Please send any questions/comments/corrections to hhao@berkeley.edu.

Note: this study guide is a *very* condensed version of the course material (which will still end up being a lot). In particular, most statements won't be proved, and I won't spend space on giving many numerical examples. The goal is rather to give intuition to some of the more difficult concepts, as well as to show how all of the concepts are interconnected (as I think the course text manages to make linear algebra seem like a much more disjointed subject than it actually is). Therefore, most of the below material is in a different order than presented in the text, and I also don't guarantee that 100% of the material on the midterm will be discussed below.

1 Linear Systems and Vector Spaces

1.1 Basics of Linear Systems

Definition 1.1.1. A *linear system* is a set of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \ldots &= \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned} \tag{1.1.1}$$

that are to be solved simultaneously (in the x_i). Here, the a_{ij} are known *coefficients*, and the b_i are also known. We say that the system is *consistent* if it has at least one solution, and *inconsistent* otherwise. We say that it is *homogeneous* if all the b_i are 0, and *inhomogeneous* otherwise.

The standard method of solving a linear system is *Gaussian elimination*, or *row reduction*. To do this, we gather the a_{ij} into a *coefficient* matrix, and the b_i into a *column vector*. We then *augment* the coefficient matrix with the column vector, to obtain a matrix of the form

$$\left[\begin{array}{ccc|c} a_{11} & \ldots & a_{1n} & b_1 \\ a_{21} & \ldots & a_{2n} & b_2 \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & \ldots & a_{mn} & b_m \end{array}\right]$$

In essence, the augmentation "bar" functions as an equals sign (so the actual coefficient matrix is the stuff to the left of the bar, and it is not really correct to think of the rightmost column as part of the matrix), and the x_i variables are suppressed. We then perform *row operations* to transform the matrix to *row echelon form*. The allowed row operations are:

- Swap two rows R_i and R_j.
- Scale a row by a scalar $r \in \mathbf{R}$; that is, $R_i \to rR_i$.
- For two different rows R_i and R_j, add a scalar multiple of the row R_i to R_j; that is, $R_j \to rR_i + R_j$ for some $r \in \mathbf{R}$ (note in particular that R_i stays the same).

Here, addition and multiplication are done componentwise, so remember that all entries in a row have to change together. To be in row echelon form, the matrix must satisfy:

- All nonzero rows are above any rows of all zeros.
- Each leading entry of a row is in a column to the right of the leading entry of the row above it (in particular the leading terms go down and to the right in a "staircase" pattern.
- All entries in a column below a leading entry are zeros.

To be in reduced row echelon form, the leading entry in a nonzero row must be 1, and each leading 1 must also be the only nonzero entry in its column. So for instance:

Example 1.1.2. The matrix

$$\begin{bmatrix} 1 & 0 & 4 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in row echelon but not reduced row echelon form.

Usually, row echelon form is sufficient to solve systems. We'll do an example:

Example 1.1.3. Consider the system of equations given by the augmented matrix

$$\left[\begin{array}{ccc|c} -1 & 3 & 4 & 2 \\ 2 & 0 & 2 & -4 \end{array}\right].$$

We can transform this matrix into row echelon form via the operations

$$\left[\begin{array}{ccc|c} -1 & 3 & 4 & 2 \\ 2 & 0 & 2 & -4 \end{array}\right] \to \left[\begin{array}{ccc|c} -1 & 3 & 4 & 2 \\ 1 & 0 & 1 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -3 & -4 & -2 \\ 1 & 0 & 1 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -3 & -4 & -2 \\ 0 & 3 & 5 & 0 \end{array}\right].$$

What is this last matrix really telling us? Well, the two rows say that $x_1 - 3x_2 - 4x_3 = -2$ and $3x_2 + 5x_3 = 0$. The second equation gives us $x_2 = -5x_3/3$, and then the first row gives us $x_1 = -2 - x_3$. Therefore we have solved the equation *parametrically*, with x_3 our parameter and x_1 and x_2 determined in terms of x_3.

One more example:

Example 1.1.4. Consider the system of equations given by the augmented matrix

$$\left[\begin{array}{cc|c} -1 & 3 & 4 \\ 2 & 3 & 1 \\ 1 & 2 & 0 \end{array}\right].$$

We reduce using row operations:

$$\left[\begin{array}{cc|c} -1 & 3 & 4 \\ 2 & 3 & -2 \\ 1 & -2 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} -1 & 3 & 4 \\ 0 & 9 & 6 \\ 0 & 1 & 4 \end{array}\right] \to \left[\begin{array}{cc|c} -1 & 3 & 4 \\ 0 & 1 & 4 \\ 0 & 3 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} -1 & 3 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & -10. \end{array}\right].$$

What does the last row actually mean? It says that $0x_1 + 0x_2 = -10$, which is clearly impossible. Therefore this system is *inconsistent*, since any solution to the original system would obtain the contradiction $0 = -10$ as a necessary deduction.

