and $d = (d_1, \ldots, d_m)$, we have $Ax = d$ if and only if $Bx = d$. This is just matrix notation for the fact that the elementary row operations preserve the solution set of a linear system.

Row equivalence is indeed an equivalence relation on $M_{m,n}$, hence it partitions $M_{m,n}$ into equivalence classes. One can motivate the main result of this note by asking for a canonical way of choosing one matrix from each equivalence class.

For $M \in M_{m,n}$, an entry m_{ij} is a **leading entry** if, reading from left to right, it is nonzero and is the first nonzero entry in its row. Thus a row has a leading entry if and only if it is not a *zero row*, and every row has at most one leading entry. Thus the number of leading entries of M is at most m, the number of rows.

A matrix $A \in M_{m,n}$ is in **row echelon form** if

(REF1) Every zero row of A lies below every nonzero row of A, and
(REF2) Every leading entry occurs to the right of every leading entry in the rows above it.

EXERCISE 2.1. *a) Show that (REF2) is equivalent to: if $1 \leq i_1 < i_2 \leq m$ and $1 \leq j_1, j_2 \leq n$, $a_{i_1 j_1}$ is a leading entry and $a_{i_2 j_2} \neq 0$, then $j_2 > j_1$.*
b) Show that (REF2) implies that every entry lying directly below a leading entry is 0. More precisely, if a_{ij} is a leading entry, then for all $i \leq i' \leq m$, $a_{i'j} = 0$.

A matrix $A \in M_{m,n}$ is in **reduced row echelon form** (or **rref**) if it is in row echelon form and moreover

(RREF1) If a column contains a leading entry, every other entry of that column is zero, and
(RREF2) Every leading entry is equal to 1.

PROPOSITION 2.1. *Every $A \in M_{m,n}$ is row equivalent to a rref matrix.*

PROOF. The proof is constructive: that is, we give an explicit procedure.
Step 1: We use elementary row operations to put $A \in M_{m,n}$ in row echelon form. We begin by looking at the first column.
Case 1: If every entry of the first column is 0, we move on to the $m \times (n-1)$ submatrix A' obtained by removing the first column of A. Any row operations that put A' in row echelon form will also put A in row echelon form (moreover if a matrix has a zero column, then so does any row equivalent matrix).
Case 2: Suppose that some entry of the first column is nonzero.
Case 2a: Suppose $a_{11} \neq 0$. Then by using the type (III) row operation, for all $2 \leq i \leq m$, we multiply row 1 by $\frac{-a_{m1}}{a_{11}}$ and add it to row i, thus making the $(i,1)$ entry equal to zero. Thus we end up with a matrix with $a_{1,1}$ nonzero (thus a leading entry) and $a_{i,1} = 0$ for all $2 \leq i \leq m$. We then proceed inward to the $(m-1) \times (n-1)$ submatrix A' formed by removing the first row and column of A, observing that any sequence of row operations that puts A' in row echelon form does the same for our matrix.[3]
Case 2b: Suppose that $a_{11} = 0$. Since we are in Case 2, there is some i such that $a_{i1} \neq 0$. For the sake of definiteness[4] take the smallest such i and perform the type (I) row operation switching the first and ith rows. This places us back in Case 2a.
We now have a smaller – either $m \times (n-1)$ or $(m-1) \times (n-1)$ – matrix to put in row echelon form, so we can apply the above procedure to this matrix. (In other words, the algorithm is *recursive*: we do something very simple and then allow the algorithm to call on itself for smaller parameter values.)
Step 2: We have now replaced A by a row equivalent matrix which in row echelon form. We may easily go further and put in reduced row echelon form. First, in each row containing a leading entry a_{ij}, we use the type (III) row operation to make all the entries in that column *above* a_{ij} equal to zero just as we did for the entries below to get to row echelon form. (It is worth thinking about why this process necessarily preserves row echelon form: e.g. how do we know we don't produce any zero rows lying above nonzero rows by doing this?) Finally, for every row containing a leading entry a_{ij} we use the type (II) row operation to multiply the ith row by $\frac{1}{a_{ij}}$, which makes the leading entry equal to 1 (and does not change which entries are zero or nonzero so preserves everything we've done so far). □

EXERCISE 2.2. *a) Suppose $A \in M_{m,n}$ has entries in the rational numbers $\mathbb{Q}$ – i.e., numbers of the form $\frac{a}{b}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$. Show that our algorithm produces a row echelon form and then a reduced row echelon form with entries in $\mathbb{Q}$.*
b) Suppose $A \in M_{m,n}$ has entries in $\mathbb{Z}$. Show that our algorithm does not *in general produce a row echelon form or a reduced row echelon form with entries in $\mathbb{Z}$.*

[3]This is straightforward to check but not, I think, immediately obvious. It is also very important...so please do check it.

[4]We do actually want to give an algorithm. An algorithm is not allowed to "make choices": it must do the same thing every time.

c) Show however that a modified algorithm will take any A with entries in $\mathbb{Z}$ and yield a row echelon form with entries in $\mathbb{Z}$. (Hint: when you want to divide by something, multiply a different row by that thing instead.) In fact, show that we can start with any A with entries in $\mathbb{Q}$ and find a row equivalent matrix in row echelon form with entries in $\mathbb{Z}$.
d) Show that if A has entries in $M_{m,n}$, then a modified algorithm will yield a row echelon form which satisfies (RREF1).[5]

Let A be a matrix in row echelon form. Then the variables corresponding to the columns which contain leading entries are called **pivot variables**, whereas the variables corresponding to the other columns are called **free variables**.

THEOREM 2.2. *The reduced row echelon form is unique. More precisely, for each $A \in M_{m,n}$, there is exactly one matrix $R \in M_{m,n}$ with $A \sim R$ and R in reduced row echelon form.*

PROOF. We follow T. Yuster [**Y84**] by inducting on n, the number of columns.

Base Case ($n = 1$): Suppose A has only one column. If A is the all zero matrix, it is row equivalent only to itself and is in reduced row echelon form. Every nonzero matrix with one column has a nonzero entry, and all such matrices have reduced row echelon form the column vector $(1, 0, \ldots, 0)$ and no other row echelon form.

Induction Step: Suppose now that $n > 1$, that the result holds for all $m \times n$ matrices, and let $A \in M_{m,n+1}$. For any $M \in M_{m,n+1}$, we let $M' \in M_{m,n}$ be obtained by removing the last column from M. Let B and C be reduced row echelon forms of A. Here is the key observation: the matrices B' and C' are in reduced row echelon form and row equivalent to A'.
By induction, we have $B' = C'$. In other words, we know that the reduced row echelon matrices B and C are equal except possibly in the last column. Seeking a contradiction we suppose that their last columns are not equal: i.e., there is some $1 \leq i \leq m$ such that $b_{i,n+1} \neq c_{i,n+1}$. Now let $x = (x_1, \ldots, x_{n+1})$ be any vector with $Bx = 0$, i.e., a solution of the associated homogeneous linear system. Because B and C are row equivalent, x is also a solution to the homogeneous system $Cx = 0$. It follows that $(B - C)x = 0$. Since the matrix $B - C$ is zero except in its last column, performing the multiplication of the ith row of $B - C$ by x simply gives $(b_{i,n+1} - c_{i,n+1})x_{n+1} = 0$. Since $b_{i,n+1} \neq c_{i,n+1}$ we deduce that $x_{n+1} = 0$. Thus x_{n+1} is not a free variable for either B or C, so in each of these matrices the last column must contain a leading entry of 1 and have all the other entries 0. Moreover, in both B and C the 1 must lie in the first zero row of B' and C'. Thus $B = C$. □

The uniqueness of reduced row echelon form has several important consequences. For now we point the following one.

COROLLARY 2.3. *Let $A \in M_{m,n}$, and let B and C be row equivalent matrices each in row echelon form. Then the pivot variables with respect to the matrix B are the same as the pivot variables with respect to the matrix C.*

PROOF. We gave an algorithm to take the matrix B and put it in *reduced* row echelon form. At every step this algorithm preserves the positions of the leading entries, so it preserves pivot variables. Thus the pivot variables with respect to B are the same as the pivot variables for some reduced row echelon form matrix R_B which is row equivalent to A. Similarly, the pivot variables with respect to C are the same as the pivot variables for some reduced row echelon form matrix R_C which is row equivalent to A. But by Theorem 2.2, $R_B = R_C$, and thus the pivot variables with respect to B are the same as the pivot variables with respect o C. □

This allows us to make the following important definition. For a matrix $A \in M_{m,n}$, we define the **rank of A** to be the number of pivot variables in any row echelon form of A and the **nullity of A** to be the number of free variables in any row echelon form of A. Corollary 2.3 ensures that this is "well-defined", i.e., independent of the row echelon form chosen. The following result follows immediately but, when translated into other contexts, is in fact important and quite powerful.

[5]Perhaps we should call this **almost reduced row echelon form?**

THEOREM 2.4 (Rank-Nullity Theorem, Version 1). *For any $A \in M_{m,n}$ we have*

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

EXERCISE 2.3. *Prove the Rank-Nullity Theorem.*

EXERCISE 2.4. *Find all rref matrices $R \in M_{2,2}$. (Hint: what are the possibilities for the first column?)*

We denote by I_n the $n \times n$ matrix with (i,i) entry equal to 1 for all 1 and other entries 0. Notice that I_n is in reduced row echelon form, and every column contains a leading entry. Thus it has n pivot variables and 0 free variables, so it has rank n. Conversely, it is easy to see that I_n is the only reduced row echelon form matrix in $M_{n,n}$ in which every column contains a leading entry.

THEOREM 2.5. *For $A \in M_{n,n}$, the following are equivalent:*
(i) $\operatorname{rref}(A) = I_n$.
(ii) For all $b \in \mathbb{R}^n$, there is a unique $x \in \mathbb{R}^n$ such that $Ax = b$.
(iii) For all $x \in \mathbb{R}^n$, if $Ax = 0$, then $x = 0$.
(iv) $\operatorname{rank} A = n$.

PROOF. (i) $\implies$ (ii): If $\operatorname{rref} A = I_n$, then for any $b \in \mathbb{R}^n$, to solve $Ax = b$ we put the augmented matrix $[A \mid b]$ in reduced row echelon form, getting $[\operatorname{rref}(A) \mid b'] = [I_n \mid b']$. (Here b' is whatever get by starting with b and doing the row reduction process.) In terms of equations this reads $I_n x = b'$, i.e., $x = b'$. So there is always a unique solution.
(ii) $\implies$ (iii): Let $b = 0$ can always solve the homogeneous system $Ax = 0$ by taking $x = 0$. Since we are assuming the solution is *unique*, we must not have any other solutions: if $Ax = 0$, then $x = 0$.
(iii) $\implies$ (iv): The number of parameters of the solution space to $Ax = 0$ is equal to the number of free variables. So if we have only the trivial solution to $Ax = 0$, we have no free variables and therefore all n variables are pivot variables: $\operatorname{rank} A = n$.
(iv) $\implies$ (i): Since $\operatorname{rank} A = n$ and n is the number of columns, $\operatorname{rref}(A)$ has n leading entries. For an $n \times n$ matrix in reduced row echelon form to have n leading entries, it must be I_n. (If you are doubtful, imagine the leading entry in the first row occurs anywhere to the right of the first column: you'll quickly see that you cannot then get $n-1$ further leading entries. Now move on to the second row; to be in row echelon form the leading entry cannot be any sooner than the second column; if it were later, then we cannot get $n-2$ further leading entries. And so forth.) □

3. Matrices and Linear Transformations

3.1. Review of Composition of Functions.

Let $f : X \to Y$ and $g : Y \to Z$. We may **compose** the functions to get

$$g \circ f : X \to Z,\ x \mapsto g(f(x)).$$

Note the somewhat confusing fact that $g \circ f$ means "first perform f, then g". This can be traced to the fact that we evaluate functions *on the right*, i.e., we write $f(x)$. If we were willing to evaluate functions *on the left* – i.e., to write $(x)f$ instead – then composition would behave less confusingly. I have seen function evaluaiton written on the left in some old textbooks, but not within the last thirty years or so (and I don't advocate it myself). It seems that we are stuck with things the way they are.

For any set A, let X denote the set of all functions $f : A \to A$. Then composition of $f, g \in X$ is always defined and gives a binary composition law on X. The composition of functions on a fixed set is surely the most important example of a binary composition law on a set, and many other important laws reduce to it.

EXAMPLE 3.1. *Given $f, g : A \to A$, we need* not *have $g \circ f = f \circ g$. That is, composition of functions is not generally* ***commutative****. One learns some form of this in elementary school when one*

is taught the order of operations. In general, the order in which procedures are performed may affect the outcome! For a simple example, suppose $f, g : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ and $g(x) = x + 1$. Then

$$g \circ f : x \mapsto g(f(x)) = g(x^2) = x^2 + 1,$$

while

$$f \circ g : x \mapsto f(g(x)) = f(x+1) = (x+1)^2 = x^2 + 2x + 1.$$

Since for all $x \neq 0$, $x^2 + 2x + 1 > x^2 + 1$, $g \circ f \neq f \circ g$.

EXERCISE 3.1. *Our above example of non-commutativity of function composition used the infinite set of real numbers. What is the* smallest *set A which admits functions $f, g : A \to A$ such that $g \circ f \neq f \circ g$? E.g. can you find such functions with $A = \{1, 2, 3, 4\}$? What about with a smaller set A?*

THEOREM 3.2. *Composition of functions is associative (when defined). That is, if $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$, then*

$$(h \circ g) \circ f = h \circ (g \circ f).$$

PROOF. Sometimes the way to show that two things are equal is simply to write out both of them and see that we get the same thing. This is one of those times. The function $(h \circ g) \circ f$ is the function which takes

$$x \mapsto ((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)),$$

while the function $h \circ (g \circ f)$ is the function which takes

$$x \mapsto (h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)).$$

No problem! □

For any set X, we denote the 1_X the function which maps each $x \in X$ to itself. This is called the **identity function**.

Portions of the following result ought to be familiar from a previous course, but we will provide a complete proof anyway.

PROPOSITION 3.3. *Let X and Y be nonempty sets; consider a function $f : X \to Y$.*
a) If X is nonempty, the following are equivalent:
(i) There is a function $g : Y \to X$ such that $g \circ f = 1_X$.
*(ii) f is **injective**: for all $x_1, x_2 \in X$, if $f(x_1) = f(x_2)$, then $x_1 = x_2$.*
b) The following are equivalent:
(i) There is a function $g : Y \to X$ such that $f \circ g = 1_Y$.
*(ii) f is **surjective**: for all $y \in Y$, there is $x \in X$ such that $f(x) = y$.*
c) The following are equivalent:
(i) There is a function $g : Y \to X$ such that $g \circ f = 1_X$ and $f \circ g = 1_Y$.
*(ii) f is **bijective**.*

PROOF. a) Suppose that $g \circ f = 1_X$, and let $x_1, x_2 \in X$ be such that $f(x_1) = f(x_2)$. Applying g we get $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$. So f is injective. Conversely, suppose that f is injective. We have several choices for $g : Y \to X$. For each $y \in Y$ which is of the form $f(x)$ for some $x \in X$, we put $g(y) = x$: this makes sense because, since f is injective, if $y = f(x)$, then it is of this form for exactly one $x \in X$. Fix an element $x_0 \in X$ (here we use $X \neq \varnothing$); if y is *not* of the form $f(x)$ for any $x \in X$, we put $g(y) = x_0$. The point is that this latter definition doesn't matter: for all $x \in X$ we have $g(f(x)) = x$, which is what we wanted.
b) Suppose that $f \circ g = 1_Y$, and let $y \in Y$. Then $f(g(y)) = y$. Thus every element of y is mapped to by some element of X: f is surjective. Conversely, suppose that f is surjective. Then for each $y \in Y$ there is at least one $x \in X$ with $f(x) = y$. We choose *any* such x and put $g(y) = x$. Then for all $y \in Y$, $f(g(y)) = f(x) = y$.
c) We simply combine parts a) and b). □

EXERCISE 3.2. *(For nullologists[6] only.)*
a) Observe that the nonemptiness of Y was never used.
b) If $X = \varnothing$, show that a(ii) always holds, whereas a(i) holds iff $Y = \varnothing$.
c) Show that part c) holds even if $X = \varnothing$.

EXAMPLE 3.4. *Let X be the set of all functions $f : \mathbb{R} \to \mathbb{R}$, with binary composition law the usual composition of functions. In this case there are certainly functions which are injective but not surjective – e.g. $f(x) = \arctan x$ – as well as functions which are surjective but not injective – e.g. $g(x) = x \sin x$. Thus it is certainly possible for an element to have a left inverse but no right inverse, or conversely.*

3.2. Linear Transformations.

A **linear transformation** $L : \mathbb{R}^n \to \mathbb{R}^m$ is a function which satisfies:

(LT1) For all $v, w \in \mathbb{R}^n$, $L(v + w) = L(v) + L(w)$.
(LT2) For all $\alpha \in \mathbb{R}$ and $v \in \mathbb{R}^n$, $L(\alpha v) = \alpha L(v)$.

PROPOSITION 3.5. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a function.*
a) If L is a linear transformation and $k \geq 2$, then for $v_1, \ldots, v_k \in \mathbb{R}^n$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$, we have

$$L(\alpha_1 v_1 + \ldots + \alpha_n v_n) = \alpha_1 L(v_1) + \ldots + \alpha_n L(v_n).$$

*b) (**One-Step Linear Transformation Test**): Suppose that for all $\alpha \in \mathbb{R}$ and $v, w \in \mathbb{R}^n$, $L(\alpha v + w) = \alpha L(v) + L(w)$. Then L is a linear transformation.*

PROOF. a) This is a clasic example of a "binary property" extending immediately to an n-ary property: c.f. [**Cl-I**, § 7]. To get a formal proof, we go by induction on k.
Base Case ($k = 2$): Applying (LT1) and then (LT2) twice we get

$$L(\alpha_1 v_1 + \alpha_2 v_2) = L(\alpha_1 v_1) + L(\alpha_2 v_2) = \alpha_1 L(v_1) + \alpha_2 L(v_2).$$

Induction Step: Suppose the result holds for $k \geq 2$. Then

$$L(\alpha_1 v_1 + \ldots + \alpha_{k+1} v_{k+1}) = L((\alpha_1 v_1 + \ldots + \alpha_k v_k) + \alpha_{k+1} v_{k+1})$$

$$= L(\alpha_1 v_1 + \ldots + \alpha_k v_k) + L(\alpha_{k+1} v_{k+1}) \stackrel{\text{IH}}{=} \alpha_1 L(v_1) + \ldots + \alpha_k L(v_k) + \alpha_{k+1} L(v_{k+1}).$$

b) Taking $\alpha = 1$ we get (LT1). Taking $w = 0$ we get (LT2). □

Remark: Don't take Proposition 3.5b) too seriously. Really it is just a way of collecting two easy things together so that we can call it one easy thing. In fact I think it saves more space in writing than it does time in thinking, so although I will use it below when proving that maps are linear transformations, when you are asked to think about whether a map is a linear transformation you may as well think in terms of (LT1) and (LT2) separately.

LEMMA 3.6. *If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $L(0) = 0$.*

PROOF. We have $L(0) = L(0 + 0) = L(0) + L(0)$. Subtracting the vector $L(0)$ from both sides yields $L(0) = 0$. □

Remark: Our statement of Lemma 3.6 is slightly sloppy: the zero on the left hand side is the zero vector in $\mathbb{R}^n$, whereas the zero vector on the right hand side is the zero vector in $\mathbb{R}^m$. In principle we should distinguish them notationally, perhaps by writing 0_m and 0_n. But in practice this adds complication without clarity.

In a way our definition of linear transformation is overly abstract and fancy. I claim that a linear

[6]Nullology: the study of the empty set. C.f. the sound of no hands clapping.

transformation is really just a vector of linear functions with zero constant terms. In other words, for $1 \leq i \leq m$ and $1 \leq j \leq n$, let a_{ij} be a real number. We define the function $L : \mathbb{R}^n \to \mathbb{R}^m$ by

$$L(x_1, \ldots, x_n) = (a_{11}x_1 + \ldots + a_{1n}x_n, \ldots, a_{m1}x_1 + \ldots + a_{mn}x_n). \tag{1}$$

PROPOSITION 3.7. *The function L defined by (1) is a linear transformation.*

PROOF. Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ and $\alpha \in \mathbb{R}$. Using the One-Step Linear Transformation Test, we have

$$\begin{aligned} L(\alpha x + y) &= L(\alpha x_1 + y_1, \ldots, \alpha x_n + y_n) \\ &= (a_{11}(\alpha x_1 + y_1) + \ldots + a_{1n}(\alpha x_1 + y_1), \ldots, a_{m1}(\alpha x_1 + y_1) + \ldots + a_{mn}(\alpha x_n + y_n)) \\ &= \alpha((a_{11}x_1 + \ldots + a_{1n}x_n, \ldots, a_{m1}x_1 + \ldots + a_{mn}x_n)) \\ &\quad +(a_{11}y_1 + \ldots + a_{1n}y_n, \ldots, a_{m1}y_1 + \ldots + a_{mn}y_n) \\ &= \alpha L(x) + L(y). \end{aligned}$$

□

The converse also holds: every linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ is of the form (1) for some numbers a_{ij}. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. First, like any function with values in a product, we may write $L = (L_1(x), \ldots, L_m(x))$. Then L is a linear transformation if and only if each component function $L_i : \mathbb{R}^n \to \mathbb{R}$ is a linear transformation. (We have tried to provide complete details for this basic but important material, but we cannot think of any way to write this claim out that is any more convincing than if you just think about it for a moment. Please do so.) Thus we have reduced to the case $m = 1$ and must show that any linear transformation $L : \mathbb{R}^n \to \mathbb{R}$ is of the form

$$L(x) = a_1x_1 + \ldots + a_nx_n.$$

To this end, let $e_j = (0, 0, \ldots, 1, \ldots, 0)$ be the vector with a 1 in the jth component and all other components zero. Then

$$L(x) = L(x_1e_1 + \ldots + x_ne_n) = x_1L(e_1) + \ldots + x_nL(e_n).$$

Since $L(e_1), \ldots, L(e_n)$ are just real numbers, we may call them $a_1, \ldots, a_n$, and then

$$L(x) = a_1x_1 + \ldots + a_nx_n.$$

Thus a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ amounts precisely to a vector of m linear functions with zero constant terms.

There is another way to view an arbitrary linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$. For $1 \leq i \leq m$ and $1 \leq j \leq n$, we let a_{ij} be the ith component of $L(e_i)$, so that

$$L(x_1, \ldots, x_n) = (a_{11}x_1 + \ldots + a_{1n}x_n, a_{21}x_1 + \ldots + a_{2n}x_n, \ldots, a_{mn}x_1 + \ldots + a_{mn}x_n)$$

as above. Let $A \in M_{m,n}$ be the matrix with $(A)_{ij} = a_{ij}$. Then

$$L(x_1, \ldots, x_n) = Ax,$$

where in the above equation we regard x as an $n \times 1$ column vector. In summary:

THEOREM 3.8. *For any $m, n \in \mathbb{Z}^+$, the following are equivalent:*
(i) A vector of m linear expressions in $x_1, \ldots, x_n$ with zero constant terms.
(ii) A linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$.
(iii) A matrix $A \in M_{m,n}$.

Let us now consider composition of linear transformations. If $L_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $L_2 : \mathbb{R}^p \to \mathbb{R}^n$, then the composition

$$L_1 \circ L_2 : \mathbb{R}^p \to \mathbb{R}^m, x \mapsto L_1(L_2(x))$$

is defined.

LEMMA 3.9. *If $L_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $L_2 : \mathbb{R}^p \to \mathbb{R}^n$ are linear transformations, then their composition $L_1 \circ L_2 : \mathbb{R}^p \to \mathbb{R}^m$ is a linear transformation.*

PROOF. Let me give away a secret: if a map is indeed a linear transformation, checking that it is is almost always trivial. We use the One-Step Linear Transformation Test and follow our noses: if $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, then

$$(L_1 \circ L_2)(\alpha x + y) = L_1(L_2(\alpha x + y)) = L_1(\alpha L_2(x) + L_2(y))$$
$$= L_1(\alpha L_2(x)) + L_1(L_2(y)) = \alpha L_1(L_2(x)) + L_1(L_2(y)) = \alpha(L_1 \circ L_2)(x) + (L_1 \circ L_2)(y).$$

□

LEMMA 3.10. *For a linear $L : \mathbb{R}^n \to \mathbb{R}^m$, the following are equivalent:*
(i) L is injective.
(ii) If $L(v) = 0$, then $v = 0$.

PROOF. (i) $\implies$ (ii): By Lemma 3.6 $L(0) = 0$, so if $L(v) = 0$ then $0 = L(v) = L(0)$. By injectivity, $v = 0$.
(ii) $\implies$ (i): Suppose $v, w \in \mathbb{R}^n$ are such that $L(v) = L(w)$. Then $0 = L(v) - L(w) = L(v - w)$. By hypothesis, this implies $v - w = 0$, so $v = w$. □

LEMMA 3.11. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a bijective linear transformation, so L admits an inverse function $g : \mathbb{R}^m \to \mathbb{R}^n$: that is, for all $x \in \mathbb{R}^n$, $g(L(x)) = x$, and for all $y \in \mathbb{R}^m$, $L(g(y)) = y$. Then g is a linear transformation.*

PROOF. Note first that L is injective: if $x_1, x_2 \in \mathbb{R}^n$, then $L(x_1) = L(x_2)$ implies $x_1 = g(L(x_1)) = g(L(x_20) = x_2$. Thus, let $y_1, y_2 \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. To show that $g(\alpha y_1 + y_2) = \alpha g(y_1) + g(y_2)$, since L is injective, it suffices to show this equality after applying L, i.e., that $L(g(\alpha y_1 + y_2)) = L(\alpha g(y_1) + g(y_2))$. But since L is linear we have

$$L(g(\alpha y_1 + y_2)) = \alpha y_1 + y_2 = \alpha L(g(y_1)) + L(g(y_2)) = L(\alpha g(y_1) + g(y_2)).$$ □

On the matrix side we can represent L_1 as multiplication by a matrix $M_1 \in M_{m,n}$ and L_2 as multiplication by a matrix $M_2 \in M_{n,p}$, and then $L_1 \circ L_2$, being a linear transformation from $\mathbb{R}^p$ to $\mathbb{R}^m$, must be represented by some matrix M_3 in $M_{m,p}$. A natural question to ask is how the entries of this "composite matrix" M_3 depend on the entries of M_1 and M_2. The answer to this question will lead us to define the fundamental operation of **matrix multiplication**.

Let $M_1 = (a_{ij})$ and $M_2 = (b_{jk})$. Then for $x = (x_1, \ldots, x_p)$,

$$(L_1 \circ L_2)(x) = M_1(M_2 x) = M_1 \begin{pmatrix} b_{11}x_1 + \ldots + b_{1p}x_p \\ \vdots \\ b_{n1}x_1 + \ldots + b_{np}x_p \end{pmatrix} =$$

$$\begin{pmatrix} a_{11}(b_{11}x_1 + \ldots + b_{1p}x_p) + a_{12}(b_{21}x_1 + \ldots + b_{2p}x_p) + \ldots + a_{1n}(b_{n1}x_1 + \ldots + b_{np}x_p) \\ \vdots \\ a_{m1}((b_{11}x_1 + \ldots + b_{1p}x_p) + a_{m2}(b_{21}x_1 + \ldots + b_{2p}x_p) + \ldots + a_{mn}(b_{mn}x_1 + \ldots + b_{np}x_p) \end{pmatrix}$$

$$= \begin{pmatrix} (a_{11}b_{11} + a_{12}b_{21} + \ldots + a_{1n}b_{n1})x_1 + \ldots + (a_{11}b_{1p} + \ldots + a_{1n}b_{np})x_p \\ \vdots \\ (a_{m1}b_{11} + \ldots + a_{mn}b_{1p})x_1 + \ldots + (a_{m1}b_{1p} + \ldots + a_{mn}b_{np})x_p \end{pmatrix} = M_3 x,$$

where M_3 is the $m \times p$ matrix whose (i, j) entry is the dot product

$$(a_{i1}, \ldots, a_{in}) \cdot (b_{1j}, \ldots, b_{nj}),$$

i.e., the dot product of the **ith row** of M_1 with the **jth column** of M_2.

This motivates the following definition: if $A \in M_{m,n}$ and $B \in M_{n,p}$, we define the **matrix product**

$AB \in M_{m,p}$ to be the matrix with (i,j) entry the dot product of the ith row of A with the jth column of B. Notice that in order for this definition to make sense we need these vector to have the same number of components; the number of components of the first vector is the number of columns of A, and the number of components of the second vector is the number of rows of B: in our setup both are equal to n, so this makes sense.

EXERCISE 3.3. *Let $A \in M_{m,n}$ and $B \in M_{n,p}$.*
a) Suppose that for some $1 \leq i \leq m$, the ith row of A is zero. Show that the ith row of AB is zero.
b) Suppose that for some $1 \leq j \leq n$, the jth column of B is zero. Show that the jth column of AB is zero.

EXAMPLE 3.12. *(Fibonacci Matrices I) Let $M = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. As advertised earlier, we will show that for all positive integers n, we have*

$$M^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix}, \tag{2}$$

*where the F_n's are the **Fibonacci numbers**:*

$$F_0 = 0,\ F_1 = F_2 = 1, \forall n \geq 2, F_n = F_{n-1} + F_{n-2}.$$

We prove this by induction on n. The base case, $n = 1$, is immediate. Now suppose that (2) holds for some $n \in \mathbb{Z}^+$. Then we have

$$M^{n+1} = M^n M = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} F_{n+1} + F_n & F_{n+1} \\ F_n + F_{n-1} & F_n \end{bmatrix} = \begin{bmatrix} F_{n+2} & F_{n+1} \\ F_{n+1} & F_n \end{bmatrix},$$

completing the induction step.

We turn now to the prospect of proving identities about Fibonacci numbers via linear algebra. Our level of success depends on how much linear algebra we know! For now we prove the Addition Formula:

$$\begin{bmatrix} F_{m+n+1} & F_{m+n} \\ F_{m+n} & F_{m+n-1} \end{bmatrix} = M^{m+n} = M^m M^n = \begin{bmatrix} F_{m+1} & F_m \\ F_m & F_{m-1} \end{bmatrix} \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix}.$$

The lower left entry of M^{m+n} is F_{m+n}. The lower left entry of $M^m M^n$ is $F_m F_{n+1} + F_{m-1} F_n$, so

$$F_{m+n} = F_m F_{n+1} + F_{m-1} F_n.$$

3.3. Matrix Products as Linear Combinations of Rows and Columns.

There is an alternate interpretation of the matrix product AB that is often underutilized. Namely, we claim that the rows of AB are linear combinations of the rows of B and that the columns of AB are linear combinations of the columns of A. To see this it suffices to work one row or column at a time, so first consider what happens if A is a $1 \times n$ matrix – i.e., a **row vector** – and B is an $n \times p$ matrix, so AB is a $1 \times p$ row vector. We may write $A = (x_1, \ldots, x_n) = x_1 e_1 + \ldots + x_n e_n$, with $e_i = (0, \ldots, 1, \ldots, 0)$ as usual. Now observe that the matrix product $e_i B$ is simply the ith row of B. It follows that $AB = x_1(e_1 B) + \ldots + x_n(e_n B)$ is a linear combination of the rows of B, as claimed. In the general case A and AB will each have m different rows, which simply means that we get m (possibly) different linear combinations of the rows of B.

The corresponding interpretation of the columns of AB as linear combinations of the columns of A is similar but more familiar. In particular, the fact that Ae_j returns the jth column of A is a key insight in analyzing the linear transformation $L_A : \mathbb{R}^m \to \mathbb{R}^n$ given by $x \mapsto Ax$.

3.4. Fundamentals of Matrix Algebra.

We have now defined addition and multiplication operations on matrices, so we have some kind of algebraic structure on them. Especially, things work out best if we restrict to **square matrices**, i.e., when the number of rows equals the number of columns, for then if $A, B \in M_{n,n}$, their product AB

is defined and is again an element of $M_{n,n}$. In other words, matrix multiplication gives a **binary composition law** on $M_{n,n}$.

3.4.1. *Identity Elements.*

Let $(X, \cdot)$ be a set endowed with a binary operation. An **identity element** is an element $1 \in X$ such that for all $x \in X$, $1 \cdot x = x \cdot 1 = x$.

LEMMA 3.13. *A set endowed with a binary operation $(X, \cdot)$ can have at most one identity element.*

PROOF. Suppose 1 and $1'$ are both identity elements. Then we have

$$1 = 1 \cdot 1' = 1'. \qquad \square$$

For $n \in \mathbb{Z}^+$, we define the **identity matrix** I_n to be the matrix with (i,i) entry 1 for all i and all other entries 0.

EXERCISE 3.4. *a) Let $A \in M_{m,n}$. Show that $AI_n = A$.*
b) Let $B \in M_{n,p}$. Show that $I_n B = B$.
c) Deduce that I_n is the unique identity element for $(M_{n,n}, \cdot)$.

3.4.2. *Absorbing Elements.*

EXERCISE 3.5. *Let $0_{m,n}$ denote the $m \times n$ matrix consisting entirely of zeros.*
a) Let $A \in M_{k,m}$. Show that $A0_{m,n} = 0_{k,n}$.
b) Let $B \in M_{n,p}$. Show that $0_{m,n}B = 0_{m,p}$.
c) Let $0 = 0_{n,n}$. Deduce that for all $A \in M_{n,n}$, $A0 = 0A = 0$.

EXERCISE 3.6. *Let $(X, \cdot)$ be a set endowed with a binary operation. An element $Z \in X$ is **absorbing** if for all $A \in X$, $ZA = AZ = Z$.*
a) Show that there can be at most one absorbing element.
b) Deduce that the zero matrix is the unique element $Z \in M_{n,n}$ such that for all $A \in M_{n,n}$, $ZA = AZ = Z$.

3.4.3. *Commutativity.*

A binary operation $\cdot$ on a set X is **commutative** if $xy = yx$ for all $x, y \in X$.

Multiplication in $M_{1,1}$ is the usual multiplication of real numbers, so it is of course commutative. However, the situation for $n \geq 2$ is quite different.

EXAMPLE 3.14. *Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then*

$$AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = B,$$

while

$$BA = 0.$$

Thus matrix multiplication is not commutative in general.

EXERCISE 3.7. *For $1 \leq i, j \leq n$, let E_{ij} be the matrix with all zero entries except for a 1 in the (i,j) entry.*
a) Let $A \in M_{n,n}$. Show that AE_{ij} has every column zero except for the jth column, and the jth column is the ith column of A.
b) Let $A \in M_{n,n}$. Show that $E_{ij}A$ has every row zero except for the ith row, and the ith row is the jth row of A.

EXERCISE 3.8. *a) Let $\alpha \in \mathbb{R}$. Show that αI_n is the matrix with (i,i) entry equal to α and (i,j) entry equal to 0 for all $i \neq j$. Such matrices are called* ***scalar matrices****.*
b) Let A be a scalar matrix. Show that A ***commutes with*** *every $B \in M_{n,n}$: for all $B \in M_{n,n}$, $AB = BA$.*

EXERCISE 3.9. *Let $A \in M_{n,n}$ be such that A commutes with every $B \in M_{n,n}$.*
a) Fix $1 \leq i \leq n$. Use $AE_{ii} = E_{ii}A$ to show that if $j \neq i$, then $a_{ij} = a_{ji} = 0$.
b) Fix $1 \leq j \leq n$. Use $AE_{1j} = E_{1j}A$ to show that $a_{11} = a_{jj}$.
c) Deduce that A is a scalar matrix.

Define the **center** of $M_{n,n}$ to be the set of all matrices $A \in M_{n,n}$ which commute with every $B \in M_{n,n}$. The preceding exercises show that the center of $M_{n,n}$ is precisely the set of all scalar matrices.

3.4.4. *Associativity.*

PROPOSITION 3.15. *Matrix multiplication is associative: if $A \in M_{m,n}$, $B \in M_{n,p}$ and $C \in M_{p,q}$, then*

$$(AB)C = A(BC).$$

PROOF. First Proof: Because matrix multiplication corresponds to composition of linear maps, and function composition is always asssocative, matrix multiplication must be associative.
Second Proof: The above argument is certainly the best explanation of why matrix multiplication is associative. On the other hand it should be possible to show directly that $(AB)C = A(BC)$, and it is. A little quiet contemplation shows that for $1 \leq i \leq m$ and $1 \leq l \leq q$, the (i,l) entry of *both* $(AB)C$ and $A(BC)$ is

$$\sum_{1 \leq j \leq n,\ 1 \leq k \leq p} a_{ij} b_{jk} c_{kl}. \qquad \square$$

Because we have shown matrix multiplication is associative, we may freely drop the parentheses, writing ABC for $(AB)C = A(BC)$. Moreover, it is a general fact about binary operations that as soon as we have associativity then we have a "generalized associativity": any n-fold product $A_1 \cdots A_n$ is well-defined independent of the parentheses. See [**Wa01**] for a nice treatment of this.

EXERCISE 3.10. *Let $M_1, \ldots, M_n$ be matrices, of dimensions such that for $1 \leq i \leq n-1$, each product $M_i M_{i+1}$ is defined. Write down an explicit expression for the general entry of $M_1 \cdots M_n$.*

3.4.5. *Inverses.*

Let $(X, \cdot)$ be a set endowed with a binary operation, and possessing an identity element 1. Let $A \in X$.

An element $A_L \in X$ is a **left inverse** to A if $A_L A = 1$.

An element $A_R \in X$ is a **right inverse** to A if $A A_R = 1$.

Let $A, B \in X$. Sometimes we will say that B is a **one-sided inverse** to A if either $AB = 1$ or $BA = 1$. Note that this is symmetric: if B is a one-sided inverse to A, then A is a one-sided inverse to B.

An element $B \in X$ is an **inverse** to A if $AB = BA = 1$.

An element $A \in X$ is **invertible** if it has an invere.

PROPOSITION 3.16. *Let $(X, \cdot)$ be a set endowed with an associative binary composition law and possessing an identity element 1. Then any element $A \in X$ has at most one inverse: if B and C are both inverses to A, then $B = C$.*

PROOF.

$$C = 1 \cdot C = (BA)C = BAC = B(AC) = B \cdot 1 = B. \qquad \square$$

Because an element has at most one inverse, it makes sense to denote the inverse of an (invertible!) element $A \in (X, \cdot)$ by A^{-1}.

EXERCISE 3.11. *Suppose that $B = A^{-1}$. Show that $A = B^{-1}$. (We sometimes say "A and B are mutually inverse.")*

EXERCISE 3.12. *Consider $f : \mathbb{R} \to [0, \infty)$ by $f(x) = x^2$.*
a) Show that f is surjective but not injective, so by Theorem 3.1 it has a right inverse g – i.e., $f \circ g = 1_{[0,\infty)}$ – but no left inverse.
b) Find one right inverse to f.
c) Find all *right inverses to f. Conclude in particular that an element can have more than one right inverse.*

EXERCISE 3.13. *Consider $f : [0, \infty) \to \mathbb{R}$ by $f(x) = x$. (Note that this is not the identity function, since the domain and codomain are different!)*
a) Show that f is injective but not surjective, so by Theorem 3.1 it has a left inverse g – i.e., $g \circ f = 1_{[0,\infty)}$ – but no right inverse.
b) Find one left inverse to f.
c) Find all *left inverses to f. Conclude in particular that an element can have more than one left inverse.*

EXERCISE 3.14. *Addition on $M_{m,n}$ is also a binary composition law.*
a) Show that there is an identity element for matrix addition: what is it?
b) Show that every matrix $A \in M_{m,n}$ has an additive inverse *matrix: what is it?*

Very often in mathematics we will have a set endowed with two binary operations, called $+$ and $\cdot$. (There are whole courses on this...) In this case, A^{-1} refers to the *multiplicative inverse* rather than the additive inverse. That goes in particular for $M_{n,n}$.

PROPOSITION 3.17. *(Shoes 'n' Socks) Let $(X, \cdot)$ be a set endowed with an associative binary composition law and possessing an identity element 1. Let $A_1, \ldots, A_n \in X$ be invertible elements. Then the product $A_1 \cdots A_n$ is invertible and*

$$(A_1 \cdots A_n)^{-1} = A_n^{-1} \cdots A_1^{-1}.$$

PROOF. Consider

$$(A_n^{-1} \cdots A_1^{-1})(A_1 \cdots A_n).$$

Working our way from the inside out we cancel $A_1^{-1}A_1$, then cancel $A_2^{-1}A_2$, and so forth, finally cancelling $A_n^{-1}A_n$ to get 1. And much the same goes for

$$(A_1 \cdots A_n)(A_n^{-1} \cdots A_1^{-1}) :$$

we first cancel $A_n A_n^{-1}$, then $A_{n-1}A_{n-1}^{-1}$, and so forth, finally cancelling $A_1 A_1^{-1}$ to get 1. $\square$

We can recapture the spirit of the statement and proof of Proposition 3.17 as follows: think of an invertible element as a process which can be reversed. (This is especially reasonable when the elements of X are functions and the binary operation is composition, because one often thinks of a function as being a procedure which takes an input, does something to it, and returns an output.) Any process which is obtained by performing several reversible processes can itself be reversed: however, to do so we must reverse the individual processes in reverse order. If that sounds like a mouthful, consider: barring a catastrophe, putting on your socks is a reversible process, as is putting on your shoes. In the morning we put on our socks first and then our shoes. In the evening we undo this composite procedure by undoing the individual components, but in order to do so we must now deal with our shoes first and our socks second.

EXERCISE 3.15. *Let $A \in M_{n,n}$. Suppose that A is invertible. Show that A* cannot *have a zero row or column.*

Let's reconsider the all-important example where $(X, \cdot)$ is the set of all functions from a set A to itself, and the binary operation is composition of functions. In this setting, Theorem 3.1 says that if f has a left inverse f_L and a right inverse f_R, then it is injective and surjective, hence bijective, hence it has an inverse function. This turns out to be a general fact about composition laws.

PROPOSITION 3.18. *Let $(X, \cdot)$ be a set endowed with an associative binary composition law and possessing an identity element 1. Let $A \in X$ have a left inverse A_L and a right inverse A_R. Then $A_L = A_R$ is the inverse of A.*

PROOF. We have

$$A_R = 1 \cdot A_R = (A_L A)A_R = A_L A A_R = A_L(AA_R) = A_L \cdot 1 = A_L.$$

It follows immediately that $A_R A = AA_R = 1$, so $A_R = A_L$ is the inverse of A. □

Here is a related result.