Remark 1.1.5. WARNING: Do not augment matrices for no good reason! Whenever you write down an augmented matrix, you should have a very clear idea of what system of equations you are trying to solve, and only then put down the "bar" in the appropriate place (i.e. functioning as an equals sign) and carry out the row reduction. Randomly augmenting and/or row reducing matrices is not likely going to help you. Row reduction should only be carried out as one of the final steps in your solution, when you have to make an actual numerical computation. The previous steps in the solution should involve general reasoning using the results presented below about vector spaces, linear maps, etc.

1.2 Vector Spaces and Linear Combinations

The above shows us how to solve any *specific* linear system. Therefore we should attempt a more significant question:

Question 1.2.1. Given a coefficient matrix with entries (a_{ij}), for which vectors $(b_1, \ldots, b_m)$ can the system (1.1.1) be solved?

It is this question, and variants thereof, that are at the heart of linear algebra. To answer this completely, we need to set up some terminology.

Definition 1.2.2. A (real) *vector* space V is a set, whose elements are called *vectors*, such that:

- V has an *addition* operation with the following properties:
  - The addition is associative and commutative.

  - There is a distinguised zero vector $0 \in V$ (the "zero vector") such that $0 + v = v + 0 = v$ for all $v \in V$.
  - For every vector $v \in V$, there is a unique additive inverse vector $-v$ such that $v - v = (-v) + v = 0$.

- V also has a *scalar multiplication* operation, denoted rv or $r \cdot v$ for a real number (scalar) r and vector v, such that:
  - The scalar multiplication is "associative": $r(sv) = (rs)v$ for scalars r, s and vector v.
  - $1 \cdot v = v$ for all v.
  - Scalar multiplication is "distributive": $r(v + w) = rv + rw$ and $(r + s)v = rv + sv$ for scalars r, s and vectors v, w.
  - From this, it follows that $0 \cdot v = 0$ for all $v \in V$. Note that the 0 on the left-hand side is the zero scalar, while the 0 on the right-hand side is the zero vector!

Example 1.2.3. The most familiar example of a vector space is *Euclidean space* $\mathbf{R}^n$. This is just the set of n-tuples of real numbers with the vector operations being componentwise addition and scalar multiplication. Other examples include: the space $\mathbf{P}_n$ of polynomials of degree at most n; $M_{m\times n}$, the space of $m \times n$ matrices (with componentwise addition and scalar multiplication); $C[0, 1]$, the continuous functions from $[0, 1]$ to $\mathbf{R}$; $C^\infty(\mathbf{R})$, the smooth (infinitely differentiable) functions $\mathbf{R} \to \mathbf{R}$.

We will usually be discussing the Euclidean spaces $\mathbf{R}^n$. On the other hand, it is useful to adopt this more general framework, since we may sometimes have to deal with subsets of $\mathbf{R}^n$ that are themselves vector spaces. We define this notion:

Definition 1.2.4. A *subspace* W of a vector space V is a subset closed under the vector space operations of addition and scalar multiplication. This means that for any $w, w' \in W$ and $r \in \mathbf{R}$, $w + w' \in W$ and $rw \in W$.

From the definition, we see that the zero vector has to be in W. Moreover, we always have a *trivial subspace* of V consisting of only the 0 vector, and V is of course a subspace of itself. For more examples:

Example 1.2.5. The set of vectors (x, y) such that $3x - y = 0$ is a subspace of $\mathbf{R}^2$, since sums and scalar multiples of such vectors satisfy the same property. The set of vectors of the form $(1, y)$ is not a subspace of $\mathbf{R}^2$, since it doesn't contain 0. The set of vectors of the form (x, x^2) is not a subspace of $\mathbf{R}^2$, since $(1, 1) + (2, 4) = (3, 5)$ is not in this set.

Let's distill down to the essence of vector spaces, using $\mathbf{R}^n$ as our mental guide if we need to. In vector spaces, we are allowed two operations: "add" and "scale". A subspace is simply a subset where, if we repeatedly apply these operations to various elements in that subset, we never "leave" (i.e. produce elements outside) that subset. It is therefore useful to formally define this notion of "applying addition and scaling over and over":

Definition 1.2.6. Suppose $\{v_1, \ldots, v_n\}$ are vectors in a vector space $\mathbf{R}^n$. A *linear combination* of the v_i is a vector of the form $c_1 v_1 + \ldots + c_n v_n$, where the c_i are scalars. The *span* of the v_i is the set of all possible linear combinations of the v_i.

Remark 1.2.7. We say that vectors $\{v_1, \ldots, v_n\}$ *span* a subspace $W \subseteq V$ if every vector in W is a linear combination of the v_i, or equivalently, that $W \subseteq \mathrm{Span}(v_1, \ldots, v_n)$. If all the v_i are in W, then we would have $W = \mathrm{Span}(v_1, \ldots, v_n)$, and in this case we say that $\{v_1, \ldots, v_n\}$ are a spanning set for W. Be careful with this multiple use of the word span!

We can therefore restate everything above in terms of linear combinations. A linear combination of some vectors is just something that can be created by repeatedly applying the operations of adding and scaling. A subspace is a subset of a vector space closed under taking linear combinations. Moreover, since the span of a set of vectors is, by definition, *all possible linear combinations* of the v_i, and linear combinations of linear combinations (of the v_i) are once again linear combinations of the v_i, we have:

Proposition 1.2.8. The span $\mathrm{Span}(v_1, \ldots, v_n)$ of any set of vectors of V is a subspace of V.

Example 1.2.9. Once again we consider vectors in $\mathbf{R}^2$. The span of $(1, 0)$ and $(0, 1)$ is clearly all of $\mathbf{R}^2$, since every vector (x, y) is a linear combination of those two: $(x, y) = x(1, 0) + y(0, 1)$. The span of $(1, 0)$ and $(1, 1)$ is also all of $\mathbf{R}^2$, since $(x, y) = (x - y)(1, 0) + y(1, 1)$. However, the span of $(1, 2)$ and $(2, 4)$ is *not* all of $\mathbf{R}^2$, since any vector (x, y) in their span must have $y = 2x$, but there are many vectors in $\mathbf{R}^2$ (e.g. $(1, 1)$) for which the second component is not twice the first.

The vector spaces we deal with in this course are all special: they can be spanned by finitely many vectors. We call such spaces *finite-dimensional*. For instance, of the vector spaces introduced in Example 1.2.3, $\mathbf{R}^n$, $\mathbf{P}_n$, and $M_{m\times n}$ are finite-dimensional. $C[0, 1]$ and $C^\infty(\mathbf{R})$ are not finite-dimensional (can you see why)?

1.3 Linear Maps

Now that we have set up the objects of our study (the vector spaces), we need to discuss functions between vector spaces (and we will see how this relates to Question 1.2.1). Since we have set up vector spaces as sets with the operations "add" and "scale", in order for maps between vector spaces to have any sort of reasonable behavior, we need to stipulate that they "preserve" these operations, so that the map $T : V \to W$ can transfer some information about the relationships between vectors in V to relationships between vectors in W (so T is not some completely random thing). So we define:

Definition 1.3.1. A $T : V \to W$ between vector spaces is *linear* if it "respects addition and scalar multiplication": that is,

- $T(v + v') = T(v) + T(v')$ for all $v, v' \in V$.
- $T(rv) = rT(v)$ for all $r \in R, v \in V$.

Example 1.3.2 (Very important!)**.** Suppose A is a $m \times n$ matrix. Then the function $\mathbf{R}^n \to \mathbf{R}^m$, $v \mapsto Av$ is linear.

Example 1.3.3. For any vector spaces V and W, the map $V \to W$ sending all elements of V to $0 \in W$ is linear, and called the *zero map.*

Example 1.3.4. The derivative map $d : C^\infty(\mathbf{R}) \to C^\infty(\mathbf{R})$ is linear, because the derivative of the sum is the sum of the derivatives, and you can pull scalars out of derivatives.

Example 1.3.5. If $\mathbf{P}$ is the vector space of all polynomials with real-valued coefficients, and a is a real number, then the map $\mathbf{P} \to \mathbf{R}$ given by evaluation at a, $p \mapsto p(a)$ is linear.

Example 1.3.6. The map $\mathbf{P} \to \mathbf{P}$ given by sending a polynomial p to the unique antiderivative P of p such that $P(0) = 0$ is linear. It is *not* linear if we instead send p to the unique antiderivative P' of p such that $P'(0) = 1$ (why?).

Example 1.3.7. Suppose l is a line in the plane $\mathbf{R}^2$ that passes through the origin. Then the map $T : \mathbf{R}^2 \to \mathbf{R}^2$ that reflects a point through the line l is linear (this one is a bit harder to see why that is the case).

Example 1.3.8. The linear maps from $\mathbf{R}$ to $\mathbf{R}$ are precisely of the form $x \mapsto cx$ for a fixed real number c.

Example 1.3.9. Let S be the vector space of all *infinite* tuples of real numbers $(x_1, x_2, \ldots)$, where the vector space operations are componentwise addition and scalar multiplication. Consider the *right-shift* map $S \to S$ that sends $(x_1, x_2, \ldots)$ to $(0, x_1, x_2, \ldots)$. This is a linear map.