PROPOSITION 3.19. *Let $(X, \cdot)$ be a set endowed with an associative binary composition law and possessing an identity element 1. Let $A, B \in X$ be such that $BA = 1$. If* either *of A or B is invertible then A and B are mutually inverse.*

PROOF. Suppose first that B is invertible. Multiplying both sides of $BA = 1$ on the left by B^{-1}, we get $A = 1 \cdot A = B^{-1}BA = B^{-1}1 = B^{-1}$. Next suppose that A is invertible. Multiplying both sides of $BA = 1$ on the right by A^{-1}, we get $B = B \cdot 1 = BAA^{-1} = A^{-1}$. □

Although matrix multiplication is an instance of function composition, it is an especially simple instance which behaves better than the general case. It turns out that a matrix $A \in M_{n,n}$ which has a left inverse must also have a right inverse and thus be invertible (and similarly, if A has a right inverse it must also have a left inverse and thus be invertible). It turns out to be difficult to prove this directly, however. The right thing to do is to take a more ambitious approach by trying to give a *characterization* of invertible matrices in terms of row reduction.

THEOREM 3.20. *For $A \in M_{n,n}$, the following are equivalent:*
(i) We have $\operatorname{rref} A = I_n$.
(ii) We have $\operatorname{rank} A = n$.
(iii): The equation $Ax = 0$ has only the trivial solution $x = 0$.
(iii′) The map $L_A : \mathbb{R}^n \to \mathbb{R}^n$ is injective.
(iv) For all $b \in \mathbb{R}^n$, the equation $Ax = b$ has a unique solution.
(iv′) The map $L_A : \mathbb{R}^n \to \mathbb{R}^n$ is bijective.
(v) The matrix A is invertible: there is $B \in M_{n,n}$ with $AB = BA = I_n$.

PROOF. Step 1: Recall that we already know (i) $\iff$ (ii) $\iff$ (iii) $\iff$ (iv): this is Theorem 2.5. The equivalence of (iii) and (iii′) is Lemma 3.1, and the equivalence of (iv) and (iv′) is immediate from the definitions: to say that L_A is bijective is exactly to say that for every $b \in \mathbb{R}^n$, there is a unique $x \in \mathbb{R}^n$ such that $Ax = L_A(x) = b$.
Step 2: It remains to show that any one of the conditions other than (v) implies (v) and that (v) implies any one of the other conditions. We will show that (iv′) implies (v) and that (v) implies (i).
(iv′) $\implies$ (v): Since $L_A : \mathbb{R}^n \to \mathbb{R}^n$ is bijective, by Theorem 3.1 it has an inverse function, i.e., there is $g : \mathbb{R}^n \to \mathbb{R}^n$ with $g \circ L_A = L_A \circ g = 1$. By Lemma 3.11 g is a linear transformation, hence by Theorem 3.8 $g = L_B$ for some matrix B. Expressing that L_A and L_B are mutually inverse functions in terms of matrices we get precisely that $BA = I_n = AB$.
(v) $\implies$ (i): If $BA = I_n$ then $L_B \circ L_A = 1_{\mathbb{R}^\times}$, so by Theorem 3.1, L_A is injective. □

The logic of the proof yields a further result. We could have stated it as an equivalent condition in Theorem 3.20, but we decided to optimize our exposition so as to make the proof of that key result as simple and clean as possible.

COROLLARY 3.21. *Let $A, B \in M_{n,n}$. If $BA = I_n$, then $AB = I_n$.*

PROOF. As in the proof of Theorem 3.20, $BA = I_n$ implies that L_A is injective. By Theorem 3.20, A is invertible. According to Proposition 3.19, its one-sided inverse B must actually be its inverse, hence also $AB = I_n$. □

This result can be restated as follows: a square matrix which has a one-sided inverse must be invertible, and any one-sided inverse is in fact the inverse.

Up until this point our discussion of inverses of matrices has been purely theoretical. We now know that we can tell whether a square matrix is invertible by putting it in reduced row echelon form, but if $\operatorname{rref}(A) = I_n$, how do we go about finding the inverse? Again by row reduction!

THEOREM 3.22. *If $\operatorname{rref}(A) = I_n$, then $\operatorname{rref}[A \mid I_n] = \left[I_n \mid A^{-1}\right]$.*

PROOF. Write B for A^{-1}, so we have the equation $AB = I_n$. Focusing in on the jth columns of both sides of this equation gives us a system of matrix equations $Ab_j = e_j$ which we are trying to solve for the vectors $b_1, \ldots, b_n$. To solve this system we proceed in the usual way: write down $[A \mid e_j]$ and put in reduced row echelon form. Since A is invertible, $\operatorname{rref} A = I_n$, the solution vector b_j is *unique*, and thus the reduced row echelon form is $[I_n \mid b_j]$. Since the columns of a matrix function quite independently under row reduction, nothing stops us from writing down the wider augmented matrix $[A \mid I_n]$. The same row operations that put A in reduced row echelon form convert the jth column on the right to the unique solution vector b_j, so $\operatorname{rref}[A \mid I_n] = \left[I_n \mid A^{-1}\right]$. □

3.5. Elementary Matrices.

Fix a positive integer m.

For $1 \leq i \neq j \leq m$, a **type I elementary matrix** $S_{i,j}$ is the $m \times m$ matrix which results from interchanging the ith and jth rows of I_m.

For $1 \leq i \leq m$, a **type II elementary matrix** $M_{i,j}(\alpha)$ is the $m \times m$ matrix which results from multiplying every entry in a single row of I_m by some nonzero $\alpha \in \mathbb{R}$.

For $1 \leq i \neq j \leq m$ and nonzero $\alpha \in \mathbb{R}$, a **type III elementary matrix** $T_{i,j}(\alpha)$ is the $m \times m$ matrix which differs from I_m precisely in having its (j, i) entry equal to α.

EXERCISE 3.16. *Show that every elementary matrix is invertible. In fact, show that the inverse of any elementary matrix is again an elementary matrix of the same type.*

In all cases an elementary matrix is the matrix you get by performing an elementary row operation on the identity matrix I_m. In fact a little more is true: in each case the above data is sufficient to describe an elementary row operation on any $m \times n$ matrix.

PROPOSITION 3.23. *Let $A \in M_{m,n}$. We perform one elementary row operation R on A to get a new matrix B. Then $B = EA$, where E is the corresponding elementary matrix.*

EXERCISE 3.17. *Prove Proposition 3.23.*

The upshot is that we can track row reduction as a (finite!) sequence of premultiplications – i.e., multiplications on the left – by elementary matrices. Suppose that these elementary matrices are called $E_1, \ldots, E_r$. Then we get the matrix equation

$$\operatorname{rref}(A) = E_r \cdots E_1 A. \tag{3}$$

Use of elementary matrices is not essential, but often allows one to make cleaner arguments involving matrices rather than row operations.

EXAMPLE 3.24. *We will use elementary matrices to give a second proof of the important fact that $A \in M_{n,n}$ is invertible if and only if* $\operatorname{rref}(A) = I_n$. *First suppose* $\operatorname{rref}(A) = I_n$. *Then (3) reads*

$$I_n = E_r \cdots E_1 A.$$

Each elementary matrix E_i is invertible, so by Shoes 'n' Socks so is $E_r \cdots E_1$, and the inverse is $E_1^{-1} \cdots E_r^{-1}$. Premultiplying both sides by this product, we get

$$E_1^{-1} \cdots E_r^{-1} = A.$$

As a product of invertible elements, A is itself invertible: to be explicit about it, we can apply Shoes 'n' Socks again to get

$$A^{-1} = E_r \cdots E_1.$$

Here is another example of a result that can be proved via row reduction considerations but is a little cleaner via elementary matrices. This time the result is a new one: we could have included it among the equivalent conditions of Theorem 3.20 but we had enough conditions to deal with at one time.

THEOREM 3.25. *Let $A \in M_{n,n}$. The equivalent conditions (i) through (v) of Theorem 3.20 are also equivalent to each of the following:*
(vi) There are no constraints: for all $b \in \mathbb{R}^n$, the linear equation $Ax = b$ is consistent, i.e., has at least one solution.
(vi′) The linear transformation $L_A : \mathbb{R}^n \to \mathbb{R}^n$ is surjective.

PROOF. Step 1: Conditions (vi) and (vi′) are equivalent to each other: indeed, unpacking the definition of the surjectivity of L_A, one gets precisely condition (vi).
Step 2: Certainly conditions (iv) implies condition (vi) (or, equally true: certainly condition (iv′) implies condition (vi′)).
Step 3: We suppose condition (vi). It suffices to show that rref A has no zero rows, for then rank $A = n$: condition (ii) of Theorem 3.20. Suppose on the contrary that the last row of rref $A = 0$. Start with $b \in \mathbb{R}^n$, and let

$$\operatorname{rref}[A \mid b] = [\operatorname{rref}(A) \mid b']. \tag{4}$$

In turn let b' be $(b_1, \ldots, b_n)$. If $b_n \neq 0$, then the last equation of the augmented rref matrix reads $0 = b_n$, a contradiction. So in particular we get an inconsistent system if $b'_n = e_n = (0, \ldots, 0, 1)$. What we need to argue for now is that there is some choice of $b \in \mathbb{R}^n$ such that the b' defined by (4) turns out to be e_n. In other words, we need to work the row reduction process backwards. This is exactly what elementary matrices make clean and easy: there are elementary matrices such that

$$E_n \cdots E_1 A = \operatorname{rref}(A),$$

and then for any $b \in \mathbb{R}^n$, performing the same row operations gives

$$E_n \cdots E_1 b = b'.$$

But $E_1, \ldots, E_n$ are all invertible, so we have

$$b = (E_n \cdots E_1)^{-1} b' = E_1^{-1} \cdots E_n^{-1} b'.$$

Taking $b' = e_n$ this shows exactly what coefficient vector b we need to start with to get $\operatorname{rref}[A \mid b] = [\operatorname{rref}(A) \mid e_n]$. Summing up: if rref A had a zero row, there is some b such that $Ax = b$ is inconsistent. This is precisely the contrapositive of what we wanted to show, so we're done. □

3.6. Diagonal Matrices.

A matrix $A \in M_{n,n}$ is **diagonal** if for all $i \neq j$, $a_{ij} = 0$. In other words the only nonzero entries lie along the **main diagonal** $a_{11}, a_{22}, \ldots, a_{nn}$.

Since we only need n numbers to specify an $n \times n$ diagonal matrix, we don't need double indexing, and we will denote the diagonal matrix with diagonal entries $a_1, \ldots, a_n$ as $\Delta(a_1, \ldots, a_n)$.

EXAMPLE 3.26. *a)* $\Delta(1, \ldots, 1)$ *is the identity matrix* I_n.
b) For any $\alpha \in \mathbb{R}$, $\Delta(\alpha, \ldots, \alpha)$ *is a scalar matrix.*

The algebra of diagonal matrices is much simpler than that of arbitary matrices.

EXERCISE 3.18. *Let* $A = \Delta(a_1, \ldots, a_n)$, $B = \Delta(b_1, \ldots, b_n)$ *be diagonal matrices.*
a) Show that $AB = \Delta(a_1 b_1, \ldots, a_n b_n)$.
b) Deduce that $AB = BA$*: diagonal matrices commute.*

PROPOSITION 3.27. *Consider a diagonal matrix* $A = \Delta(a_1, \ldots, a_n)$.
a) The following are equivalent:
(i) A *is invertible.*
(ii) $a_1, \ldots, a_n$ *are all nonzero.*
b) If the equivalent conditions of part a) hold, then $A^{-1} = \Delta(a_1^{-1}, \ldots, a_n^{-1})$.

PROOF. Step 1: To show (i) $\implies$ (ii) in part a) we will verify the contrapositive: not (ii) $\implies$ not (i). The negation of (ii) is that for some i, $a_i = 0$. If this is the case then the ith row of A is zero, so $\operatorname{rref} A \neq I_n$ and A is not invertible.
Step 2: Suppose $a_1, \ldots, a_n$ are all nonzero. By Exercise 3.18a), we have

$$\Delta(a_1, \ldots, a_n)\Delta(a_1^{-1}, \ldots, a_n^{-1}) = \Delta(1, \ldots, 1) = I_n.$$

This shows that $\Delta(a_1, \ldots, a_n)$ is invertible and that its inverse is $\Delta(a_1^{-1}, \ldots, a_n^{-1})$. Hence we get (ii) $\implies$ (i) in part a) and also the result of part b): we're done. □

3.7. Triangular Matrices.

A matrix $A \in M_{n,n}$ is **upper triangular** if for all $i > j$, $a_{ij} = 0$.

A matrix $A \in M_{n,n}$ is **lower triangular** if for all $i < j$, $a_{ij} = 0$.

A matrix $A \in M_{n,n}$ is diagonal if and only if it is both upper and lower triangular. Above we introduced the idea that diagonal matrices are ideally simple and easy to work with that we wish that every matrix could be diagonal. In a precise sense that we have not yet encountered – that of **similarity** – it is not possible to make every matrix diagonal, but (at least if we are able to use complex numbers as scalars, as we eventually will!) in this same sense it will turn out that every matrix is similar to an upper triangular matrix.

For an $n \times n$ matrix the entries $a_{11}, \ldots, a_{nn}$ are said to lie on the **main diagonal**. Thus $A \in M_{n,n}$ is upper triangular if all the entries lying *below* the main diagonal are zero, is lower triangular if all the entries lying *above* the main diagonal are zero, and is diagonal if all the entries are zero except possibly those on the main diagonal.

EXERCISE 3.19. *Let* $A \in M_{n,n}$.
a) Suppose A *is in row echelon form. Show that* A *is upper triangular.*
b) Suppose A *is upper triangular. Must* A *be in row echelon form?*

PROPOSITION 3.28. *Let* $A, B \in M_{n,n}$.
a) If A *and* B *are both upper triangular, then* AB *is upper triangular.*
b) If A *and* B *are both lower triangular, then* AB *is lower triangular.*

PROOF. We have, as always, that

$$(AB)_{ij} = \sum_{j=1}^{n} a_{ik} b_{kj}. \tag{5}$$

a) Suppose A and B are upper triangular and $i > j$. If $i > k$ we have $a_{ik} = 0$ since A is upper triangular. On the other hand, if $k > j$ we have $b_{kj} = 0$ since B is upper triangular. But no matter

what k is, at least one of these two conditions must hold, because if both fail we have $j \leq k \leq i$, contradicting our assumption that $i > j$. Thus each term in the sum of (5) is the product of two numbers, at least one of which is zero. So $(AB)_{ij} = 0$.
b) I leave this to you: you'll know that you've understood the proof of part a) if you can supply the proof of part b). □

EXERCISE 3.20. *a) Let $A, B \in M_{n,n}$ be upper triangular. Show that for all $1 \leq i \leq n$, $(AB)_{ii} = a_{ii}b_{ii}$.*
b) Does the conclusion of part a) hold if instead of being upper triangular, A and B are both lower triangular?

EXERCISE 3.21. *Suppose $A \in M_{n,n}$ is upper triangular with $a_{ii} \neq 0$ for all $1 \leq i \leq n$. Show that A is in row echelon form.*

PROPOSITION 3.29. *For an upper triangular $A \in M_{n,n}$, the following are equivalent:*
(i) For all $1 \leq i \leq n$, $a_{ii} \neq 0$.
(ii) A is invertible.

PROOF. (i) $\implies$ (ii): By Exercise 3.21, the matrix A is in row echelon form. Since every row has the leading entry a_{ii}, A has rank n and is thus (by Theorem 3.20) invertible.
(ii) $\implies$ (i): We will work our way from the bottom right to the top left. First, we must have $a_{nn} \neq 0$, since otherwise the last row is zero and A cannot be invertible. Since $a_{nn} \neq 0$, we zero out the entries of the last column. Because a_{nn} is the only nonzero entry in the last row, this process does not any of the entries of the other column, so in particular does not change the diagonal entries: it is enough to prove that the diagonal entries are all nonzero. Now if $a_{n-1,n-1}$ were equal to zero, the $(n-1)$st row would be zero, again contradicting invertibility. Now we can zero out the other entries of the $(n-1)$st column without disturbing any of the diagonal entries. And so forth: we get that $a_{nn} \neq 0$, $a_{n-1n-1} \neq 0, \ldots, a_{11} \neq 0$. □

PROPOSITION 3.30. *Let $A \in M_{n,n}$ be an invertible matrix. Then:*
a) If A is upper triangular, then A^{-1} is upper triangular.
b) If A is lower triangular, then A^{-1} is lower triangular.

PROOF. a) If A is upper triangular and invertible, then by Proposition 3.29, all the entries along the main diagonal are nonzero. Thus to put it in reduced row echelon form we only need to perform type (II) row operations to make the leading entries 1 and type (III) row operations of the form of adding a multiple of row j to row i with $i < j$ to zero out the entries above each leading entry. The elementary matrices E_i corresponding to each of these row operations is upper triangular, and

$$I_n = \text{rref}(A) = E_r \cdots E_1 A,$$

so $A^{-1} = E_r \cdots E_1$ is a product of upper triangular matrices and hence, by Proposition 3.28, upper triangular.
b) The argument is very similar to that of part a) and we leave it to the reader. □

EXERCISE 3.22. *Prove part b) of Theorem 3.30. (One thing to convince yourself of is the fact that an invertible lower triangular matrix is generally* not *in row echelon form is not a problem.)*

3.8. The Transpose Matrix.

Before we leave the realm of basic matrix algebra for more exalted terrain, we want to discuss one last operation on matrices. In a way this last operation is the easiest – certainly you could explain it to anyone, regardless of their mathematical background – but it is a little less clear why this operation should be important in linear algebra. The latter question, unfortunately, will only be fully addressed later in the course.

Let $A \in M_{m,n}$. We define the **transpose matrix** $A^T \in M_{n,m}$ as the matrix whose (i,j) entry is

a_{ji}. In other words, the rows of A become the columns of A^T, and vice versa. One can also think of taking the transpose as reflecting the entries of A across the main diagonal.

PROPOSITION 3.31. *Let $A, B \in M_{m,n}$, $C \in M_{n,p}$ and $\alpha \in \mathbb{R}$. Then:*
a) $(A^T)^T = A$.
b) $(\alpha A)^T = \alpha A^T$.
c) $(A+B)^T = A^T + B^T$.
d) $(AC)^T = C^T A^T$.
e) If A is invertible, so is A^T and $(A^T)^{-1} = (A^{-1})^T$.

PROOF. a) To get from a matrix to its transpose we interchange the rows and the columns. Doing this twice gets us back to the original matrix.
b) Both $(\alpha A)^T$ and αA^T have (i,j) entry αa_{ji}.
c) Both $(A+B)^T$ and $A^T + B^T$ have (i,j) entry $a_{ji} + b_{ji}$.
d) The (k,i) entry of $C^T A^T$ is the dot product of the kth row of C^T with the ith column of A^T. This is also the dot product of the ith row of A with the jth column of C, hence it is the (k,i) entry of AC and thus the (i,k) entry of $(AC)^T$.
e) Using part d), we have $(A^{-1})^T A^T = (AA^{-1})^T = I_n^T = I_n$. □

In the above proof we used that the identity matrix I_n has the property that $I_n^T = I_n$. A matrix which has this property must be square. This turns out to be an interesting and important class of square matrices: we say $A \in M_{n,n}$ is **symmetric** if $A^T = A$. We also say that $A \in M_{n,n}$ is **skew symmetric** if $A^T = -A$.

Every symmetric matrix can be built as follows: we fill in all $1 + 2 + \ldots + n = \frac{n(n+1)}{2}$ entries lying on our above the main diagonal arbitrarily. Then the symmetry condition tells us that each entry a_{ij} for $i < j$ is equal to the corresponding entry a_{ji}. In particular, in a natural sense that we will later make precise, the set $S_{n,n}$ of all $n \times n$ symmetric matrices can be paramterized in terms of $\frac{n(n+1)}{2}$ parameters.

EXERCISE 3.23. *Give a similar description of an $n \times n$ skew symmetric matrix. How many parameters does it take to specify such a matrix?*

EXERCISE 3.24. *a) Let $A \in M_{n,n}$. Suppose $A = A^T = -A$. Show: $A = 0$.*
b) Let $A \in M_{n,n}$. Show that $A + A^T$ is symmetric and $A - A^T$ is skew symmetric.
c) Let $A \in M_{n,n}$. Show that there are unique *matrices $A_s, A_{\mathrm{ss}} \in M_{n,n}$ such that: A_s is symmetric, A_{ss} is skew symmetric, and $A = A_s + A_{\mathrm{ss}}$.*

One merit of the transpose matrix is to give us a formalism between switching between "row vectors" – i.e., elements of $M_{1,n}$ – and "column vectors" – i.e., elements of $M_{m,1}$: namely, take the transpose. We reiterate our standard convention that when viewing a vector $v \in \mathbb{R}^n$ as a matrix we view it as a column vector, not a row vector. With this convention, we can reinterpret the dot product as itself being a matrix multiplication:

$$\forall v, w \in \mathbb{R}^n, v \cdot w = v^T w.$$

PROPOSITION 3.32. *Let $A \in M_{m,n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Then*

$$(Ax) \cdot y = x \cdot (A^T y).$$

PROOF. Much of the work in this result is appreciating that both sides are well-defined, even though the left hand side is a dot product of vectors in $\mathbb{R}^m$ and the right hand side is a dot product of vector in $\mathbb{R}^n$. Once you agree that both expressions are well-defined, we can move on to the proof:

$$(Ax) \cdot y = (Ax)^T y = x^T A^T y = x^T (A^T y) = x \cdot (A^T y).$$ □

4. Subspaces, Bases and Dimension

4.1. Spans in $\mathbb{R}^n$.

Let $v_1, \ldots, v_m$ be an ordered list of vectors in $\mathbb{R}^n$. By definition, a **linear combination** of $v_1, \ldots, v_m$ is given by choosing $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ and forming

$$\alpha_1 v_1 + \ldots + \alpha_m v_m.$$

This is not meant to be a deep definition. In general, when you have a finite list of things for which it makes sense to add them and to scale them by real numbers, you can form a linear combination. More interesting is: the **span** of $v_1, \ldots, m$ is

$$\{\alpha_1 v_1 + \ldots + \alpha_m v_m \mid \alpha_1, \ldots, \alpha_m \in \mathbb{R}\},$$

that is, the span is the set of all possible linear combinations of $v_1, \ldots, v_m$. We denote this span by $\langle v_1, \ldots, v_m \rangle$.

EXAMPLE 4.1. *Let $v_1,\ v_2 \in \mathbb{R}^2$ be nonzero vectors. We claim that if v_1 and v_2 both lie on a line ℓ through the origin, then $\langle v_1, v_2 \rangle = \ell$, and otherwise we have $\langle v_1, v_2 \rangle = \mathbb{R}^2$. COMPLETE ME!!*

Evidently the span of an ordered list of vectors does not change if we reorder the list – this is simply because vector addition is commutative.

PROPOSITION 4.2. *Let $v_1, \ldots, v_m, w \in \mathbb{R}^N$. The following are equivalent:*
(i) We have $w \in \langle v_1, \ldots, v_m \rangle$.
(ii) We have $\langle v_1, \ldots, v_m \rangle = \langle v_1, \ldots, v_m, w \rangle$.

PROOF. (i) $\implies$ (ii): Suppose there are $\beta_1, \ldots, \beta_r$ such that $w = \beta_1 v_1 + \ldots + \beta_m v_m$. Then for all $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ we have

$$\begin{aligned}\alpha_1 v_1 + \ldots + \alpha_m v_m + \alpha_{m+1} w &= \alpha_1 v_1 + \ldots + \alpha_m v_m + (\alpha_{m+1}\beta_1 v_1 + \ldots + \alpha_{m+1}\beta_m v_m) \\ &= (\alpha_1 + \alpha_{m+1}\beta_1)v_1 + \ldots + (\alpha_m + \alpha_{m+1}\beta_m)v_m \in \langle v_1, \ldots, v_m \rangle,\end{aligned}$$

which shows that $\langle v_1, \ldots, v_m, w \rangle \subset \langle v_1, \ldots, v_m \rangle$. The containment $\langle v_1, \ldots, v_m \rangle \subset \langle v_1, \ldots, v_m, w \rangle$ is easier: for all $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ we have

$$\alpha_1 v_1 + \ldots + \alpha_m v_m = \alpha_1 v_1 + \ldots + \alpha_m v_m + 0w.$$

(ii) $\implies$ (ii): If $\langle v_1, \ldots, v_m \rangle = \langle v_1, \ldots, v_m, w \rangle$, then since $w = 0v_1 + \ldots + 0v_m + 1w$ we have $w \in \langle v_1, \ldots, v_m, w \rangle = \langle v_1, \ldots, v_m \rangle$. □

Thus when we append a vector at the end of an ordered list, we enlarge the span if and only if that vector was not already in the span. This shows in particular that although it is permissible for our ordered list to have repeated vectors, there is no advantage in doing so: removing all but the first instance of any given vector in an ordered list yields an ordered list without repeated vectors with the same span. Since also the order is important, this shows that the span is not *really* a property of an ordered list of vectors but actually a property of a finite *set* $S = \{v_1, \ldots, v_m\}$ of vectors. It is better for most theoretical purposes to think of span in terms of subsets, but it can also be convenient to think of spans in terms of ordered lists, especially for computations.

We now enlarge our definition of linear combination and span, as follows: let S be any subset of $\mathbb{R}^n$ (possibly infinite).

- If $S = \varnothing$, we define its span to be $\{0\}$.[7]
- If S is nonempty, then by **a linear combination from** S we mean that we choose a finite sequence $v_1, \ldots, v_m$ of elements of S and a finite sequence $\alpha_1, \ldots, \alpha_m$ of real numbers and form

$$\alpha_1 v_1 + \ldots + \alpha_m v_m.$$

[7]There is not much content here; it's just to keep things tidy.

Note well: when S is infinite, we do *not* entertain infinite linear combinations of vectors here. One could do so, but one would need to use a notion of *limit*, making this a part of topology and/or analysis. We define the **span of S**, denoted $\langle S \rangle$, as the set of all (finite!) linear combinations from S.

EXERCISE 4.1. *Let S and T be subsets of $\mathbb{R}^n$. Show:*
a) We have $S \subset \langle S \rangle$.
b) If $S \subset T$ then $\langle S \rangle \subset \langle T \rangle$.
c) We have $\langle \langle S \rangle \rangle = \langle S \rangle$.

4.2. Linear independence and spanning in $\mathbb{R}^n$.

Let $v_1, \ldots, v_m$ be an ordered list of m vectors in $\mathbb{R}^n$. We say that $v_1, \ldots, v_m$ are **linearly independent** if for all $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$, if $\alpha_1 v_1 + \ldots + \alpha_m v_m = 0$, then $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$. We say that $v_1, \ldots, v_m$ are **linearly dependent** if they are not linearly independent.

Let us write $v_1 = (a_{11}, \ldots, a_{n1})$, $v_2 = (a_{21}, \ldots, a_{n2})$,..., $v_m = (a_{m1}, \ldots, a_{mn})$, and let $A \in M_{n,m}(\mathbb{R})$ be the matrix with (i,j) entry a_{ij}: in other words, for all $1 \leq j \leq m$, the jth **column** of A is the vector v_j. Then the vector equation

$$\alpha_1 v_1 + \ldots + \alpha_m v_m = 0$$

is really the matrix equation

$$A(\alpha_1, \ldots, \alpha_m)^T = (0, \ldots, 0)^T.$$

Thus from a computational perspective, we already understand linear independence: it means that the associated homogeneous linear system has only the trivial solution. Nevertheless the concept is of paramount theoretical importance, as we now develop.

4.3. Subspaces.

A subset V of $\mathbb{R}^n$ is a **linear subspace** if all of the following hold:

(SS1) $0 \in V$.
(SS2) For all $v, w \in V$, $v + w \in V$.
(SS3) For all $v \in V$ and all $\alpha \in \mathbb{R}$, $\alpha v \in V$.

REMARK 4.3. *Since we will not be considering any other kind of subspace of $\mathbb{R}^n$ in this course, we will omit the "linear" from "linear subspace"...but you should be aware that subspaces of various sorts are ubiquitous throughout mathematics.*

We claim that in the presence of (SS2) and (SS3), we could replace (SS1) with

(SS1′) $V \neq \varnothing$ (i.e., V is nonempty).

Indeed, it is clear that (SS1) $\implies$ (SS1′). Conversely, suppose (SS1′) holds: thus there is some element $v \in V$. By (SS3), $0 \cdot v = 0 \in V$.

We prefer to give the axioms for a subspace in this form because (i) in practice it is certainly no harder to check (SS1) than (SS1′) and (ii) the formulation (SS1′) is so innocuous that it is easy to forget.

EXAMPLE 4.4. *The subset $\{0\}$ is a subspace of $\mathbb{R}^n$: it may seem too obvious to be worth mentioning, but this is important for bookkeeping purposes. Yet more obviously, $\mathbb{R}^n$ is a subspace of itself.*

EXAMPLE 4.5. *Let ℓ be a line in $\mathbb{R}^n$. We claim that ℓ is a subspace iff ℓ passes through the origin. Indeed the condition $0 \in \ell$ is necessary by (SS1). Conversely, if $0 \in \ell$ then we may express ℓ*

parametrically as $\{tx \mid t \in \mathbb{R}\}$ *for some nonzero vector* $x \in \mathbb{R}^n$. *We can now easily check (SS1) and (SS2): if* $v, w \in \ell$ *then* $v = t_1 x$ *and* $w = t_2 x$ *for some* $t_1, t_2 \in \mathbb{R}$. *Then*

$$v + w = t_1 x + t_2 x = (t_1 + t_2)x$$

lies in ℓ. *Similarly but yet more easily, if* $v = t_1 x$ *and* $\alpha \in \mathbb{R}$, *then* $\alpha v = (\alpha t_1)x \in \ell$.

EXERCISE 4.2. *Try to convince yourself that the subspaces of* $\mathbb{R}^2$ *are precisely:* $\{0\}$, *the lines through the origin, and* $\mathbb{R}^2$. *(Suggestion: the key is to believe if a subspace* V *of* $\mathbb{R}^2$ *contains two vectors* v *and* w *which do not lie on the same line, we must have* $V = \mathbb{R}^2$. *You can try to prove this now if you like, but we will carefully prove this and more general facts later on. The point of this exercise is to acquire intuition that it should be true.)*

EXAMPLE 4.6. *Let* P *be a plane in* $\mathbb{R}^3$. *We claim that* P *is a subspace iff* P *passes through the origin. Again the condition is certainly necessary. We can show the converse in several different ways, depending upon how we choose to represent* P: *recall that we know at least two ways to do so. Suppose first that we use the* parametric expression *of* P: *namely there are vectors* x *and* y – *which are not scalar multiples of one another* – *such that*

$$P = \{sx + ty \mid s, t \in \mathbb{R}\}.$$

Now let $v = s_1 x + t_1 y$ *and* $w = s_2 x + t_2 y$ *be vectors in* P. *Then*

$$v + w = s_1 x + t_1 y + s_2 x + t_2 y = (s_1 + s_2)x + (t_1 + t_2)y \in P.$$

Similarly but more easily, if $v = sx + ty \in P$ *and* $\alpha \in \mathbb{R}$, *then*

$$\alpha v = \alpha(sx + ty) = (\alpha s)x + (\alpha t)y \in P.$$

Thus P *is a subspace.*

EXAMPLE 4.7. *Let's revisit the previous example of a plane* $P \subset \mathbb{R}^3$ *passing through the origin, but this time we represent* P *using a normal vector* $n = (a, b, c)$, *namely*

$$P = \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \mid 0 = n \cdot x = ax_1 + bx_2 + cx_3 = 0\}.$$

Once again we have rigged things so that $0 \in P$. *Further, if* $x = (x_1, x_2, x_3), y = (y_1, y_2, y_3) \in P$ *then*

$$ax_1 + bx_2 + cx_3 = ay_1 + by_2 + cy_3 = 0$$

and thus

$$a(x_1 + y_1) + b(x_2 + y_2) + c(x_3 + y_3) = (ax_1 + bx_2 + cx_3) + (ay_1 + by_2 + cy_3) = 0 + 0 = 0,$$

so $x + y \in P$. *Finally, if* $x = (x_1, x_2, x_3) \in P$ *and* $\alpha \in \mathbb{R}$, *then*

$$ax_1 + bx_2 + cx_3 = 0,$$

hence

$$0 = \alpha(ax_1 + bx_2 + cx_3) = (\alpha a)x_1 + (\alpha b)x_2 + (\alpha c)x_3 = 0.$$

EXERCISE 4.3. *I claim the subspaces of* $\mathbb{R}^3$ *are:* $\{0\}$, *the lines through the origin, the planes through the origin, and* $\mathbb{R}^3$. *Try to convince yourself that this is true.*

The above two techniques of showing that a plane through the origin in $\mathbb{R}^3$ is a subpace each generalizes in a different way.

EXERCISE 4.4. *Recall: for any* $n \geq 2$, *a* ***plane*** P *in* $\mathbb{R}^n$ *is a subset of the form* $\{sx+ty+z \mid s, t \in \mathbb{R}\}$, *where* $x, y, z \in \mathbb{R}^n$ *and* x, y *are not scalar multiples of each other. Show that* P *is a subspace if and only if it passes through the origin, i.e.,* $0 \in P$.

EXERCISE 4.5. *Recall: for any* $n \geq 2$, *a* ***hyperplane*** H *in* $\mathbb{R}^n$ *is a subset of the form* $\{x \in \mathbb{R}^n \mid n \cdot x = c\}$ *for any nonzero vector* n *and* $c \in \mathbb{R}$. *Show that a hyperplane is a subspace of* $\mathbb{R}^n$ *if and only if it passes through the origin (if and only if* $c = 0$*).*

EXERCISE 4.6. *Try to convince yourself that the subspaces of* $\mathbb{R}^4$ *are:* $\{0\}$, *lines through the origin, planes through the origin, hyperplanes through the origin, and* $\mathbb{R}^4$.

EXERCISE 4.7. *Show there are more subspaces of $\mathbb{R}^5$ than just: $\{0\}$, lines through the origin, planes through the origin, hyperplanes through the origin, and $\mathbb{R}^5$. (Hint: we are missing "three-dimensional subspaces", whatever that means. But try to write one down.)*

We should also be sure to give some examples of subsets of $\mathbb{R}^n$ which are *not* subspaces. Of course any subset which does not contain the origin is such an example. Having said that, we may as well consider subsets which have this property.

EXAMPLE 4.8. *Let $Q = \{(x, y) \in \mathbb{R}^2 \mid x, y \geq 0\}$ be the first quadrant in $\mathbb{R}^2$. Then $0 \in Q$, so Q satisfies (SS1). Moreover, it is easy to see that if $v, w \in Q$, so is $v + w$. However, Q does not satisfy (SS3): $(1, 1) \in Q$ but $(-1, -1) = -1 \cdot (1, 1) \notin Q$. So Q is not a subspace. (However it is rather close to being a subspace in the sense that it satisfies (SS1), (SS2) and (SS3) for all $\alpha \geq 0$. Such subsets of $\mathbb{R}^n$ are called* ***cones****. They too show up throughout mathematics.)*

EXAMPLE 4.9. *Let $S = \{(x, y) \in \mathbb{R}^2 \mid x = 0 \text{ or } y = 0\}$. Then S satisfies (SS1) and (SS3) but not (SS2): $e_1 = (1, 0)$ and $e_2 = (0, 1)$ lie in S, but $(1, 1) = e_1 + e_2 \notin S$.*

Notice that the previous (non)example is precisely the union of two lines through the origin. We conclude that **the union of two subspaces of $\mathbb{R}^n$ need not be a subspace**. The following exercise purues this phenomenon more closely.

EXERCISE 4.8. *Let V, W be subspaces of $\mathbb{R}^n$.*
a) Suppose that $V \subset W$ or $W \subset V$. Show that $V \cup W$ is a subspace.
(Hint: this is a triviality.)
b) Suppose that $V \cup W$ is a subspace. Show that either $V \subset W$ or $W \subset V$.
(This is not. Suggestion: work by contradiction and suppose that neither $V \subset W$ nor $W \subset V$. Thus there is $x \in V \backslash W$ and $y \in W \backslash V$. Since $V \cup W$ is assumed to be a subspace, we must have $x + y \in V \cup W$. Deduce a contradiction.)

In particular, $\mathbb{R}^n$ is never the union of two proper subspaces. One can carry this argument further, but we will wait until we know a bit more about subspaces.

Subspaces behave much better with respect to intersection.

PROPOSITION 4.10. *Let $V, W \subset \mathbb{R}^n$ be subspaces. Then $V \cap W$ is a subspace of $\mathbb{R}^n$.*

PROOF. The key idea is that no ideas are necessary (!!): we just follow our nose and check the properties. First, since $0 \in V$ and $0 \in W$, $0 \in V \cap W$. Second, let $x, y \in V \cap W$. Thus $x, y \in V$ and $x, y \in W$. Since V is a subspace, $x + y \in V$; since W is a subspace, $x + y \in W$. Thus $x + y \in V \cap W$. Finally, let $x \in V \cap W$ and $\alpha \in \mathbb{R}$. Since V is a subspace, $\alpha x \in V$; since W is a subspace, $\alpha x \in W$. Thus $\alpha x \in V \cap W$. We're done! □

EXERCISE 4.9. *a) Let $V_1, \ldots, V_k$ be subspaces of $\mathbb{R}^n$. Show that the common intersection $V_1 \cap \ldots \cap V_k$ is a subspace of $\mathbb{R}^n$.*
b) Suppose that I is a nonempty set and that for each $i \in I$ we are given a subspace V_i of $\mathbb{R}^n$. (Thus we have $\{V_i\}_{i \in I}$, an ***indexed family*** *of subspaces.) Show that the common intersection $\bigcap_{i \in I} V_i$ – i.e., the set of $x \in \mathbb{R}^n$ which lie in V_i for all i – is a subspace of $\mathbb{R}^n$.*

4.4. Universal Examples of Subspaces.

In this section we give further examples of subspaces. In contrast to the examples given above, each of these examples will turn out to be a *universal example*: that is, every subspace of $\mathbb{R}^n$ arises as a case of *each* of our examples.[8]

[8]The idea of giving more than one *universal example* of something may seem strange at first. However in mathematics there is often more than one way to "understand everything," and pursuing these multiple understandings can be very fruitful.

EXAMPLE 4.11. *Let $A \in M_{m,n}$. Then the* ***null space***

$$N(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$

is a subspace of $\mathbb{R}^n$. Indeed: since $A0 = 0$, $0 \in N(A)$. If $x, y \in N(A)$ then $Ax = Ay = 0$, so $A(x + y) = Ax + Ay = 0$. Finally, if $x \in N(A)$ and $\alpha \in \mathbb{R}$ then $A(\alpha x) = \alpha Ax = \alpha \cdot 0 = 0$.

EXAMPLE 4.12. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then the* ***kernel***

$$\operatorname{Ker} L = \{x \in \mathbb{R}^n \mid Lx = 0\}$$

is a subspace of $\mathbb{R}^n$. Indeed, $L0 = 0$, so $0 \in \operatorname{Ker} L$. If $x, y \in \operatorname{Ker} L$, then $Lx = Ly = 0$, so $L(x+y) = 0$. Finally, if $x \in \operatorname{Ker} L$ and $\alpha \in \mathbb{R}$ then $L(\alpha x) = \alpha L(x) = \alpha \cdot 0 = 0$.

REMARK 4.13. *Since every linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ is of the form $x \mapsto Ax$ for a unique matrix $A \in M_{m,n}$, Examples 4.11 and 4.12 are really the same example: the null space of a matrix A is the kernel of the corresponding linear transformation L_A, and conversely. Nevertheless both terms are commonly used.*

4.5. Reducing Spanning Sets to Bases.

THEOREM 4.14. *Let $S = \{v_1, \ldots, v_k\}$ be a finite set of vectors in $\mathbb{R}^n$, and let $V = \operatorname{span}\{v, \ldots, v_k\}$. Then there is a subset $T \subset S$ such that T is a basis for V. In brief:* ***Every finite spanning set can be reduced to a basis****.*

PROOF. The idea is simple: we write out the vectors in our spanning set $v_1, \ldots, v_k$ and work from left to right: whenever we get a vector v_{i+1} which is a linear combination of the previous vectors $v_1, \ldots, v_i$, then we may remove it from S without changing the span. (In particular, we remove v_1 if and only if it is the zero vector.) We are left with a subset $T \subset S$ of vectors which still spans V, and for which none of which can be written as a linear combination of the previous vectors, hence it is linearly independent and also spans V, so it is a basis for V. □

The previous proof could hardly have been simpler, but maybe it seems a bit too theoretical for its own good: if I give you actual, numerical vectors, e.g.

$$S = \{(1, 3, 5, 7), (-2, 3, 0, 4), (1, 1, 1, 1), (0, 9, 10, 18)\},$$

how do we actually find a subset T which is a basis for $\operatorname{span} S$?

Algorithm: Given $S = \{v_1, \ldots, v_k\}$ be a finite subset of $\mathbb{R}^n$. Let $M \in M_{n,k}$ be the matrix with columns $v_1, \ldots, v_k$. Put M in reduced row echelon form. Let $\mathfrak{t} \subset \{1, \ldots, k\}$ be the set of indices i such that the ith column of $\operatorname{rref} M$ contains a leading entry. Then $T = \{v_i \mid i \in \mathfrak{t}\}$ is a basis for $\operatorname{span} S$.

For $1 \leq i \leq k$, let M_i be the matrix obtained by taking the first i columns of M. The algorithm works because:
v_i is a linear combination of $v_1, \ldots, v_{i-1}$
$\iff$ there are $\alpha_1, \ldots, \alpha_i \in \mathbb{R}$ with $\alpha_i \neq 0$ such that $\alpha_1 v_1 + \ldots + \alpha_i v_i = 0$
$\iff$ in $\operatorname{rref} M_i$, the ith column is a free variable.
Since elementary row operations work independently on the columns, we can test this all at once by looking at $\operatorname{rref} M$.