By combining adding and scaling, we see that **a map is linear if it respects linear combinations**:

$$T(c_1 v_1 + \ldots + c_n v_n) = c_1 T(v_1) + \ldots + c_n T(v_n). \tag{1.3.1}$$

The most important application of this is when the v_i already span the vector space V. In that case, **knowing the effect of the map T on the v_i determines the entire map T.** The slogan is:

A linear map $T : V \to W$ is determined on a spanning set for T.

In other words, I can reconstruct the map T just by knowing what it does to the v_i. Indeed, since any vector in V can be written as a linear combination $c_1v_1 + \ldots + c_nv_n$, Equation (1.3.1) tells us how to evaluate T on such a vector.

Example 1.3.10. Let $V = \mathbf{R}^n$. Let e_i be the vector in $\mathbf{R}^n$ with a 1 in the ith component, and 0s elsewhere. We call the set $\{e_1, \ldots, e_n\}$ the *standard basis* of $\mathbf{R}^n$ (we will explain the name later). Suppose now that $n = 3$, and we have a linear map $T : \mathbf{R}^3 \to \mathbf{R}^2$ such that $T(e_1) = (1, 0)$, $T(e_2) = (1, 1)$, and $T(e_3) = (0, 2)$. Then we can calculate, for instance:

$$T(2, 3, -1) = T(2e_1 + 3e_2 - e_3) = 2T(e_1) + 3T(e_2) - T(e_3) = (-1, 1).$$

In general,

$$T(x, y, z) = T(xe_1 + ye_2 + ze_3) = (x + y, y + 2z).$$

We now come to the all-important result:

Proposition 1.3.11. Linear maps $\mathbf{R}^n \to \mathbf{R}^m$ are "the same thing" as $m \times n$ matrices A. More precisely, every such A induces a linear map by multiplication against an n-vector (this is Example 1.3.2), and more importantly, every linear map $T : \mathbf{R}^n \to \mathbf{R}^m$ is the same as the map $v \mapsto A_Tv$ for a unique $m \times n$ matrix A_T (so $A_Tv = T(v)$ for all v).

In fact, the discussion in Example 1.3.10 shows how A_T is determined from T. One can show from the definition of matrix multiplication that if A has column vectors $a_1, \ldots, a_n \in \mathbf{R}^m$, then for a column vector $v = (x_1, \ldots, x_n)^T$, we have

$$Av = x_1a_1 + \ldots + x_na_n. \tag{1.3.2}$$

But since $v = x_1e_1 + \ldots + x_ne_n$ and $T(v) = x_1T(e_1) + \ldots + x_nT(e_n)$, if we require that the ith column a_i of A to equal $T(e_i)$, we would have $Av = T(v)$ for any v. **This is the procedure to (re)construct the matrix of a linear transformation.**

Example 1.3.12. Let T be as in Example 1.3.10. Then T has the matrix $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$. Indeed, we have

$$A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + y \\ y + 2z \end{bmatrix}$$

as previously calculated.

Example 1.3.13. Let T be the map $\mathbf{R}^2 \to \mathbf{R}^2$ that rotates vectors counterclockwise in the plane by θ radians. One can show that T is a linear map. Then to determine the matrix for T, we need to find the images of e_1 and e_2 under T. Basic trigonometry shows that $T(e_1) = (\cos(\theta), \sin(\theta))$ and $T(e_2) = (-\sin(\theta), \cos(\theta))$, so that the matrix for T is

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

1.4 Linear Independence and Bases

In every situation above, we have been handed a linear map T, which we then analyze. Suppose we are now faced with the task of *constructing* a linear map from $V \to W$. Ideally, we would like to mimic the above ideas but in reverse: supposing that $\{v_1, \ldots, v_n\}$ span V, we would choose $w_1, \ldots, w_n$ to be the images of the v_i under T, and then stipulate that

$$T(c_1v_1 + \ldots + c_nv_n) = c_1T(v_1) + \ldots + c_nT(v_n) = c_1w_1 + \ldots + c_nw_n \tag{1.4.1}$$

to give the effect of T on any element of V (this procedure is known as "extending by linearity"). However, we may run into the following problem. Suppose for another set of scalars $d_1, \ldots, d_n$, where the c_i are not all equal to the corresponding d_i, we had

$$c_1v_1 + \ldots + c_nv_n = d_1v_1 + \ldots + d_nv_n.$$

Then the above construction would force

$$c_1w_1 + \ldots + c_nw_n = d_1w_1 + \ldots + d_nw_n,$$

but there is no guarantee that this actually occurs for the w_i that we chose, so we may have run ourselves into a contradiction. Therefore, we are motivated to make the following definition:

Definition 1.4.1. A set of vectors $\{v_1, \ldots, v_n\}$ in V is *linearly independent* if

$$c_1v_1 + \ldots + c_nv_n = 0 \tag{1.4.2}$$

implies $c_1, \ldots, c_n$ are all 0. In other words, the only linear combination of the v_i that equals 0 is the trivial linear combination, where all the scalars are 0. Otherwise the v_i are *linearly dependent*.

This condition really is what we wanted:

Proposition 1.4.2. The following are equivalent:

1. Vectors $\{v_1, \ldots, v_n\}$ in V are linearly independent.
2. For any $w \in \text{Span}(v_1, \ldots, v_n)$, w can be written as a linear combination of the v_i in *exactly one way*. In other words, if $c_1v_1 + \ldots + c_nv_n = d_1v_1 + \ldots + d_nv_n$, then $c_i = d_i$ for all i.

Similarly,

Proposition 1.4.3. The following are equivalent:

1. Vectors $\{v_1, \dots, v_n\}$ in V are linearly dependent.
2. There is some v_j that can be expressed as a linear combination of the other v_i (not including v_j).
3. There is some $w \in \operatorname{Span}(v_1, \dots, v_n)$ such that w can be written as a linear combination of the v_i in more than one way.

Here are some examples to build some intuition for linear (in)dependence.

Example 1.4.4. We check whether a specific set of vectors in $\mathbf{R}^n$ is linearly independent by solving the homogeneous system given by Equation (1.4.2), and seeing whether the zero solution is the only solution to that system. For instance, the vectors $(1, 1, 2), (1, 2, -1), (-1, 1, 1)$ in $\mathbf{R}^3$ are linearly independent, since the system

$$\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ x_1 + 2x_2 + x_3 &= 0 \\ 2x_1 - x_2 + x_3 &= 0 \end{aligned}$$

only has the trivial solution.

Example 1.4.5. In a more abstract situation, we will have to reason more generally. For instance, suppose that v_1, v_2, v_3 are linearly independent vectors in some space V. Then we claim that $v_1 + v_2, v_2 + v_3, v_1 + v_3$ are also linearly independent. Indeed, if there was a linear combination $c_1(v_1 + v_2) + c_2(v_2 + v_3) + c_3(v_1 + v_3)$ that was 0, then we also have

$$(c_1 + c_3)v_1 + (c_1 + c_2)v_2 + (c_2 + c_3)v_3 = 0,$$

so linear independence of v_1, v_2, v_3 gives $c_1 + c_3 = c_1 + c_2 = c_2 + c_3 = 0$. Solving this homogeneous system yields $c_1 = c_2 = c_3 = 0$, so that $\{v_1 + v_2, v_2 + v_3, v_1 + v_3\}$ is a linearly independent set of vectors by definition.

Example 1.4.6. Any set of vectors containing the 0 vector must be linearly dependent, since if the set is $\{v_1, \dots, v_n\}$ and $v_1 = 0$, then we have a nontrivial linear combination $v_1 + 0 \cdot v_2 + \dots + 0 \cdot v_n$ that equals 0.

Example 1.4.7. By item (2) of Proposition 1.4, it is true that *two* vectors are linearly dependent only if one is a scalar multiple of the other. But this is ***not*** true for three or more vectors: the vectors $(1, 0), (0, 1), (-1, -1)$ are linearly dependent as their sum is 0, but none of the three is a scalar multiple of another.

Example 1.4.8. Notice that we are very particular to say "some v_j" in item (2) of Proposition 1.4. It is ***not*** true that if $\{v_1, \dots, v_n\}$ are linearly dependent, then *every* v_j can be expressed as a linear combination of the other v_i. Indeed, the vectors $(1, 0), (0, 1), (2, 0)$ are linearly dependent, and $(2, 0)$ is a linear combination of $(1, 0)$ and $(0, 1)$, but $(0, 1)$ is not a linear combination of $(1, 0)$ and $(2, 0)$.

Remark 1.4.9. Linear (in)dependence is a notion that applies to a *set* of (multiple) vectors. It does not make sense to refer to a single vector as "linearly independent" (from what?), or refer to any non-vector object as "linearly (in)dependent."

We now investigate the relationship between the ideas of "spanning set" and "linear independence". Here is the first important result:

Proposition 1.4.10. Suppose a vector space V can be spanned by n vectors. Then if $m > n$, any collection of m vectors in V are linearly dependent.

As an application, any $n+1$ vectors in $\mathbf{R}^n$ are linearly dependent. This makes sense, as $\mathbf{R}^n$ only has n "degrees of freedom", so $n+1$ vectors should have some dependency going on among them. Similarly, $\mathbf{R}^n$ cannot be spanned by *less* than n vectors, because $\mathbf{R}^n$ contains the n vectors $e_1, \ldots, e_n$, which are linearly independent.