EXAMPLE 4.15. *As above take $S = \{(1, 3, 5, 7), (-2, 3, 0, 4), (1, 1, 1, 1), (0, 9, 10, 18)\}$. The reduced row echelon form of*

$$A = \begin{bmatrix} 1 & 3 & 5 & 7 \\ -2 & 3 & 0 & 4 \\ 1 & 1 & 1 & 1 \\ 0 & 9 & 10 & 18 \end{bmatrix}$$

is

$$\operatorname{rref}(A) = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

So $\{(1,3,5,7),\ (-2,3,0,4),\ (1,1,1,1)\}$ *is a basis for the subspace spanned by* S*.*

4.6. Enlarging Linearly Independent Sets to Bases. For a finite set S, $\#S$ denotes the number of elements of S; e.g. $\#\{2,4,6,8,10\} = 5$.

To be honest, for stylistic/aesthetic reasons I would prefer to defer the proof of the following result until we discuss the **Exchange Lemma**, at which point it comes for free. But giving the proof now will enable us to quickly establish a result which is the natural complement to the result of the last section.

LEMMA 4.16. *Let* $S \subset \mathbb{R}^n$ *be linearly independent. Then* S *is finite, and* $\#S \leq n$*.*

PROOF. It is enough to show that any set of $n+1$ vectors in $\mathbb{R}^n$ is linearly dependent. Recall that any set $\{v_1, \ldots, v_k\}$ of vectors in $\mathbb{R}^n$ is linearly dependent if and only if the matrix $M \in M_{n,k}$ with $v_1, \ldots, v_k$ as its columns has a nontrivial null space. But if $k = n+1$, then $M \in M_{n,n+1}$, i.e., M has more columns than rows, so must have at least one free variable and thus a nontrivial null space. □

THEOREM 4.17. *Let* V *be a subspace of* $\mathbb{R}^n$*, and let* $S \subset V$ *be a linearly independent set. Then there is a finite subset* $T \supset S$ *which is a basis for* V*.*

PROOF. Again the idea is very simple: S is a linearly independent subset of V, so it is *not* a basis precisely if there is some vector $v_1 \in V$ which is not in the span of S. If so, $S_1 = S \cup \{v_1\} \subset V$ is still linearly independent. If this larger set S_1 does not span V then there is some $v_2 \in V$ which is not in the span of S, so $S_2 = S_1 \cup \{v_2\} = S \cup \{v_1, v_2\}$ is linearly independent. And so on. This process stops precisely when we get a basis for V. And it must stop eventually. Indeed, it must stop after at most n steps: otherwise we get a linearly independent subset of $\mathbb{R}^n$ consisting of more than n vectors, contradicting the previous result. □

COROLLARY 4.18. *Every subspace* V *of* $\mathbb{R}^n$ *has a finite basis.*

PROOF. Apply Theorem 4.17 to $S = \varnothing$. □

It is not quite as clear how to make Theorem 4.17 concrete as we did for Theorem 4.14. One difference is that in Theorem 4.14 what we are given is completely concrete: a finite set of vectors in $\mathbb{R}^n$. In Theorem 4.17, we are instead given a subspace $V \subset \mathbb{R}^n$, an object which, as we have seen and will continue to see, is a bit more abstract and can be concretely realized in several different ways. But let us suppose for instance that $S = \{v_1, \ldots, v_k\}$ and that V is given to us as the span of a finite set of vectors:

$$V = \operatorname{span} w_1, \ldots, w_\ell.$$

Now we can be concrete about how to enlarge S to a finite basis for V: we use the algorithm given as the concrete form of Theorem 4.14! We start with the ordered list $v_1, \ldots, v_k, w_1, \ldots, w_\ell$, which is certainly a spanning set of V. Then we *reduce* this ordered list to a basis as we did before: we form the matrix $M \in M_{n,k+\ell}$ with columns $v_1, \ldots, v_k, w_1, \ldots, w_\ell$, and we put in reduced row echelon form. Then we keep the vectors formed by the columns of M such that the corresponding columns of rref M have leading entries. Because $v_1, \ldots, v_k$ is linearly independent, we will necessarily keep all of the first k columns, so we have expanded S to a basis of V.

4.7. The Exchange Lemma.

In the previous two sections we established a kind of "duality" between spanning sets and linearly independent sets: spanning sets have one of the properties of a basis but are in general "too large" to have the other property (linear independence). But any finite spanning set can be *reduced* so as to gain the linear independence property without losing the spanning property. Dually, linearly independent sets have the other property of a basis but are in general "too small" to have the first property (spanning). However, any linearly independent set can be *enlarged* so as to gain the spanning property without losing the linear independence property.

There is however a further "largeness" property of spanning sets that we have yet to establish and a further "smallness" property of linearly independent sets that we have as yet seen only in a weak form. Namely, we know that every linearly independent set of vectors in $\mathbb{R}^n$ has at most n elements. That's nice, but we want more. For instance, suppose that I have a linearly independent subset S not only of $\mathbb{R}^3$ but of some plane P in $\mathbb{R}^3$. We would then like to say not only that $\#S \leq 3$ but that in fact $\#S \leq 2$: it seems geometrically clear that a set of three linearly independent vectors in $\mathbb{R}^3$ should not lie in any plane but rather span all of $\mathbb{R}^3$. We feel this way because we think of a plane as a **two-dimensional object**: it is given as the span of two linearly independent vectors. However, so far as we've shown so far, a plane in $\mathbb{R}^3$ might also be the span of a linearly independent set of three vectors, and a plane in $\mathbb{R}^{17}$ might also be the span of a linearly independent set of up to 17 vectors.

In crisper terms, so far as we know, a plane $P \subset \mathbb{R}^n$ is defined as a subspace with a basis $\mathcal{B}$ with $\#\mathcal{B} = 2$, but how do we know that it does not have a different basis $\mathcal{B}'$ – recall that every nonzero subspace has infinitely many bases – with $\#\mathcal{B}' = 3$? We hope that this cannot happen. In fact, to have a notion of "dimension" of a subspace V of $\mathbb{R}^n$, what we need is that any two bases of V have the same number of elements.

The following result will allow us to show this and more.

LEMMA 4.19. *(Steinitz Exchange Lemma) Let V be a subspace of $\mathbb{R}^N$. Let $(v_1, \ldots, v_m)$ be a linearly independent sequence of vectors in V, and let $(w_1, \ldots, w_n)$ be a sequence of vectors in V with $V = \operatorname{span} w_1, \ldots, w_n$. Then $m \leq n$, and – after reordering the w_i's, if necessary – we have $V = \operatorname{span} v_1, \ldots, v_m, w_{m+1}, \ldots, w_n$.*

This is an archetypical example of a "lemma": a result which looks a bit too technical to be front page news but does the lion's share of the work of the flashier theorem that it is used to prove. And indeed, although the proof of Lemma 4.19 is not so bad, it is a little technical, so before we give it let's see the remarkable consequences that it has. It *immediately* implies the following fundamental result.

THEOREM 4.20. *Let $V \subset \mathbb{R}^n$ be a subspace.*
a) Let $S \subset V$ be a finite linearly independent subset, and let $T \subset V$ be a finite spanning set. Then $\#S \leq \#T$.
b) Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases for V. Then both $\mathcal{B}_1$ and $\mathcal{B}_2$ are finite sets, and

$$\#\mathcal{B}_1 = \#\mathcal{B}_2.$$

Part a) of Theorem 4.20 is precisely what the conclusion $m(= \#S) \leq n(= \#T)$ of the Exchange Lemma is telling us. As for part b): first, we see again that any linearly independent subset $S = \mathbb{R}^n$ has at most n elements, by applying the Exchange Lemma with $T = \{e_1, \ldots, e_n\}$. Second, a basis of V is precisely a linearly independent, spanning subset of V, so if $\mathcal{B}_1$ and $\mathcal{B}_2$ are bases of V, we can apply the Exchange Lemma with linearly independent subset $\mathcal{B}_1$ and spanning set $\mathcal{B}_2$ to get

$$\#\mathcal{B}_1 \leq \#\mathcal{B}_2$$

and then we can turn things around, applying the Exchange Lemma with linearly independent subset $\mathcal{B}_2$ and spanning set $\mathcal{B}_1$ to get

$$\#\mathcal{B}_2 \leq \#\mathcal{B}_1.$$

We conclude $\#\mathcal{B}_1 = \#\mathcal{B}_2$.

Now that we are fully invested, we turn to the *Proof of the Steinitz Exchange Lemma.* We will show in fact that for any $1 \leq i \leq m$, then $i \leq n$, and after reordering the w's if necessary, we have

$$\operatorname{span} v_1, \ldots, v_i, w_{i+1}, w_n = V.$$

Taking $i = m$ gives us the result we want. Now we proceed by induction on i.
Base Case: $i = 1$. Sure, $1 \leq n$. So we just need to "exchange" v_1 for one of the w's. We do this in two steps: first we simply put v_1 into our list: $v_1, w_1, \ldots, w_n$. But now we have too many w's: we need to take one out and still get a spanning set. This is really what we want to show: if we can do that, then we just reorder the remaining $n-1$ w's and call them $w_2, \ldots, w_n$. For this: $w_1, \ldots, w_n$ is a spanning set for V and $v_1 \in V$, so we can write $v_1 = \alpha_1 w_1 + \ldots + \alpha_n w_n$ for some $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. Moreover, not all the α_i's can be zero: if so, $v_1 = 0$, but v_1 was an element of a linearly independent sequence, so it can't be 0. Suppose for instance that $\alpha_j \neq 0$. Then we can write w_j as a linear combination of the other w's and v:

$$\alpha_j w_j = v - \alpha_1 w_1 - \ldots - \alpha_{j-1} w_{j-1} - \alpha_{j+1} w_{j+1} - \ldots - \alpha_n w_n,$$

so

$$w_i = \frac{1}{\alpha_j} v_1 - \frac{\alpha_1}{\alpha_j} w_1 - \ldots - \frac{\alpha_{i-1}}{\alpha_j} w_{j-1} - \frac{\alpha_{j+1}}{\alpha_j} w_{j+1} - \ldots - \frac{\alpha_n}{\alpha_j} w_n.$$

Thus indeed can remove w_j without changing the span, getting the n element spanning sequence $v, w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_n$. The bit about reordering is just that we will change the indices on the w's around so as to write this as $v, w_2, \ldots, w_n$.
Induction Step: Let $1 \leq i < m$, and suppose (inductively) that $i \leq n$ and after reordering we have a spanning sequence $v_1, \ldots, v_i, w_{i+1}, \ldots, w_n$. We need to show that $i+1 \leq n$ and that we can exchange v_{i+1} for one of the w's. First: since $i \leq n$, we need to rule out the possibility that $i = n$ (if so, $i < n$, so $i+1 \leq n$). If $i = n$, then this means we have already exchanged out all the w's, so $v_1, \ldots, v_i$ is a spanning set for V. But since $i < m$, $i+1 \leq m$, so we have another vector v_{i+1} in V, which must then be a linear cominbation of $v_1, \ldots, v_i$, contradicting the assumed linear independence of $v_1, \ldots, v_m$. Having negotiated that slightly tricky part, the rest of the argument is the same as the base case: first add in the next v vector, getting a spanning sequence $v_1, \ldots, v_i, v_{i+1}, w_{i+1}, \ldots, w_n$. Now we need to remove one of the w's. Since we already had a spanning sequence, there are $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$v_{i+1} = \alpha_1 v_1 + \ldots + \alpha_i v_i + \alpha_{i+1} w_{i+1} + \ldots + \alpha_n w_n.$$

Moreover, it cannot be that $\alpha_{i+1}, \ldots, \alpha_n$ are all 0: if so, we would have

$$v_{i+1} = \alpha_1 v_1 + \ldots + \alpha_i v_i,$$

and again this contradicts the linear independence of $v_1, \ldots, v_m$. Thus there is a j with $i+1 \leq j \leq n$ such that $\alpha_j \neq 0$, so as above we can write

$$w_j = \frac{1}{\alpha_j} v_{i+1} - \frac{\alpha_1}{\alpha_j} v_1 - \ldots - \frac{\alpha_i}{\alpha_j} v_i - \frac{\alpha_{i+1}}{\alpha_j} w_{i+1} - \ldots - \frac{\alpha_{j-1}}{\alpha_j} w_{j-1} - \frac{\alpha_{j+1}}{\alpha_j} w_{j+1} - \ldots - \frac{\alpha_n}{\alpha_j} w_n.$$

This shows that we can remove w_j and still get a spanning sequence

$$v_1, \ldots, v_{i+1}, w_{i+1}, \ldots, w_{j-1}, w_{j+1}, \ldots, w_n.$$

Changing the indices on the w's, we write this as

$$v_1, \ldots, v_{i+1}, w_{i+2}, \ldots, w_n,$$

and we're done.

I don't know as much about the historical development of linear algebra as I should. There is the

following complicating factor: some time in the early 20th century, the algebraic parts of mathematics became much more "abstract" following work of the golden gods Hilbert, Noether and Artin, among others. This abstract approach to mathematics has proven to be both very powerful and in many ways simpler than the previous, more numerical/concrete approach. Of course it is difficult for students to grasp at first: in a way, the difficulties you have grappling with mathematical abstraction recapitulate those of many contemporaries of those aforementioned golden gods: famously, a very eminent but "old-fashioned" mathematician Paul Gordan was alarmed at the way Hilbert proved vastly more general results about certain "rings of invariants" than Gordan had over the course of a long career. The gist of it is very similar to the dichotomy between showing that a subspace of $\mathbb{R}^n$ has a finite spanning set by a theoretical argument versus giving an algorithmic procedure for actually producing such a finite set. (Compare especially the proof of Corollary 4.18 to the more explicit procedure which is given just afterwards.) Hilbert was one of the first mathematicians to realize that it can be *much easier* to prove that something like a finite spanning set exists than to give an explicit recipe for writing one down, and he exploited this brilliantly. Gordan remarked: "This is not mathematics; this is theology." History has proven that Gordan was wrong: the abstract approach is most certainly mathematics. Proponents of the abstract approach were similarly disdainful of Gordan: "Er war ein Algorithmiker," wrote Max Noether (Emmy Noether's father, and a great mathematician in his own right...though not quite as good as his daughter) in Gordan's obituary. Nowadays – and especially with the ubiquitousness of modern computers – mathematics well understands that "algorithmikers" (i.e., algorithm-makers) can be leading mathematicians too. In our golden age both the abstract and the concrete approaches are extremely important.

Steinitz's Exchange Lemma was developed by Ernst Steinitz, a German mathematician who lived from 1871 to 1928.[9] Like much of Steinitz's work, it seems to have been somewhat neglected in his own time but firmly embraced in ours. I feel that more core linear algebra content resides in the Steinitz Exchange Lemma than any other single result in this course.

4.8. The Dimension of a Subspace.

Let V be a subspace of $\mathbb{R}^n$. By Theorem 4.20, there is a fiite basis for V. Moreover every basis of V is finite; and any two bases of V have the same number of elements. We define the **dimension** of V to be the number of elements in any basis for V.

PROPOSITION 4.21. *Let $V \subset W \subset \mathbb{R}^n$ be subspaces. If $\dim V = \dim W$, then $V = W$.*

PROOF. Let $\mathcal{B}_V = \{v_1, \ldots, v_d\}$ be a basis for V. Then $\mathcal{B}_V$ is a linearly independent subset of W, so by Theorem 4.17 there is a basis $\mathcal{B}_W$ for W containing $\mathcal{B}_V$. Since $\mathcal{B}_V \subset \mathcal{B}_W$ are finite sets with the same cardinality, $\mathcal{B}_V = \mathcal{B}_W$, and thus

$$V = \operatorname{span} \mathcal{B}_V = \operatorname{span} \mathcal{B}_W = W. \qquad \square$$

EXAMPLE 4.22. *a) The zero-subspace has dimension 0: its only basis is $\varnothing$.*
b) $\mathbb{R}^n$ has dimension n, since $e_1, \ldots, e_n$ is a basis.
c) A line in $\mathbb{R}^n$ (passing through the origin) can be formally defined as a 1-dimensional subspace of $\mathbb{R}^n$.
d) A plane in $\mathbb{R}^n$ (passing through the origin) can be formally defined as a 2-dimensional subspace of $\mathbb{R}^n$.

PROPOSITION 4.23. *Let $A \in M_{m,n}$. Then the dimension of the null space of A is precisely the number of free variables in* rref A. *Thus*

$$\dim \operatorname{nullity} A + \operatorname{rank} A = n.$$

[9]This was not the same Steinitz who was the first official world chess champion, although the next world chess champion, Emanuel Lasker, was a student of Emmy Noether.

EXAMPLE 4.24. *Recall that a* ***hyperplane*** *in $\mathbb{R}^n$ is a subspace of the form*

$$H = \{x \in \mathbb{R}^n \mid x \cdot n = 0\}$$

for some nonzero vector n. We claim that the hyperplanes are precisely the $(n-1)$-dimensional subspaces of $\mathbb{R}^n$.
Step 1: Let H be a hyperplane. Then H is the null space of the $1 \times n$ matrix A with n as its row. Since $n \neq 0$, M has rank one, and nullity $A = n - 1$.

4.9. Dimensions of Intersections; Independent Subspaces.

THEOREM 4.25. *Let V and W be subspaces of $\mathbb{R}^n$. Then*

$$\dim V + W = \dim V + \dim W - \dim V \cap \dim W.$$

PROOF. Let $r = \dim V \cap W$, $k = \dim V$, $\ell = \dim W$. Let $v_1, \ldots, v_r$ be a basis for $V \cap W$, and extend it to a basis $v_1, \ldots, v_r, u_{r+1}, \ldots, u_k$ for V and again to a basis $v_1, \ldots, v_r, w_{k+1}, \ldots, w_\ell$ for W. We claim that

$$\mathcal{B} = \{v_1, \ldots, v_r, u_{r+1}, \ldots, u_k, w_{r+1}, \ldots, w_\ell\}$$

is a basis for $V + W$. If so, then indeed

$$\dim V + W = \dim V + \dim W - \dim V \cap W$$

and the result follows. Since $\mathcal{B}$ is obtained as the union of spanning sets for V and W, it is a spanning set for $V + W$, so it remains to show that $\mathcal{B}$ is linearly independent. Let $\alpha_1, \ldots, \alpha_\ell \in \mathbb{R}$ be such that

$$\alpha_1 v_1 + \ldots + \alpha_r v_r + \alpha_{r+1} u_{r+1} + \ldots + \alpha_k u_k + \alpha_{k+1} w_{k+1} + \ldots + \alpha_\ell w_\ell = 0.$$

We rewrite this as

$$z = \alpha_1 v_1 + \ldots + \alpha_r v_r + \alpha_{r+1} u_{r+1} + \ldots + \alpha_k u_k = -\alpha_{k+1} w_{k+1} - \ldots - \alpha_\ell w_\ell.$$

The left hand side lies V and the right hand side lies in W, so $z \in V \cap W$. Since $v_1, \ldots, v_r, u_{r+1}, \ldots, u_k$ is linearly independent, this implies $\alpha_{r+1} = \ldots = \alpha_k = 0$. Since $v_1, \ldots, v_w, w_1, \ldots, w_\ell$ is linearly independent, this implies $\alpha_1 = \ldots = \alpha_r = \alpha_{k+1} = \ldots = \alpha_\ell = 0$. □

THEOREM 4.26. *For subspaces V and W of $\mathbb{R}^n$, the following are equivalent:*
(i) $\dim V + W = \dim V + \dim W$.
(ii) $V \cap W = \{0\}$.
(iii) If $v \in V \setminus \{0\}$ and $w \in W \setminus \{0\}$, then the sequence (v, w) is linearly independent.
(iv) If L_1 is a linearly independent list in V and L_2 is a linearly independent list in W, then (L_1, L_2) is a linearly independent list.
(v) If $\mathcal{B}_1$ is an ordered basis for V and $\mathcal{B}_2$ is an ordered basis for W, then $(\mathcal{B}_1, \mathcal{B}_2)$ is an ordered basis for $V + W$.

PROOF. (i) $\iff$ (ii) follows from Theorem 4.25: since $\dim V + \dim W - \dim(V+W) = \dim V \cap W$, we have $\dim V + \dim W = \dim V + W$ if and only if $\dim V \cap W = 0$.
(ii) $\iff$ (iii): If v is a nonzero vector in V and w is a nonzero vector in W, then (v, w) is linearly dependent if and only if $w = \alpha v$ for some nonzero $\alpha \in \mathbb{R}$. If this happens, then w is a nonzero vector in $V \cap W$, so $\dim V \cap W \geq 1$. Conversely, if $\dim V \cap W \geq 1$, then taking any nonzero $v \in V \cap W$ we get a linearly dependent sequence (v, v) with $v \in V$ and $v \in W$.
(ii) $\implies$ (iv): Let $L_1 = (v_1, \ldots, v_k)$ and $L_2 = (w_1, \ldots, w_\ell)$, and suppose $\alpha_1, \ldots, \alpha_{k+\ell}$ are real numbers such that

$$\alpha_1 v_1 + \ldots + \alpha_k v_k + \alpha_{k+1} w_1 + \ldots + \alpha_{k+\ell} w_\ell = 0.$$

Equivalently,

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = -\alpha_{k+1} w_1 - \ldots - \alpha_{k+\ell} w_\ell.$$

If the left hand side is zero, then since L_1 is linearly independent, $\alpha_1 = \ldots = \alpha_k = 0$, and then also the right hand side is zero and by linear independence of L_2, $\alpha_{k+1} = \ldots = \alpha_{k+\ell} = 0$. Similarly if the right hand side is zero. The only other possibility is that both sides are nonzero, and thus we get a nonzero vector in $V \cap W$, contradicting (ii).

(iv) $\implies$ (v): By applying (iv) we get that $(\mathcal{B}_1, \mathcal{B}_2)$ is linearly independent. And we always get a spanning set for $V + W$ by taking the union of a spanning set of V and a spanning set of W.
(v) $\implies$ (i): By contraposition: if $V \cap W$ were nonempty, then any nonzero vector would be part of an ordered basis $\mathcal{B}_1$ for V and an ordered basis $\mathcal{B}_2$ for W and then $(\mathcal{B}_1, \mathcal{B}_2)$ is a linearly dependent list, contradicting our assumption. □

We describe the equivalent conditions of Theorem 4.26 by saying that V and W are **independent subspaces**, and in this case we sometimes write $V \oplus W$ instead of $V + W$. This notation is analogous to writing $S \coprod T$ for the union of two sets which are known to be disjoint.

Suppose now that we have k subspaces $V_1, \ldots, V_k \subset \mathbb{R}^n$. We would like an analogous notion of independence. Here however we need to be more careful: to have the subspaces be pairwise disjoint – i..e, $V_i \cap V_j = \{0\}$ for all $i \neq j$ – is not enough. For instance, any family of lines through the origin in $\mathbb{R}^2$ will satisfy this condition, but if there are more than two of them we do not wish to consider this family as independent.

The correct condition can be motivated, as follows: suppose we have three subspaces V_1, V_2, V_2 of $\mathbb{R}^n$ and are trying to show that

$$\dim(V_1 + V_2 + V_3) = \dim V_1 + \dim V_2 + \dim V_3.$$

It is natural to proceed inductively and thus to consider the subspaces $V_1 + V_2$ and V_3. If V_1 and V_2 are disjoint, then $\dim V_1 + V_2 = \dim V_1 + \dim V_2$. But now we want V_3 to be disjoint, not just from V_1 and V_2 but from their sum $V_1 + V_2$. If so, then again we get $\dim(V_1 + V_2 + V_3) = \dim V_1 + \dim V_2 + \dim V_3$. So in general we want each subspace V_i to be disjoint from the subspace spanned by all the others.

THEOREM 4.27. *For subspaces $V_1, \ldots, V_k$ of $\mathbb{R}^n$, the following are equivalent:*
(i) We have $\dim(V_1 + \ldots + V_k) = \dim V_1 + \ldots + \dim V_k$.
(ii) For all $1 \leq i \leq k$, *we have that* $V_i \cap (V_1 + \ldots + V_{i-1} + V_{i+1} + \ldots + V_k) = \{0\}$.
(iii) If we choose for all $1 \leq i \leq k$ *a nonzero vector* $w_i \in V_i$, *then* $w_1, \ldots, w_k$ *is a linearly independent list.*
(iv) If we choose for all $1 \leq i \leq k$ *a linearly independent list* L_i *in* V_i, *then the list* $(L_1, \ldots, L_k)$ *is linearly independent.*
(v) If we choose for all $1 \leq i \leq k$ *an ordered basis* $\mathcal{B}_i$ *for* V_i, *then* $(\mathcal{B}_1, \ldots, \mathcal{B}_k)$ *is an ordered basis for* $\mathbb{R}^n$.
We say $V_1, \ldots, V_k$ *are an* ***independent family of subspaces*** *when these conditions are satisfied.*

PROOF. We have that (iv) $\implies$ (v) $\implies$ (i) and also (iv) $\implies$ (iii).
(ii) $\implies$ (iii): For $1 \leq i \leq k$ choose a nonzero vector w_i in V_i. If $w_1, \ldots, w_k$ is linearly dependent, then some vector W_i can be written as a linear combination of the others, which gives a nonzero vector in $V_i \cap (V_1 + \ldots + V_{i-1} + V_{i+1} + \ldots + V_k)$.
(iii) $\implies$ (ii): This is almost identical to the proof of (ii) $\implies$ (iii) and is left to the reader.
(ii) $\implies$ (i): We show this by induction on k, the case of $k = 1$ being trivial and the case of $k = 2$ being Theorem 4.26. Suppose $k \geq 3$ and the result holds for any family of fewer than k subspaces. Put $W \coloneqq (V_1 + \ldots + V_{k-1}$. Then since each V_i is disjoint from the sum of the others, by our inductive hypothesis we have $\dim W = \dim V_1 + \ldots + \dim V_{k-1}$. Our hypothesis also implies that W and V_k are disjoint, so applying the result for $k = 2$ gives

$$\dim(V_1 + \ldots + V_k) = \dim(W + V_k) = \dim W + \dim V_k = \dim V_1 + \ldots + \dim V_{k-1} + \dim V_k.$$

(i) $\implies$ (iv): Observe that if $\dim(V_1 + \ldots + V_k) = \dim V_1 + \ldots + \dim V_k$ then the same holds for any nonempty subset of $V_1, \ldots, V_k$: this is because for any subspace Z we have

$$\dim(Z + V_i) \leq \dim Z + \dim V_i.$$

For $1 \leq i \leq k$ let L_i be a linearly independent list in V_i, and let $L' = (L_1, \ldots, L_{k-1})$. If we put $W = V_1 + \ldots + V_{k-1}$, then induction gives us that the list L' is linearly indpendent, and then applying the $k = 2$ case to W and V_k gives that $(L', L_k) = (L_1, \ldots, L_k)$ is linearly independent. □

4.10. Rank Revisited.

For any matrix $A \in M_{m,n}$, we defined $\operatorname{rank} A$ to be the number of leading entries in $\operatorname{rref} A$ (or in any row echelon form of A). And we have seen the uses of that definition in solving systems of linear equations and elsewhere. However, one must admit that our definition of rank is not the most graceful one: it is a number that we associate to a matrix after performing a certain algorithm (Gaussian reduction) on it. In this section we pursue other, more intrinsic definitions of rank, with applications to an improved Rank Nullity Theorem and the important result that $\operatorname{rank} A = \operatorname{rank} A^T$ for all A.

Recall that the **row space** $R(A)$ of $A \in M_{m,n}$ is the subspace of $\mathbb{R}^n$ spanned by the rows of A. Elementary row operations change the rows but not the row space: one way to see this is to think of an elementary row operation as premultiplication by an elementary matrix: $A \mapsto EA$. In general, the rows of BA are linear combinations of the rows of A – c.f. § 2.3 – so this shows that the row space of EA is contained in the row space of A. But E is invertible, and applying the same argument with E^{-1} in place of E shows that the row space of $A = E^{-1}(EA)$ is contained in the row space of EA, so the row spaces of A and EA are equal.

Perhaps the argument of the preceding paragraph is too slick for its own good. The reader may prefer a more hands-on approach:

EXERCISE 4.10. *Check more concretely that the row space of EA is equal to the row space of A for every elementary matrix by considering each of the three elementary row operations separately, and for each one, explicitly writing each row of EA as a linear combination of the rows of A, and explicitly writing each row of A as a linear combination of the rows of EA.*

Anyway, since there are elementary matrices $E_1, \ldots, E_r$ such that

$$\operatorname{rref}(A) = E_r \cdots E_1 A,$$

we deduce the following.

PROPOSITION 4.28. *For any $A \in M_{m,n}$, the row space of A is equal to the row space of* $\operatorname{rref} A$.

Along with the row space of $A \in M_{m,n}$ we can also consider its **column space** $C(A)$, the subspace of $\mathbb{R}^m$ spanned by the columns of A. But beware:

Warning: Elementary row operations *need not* preserve the column space!

EXAMPLE 4.29. *Take $A \in M_{m,1}$, i.e., a column vector. Then so long as A is not the zero matrix,* $\operatorname{rref} A = e_1 = (1, \ldots, 0)$. *Thus the column space of* $\operatorname{rref} A$ *is the line spanned by e_1. But the matrix A we started with could be given by any nonzero vector, so its span need not be the span of e_1.*

Back to the row space and the rank. Here is the first important result.

THEOREM 4.30. *For all $A \in M_{m,n}$, we have*

$$\operatorname{rank}(A) = \dim R(A).$$

That is, the rank is equal to the dimension of the row space.

PROOF. Since A and $\operatorname{rref} A$ have the same rank (by definition) and the same row space (by Proposition 4.28), it is enough to show that the number of leading entries in $\operatorname{rref} A$ is equal to $\operatorname{rank} \operatorname{rref} A$. I claim that a basis for $\operatorname{rref} A$ is obtained by taking the nonzero rows. Indeed, the nonzero rows certainly span the row space: the zero rows contribute nothing to the span. Moreover they are linearly independent because each row contains a leading entry: we can be sure that a finite set $v_1, \ldots, v_k$ of vectors is linearly independent if, as we move from left to right, each v_i has a nonzero entry in a coordinate where all the vectors to the left of it have zero entries (this observation is formalized in the following exercise and you are asked to prove it: it's not hard!). So indeed the nonzero rows of $\operatorname{rref} A$ form a basis for the row space. Since the number of nonzero rows of $\operatorname{rref} A$ is precisely $\operatorname{rank} A$, we're done. □

EXERCISE 4.11. *Prove the **Eyeball Criterion for Linear Independence**: Let $v_1, \ldots, v_k$ be vectors in $\mathbb{R}^n$. Suppose that for all $1 \leq i \leq k$, there is some coordinate (i.e, $1 \leq j \leq n$) such that v_i is nonzero in the jth coordinate, but all $v_{i'}$ with $i' < i$ are zero in the jth coordinate. Show that $\{v_1, \ldots, v_k\}$ is a linearly independent set.*

Thus we can think of row reduction as taking the hard-to-see property of linear independence of a set of vectors and reworking it until it is visible to the naked eye. This also gives a method for testing a finite list of vectors $v_1, \ldots, v_k$ for linear independence: make a matrix with these vectors as *row vectors* and row reduce: if the rref has rank k, the vectors are linearly independent; otherwise they are linearly dependent. Is this interesting? We already know how to row reduce matrices to test for linear independence.

But, wait – this is a **different test** from the one we've seen before: according to the definition of linear independence, $v_1, \ldots, v_k$ are linearly independent if and only if we when we make a matrix with these vectors as *column vectors*, then the rref has no free variables. That's not the same test, and we have begun to uncover the deep relationship between the rows and columns of a matrix.

Let's push it farther: recall that the **nullity** $\operatorname{nullity} A$ of $A \in M_{m,n}$ is the number of free variables in $\operatorname{rref} A$. Since every variable is either a free variable or a pivot variable, we get the Rank-Nullity Theorem:

$$\forall A \in M_{m,n},\ \operatorname{rank} A + \operatorname{nullity} A = n.$$

Let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^n$; let $A \in M_{m,n}$ be the matrix with ith row v_i. Then:

$$\begin{aligned} &\operatorname{rank} A = m \\ \iff &v_1, \ldots, v_m \text{ is linearly independent} \\ \iff &\operatorname{nullity}(A^T) = 0 \\ \iff &\operatorname{rank} A^T = m. \end{aligned}$$

Thus $A \in M_{m,n}$ has rank m iff A^T has rank m.

Let's go even farther: let $A \in M_{m,n}$ have rank k. Then there are some k rows of A which are linearly independent: let $B \in M_{k,n}$ be the matrix formed by these rows. By what we just said, $\operatorname{rank} B^T = k$. But sine B^T simply consists of some of the rows of A^T, we must have

$$\operatorname{rank} A = k = \operatorname{rank} B = \operatorname{rank} B^T \leq \operatorname{rank} A^T.$$

So for *any* matrix, $\operatorname{rank} A \leq \operatorname{rank} A^T$. Applying this inequality to A^T we get $\operatorname{rank} A^T \leq \operatorname{rank}(A^T)^T = A$. Thus, we've proven:

THEOREM 4.31. *For any matrix $A \in M_{m,n}$,*

$$\operatorname{rank} A = \operatorname{rank} A^T.$$

Equivalently,

$$\dim R(A) = \dim C(A).$$

The rank of A^T is often called the **column rank** of A (as seems reasonable). Thus Theorem 4.31 is often abbreviated as **row rank equals column rank**. This is actually one of the least obvious of the fundamental theorems relating dimensions of subspaces, because the row space $R(A)$ and the column space $C(A)$ of A are usually not in any sense "the same" subspace: indeed, $R(A)$ is a subspace of $\mathbb{R}^n$ and $C(A)$ is a subspace of $\mathbb{R}^m$. Nevertheless these subspaces of *different Euclidean spaces* always have the same dimension. That's quite a deep result!

We want to give yet another interpretation of $\operatorname{rank} A$. Namely, consider the associated linear transformation $L_A : \mathbb{R}^n \to \mathbb{R}^m$, $x \mapsto Ax$.

THEOREM 4.32. *For any $A \in M_{m,n}$, $\operatorname{rank} A = \dim \operatorname{Image} L_A$.*

PROOF. The proof will use the (very easy) fact that the null space of A is the kernel of L_A, so that by Rank-Nullity, $\dim \operatorname{Ker} L_A + \operatorname{rank} A = n$ and thus

$$\operatorname{rank} A = n - \dim \operatorname{Ker} L_A. \tag{6}$$

Really what we will show is

$$\dim \operatorname{Image} L_A = n - \dim \operatorname{Ker} L_A. \tag{7}$$

And of course, combining (6) and (7) gives $\operatorname{rank} A = \dim \operatorname{Image} L_A$. To show (7) we argue as follows: let $k = \dim \operatorname{Ker} L_A$, so there is a basis $v_1, \ldots, v_k$ of $\operatorname{Ker} L_A$. Like any basis for a subspace, we can extend this to a basis for all of $\mathbb{R}^n$, say $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$. Now I claim that $L(v_{k+1}), \ldots, L(v_n)$ is a basis for $\operatorname{Image} L_A = L_A(\mathbb{R}^n)$: if so, we're done, because we have $n-k$ vectors in our basis, so $\dim \operatorname{Image} L_A = n - k = n - \dim \operatorname{Ker} L_A$. Now $\operatorname{Image} L_A$ is spanned by $L(v_1), \ldots, L(v_n)$, and since $L(v_1) = \ldots = L(v_k) = 0$, it is certainly also spanned by $L(v_{k+1}), \ldots, L(v_n)$. So it remains to check the linear independence: suppose we have $\alpha_{k+1}, \ldots, \alpha_n$ such that

$$\alpha_{k+1} L(v_{k+1}) + \ldots + \alpha_n L(v_n) = 0.$$

Then

$$0 = \alpha_{k+1} L(v_{k+1}) + \ldots + \alpha_n L(v_n) = L(\alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n),$$

so $\alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n \in \operatorname{Ker} L_A$. Since $\operatorname{Ker} L_A$ is spanned by $v_1, \ldots, v_k$, there are $\alpha_1, \ldots, \alpha_k$ such that

$$\alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n = \alpha_1 v_1 + \ldots + \alpha_k v_k,$$

or

$$(-\alpha_1) v_1 + \ldots + (-\alpha_k) v_k + \alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n.$$

Since $v_1, \ldots, v_n$ is linearly independent this means $-\alpha_1 = \ldots = -\alpha_k = \alpha_{k+1} = \ldots = \alpha_n = 0$; this shows that $L(v_{k+1}), \ldots, L(v_n)$ is linearly independent. □

Here is an equivalent statement.

THEOREM 4.33. *(Dimension Theorem) For any linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$, we have*

$$\dim \operatorname{Ker} L + \dim \operatorname{Image} L = \dim \mathbb{R}^n.$$

EXERCISE 4.12. *Deduce Theorem 4.33 from Theorem 4.32.*

The Rank-Nullity Theorem and the Dimension Theorem are really *the same result*; the former is couched in matrix language, the latter in the language of linear transformations. This is a common theme in linear algebra: many if not most results can be expressed either way. This is useful because linear transformations are more conceptual and thus ultimately more useful for theorems and proofs, whereas it is critical to be able to phrase things in terms of matrices to do calculations.

As an example of this duality, let us use Theorem 4.32 (and Proposition 4.28) to give another proof that row rank equals column rank. Namely, for $A \in M_{m,n}$,

$$\dim R(A) = \operatorname{rank} A = \dim \operatorname{Image} L_A = \dim \operatorname{span} L_A(e_1), \ldots, L_A(e_n) = \operatorname{span} A(e_1), \ldots, A(e_n).$$

But $A(e_1), \ldots, A(e_n)$ are precisely the columns of A, so

$$\dim \operatorname{span} A(e_1), \ldots, A(e_n) = \dim C(A) = \operatorname{rank} A^T.$$

4.11. Sylvester's Law of Nullity.

THEOREM 4.34. *(Sylvester's Law of Nullity) Let $A \in M_{m,n}$ and $B \in M_{n,p}$. Then we have*

$$\text{nullity } AB \leq \text{nullity } A + \text{nullity } B. \tag{8}$$

PROOF. If $Bv = 0$ then $ABv = A0 = 0$, so nullity $B \subset$ nullity AB. Let $v_1, \ldots, v_a$ be a basis for B, and extend it to a basis $v_1, \ldots, v_a, w_1, \ldots, w_b$ for AB. We claim that $B(w_1), \ldots, B(w_b)$ is linearly independent. The argument for this is essentially the same one used in the proof of Theorem 4.32: let $\alpha_1, \ldots, \alpha_b \in \mathbb{R}$ be such that $\alpha_1 B(w_1) + \ldots + \alpha_b B(w_b) = 0$. Then

$$0 = \alpha_1 B(w_1) + \ldots + \alpha_a B(w_b) = B(\alpha_1 w_1 + \ldots + \alpha_b w_b),$$

so $\alpha_1 w_1 + \ldots + \alpha_b w_b \in B$. Thus there are $\beta_1, \ldots, \beta_a \in \mathbb{R}$ such that

$$\alpha_1 w_1 + \ldots + \alpha_b w_b = \beta_1 v_1 + \ldots + \beta_a v_a,$$

and by linear independence of $v_1, \ldots, v_a, w_1, \ldots, w_b$, this gives $0 = \alpha_1 = \cdots = \alpha_b (= \beta_1 = \ldots = \beta_a$, though we don't need this). Since $B(w_1), \ldots, B(w_b) \subset A$, this gives

$$\text{nullity } A \geq b,$$

and thus

$$\text{nullity } A + \text{nullity } B \geq b + \text{nullity } B = b + a = \text{nullity } AB. \qquad \square$$

EXERCISE 4.13. *Maintain the notation of Theorem 4.34. Use the Rank-Nullity Theorem to deduce the following additional inequalities.*
*a) (**Sylvester's Rank Inequality**) We have* $\text{rank } AB \geq \text{rank } A + \text{rank } B \leq \text{rank } AB + n$.
b) We have $\text{rank } B \leq \text{rank } AB + \text{nullity } A$.

COROLLARY 4.35.
a) Let $A_1, \ldots, A_N$ be matrices such that the product $A_1 \cdots A_N$ is defined. Then we have

$$\text{nullity } A_1 \cdots A_N \leq \sum_{i=1}^{N} \text{nullity } A_i.$$

b) Let $A_1, \ldots, A_N \in M_{n,n}$ be such that $A_1 \cdots A_N = 0$. Then $\sum_{i=1}^{N} \text{nullity } A_i \geq n$.

EXERCISE 4.14. *Prove Corollary 4.35.*

4.12. Ackerson's Theorem.

THEOREM 4.36 (Ackerson [**Ac55**]). *Let $A \in M_{n,n}$. Let K be the null space of A, let R be the image of A, let S be the image of A^2, and let $I \coloneqq K \cap R$. Then*

$$\dim I = \dim R - \dim S.$$

PROOF. Let $k = \dim I$, and let $z_1, \ldots, z_k$ be a basis for I. Since I is a subspace of the image of A, there are $x_1, \ldots, x_k \in \mathbb{R}^n$ such that $z_i = Ax_i$ for $1 \leq i \leq k$. If $r = \text{rank } A$, then we may extend to a basis $Ax_1, \ldots, Ax_r$ of R.

Let $x \in \mathbb{R}^n$. Then there are $\alpha_1, \ldots, \alpha_r$ such that

$$Ax = \alpha_1 Ax_1 + + \ldots + \alpha_k Ax_k + \alpha_{k+1} Ax_{k+1} + \ldots + \alpha_r Ax_r,$$

so

$$\begin{aligned} A^2x &= \alpha_1 A^2 x_1 + \ldots + \alpha_k A^2 x_k + \alpha_{k+1} A^2 x_{k+1} + \ldots + \alpha_r A^2 x_r \\ &= \alpha_{k+1} A^2 x_{k+1} + \ldots + \alpha_r A^2 x_r. \end{aligned}$$

This shows that $A^2x_{k+1}, \ldots, A^2x_r$ span S. Suppose there are $\beta_{k+1}, \ldots, \beta_r \in \mathbb{R}$ such that

$$\beta_{k+1} A^2 x_{k+1} + \ldots + \beta_r A^2 x_r = 0.$$

Then

$$A(\beta_{k+1} Ax_{k+1} + \ldots + \beta_r Ax_r) = 0,$$

so

$$y := \beta_{k+1}Ax_{k+1} + \ldots + \beta_r Ax_r \in K \cap R = I.$$

This means that y is a linear combination of $Ax_1, \ldots, Ax_k$. Since $Ax_1, \ldots, Ax_k, Ax_{k+1}, \ldots, Ax_r$ is linearly independent, it follows that $\beta_{k+1} = \ldots = \beta_r = 0$. Thus $A^2x_{k+1}, \ldots, A^2x_r$ form a basis for S and we get

$$\dim S = r - k = \dim R - \dim I$$

and thus

$$\dim I = \dim R - \dim S.$$ □

COROLLARY 4.37. *Maintain the above notation.*
a) We have $K \subset R$ iff $\dim K = \dim R - \dim S$.
b) We have $R \subset K$ iff $A^2 = 0$.
c) We have $R = K$ iff $A^2 = 0$ and $\dim K = \dim R$.