Corollary 1.4.11. If two finite linearly independent sets of vectors $\{v_1, \ldots, v_r\}$ and $\{w_1, \ldots, w_s\}$ both have the same span V' inside V, then $r = s$.

In the special case that $V' = V$, we conclude that any linearly independent spanning set of V has the same (finite) size. Therefore we are motivated to define

Definition 1.4.12. A *basis* of a vector space V is a linearly independent spanning set. The *dimension* of a finite-dimensional vector space is the size of any basis, and any spanning set for V has at least $\dim(V)$ elements.

It is a fact that every vector space has a basis. In the finite-dimensional case, this may be seen as follows. Take a finite set $\{v_1, \ldots, v_n\}$ that spans V. Then if any of the v_i, say v_n, is a linear combination of the others, remove it from the set; $\{v_1, \ldots, v_{n-1}\}$ will still span V (because v_n was "redundant" and did not allow us to produce any new linear combinations that we could not have achieved using only $v_1, \ldots, v_{n-1}$). If we continue this procedure of removing vectors that are linear combinations of the others, until none of the vectors are linear combinations of the others, then the remaining set will be linearly independent. Since the span remained the same, this truncated set will be a basis for V.

It also follows from Proposition 1.4.10 that

Corollary 1.4.13. If W is a subspace of V, then $\dim(W) \subseteq \dim(V)$, and if this inequality is an equality, than $W = V$.

Therefore the notion of dimension behaves in a reasonable manner.

In the familiar case of Euclidean space, we have:

Example 1.4.14. The standard basis vectors e_i form a basis for $\mathbf{R}^n$, so $\mathbf{R}^n$ has dimension n. The set $\{(1, 1), (-1, 1)\}$ is a different basis (of size 2!) for $\mathbf{R}^2$.

Proposition 1.4.15. Suppose V is finite-dimensional with $\dim(V) = n$. Then any linearly independent set of n vectors in V is a basis.

This is useful, since it is usually easier to check that a set of vectors is linearly independent, than to check that it spans some vector space.

Remark 1.4.16. It is good to think of dimension as how much "information" you need to specify a vector in a vector space V. The intuition is that we need $n := \dim(V)$ pieces of information, since if we have a basis $\{v_1, \ldots, v_n\}$ for V, then a vector is specified by giving n real numbers that serve as the coefficients in the linear combination $c_1v_1 + \ldots + c_nv_n$. In the case of Euclidean space with the standard basis, these n pieces of information are just the components in the n-tuple representation of a vector.

It is also useful to be able to extend linearly independent subsets of a vector space to a basis.

Proposition 1.4.17. Suppose V is finite-dimensional and $\{v_1, \ldots, v_k\}$ is a linearly independent set of vectors in V (so necessarily $k \leq \dim(V)$). Then we may always "complete" this set to a basis of V: there are $\dim(V) - k$ vectors $w_1, \ldots, w_{\dim(V)-k}$ in V such that the set $\{v_1, \ldots, v_k, w_1, \ldots, w_{\dim(V)-k}\}$ is a basis for V.

Example 1.4.18. Suppose we have the two linearly independent vectors $(1, 1, 2, 3)$ and $(0, 0, 1, 1)$ in $\mathbf{R}^4$, which has dimension 4. Then by adding the $4 - 2 = 2$ vectors $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$, we obtain a basis of $\mathbf{R}^4$.

Finally, we come back to the original motivation for a basis. The mantra is:

To give a linear map $T : V \to W$ is equivalent to specifying the image of a V-basis in W.

By this, we mean the following. To construct a linear map $T : V \to W$ (assuming V is finite-dimensional), we simply need to give a basis $\{v_1, \ldots, v_n\}$ of V, pick out elements $w_1, \ldots, w_n$, and stipulate that $T(v_i) = w_i$ for each i. Then we can extend the effect of T via linearity to all elements of V, using the rule of (1.4.1). Since every element v of V can be *uniquely* expressed as a linear combination $c_1v_1 + \ldots + c_nv_n$ (this is where linear independence comes into play!), we are assured that there is no ambiguity in my choice of the representation of v as a linear combination of the v_i (such ambiguity caused the problem discussed at the beginning of this subsection).

Remark 1.4.19. I strongly recommend to think in terms of linear maps, rather than always relying on the crutch of matrix representations. The main reason is that matrix representations depend on the choice of a specific basis/coordinate system, and sometimes you do not have the luxury of writing down such a choice (e.g. if you are working with some subspace inside $\mathbf{R}^n$). Direct matrix computations also frequently obfuscate the underlying geometry of the vector spaces and linear maps that appear, so that it's very easy to get lost in computations without having any idea of what is actually going on.