PROOF. a) We have $K \subset R$ iff $K \cap R = K$ iff $\dim K = \dim R - \dim S$.
b) We have $R \subset K$ iff $K \cap R = R$ iff $\dim R = \dim R - \dim S$ iff $\dim S = 0$ iff $A^2 = 0$.
c) By parts a) and b) we ave $R = N$ iff $A^2 = 0$ and $\dim K = \dim R - \dim S = \dim R$.

□

COROLLARY 4.38. *Maintain the above notation. The following are equivalent:*
(i) $I = 0$.
(ii) $R = S$.
(iii) $\mathbb{R}^n = K \oplus R$.

PROOF. (i) $\implies$ (ii): Since $I = 0$, we have $\dim R = \dim S$. Since $S \subset R$, we get $R = S$.
(ii) $\implies$ (iii) By the Dimension Theorem we have

$$n = \dim K + \dim R.$$

If $R = S$, then $\dim I = \dim R - \dim S = 0$, so $I = 0$ and thus R and K are independent subspaces, so $\dim\langle R, K\rangle = \dim R + \dim K = n$, and thus $\langle R, K\rangle = \mathbb{R}^n$.
(iii) $\implies$ (i): This is immediate: $I = K \cap R = 0$. □

5. Some Linear Transformations

5.1. Permutations.

5.2. Projections.

Let V, W be subspaces of $\mathbb{R}^n$ with $V \oplus W = \mathbb{R}^n$. Recall this means: $V + W = \mathbb{R}^n$ and $V \cap W = \{0\}$. We define a linear transformation $\pi_{V,W}$, the **projection onto V with respect to W**, as follows: the complementarity of V and W means precisely that every $x \in \mathbb{R}^n$ can be written uniquely as $x = v + w$ for $v \in v$, $w \in W$. Then we define

$$\pi_{V,W}(x) = \pi_{V,W}(v + w) = v.$$

In ohter words, the map $\pi_{V,W}$ resolves a vector into the sum of its V component and its W component, keeps the V component and kills the W-component.

PROPOSITION 5.1. *The projection map $\pi_{V,W} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation.*

PROOF. As usual, this is straightforward. Namely, let $x_1, x_2 \in \mathbb{R}^n$ and write $x_1 = v_1 + w_1$, $x_2 = v_2 + w_2$ with $v_i \in V$ and $w_i \in W$ for $i = i, 2$. Then $x_1 + x_2 = (v_1 + v_2) + (w_1 + w_2)$. It follows that

$$\pi_{V,W}(x_1 + x_2) = v_1 + v_2 = \pi_{V,W}(x_1) + \pi_{V,W}(x_2).$$

Similarly, for any $\alpha \in \mathbb{R}$, if $x = v + w$ then

$$\alpha x = \alpha v + \alpha w,$$

so

$$\pi_{V,W}(\alpha x) = \pi_{V,W}(\alpha v + \alpha w) = \alpha v = \alpha \pi_{V,W}(v).$$

□

EXAMPLE 5.2. *Let V be the span of e_1 in $\mathbb{R}^2$, i.e., the line $y = 0$. Let w be any vector in $\mathbb{R}^2 \setminus V$, and put $W = \langle w \rangle$. Then $V \cap W = \{0\}$, hence*

$$\dim V + W = \dim V + \dim W \setminus \dim V \cap W = 1 + 1 - 0 = 2,$$

and thus $V + W = \mathbb{R}^2 = V \oplus W$. Let's find the standard matrix representation of $\pi_{V,W}$, i.e., the 2×2 matrix with columns $\pi_{V,W}(e_1)$ and $\pi_{V,W}(e_2)$. First,

$$\pi_{V,W}(e_1) = e_1.$$

To find $\pi_{V,W}(e_2)$, the key idea is to express e_2 as a linear combination of e_1 and w. For this, write $w = (x, y)$; since $w \notin \langle e_1 \rangle$, we must have $y \neq 0$. Then

$$w = xe_1 + ye_2,$$

and we solve for e_2, getting

$$e_2 = \frac{1}{y}w - \frac{x}{y}e_1$$

and thus

$$\pi_{V,W}(e_2) = \frac{1}{y}\pi_{V,W}(w) - \frac{x}{y}\pi_{V,W}(e_1) = 0 - \frac{x}{y} = \frac{-x}{y}.$$

So the standard matrix is

$$A = \begin{bmatrix} 1 & \frac{-x}{y} \\ 0 & 0 \end{bmatrix}.$$

Some things to notice about this matrix are: (i) it has a row of zeros so is not invertible. This is not surprising, since the linear transformation had a nontrivial kernel: it killed w. We also see that A takes the simplest form if $x = 0$, in other words if $W = \langle (0, y) \rangle = \langle e_2 \rangle$ is just the y-axis. In this special case we have

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

a diagonal matrix. We get this case by taking w to be perpendicular to $v = e_1$.

There is a nice way of thinking about projection operators in terms of splitting a basis. Namely, let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for $\mathbb{R}^n$. We split it in two, namely we choose subsets $\mathcal{B}_1, \mathcal{B}_2 \subset \mathcal{B}$ such that $\mathcal{B}_1 \cup \mathcal{B}_2 = \mathcal{B}, \mathcal{B}_1 \cap \mathcal{B}_2 = \varnothing$. Then $\mathcal{B}_1$ and $\mathcal{B}_2$ are both subsets of a linearly independent set, hence linearly independent, hence $\mathcal{B}_1$ is a basis for $V_1 = \operatorname{span} \mathcal{B}_1$ and $\mathcal{B}_2$ is a basis for $V_2 = \operatorname{span} \mathcal{B}_2$.

EXERCISE 5.1. *With notation as above, show: $\mathbb{R}^n = V_1 \oplus V_2$.*

Thus we can define a projection operator π_{V_1,V_2} associated to the splitting of the basis. Moreover, like any linear transformation, π_{V_1,V_2} is uniquely specified by what it does to the elements of any basis, so let's see what it does to $\mathcal{B}$. Well, we have split $\mathcal{B}$ into $\mathcal{B}_1$ and $\mathcal{B}_2$, and π_{V_1,V_2} keeps the elements of $\mathcal{B}_1$ and kills the elements of $\mathcal{B}_\in$:

$$\forall b_1 \in \mathcal{B}_1,\ \pi_{V_1,V_2}(b_1) = b_1,$$
$$\forall b_2 \in \mathcal{B}_2,\ \pi_{V_1,V_2}(b_2) = 0.$$

Next we observe that any projection has the property that doing it and then doing it again is the same as doing it once:

$$\pi_{V,W} \circ \pi_{V,W} = \pi_{V,W}. \tag{9}$$

Indeed, for any $x = v + w \in \mathbb{R}^n$,

$$(\pi_{V,W} \circ \pi_{V,W})(x) = \pi_{V,W}(\pi_{V,W}(v + w)) = \pi_{V,W}(v) = v = \pi_{V,W}(x).$$

Let A be the standard matrix of $\pi_{V,W}$. Since composition of linear transformations corresponds to multiplication of matrices, the matrix version of (10) is

$$A^2 = A. \tag{10}$$

A matrix $A \in M_{n,n}$ satisfying (10) is called **idempotent**. The equation $A^2 = A$ looks a little silly from the perspective of high school algebra: if A is a real number it has precisely the solutions $A = 0$ and $A = 1$. Equivalently, these are the only idempotent 1×1 matrices. However, for any $n \geq 2$ there are many, many more solutions $A \in M_{n,n}$: namely the standard matrix associated to any projection operator.

EXERCISE 5.2. *Show: if A is idempotent, so is $1 - A$.*

EXERCISE 5.3. *a) It follows from our discussion so far that if $A = \begin{bmatrix} 1 & b \\ 0 & 0 \end{bmatrix}$ for any $b \in \mathbb{R}$, then $A^2 = A$. Check this directly.*
b) Write down an idempotent matrix $A \in M_{3,3}$ different from 0 and 1.
c) Write down a nondiagonal idempotent matrix $A \in M_{3,3}$.

That the standard matrix of a projection operator is idempotent is an instance of geometry (linear transformations) governing algebra (matrix operations). It seems a bit more interesting that the converse is also true.

THEOREM 5.3. *Let $A \in M_{n,n}$ be an idempotent matrix: $A^2 = A$. Then:*
a) We have $\mathrm{Ker}(1 - A) = \mathrm{Image}(A)$ and $\mathrm{Ker}\, A = \mathrm{Image}(1 - A)$.
b) We have $\mathbb{R}^n = \mathrm{Image}(A) \oplus \mathrm{Image}(1 - A)$.
c) A is the standard matrix of the projection operator $\pi_{\mathrm{Image}(A),\mathrm{Image}(1-A)}$.

PROOF. a) If $v \in \mathrm{Ker}(1-A)$, then $0 = (1-A)v = v - Av$, so $Av = v$, so $v \in \mathrm{Image}(A)$. Conversely, if $v \in \mathrm{Image}(A)$ then $v = Aw$ for some $w \in \mathbb{R}^n$, and then $(1 - A)v = (1 - A)(Aw) = Aw - A^2w = Aw - Aw = 0$. This shows

$$\mathrm{Ker}(1 - A) = \mathrm{Image}(A).$$

The equality $\mathrm{Ker}\, A = \mathrm{Image}(1 - A)$ can be shown similarly, or by applying the above argument with $1 - A$ in place of A, which is valid since $1 - A$ is also idempotent by Exercise 5.2.
b) Suppose $v \in \mathrm{Image}(A) \cap \mathrm{Image}(1 - A) = \mathrm{Image}(A) \cap \mathrm{Ker}(A)$. Then $v = Aw$ for some $w \in \mathbb{R}^n$ and also $Av = 0$, so

$$0 = Av = A(Aw) = A^2w = Aw = v.$$

This shows that $\mathrm{Image}(A) \cap \mathrm{Image}(1 - A) = 0$. To see that $\mathrm{Image}(A) + \mathrm{Image}(1 - A) = \mathbb{R}^n$ is even easier: we may write any $x \in \mathbb{R}^n$ as

$$x = Ax + (1 - A)x.$$

c) It is enough to see that if $x \in \mathrm{Image}(A)$ then $Ax = x$ and if $x \in \mathrm{Image}(1 - A)$ then $Ax = 0$. Really we've done this already, but nce again: if $x = Aw$ for some $w \in \mathbb{R}^n$ then $Ax = A(Aw) = A^2w = Aw = x$, and if $x = (1 - A)w$,then $Ax = A(1 - A)w = (A - A^2)w = 0$. □

5.3. Reflections. The above discussion of projections is easily modified so as to apply to another class of operators, the **reflections**. We begin in the same way, with a decomposition of $\mathbb{R}^n$ into complementary subspaces $\mathbb{R}^n = V_1 \oplus V_2$ and choices of bases $\mathcal{B}_1$ of V_1 and $\mathcal{B}_2$ of V_2, so that

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$$

is a basis for $\mathbb{R}^n$. Whereas to get π_{V_1,V_2} we *kept* the elements of $\mathcal{B}_1$ and *killed* the elements of $\mathcal{B}_2$, we now define a **reflection** τ_{V_1,V_2} by:

$$\forall b_1 \in \mathcal{B}_1,\ \tau(b_1) = b_1,$$
$$\forall b_2 \in \mathcal{B}_2,\ \tau(b_2) = -b_2.$$

That is, instead of killing the basis elements of V_2, we **flip** them.
In more algebraic terms, we write any x in $\mathbb{R}^n$ uniquely as $v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$ and put

$$\tau_{V_1,V_2}(x) = \tau_{V_1,V_2}(v_1 + v_2) = v_1 - v_2.$$

Now let A be the standard matrix of the reflection τ_{V_1,V_2}. Above we found that the geometry of projection was faithfully recorded in the simple algebraic equation $A^2 = A$, so it is natural to ask whether the same kind of thing will hold for reflections. The answer is yes.

PROPOSITION 5.4. *a) Let $\mathbb{R}^n = V_1 \oplus V_2$, and let τ_{V_1,V_2} be the corresponding reflection operator. Then*

$$\tau_{V_1,V_2} \circ \tau_{V_1,V_2} = 1_{\mathbb{R}^\times}. \tag{11}$$

Equivalently, if A is the standard matrix of τ_{V_1,V_2}, then

$$A^2 = 1. \tag{12}$$

b) Conversely, if $A \in M_{n,n}$ is such that $A^2 = 1$, then A is the standard matrix of a projection operator π_{V_1,V_2}, with

$$V_1 = \{x \in \mathbb{R}^n \mid Ax = x\}$$

and

$$V_2 = \{x \in \mathbb{R}^n \mid Ax = -x\}.$$

PROOF. a) For $x \in \mathbb{R}^n$, write $x = v_1 + v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$. Then

$$(\tau_{V_1,V_2} \circ \tau_{V_1,V_2})(x) = \tau_{V_1,V_2}(\tau_{V_1,V_2}(v_1 + v_2)) = \tau_{V_1,V_2}(\tau_{V_1,V_2}(v_1) + \tau_{V_1,V_2}(v_2))$$

$$= \tau_{V_1,V_2}(v_1 - v_2) = \tau_{V_1,V_2}(v_1) - \tau_{V_1,V_2}(v_2) = v_1 - (-v_2) = v_1 + v_2 = x.$$

This shows (11); since composition of linear operators corresponds to multiplication of matrices, (12) follows immediately.
b) Notice first that we defined V_1 to be the set of vectors "kept" (or, in more common parlance, "fixed") by A and V_2 to be the set of vectors "flipped" by A. A little thought shows that if A is the standard matrix of a reflection operator then V_1 and V_2 have to be as we defined them, and what is left to show is that $\mathbb{R}^n = V_1 \oplus V_2$. If $x \in V_1 \cap V_2$, then $Ax = x$ and also $Ax = -x$, so $x = Ax = -x$ and thus $2x = 0$ so $x = 0$. Now let $x \in \mathbb{R}^n$. Here's a trick: put $y = x + Ax$ and $z = x - Ax$. Then

$$Ay = A(x + Ax) = Ax + A^2x = Ax + x = y,$$

so $y \in V_1$. Also

$$Az = A(x - Ax) = Ax - A^2x = Ax - x = -(x - Ax) = -z,$$

so $z \in V_2$. Finally,

$$x = \frac{1}{2}y + \frac{1}{2}z \in V_1 + V_2.$$

□

Remark: In more advanced linear algebra one works with vector spaces not only over $\mathbb{R}$ but over an arbitrary *field* of scalars. *Most* of linear algebra by its nature carries over to this general context with no changes in the statements of proofs, but there are some exceptions. Here we fundamentally used that $2 = 1 + 1 \neq 0$. If we took as our scalars the field $\mathbb{F}_2$ of two elements (or any field containing it), then we would in fact have $1 + 1 = 0$ and the above argument breaks down. In fact the result becomes false: there are matrices with $A^2 = 1$ which do not correspond to projection operators. The simplest such example is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

6. Determinants

THEOREM 6.1. *For $A \in M_{n,n}$, the following are equivalent:*
(i) The matrix A is singular.
(ii) We have $\det A = 0$.

THEOREM 6.2. *For $A \in M_{n,n}$ we have* $\det A = \det A^T$.

7. Orthogonality

7.1. Projection of a vector onto a line. Let ℓ be a line in $\mathbb{R}^n$ passing through the origin, so ℓ is the span of a nonzero vector $u \in \mathbb{R}^n$. For any vector v in $\mathbb{R}^n$ we define the **projection** of v onto $\ell = \langle u \rangle$ as

$$\mathrm{proj}_\ell\, v = \mathrm{proj}_{\langle u \rangle}\, v = \frac{v \cdot u}{u \cdot u} u.$$

The first thing to check is that this is well-defined: we say the projection is onto the line ℓ, but to define it we chose a nonzero vector u on ℓ. What happens if we chose a different nonzero vector u' on ℓ? Then there is $\alpha \in \mathbb{R} \setminus \{0\}$ such that $u' = \alpha u$, so

$$\frac{v \cdot u'}{u' \cdot u'} u' = \frac{v \cdot (\alpha u)}{(\alpha u) \cdot (\alpha u)} \alpha u = \frac{\alpha^2}{\alpha^2} \frac{v \cdot u}{u \cdot u} = \frac{v \cdot u}{u \cdot u} u.$$

It is easy to see that the map

$$A : v \in \mathbb{R}^n \mapsto \mathrm{proj}_\ell\, v \in \mathbb{R}^n$$

is a linear transformation of $\mathbb{R}^n$. We claim that it is indeed a projection onto the subspace ℓ in the sense of §X.X. First of all the image of this map is ℓ: visibly it is contained in ℓ and for every nonzero $u \in \ell$ the formula gives

$$\mathrm{proj}_\ell\, u = u.$$

This also shows that the transformation is idempotent – i.e., $A^2 = A$ since indeed idempotent linear transformations are those that, upon restriction to their image, are the identity. As we saw, being idempotent is a characteristic property of projection maps. Evidently the kernel is

$$H \coloneqq \{v \in \mathbb{R}^n \mid v \cdot u = 0\},$$

i.e., the set of all vectors perpendicular to ℓ. Since A has rank 1, its nullity is $n-1$, and thus (as we've seen before) H is a linear subspace of dimension $n-1$, i.e., a hyperplane.

Later we will see that for every subspace W of $\mathbb{R}^n$ we can define a special projection onto W, called **orthogonal projection**, which is a projection operator with image W and for which the kernel is the set of vectors perpendicular to every element of W.

7.2. Orthogonal Bases and Gram-Schmidt. A set of vectors $S \subset \mathbb{R}^n$ is **orthogonal** if $0 \notin S$ and for all $v \neq w \in S$ we have $v \cdot w = 0$. A set of vectors $S \subset \mathbb{R}^n$ is **orthonormal** if it is orthogonal and moreover for all $v \in S$ we have $v \cdot v = 1$.

To every nonzero vector $v \in \mathbb{R}^n$ we associate a unit vector $u_v = \frac{v}{||v||} = \frac{v}{\sqrt{v \cdot v}}$.

Exercise 7.1. *Let $S \subset \mathbb{R}^n \setminus \{0\}$ be a set of nonzero vectors in $\mathbb{R}^n$. Put*

$$U(S) \coloneqq \{u_s = \frac{s}{\sqrt{s \cdot s}} \mid s \in S\},$$

the set of associated unit vectors of S.
a) Show: $\langle S \rangle = \langle U(S) \rangle$, i.e., replacing S by its set of associated unit vectors does not change its span.
b) Show: S is linearly independent iff $U(S)$ is linearly independent.
c) Show: S is orthogonal iff $U(S)$ is orthonormal.

Because of Exercise 7.2 the distinction between orthogonal subsets and orthonormal subsets is always a minor one. Here is a more significant result.

Proposition 7.1. *An orthogonal subset $S \subset \mathbb{R}^n$ is linearly independent.*

Proof. We go by contraposition: since $0 \notin S$, if S is linearly dependent then there are distinct vectors $v_1, \ldots, v_n \in S$ and $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$ such that

$$v_n = \alpha_1 v_1 + \ldots + \alpha_{n-1} v_{n-1}.$$

Taking the dot product of both sides with v_n we get

$$||v_n||^2 = \alpha_1 v_1 \cdot v_n + \ldots + \alpha_{n-1} v_{n-1} \cdot v_n = \alpha_1 0 + \ldots + \alpha_{n-1} 0 = 0,$$

a contradiction. □

It follows that an orthogonal subset of $\mathbb{R}^n$ has size at most n, and that an orthogonal subset of $\mathbb{R}^n$ has size n iff it is a basis for $\mathbb{R}^n$.

Do orthogonal bases exist? Well, sure: the standard basis $e_1 = (1, \ldots, 0), \ldots, e_n = (0, \ldots, 1)$ of $\mathbb{R}^n$ is an orthonormal basis. A more interesting question is whether every subspace W of $\mathbb{R}^n$ admits an orthogonal basis B. If so, then by Exercise 7.2 the associated set of unit vectors $U(B)$ is an orthonormal basis.

This is possible, indeed in a canonical way.

THEOREM 7.2 (Gram-Schmidt Process). *Let $v_1, \ldots, v_k$ be linearly independent vectors in $\mathbb{R}^n$, and let $V = \langle v_1, \ldots, v_k \rangle$. We define vectors $w_1, \ldots, w_k$ as follows:*
- $w_1 \coloneqq v_1$.
- $w_2 \coloneqq v_2 - \text{proj}_{w_1} v_2$.
- *Having defined $w_1, \ldots, w_i$, we put*

$$w_{i+1} \coloneqq v_{i+1} - \text{proj}_{w_1} v_{i+1} - \text{proj}_{w_2} v_{i+1} - \ldots - \text{proj}_{w_i} v_{i+1}.$$

Then $w_1, \ldots, w_k$ is an orthogonal basis for V.

PROOF. Let $1 \leq i \leq k$. We will show, by induction on i, that $w_1, \ldots, w_i$ is an orthogonal basis for $\langle v_1, \ldots, v_i$. Taking $i = k$ gives the result.
Base Case: Since v_1 is part of a linearly independent set, it is nonzero, and thus $w_1 = v_1$ is nonzero hence is an orthogonal basis for $\langle v_1 \rangle$.
Induction Step: Let $1 \leq i \leq n-1$. Suppose $w_1, \ldots, w_i$ is an orthogonal basis for $\langle v_1, \ldots, v_i \rangle$. We have

$$w_{i+1} = v_{i+1} - \sum_{j=1}^{i} \text{proj}_{w_i} v_{i+1}.$$

If $w_{i+1} = 0$ then $v_{i+1} \in \langle w_1, \ldots, w_i \rangle = \langle v_1, \ldots, v_i \rangle$, contradicting linear independence. So $w_{i+1} \neq 0$. Since $w_1, \ldots, w_i$ are orthogonal, it is enough to show that for all $1 \leq j \leq i$ we have

$$w_j \cdot w_{i+1} = 0.$$

Since $\text{proj}_{w_l} v_{i+1}$ is a scalar multiple of w_l, if $1 \leq j, l \leq i$ and $j \neq l$ then $w_j \cdot \text{proj}_{w_l} v_{i+1} = 0$. Thus

$$w_j \cdot w_{i+1} = w_j \cdot v_{i+1} - w_j \cdot (\text{proj}_{w_j} v_{i+1})$$

$$= w_j \cdot v_{i+1} - w_j \cdot (\frac{w_j \cdot v_{i+1}}{w_j \cdot w_j} w_j) = w_j \cdot v_{i+1} - w_j \cdot v_{i+1} = 0.$$ □

The approach taken to the proof of Theorem 7.2 is clean but not so geometrically enlightening. We will revisit the construction after the notion of orthogonal projection onto a subspace has been introduced.

As we have seen, a basis of $\mathbb{R}^n$ gives an alternate coordinate system. The following result shows that an orthogonal basis gives an especially nice coordinate system and that an orthonormal basis gives an extraspecially nice coordinate system.

THEOREM 7.3. *Let $v_1, \ldots, v_n$ be an orthogonal basis of $\mathbb{R}^n$.*
a) For all $v \in \mathbb{R}^n$ we have

$$v = \sum_{i=1}^{n} \text{proj}_{v_i} v.$$

b) If $v_1, \ldots, v_n$ is an orthonormal basis of $\mathbb{R}^n$, for all $v \in \mathbb{R}^n$ we have

$$v = \sum_{i=1}^n (v \cdot v_i) v_i.$$

PROOF. Since $v_1, \ldots, v_n$ is a basis for $\mathbb{R}^n$ there are unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$v = \sum_{i=1}^n \alpha_i v_i.$$

Taking the dot product with v_i gives

$$v \cdot v_i = \alpha_i v_i \cdot v_i.$$

Thus

$$\alpha_i = \frac{v \cdot v_i}{v_i \cdot v_i}$$

and

$$\alpha_i v_i = \text{proj}_{v_i} v.$$

b) For all $1 \leq i \leq n$, since v_i is a unit vector we have $\text{proj}_{v_i} v = (v \cdot v_i) v_i$. □

7.3. Orthogonal Complements. For a subset $S \subset \mathbb{R}^n$ we put

$$S^\perp := \{v \in \mathbb{R}^n \mid v \cdot s = 0 \ \forall s \in S\}.$$

That is, $S^\perp$ is the set of vectors in $\mathbb{R}^n$ that are perpendicular to every vector in S.

PROPOSITION 7.4. *Let S and T be subsets of $\mathbb{R}^n$.*
a) We have that $S^\perp$ is a subspace of $\mathbb{R}^n$.
b) If $S \subset T$ then $T^\perp \subset S^\perp$.
c) If W is the subspace spanned by S, we have $S^\perp = W^\perp$.
d) We have $S \subset (S^\perp)^\perp$.

PROOF. a) Let $v, w \in S^\perp$ and let $\alpha \in \mathbb{R}$. Then for all $s \in S$ we have

$$(\alpha v + w) \cdot s = \alpha(v \cdot s) + w \cdot s = \alpha \cdot 0 + 0 = 0.$$

b) Suppose $S \subset T$, and let $v \in T^\perp$. Then for all $s \in S$ we have $s \in T$ so $v \cdot s = 0$, so $v \in S^\perp$.
c) Since $S \subset W$ it follows from part b) such that $W^\perp \subset S^\perp$. Conversely, let $v \in S^\perp$ and let $w \in W$. Then there are $s_1, \ldots, s_n \in S$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$w = \sum_{i=1}^n \alpha_i s_i,$$

and thus we have

$$v \cdot w = \sum_{i=1}^n \alpha_i (v \cdot s_i) = \sum_{i=1}^n \alpha_i \cdot 0 = 0,$$

so $v \in W^\perp$.
d) If $s \in S$ and $v \in S^\perp$ then $s \cdot v = 0$, so $s \in (S^\perp)^\perp$. Otherwise put: indeed every element of S is perpendicular to every vector that is perpendicular to every element of S! □

From now on we will abbreviate $(S^\perp)^\perp$ to $S^{\perp\perp}$.

EXERCISE 7.2. *Show $\{0\}^\perp = \mathbb{R}^n$ and $(\mathbb{R}^n)^\perp = \{0\}$.*

We say that subspaces V, W of $\mathbb{R}^n$ are **orthogonal** if for all $v \in V$ and $w \in W$ we have $v \cdot w = 0$. It follows that $V \cap W = \{0\}$: indeed, if $x \in V \cap W$ then $x \cdot x = 0$, so $x = 0$. Thus V and W are independent subspaces in the sense of Theorem 4.26, so we have $V + W = V \oplus W$. In this situation we will write $V \perp W$ in place of $V \oplus W$ and refer to it as an "orthogonal direct sum decomposition."

More generally we say that subspaces $V_1, \ldots, V_k$ of $\mathbb{R}^n$ are **orthogonal** if for all $1 \leq i, j \leq k$, if $v_i \in V_i$ and $v_j \in V_j$ then $v_i \cdot v_j = 0$. It then follows that if we choose any nonzero vectors $v_1 \in V_1, \ldots, v_k \in V_k$

then $\{v_1, \ldots, v_k\}$ is orthogonal. It follows that the subspaces $V_1, \ldots, V_k$ are **independent** in the sense of Theorem 4.27. In this situation we write

$$\langle V_1, \ldots, V_k \rangle = V_1 \perp V_2 \perp \ldots \perp V_k.$$

THEOREM 7.5. *Let W be a subspace of $\mathbb{R}^n$.*
a) We have $\mathbb{R}^n = W \perp W^\perp$.
b) We have $\dim W^\perp = n - \dim W$.
c) We have $W^{\perp\perp} = W$.

PROOF. a) Let $m = \dim W$. The cases $m = 0$ and $m = n$ are handled by Exercise 7.2, we will assume that $1 \leq m \leq n-1$. Let $v_1, \ldots, v_m$ be a basis for W, and extend it to a basis $v_1, \ldots, v_n$ for $\mathbb{R}^n$. Applying the Gram-Schmidt process we get an orthonormal basis $b_1, \ldots, b_n$ for $\mathbb{R}^n$ such that $b_1, \ldots, b_m$ is an orthonormal basis for W. Let U be the subspace spanned by $b_{m+1}, \ldots, b_n$, so $\mathbb{R}^n = W \perp U$. It is clear that $U \subset W^\perp$. Conversely, let $v \in W^\perp$. Write $v = \sum_{i=1}^n \alpha_i b_i$. Then for all $1 \leq i \leq m$ we have

$$0 = v \cdot b_i = \alpha_i,$$

so $v = \sum_{i=m+1}^n \alpha_i b_i \in U$.
b) This is immediate from part a): indeed, we saw that if $\dim W = m$ then $\dim W^\perp = \dim U = n - m$.
c) By Proposition 7.4 we have $W \subset W^{\perp\perp}$. Also we have

$$\dim W^{\perp\perp} = n - \dim W^\perp = n - (n - \dim W) = \dim W,$$

and this implies that $W = W^{\perp\perp}$. □

7.4. Orthogonal Projection. Let W be a subspace of $\mathbb{R}^n$. By Theorem 7.5 we have

$$\mathbb{R}^n = W \perp W^\perp.$$

Therefore we can define a linear transformation

$$\mathrm{proj}_W = \pi_{W,W^\perp} : \mathbb{R}^n \to \mathbb{R}^n,$$

orthogonal projection onto W.

Let $\ell \subset \mathbb{R}^n$ be a one-dimensional subspace, and let w be a nonzero vector in ℓ. The notation $\mathrm{proj}_\ell v$ has now been defined twice: in §6.1 is given the definition

$$\mathrm{proj}_\ell v = \frac{v \cdot w}{w \cdot w} w,$$

and just now we defined it as $\mathrm{proj}_{\ell,\ell^\perp} v$. Fortunately these are the same transformation, as was shown in §6.1: we have that $v \mapsto \frac{v \cdot w}{w \cdot w} w$ is an idempotent linear map with kernel $\ell^\perp$.

PROPOSITION 7.6. *a) Let $V_1, \ldots, V_k$ be orthogonal subspaces of $\mathbb{R}^n$. Then we have*

$$\mathrm{proj}_{V_1 \perp \ldots \perp V_k} = \sum_{i=1}^k \mathrm{proj}_{V_i}.$$

b) Let V be a subspace of $\mathbb{R}^n$, and let $w_1, \ldots, w_k$ be an orthogonal basis of V. Then we have

$$\mathrm{proj}_V = \sum_{i=1}^k \mathrm{proj}_{\langle w_i \rangle}.$$

PROOF. a) If we put $W \coloneqq (V_1 \perp \ldots \perp V_k)^\perp$, then we have

$$\mathbb{R}^n = V_1 \perp \ldots \perp V_k \perp W.$$

By Gram-Schmidt, we may choose orthogonal bases for $V_1, \ldots, V_k, W$ respectively, and combining these bases (in some order) gives an orthogonal basis of $\mathbb{R}^n$. Writing down the matrices for both $\mathrm{proj}_{V_1 \perp \ldots \perp V_k}$ and $\sum_{i=1}^k \mathrm{proj}_{V_i}$ in this basis, we get that both sides are diagonal matrices with diagonal entries one in the places corresponding to the basis elements of V_i for some i and 0 in the places corresponding to

the basis elements of W.
b) Since $V = \langle w_1 \rangle \perp \ldots \perp \langle w_k \rangle$, this is a special case of part a). □

We can now revisit the Gram-Schmidt process. Let $v_1, \ldots, v_k$ be linearly independent in $\mathbb{R}^n$. For $1 \leq i \leq k$, let $V_i \coloneqq \langle v_1, \ldots, v_k \rangle$. Then we have

$$w_1 = v_1,$$

$$w_2 = v_2 - \text{proj}_{\langle w_1 \rangle} v_2 = v_2 - \text{proj}_{V_1} v_2 = \text{proj}_{V_1^\perp} v_2,$$

$$w_3 = v_3 - \text{proj}_{\langle w_1 \rangle} v_3 - \text{proj}_{\langle w_2 \rangle} v_3 = v_3 - \text{proj}_{\langle w_1, w_2 \rangle} v_3 = v_3 - \text{proj}_{V_2} v_3 = \text{proj}_{V_2^\perp} v_3,$$

and in general, for $1 \leq i \leq k-1$ we have

$$w_{i+1} = v_{i+1} - \text{proj}_{V_i} v_{i+1} = \text{proj}_{V_i^\perp} v_{i+1}.$$

7.5. Orthogonal Matrices.

LEMMA 7.7. *(Polarization Identity) For all $v, w \in \mathbb{R}^n$ we have*

$$v \cdot w = \frac{||v+w||^2 - ||v||^2 - ||w||^2}{2}. \tag{13}$$

EXERCISE 7.3. *Prove Lemma 7.7.*
(Hint: no need to think: just calculate the right hand side.)

Notation: For $1 \leq i, j \leq n$, we set $\delta_{i,j}$ to be 1 if $i = j$ and 0 otherwise.

We say that $v_1, \ldots, v_n \in \mathbb{R}^n$ forms an **orthonormal basis** if for all $1 \leq i, j \leq n$, $v_i \cdot v_j = \delta_{i,j}$. This is just a compact way of saying that we have an orthogonal basis of unit vectors.

THEOREM 7.8. *For a matrix $A \in M_{n,n}$, the following are equivalent:*
(i) For all $v \in \mathbb{R}^n$, we have $||Av|| = ||v||$.
(ii) For all $v, w \in \mathbb{R}^n$, we have $Av \cdot Aw = v \cdot w$.
(iii) For every orthonormal basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$, $Av_1, \ldots, Av_n$ is an orthonormal basis.
(iv) We have that $Ae_1, \ldots, Ae_n$ is an orthonormal ordered basis for $\mathbb{R}^n$.
(v) $A^T A = AA^T = 1$.

PROOF. (i) $\implies$ (ii): The follows from the polarization identity:

$$Av \cdot Aw = \frac{||Av+Aw||^2 - ||Av||^2 - ||Aw||^2}{2} = \frac{||v+w||^2 - ||v||^2 - ||w||^2}{2} = v \cdot w.$$

(ii) $\implies$ (iii): This is immediate: for all $1 \leq i, j \leq n$, we have

$$Av_i \cdot Av_j = v_i \cdot v_j = \delta_{i,j}.$$

(iii) $\implies$ (iv): Since $e_1, \ldots, e_n$ is an orthonormal ordered basis of $\mathbb{R}^n$, this is a special case of (iii).
(iv) $\implies$ (v): First recall that for any $A, B \in M_n$, if $AB = 1$ then also $BA = 1$. So it is enough to assume (iv) and show that $A^T A = 1$. The (i,j) entry of AA^T is the dot product of the ith row of A^T with the jth column of A, which is the dot product of the ith and jth columns of A, which is $Ae_i \cdot Ae_j$. Since the (i,j) entry of the identity matrix is δ_{ij}, this shows $A^T A = 1$.
(v) $\implies$ (i): It's equivalent to show that for all $v \in \mathbb{R}^n$, $Av \cdot Av = v \cdot v$. For this, we have

$$Av \cdot Av = (Av)^T Av = v^T A^T Av = v^T v = v \cdot v.$$ □

A matrix which satisfies the equivalent properties of Theorem 7.8 is called an **orthogonal matrix**. We denote by O_n the set of all $n \times n$ orthogonal matrices.

EXERCISE 7.4. *Show: $A \in M_n$ is orthogonal $\iff$ A is invertible and $A^T = A^{-1}$.*

EXERCISE 7.5. *Show that any permutation matrix is orthogonal.*

LEMMA 7.9. *If $A \in M_{n,n}$ is orthogonal, then* $\det A \in \{\pm 1\}$.

PROOF. Using Theorem 6.2 we get

$$1 = \det(I_n) = \det(AA^T) = \det A \det A^T = (\det A)^2,$$

so $\det A \in \{\pm 1\}$. □

We write SO_n for the set of all orthogonal matrices with determinant 1.

EXERCISE 7.6. *Let $\sigma \in S_n$ be a permutation, and let $P_\sigma \in M_{n,n}$ be the corresponding permutation matrix. Show:*

$$\det P_\sigma = \operatorname{sgn} \sigma.$$

EXAMPLE 7.10. *For $\theta \in \mathbb{R}$, let*

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

be the corresponding rotation matrix. Then $(\cos\theta, \sin\theta)$, $(-\sin\theta, \cos\theta)$ is an orthonormal basis for $\mathbb{R}^2$, so R_θ is orthogonal. It has determinant $\cos^2\theta + \sin^2\theta = 1$, hence $R_\theta \in \mathrm{SO}_2$.

Conversely, let $A \in \mathrm{SO}_2$. Let v_1 and v_2 be the two columns of A. Since v_1 is a unit vector, it lies on the unit circle and thus we can write $v_1 = (\cos\theta, \sin\theta)$ for a unique $\theta \in [0, 2\pi)$. Since $v_1 \cdot v_2 = 0$, we have that $v_2 \in \langle v_1 \rangle^\perp$, which is a line spanned by $(-\sin\theta, \cos\theta)$. The two unit vectors on that line are $\pm(-\sin\theta, \cos\theta)$, so we have $v = \pm(-\sin\theta, \cos\theta)$. The choice that leads to a matrix of determinant 1 is $v_2 = (-\sin\theta, \cos\theta)$.

Thus SO_2 consists precisely of rotation matrices.

LEMMA 7.11. *a) If $A, B \in O_n$, then we have $A^{-1} \in O_n$ and $AB \in O_n$.*
b) If $A, B \in \mathrm{SO}_n$, then we have $A^{-1} \in \mathrm{SO}_n$ and $AB \in \mathrm{SO}_n$.

EXERCISE 7.7. *Prove Lemma 7.11.*

8. Invariant Subspaces

For a matrix $A \in M_{n,n}$ and a subspace V of $\mathbb{R}^n$, we say that V is invariant under A (or A-invariant) if $A(V) \subset V$: that is, for all $v \in V$ we have $av \in V$.

The subspaces $\{0\}$ and $\mathbb{R}^n$ are invariant under A (no matter what A is). We call these invariant subspaces **trivial** and every other subspace **nontrivial**.

Let $\theta \in (0, \pi)$ and let $A = R_\theta$ be the rotation matrix. Every other subspace of $\mathbb{R}^2$ is a line through the origin, and rotation through θ does not preserve any lines (since θ is not a multiple of π), so A has no nontrivial invariant subspaces.

EXERCISE 8.1. *Let $A \in M_{n,n}$. Let $V_1, \ldots, V_k \subset \mathbb{R}^n$ be A-invariant subspaces.*
a) Show: $V_1 + \ldots + V_k$ is A-invariant.
b) Show: $\bigcap_{i=1}^k V_i$ is A-invariant.

PROPOSITION 8.1. *Let $A \in M_{n,n}$ be a nonsingular matrix, and let $V \subset \mathbb{R}^n$ be invariant under A. Then V is also invariant under A^{-1}.*

PROOF. Since A is injective, we have $\dim A(V) = \dim V$, and since $A(V) \subset V$ is an inclusion of subspaces of the same dimension, we must have equality: $A(V) = V$. Thus every $v \in V$ can be written (uniquely) in the form $v = Av'$ for some $v' \in V$. Then

$$A^{-1}v = A^{-1}(Av') = v' \in V.$$

□

THEOREM 8.2. *Let $A \in M_{n,n}$, and let V be an A-invariant subspace of $\mathbb{R}^n$.*
a) The subspace $V^\perp$ is A^T-invariant.
b) If A is either symmetric or orthogonal, then $V^\perp$ is A-invariant.

PROOF. a) We must check that if $w \in V^\perp$ then $A^T w \cdot v = 0$ for all $v \in V$. We have

$$A^T w \cdot v = w \cdot (Av) = 0$$

since $Av \in V$ and $w \in V^\perp$.
b) If A is symmetric then $A^T = A$, so this is immediate. If A is orthogonal, then by Example 8.1 V is invariant under $A^{-1} = A^T$, so $V^\perp$ is invariant under $(A^T)^T = A$. □

8.1. The Fitting Decomposition.

THEOREM 8.3 (Fitting Decomposition). *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. There are unique A-invariant subspaces V_0, V_1 of $\mathbb{R}^n$ such that*
(i) $A|_{V_0}$ is nilpotent,
(ii) $A|_{V_1}$ is invertible, and
(iii) $\mathbb{R}^n = V_0 \oplus V_1$.

PROOF. Observe that if W_1 and W_2 are invariant subspaces on which A is nilpotent, then A remains nilpotent on $\langle W_1, W_2 \rangle$: indeed, if $w_1 \in W_1$ and $w_2 \in W_2$ then there are positive integers n_1, n_2 such that $A^{n_1} w_1 = 0$ and $A^{n_2} w_2 = 0$ and then $A^{n_1+n_2}(w_1 + w_2) = 0$. So there is a unique largest invariant subspace V_0 on which A is nilpotent: can we describe it in a more useful way?

Well, being nilpotent means that some power of A is 0. In fact, a linear transformation is nilpotent iff any power of it is nilpotent. So the subspace V_0 found above for A is the same as for A^n for any $n \in \mathbb{Z}^+$. On the other hand, if $v \in V_0$ then v lies in the kernel of A^n for all sufficiently large n. A little thought then shows that we have an increasing sequence of invariant subspaces

$$0 \subset \operatorname{Ker} A \subset \operatorname{Ker} A^2 \subset \ldots \subset \operatorname{Ker} A^n \subset \ldots \subset \mathbb{R}^n.$$

Because $\mathbb{R}^n$ is finite-dimensional, this sequence must stabilize at some point. More precisely, if we have $\operatorname{Ker} A^n = \operatorname{Ker} A^{n+1}$ then we must have $\operatorname{Ker} A^n = \operatorname{Ker} A^{n+k}$ for all $k \geq 1$: inductively, it is enough to show this for $k = 2$ and if $A^{n+2} v = 0$, then $A^{n+1}(Av) = 0$, so $Av \in \operatorname{Ker} A^{n+1} = \operatorname{Ker} A^n$, so $0 = A^n(Av) = A^{n+1} v$. So let N_0 be the least positive integer such that $\operatorname{Ker} A^{N_0} = \operatorname{Ker} A^{N_0+1}$ and put

$$V_0 \coloneqq \operatorname{Ker} A^{N_0}.$$

What about V_1? Well, for any $n \in \mathbb{Z}^+$ we have that A is invertible iff A^n is invertible. This time we get a descreasing sequence of invariant subspaces

$$\mathbb{R}^n \supset A(\mathbb{R}^n) \supset A^2(\mathbb{R}^n) \supset \ldots \supset A^n(R^n) \supset \ldots \supset \{0\},$$

and we have a similar discussion to the above: the containments must be strict until the first instance of $A^{n+1}(\mathbb{R}^n) = A^n(\mathbb{R}^n)$ at which point we must have $A^{n+2}(\mathbb{R}^n) = A(A^{n+1}(\mathbb{R}^n)) = A(A^n(\mathbb{R}^n)) = A^{n+1}(\mathbb{R}^n)$ and so forth. Let N_1 be the least such N such that $A^{N_1}(\mathbb{R}^n) = A^{N_1+1}(\mathbb{R}^n)$ and put

$$V_1 \coloneqq A^{N_1}(\mathbb{R}^n).$$

We have $A(V_1) = A^{N_1+1}(\mathbb{R}^n) = A^{N_1}(\mathbb{R}^n) = V_1$, so $A|_{V_1}$ is invertible. If W is any invariant subspace on which A is invertible, then

$$W = A^{N_1}(W) \subset A^{N_1}(V) = V_1,$$

so indeed V_1 is the unique largest invariant subspace on which A is invertible.

The subspaces V_0 and V_1 must be independent, since on their intersection A is both nilpotent and invertible. Thus $\dim\langle V_0, V_1 \rangle = \dim V_0 + \dim V_1$. Finally, taking $N_2 = \max(N_0, N_1)$ the Dimension Theorem gives

$$n = \dim \mathbb{R}^n = \dim \operatorname{Ker} A^{N_2} + \dim A^{N_2}(V) = \dim V_0 + \dim V_1,$$

so we must have $\mathbb{R}^n = V_0 \oplus V_1$. □

EXERCISE 8.2. *In the notation of the proof of Theorem 8.3, show: $N_0 = N_1$.*

EXERCISE 8.3. *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Let $V_1, V_2 \subset \mathbb{R}^n$ be subspaces such that $A|_{V_1}$ is injective and $A|_{V_2}$ is injective.*

a) *Suppose that V_1 and V_2 are A-invariant. Show: $V_1 + V_2$ is an A-invariant subspace on which A is invertible.*
b) *Show that in general $A|_{V_1+V_2}$ need not be injective.*

EXERCISE 8.4. *Let $A : \mathbb{R}^n \to \mathbb{R}^n$ and suppose that $\operatorname{Ker} A \cap A(\mathbb{R}^n) = (0)$. Show that, in the notation of Theorem 8.3, we have $\operatorname{Ker} A = V_0$ and $A(\mathbb{R}^n) = V_1$.*

8.2. Existence of Invariant Subspaces.

THEOREM 8.4. *Let $n \in \mathbb{Z}^+$, and let $A \in M_{n,n}$ be a matrix.*
a) If n is odd, then A has a one-dimensional invariant subspace.
b) If n is even, then A has an invariant subspace of dimension two.

PROOF. a) Every one-dimensional A-invariant subspace is spanned by an eigenvector of A, so A has a one-dimensional invariant subspace iff it has an eigenvalue. If n is odd, then $\chi_A(t)$ is a polynomial with real coefficients of odd degree, hence has a real root by the Intermediate Value Theorem.
b) Let $P(t)$ be the minimal polynomial of A.
Case 1: Suppose that P has degree 1. Then $P(t) = t - \lambda$ for some $\lambda \in \mathbb{R}$ and A is the scalar matrix λI_n for which every subspace is invariant. Since n is even, it is at least two, and so indeed there is a 2-dimensional A-invariant subspace.
Case 2: Suppose P admits an irreducible quadratic factor $Q(t) = t^2 + bt + c$, and write $P = QR$. Since $\deg R < \deg P$ we must have $R(A) \neq 0$; let $v \in \mathbb{R}^n$ be such that $R(A)v \neq 0$. Put $w := R(A)v$ and $V := \langle w, Aw \rangle$. We claim that V is a two-dimensional A-invariant subspace. First, it is A-invariant since $A^2w = -bAw - cw$. Second, if V were one-dimensional then w would be an eigenvector for A, say with eigenvalue λ. But then

$$0 = (A^2 + bA + cI_n)w = (\lambda^2 + b\lambda + c)w,$$

and since $w \neq 0$ we have $0 = Q(\lambda)$, contradicting the fact that Q is irreducible quadratic.
Case 3: Suppose P has linear factors $(t - \lambda_1)$ and $(t - \lambda_2)$ for $\lambda_1 \neq \lambda_2 \in \mathbb{R}$. Then the λ_1 and λ_2 eigenspaces are each nontrivial. If v_1 is an eigenvector with eigenvalue λ_1 and v_2 is an eigenvector with eigenvalue λ_2 then $V := \langle v_1, v_2 \rangle$ is a two-dimensional A-invariant subspace.
Case 4: If none of the above hold, then we must have $P(t) = (t - \lambda)^k$ for some $\lambda \in \mathbb{R}$ and $k \geq 1$. There must be $v \in \mathbb{R}^n$ such that $(A - \lambda)^{k-1}v \neq 0$, for otherwise A satisfies the polynomial $(t - \lambda)^{k-1}$, which has degree less than the minimal polynomial. Let $w := (A - \lambda)^{k-2}(v)$. Then $(A - \lambda)w \neq 0$ but $(A - \lambda)^2 w = 0$, so the subspace $V := \langle w, Aw \rangle$ is two-dimensional and A-invariant. Indeed, by construction w is not an eigenvector with eigenvalue λ, and by Proposition 9.27 since λ is the only root of P it is the only eigenvalue for A. □

9. Eigenvectors and Diagonalization

9.1. Diagonalization.

9.2. Eigenvectors, Eigenvalues and Eigenspaces.

A vector $v \in \mathbb{R}^n$ is an **eigenvector for a linear transformation** $L : \mathbb{R}^n \to \mathbb{R}^n$ if

(EV1) $v \neq 0$, and
(EV2) There is $\lambda \in \mathbb{R}$ such that $L(v) = \lambda v$.

Thus an eigenvector is a nonzero v such that $L(v)$ is a scalar multiple of v.

If v is an eigenvector, then the scalar λ such that $L(v) = \lambda v$ is unique: if

$$\lambda_1 v = L(v) = \lambda_2 v \implies (\lambda_1 - \lambda_2)(v) = 0.$$

Since $v \neq 0$, this forces $\lambda_1 = \lambda_2$. This scalar is called the **eigenvalue** of v. Moreover, a scalar $\lambda \in \mathbb{R}$ is called an **eigenvalue for L** if there is some eigenvector v with eigenvalue λ.

EXAMPLE 9.1. *Let D be a diagonal matrix with diagonal entries $d_1, \ldots, d_n$. Then each of the standard basis vectors $e_1, \ldots, e_n$, and the eigenvalues are (respectively) $d_1, \ldots, d_n$. In particular there is a basis – the standard basis! – of $\mathbb{R}^n$ consisting of eigenvectors of D.*

REMARK 9.2. *Although the zero* vector *is not allowed to be an eigenvector, the zero* scalar *is allowed to be an eigenvalue, and this is an important case: 0 is an eigenvalue for L if and only if there is $0 \neq v$ such that $L(v) = 0v = 0$. Thus the eigenvectors with eigenvalue 0 are precisely the nonzero vectors in the kernel (or null space) of T, and L has 0 has an eigenvalue if and only if it is singular.*

For any $\lambda \in \mathbb{R}$ we define the **λ-eigenspace**

$$V_\lambda := \{v \in \mathbb{R}^n \mid L(v) = \lambda v\}.$$

In other words, V_λ consists of the eigenvectors for v with eigenvalue λ (if any) along with the zero vector. We also define the **geometric multiplicity** of λ as $\dim V_\lambda$.[10]

The following exercise is very easy but all-important: it tells us that computing eigenspaces is a special case of our favorite linear algebraic computation.

EXERCISE 9.1. *Show: for all $\lambda \in \mathbb{R}$, V_λ is the null space of $\lambda I_n - A$.*

EXAMPLE 9.3. *We return to the case of a diagonal matrix $D \in M_{n,n}$ with diagonal entries $d_1, \ldots, d_n$. Earlier we saw that the standard basis vectors $e_1, \ldots, e_n$ are eigenvectors, with corresponding eigenvalues $d_1, \ldots, d_n$. Now we want to go further by computing all the eigenspaces. First, suppose $v = (x_1, \ldots, x_n)$ is an eigenvector for D. Then there is $\lambda \in \mathbb{R}$ such that*

$$\lambda v = (\lambda x_1, \ldots, \lambda x_n) = Dv = (d_1 x_1, \ldots, d_n x_n).$$

Thus for all $1 \leq i \leq n$, we have $\lambda x_i = d_i x_i$, so if $x_i \neq 0$ then $\lambda = d_i$. By definition of eigenvectors, $v \neq 0$ hence at least one x_i is nonzero, so $\lambda = d_i$. This shows that the only eigenvalues of D are the diagonal entries. Moreover, if for $1 \leq i \neq j \leq n$ we have both $x_i \neq 0$ and $x_j \neq 0$, then $d_i = \lambda = d_j$. In other words, if several components of v are nonzero, then the corresponding diagonal entries must all be equal; conversely when this happens we do indeed have $Dv = \lambda v$. This shows:

$$V_{d_i} = \operatorname{span}_{1 \leq j \leq n}\{e_j \mid d_j = d_i\}.$$

The dimension of V_{d_i} is the number of $1 \leq j \leq n$ such that $d_j = d_i$.
*(**Subexample**: For instance, if*

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix},$$

then $V_1 = \operatorname{span} e_1, e_3$, $V_2 = \operatorname{span} e_2$, $v_3 = \operatorname{span} e_4$.)
Further, the eigenspaces are independent and the sum of their dimensions is n.

EXERCISE 9.2. *As the previous example indicates, cleaner bookkeeping arises for diagonal matrices if we assume that repeated diagonal entries occur in blocks of consecutive terms (unlike the subexample above, in which the two 1's occur nonconsecutively). Show that any diagonal matrix is similar to a diagonal matrix with this property.*

PROPOSITION 9.4. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation.*
a) For any $\lambda \in \mathbb{R}$, V_λ is a subspace of $\mathbb{R}^n$.
b) $V_\lambda \supsetneq \{0\} \iff \lambda$ is an eigenvalue of T.

[10] Yes, this is a lot to swallow. Our pedagogical strategy here is to put all the basic definitions in one place for easy reference, and then explore the consequences of these definitions in a more leisurely manner.

PROOF. It is tempting to leave this as an exercise – it is quite straightforward – but because of its importance to our narrative we prefer to give a complete proof.
a) As usual, this is easy: $0 \in V_\lambda$. Further, if $v, w \in V_\lambda$ and $\alpha \in \mathbb{R}$, then

$$L(\alpha v + w) = \alpha L(v) + L(w) = \alpha\lambda v + \lambda w = \lambda(\alpha v + w),$$

so $\alpha v + w \in V_\lambda$. By the One-Step Linear Transformation Test, V_λ is a subspace.
b) According to our definitions, λ is an eigenvalue for T if and only if there is a nonzero vector $v \in \mathbb{R}^n$ with $T(v) = \lambda v$; this occurs if and only if $V_\lambda \neq \{0\}$. □

PROPOSITION 9.5. *Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and let $\lambda_1, \ldots, \lambda_n$ be distinct real numbers. Then the eigenspaces $V_{\lambda_1}, \ldots, V_{\lambda_n}$ are independent:*

$$V_{\lambda_1} + \ldots + V_{\lambda_n} = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_n}.$$

PROOF. According to Theorem **??** it suffices to show: for any nonzero vectors $v_i \in V_{\lambda_i}$, the set $\{v_1, \ldots, v_n\}$ is linearly independent. Suppose not. Then after reordering the vectors there is some $2 \leq k \leq n$ such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k = 0 \tag{14}$$

with *every* $\alpha_k \neq 0$ (we cannot have $k = 1$ because each v_i is nonzero) and among all such relations we may choose one with k as small as possible. If we can conjure up a similar linear dependence relation among $k-1$ vectors, we get a contradiction and we'll be done. Well, the big idea is to apply L to (14), getting

$$\alpha_1\lambda_1 v_1 + \alpha_2\lambda_2 v_2 + \ldots + \alpha_k\lambda_k v_k = 0. \tag{15}$$

Multiplying (14) by λ_1 and subtracting what we get from (15), we obtain

$$\alpha_2(\lambda_2 - \lambda_1)v_2 + \alpha_3(\lambda_3 - \lambda_1)v_3 + \ldots + \alpha_k(\lambda_n - \lambda_1)v_k = 0. \tag{16}$$

Since the λ_i's are distinct, for all $2 \leq i \leq k$, $\lambda_i - \alpha_1 \neq 0$, and thus (16) is a linear dependence relation with all nonzero coefficients but with $k-1$ terms instead of k terms: contradiction. □

EXERCISE 9.3. *Give a much shorter proof of Proposition 9.5 when $n = 2$.*

COROLLARY 9.6. *A linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ has at most n eigenvalues.*

PROOF. If we had $n+1$ eigenvalues $\lambda_1, \ldots, \lambda_{n+1}$, then $V_{\lambda_1} + \ldots + V_{\lambda_{n+1}}$ would be a subspace of $\mathbb{R}^n$ of dimension $\sum_{i=1}^{n+1} \dim V_{\lambda_i} \geq \sum_{i=1}^{n+1} 1 = n+1$. □

Why do we care about all this eigenstuff anyway?? Because of the following result.

THEOREM 9.7. *Let $A \in M_{n,n}$, and let $P \in M_{n,n}$ be invertible.*
a) The following are equivalent:
(i) $P^{-1}AP$ is diagonal.
(ii) The columns of P are eigenvectors for L_A.
b) The following are equivalent:
(i) A is diagonalizable.
(ii) There is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors for A.
(iii) There is an eigenvalue for A, and if $\lambda_1, \ldots, \lambda_k$ are the eigenvalues, then

$$\mathbb{R}^n = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}.$$

PROOF. a) (i) $\implies$ (ii): Suppose that $P^{-1}AP = D$, where D is diagonal with diagonal entries $\lambda_1, \ldots, \lambda_n$. Let v_i be the ith column of P. Then for all $1 \leq i \leq n$,

$$\lambda_i e_i = De_i = P^1 APe_i = P^{-1}Av_i.$$

Multiplying on the left by P gives

$$\lambda_i v_i = \lambda_i P(e_i) = P(\lambda_i e_i) = PP^{-1}Av_i = Av_i,$$

so $v_i \in V_{\lambda_i}$. Since P is invertible, each v_i is nonzero, so v_i is an eigenvector for L_A.
(ii) $\implies$ (i): Suppose that for all $1 \leq i \leq n$, there is $\lambda_i \in \mathbb{R}$ such that $Av_i = \lambda_i v_i$, and let $\mathcal{B} = (v_1, \ldots, v_n)$. Then the change of basis formula gives

$$A_{L,\mathcal{B}} = P^{-1}AP.$$

Moreover, since $Av_i = \lambda_i v_i$, the ith column of $A_{L,\mathcal{B}}$ is $\lambda_i e_i$: thus $A_{L,\mathcal{B}}$ is a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$.
b) The equivalence of (i) and (ii) follows immediately from part a). Moreover, if we have a basis of $\mathbb{R}^n$ consisting of eigenvectors for L, then breaking it apart into subsets $\mathcal{B}_i$ consisting of eigenvectors for the eigenvalue λ_i gives a direct sum decomposition $\mathbb{R}^n = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_k}$. And conversely: given such a direct sum decomposition, we take $\mathcal{B}_i$ to a basis for V_{λ_i}, and then $\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_k$ is a basis for $\mathbb{R}^n$ consisting of eigenvectors. □

PROPOSITION 9.8. *Let $A \in M_{n,n}$ be nonsingular. Then 0 is not an eigenvalue of A, and for all nonzero $\lambda \in \mathbb{R}$, the λ-eigenspace of A is the λ^{-1}-eigenspace of A^{-1}.*

PROOF. $Av = \lambda v$. If $\lambda = 0$, then v lies in the null space of A so A is singular, contradiction. Applying A^{-1} to the equation $Av = \lambda v$ we get

$$v = A^{-1}\lambda v = \lambda A^{-1}v,$$

or

$$A^{-1}v = \lambda^{-1}v,$$

which shows that v lies in the λ^{-1} eigenspace of A^{-1}. Applying the same claim with (A^{-1}, λ^{-1}) in place of (A, λ) shows that every vector lying in the λ^{-1} eigenspace of A^{-1} also lies in the λ eigenspace of A. □

EXERCISE 9.4. *a) Show: $A^2 = A$, then A is diagonalizable.*
b) Show: if $A^2 = 1$, then A is diagonalizable.[11]

Recall that the determinant of a matrix is a **similarity invariant**: in concrete terms, this means that if $A, B \in M_{n,n}$ are similar – i.e., $B = PAP^{-1}$ for some invertible P – then $\det A = \det B$. But there is also a richer perspective: as we have seen, similarity of matrices is an equivalence relation on $M_{n,n}$, and thus it partitions $M_{n,n}$ into equivalence classes. However, if I give you two matrices $A, B \in M_{n,n}$, it is usually not so easy to tell whether they lie in the same equivalence class: in principle we would have to try conjugating A by *every* invertible matrix P to see whether we get B, but there are infinitely many such matrices so it is not clear that this can always be done in a practical manner. This is common in higher mathematics: for an interesting equivalence relation $\sim$ on a set X, it is not always clear how to check in practice whether two objects are equivalent. One strategy is to find an **invariant** of the equivalence relation. One can think of this as a function $f : X \to Y$ defined on the entire set such that if $x_1 \sim x_2$ then $f(x_1) = f(x_2)$. Then, if we have two objects x_1, x_2 such that $f(x_1) \neq f(x_2)$, we knw that they can't be equivalent. The determinant is such a function: if $\det A \neq \det B$, then A and B cannot be similiar. Unfortunately the converse does not hold, as we have seen. Thus we want further such invariants. If we are lucky, then eventually we will find a **complete set of invariants**, such that if all the invariants of x_1 and x_2 agree, then indeed $x_1 \sim x_2$.

In the case of similarity of matrices this can indeed be done, but unfortunately the end of this story lies beyond the end of this course. But here are some further important invariants.

THEOREM 9.9. *Let $A, P \in M_{n,n}$, and suppose P is invertible.*
a) For $v \in \mathbb{R}^n$, if $Av = \lambda v$, then

$$(P^{-1}AP)(P^{-1}v) = \lambda(P^{-1}v).$$

[11] In this exercise it is important that our scalar field is $\mathbb{R}$.

b) For all $\lambda \in \mathbb{R}$, let V_λ be the λ-eigenspace for A, and let W_λ be the λ-eigenspace for $P^{-1}AP$. Then

$$P^{-1}V_\lambda = W_\lambda.$$

c) For all $\lambda \in \mathbb{R}$, the geometric multiplicity of λ for A is equal to the geometric multiplicity of λ for $P^{-1}AP$.

PROOF. a) This is a straightforward computation that we leave to the reader.
b) Part a) says that if v is an eigenvector for A, then $P^{-1}v$ is an eigenvector for $P^{-1}AP$. In other words, we have

$$P^{-1}V_\lambda \subset W_\lambda.$$

Conversely, if $w \in W_\lambda$ then $P^{-1}APw = \lambda w$, so $A(Pw) = P(\lambda w) = \lambda(Pw)$ and thus $Pw \in V_\lambda$: thus $PW_\lambda \subset V_\lambda$; applying P^{-1} to both sides gives $W_\lambda \subset P^{-1}V_\lambda$. Thus $P^{-1}V_\lambda = W_\lambda$.
c) Let $(v_1, \ldots, v_k)$ be an ordered basis for V_λ. Then $(P^{-1}(v_1), \ldots, P^{-1}(v_k))$ is an ordered basis for $P^{-1}V_\lambda = W_\lambda$, so $\dim V_\lambda = \dim W_\lambda$. □

Thus similar matrices have the same eigenvalues and the same geometric multiplicities. Thus for all $\lambda \in \mathbb{R}$, the geometric multiplicity of λ (which will be 0 if λ is not an eigenvalue) is a similarity invariant.

EXERCISE 9.5. *Theorem 9.9 does* not *say that similar matrices have* equal *λ-eigenspaces. Give an explicit example of similar matrices with distinct λ-eigenspaces for some λ.*
(Suggestion: any matrix that is diagonalizable but not diagonal will give rise to an example.)

EXERCISE 9.6. *a) Let $A \in M_{n,n}$, and let $N \in \mathbb{Z}^+$. If v is an eigenvector for A with eigenvalue λ, show that v is an eigenvector for A^N with eigenvalue λ^N.*
b) Recall that a matrix $A \in M_{n,n}$ is ***nilpotent*** *if $A^N = 0$ for some $N \in \mathbb{Z}^+$. Show that a nilpotent matrix has precisely one eigenvalue:* 0.
c) In the setting of part a), show that it is possible that the Nth powers of the eigenvalues of A do not give all of the eigenvalues of A^N. For instance, exhibit an A which has no eigenvalues but A^2 does.

COROLLARY 9.10. *If $A \in M_{n,n}$ has n eigenvalues, then it is diagonalizable.*

PROOF. Each eigenvalue contributes at least one dimension to the sum of the eigenspaces, hence if we have n eigenvalues then the sum of the eigenspaces is all of $\mathbb{R}^n$. □

It is part of Theorem 9.7 that in order for a matrix $A \in M_{n,n}$ to be diagonalizable it must have at least one eigenvalue. This also follows because similar matrices have the same eigenvalues, and diagonal matrices have their diagonal entries as eigenvalues, hence at least one. For any $n \geq 1$, an $n \times n$ matrix may well have only a single eigenvalue and still be diagonalizable.

If $A \in M_{n,n}$ has at least one and fewer than n eigenvalues, whether it is diagonalizable or not depends upon the geometric multiplicities: for A to be diagonalizable, it is necessary and sufficient that the geometric multiplicities sum to n. Thus, the fewer eigenvalues we have, the larger each of their geometric multiplicities must be in order for the matrix to be diagonalizable. Here is the extreme case:

EXAMPLE 9.11. *Let $\lambda \in \mathbb{R}$. The scalar matrix λI_n has λ as an eigenvalue. In fact, $\mathbb{R}^n = V_\lambda$; in particular λ is the only eigenvalue. Conversely, this property is evidently characteristic of scalar matrices: if $\mathbb{R}^n = V_\lambda$ for some λ, this precisely means that $Av = \lambda v$ for all $v \in \mathbb{R}^n$. In particular this holds for $v = e_1, \ldots, e_n$, so $A = \lambda I_n$ is a scalar matrix.*

Thus a matrix $A \in M_{n,n}$ with exactly one eigenvalue λ is diagonalizable if and only if it is the scalar matrix λI_n. In particular, a nondiagonal matrix with at most one eigenvalue cannot be diagonalizable.

EXERCISE 9.7. *Let $A \in M_{n,n}$ be a singular matrix with a single eigenvalue. Show that A is diagonalizable if and only if $A = 0$.*

EXAMPLE 9.12. *Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation through an angle $\theta \in [0, 2\pi)$. If $\theta \neq 0, \pi$, then there are no eigenvectors: rotating a nonzero vector through an angle of θ changes the line that it lies on. (Rotating a vector through an angle of* 0 *fixes it, and rotating a vector through an angle of π scales it by -1.) Having no eigenvectors –equivalently, having no eigenvalues) – R is* not *diagonalizable.*

EXERCISE 9.8. *For any even positive integer n, construct a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ with no eigenvectors.*

9.3. The Characteristic Polynomial.

So far we have developed just the beginning of the theory of eigenvectorss and diagonalization: there are many (many!) results that give necessary and/or sufficient conditions for a linear transformation (or a matrix) to admit a basis of eigenvectors. However, before we press on in this direction we should first address a computational issue.

QUESTION 9.13. *Let $A \in M_{n,n}$. How do we compute the eigenvalues of A?*

Notice that computing the eigenvalues is the key to computing the eigenspaces. Indeed, recall Exercise 9.1: for any $\lambda \in \mathbb{R}$, the eigenspace V_λ is simply the null space of $\lambda I_n - A$, so we can compute it via row reduction. However, there are of course infinitely many real numbers, so we can't simply compute the null spacess of all $\lambda I_n - A$. In some cases one can successfully guess (or know) some good candidates for λ. But in general this would be difficult.

EXAMPLE 9.14. *Let $M = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. We claim that $\lambda = \frac{1\pm\sqrt{5}}{2}$ are eigenvalues. Indeed,*

$$(\frac{1+\sqrt{5}}{2})I_2 - M = \begin{bmatrix} \frac{1+\sqrt{5}}{2} - 1 & -1 \\ -1 & \frac{1+\sqrt{5}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}-1}{2} & -1 \\ -1 & \frac{\sqrt{5}+1}{2} \end{bmatrix}.$$

Multiplying the first row by $\frac{2}{\sqrt{5}-1}$ and adding it to the second row, we get

$$\begin{bmatrix} \frac{\sqrt{5}-1}{2} & -1 \\ 0 & \frac{\sqrt{5}+1}{2} - \frac{2}{\sqrt{5}-1} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}-1}{2} & -1 \\ 0 & 0 \end{bmatrix}.$$

So the null space is nontrivial and $v_1 = (\frac{2}{\sqrt{5}-1}, 1) = (\frac{\sqrt{5}+1}{2}, 1)$ is an eigenvector. Similarly,

$$(\frac{1-\sqrt{5}}{2})I_2 - M = \begin{bmatrix} \frac{-1-\sqrt{5}}{2} & -1 \\ -1 & \frac{1-\sqrt{5}}{2} \end{bmatrix}.$$

Multiplying the first row by $\frac{2}{-1-\sqrt{5}}$ and adding it to the second row, we get

$$\begin{bmatrix} \frac{-1-\sqrt{5}}{2} & -1 \\ 0 & \frac{2}{1+\sqrt{5}} + \frac{1-\sqrt{5}}{2} \end{bmatrix} = \begin{bmatrix} \frac{-1-\sqrt{5}}{2} & -1 \\ 0 & 0 \end{bmatrix}.$$

Again the null space is nontrivial, and $v_2 = (\frac{1-\sqrt{5}}{2}, 1)$ is an eigenvector.

Probably few of us would have guessed taking $\lambda = \frac{1\pm\sqrt{5}}{2}$ without some extra insight/information. We need a technique for computing eigenvalues.

Now determinants come to our rescue: for any $A \in M_{n,n}$ and $\lambda \in \mathbb{R}$, we have:

$$\lambda \text{ is an eigenvalue} \iff \text{Ker}(\lambda I_n - A) \supsetneq \{0\} \iff \lambda I_n - A \text{ is singular} \iff \det(\lambda I_n - A) = 0.$$

Thus the eigenvalues are precisely the real numbers λ such that $\det(\lambda I_n - A) = 0$. This is useful because of the following result.

PROPOSITION 9.15. *For any $A \in M_{n,n}$ and $\lambda \in \mathbb{R}$, $\det(\lambda I_n - A)$ is a monic[12] polynomial of degree n with real coefficients.*

[12]A polynomial is **monic** if its highest order term has leading coefficient 1.

PROOF. The (i,j) entry of $\lambda I_n - A$ is $\lambda\delta(i,j) - a_{ij}$: when $i = j$ this is a linear polynomial in λ; otherwise it is a real number. Because the determinant of any matrix is a certain polynomial expression involving the matrix entries, $\det \lambda I_n - A$ is certainly a polynomial in λ. More precisely each of the $n!$ terms in the determinant is, up to ± 1, obtained by multiplying a choice of one entry from each row and column of the matrix, hence each term is a product of n factors each of which is either a constant or a linear polynomial, so each term is a polynomial of degree at most n. In order to get a degree n polynomial we must have a factor of λ every time, and this happens precisely when we choose the diagonal entries (or, if you like, the identity permutation): this term contributes $(\lambda - a_{11}) \cdots (\lambda - a_{nn})$, which is a monic degree n polynomial in λ. If we add a monic polynomial of degree n to a polynomial of smaller degree, the leading term cannot change so we get another monic polynomial of degree λ. □

We define the **characteristic polynomial of A**

$$\chi_A(t) \coloneqq \det(tI_n - A).$$

(Why did we switch from λ to t? This is a fastidiousness on my part: I want to distinguish between a polynomial and the numbers that we plug into it.)

The characteristic polynomial $\chi_A(t)$ is precious to us because (i) we know how to compute it, by computing the determinant (e.g. by row reduction to upper triangular form) and (ii) as we have seen, the eigenvalues of A are precisely the **real roots** of $\chi_A(t)$, i.e., the real numbers λ such that $\chi_A(\lambda) = 0$.

EXERCISE 9.9. *Let $A \in M_{n,n}$.*
a) Suppose each a_{ij} is a rational number. Show that all the coefficients of $\chi_A(t)$ are rational numbers.
b) Supose each a_{ij} is an integer. Show that all the coefficients of $\chi_A(t)$ are integers.

The characteristic polynomial is the final tool we need for a complete computational method for determining whether $A \in M_{n,n}$ is diagonalizable and if so finding a matrix P such that $P^{-1}AP$ is diagonal.

Step 1: We compute the characteristic polynomial $\chi_A(t)$ and find all the real roots $\lambda_1, \ldots, \lambda_k$. These are the eigenvalues of A. If not all of the roots of $\chi_A(t)$ are real then A cannot be diagonalized (over $\mathbb{R}$). If all the roots are real, we proceed to Step 2.

Step 2: For each eigenvalue λ, we compute the eigenspace $V_\lambda = \operatorname{Ker}(\lambda I_n - A)$.

Step 3: A is diagonalizable iff $\sum_{i=1}^k \dim V_\lambda = n$. If so, we find a basis $\mathcal{B}_i$ for each V_{λ_i}, and let P be the matrix with columns the elements of the $\mathcal{B}_i$'s.

Although we are in theory working with the real numbers as our "scalars," in practice most of our matrices have had rational numbers as entries. Gaussian reduction applied to a matrix with rational entries will yield a rref matrix also with rational entries, and thus a basis of the null space consisting of vectors with rational entries can always be found. Similarly orthogonalization does not introduce irrational numbers (ortho*normalization* does, but only square roots). This explains why most of our calculations have involved only rational numbers. It would be great if the characteristic polynomial of a matrix with rational entries necessarily had only rational numbers as roots, because then we can find all of them easily using the following high school result.

THEOREM 9.16. *(Rational Roots Theorem) Let $P(t) = a_n t^n + \ldots + a_1 t + a_0$ be a polynomial with integer coefficients a_i. Then the only possible nonzero rational numbers r such that $P(r) = 0$ are of the form $\pm\frac{c}{d}$ where c and d are nonzero integers, $d \neq 0$, a_0 is divisible by c and a_n is divisible by d.*

As Example 9.14 shows, the characteristic polynomial can have rational (even integral) coefficients but still have irrational numbers as roots. In fact a polynomial with integral coefficents will have irrational roots most of the time: consider for instance the case of a quadratic polynomial $P(t) = t^2 + bt + c$, in which the quadratic formula gives us a square root. You have probably solved enough quadratic

equations in your time to understand that if you choose integers b and c at random, the discriminant $b^2 - 4ac$ is very unlikely to be a perfect square, so the roots are very likely to be irrational numbers.

EXERCISE 9.10. *a) Let $a_0, a_1, a_2 \in \mathbb{R}$, and let $A = \begin{bmatrix} 0 & 0 & -a_0 \\ 1 & 0 & -a_1 \\ 0 & 1 & -a_2 \end{bmatrix}$. Show:*

$$\chi_A(t) = (t - a_0)(t - a_1)(t - a_2).$$

b) Let n be a positive integer, let $a_0, \ldots, a_{n-1} \in \mathbb{R}$, and let $p(t) = t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + a_0$, i.e., an arbitrary monic degree n polynomial. Let

$$A_p = \begin{bmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & \ldots & 0 & -a_2 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & -a_{n-1} \end{bmatrix}.$$

*Show that $\chi_{A_p}(t) = p(t)$. (A_p is called the **companion matrix** of p.)*

EXERCISE 9.11. *Let A be the companion matrix of t^n.*
a) Show: A is nilpotent.
b) Show: $A^n = 0$.
c) Show: $A^{n-1} \neq 0$.

EXERCISE 9.12. *Let $A \in M_{n,n}$ be nilpotent.*
a) Let $k \geq 1$ be such that $A^k \neq 0$. Thus there is $v \in \mathbb{R}^n$ such that $A^k v \neq 0$. Show that there is $\ell \geq 0$ such that $A^{k+\ell} v \neq 0$ and $A^{k+\ell+1} v = 0$.
b) Show that $A^\ell v \in \operatorname{Ker} A^{k+1} \setminus \operatorname{Ker} A^k$.
c) Deduce: $A^n = 0$.

PROPOSITION 9.17. *Let $A, P \in M_{n,n}$ with P invertible. Then*

$$\chi_{P^{-1}AP}(t) = \chi_A(t).$$

In other words, the characteristic polynomial is a similarity invariant.

PROOF. We use the fact that scalar matrices commute with all matrices:

$$\begin{aligned} \chi_{P^{-1}AP}(t) &= \det(tI_n - P^{-1}AP) = \det(tP^{-1}I_nP - P^{-1}AP) \\ &= \det(P^{-1}(tI_n - A)P) = \det(P^{-1}) \det(tI_n - A) \det P \\ &= \chi_A(t) \det(P)^{-1} \det P = \chi_A(t). \end{aligned}$$

□

EXAMPLE 9.18. *Let $A \in M_{n,n}$ be (upper or lower) triangular. Then $tI_n - A$ is also triangular, so its determinant is the product of the diagonal entries:*

$$\chi_A(t) = (t - a_{11}) \cdots (t - a_{nn}).$$

*In particular $\chi_A(t)$ is **split**: that is, it factors into a product of linear polynomials. Not every polynomial is split: e.g. $t^2 + 1$ is not, hence neither is any polynomial of the form $(t^2 + 1)g(t)$ for a nonzero polynomial $g(t)$.*

THEOREM 9.19. *Let $A \in M_{n,n}$. If A is diagonalizable – or even **triangularizable**, i.e., similar to a triangular matrix – then $\chi_A(t)$ is split.*

PROOF. This follows immediately from the two previous results. □

Perhaps the previous result looks a bit abstruse. In fact it is given here to help us out: i.e., to tell us that in certain situations diagonalization is hopeless so we need not compute the eigenspaces.

EXAMPLE 9.20. *Let $A \in M_{3,3}$ have characteristic polynomial $t(t^2+1)$. (Recall that by Exercise 9.10 there is at least one $n \times n$ matrix with any given monic degree n polynomial as its characteristic polynomial.) Then the only eigenvalue of A is $\lambda = 0$, so the only way that A could be diagonalizable is if it is the zero matrix. But the zero matrix has characteristic polynomial t^3. So A is not diagonalizable.*

Let $A \in M_{4,4}$ have characteristic polynomial $(t-1)(t-2)(t^2+1)$. Then A has two eigenvalues, $\lambda = 1$ and $\lambda = 2$. The argument of the preceding paragraph does not apply. But it would be a waste of time to compute the eigenspaces V_1 and V_2: because $\chi_A(t)$ is not split, the matrix cannot be diagonalizable.

Having a split characteristic polynomial is not enough for a matrix to be diagonalizable: take a nonzero nilpotent matrix. However, the following weaker result is true.

THEOREM 9.21. *For $A \in M_{n,n}$, the following are equivalent:*
a) The characteristic polynomial $\chi_A(t)$ is split.
b) A is similar to an upper triangular matrix.
c) A is similar to a lower triangular matrix.

PROOF. a) $\implies$ b): We go by induction on n, the case $n = 1$ being clear. Suppose the result holds for all $B \in M_{n-1,n-1}$, and consider $A \in M_{n,n}$. By assumption, we may write $\chi_A(t) = (t-\lambda)g(t)$ where $g(t) \in \mathbb{R}[t]$ is a split polynomial of degree $n-1$. Let $v_1 \in \mathbb{R}^n$ be a nonzero vector such that $Av_1 = \lambda v_1$, and complete to a basis $v_1, \ldots, v_n$. The matrix A' of the linear transformation with respect to this basis has first column $(\lambda, 0, \ldots, 0)^T$, so has the form

$$\begin{bmatrix} \lambda & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2n} \\ \vdots & & & \\ 0 & a_{n2} & \ldots & a_{nn} \end{bmatrix}.$$

Schematically it has the form

$$\begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix}.$$

(Here we write $*$ for a block of entries in the matrix that we do not care about. When we write $*$ from one equation to the next, it does not mean that the entries are the same as before.) Then we have

$$(t-\lambda)g(t)\chi_A(t) = \chi_{A'}(t) = \det \begin{bmatrix} t-\lambda & * \\ 0 & tI_{n-1} - B \end{bmatrix} = (t-\lambda)\det(tI_{n-1} - B) = (t-\lambda)\chi_B(t).$$

Thus $\chi_B(t) = g(t)$ is split. By induction there is $Q \in GL_{n-1}$ such that QBQ^{-1} is upper triangular. Let P be the block diagonal matrix $I_1 \oplus Q$. Then

$$PA'P^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} 1 & * \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix} = \begin{bmatrix} 1 & * \\ 0 & QB \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & Q^{-1} \end{bmatrix} = \begin{bmatrix} 1 & * \\ 0 & QBQ^{-1} \end{bmatrix}$$

is upper triangular.
b) $\implies$ c): Let σ be the permutation $i \mapsto n+1-i$, and let $P(\sigma)$ be the corresponding permutation matrix. If T is upper triangular, then $P(\sigma)AP(\sigma)^{-1}$ is lower triangular: indeed, the jth column of $B = P(\sigma)AP(\sigma)^{-1}$ expresses the linear transformation on e_{n+1-j} in terms of $e_n, e_{n-1}, \ldots, e_1$: in other words, the jth column of B is obtained from the $(n+1-j)$th column of A by writing the entries in the reverse order. Working from right to left, we see that if A is upper triangular, then B is lower triangular.
c) $\implies$ a): This is Theorem 9.19. □

We now want to identify a further situation in which diagonalizability is hopeless. Let D be a diagonal matrix with diagonal entries $d_1, \ldots, d_n$. As a special case of Example 9.18, we know that $\chi_D(t) = (t-d_1)\cdots(t-d_n)$. Not only is $\chi_D(t)$ split with roots precisely the eigenvalues of D, but that the **multiplicity** of each root λ – i.e., the number of occurrences of the linear factor $t-\lambda$ – is equal to $\dim V_\lambda$, i.e., to the geometric multiplicity of λ. This motivates the following definition.

For an eigenvalue λ of $A \in M_{n,n}$, the **algebraic multiplicity** of λ is equal to its multiplicity as a root of the characteristic polynomial $\chi_A(t)$. Since $\chi_A(t)$ is a similarity invariant, so are the algebraic multiplicities of the eigenvalues. Moreover we deduce the following result.

PROPOSITION 9.22. *Let $A \in M_{n,n}$. Then A is diagonalizable if and only if $\chi_A(t)$ is split and for all eigenvalues λ, the geometric multiplicity of λ is equal to the algebraic multiplicity of λ.*

PROOF. We have seen that if A is diagonal, then $\chi_A(t)$ is split and the algebraic and geometric multiplicities coincide. Since all of these are similarity invariants, this is a necessary condition for diagonalizability. Conversely, if $\chi_A(t)$ is split then it factors as a product of n (not necessarily distinct) linear factors, and if all of the geometric multiplicities are equal to the algebraic multiplicities, then the sum of the geometric multiplicities is equal to the sum of the algebraic multiplicities, which (since $\chi_A(t)$ is split!) is equal to n. Thus A is diagonalizable. □

EXAMPLE 9.23. *Suppose that $A \in M_{4,4}$ has characteristic polynomial $\chi_A(t) = (t-1)^2(t-3)(t-4)$. Since $\chi_A(t)$ is split, A may or may not be diagonalizable: we need to do some computations. We begin by computing V_1 and its dimension. If $\dim V_1 = 1$, then the geometric multiplicity of $\lambda = 1$ is less than its algebraic multiplicity, so there is no need to compute V_3 and V_4: A is not diagonalizable. Conversely if $\dim V_2 = 2$ then since $\dim V_3 \geq 1$ and $\dim V_4 \geq 1$ we must have $\dim V_2 = 2$, $\dim V_3 = \dim V_4 = 1$ and thus $\dim V_2 + \dim V_3 + \dim V_4 = 4 = \dim \mathbb{R}^4$, so A is diagonalizable.*

In our discussion we have saved the following result for last: it seems more technical and less useful than the others.

THEOREM 9.24. *The geometric multiplicity is always less than or equal to the algebraic multiplicity. More precisely: let λ be an eigenvalue for $A \in M_{n,n}$. Then $\dim V_\lambda$ is less than or equal to the multiplicity of λ as a root of $\chi_A(t)$.*

PROOF. Let $v_1, \ldots, v_k$ be a basis for V_λ and extend it to a basis $v_1, \ldots, v_n$ for $\mathbb{R}^n$. The matrix of A with respect to this basis has the block form

$$B = \begin{bmatrix} \lambda I_k & * \\ 0 & A' \end{bmatrix},$$

where $A' \in M_{n-k,n-k}$. Since B is similar to A, we have

$$\chi_A(t) = \chi_B(t) = \det(tI_n - B).$$

In order to compute this determinant we need only row reduce to get an upper triangular matrix. We can do so by performing row operations on the last $n-k$ rows only, so as to make the lower right corner the identity matrix; doing so we acquire a factor of $\chi_{A'}(t)$, so that

$$\chi_A(t) = \chi_{A'}(t) \begin{bmatrix} (t-\lambda)I_k & * \\ 0 & I_{n-k} \end{bmatrix} = (t-\lambda)^k \chi_{A'}(t).$$

So the multiplicity of λ as a root of $\chi_A(t)$ is at least k. □

9.4. An Alternative to the Characteristic Polynomial.

We denote by $\mathbb{R}[t]$ the set of all polynomials $p(t) = a_n t^n + \ldots + a_1 t + a_0$ with real coefficients. Polynomials are ubiquitous and flexible algebraic objects (much like vector spaces and matrices). We can think of them on one hand as formal expressions which can be added and multiplied. On the other hand, we can "plug things into them". E.g. in calculus a polynomial is usually thought of as a function $\mathbb{R} \to \mathbb{R}$, $x \in \mathbb{R} \mapsto p(x)$. But we can also plug in an $n \times n$ matrix, with the convention that the constant term a_0 denotes the scalar matrix $a_0 I_n$. Thus e.g. if $p(t) = t^2 + t + 1$, then

$$p(A) = A^2 + A + I_n.$$

We say that a matrix A **satisfies** a polynomial $p(t)$ if $p(A) = 0$.