1.5 Kernel (nullspace), Image (column space), and back to linear systems

To every linear map, we may associate two important vector (sub)spaces:

Definition 1.5.1. Let $T : V \to W$ be a linear map, with associated matrix A if the map is between Euclidean spaces $\mathbf{R}^n \to \mathbf{R}^m$. The *nullspace* or *kernel* of T, denoted $\ker(T)$ or $N(A)$, is the subspace of V consisting of vectors v such that $T(v) = 0$. The *column space* or *image* of T, denoted $\operatorname{Im}(T)$ or $C(A)$, is the subspace of W consisting of vectors of the form $T(v)$ for some $v \in V$. That is, W consists of the elements in W that are "hit" by T. Convince yourself that these are actually subspaces of V and W respectively!

Remark 1.5.2. It's important to keep track of where the kernel and image "live". The kernel lives inside the domain of the linear map, and the image lives inside the codomain, so unless the domain is the codomain (i.e. $W = V$), the domain and codomain will not "interact" (so you shouldn't be adding vectors in the kernel to vectors in the image, for instance!).

Let's specialize to the case when T is a map $\mathbf{R}^n \to \mathbf{R}^m$, so we can represent it by a matrix A. Then by Equation (1.3.2), it follows that

Proposition 1.5.3. The image of T is *exactly* the span of the columns of A (hence the name "column space"). The kernel of T is exactly the set of solutions $x \in \mathbf{R}^n$ to the homgeneous system $Ax = 0$ (which can be computed by row reduction).

Therefore we can now answer the Question 1.2.1 that started this whole discussion. Suppose we have a coefficient matrix A with entries (a_{ij}), and we ask for which vectors $b = (b_1 \ldots, b_m) \in \mathbf{R}^m$ can the system (1.1.1) be solved by a vector $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. But the definition of matrix multiplication shows that (1.1.1) is simply asking us to find x such that $Ax = b$. So:

Proposition 1.5.4. The linear system (1.1.1) can be solved exactly when b is in the column space of A, or equivalently, when b is in the span of the columns of A.

Corollary 1.5.5. The linear system (1.1.1) can be solved *for any* $b \in \mathbf{R}^m$ when $C(A) = \mathbf{R}^m$, or equivalently, when the columns of A span $\mathbf{R}^m$.

It is useful to give a word to the situation described in the previous corollary. Namely:

Definition 1.5.6. A linear map $T : V \to W$ is called *onto*, or *surjective*, if $\operatorname{Im}(T) = W$.

For linear maps $T : \mathbf{R}^n \to \mathbf{R}^m$, being surjective implies that $n \geq m$, or that the corresponding matrix A has at least as many columns as rows and that the columns span $\mathbf{R}^m$.

This is a corollary of the fact that it requires at least m vectors to span $\mathbf{R}^m$ (cf. Definition 1.4.12).

Having answered Question 1.2.1 in some sense, we can now ask the "dual" question. With the $Ax = b$ setup as above, we now ask when this system has *at most* one solution (instead of "at least one solution", as before), so that a solution is unique if it exists. This question is answered by the nullspace:

Proposition 1.5.7. The linear system (1.1.1) has at most 1 solution $x \in \mathbf{R}^n$ for any $b \in \mathbf{R}^m$ exactly when $N(A) = 0$, or equivalently, when the columns of A are linearly independent.

It is again useful to make a definition based on this proposition:

Definition 1.5.8. A linear map $T : V \to W$ is called *one-to-one*, or *injective*, if $\ker(T) = 0$. This is the same as saying that $T(v) = T(v')$ only if $v = v'$; that is, T sends distinct elements of V to distinct elements of W.

For linear maps $T : \mathbf{R}^n \to \mathbf{R}^m$, being injective implies that $n \leq m$, or that the corresponding matrix A has at least as many rows as columns and that the columns are linearly independent in $\mathbf{R}^m$. This is a corollary of the fact that no more than n vectors can be linearly independent in $\mathbf{R}^n$ (cf. Proposition 1.4.10).

Remark 1.5.9. Notice that if T is both injective and surjective (in which case we call T *bijective*), we must have $n = m$. In this case, if T is a linear map $\mathbf{R}^n \to \mathbf{R}^n$ with associated matrix A, we can interpret this condition in the context of linear systems, in that the linear system (1.1.1) has *exactly* 1 solution $x \in \mathbf{R}^n$ for any $b \in \mathbf{R}^m$. This very interesting case will be discussed further in the section about matrix inversion.

We should notice the duality between linear independence and span, nullspace and column space, manifested in the solutions of linear systems (at most one solution versus at least one solution).