EXERCISE 9.13. *Let $A, B \in M_{n,n}$ be similar matrices, and let $p(t)$ be a polynomial. Suppose A satisfies $p(t)$. Show: B satisfies $p(t)$.*

THEOREM 9.25. *Let $A \in M_{n,n}$. Then there is a monic polynomial $p(t)$ of degree at most n^2 such that $p(A) = 0$.*

PROOF. We identify $M_{n,n}$ with $\mathbb{R}^{n^2}$, by sending the matrix $A = (a_{ij})$ to the vector

$$(a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{nn}).$$

In this way we consider the powers $I = A^0$, A^1, A^2, ... as elements of $\mathbb{R}^{n^2}$. Because $\dim \mathbb{R}^{n^2} = n^2$, there must be some nontrivial linear relation among $I_n = A^0, \ldots, A^{n^2}$. Arguing in the usual manner there is a $k \leq n^2$ such that $A^0, \ldots, A^{k-1}$ are linearly independent and

$$A^k = c_{k-1}A^{k-1} + \ldots + c_1 A + C_0 I_n.$$

Then if $p(t) = t^k - c_{k-1}t^{k-1} - \ldots - c_1 t - c_0$, we have $p(A) = 0$. □

In the above proof we found a monic polynomial $p(t)$ of *least degree* k such that $p(A) = 0$. In fact this polynomial is unique. This may seem surprising at first, but the argument is simple: suppose $q(t)$ is another monic polynomial of minimal degree such that $q(A) = 0$. Put $r(t) = p(t) - q(t)$. Then $r(A) = p(A) - q(A) = 0 - 0 = 0$. Since $p(t)$ and $q(t)$ are both monic of the same degree, the highest order terms cancel out and r has smaller degree. If r is not the zero polynomial then we may write

$$r(t) = d_\ell t^\ell + \ldots + d_1 t + d_0$$

and $d_\ell \neq 0$. Then $\frac{1}{d_\ell} r(t)$ is a monic polynomial of degree $\ell < k$ and $\frac{1}{d_\ell} r(A) = \frac{1}{d_\ell} 0 = 0$: this contradicts the minimality of $p(t)$. Thus it must be that $r(t)$ is the zero polynomial: i.e., $p(t) = q(t)$.

We call this unique monic polynomial of least degree satisfied by A the **minimal polynomial of A** and write it as $m_A(t)$.

PROPOSITION 9.26. *Let $A \in M_{n,n}$, and let $p(t)$ be a polynomial satisfied by A. Then $m_A(t)$ divides $p(t)$: there is a polynomial $q(t)$ such that $p(t) = m_A(t)q(t)$.*

PROOF. We use polynomial division with remainder: there are polynomials $q(t)$ and $r(t)$ such that

$$p(t) = m_A(t)q(t) + r(t)$$

and $\deg r < \deg m_A$. Now plug in A:

$$0 = p(A) = m_A(A)q(A) + r(A) = 0 \cdot q(A) + r(A) = r(A).$$

Thus $r(t)$ is a polynomial of smaller degree than $m_A(t)$ satisfied by A; as we saw above, this means that $r(t)$ is the zero polynomial and thus $p(t) = m_A(t)q(t)$. □

PROPOSITION 9.27. *Let $A \in M_{n,n}$, and let $p(t)$ be a monic polynomial satisfied by A. Then for every eigenvalue λ of A, we have $p(\lambda) = 0$.*

PROOF. Let v be an eigenvector for λ, i.e., a nonzero vector in $\mathbb{R}^n$ such that $Av = \lambda v$. Then since $A^k v = \lambda^k v$, adding these up we find that

$$0v = p(A)v = p(\lambda)v.$$

Since $v \neq 0$, we must have $p(\lambda) = 0$. □

A monic polynomial $p(t)$ is **squarefree split** if it is a product of *distinct* linear factors. Thus e.g. $t^2 + t = t(t+1)$ is squarefree split and $t^3 + t^2 = t^2(t+1)$ is split but not squarefree split.

EXERCISE 9.14. *Let $f(t)$ and $g(t)$ be monic polynomials. Show: if g is squarefree split and f divides g, then f is squarefree split.*

THEOREM 9.28. *For $A \in M_{n,n}$, the following are equivalent:*
(i) A is diagonalizable.
(ii) There is a squarefree split polynomial $p(t)$ such that $p(A) = 0$.
(iii) The minimal polynomial $m_A(t)$ is squarefree split.

PROOF. (i) $\implies$ (ii): By Exercise 9.13, the set of polynomials satisfied by a matrix is a similarity invariant, so we may as well assume that A is diagonal. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of A (i.e., the distinct real numbers which comprise the diagonal entries of A). Let $p(t) = (t-\lambda_1)\cdots(t-\lambda_k)$, and observe that $p(t)$ is squarefree split. We claim that A satisfies $p(t)$. Indeed,

$$p(A) = (A - \lambda_1 I_n)\cdots(A - \lambda_k I_n)$$

is a diagonal matrix with ith diagonal entry equal to $(a_{ii} - \lambda_1)\cdots(a_{ii} - \lambda_k) = 0$ since each $a_i i$ is equal to one of the λ's.
(ii) $\implies$ (iii): Since $p(A) = 0$ and $m_A(t)$ is the minimal polynomial, we have $m_A(t)$ divides $p(t)$. By Exercise 9.14 the polynomial $m_A(t)$ is squarefree split.
(iii) $\implies$ (ii) is immediate.
(ii) $\implies$ (i): Let $p(t) = (t - r_1)\cdots(t - r_k)$ with $r_1, \ldots, r_k$ distinct be a squarefree split polynomial satisfied by A: thus

$$(A - r_1 I_n)\cdots(A - r_k I_n) = 0.$$

By Corollary 4.35, we have

$$\sum_{i=1}^{k} \dim V_{r_i} = \sum_{i=1}^{k} \text{nullity}(A - r_i I_n) \geq n,$$

so A is diagonalizable by Theorem 9.7. □

Theorem 9.28 seems in many ways more insightful than the characterization of diagonalization in terms of algebraic and geometric multiplicities. For one thing, if A has rational entries, then $m_A(t)$ has rational coefficients and its computation requires only row operations with matrices with rational entries. If our primary goal is to determine whether A is diagonalizable then computin $m_A(t)$ and determining whether it is squarefree split is faster and more straightforward than computing eigenspaces.

The result also gives us insight into some of our previous results. For instance, we saw that every (not necessarily orthogonal) projection and reflection was diagonalizable. From our present persepctive, this is immediate: a matrix A is a projection if $A^2 = A$, i.e., if A satisfies the squarefree split polynomial $t^2 - t = t(t-1)$. Similarly, A is a reflection if $A^2 = I_n$, i.e., if A satisfies the squarefree split polynomial $t^2 - 1 = (t+1)(t-1)$.

There is a natural question that our discussion so far has (somewhat nontraditionally) avoided. Namely, we have defined two polynomials attached to A, the minimal polynomial $m_A(T)$ and the characteristic polynomial $\chi_A(t)$. How are they related?

Here is one case: if A is diagonalizable, then the characteristic polynomial is split, so is of the form $\prod_{i=1}^{k}(t - \lambda_i)^{r_i}$. In the proof of Theorem 9.28 we showed that the squarefree part of this polynomial, namely $p(t) = \prod_{i=1}^{k}(t - \lambda - i)$, is satisfied by A. Thus the minimal polynomial $m_A(T)$ divides $\prod_{i=1}^{k}(t - \lambda_i)$. Looking back at that calculation, we see that we need each factor $t - \lambda_i$ to kill all the diagonal entries of $p(A)$, so that $m_A(t) = \prod_{i=1}^{k}(t-\lambda_i)$. In particular $m_A(t)$ divides $\chi_A(t)$: equivalently, A satisfies its characteritic polynomial $\chi_A(t)$ in this case.

The following is an extremely elegant theorem whose proof we defer until the next chapter.

THEOREM 9.29. *(Cayley-Hamilton) Let $A \in M_{n,n}$.*
a) A satisfies its characteristic polynomial: $\chi_A(A) = 0$. Equivalently, the minimal polynomial divides the characteristic polynomial.

b) The minimal polynomial and the characteristic polynomial have the same irreducible factors (although the characteristic polynomial may have them with larger multiplicities).

EXAMPLE 9.30. *By pure brute force we will verify the Cayley-Hamilton Theorem for $n = 2$. Namely, let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then as we know, $\chi_A(t) = t^2 - (a+d)t + (ad - bc)$. We compute*

$$\chi_A(A) = A^2 - (a+d)A + (ad-bc)I_2$$
$$= \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} - \begin{bmatrix} a^2+da & ab+bd \\ ac+cd & ad+d^2 \end{bmatrix} + \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

EXERCISE 9.15. *Use bruter force to verify the Cayley-Hamilton Theorem for $n = 3$.*

Suppose that $\chi_A(t)$ is split. Then the irreducible factors of the characteristic polynomial are precisely $t - \lambda$ as λ runs through the eigenvalues of A. By part a) of the Cayley-Hamilton Theorem, the minimal polynomial $m_A(t)$ is also split, and by Proposition 9.27 (and the Root Factor Theorem: if $p(\lambda) = 0$ then $(t - \lambda)$ divides $p(t)$) for every eigenvalue λ, $t - \lambda$ divides $m_A(t)$. Thus in this case part b) follows from part a).

From our perspective, the new content of part b) of the Cayley-Hamilton Theorem lies in the case where the characteristic polynomial has irreducible quadratic factors, e.g. $t^2 + 1$. In this chapter we are working over the real numbers as a scalar field. If instead we admitted complex numbers as scalars then the "splitness" issue would evaporate and we would only need part a) of the Cayley-Hamilton Theorem.

COROLLARY 9.31. *For $A \in M_{n,n}$, the minimal polynomial $m_A(t)$ has degree at most n.*

EXERCISE 9.16. *Prove Corollary 9.31.*

Some history: The Cayley-Hamilton Theorem was stated by Arthur Cayley in 1858. Arthur Cayley (1821-1895) was the greatest English mathematician since Newton. In fact, in the long years in between Newton's death (1727) and Cayley's ascendancy (circa 1860), English mathematics lay remarkably fallow. The high esteem in which British pure mathematics has been held for the last century or so is probably due more to Cayley than anyone else. William Rowan Hamilton (1805-1865) was a leading Irish mathematician who deeply studied rotations in three-dimensional space and invented quaternions, among other things. By modern standards of pure mathematics, the following seems somewhere between amusing and scandalous: neither Cayley nor Hamilton even attempted a proof of the Cayley-Hamilton Theorem in the general case! In 1858 Cayley checked the $n = 2$ case – as we did in Example 9.30 – and in the $n = 3$ case (as we assigned as an exercise). On the basis of these calculations he was quite confident of the general case. Five years earlier, Hamilton had checked the result for rotations in $\mathbb{R}^3$. On this basis the result is named after them! It seems that the first proof of the general case was given by Georg Frobenius (a German mathematician and one of the true founders of modern algebra) in 1878.

Many proofs have since been given. However, elementary proofs of the Cayley-Hamilton theorem tend not to be very insightful or rewarding. We gave a proof in the diagonalizable case. Once one reaches a certain level of algebraic sophistication it is possible to explain by "pure thought" why the general case follows from this. In the next chapter we will give a proof built around the fact that any matrix is similar to a block diagonal matrix each of whose blocks is a companion matrix.

9.5. The Spectral Theorem.

Let $A \in M_{n,n}$. So far we have studied the question of whether there is an invertible $P \in M_{n,n}$ such that $P^{-1}AP$ is diagonal. One interpretation of this is that the columns of B form a new coordinate system for $\mathbb{R}^n$ with respect to which the linear transformation has a very simple structure.

However, not all coordinate systems in $\mathbb{R}^n$ are created equal. As we saw, the standard basis has a property that most other bases lack: it is an orthonormal basis, and this explains why if $v = a_1e_1 + \ldots + a_ne_n$, then the coefficient a_i is simply $v \cdot e_i$. Although from an algebraic perspective it is certainly very helpful to have any basis of eigenvectors, from a geometric perspective it would be more natural to have an orthonormal basis of eigenvectors. This motivates the following definition.

A matrix $A \in M_{n,n}$ is **orthogonally diagonalizable** if there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

EXERCISE 9.17. *Show: for $A \in M_{n,n}$, the following are equivalent.*
(i) A is orthogonally diagonalizable.
(ii) A admits an orthonormal basis of eigenvectors.
(iii) A admits an orthogonal basis of eigenvectors.

It is perhaps not immediately clear that orthogonal diagonalizability is really a stronger condition than mere diagonalizability. Up until this point, our take on orthonormal bases is that they are nice but nothing special: if you have a basis and want an orthonormal basis, no problem: apply the Gram-Schmidt process. However, the Gram-Schmidt process usually does not preserve eigenvectors, and indeed it is not hard to come up with examples of matrices that admit bases for eigenvectors but no orthonormal bases. Indeed, consider $A \in M_{2,2}$ which has distinct eigenvalues λ_1, λ_2. Then the eigenspaces V_{λ_1}, V_{λ_2} are both lines in the plane. These lines might be orthogonal to each other and they might not: if not, there is no orthogonal basis of eigenvectors.

EXAMPLE 9.32. *Let $A = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$. The characteristic polynomial is $\chi_A(t) = t^2 + t$, so the eigenvalues are $\lambda = 0, l$. Solving for the nullspaces, we find $V_0 = \text{span}(1,1)$ and $V_1 = \text{span}(1,0)$. The vectors $(1,0)$ and $(1,1)$ are simply not perpendicular to each other, so – although there is a basis of eigenvectors and A is diagonalizable – there is on orthogonal basis of eigenvectors, so A is not orthgonally diagonalizable. Notice that A is a projection operator: $A^2 = A$, and for all such operators $\mathbb{R}^n = V_1 \oplus V_0$. However it is not an* orthogonal *projection: this means precisely that V_0 and V_1 are not orthogonal subspaces.*

EXERCISE 9.18. *Let $A \in M_{n,n}$ be a projection: $A^2 = A$, so that $\mathbb{R}^n = V_1 \oplus V_0$ and A is diagonalizable. Show that A is orthogonally diagonalizable if and only if $V_1 \perp V_0$, i.e., if and only if A is an orthogonal projection.*

I claim that in fact I can tell immediately upon looking at $\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$ that it is not orthogonally diagonalizable. Why is that? Because of the following result.

PROPOSITION 9.33. *If $A \in M_{n,n}$ is orthogonally diagonalizable then A is symmetric: $A^T = A$.*

PROOF. Suppose that there is an orthogonal matrix P such that $P^{-1}AP = D$ is diagonal. Then $A = PDP^{-1}$, so

$$A^T = (PDP^{-1})^T = (P^{-1})^T D^T P^T = PDP^{-1} = PDP^{-1} = A. \qquad \square$$

EXERCISE 9.19. *Show: if A is symmetric and P is orthogonal, then $P^{-1}AP$ is symmetric.*

Well, that was easy. In general in mathematics when you learn a result of the form $A \implies B$, you should immediately inquire about the converse. Sometimes the proof that you gave of $A \implies B$ can be easily turned around to give a proof of $B \implies A$. Not always, of course: often enough the converse is false.[13] There is a third possibility: sometimes the converse is also true, but the proof of $B \implies A$ has nothing to do with the proof of $B \implies A$. Sometimes the innocent question "Is the converse also true?" leads us to some deep results. That is the case here.

[13]All squares are rectangles, but not all rectangles are squares.

THEOREM 9.34 (Spectral Theorem). *Every symmetric matrix $A \in M_{n,n}(\mathbb{R})$ is orthogonally diagonalizable.*

We will prove the Spectral Theorem in the next section: it will take some doing.

9.6. Proof of the Spectral Theorem.

EXAMPLE 9.35. *Let v be an eigenvector for A, with eigenvalue λ, and let $V = \operatorname{span} v$. Then every element w of V is of the form $w = \alpha v$ for some $\alpha \in \mathbb{R}$, so $Aw = A(\alpha v) = \alpha Av = \alpha\lambda v \in V$. Thus V is a one-dimensional invariant subspace. Conversely, let $V = \operatorname{span} v$ be a one-dimensional invariant subspace. Then $Av \in V$, and every element of V is of the form αv for some $\alpha \in \mathbb{R}$, so $Av = \alpha v$ and v is an eigenvector. We deduce that the one-dimensional invariant subspaces are precisely the lines spanned by eigenvectors.*

PROPOSITION 9.36. *Let $A \in M_{n,n}$ be either symmetric or orthongal. Then the eigenspace are orthogonal: if $\lambda_2 \neq \lambda_2$ are eigenvalues of A, then $V_{\lambda_1} \perp V_{\lambda_2}$.*

PROOF. Let $v \in V_{\lambda_1}$ and $w \in V_{\lambda_2}$.

First we suppose that A is symmetric. Then we have

$$\lambda_1 v \cdot w = (\lambda_1 v) \cdot w = (Av) \cdot w = v \cdot (A^T w) = v \cdot Aw = v \cdot (\lambda_2 w) = \lambda_2 v \cdot w,$$

so $(\lambda_1 - \lambda_2)(v \cdot w) = 0$. Since $\lambda_1 \neq \lambda_2$ we have $v \cdot w = 0$.

Now suppose that A is orthgonal, and let λ be an eigenvalue of A. Then there is a nonzero $v \in \mathbb{R}^n$ such that $Av = \lambda v$. Since A is orthogonal we have

$$||v|| = ||Av|| = ||\lambda v|| = |\lambda|||v||,$$

so $\lambda \in \{\pm 1\}$. Thus without loss of generality we have $\lambda_1 = 1$ and $\lambda_2 = -1$, so

$$v \cdot w = (Av) \cdot w = v \cdot (A^T w) = v \cdot (A^{-1} w) = v \cdot (-w) = -(v \cdot w).$$

Thus $2(v \cdot w) = 0$, so $v \cdot w = 0$. □

In light of Proposition 9.36b), in order to prove the Spectral Theorem it is enough to show that every symmetric matrix has a basis of eigenvalues.

LEMMA 9.37. *Let $A \in M_{n,n}(\mathbb{R})$ be symmetric, and let $\alpha, \beta \in \mathbb{R}$ be such that $\alpha^2 < 4\beta$. Then $A^2 + \alpha A + \beta I_n$ is invertible.*

PROOF. We claim that for all $0 \neq v \in \mathbb{R}^n$,

$$(A^2 + \alpha A + \beta I_n)v \cdot v > 0.$$

If so then the null space of $A^2 + \alpha A + \beta I_n$ is $\{0\}$ and thus $A^2 + \alpha A + \beta I_n$ is invertible.

Step 1: Recall the Cauchy-Schwarz inequality: for $v, w \in \mathbb{R}^n$,

$$|v \cdot w| \leq ||v||||w||.$$

Thus for any $\alpha \in \mathbb{R}$,

$$-\alpha(v \cdot w) \leq |-\alpha||v \cdot w| = |\alpha||v \cdot w| \leq |\alpha|||v||||w||;$$

multiplying through by -1 gives

$$\alpha(v \cdot w) \geq -|\alpha|||v||||w||.$$

Using this we find

$$\begin{aligned}
(A^2 + \alpha A + \beta I_n)v \cdot v &= (A^2 v) \cdot v + \beta(v \cdot v) \\
&= Av \cdot A^T v + \alpha(Av \cdot v) + \beta||v||^2 \\
&= Av \cdot aV + \alpha(Av \cdot v) + \beta||v||^2 \\
&= ||Av||^2 + \alpha(Av \cdot v) + \beta||v||^2 \\
&\geq ||Av||^2 - |\alpha|||Av||||v|| + \beta||v||^2 \\
&= \left(||Av|| - \frac{|\alpha|||v||}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right)||v||^2 > 0.
\end{aligned}$$

□

PROPOSITION 9.38. *Let $A \in M_{n,n}(\mathbb{R})$ be symmetric. Then A has an eigenvalue.*

PROOF. Let $0 \neq v \in \mathbb{R}^n$. Then the set $\{v, Av, \ldots, A^n v\}$ consists of $n+1$ vectors in $\mathbb{R}^n$, so they are linearly dependent: there are $a_0, \ldots, a_n \in \mathbb{R}$, not all zero, such that

$$a_n A^n v + \ldots + a_1 Av + a_0 v = 0.$$

There is $N \leq n$ such that $a_N \neq 0$ and

$$a_N a^N v + \ldots + a_1 Av + a_0 v = 0.$$

Let $p(t) = a_N t^N + \ldots + a_1 t + a_0$, so $p(A)v = 0$. We factor $p(t)$ as

$$p(t) = a_N(t^2 + \alpha_1 t + \beta_1) \cdots (t^2 + \alpha_s t + \beta_s) \cdot (t - \lambda_1) \cdots (t - \lambda_r),$$

where the quadratic polynomials $t^2 + \alpha_i t + \beta_i$ have no real roots – equivalently by the quadratic formula, $\alpha_i^2 < 4\beta_i$ for all i. Since

$$0 = p(A)v = a_N(A^2 + \alpha_1 A + \beta_1 I_n) \cdots (A^2 + \ldots + \alpha_s A + \beta_s I_n)(A - \lambda_1 I_n) \cdots (A - \lambda_r I_n)v.$$

By Lemma 9.37 each matrix $A^2 + \alpha_i A + \beta_i I_n$ is invertible, so multiplying by their inverses gives

$$0 = (A - \lambda_1 I_n) \cdots (A - \lambda_r I_n)v.$$

If $(A - \lambda_r I_n)v = 0$, then λ_r is an eigenvalue. If not, then

$$v' \coloneqq (A - \lambda_r I_n)v \neq 0,$$

so if $(A\lambda_{r-1} I_n)v' = 0$ then λ_{r-1} is an eigenvalue. And so forth: since the product is zero, at some point multiplying by $(A - \lambda_i I_n)$ must convert a nonzero vector to the 0 vector, so one of $\lambda_1, \ldots, \lambda_r$ is an eigenvalue for A. □

Proof of the Spectral Theorem: We go by induction on n, the case $n = 1$ being trivial. So suppose $n \geq 2$ and every symmetric matrix $B \in M_{n-1,n-1}(\mathbb{R})$ is diagonalizable. Let $A \in M_{n,n}(\mathbb{R})$ be a symmetric matrix. By Proposition 9.38, there is an eigenvector v for A, say with eigenvalue λ. By rescaling v we may choose v to be a unit vector. Let $v, v_2, \ldots, v_n$ be an orthonormal basis of $\mathbb{R}^n$ (extend v to a basis then apply Gram-Schmidt to get an orthonormal basis). Let P be the (orthogonal!) matrix with columns $v, v_2, \ldots, v_n$, and let $A' = P^{-1}AP$. Since P is orthogonal, by Exercise 9.19, the matrix A' is again symmetric. The matrix A' is block diagonal, with upper left 1×1 block λ and bottom right $n-1 \times n-1$ block B, say. Since A' is symmetric, so is B. By induction, B is orthogonally diagonalizable: there is an orthogonal matrix $Q_1 \in M_{n-1,n-1}$ such that $Q_1^{-1} B Q_1$ is diagonal. Thus if Q is the block matrix $1 \oplus Q_1$, then Q is orthogonal and

$$Q^{-1}A'Q = Q^{-1}P^{-1}APQ = (PQ)^{-1}A(PQ)$$

is diagonal. Since P and Q are orthogonal, so is PQ, so A is orthogonally diagonalizable.

COROLLARY 9.39. *If $A \in M_{n,n}$ is symmetric, then its characteristic polynomial $\chi_A(t)$ has n real roots.*

PROOF. This follows from Theorems 9.34 and 9.19. □

9.7. Canonical Form for Orthogonal Matrices.

PROPOSITION 9.40. *Let $A \in \mathrm{SO}_3$. Then there is a nonzero vector $v \in \mathbb{R}^3$ such that $Av = v$.*

PROOF. Equivalently, we must show that 1 is an eigenvalue of A. Let $\chi_A(t) = t^3 + a_2 t^2 + a_1 t + a_0$ be the characteristic polynomial of A. Recall that by the Intermediate Value Theorem, every cubic polynomial has at least one real root. Thus $\chi_A(t)$ either factors as $(t - \lambda)q(t)$ for an irreducible quadratic polynomial q or factors as $(t - \lambda_1)(t - \lambda_2)(t - \lambda_3)$.

Suppose first that we have three linear factors. Then

$$-\det A = (-1)^3 \det A = \det(-A) = (-\lambda_1)(-\lambda_2)(-\lambda_3),$$

so

$$1 = \det A = \lambda_1 \lambda_2 \lambda_3.$$

Since an orthogonal matrix preserves lengths of vectors, if v is an eigenvector for A with eigenvalue λ, then we we have

$$||v|| = ||Av|| = ||\lambda v|| = |\lambda|||v||,$$

so $|\lambda| = 1$ and $\lambda = \pm 1$. Since $\lambda_1\lambda_2\lambda_3 = 1$ we must have some $\lambda_i = 1$.

Now suppose that $\chi_A(t) = (t - \lambda_1)q(t)$, and write

$$q(t) = t^2 + Bt + C.$$

Since q has no root in $\mathbb{R}$ we must have $B^2 < 4C$ and thus $C > 0$. Setting $t = 0$ we get

$$-1 = \det(-A) = \chi_A(0) = -\lambda C,$$

and thus $\lambda = \frac{1}{C} > 0$, so $\lambda = 1$. □

We will now show that every $A \in \mathrm{SO}_3$ is a rotation through a fixed axis ℓ. The proof will in particular clarify what this means!

THEOREM 9.41 (Principal Axis Theorem). *Let $A \in \mathrm{SO}_3$. Then there is a line through the origin $\ell \subset \mathbb{R}^3$ such that A consists of a rotation with axis ℓ.*

PROOF. By Proposition 9.40 there is a nonzero vector $v_1 \in \mathbb{R}^3$ with $Av_1 = v_1$: we take $\ell = \langle v \rangle$. Let $P = \ell^\perp$, so P is the "normal plane" to ℓ, so $\mathbb{R}^3 = \ell \perp P$. By Theorem 8.2, the plane P is also invariant under A. Let v_2, v_3 be an orthogonal basis for P, and let $R = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be the matrix for $A|_P$ with respect to this basis. Then v_1, v_2, v_3 is an orthogonal basis for $\mathbb{R}^3$, and the matrix of P with respect to this basis is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{bmatrix}.$$

From this we see that

$$1 = \det A = 1 \cdot \det R,$$

so $\det P = 1$. The matrix R must be orthogonal, since a matrix is orthogonal iff the corresponding linear transformation preserves lengths of vectors and since A has this property on all of $\mathbb{R}^3$ it certainly has it on the subspace P. Thus $R \in \mathrm{SO}_2$ so by Example 7.10 it is rotation through an angle of θ. □

We will now use these results to give a "canonical form" for orthogonal matrices.

THEOREM 9.42. *Let $A \in O_n$ be an orthogonal matrix. We have an orthogonal sum decomposition*

$$\mathbb{R}^n = V_1 \perp \ldots \perp V_k$$

such that for all $1 \leq i \leq k$ we have that:
(i) The subspace V_i is A-invariant,
(ii) The dimension of V_i is 1 or 2,
(iii) If $\dim V_i = 1$ then $A|_{V_i} = \pm 1$, and
(iv) If $\dim V_i = 2$ then there is an orthonormal basis $e_{i,1}, e_{i,2}$ of V_i with respect to which the matrix of $A|V_i$ is the rotation matrix $R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

PROOF. Let V_1 and V_{-1} be the 1 and -1 eigenspaces of A, respectively. Then we have $\mathbb{R}^n = V_1 \perp V_{-1} \perp W$, where $W = (V_1 \perp V_{-1})^\perp$. There is a basis of $V_1 \oplus V_{-1}$ on which A is diagonal with eigenvalues ± 1. The map $A_{|W}$ has no eigenvectors and thus no one-dimensional invariant subspaces. By Theorem 8.4 and induction, W is an orthogonal sum of two-dimensional invariant subspaces W_i, and on each W_i the characteristic polynomial $\chi_i(t) = t^2 + b_i t + c_i$ is irreducible quadratic, so $\det A|_{W_i} = c_i > 0$. Thus $c_i = 1$, and by Example 7.10 there is an orthonormal basis of W_i with respect to which the matrix of $A|_{W_i}$ is a rotation matrix R_θ. □

10. Complex Scalars

10.1. Vector Spaces Over a Scalar Field. A **field** is a set F endowed with binary operations

$$+ : F \times F \to F, \ \cdot : F \times F \to F,$$

called addition and multiplication, that are required to satisfy the following "field axioms":

(F1) Commutativity of Addition: $\forall x, y \in F, \ x + y = y + x$
(F2) Associativity of Addition: $\forall x, y, z \in F, \ (x + y) + z = x + (y + z)$
(F3) Existence of an Identity for Addition: there is $0 \in F$ such that for all $x \in F$, we have $0 + x = x + 0 = x$.
(F3) Existence of Inverses for Addition: for all $x \in F$ there is $y \in F$ such that $x + y = y + x = 0$.
(F4) Commutativity of Multiplication: $\forall x, y \in F$ we have $x \cdot y = y \cdot x$.
(F5) Associativity of Multiplication: $\forall x, y, z \in F$ we have $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.
(F6) Existence of an Identity for Multiplication: there is $1 \in F$ such that for all $x \in F$, we have $1 \cdot x = x \cdot 1 = x$.
(F7) Existence of Multiplicative Inverses for Nonzero Elements: for all $0 \neq x \in F$, there is $y \in F$ such that $x \cdot y = y \cdot x = 1$.
(F8) Distributivity of Multiplication Over Addition: for all $x, y, z \in F$ we have

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z) \text{ and } (x + y) \cdot z = (x \cdot z) + (y \cdot z).$$

(F9) Nondegeneracy: We have $0 \neq 1$.

LEMMA 10.1. *For all $x \in F$ we have $0 \cdot x = 0$.*

PROOF.

$$0 \cdot x = (0 + 0) \cdot x = (0 \cdot x) + (0 \cdot x).$$

If y is the additive inverse to $0 \cdot x$, then adding y to both sides gives

$$0 = 0 \cdot x.$$ □

Suppose that (F1) through (F8) holds but (F9) fails: then $0 = 1$, so for all $x \in F$ we have

$$0 = 0 \cdot x = 1 \cdot x = x,$$

and thus F consists of a single element, called both 0 and 1. This is precisely the "degeneracy" that we are not allowing.

In practice we usually abbreviate $x \cdot y$ to xy.

EXAMPLE 10.2.
a) *The real numbers $\mathbb{R}$ form a field. Indeed this is the "scalar field" over which we have been doing linear algebra thus far. Exactly what that means will become clear soon.*
b) *The rational numbers $\mathbb{Q}$ form a field.*
c) *The complex numbers $\mathbb{C}$ form a field. We may view $\mathbb{C}$ as the vector space $\mathbb{R}^2$ together with the usual addition of vectors and the following "exotic" multiplication.*

$$(x_1, y_1) \cdot (x_2, y_2) \coloneqq (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1).$$

In practice though we identify $\mathbb{R}$ as a subset of $\mathbb{C}$ via $x \mapsto (x, 0)$, put $i \coloneqq \sqrt{-1}$, so that

$$(x, y) = x + iy.$$

Note that

$$i^2 = (0, 1) * (0, 1) = (-1, 0) = -1.$$

Then axioms (F1) through (F3) follow from the vector space axioms. We still need to check (F4) through (F9). In this regard: (F4) and (F9) are virtually immediate; (F5), (F6), (F8) and (F9) are straightforward calculations that we leave to the reader. To show (F7) on the other hand we need to do something: if $x, y \in \mathbb{R}$, not both zero, then we must find $w \in \mathbb{C}$

such that $(x+yi)w = 1$. *For this we first introduce, for any* $z = x+iy \in \mathbb{C}$, *the* ***complex conjugate***

$$\overline{z} = \overline{x+iy} \coloneqq x - iy.$$

Then we have the useful identity

$$z\overline{z} = (x+iy)(x-iy) = x^2 + ixy - ixy - i^2y^2 = x^2 + y^2.$$

Thus $z\overline{z}$ *is a non-negative real number and is strictly positive iff* $z \neq 0$. *So if* $z \neq 0$ *then*

$$z \cdot \left(\frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2} \right) = 1,$$

so

$$z^{-1} = \frac{\overline{z}}{x^2+y^2}.$$

d) *If there is a field with two elements, then the elements must be* 0, *the additive identity, and* 1, *the multiplicative identity. By the nature of identity elements we must have* $0+0=0$, $0+1=1$, $1\cdot 0 = 0$, $1\cdot 1 = 1$. *By Lemma 10.1 we must have* $0\cdot 0 = 0$. *This leaves* $1+1$. *If we had* $1+1=1$, *then adding the additive inverse of* 1 *to both sides yields* $1 = 0$, *contradicting (F9). So we must have* $1+1=0$. *One still has to check that these binary operations satisfy the field axioms, and we leave that to the reader.*

It is now time to reveal that **most of** linear algebra can be done with respect to any field of scalars rather than just $\mathbb{R}$. (There are, however, some exceptions where we used special properties of $\mathbb{R}$. We will explain exactly what these are.) Here are the main points:

It makes sense to consider linear equations $a_1x_1 + \ldots + a_1x_n = b$ over a field F: here $a_1, \ldots, a_n, b$ are fixed elements of F and $x_1, \ldots, x_n$ are variables, so that given any $(x_1, \ldots, x_n) \in F^n$ either $a_1x_1 + \ldots + a_nx_n = b$ holds or it doesn't. We may similarly consider systems of linear equations

$$a_{1,1}x_1 + \ldots + a_{1,n}x_n = b_1$$
$$\vdots$$
$$a_{m,1}x_1 + \ldots + a_{m,n}x_n = b_n$$

and consider the set of $(x_1, \ldots, x_n) \in F^n$ that solve all of these equations. Moreover the techniques we used to solve linear systems with real variables and coefficients adapt verbatim to the case of any scalar field F. In particular we may consider matrices with entries in F. There is a notion of row reduction, row echelon form, and reduced row echelon form, and all the results of §2 carry over verbatim.

We can define a dot product on F^n the same way:

$$\forall x = (x_1, \ldots, x_n),\ y = (y_1, \ldots, y_n) \in F^n,\ x \cdot y = (x_1 \cdots x_n) \cdot (y_1, \ldots, y_n) \coloneqq x_1y_1 + \ldots + x_ny_n.$$

However we need to be careful with this dot product because it lacks some properties that it does over $\mathbb{R}$: namely, over $\mathbb{R}$ for any nonzero $x = (x_1, \ldots, x_n) \in R^n$ we have

$$x \cdot x = x_1^2 + \ldots + x_n^2 > 0.$$

In particular, if $x \cdot x = 0$ then $x = 0$. This property holds for some scalar fields F and not for others: it also holds, for instance, for $F = \mathbb{Q}$ since $\mathbb{Q}$ is a subfield of $\mathbb{R}$ – that is, $\mathbb{Q} \subset \mathbb{R}$ and the addition and multiplication on $\mathbb{Q}$ are those on $\mathbb{R}$ restricted to $\mathbb{Q}$ – and indeed it holds for any subfield of $\mathbb{R}$, of which there are many. It does not hold over $\mathbb{F}_2$:

$$(1,1) \cdot (1,1) = 1 \cdot 1 + 1 \cdot 1 = 1 + 1 = 0.$$

Moreover it does not hold over $\mathbb{C}$:

$$(1,i) \cdot (1,i) = 1 \cdot 1 + i \cdot i = 1 + (-1) = 0.$$

One says a field F is **formally real** if for all $n \in \mathbb{Z}^+$, if $x_1, \ldots, x_n \in F$ are such that $x_1^2 + \ldots + x_n^2 = 0$ then $x_1 = \cdots = x_n = 0$.

EXAMPLE 10.3 (Formally Real Fields and Ordered Fields). *COMPLETE ME!*

We defined the length of a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ as

$$||x|| = \sqrt{x_1^2 + \ldots + x_n^2}.$$

This works first because $\mathbb{R}$ is formally real, but also because in $\mathbb{R}$ every positive number has a square root. Even most ordered fields lack this latter property: e.g. because $\sqrt{2}$ is irrational, 2 is not a square in $\mathbb{Q}$ and thus we cannot define $||(1,1)||$ as an element of $\mathbb{Q}$.

It turns out that the notion of a length of a vector simply is not part of linear algebra over a general scalar field F. We do have this notion when $F = \mathbb{C}$, and it even arises from a variation on the standard inner product, as we will see in the next section.

For any field F, we may consider $M_{m,n}(F)$, the set of $m \times n$ matrices with entries in F. Addition and scalar multiplication of matrices holds verbatim. Moreover, if we have $A \in M_{m,k}(F)$ and $B \in M_{k,n}(F)$ then we define $AB \in M_{m,n}(F)$ in exactly the same way: namely, the (i,j) entry is $\sum_{l=1}^k a_{il}b_{lj}$. Although we warned about being careful with dot products, indeed here it is still the case that the (ij) entry of AB is the dot product of the ith row of A with the jth column of B.

For $m, n \in \mathbb{Z}^+$ a **linear transformation** $L : F^n \to F^m$ is a function that satisfies:

(LT1) For all $v, w \in F^n$ we have $L(v + w) = L(v) + L(w)$, and
(LT2) For all $\alpha \in F$ and $v \in F^n$, we have $L(\alpha v) = \alpha L(v)$.

The correspondence between linear transformations and matrices holds verbatim over any scalar field F, as do the rest of the results of §3.

The notions of linear indepenence, spanning, subspace and basis hold verbatim over any scalar field F, as do all the results of §4. We lose the direct geometric interpretations: e.g. a 2-dimensional subspace of F^3 is no longer literally a plane in anything approaching the sense of Euclidean geometry, but it is still extremely useful to maintain this geometric language and intuition. Thus for instance one speaks of one-dimensional subspaces of F^n as "lines," two-dimensional subspaces of F^n as "planes" and $n-1$-dimensional subspaces of F^n as "hyperplanes."

The results of §5 go through verbatim over any scalar field F.

Most of §6 on determinants goes through verbatim over any scalar field: we lose, however, the geometric interpretation of the determinant as the signed change of volume and also the dichotomy that a nonsingular matrix must have either positive or negative determinant. One says a field F has **characteristic 2** if $1 + 1 = 0$ in F. We saw that the field with two elements has characteristic 2: there are many others. If we regard the sign $\text{sgn}(\sigma)$ of a permutation as an element of a field F of characteristic 2, then because $-1 = 1$ in F this becomes trivial: $\sigma(\sigma) = 1$ for all σ. In particular we cannot use the determinant of a permutation matrix to compute the sign of a permutation in characteristic 2; otherwise we can.

We lose all of §7 over a general field F: orthogonal projection onto a line, orthogonal bases and Gram-Schmidt all need the field F to be formally real to even make sense. If the field F is formally real we can do the Gram-Schmidt process to convert any basis to an orthogonal basis; however, we cannot in general rescale to get unit vectors, so we cannot in general get an orthonormal basis.

Over any field F we can still define a matrix to be orthogonal if $A^T = A^{-1}$ and this still gives a

class of invertible matrices with determinant ± 1, but the geometric interpretation of orthogonal matrices as corresponding to distance-preserving linear transformations is lost.

Proposition 8.1 and Theorem 8.3 hold over any scalar field F. For Theorem 8.2 we need the scalar field to be formally real. In Theorem 8.4 we are using the fact that every odd degree polynomial $f \in \mathbb{R}[t]$ has a root in $\mathbb{R}$, so this result breaks down completely over an arbitrary scalar field: e.g. when $F = \mathbb{Q}$, for all $n \in \mathbb{Z}^+$ there is $A \in M_{n,n}(\mathbb{Q})$ such that no subspace $\{0\} \subsetneq V \subsetneq \mathbb{Q}^n$ is A-invariant.

In §9, an eigenvalue for $A \in M_{n,n}(F)$ is an element $\lambda \in F$ such that there is a nonzero $v \in F^n$ such that $Av = \lambda v$. All of the results in §9.1 through 9.4 continue to hold. The Spectral Theorem breaks down completely: over an arbitrary F a symmetric matrix need not be diagonalizable at all, let alone orthogonally diagonalizable.

EXAMPLE 10.4 (Speyer). *Let $F = \mathbb{C}$ and consider the symmetric matrix*

$$A = \begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}.$$

The trace of A is 0 and $\det A = -1 - (i)(i) = 0$, so $\chi_A(t) = t^2$ and the only eigenvalue is 0. Therefore if A were diagonalizable it would be the scalar matrix 0, which it isn't.

In [**MSV93**] necessary and sufficient conditions on a field F are given such that every symmetric matrix $A \in M_{n,n}(F)$ are diagonalizable. In particular such fields must be formally real: see `https://mathoverflow.net/questions/118680`

10.2. The Complex Inner Product.

LEMMA 10.5 (Polarization Identity). *For all $v, w \in \mathbb{C}^n$ we have*

$$\langle v, w \rangle = \frac{1}{4}\left(||v+w||^2 - ||v-w||^2 + i||v-iw||^2 - i||v+iw||^2 \right).$$

EXERCISE 10.1. *Prove Lemma 10.5.*

10.3. Normal Operators and the Spectral Theorem.

PROPOSITION 10.6. *Let $A \in M_{n,n}(\mathbb{C})$. There is a unique matrix A^* such that for all $v, w \in \mathbb{C}^n$ we have*

$$\langle Av, w \rangle = \langle v, A^* w \rangle.$$

Explicitly, the (i,j) entry of A^* is $\overline{a_{j,i}}$ of A. Thus $A^* = (\overline{A})^T = \overline{A^T}$ is the conjugate transpose of A.

PROOF. For $v, w \in \mathbb{C}^n$ we have $\langle v, w \rangle = v^T \overline{w}$, so

$$\langle Av, w \rangle = (Av)^T \overline{w} = v^T A^T \overline{w}$$

and

$$\langle v, A^* w \rangle = v^T (\overline{A^* w}) = v^T \overline{A^*} \overline{w}.$$

If for $1 \leq i, j \leq n$ we take $v = e_i$, $w = e_j$ then we get

$$\langle Ae_i, e_j \rangle = a_{ji}$$

and

$$\langle e_i, A^* e_j \rangle = \overline{a_{ij}^*},$$

showing that we must have

$$a_{ij}^* = \overline{a_{ji}}$$

and thus $A^* = \overline{A^T}$. Conversely, with this choice of A^* it is clear that $v^T A^T \overline{w} = v^T \overline{A^*} \overline{w}$ for all $v, w \in \mathbb{C}^n$. □

We refer to A^* as the **adjoint** of A.

Observe that we could run the entire discussion in $\mathbb{R}^n$ rather than $\mathbb{C}^n$. It then simplifies because complex conjugation is trivial, and we get that for $A \in M_{n,n}(\mathbb{R})$, $A^* = A^T$: i.e., the adjoint is just the transpose.

EXERCISE 10.2. *Show that for all $A \in M_{n,n}(\mathbb{C})$ we have $(A^*)^* = A$.*

LEMMA 10.7. *For a subspace W of $\mathbb{C}^n$, put*

$$\overline{W} = \{\overline{(z_1, \ldots, z_n)} := (\overline{z_1}, \ldots, \overline{z_n}) \in \mathbb{C}^n \mid (z_1, \ldots, z_n) \in W\}.$$

a) *If $v_1, \ldots, v_m$ is a linearly independent list in $\mathbb{C}^n$, then so is $\overline{v_1}, \ldots, \overline{v_m}$.*
b) *If $v_1, \ldots, v_m$ is a spanning set for W, then $\overline{v_1}, \ldots, \overline{v_m}$ is a spanning set for $\overline{W}$.*
c) *If $b_1, \ldots, b_m$ is a basis for W, then $\overline{b_1}, \ldots, \overline{b_m}$ is a basis for $\overline{W}$.*
d) *We have $\dim_{\mathbb{C}} \overline{W} = \dim_{\mathbb{C}} W$.*

PROOF. a) Let $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$ be such that $\alpha_1\overline{v_1} + \ldots + \alpha_m\overline{v_m} = 0$. For $1 \leq i \leq m$, put $\beta_i = \overline{\alpha_i}$. Then we have

$$0 = \alpha_1\overline{v_1} + \ldots + \alpha_m\overline{v_m} = \overline{\beta_1 v_1 + \ldots + \beta_m v_m},$$

which implies

$$\beta_1 v_1 + \ldots + \beta_m v_m = 0,$$

and then by linear independence we get $\beta_1 = \overline{\alpha_1} = \ldots = \beta_m = \overline{\alpha_m} = 0$, which finally implies that $\alpha_1 = \ldots = \alpha_m = 0$.
b) If $w' \in \overline{W}$ then $\overline{w'} \in W$, so there are $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$ such that

$$\overline{w'} = \alpha_1 v_1 + \ldots + \alpha_m v_m,$$

and then

$$w' = \overline{\alpha_1 v_1} + \ldots + \overline{\alpha_m v_m},$$

so $w' \in \langle \overline{v_1}, \ldots, \overline{v_m} \rangle$.
Part c) follows immediately from parts a) and b), and part d) follows immediately from part c). □

PROPOSITION 10.8. *Let $A \in M_{n,n}(\mathbb{C})$. Then we have*

$$(\text{Image } A^*) = (\text{Ker } A)^{\perp} \tag{17}$$

and

$$(\text{Ker } A^*) = (\text{Image } A)^{\perp}. \tag{18}$$

PROOF. Applying (17) with A^* in place of A and using $(A^*)^* = A$, we get

$$(\text{Image } A)^{\perp} = (\text{Image}(A^*)^*)^{\perp} = (\text{Ker } A^*)^{\perp\perp} = \text{Ker } A^*,$$

so it suffices to show (17). Let $u, v \in \mathbb{C}^n$ be such that $Au = 0$. Then

$$0 = \langle Au, v \rangle = \langle u, A^* v \rangle.$$

This shows that $(\text{Image } A^*) \perp \text{Ker } A$, so $\text{Image } A^* \subset (\text{Ker } A)^{\perp}$. We know that $\text{rank } A = \text{rank } A^T$. Moreover, for any $B \in M_{n,n}(\mathbb{C})$, if W is the image of B then $\overline{W}$ is the image of $\overline{B}$, so by Lemma 10.7d) we have $\text{rank } B = \text{rank } \overline{B}$. Combining this with the Dimension Theorem we get

$$\dim \text{Image } A^* = \dim \text{Image } \overline{A} = \dim \text{Image } A = n - \dim \text{Ker } A = \dim(\text{Ker } A)^{\perp}.$$

It follows that $\text{Image } A^* = (\text{Ker } A)^{\perp}$. □

A matrix $A \in M_{n,n}(\mathbb{C})$ is **Hermitian** (or **self-adjoint**) if $A^* = A$. Thus a real matrix is self-adjoint iff it is symmetric.

If $A \in M_{n,n}(\mathbb{R})$ then $\chi_A(t)$ has n real roots: Corollary 9.39. We worked quite hard to prove this: it came as a byproduct of the Spectral Theorem, that A is orthogonally diagonalizable. It is a little weird how much easier it is to prove the following (stronger!) result.

PROPOSITION 10.9. *All the eigenvalues of a Hermitian matrix $A \in M_{n,n}(\mathbb{C})$ are real.*

PROOF. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A: so there is $0 \neq v \in \mathbb{C}^n$ such that $Av = \lambda v$. Then

$$\lambda||v||^2 = \langle \lambda v, v\rangle = \langle Av, v\rangle = \langle v, A^*v\rangle = \langle v, Av\rangle$$
$$= \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle = \overline{\lambda}||v||^2.$$

Since $v \neq 0$ we have $||v||^2 \neq 0$ and thus $\lambda = \overline{\lambda}$. □

EXERCISE 10.3. *A matrix $A \in M_{n,n}(\mathbb{C})$ is **skew-Hermitian** if $A^* = -A$.*

(i) *Show: if A is skew-Hermitian, then every eigenvalue of A is purely imaginary, i.e., is of the form ib for some $b \in \mathbb{R}$.*

(ii) *Show: A is skew-Hermitian iff iA is Hermitian.*

Let's reflect on the weird state of affairs a bit. The problem is that over the scalar field $\mathbb{R}$ a matrix need not have any eigenvalues at all, whereas any $A \in M_{n,n}(\mathbb{C})$ certainly has n eigenvalues (counting multiplicity). It is much easier to show that a complex number is real than to show that something that might not exist actually does. As a corollary to this, since $\mathbb{R} \subset \mathbb{C}$ it is often useful to think of the eigenvalues of a real matrix as complex numbers $\lambda_1, \ldots, \lambda_n$ as complex numbers that may or may not be real. Of course we must keep in mind that a necessary condition for diagonalizability (or even triangularizability) is that all the eigenvalues be real.

A matrix $A \in M_{n,n}(\mathbb{C})$ is **unitary** if $AA^* = I_n$. Notice that a real matrix is unitary iff it is orthogonal.

A matrix $A \in M_{n,n}(\mathbb{C})$ is **normal** (or a **normal operator**) if $AA^* = A^*A$.

LEMMA 10.10. *Let $A \in M_{n,n}(\mathbb{C})$ be such that $\langle Av, v\rangle = 0$ for all $v \in \mathbb{C}^n$. Then $A = 0$.*

PROOF. For $v, w \in \mathbb{C}^n$ we have

$$\langle A(v+w), v+w\rangle - \langle A(v-w), v-w\rangle + i\langle A(v+iw), v+iw\rangle - i\langle A(v-iw), v-iw\rangle$$
$$= 4\langle Av, w\rangle.$$

Thus if $\langle Av, v\rangle = 0$ for all $v \in \mathbb{C}^n$ then $\langle Av, w\rangle = 0$ for all $v, w \in \mathbb{C}^n$. In particular, for all $v \in \mathbb{C}^n$, taking $w = Av$ we get $\langle Av, Av\rangle = 0$ and thus $Av = 0$. So $A = 0$. □

Lemma 10.10 fails over the scalar field $\mathbb{R}$: if $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is the matrix of a rotation through $\frac{\pi}{2}$, then $A \neq 0$ but $\langle Av, v\rangle = 0$ for all $v \in \mathbb{R}^2$.

PROPOSITION 10.11. *For $A \in M_{n,n}(\mathbb{C})$, the following are equivalent:*

(i) *We have $||Av|| = ||A^*v||$ for all $v \in \mathbb{C}^n$.*

(ii) *The operator A is normal.*

PROOF. The matrix A is normal iff $A^*A - AA^* = 0$. By Lemma 10.10 this holds iff

$$\forall v \in V,\ \langle (A^*A - AA^*)v, v\rangle = 0 \iff$$
$$\forall v \in V,\ \langle AA^*v, v\rangle = \langle A^*Av, v\rangle \iff$$
$$\forall v \in V,\ \langle Av, Av\rangle = \langle A^*v, A^*v\rangle \iff$$
$$\forall v \in V,\ ||Av|| = ||A^*v||.$$

□

COROLLARY 10.12. *Let $A \in M_{n,n}(\mathbb{C})$ be a normal operator. Then*

$$\mathbb{C}^n = (\operatorname{Ker} A) \perp (\operatorname{Image} A).$$

PROOF. By Proposition 10.11, for $v \in \mathbb{C}^n$ we have $v \in \operatorname{Ker} A \iff ||Av|| = 0 \iff ||A^*v|| = 0 \iff v \in \operatorname{Ker} A^*$, so by Proposition 10.8 we have

$$\mathbb{C}^n = (\operatorname{Ker} A^*) \perp (\operatorname{Image} A) = (\operatorname{Ker} A) \perp (\operatorname{Image} A).$$

□

EXERCISE 10.4. *Let $A \in M_{n,n}(\mathbb{C})$. Let $p(t) = a_n t^n + \ldots + a_1 t + a_0 \in \mathbb{C}[t]$ be a polynomial. Put $\overline{p}(t) \coloneqq \overline{a_n} t^n + \ldots + \overline{a_1} t + \overline{a_0}$.*

a) *Show: We have $(p(A))^* = \overline{p}(A^*)$.*
b) *Show: if A is normal, then so is $p(A)$.*

LEMMA 10.13. *Let $A, U \in M_{n,n}(\mathbb{C})$. If A is normal and U is unitary, then $U^{-1}AU$ is normal.*

PROOF. We have $(U^{-1}AU)^* = U^*A^*(U^{-1})^* = U^{-1}A^*U$. Since unitary matrices preserve lengths of vectors, for $v \in \mathbb{C}^n$ we have

$$||(U^{-1}AU)^*v|| = ||U^{-1}A^*Uv|| = ||A^*(Uv)|| = ||A(Uv)|| = ||U^{-1}AUv||.$$

So $U^{-1}AU$ is normal by Proposition 10.11. □

THEOREM 10.14. *Let $A \in M_{n,n}(\mathbb{C})$ be a normal operator.*

a) *For $\lambda \in \mathbb{C}$, let V_λ denote the λ-eigenspace for A, and let W_λ denote the λ-eigenspace for A^*. Then for all $\lambda \in \mathbb{C}$ we have*
$$V_\lambda = W_{\overline{\lambda}}.$$
b) *If $\lambda_1 \neq \lambda_2$ then $V_{\lambda_1} \perp V_{\lambda_2}$.*
c) *Let $v_1, \ldots, v_n \in \mathbb{C}^n$. Then $v_1, \ldots, v_n$ is an orthonormal basis of eigenvectors for A iff it is an orthonormal basis of eigenvectors for A^*.*

PROOF. a) For $\lambda \in \mathbb{C}$, by Exercise 10.4 we have that since A is normal, so is $A - \lambda I_n$. So for $v \in \mathbb{C}^n$ we have

$$v \in V_\lambda \iff ||(A - \lambda I_n)v|| = 0 \iff ||(A - \lambda I)^*v|| = 0 \iff Av = \overline{\lambda}v \iff v \in W_{\overline{\lambda}}.$$

b) Let $v \in V_{\lambda_1}$ and $w \in V_{\lambda_2}$. Then we have

$$(\lambda_1 - \lambda_2)\langle v, w\rangle = \langle \lambda v, w\rangle - \langle v, \overline{\lambda_2} w\rangle = \langle Av, w\rangle - \langle v, A^*w\rangle = 0.$$

Since $\lambda \neq \lambda_2$, this forces $\langle v, w\rangle = 0$.
c) Whether $v_1, \ldots, v_n$ is an orthonormal basis does not involve the matrices at all. By part a), each v_i is an eigenvector for A iff it is an eigenvector for A^*. □

THEOREM 10.15 (Schur). *Every $A \in M_{n,n}(\mathbb{C})$ is unitarily triangularizable: there is $U \in U_n$ such that $U^{-1}AU$ is upper triangular.*

PROOF. If $u_1, \ldots, u_n$ is an orthonormal basis of $\mathbb{C}^n$ and $U = (u_1|\cdots|u_n)$, then $U^{-1}AU$ is the matrix of the linear transformation $A\bullet$ with respect to the basis $u_1, \ldots, u_n$. So it suffices to show that $A\bullet$ is upper triangular with respect to some orthonormal basis. By Theorem 9.21 $A_\bullet$ is upper triangular with respect to some basis $v_1, \ldots, v_n$. Let $u_1, \ldots, u_n$ be obtained from $v_1, \ldots, v_n$ be the Gram-Schmidt process. Then for all $1 \leq i \leq n$ we have $\langle v_1, \ldots, v_i\rangle = \langle u_1, \ldots, u_i\rangle$, which means that the matrix with respect to $u_1, \ldots, u_n$ is also upper triangular. □

EXERCISE 10.5. *Show: for every $A \in M_{n,n}(\mathbb{C})$ there is $U \in U_n$ such that $U^{-1}AU$ is lower triangular.*

LEMMA 10.16. *Let $A \in M_{n,n}(\mathbb{C})$ be a normal operator. If A is upper triangular, then A is diagonal.*

PROOF. Write $A = (a_{ij})$. For $1 \leq i \leq n$ let c_i denote the (i,i) entry of AA^* and let d_i denote the (i,i) entry of A^*A. Since A is normal we have $c_i = d_i$. More explicitly, let r_i denote the ith row of A and let s_i denote the ith row of A^*. Then

$$c_i = \langle r_i, r_i\rangle = |a_{ii}|^2 + \sum_{i<j\leq n} |a_{ij}|^2.$$

$$d_i = \langle s_i, s_i\rangle = |a_{ii}|^2 + \sum_{1\leq j<i} |a_{ji}|^2.$$

Since $c_i = d_i$, we have

$$\sum_{i<j\leq n} |a_{ij}|^2 = \sum_{1\leq j<i} |a_{ji}|^2.$$

Taking $i = 1$, we get that $a_{1j} = 0$ for all $j \geq 2$. Now taking $i = 2$, we get $a_{2j} = 0$ for all $j \geq 3$. Proceeding inductively, we eventually find that $a_{ij} = 0$ for all $i < j$, so A is diagonal. □

THEOREM 10.17 (Spectral Theorem). *For $A \in M_{n,n}(\mathbb{C})$ the following are equivalent:*

(i) *A is unitarily diagonalizable: there is a unitary matrix $U \in U_n$ and a diagonal matrix $D \in M_{n,n}(\mathbb{C})$ such that $U^{-1}AU = D$.*
(ii) *There is an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors for A.*
(iii) *A is a normal operator.*

PROOF. (i) $\iff$ (ii): We have seen this before: over any scalar field F, for any matrix A and invertible matrix P, the matrix $P^{-1}AP$ is diagonal iff the columns of P give a basis of eigenvectors for A. Since a matrix is unitary iff its columns form an orthonormal basis for $\mathbb{C}^n$, the equivalence follows.
(ii) $\implies$ (iii): Let $v_1, \ldots, v_n$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors for A. By 10.14c) it is also an orthonormal basis of eigenvectors for A^*. So $U \coloneqq (v_1|\cdots|v_n) \in U_n$ and

$$D_1 \coloneqq U^{-1}AU,\ D_2 \coloneqq U^{-1}A^*U$$

are both diagonal. All diagonal matrices commute with each other, so

$$U^{-1}AA^*U = (U^{-1}AU)(U^{-1}A^*U) = D_1D_2 = D_2D_1 = (U^{-1}A^*U)(U^{-1}AU) = U^{-1}A^*AU,$$

and multiplying by U on the left and U^{-1} on the right, we get

$$AA^* = A^*A.$$

(iii) $\implies$ (i): Suppose $A \in M_{n,n}(\mathbb{C})$ is a normal operator. By Theorem 10.15, for *any* $A \in M_{n,n}(\mathbb{C})$ there is a unitary $U \in U_n$ such that $U^{-1}AU$ is upper triangular. By Lemma 10.13, $U^{-1}AU$ is again normal, so by Lemma 10.16, the matrix $U^{-1}AU$ is diagonal. □

COROLLARY 10.18. *For $A \in M_{n,n}(\mathbb{C})$, the following are equivalent:*

(i) *The matrix A is Hermitian: $A^* = A$.*
(ii) *The matrix A is unitarily diagonalizable and has real eigenvalues.*

PROOF. (i) $\implies$ (ii): Suppose A is Hermitian. Then $AA^* = A^*A$, so by Theorem 10.17 A is unitarily diagonalizable. By Proposition 10.9, the matrix A has real eigenvalues.
(ii) $\implies$ (i): For $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$, we denote by $D(\lambda_1, \ldots, \lambda_n)$ the diagonal matrix in $M_{n,n}(\mathbb{C})$ with diagonal entries $\lambda_1, \ldots, \lambda_n$. We have that

$$(D(\lambda_1, \ldots, \lambda_n))^* = D(\overline{\lambda_1}, \ldots, \overline{\lambda_n}).$$

Now suppose there is a unitary matrix U and real numbers $\lambda_1, \ldots, \lambda_n$ such that $U^{-1}AU = D(\lambda_1, \ldots, \lambda_n)$. Then

$$A = UD(\lambda_1, \ldots, \lambda_n)U^{-1},$$

so

$$A^* = (U^{-1})^*D(\lambda_1, \ldots, \lambda_n)^*U^* = UD(\lambda_1, \ldots, \lambda_n)U^{-1} = A.$$ □

10.4. Gershgorin's Theorem. Let $A = (a_{ij}) \in M_{n,n}(\mathbb{C})$. If A is diagonal, then the eigenvalues are precisely the diagonal entries $a_{1,1}, \ldots, a_{n,n}$. On the other hand, as A varies over all $n \times n$ complex matrices, then the eigenvalues should vary continuously with the coefficients of A. (This actually takes some work to pin down precisely, the main difficulty being that the eigenvalues form an unordered set with multiplicities. But it turns out to be true!) From this it stands to reason that if A is "almost" diagonal – i.e., is of the form $D + S$ where $D = (d_{i,i})$ is diagonal and $S = (s_{ij})$ with $|s_{ij}| \leq \epsilon$ for all i, j and some small $\epsilon > 0$, then the eigenvalues of A should be "close" to the diagonal entries $d_{1,1}, \ldots, d_{n,n}$ somehow in terms of ϵ and n. The following beautiful result of Gershgorin[14] accomplishes this.

[14]Semyon Aronovich Gershgorin, 1901-1933.

THEOREM 10.19 (Gershgorin Disk Theorem [**Ge31**]). *Let $A \in M_{n,n}(\mathbb{C})$. For $1 \leq i \leq n$, put*

$$R_i \coloneqq \sum_{j \neq i} |a_{ij}|,$$

let $B^{\bullet}_{R_i}(a_{i,i})$ be the closed disk with center $a_{i,i}$ and radius R_i, and let λ be an eigenvalue of A. Then

$$\lambda \in \bigcup_{i=1}^{n} B^{\bullet}_{R_i}(a_{i,i}).$$

We will give the (easy!) proof presently, but first a bit of discussion. First: A is diagonal iff $R_i = 0$ for all i iff each of the "Gershgorin disks" $B^{\bullet}_{R_i}(a_{i,i})$ just consists of the single point $a_{i,i}$, and we recover the statement that the eigenvalues of a diagonal matrix are its diagonal entries. Second: suppose that for some $\epsilon > 0$ we have that every off-diagonal entry has absolute value at most ϵ. Then we have $R_i \leq (n-1)\epsilon$ for all i and we get that every eigenvalue distance at most $(n-1)\epsilon$ from some diagonal entry. So this is indeed a quantitative version of our desired qualitative statement.

PROOF. Let λ be an eigenvalue of A, and let $v = (v_1, \ldots, v_n)$ be a corresponding eigenvector – recall that by definition v is nonzero. We may scale v such that there is $1 \leq i \leq n$ such that $v_i = 1$ and $|v_j| \leq 1$ for all $j \neq i$. The ith coordinate of the identity $Av = \lambda v$ is

$$\lambda = \lambda v_i = \sum_{j=1}^{n} a_{ij} v_j = a_{ii} + \sum_{j \neq i} a_{ij} v_j,$$

so

$$|\lambda - a_{ii}| = |\sum_{j \neq i} a_{ij} v_j| \leq \sum_{j \neq i} |a_{ij}||v_j| \sum_{j \neq i} |a_{ij}| = R_i. \qquad \square$$

A matrix $A \in M_{n,n}(\mathbb{C})$ is **strictly diagonally dominant** if for all $1 \leq i \leq n$ we have

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|.$$

COROLLARY 10.20 (Levy-Desplanques). *A strictly diagonally dominant matrix $A \in M_{n,n}(\mathbb{C})$ is nonsingular.*

EXERCISE 10.6. *Prove Corollary 10.20.*

We mention some refinements of the Gershgorin Disk Theorem.

THEOREM 10.21. *With the setup as in the Gershgorin Disk Theorem, suppose that for some $1 \leq k < n$ the union S_k of some k of the disks is disjoint from the union T_{n-k} of the remaining $n-k$ disks. Then precisely k of the eigenvalues (with multiplities counted) lie in S_k and the remaining $n-k$ lie in T_{n-k}. In particular if the disks are pairwise disjoint, each contains a single eigenvalue.*

THEOREM 10.22 (Marsli-Hall [**MH13**]). *Let $A \in M_{n,n}(\mathbb{C})$. If λ is an eigenvalue of A with geometric multiplicity k, then λ lies in at least k of the Gershgorin disks.*

CHAPTER 2

Theory

1. Linear independence and bases in infinite-dimensional vector spaces

Let F be a field, and let V be an F-vector space. We now fully engage with the possibility that V need not be finite-dimensional.

For any subset $S \subset V$, the **span** $\langle S \rangle$ is the set of all F-linear combinations from S: that is, we choose $\alpha_1, \ldots, \alpha_m \in F$ and $v_1, \ldots, v_m \in V$ and form $\alpha_1 v_1 + \ldots + \alpha_m v_m$. For any S, $\langle S \rangle$ is an F-linear subspace of V. If S is a subset of V and W is an F-linear subspace of V, we say that S **spans** W if $\langle S \rangle = W$.

A finite subset $S = \{v_1, \ldots, v_m\} \subset V$ is **linearly independent** if

$$\forall \alpha_1, \ldots, \alpha_m \in F,\ \alpha_1 v_1 + \ldots + \alpha_m v_m = 0 \implies \alpha_1 = \ldots = \alpha_m = 0.$$

An infinite subset S of V is linearly independent if every finite subset $T \subset S$ is linearly independent.

We say that taking $\alpha_1 = \ldots = \alpha_r = 0$ gives the **trivial linear combination** of $v_1, \ldots, v_m$ and any other choice of α_i's is **nontrivial**. Thus we can reword linear independence as: a nontrivial linear combination of $v_1, \ldots, v_m$ is never 0.

A **basis** for V is a linearly independent subset S of V that spans V.

The collection of linearly independent subsets of an F-vector space V is partially ordered under inclusion. By a **maximal** linearly independent subset of V we mean a maximal element S in this partially ordered set: in plainer terms, this means that S is linearly independent and is not properly contained in any other linearly independent subset of V.

THEOREM 1.1. *Let V be an F-vector space.*
a) A basis of V is precisely a maximal linearly independent subset of V.
b) Every linearly independent subset of V is contained in a basis of V.

PROOF. a) Let B be a basis for V. By definition, B is a linearly independent subst of V that spans V. It suffices to show that for any $v \in V \setminus B$, the set $B \cup \{v\}$ is linearly dependent. Indeed, since B spans V there are vectors $w_1, \ldots, w_m \in B$ and scalars $\alpha_1, \ldots, \alpha_m \in F$ such that

$$v = \alpha_1 w_1 + \ldots + \alpha_m w_m.$$

But then

$$\alpha_1 w_1 + \ldots + \alpha_m w_m - v = 0$$

is a nontrivial linear combination of $w_1, \ldots, w_m, v$ that is 0, so indeed $B \cup \{v\}$ is linearly dependent.

Now let S be a maximal linearly independent subset of V, and let $v \in V$. If $v \notin \langle S \rangle$ then $S \cup \{v\}$ remains linearly independent: indeed, suppose $\alpha_1, \ldots, \alpha_m, \beta \in F$ are such that

$$\alpha_1 v_1 + \ldots + \alpha_m v_m + \beta w = 0.$$

If $\beta \neq 0$ then

$$w = \frac{-\alpha_1}{\beta} v_1 + \ldots + \frac{-\alpha_m}{\beta} \in \langle S \rangle,$$

a contradiction. So $\beta = 0$, and then we must have $\alpha_1 = \ldots = \alpha_m = 0$ because V is lienearly independent. Thus the maximality of S as a linearly independent subset implies $\langle S \rangle = V$.
b) In view of part a) it is equivalent to show that every linearly independent subset S of V is contained in a maximal linearly independent subset. For this we will apply Zorn's Lemma in the partially ordered set $\mathcal{S}$ of linearly independent subsets T with $S \subset T \subset V$. Since $S \in \mathcal{S}$, certainly $\mathcal{S}$ is nonempty. Now let $\{T_i\}_{i \in I}$ be a chain of linearly independent subsets of V, each containing S. Then $T \coloneqq \bigcup_{i \in I} T_i$ contains each T_i and thus also contains S, so it suffices to show that T remains linearly independent. But by definition T is linearly independent iff each finite subset is linearly independent, and each finite subset of T lies in some T_i: indeed, if $v_1, \ldots, v_m \in T$ then there are $i_1, \ldots, i_m \in I$ such that $v_j \in i_j$ and then $v_1, \ldots, v_m \in T_{\max i_j}$. Thus we have showed that every chain in $\mathcal{S}$ has an upper bound in $\mathcal{S}$, so $\mathcal{S}$ has a maximal element. □

REMARK 1.2. *The appeal to Zorn's Lemma is* not *overkill. Zorn's Lemma is equivalent to the Axiom of Choice, and in a certain precise sense in which we will not get into here, the assumption that every vector space has a basis implies the Axiom of Choice and thus also Zorn's Lemma.*

THEOREM 1.3. *Let V be an F-vector space, and let S, T be subsets of V. If S is linearly independent and T spans V then $\#S \leq \#T$.*

PROOF. Case 1: Suppose that T is finite. Then the Steinitz Exchange Lemma holds in V: in Lemma 4.19 this is stated for a subspace V of $\mathbb{R}^N$ that admits a finite spanning set. However the same argument holds verbatim for a vector space over any field F that admits a finite spanning set, as we ask the reader to show. As recorded in Theorem 4.20, this immediately implies the result in this case.
Case 2: Suppose that T is infinite. If S is finite then indeed we have $\#S \leq \#T$, so we may assume that S is also infinite. By Theorem 1.1, there is a basis B of V such that B contains S, and since then we have $\#S \leq \#B$, it is enough to show that $\#B \leq \#T$. Write $B = \{v_i \mid i \in I\}$ and $T = \{w_j \mid j \in J\}$; seeking a contradiction we assume that $\#J < \#I$. For all $j \in J$, there is a finite subset $E_j \subset I$

$$w_j = \sum_{i \in E_j} \alpha_{i,j} v_i \text{ for some } \alpha_{i,j} \in F.$$

Sine J is infinite, we have $\#\bigcup_{j \in J} E_j \leq \#J < \#I$, so there is $i_\bullet \in I \setminus \bigcup_{j \in J} E_j$. Since T is a spanning set, we can write $v_{i_\bullet}$ as a linear combination of the vectors w_j, each of which in turn is a linear combination of vectors v_i with $i \neq i_\bullet$. This shows that $v_{i_\bullet} \in \langle v_i \mid i \neq i_\bullet \rangle$, contradicting the linear independence of B. □

We immediately deduce:

COROLLARY 1.4. *If B_1 and B_2 are two bases for an F-vector space V, then $\#B_1 = \#B_2$.*

Thus for any F-vector space V, we can define its **dimension** $\dim V$ as the cardinality of any basis. In this case $\dim V$ is a cardinal number, possibly infinite. For any cardinal number κ there is an F-vector space of dimension κ: as soon as we discuss direct sums (and it will be soon), we will see that we can just take the direct sum of κ copies of F itself.

2. Linear Maps

Let V and W be F-vector spaces. An **F-linear map** (or just a **linear map** if F is understood; or **linear transformation**) is a function $L : V \to W$ such that

(L1) For all $v_1, v_2 \in V$ we have $L(v_1 + v_2) = L(v_1) + L(v_2)$, and
(L2) For all $\alpha \in F$ and all $x \in V$ we have $L(\alpha v) = \alpha L(v)$.

EXERCISE 2.1. *Let V and W be F-vector spaces. State and prove an analogue of Proposition 3.5 for a function $f : V \to W$.*

An **isomorphism** of vector spaces is a linear map $L : V \to W$ for which there is a linear map $L' : W \to V$ that is a two-sided inverse to L: that is, we have

$$L' \circ L = 1_V,\ L \circ L' = 1_W.$$

LEMMA 2.1. *For a linear map $L : V \to W$, the following are equivalent:*
(i) The map L is an isomorphism of vector spaces.
(ii) The map L is a bijection.

PROOF. Any function $f : X \to Y$ admits an inverse function iff f is bijective, in which case the inverse is unique. So the condition that $L : V \to W$ is an isomorphism of vector spaces amounts to saying that L is bijective and the inverse function L^{-1} is again an F-linear map. This shows (i) $\implies$ (ii) and clarifies that to show (ii) $\implies$ (i) we need to show that the inverse function $L^{-1} : W \to V$ is a linear map. Let $w, w_1, w_2 \in W$ and $\alpha \in F$. We have

$$L(L^{-1}(w_1 + w_2)) = w_1 + w_2 = L(L^{-1}(w_1)) + L(L^{-1}(w_2)) = L(L^{-1}(w_1) + L^{-1}(w_2)).$$

Since L is an injection, this gives $L^{-1}(w_1 + w_2) = L^{-1}(w_1) + L^{-1}(w_2)$. Similarly we have

$$L(L^{-1}(\alpha w)) = \alpha w = \alpha L(L^{-1}(w)) = L(\alpha L^{-1}(w)),$$

and injectivity of L gives $L^{-1}(\alpha w) = \alpha L^{-1}(w)$. □

PROPOSITION 2.2. *Let $L : V \to W$ be a linear map, and let $S \subset V$.*

a) *Suppose that $L|_S$ is injective and that $L(S)$ is linearly independent in W. Then S is linearly independent in V.*
b) *Suppose S spans V. Then $L(S)$ spans $L(V)$ and thus spans W iff L is surjective.*
c) *If L is injective and S is linearly independent in V, then $L(S)$ is linearly independent in W. If L is not injective, then there is a linearly independent subset R of V such that $L(R)$ is linearly dependent in W.*
d) *If L is an isomorphism and S is a basis for V, then $L(S)$ is a basis for W.*

PROOF. a) Let $s_1, \ldots, s_n \in S$ and let $\alpha_1, \ldots, \alpha_n \in F$ be such that $\alpha_1 s_1 + \ldots + \alpha_n s_n = 0$. Then $0 = L(0) = L(\alpha_1 s_1 + \ldots + \alpha_n s_n) = \alpha_1 L(s_1) + \ldots + \alpha_n L(s_n)$. Because $L(S)$ is linearly independent, we have $\alpha_1 = \ldots = \alpha_n = 0$.
b) If $w \in L(V)$ then $w = L(v)$ for some $v \in V$. Since S spans V there are $s_1, \ldots, s_n \in S$ and $\alpha_1, \ldots, \alpha_n \in F$ such that $\alpha_1 s_1 + \ldots + \alpha_n s_n = v$ and thus $w = \alpha_1 L(s_1) + \ldots + \alpha_n L(s_n) \in \langle L(S)$. Since L is surjective iff $L(V) = W$, the second statement is clear.
c) Let $s_1, \ldots, s_n$ be distinct elements of S, and let $\alpha_1, \ldots, \alpha_n \in F$ be such that $\alpha_1 L(s_1) + \ldots + \alpha_n L(s_n) = 0$. Then

$$0 = L(\alpha_1 s_1 + \ldots + \alpha_n s_n),$$

so $\alpha_1 s_1 + \ldots + \alpha_n s_n = 0$ because L is injective. By linear independence of S we get $\alpha_1 = \ldots = \alpha_n = 0$.
d) If L is an isomorphism, then it follows from parts b) and c) that $L(S)$ is a basis for W. Suppose that L is not an isomorphism; then it fails to be either injective or surjective. □

EXAMPLE 2.3. *Let $\pi_1 : F^2 \to F$ by $(x, y) \mapsto x$. Let $b_1 = (1, 0)$ and $b_2 = (1, 1)$. Then $\mathcal{B} = \{b_1, b_2\}$ is a basis for F^2 and $\pi_1(\mathcal{B}) = \{1\}$ is a basis for F, even though π_1 is not injective. This shows why in part a) it was necessary to assume that $L|_S$ is injective and why in part d) we did not include the converse statement that if a linear map preserves bases then it must be an isomorphism.*

For F-vector spaces V and W, we write $\mathrm{Hom}_F(V, W)$ for the set of all F-linear maps $L : V \to W$. (When the scalar field F is understood, we may abbreviate this to $\mathrm{Hom}(V, W)$.) We may endow $\mathrm{Hom}_F(V, W)$ with the structure of an F-vector space, as follows:

$$L_1 + L_2 : v \mapsto L_1(v) + L_2(v).$$

$$\alpha L : v \mapsto \alpha L(v).$$

3. Direct products and Direct Sums

Let I be a set, and for each $i \in I$, let V_i be an F-vector space. We define two vector spaces that combine the V_i's into a single space. The **direct product** $\prod_{i\in I} V_i$ has underlying set the Cartesian product of the V_i's, i.e., an element of $\prod_{i\in I} V_i$ is an I-tuple $(v_i)_{i\in I}$ such that $v_i \in V_i$ for all $i \in I$. This gets the structure of an F-vector space "coordinatewise": that is, we define

$$\forall (v_i),\ (w_i) \in \prod_{i\in I} V_i,\ (v_i) + (w_i) \coloneqq (v_i + w_i),$$

$$\forall (v_i) \in \prod_{i\in I} V_i,\ \alpha \in F, \alpha(v_i) \coloneqq (\alpha v_i).$$

The **direct sum** $\bigoplus_{i\in I} V_i$ is a subspace of the direct product $\prod_{i\in I} V_i$: it consists of tuples (v_i) such that $\{i \in I \mid v_i \neq 0\}$ is finite. Then we have

$$\bigoplus_{i\in I} V_i = \prod_{i\in I} V_i \iff I \text{ is finite.}$$

For $i \in I$, we define the **inclusion map** $\iota_i : V_i \to \bigoplus_{i\in I} V_i$ by sending $v_i \in V$ to the element whose ith coordinate is v_i and all other coordinates are 0. This is an injective linear map, and thus

$$\iota_i : V_i \xrightarrow{\sim} \iota_i(V_i).$$

Moreover we have

$$\bigoplus_{i\in I} V_i = \langle \{\iota_i(V_i)\}_{i\in I} \rangle. \tag{19}$$

PROPOSITION 3.1. *Let I be a set, and let $\{V_i\}_{i\in I}$ be an indexed family of F-vector spaces.*

a) *For each $i \in I$, let $\{b_i j\}_{j\in J_i}$ be a basis of V_i, and let B_{ij} be the element of $\bigoplus_{i\in I}$ with ith component b_{ij} and with all other components 0. Then $\mathcal{B} \coloneqq \{B_{ij}\}_{i\in I,\ j\in J_i}$ is a basis for $\bigoplus_{i\in I} V_i$.*

b) *We have*

$$\dim \bigoplus_{i\in I} V_i = \sum_{i\in I} \dim V_i.$$

PROOF. a) For all $i \in I$, every element of V_i is a finite F-linear combination of the b_{ij}, so every element of $\iota_i(V_i)$ is a finite F-linear combination of the B_{ij}. By (19) every element of $\bigoplus_{i\in I} V_i$ is a finite F-linear combination of elements from the $\iota_i(V_i)$. Thus $\mathcal{B}$ spans $\bigoplus_{i\in I} V_i$. As for linear independence: if not, then some B_{ij} is a finite F-linear combination of the others, and without loss of generality we may assume that all the scalars in the linear combination are nonzero. But since B_{ij} is 0 except in the ith coordinate, B_{ij} could only be a linear combination of other $B_{ij'}$'s, which is not possible – since $\{B_{ij}\}_{j\in J_i}$ is the image of the basis $\{b_{ij}\}_{j\in J_i}$ under the injective linear map ι_i, it is a linearly independent set. Thus $\mathcal{B}$ is a basis for $\bigoplus_{i\in I} V_i$.
b) We have

$$\dim \bigoplus_{i\in I} V_i = \#\mathcal{B} = \#\bigcup_{i\in I} \{B_{ij}\}_{j\in J_i} = \sum_{i\in I} \#\{B_{ij}\}_{j\in J_i} = \sum_{i\in I} \#\{b_{ij}\}_{j\in J_i} = \sum_{i\in I} \dim V_i. \qquad \square$$

THEOREM 3.2. *(Universal Property of the Direct Sum) Let I be a set, and let $\{V_i\}_{i\in I}$ be an indexed family of F-vector spaces. For $i \in I$, let $\iota_i : V_i \to \bigoplus_{i\in I} V_i$ be the inclusion map as above. Let W be an F-vector space, and for each $i \in I$, let $L_i : V_i \to W$ be an F-linear map. Then there is a unique F-linear map $L : \bigoplus_{i\in I} V_i \to W$ such that for all $i \in I$ we have $L_i = L \circ \iota_i$.*

PROOF. First we will show that there is at most one such map L. Indeed, for all $i \in I$ the identity $L_i = L \circ \iota_i$ means that for all $v_i \in V_i$ we have $L(\iota_i(v_i)) = L_i(v_i)$. This means that the behavior of L is determined on $\iota_i(V_i)$. Since $\langle \{\iota_i(V_i)\}_{i\in I} \rangle = \bigoplus_{i\in I}$, the behavior of L is determined on a spanning set for $\bigoplus_{i\in I} V_i$ and thus it is determined on all of V_i.

The question remains whether there is such a linear map L! A little thought shows that we can take $L : \bigoplus_{i \in I} V_i \to W$ to be

$$(v_i) \mapsto \sum_{i \in I} L_i(v_i).$$

Because each (v_i) is zero except in finitely many coordinates, the above sum is actually a finite sum, so L is well-defined. That L is an F-linear map is almost immediate and left to the reader. Let $i \in I$ and $v_i \in V_i$. Then $\iota_i(v_i)$ is the element of $\bigoplus_{i \in I}$ that is v_i in the ith coordinate and 0 in all other coordinates, so $L(\iota_i(v_i)) = L_i(v_i)$, showing that $L_i = L \circ \iota_1$. □

Let I be a set, and let W be an F-vector space. For all $i \in I$, let $e_i \in \bigoplus_{i \in I} F$ be the element that is 1 in the ith coordinate and 0 in all other coordinates. Then $\{e_i\}_{i \in I}$ is a basis for $\bigoplus_{i \in I} F$: this is a special case of Proposition 3.1. Given an F-vector space W and a linear map $L : \bigoplus_{i \in I} F \to W$, evaluating on the basis gives us an I-tuple of elements of W: $\{L(e_i)\}_{i \in I}$. Conversely, given an I-tuple (w_i) of elements of W, there is a unique linear map $L : \bigoplus_{i \in I} F \to W$ such that $L(e_i) = w_i$ for all $i \in I$. This is clear in its own right, because all linear maps from a vector space V to a vector space W come from mapping all elements of a basis of V to arbitrary elements of W; it is also a special case of Theorem 3.2. In particular:

COROLLARY 3.3. *Let V be an F-vector space, and let $\mathcal{B}$ be a basis for V. Then there is a canonical isomorphism*

$$\bigoplus_{b \in \mathcal{B}} F \xrightarrow{\sim} V.$$

PROOF. For all $b \in \mathcal{B}$, let $e_b \in \bigoplus_{b \in \mathcal{B}} F$ be the element whose bth coordinate is 1 and all other coordinates are 0, so $\{e_b\}_{b \in \mathcal{B}}$ is a basis for $\bigoplus_{i \in I} F$. Therefore there is a unique F-linear map $\bigoplus_{b \in \mathcal{B}} F \to V$ that maps e_b to b, and since this maps a basis to a basis, it is an isomorphism. □

4. Quotients

Let V be an F-vector space, and let W be an F-linear subspace of V. For $v \in V$, we put

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

The subsets $v + W$ are called **cosets of W in V**.

LEMMA 4.1. *The set of cosets of W in V $\{v + W\}_{v \in V}$ forms a partition of V.*

PROOF. Every coset $v + W$ contains $v + 0 = v$ and thus is nonempty, and for all $v \in V$, v lies in $v + W$, so it remains to check that distinct cosets are pairwise disjoint. So, let $v_1, v_2 \in V$ be such that there is

$$v_3 \in (v_1 + W) \cap (v_2 + W).$$

That is, there are $w_1, w_2 \in W$ such that

$$v_1 + w_1 = v_3 = v_2 + w_2.$$

This means that $w' \coloneqq v_1 - v_2 \in W$ and thus for all $w \in W$ we have

$$v_1 + w = v_1 + (v_2 - v_1) + w = v_2 + (w - w') \in v_2 + W,$$

so

$$v_1 + W \subset v_2 + W,$$

and similarly

$$v_2 + w = v_2 + (v_1 - v_2) + w = v_1 + (w + w') \in v_1 + W,$$

so

$$v_2 + W \subset v_1 + W.$$

Thus $v_1 + W = v_w + W$, and we're done. □

The following result was essentially shown in the above proof, but it is so important that we ask the reader to check it carefully once again.

EXERCISE 4.1. *Let W be a subspace of a vector space V. For $v_1, v_2 \in V$, the following are equivalent:*
(i) We have $v_1 - v_2 \in W$.
(ii) We have $v_1 + W = v_2 + W$.

For any partition $\mathcal{P}$ on a set S we have an associated quotient map

$$q : S \to \mathcal{P}$$

in which we send $s \in S$ to $[s]$, the unique element of $\mathcal{P}$ that contains s. This map is surjective and has the "tautological" property that the preimage $q^{-1}([s]) = [s]$.

For our subspace $W \subset V$, we denote the the set of cosets of W in V by V/W, and as above let $q : V \to V/W$ be the quotient map:

$$q : v \mapsto v + W.$$

PROPOSITION 4.2. *There is exactly one way to make V/W into an F-vector space such that the quotient map $q : V \to V/W$ is F-linear.*

PROOF. First we want an addition operation on cosets with the property that for all $v_1, v_2 \in V$ we have

$$(v_1 + v_2) + W = q(v_1 + v_2) = q(v_1) + q(v_2) = (v_1 + W) + (v_2 + W).$$

But reading this equation from right to left tells us exactly how this operation must be defined:

$$(v_1 + W) + (v_2 + W) \coloneqq (v_1 + v_2) + W.$$

It remains to check that it is *well-defined*, i.e., that it did not depend on the representative elements v_1 and v_2 chosen. If we have $v_1', v_2' \in V$ such that $v_1 + W = v_1' + W$ and $v_2 + W = v_2' + W$, then $v_1' - v_1,\ v_2' - v_2 \in W$, so

$$(v_1' + v_2') + W = v_1 + v_2 + (v_1' - v_1 + v_2' - v_2) + W = v_1 + v_2 + W.$$

So addition of cosets is well-defined. That it makes V/W into a commutative group is almost immediate and left to the reader. To be sure, the identity is $0 + W = W$, and the inverse of $v + W$ is $-v + W$.

Next we want a scalar multiplication on cosets such that for all $\alpha \in F$ and all $v \in V$ we have

$$\alpha v + W q(\alpha v) = \alpha q(v) = \alpha(v + W).$$

Once again, reading right to left shows that we *must* have

$$\alpha(v + W) \coloneqq \alpha v + W.$$

This time we will leave to the reader the verification that this is a well-defined operation on cosets. We also leave the verification that V/W satisfies all the vector space axioms. □

THEOREM 4.3 (Fundamental Isomorphism Theorem). *Let $L : V \to W$ be an F-linear map, and let $q : V \to V/\operatorname{Ker} L$ be the quotient map.*
a) There is a unique linear map $\overline{L} : V/\operatorname{Ker} L \to W$ such that $L = \overline{L} \circ q$.
b) The map $\overline{L}$ is injective and has image $L(V)$, and thus

$$\overline{L} : V/\operatorname{Ker} L \to L(V)$$

is an isomorphism of vector spaces.

PROOF. a) Suppose there is such an $\overline{L}$ such that $L = \overline{L} \circ q$. Then for all $v \in V$ we have

$$L(v) = \overline{L}(q(v)) = \overline{L}(v + \operatorname{Ker} L).$$

Once again this tells us what $\overline{L}$ has to be: it must map the coset $v + \operatorname{Ker} L$ to $L(v)$. And once again we need to check that this map is well-defined; if $v + \operatorname{Ker} L = v' + \operatorname{Ker} L$, then $v - v' \in \operatorname{Ker} L$, so $L(v - v') = 0$, so $L(v) = L(v')$. Thus we do get a well-defined linear map $\overline{L} : V/\operatorname{Ker} L \to W$.
b) Suppose $v + \operatorname{Ker} L$ lies in the kernel of $\overline{L}$. Then

$$0 = \overline{L}(v + \operatorname{Ker}) = L(v),$$

so $v \in \operatorname{Ker} L$, so $v + \operatorname{Ker} L = \operatorname{Ker} L$ is the zero element of $V/\operatorname{Ker} L$. This shows that $\overline{L}$ is injective. It is clear from the definition of $\overline{L}$ that its image is $\operatorname{Image}(L)$, so $\overline{L} : V/\operatorname{Ker} L \to L(V)$ is an isomorphism. □

If W is a subspace of an F-vector space V, then we define the **codimension** $\operatorname{codim} W$ to be the dimension of the quotient space V/W.

Let $\mathcal{B}_1$ be a basis for W, extend it to a basis $\mathcal{B}$ for V, and put $\mathcal{B}_2 \coloneqq \mathcal{B} \setminus \mathcal{B}_1$. Let $q : V \to V/W$ be the quotient map. We claim $\{q(b) \mid b \mathcal{B}_2\}$ is a basis for V/W. First, every $v \in V$ is of the form $\sum_i \alpha_i b_i + \sum_j \beta_j b_j$ such that $\alpha_i, \beta_j \in F$ and all but finitely many are zero, so every element of V/W is of the form

$$(\sum_i \alpha_i b_i + \sum_j \beta_j b_j) + W = (\sum_j \beta_j b_j) + W = \sum_j \beta_j q(b_j),$$

so we get a spanning set. Now suppose that we have $\beta_j \in F$ such that $\sum_j \beta_j q(b_j) = 0$. Equivalently, we have $(\sum_j \beta_j b_j) + W = W$, so $\sum_j \beta_j b_j \in W$ and thus there are $\alpha_i \in F$ such that $\sum_i \alpha_i b_i + \sum_j \beta_j b_j = 0$. Since $\mathcal{B}_1 \cup \mathcal{B}_2 = \mathcal{B}$ is a basis, this gives $\alpha_i = \beta_j = 0$ for all i and j, so the $q(b_j)$'s are linearly independent. This shows:

$$\dim V = \dim W + \operatorname{codim} W.$$

This is an (important) special case of our next result, the Dimension Theorem.

5. The Dimension Theorem

THEOREM 5.1 (Dimension Theorem). *Let $L : V \to W$ be an F-linear map. Then we have*

$$\dim V = \dim \operatorname{Ker} L + \dim \operatorname{Image} L.$$

PROOF. Let $\mathcal{C} = \{c_j\}_{j \in J}$ be a basis for $\operatorname{Image} L$. For each $j \in J$, choose $b_j \in L^{-1}(c_j)$. Then $\mathcal{B} \coloneqq \{b_j\}_{j \in J}$ is a linearly independent subset of V by Proposition 2.2b). Let $Y \coloneqq \langle \mathcal{B} \rangle$. We claim that $V = \operatorname{Ker} L \bigoplus Y$. If so we're done, since then by Proposition 3.1b) we have

$$\dim V = \dim \operatorname{Ker} L + \dim Y = \dim \operatorname{Ker} L + \#\mathcal{B} = \dim \operatorname{Ker} L + \#\mathcal{C} = \dim \operatorname{Ker} L + \dim \operatorname{Image} L.$$

First, if $v \in \operatorname{Ker} L \cap Y$ then $L(v) = 0$. Write $v = \sum_{j \in J} \alpha_j b_j$ and then

$$0 = L(v) = \sum_{j \in J} \alpha_j L(b_j) = \sum_{j \in J} \alpha_j c_j.$$

But since the c_j's are linearly independent this means that $\alpha_j = 0$ for all $j \in J$ and thus $v = 0$. Thus $\operatorname{Ker} L \cap Y = \{0\}$. Next, let $v \in V$. Since the restriction of L to Y maps Y isomorphically onto $\operatorname{Image} V$, there is $y \in Y$ such that $L(v) = L(y)$. Thus $v = y + (v - y)$ with $y \in Y$ and $L(v - y) = 0$, so $v - y \in \operatorname{Ker} L$. □

For an F-linear map $L : V \to W$, we define the **cokernel** $\operatorname{Coker} L$ as $W/\operatorname{Image}(L)$.

EXERCISE 5.1. *Let $L : V \to W$ be an F-linear map, and let $U \subset V$ be an F-subspace.*

a) *Let $\iota : U \hookrightarrow V$ be the inclusion map. Show:*

$$\operatorname{codim} U = \operatorname{Coker} \iota.$$

b) *Show: L is surjective iff* $\operatorname{Coker} L = \{0\}$.

c) *Show:*

$$\dim \operatorname{Ker} L + \dim W = \dim V + \dim \mathit{commutativegroup} L. \tag{20}$$

d) *Deduce: if $V = W$ is finite-dimensional, then* $\dim \operatorname{Ker} L = \dim \operatorname{Coker} L$.

EXERCISE 5.2. *Let V and W be F-vector spaces. An F-linear map $L : V \to W$ is **Fredholm** if both* $\operatorname{Ker} L$ *and* $\operatorname{Coker} L$ *are finite-dimensional. For such a map we define the **index***

$$\operatorname{ind} L \coloneqq \dim \operatorname{Ker} L - \dim \operatorname{Coker} L.$$

a) *Show: if V and W are finite-dimensional, every $L \in \operatorname{Hom}(V, W)$ is Fredholm.*

b) *Show: if $V = W$ is finite-dimensional, then* $\operatorname{ind} L = 0$.

c) *(Hadwin* [**Ha94**]*) Suppose $L_1 : V \to W$ and $L_2 : W \to X$ are linear maps. Show:*

$$\dim \operatorname{Ker}(L_2 \circ L_1) + \dim \operatorname{Coker} L_1 + \dim \operatorname{Coker} L_2 = \dim \operatorname{Coker}(L_2 \circ L_1) + \dim \operatorname{Ker} L_1 + \dim \operatorname{Ker} L_2.$$

d) *Show: if $L_1 : V \to W$ and $L_2 : W \to X$ are Fredholm, then so is $L_2 \circ L_1$, and we have*

$$\operatorname{ind} L_2 \circ L_1 = \operatorname{ind} L_1 + \operatorname{ind} L_2.$$

5.1. Independent Families of Subspaces. Let V be an F-vector space, let I be a set, and let $\{W_i\}_{i\in I}$ be an indexed family of subspaces of V. For all $i \in I$, let $L_i : W_i \hookrightarrow V$ be the inclusion map. The universal property of direct sums gives us a map

$$L : \bigoplus_{i\in I} W_i \to V, (v_i) \mapsto \sum_{i\in I} L_i(v_i) = \sum_{i\in I} v_i.$$

The image of L is $W \coloneqq \langle W_i \rangle$, the subspace spanned by the W_i's. We say that the family of subspaces $\{W_i\}_{i\in I}$ is **independent** if L is an injection; equivalently, if it induces an isomorphism

$$\bigoplus_{i\in I} W_i \xrightarrow{\sim} \langle W_i \rangle.$$

THEOREM 5.2. *For a family $\{W_i\}_{i\in I}$ of subspaces of a vector space V, the following are equivalent:*
(i) The family $\{W_i\}_{i\in I}$ is independent.
(ii) For each $i \in I$, choose $w_i \in W_i$ such that $w_i = 0$ for all but finitely many $i \in I$. If $\sum_{i\in I} w_i = 0$, then $w_i = 0$ for all i.
(iii) If for all $i \in I$ we choose a nonzero $w_i \in W_i$, then $\{w_i\}_{i\in I}$ is a linearly independent subset of V.
(iv) For all $i \in I$ we have $W_i \cap \langle \{W_j\}_{j\in I\setminus\{i\}} \rangle = \{0\}$.
(v) For each $i \in I$, choose a linearly independent subset S_i of W_i. Then $S \coloneqq \bigcup_{i\in I} S_i$ is a linearly independent subset of V.

PROOF. (i) $\iff$ (ii): This is immediate, since the kernel of the natural linear map $L : \bigoplus_{i\in I} W_i \to V$ defined above is the set of (w_i) such that $\sum_{i\in I} w_i = 0$.
(ii) $\iff$ (iii): If (ii) fails then it gives a linear dependence relation among nonzero vectors $w_i \in W_i$. If (iii) fails then we can choose for each $i \in I$ a nonzero $w_i \in W_i$ such that $\{w_i\}_{i\in I}$ are linearly dependent, then there are distinct $i_0, i_1, \ldots, i_n \in I$ such that

$$w_i = \sum_{j=1}^{n} \alpha_j w_{i_j}$$

for $\alpha_j \in F$; since $w_i \neq 0$, after decreasing n if necessary we may assume that each α_j is nonzero. But then we have

$$w_i + \sum_{j=1}^{n} (-\alpha_j w_{i_j}) = 0,$$

and each $-\alpha_j w_{i_j}$ is a nonzero vector in W_{i_j}, so (ii) fails.
(iii) $\iff$ (iv): Condition (iii) fails iff there is some $i \in I$ and some nonzero $w_i \in W_i$ that lies in the span of the W_j's for $j \neq i$, which is precisely the negation of condition (iv) (ii) $\implies$ (v): Suppose that (v) fails. Then there is a finite nonempty subset $J = \{i_1, \ldots, i_m\} \subset I$ and for each $i \in J$ we have a linearly independent subset $w_{i_1,1}, \ldots, w_{i_1,n_{i_1}}$ such that their union is linearly dependent, i.e., there is a nontrivial linear dependence relation of the form

$$\sum_{j=1}^{m} \alpha_{j,1} w_{i_j,1} + \ldots + \alpha_{j,n_{i_j}} w_{i_j,n_{i_j}} = 0.$$

For $1 \leq j \leq m$ put

$$w_j \coloneqq \alpha_{j,1} w_{i_j,1} + \ldots + \alpha_{j,n_{i_j}} w_{i_j,n_{i_j}}.$$

Because for each fixed j the vectors $w_{i_j,1} \ldots, w_{i_j,n_{i_j}}$ are linearly independent, we have $w_j = 0$ iff $\alpha_{j,1} = \ldots = \alpha_{j,n_j} = 0$; by assumption, this does not hold for at least one j. Shrinking to a k-element

subset $K = \{i_1, \ldots, i_k\} \subset J$, we get that $w_{i_1} + \ldots + w_{i_k} = 0$ where each w_{i_j} is a nonzero vector in W_{i_j}: this contradicts (ii).
(v) $\implies$ (ii): Condition (ii) is the special case of condition (v) in which each S_i consists of a single, necessarily nonzero, element. □

COROLLARY 5.3. *Let W_1, W_2 be subspaces of an F-vector space V.*
a) We have

$$\dim(W_1 + W_2) + \dim W_1 \cap W_2 = \dim W_1 + \dim W_2.$$

b) The following are equivalent:
(i) The natural map $L : W_1 \oplus W_2 \to V$ given by $(w_1, w_2) \mapsto w_1 + w_2$ is an injection.
(ii) We have $W_1 \cap W_2 = \{0\}$.

PROOF. a) We have a surjective F-linear map $\Phi : W_1 \oplus W_2 \to W_1 + W_2$ given by $(w_1, w_2) \mapsto w_1 + w_2$. Its kernel is the set of all pairs $(w_1, w_2) \in W_1 \oplus W_2$ such that $w_1 + w_2 = 0$, or in other words, the set of $(w_1, -w_1) \in W_1 \oplus W_2$. It follows that the linear map

$$W_1 \cap W_2 \to \operatorname{Ker} L,\ w \mapsto (w, -w)$$

is an isomorphism, so

$$\dim \operatorname{Ker} L = \dim W_1 \cap W_2.$$

Using the Dimension Theorem and Proposition 3.1b) we get

$$\dim W_1 + \dim W_2 = \dim W_1 \oplus W_2 = \dim \operatorname{Ker} \Phi + \dim \operatorname{Image} \Phi = \dim(W_1 \cap W_2) + \dim(W_1 + W_2).$$

b) This is immediate from part a) and also from the equivalence of (i) and (iv) in Theorem 5.2. □

6. Dual Spaces

For a set S, let F^S be the set of all functions $f : S \to F$. For $f, g \in F^S$ and $\alpha \in F$, we put

$$\alpha f + g : s \in S \mapsto \alpha f(s) + g(s).$$

This makes F^S into an F-vector space.

EXERCISE 6.1. *a) For $s \in S$, let δ_s be the function $t \mapsto \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases}$. Show: $\Delta := \{\delta_s \mid s \in S\}$ is a linearly independent subset of F^S.*
b) Show: Δ is a basis for F^S iff S is finite.

Let V be an F-vector space. Let

$$V^\vee := \{f \in F^V \mid f \text{ is linear}\}.$$

EXERCISE 6.2. *Show: $V^\vee$ is a subspace of F^V.*

We call $V^\vee$ the **dual space** of V. The elements $f : V \to F$ of V are called **linear functionals** on V.

For $\ell \in V^\vee$ and $v \in V$, we put $\langle \ell, v \rangle := \ell(v)$.

Let $\mathcal{B} = \{e_i\}_{i \in I}$ be a basis of V. We define a subset $\mathcal{B}^\vee = \{e_i^\vee\}_{i \in I}$ of $V^\vee$ as follows:

$$\forall i \in I, \langle e_i^\vee, e_j \rangle := \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}.$$

Since as usual, every map on a basis extends uniquely to a linear map on the space, the above formula serves to uniquely define $e_i^\vee$. We call $\mathcal{B}^\vee$ the **dual basis** to $\mathcal{B}$. However, please read on!

PROPOSITION 6.1. *Let V be a vector space with basis $\mathcal{B} = \{e_i\}_{i \in I}$, and let $\mathcal{B}^\vee = \{e_i^\vee\}_{i \in I}$ be the dual basis of $V^\vee$.*

a) *The set $\mathcal{B}^\vee$ is linearly independent in $V^\vee$.*
b) *The set $\mathcal{B}^\vee$ is a basis for $V^\vee$ iff $\dim V$ is finite.*

PROOF. a) Let $\alpha_1, \ldots, \alpha_n \in F$, let $i_1, \ldots, i_n \in I$ and suppose that $\alpha_1 e_{i_1}^\vee + \ldots + \alpha_n e_{i_n}^\vee = 0$. That is, for all $v \in V$ we have

$$\alpha_1 \langle e_{i_1}^\vee, v \rangle + \ldots + \alpha_n \langle e_{i_n}^\vee, v \rangle = 0.$$

For $1 \leq j \leq n$, taking $v = e_j$ gives $\alpha_j = 0$.
b) Suppose $\dim V$ is finite, in which case we may take $I = \{1, \ldots, n\}$. Let $\ell \in V^\vee$ and let $v = \alpha_1 e_1 + \ldots + \alpha_n e_n \in V$. We claim that for $\ell \in V^\vee$ we have

$$\ell = \ell(e_1) e_1^\vee + \ldots + \ell(e_n) e_n^\vee.$$

Indeed, for all $1 \leq i \leq n$, both ℓ and $\ell(e_1) e_1^\vee + \ldots + \ell(e_n) e_n^\vee$ upon evaluation at e_i give $\ell(e_i)$, so the maps agree. Thus $\mathcal{B}^\vee$ is a basis for $V^\vee$.

Suppose $\dim V$ is infinite; equivalently, I is infinite. Let $W := \langle \mathcal{B}^\vee \rangle$. For all $\ell \in W$, we have $\ell(e_i) = 0$ for all but finitely many $i \in I$, so the linear functional λ that maps each e_i to 1 does not lie in W. □

So **beware**: when $\dim V$ is infinite, the dual basis $\mathcal{B}^\vee$ is not actually a basis for the dual space $V^\vee$!

EXERCISE 6.3. *Let V be a vector space, and put $V^\vee := (V^\vee)^\vee$, the "second dual space" of V. Let*

$$\iota : V \to V^{\vee\vee}, \ \iota(v) : \ell \mapsto \langle \ell, v \rangle.$$

a) Show: ι is an injective linear map.
*b) We say that V is **reflexive** if ι is an isomorphism. Show: V is reflexive iff it is finite-dimensional.*

Thus the notion of a reflexive vector space is just a fancy way of saying it is finite-dimensional: we must admit that this is not so interesting. However the notion of reflexivity carries over to other contexts in a richer way. We mention two examples in passing:
(i) If R is a commutative ring and M is an R-module, then we can define $M^\vee = \mathrm{Hom}_R(M, R)$ to be the collection of R-module maps $f : M \to R$. In this context the map ι is still defined but need not be injective; again, one says M is reflexive when ι is an isomorphism. When e.g. $R = \mathbb{Z}$, being finitely generated is neither necessary nor sufficient for reflexivity.
(ii) If V is a Banach space over $\mathbb{R}$ (say), one takes $V^* = \mathrm{Hom}_c(V, \mathbb{R})$ to be the *continuous* linear functionals. The set V^* can naturally be given the structure of a Banach space. In this context $\iota : V \to V^{**}$ is always an isometric embedding of Banach spaces but need not be an isomorphism: a Banach space is called reflexive when ι is an isomorphism. This is a very important class of Banach spaces, including for instance the L^p-spaces for $p \in (1, \infty)$.

PROPOSITION 6.2. *Let $\mathcal{B} = \{e_i \mid i \in I\}$ be a basis for the vector space V. Then there is an isomorphism of vector spaces*

$$\Phi : V^\vee \to F^I, \ \ell \mapsto (\langle \ell, e_i \rangle).$$

PROOF. For any $v \in V$, we get a map

$$\langle \cdot, v \rangle : V^\vee \to F, \ \lambda \in V^\vee \mapsto \langle \lambda, v \rangle,$$

and one checks immediately that this is F-linear. In particular this holds for $v = e_i$, and the map Φ comes from this family of maps via the universal property of the direct product. (This is fancier than necessary: no problem to check directly that Φ is F-linear.) The bijectivity of Φ is really nothing else than the statement that a linear map is uniquely and freely determined by what it does to a basis. □

Let $f : V \to W$ be an F-linear map. We define an induced F-linear map $f^\vee : W^\vee \to V^\vee$, as follows:

$$\ell \in W^\vee = \mathrm{Hom}_F(W, F) \mapsto \ell \circ f \in \mathrm{Hom}_F(V, F) = V^\vee.$$

It is natural to wonder whether and how properties of f induce properties of $f^\vee$: for instance, if f is injective, must $f^\vee$ also be? Well, it is always reasonable to guess, but as guesses go this one is not the greatest: for instance if V and W are finite-dimensional with $\dim V < \dim W$ then there is an injective linear map $f : V \to W$ but no injective linear map from $W^\vee$ to $V^\vee$. Similarly, if V and W are finite-dimensional with $\dim V > \dim W$ then there is a surjective linear map $f : V \to W$ but not

surjective linear map from $W^\vee$ to $V^\vee$. Reflecting a bit more about the dimension obstructions one might be led to guess the following result.

PROPOSITION 6.3. *Let $f : V \to W$ be a linear map, with dual map $f^\vee : W^\vee \to V^\vee$.*
a) If f is injective, then $f^\vee$ is surjective.
b) If f is surjective, then $f^\vee$ is injective.
c) If f is an isomorphism, then so is $f^\vee$.

PROOF. a) Since f is injective, it defines an isomorphism $f : V \to f(V)$; let $f^{-1} : f(V) \to V$ be the inverse map. Now let $\ell_V \in \mathrm{Hom}_F(V, F)$; we want to find $\ell_W \in \mathrm{Hom}_F(W, F)$ such that $\ell_W \circ f = \ell_V$. We claim that ℓ_W restricted to $f(V)$ is uniquely defined: indeed, for all $v \in V$ we must have

$$\ell_W(f(v)) = \ell_V(v);$$

since for all $v \in V$ there is a unique $w \in f(V)$ such that $v = f^{-1}(w)$, we have

$$\ell_W(w) = \ell_V(f^{-1}(w)).$$

This defines ℓ_W uniquely on $f(V)$. We can then extend ℓ_W to a linear map defined on W, in general in many ways, and any such extension will satisfy $\ell_W \circ f = \ell_V$.
b) Suppose f is surjective, and let $\ell \in \mathrm{Hom}_F(W, F)$ be such that $f^\vee(\ell) = \ell \circ f = 0$. That is, for all $v \in V$ we have $\ell(f(v)) = 0$. But since f is surjective this implies that $\ell(w) = 0$ for all $w \in W$, and thus $\ell = 0$.
c) This follows immediately from parts a) and b). □

7. Dimensions of Direct Products and Dual Spaces

In this section we follow a MathOverflow discussion [**MO**] and use in particular the answers of Pierre-Yves Gaillard and Todd Trimble.

PROPOSITION 7.1. *Let F be a field, and let V be a nonzero F-vector space.*

a) *If V is finite (as a set), then $F \cong \mathbb{F}_q$ is a finite field, V is finite-dimensional and $\#V = q^{\dim V}$.*
b) *If V is infinite (as a set), then we have $\#V = \max(\dim V, \#F)$.*

PROOF. a) If V is a nontrivial vector space over an infinite field F, then let $v \in V \setminus \{0\}$. Then $F \hookrightarrow V$, $\alpha \mapsto \alpha v$ is an injection, so V is infinite. So if V is finite then F must be finite and thus have q elements for some prime power q. If $\dim V$ were infinite then for all $d \in \mathbb{Z}^+$ there would be a subspace isomorphic to $\mathbb{F}_q^d$, so $\#V \geq q^d$ for all d and thus V is infinite. So V must be finite-dimensional and thus isomorphic to $\mathbb{F}_q^{\dim V}$.
b) If G is a group and $S \subset G$ is an infinite generating set, then $\#S = \#G$. Indeed, since S is a subset of G we have $\#S \leq \#G$. Without changing $\#S$ we may assume that S is closed under inversion, and then there is a natural map $\coprod_{n=1}^\infty S^n \to G$ obtained by $(s_1, \ldots, s_n) \mapsto s_1 \cdots s_n$, and this map is surjective since S generates G. Since S is infinite we have $\#\coprod_{n=1}^\infty S^n = \#S$, and thus $\#S \geq \#G$.

Certainly we have $\dim V \leq \#V$, and as above we have $\#F \leq \#V$. Let $\mathcal{B}$ be a basis for V. Then $S := \{\alpha b \mid \alpha \in F,\ b \in \mathcal{B}\}$ generates V as an additive group, and S has cardinality at most $\dim V \cdot \#F = \max(\dim V, \#F)$, so $\#V \leq \max(\dim V, \#F)$. □

Following Gaillard, we will say that an F-vector space V is **large** if $\dim V \geq \max(\#F, \aleph_0)$. Proposition implies that V is large iff $\dim V = \#V$.

THEOREM 7.2. *Let I be a nonempty set, for all $i \in I$, let V_i be a nonzero F-vector space, and put $V := \prod_{i \in I} V_i$.*

a) *If I is finite, then $V = \bigoplus_{i \in I} V_i$, so $\dim V = \sum_{i \in I} \dim V_i$.*
b) *If I is infinite, then V is large.*
c) *In particular we have $\dim F^I = (\#F)^{\#I}$.*

PROOF. a) This is Proposition 3.1a).
b) STEP 1: Let $\mathbb{E} := F^{\aleph_0} = \prod_{i=1}^{\infty} F$. We show that V is large. Since the subspace $\bigoplus_{i=1}^{\infty} F$ of V has dimension $\aleph_0$, it suffices to show that $\dim V \geq \#F$. Seeking a contradiction, we suppose that there is an F-basis $\mathcal{B}$ of $\mathbb{E}$ with $\#B < \#F$. This implies that F is uncountably infinite. Let F_0 be the prime subfield of F. Let F_1 be the subfield of F obtained by adjoining to F_0 the coordinates in F of every $b \in B$. We have $\#F_1 \leq \#B < \#F$, so the field F is a large F_1-vector space: we have $\dim_{F_1} F = \#F > \aleph_0$, and therefore there is $x \in \mathbb{E}$ whose coordinates are F_1-linearly independent. Let $b_1, \ldots, b_n \in \mathcal{B}$ be such that x lies in the F-span of $b_1, \ldots, b_N$. Consider the matrix $M = (m_{ij}) \in M_{n,n+1}(F_1)$ with m_{ij} equal to the jth coordinate of b_i. The nullspace of M must be nontrivial: let $0 \neq \lambda = (\lambda_1, \ldots, \lambda_{n+1})^T \in F^{n+1}$ be such that $M\lambda = 0$: thus

$$\forall 1 \leq i \leq n,\ \sum_{j=1}^{n+1} \lambda_j m_{i,j} = 0.$$

It then follows that

$$\sum_{j=1}^{n+1} \lambda_j x_j = 0,$$

contradicting the choice of x.
STEP 2: There is an injective F-linear map $\iota : \mathbb{E} \to V$, so $\dim V \geq \dim \mathbb{E} \geq \max(\aleph_0, \#F)$ and thus V is also large.
c) Since $\#F^I = (\#F)^{\#I}$, this is immediate from part b). □

COROLLARY 7.3. *Let V be a vector space over a field F. If $\dim V$ is infinite, then $\dim V < \dim V^\vee$.*

PROOF. By Proposition 6.2 we have $V^\vee \cong F^{\dim V}$, so by Theorem 7.2 and Cantor's Theorem [**Cl-STI**, Thm. 12] we have

$$\dim V^\vee = (\#F)^{\dim V} \geq 2^{\dim V} > \dim V.$$ □

Corollary 7.3 immediately implies that no infinite-dimensional vector space is reflexive, which is the more challenging direction of Exercise 6.3b). (It is possible to solve the exercise without this result!)

EXERCISE 7.1. *Let I be an infinite index set; for all $i \in I$ let V_i be a nonzero F-vector space. Put*

$$\mu := \max(\aleph_0, \#F),\ \alpha := \#\{i \in I \mid \dim V_i < \mu\}.$$

Show:

$$\dim \prod_{i \in I} V_i = (\#F)^\alpha \prod_{\dim V_i \geq \mu} \dim V_i.$$

8. Change of Basis

Let V be an m-dimensional vector space, let W be an n-dimensional vector space, and let $L : V \to W$ be a linear transformation. Then L is a rather abstract object, but we can study it quite concretely by means of bases. Namely, let $\mathcal{B}_V = v_1, \ldots, v_m$ be an ordered basis for V, and let $\mathcal{B}_W = w_1, \ldots, w_n$ be an ordered basis for W. Then there are unique scalars $a_{ij} \in F$ such that

$$L(v_i) = \sum_{j=1}^{n} a_{ij} w_j.$$

In what follows, we let $e_1, \ldots, e_n$ be the standard basis of F^n.

THEOREM 8.1. a) *Let $A \in M_{n,m}(F)$.*
The map

$$A\bullet : (x_1, \ldots, x_m) \mapsto A(x_1, \ldots, x_m)^T$$

is a linear map from F^m to F^n.

b) *Let $L : F^m \to F^n$ be a linear map. There is a unique $A = (a_{ij}) \in M_{m,n}(F)$ such that for all $1 \leq i \leq m$ we have*

$$L(e_i) = \sum_{j=1}^{n} a_{ij} e_j.$$

Then $L = A^T \bullet$.

PROOF. a) This is left to the reader.
b) The existence and uniqueness of A is just because $e_1, \ldots, e_n$ is a basis of F^n. By part a), $A^T \bullet : F^m \to F^n$ is a linear map. For all $1 \leq i \leq m$ we have

$$A^T e_i = (a_{i,1}, \ldots, a_{i,n})^T = \sum_{j=1}^{n} a_{i,j} e_j = T(e_i).$$

Since a linear map is determined by its action on a basis, we have $A^T \bullet = L$. □

The transpose in the above result is a bit annoying.[1] The easiest way to fix it turns out to be just to use a less-immediately-appealing indexing convention: namely, instead of writing

$$L(v_i) = \sum_{j=1}^{n} a_{ij} w_j,$$

if we write

$$L(v_i) = \sum_{j=1}^{n} a_{ji} w_j,$$

then the transpose goes away. A better way to say this is that we write down the rectangular array of scalars whose ith *column* expresses $L(v_i)$ in terms of the w_j's.

We return to the case of $L : V \to W$ as above. To L we associated a rectangular array of numbers, so we should sharpen it up in terms of matrix multiplication. This is done as follows: the ordered basis $v_1, \ldots, v_m$ of V gives us an isomorphism

$$\mathcal{B}_V : F^m \to V,$$

and the ordered basis $w_1, \ldots, w_n$ of W gives us an isomorphism

$$\mathcal{B}_W : F^n \to W.$$

A key idea: composing with an isomorphism (or its inverse!) does not really change anything; it just induces a "relabelling." In this case, we may replace $L : V \to W$ by

$$L_{\mathcal{B}_V, \mathcal{B}_W} : F^m \overset{\mathcal{B}_V}{\to} V \overset{L}{\to} W \overset{\mathcal{B}_W^{-1}}{\to} F^n.$$

By Theorem 8.1a) we see that $L_{\mathcal{B}_V, \mathcal{B}_W}$ is given by multiplication by a matrix $A \in M_{n,m}(F)$, and it is a minor variant of Theorem 8.1b) to see that the (i,j) entry of A is a_{ji}.

Now the plot thickens: we can change the basis on V and/or on W. Let $v_1', \ldots, v_m'$ be another ordered basis for V and let $w_1', \ldots, w_n'$ be another ordered basis for W. Then the linear map

$$L_{\mathcal{B}_V', \mathcal{B}_W'} : F^m \overset{\mathcal{B}_V'}{\to} V \overset{L}{\to} W \overset{\mathcal{B}_W'^{-1}}{\to} F^n$$

is given by $A' \bullet$, where $A' \in M_{n,m}(F)$ is the matrix whose (i,j)th entry is a_{ji}', where for all $1 \leq i \leq m$ and $1 \leq j \leq n$ we have

$$L(v_i') = \sum_{j=1}^{n} a_{ji}' w_j'.$$

[1] It ruined one of my lectures: Math 7900, October 7, 2019.

The question is now: what is the relationship between A and A'?

The evident thing to do is to write the new bases in terms of the old bases, so let's start there: for $1 \leq i, j \leq m$ there are unique $\alpha_{i,j} \in F$ such that for all $1 \leq i \leq m$ we have

$$v_i' = \sum_{j=1}^{m} \alpha_{ji} v_j.$$

Put $P \coloneqq (\alpha_{ij}) \in \mathrm{GL}_m(F)$. Similarly, for $1 \leq i, j \leq n$ there are unique $\beta_{i,j} \in F$ such that for all $1 \leq i \leq n$ we have

$$w_i' = \sum_{j=1}^{n} \beta_{ji} w_j.$$

Put $Q \coloneqq (\beta_{ij}) \in \mathrm{GL}_n(F)$. So, more precisely, we want to describe A' in terms of A, P and Q.

THEOREM 8.2. *With notation as above we have*

$$A' = Q^{-1}AP. \tag{21}$$

PROOF. Step 1: We claim that as linear transformations we have

$$\mathcal{B}_V' = \mathcal{B}_V \circ (P\bullet) \tag{22}$$

and

$$\mathcal{B}_W'^{-1} = (Q^{-1}\bullet) \circ \mathcal{B}_W^{-1}. \tag{23}$$

To see (22): for $1 \leq i \leq m$ we have

$$\mathcal{B}_V(Pe_i) = \mathcal{B}_V(\sum_{j=1}^{m} \alpha_{ji} e_j) = \sum_{j=1}^{m} \alpha_{ji} v_j = v_i' = \mathcal{B}_V'(e_i).$$

As for (23): taking inverses gives $\mathcal{B}_W' = \mathcal{B}_W \circ (Q\bullet)$, and this is proved identically to (22).
Step 2: We have

$$A'\bullet = \mathcal{B}_W'^{-1} \circ L \circ \mathcal{B}_V' = \big((Q^{-1}\bullet) \circ \mathcal{B}_W\big) \circ L \circ (\mathcal{B}_V \circ (P\bullet))$$

$$= Q^{-1} \bullet \circ (\mathcal{B}_W \circ L \circ \mathcal{B}_V) \circ P\bullet = (Q^{-1}\bullet) \circ (A\bullet) \circ (P\bullet) = Q^{-1}AP\bullet,$$

and thus

$$A' = Q^{-1}AP. \qquad \square$$

Reflecting on the above proof one sees that most of it is not really about the linear transformation at all! Rather it concerns the phenomenon on having two bases $\mathcal{B} = \{b_i \mid i \in I\}$ and $\mathcal{B}' = \{b_i' \mid i \in I\}$ for the same vector space V. (Since any two bases must have the same cardinality, it is no loss of generality to index them by the same set.) This gives two different isomorphisms

$$\mathcal{B}, \mathcal{B}' : \bigoplus_{i \in I} F \to V.$$

(Here we are using the same notation for the basis as the isomorphism. This seems justified, since the basis and the isomorphism correspond uniquely to each other!) There is then a unique automorphism $P_{\mathcal{B}',\mathcal{B}}$ of $\bigoplus_{i\in I} F$ such that

$$\mathcal{B}' = \mathcal{B} \circ P_{\mathcal{B}',\mathcal{B}};$$

indeeed we must have $P_{\mathcal{B}',\mathcal{B}} = \mathcal{B}^{-1} \circ \mathcal{B}'$. Assuming now that I is finite and replacing $\bigoplus_{i\in I}$ by F^n, the map $P_{\mathcal{B}',\mathcal{B}}$ is given by an invertible $n \times n$ matrix. We already figured out what this matrix is: its ith column expresses the ith vector b_i' of $\mathcal{B}'$ as a linear combination of the vectors $b_1, \ldots, b_n$. We think of this matrix as "transforming us from $\mathcal{B}'$-coordinates to $\mathcal{B}$-coordinates."

EXERCISE 8.1. *Let V be an finite-dimensional F-vector space, and let $\mathcal{B}_1, \mathcal{B}_2$, $\mathcal{B}_3$ be bases of V.*
a) Show:

$$P_{\mathcal{B}_2,\mathcal{B}_1} = P^{-1}_{\mathcal{B}_1,\mathcal{B}_2}.$$

b) Show:

$$P_{\mathcal{B}_3,\mathcal{B}_1} = P_{\mathcal{B}_2,\mathcal{B}_1} P_{\mathcal{B}_3,\mathcal{B}_2}.$$

EXERCISE 8.2. *Suppose $V = F^n$ with its canonical basis $\mathcal{E} = e_1, \ldots, e_n$.*
a) Let $\mathcal{B}_1 = v_1, \ldots, v_n$ be a basis of V. Show that $P_{\mathcal{B},\mathcal{E}}$ is the matrix whose ith column is the vector v_i.
b) Let $\mathcal{B}_2 = w_1, \ldots, w_n$ be another basis of V. Show:

$$P_{\mathcal{B}_2,\mathcal{B}_1} = [v_1 \mid \cdots \mid v_n]^{-1}[w_1 \mid \cdots \mid w_n].$$

Bibliography

[A] S. Axler, *Linear algebra done right*. Second edition. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1997.

[Ac55] R.H. Ackerson, *A note on vector spaces.* Amer. Math. Monthly 62 (1955), 721–722.

[Ax95] S. Axler, *Down with determinants!* Amer. Math. Monthly 102 (1995), no. 2, 139–154.

[BB03] K. Boulabiar and G. Buskes, *After the determinants are down: a criterion for invertibility*. Amer. Math. Monthly 110 (2003), no. 8, 737–741.

[Be18] C. Bernhardt, *Powers of positive matrices*. Math. Mag. 91 (2018), 218–227.

[Bu73] M.D. Burrow, *The minimal polynomial of a linear transformation*. Amer. Math. Monthly 80 (1973), 1129–1131.

[BW06] T. Brady and C. Watt, *On products of Euclidean reflections.* Amer. Math. Monthly 113 (2006), 826–829.

[Cl12] P.L. Clark, *Covering numbers in linear algebra*. Amer. Math. Monthly 119 (2012), 65–67.

[Cl-I] P.L. Clark, *Lecture notes on mathematical induction*. `http://www.math.uga.edu/∼pete/3200induction.pdf`

[Cl-IS] P.L. Clark, *Linear algebra: invariant subspaces*. `http://math.uga.edu/~pete/invariant_subspaces.pdf`

[Cl-STI] P.L. Clark, *Set Theory: Part I*. `http://math.uga.edu/~pete/settheorypart1.pdf`

[De03] H. Derksen, *The fundamental theorem of algebra and linear algebra*. Amer. Math. Monthly 110 (2003), 620–623.

[FHM13] M. Fiedler, F.J. Hall and R. Marsli, *Geršgorin discs revisited*. Linear Algebra Appl. 438 (2013), 598–603.

[Ge31] S. Gerschgorin, *Über die Abgrenzung der Eigenwerte einer Matrix*. Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk 6 (1931), 749–754.

[Gi91] G.T. Gilbert, *Positive definite matrices and Sylvester's criterion*. Amer. Math. Monthly 98 (1991), 44–46.

[Ha94] D. Hadwin, *An algebraic version of the multiplication property of the Fredholm index.* Linear Algebra Appl. 208/209 (1994), 229–230.

[HC] P.L. Clark, *Honors Calculus./..*

[Hu02] C. Huneke, *The Friendship Theorem*. Amer. Math. Monthly 109 (2002), 192–194.

[Ja73] H.G. Jacob, *Another proof of the rational decomposition theorem*. Amer. Math. Monthly 80 (1973), 1131–1134.

[La14] K. Lange, *Hadamard's determinant inequality*. Amer. Math. Monthly 121 (2014), 258–259.

[MH13] R. Marsli and F.J. Hall, *Geometric multiplicities and Geršgorin discs*. Amer. Math. Monthly 120 (2013), 452–455.

[MO] `https://mathoverflow.net/questions/49551/dimension-of-infinite-product-of-vector-spaces`

[Mo14] J.T. Moore, *A Zorn's lemma proof of the dimension theorem for vector spaces.* Amer. Math. Monthly 121 (2014), 260–262.

[MP12] C.D. Martin and M.A. Porter, *The extraordinary SVD*. Amer. Math. Monthly 119 (2012), 838–851.

[MSV93] D. Mornhinweg, D.B. Shapiro, and K.G. Valente, *The principal axis theorem over arbitrary fields.* Amer. Math. Monthly 100 (1993), 749–754.

[PP07] B. Palais and R. Palais, *Euler's fixed point theorem: The axis of a rotation*. J. Fixed Point Theory Appl. 2 (2007), 215–220.

[SA] T. Shifrin and M. Adams, *Linear Algebra: A Geometric Approach*.

[Sa87] D. Sarason, *The multiplication theorem for Fredholm operators.* Amer. Math. Monthly 94 (1987), 68–70.

[SdeS09] A. Sandovici and H. de Snoo, *An index formula for the product of linear relations.* Linear Algebra Appl. 431 (2009), 2160–2171.

[St93] G. Strang, *The fundamental theorem of linear algebra*. Amer. Math. Monthly 100 (1993), 848–855.

[Wa01] W.P. Wardlaw, *A Generalized General Associative Law*. Mathematics Magazine 74 (2001), 230-233.

[Wa80] W. Watkins, *Similarity of matrices*. Amer. Math. Monthly 87 (1980), 300.

[Y84] T. Yuster, *The Reduced Row Echelon Form of a Matrix is Unique: A Simple Proof*. Math. Mag. 57 (1984), 93–94.