2 Matrix Properties

2.1 Transpose

Definition 2.1.1. The transpose A^T of a $m \times n$ matrix A is an $n \times m$ matrix given by turning the rows of A into the columns of A^T.

Example 2.1.2. If

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

then

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

The actual geometric interpretation of the transpose is mysterious: it tells us the *dual linear map* between the *dual vector spaces*. We won't need this interpretation for the course, so it's fine to only remember the definition of the transpose. We also have the useful formula

$$(A_1 \dots A_n)^T = A_n^T \dots A_1^T,$$

which is to say that the transpose of a product is the product of the transposes in the reverse order.

2.2 Multiplication and Inversion

The most important thing to remember with matrix multiplication is that

Matrix multplication corresponds to composition of linear transformations.

By this, we mean that if $T : \mathbf{R}^n \to \mathbf{R}^m$ and $S : \mathbf{R}^m \to \mathbf{R}^p$ are linear transformations with corresponding matrices A_T and A_S, then the matrix corresponding to the map $S \circ T : \mathbf{R}^n \to \mathbf{R}^p$ is given by $A_S A_T$. This makes it very apparent why ***matrix multiplication is not in general commutative***, since composition of functions is not commutative (usually $f \circ g \neq g \circ f$). This also shows why, for instance, that $AB = 0$ does not imply $A = 0$ or $B = 0$ (two nonzero functions can compose to the 0 function), and why $AB = AC$ does not imply $B = C$. Also, since the composition of one-to-one or onto functions is also one-to-one or onto, we conclude the same is true for products of such matrices.

Example 2.2.1. Suppose T is the linear transformation $\mathbf{R}^2 \to \mathbf{R}^2$ that rotates vectors by $\pi/2$ radians counterclockwise, and S is the linear transformation that reflects vectors across the line $y = x$. We will find the matrix A_{ST} for $S \circ T$ in two different ways. First, we can directly compute this matrix by finding $S(T(e_1))$ and $S(T(e_2))$. We have $S(T(e_1)) = S(0, 1) = (1, 0)$ and $S(T(e_2)) = S(-1, 0) = (0, -1)$, so the matrix for ST is

$$A_{ST} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We can also compute A_T and A_S individually, and multiply them. Since $T(e_1) = (0, 1)$ and $T(e_2) = (-1, 0)$, $A_T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Similarly, $S(e_1) = (0, 1)$ and $S(e_2) = (1, 0)$, so $A_S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then

$$A_S A_T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

which is precisely A_{ST}.

We now turn to the problem of when a linear transformation is *invertible* in some sense. Suppose we have a linear map $T : \mathbf{R}^n \to \mathbf{R}^m$ represented by a matrix A. We ask when there is a map $S : \mathbf{R}^m \to \mathbf{R}^n$ such that $S \circ T$ is the identity map on $\mathbf{R}^n$, in which case we call T *left-invertible*, and we call S a left inverse of T. Equivalently, we ask when there is a $n \times m$ matrix B such that $BA = I$, in which case we call A left-invertible as well (and B the left inverse of A). This is true when:

Proposition 2.2.2. A linear map $T : \mathbf{R}^n \to \mathbf{R}^m$, or the corresponding matrix A, is left-invertible exactly when it is *injective*.

In particular, if a $m \times n$ matrix is left-invertible, then we must have $m \geq n$.

We can ask the same question about the dual notion of right-inveribility. Suppose we have a linear map $T : \mathbf{R}^n \to \mathbf{R}^m$ with corresponding matrix A. We ask when there is a map $S : \mathbf{R}^m \to \mathbf{R}^n$ such that $T \circ S$ is the identity map on $\mathbf{R}^m$, in which case we call T *right-invertible* and S the right inverse of T (note the difference from left-invertibility!). Equivalently, we ask when there is a $n \times m$ matrix B such that $AB = I$, in which case we call A right-invertible as well. This is true when:

Proposition 2.2.3. A linear map $T : \mathbf{R}^n \to \mathbf{R}^m$, or the corresponding matrix A, is right-invertible exactly when it is *surjective*.

In particular, if a $m \times n$ matrix is right-invertible, then we must have $m \leq n$.

Definition 2.2.4. A linear map $T : \mathbf{R}^n \to \mathbf{R}^m$, or the corresponding matrix A, is *invertible* if it is both left and right-invertible.

Putting together the above discussion, we see that if $T : \mathbf{R}^n \to \mathbf{R}^m$ is an invertible linear map, when $m \geq n$ and $m \leq n$, so that $m = n$ and the associated matrix A must be square. Moreover, T must be both injective and surjective; we call such maps *bijective*. We saw in Remark 1.5.9 that this can be interpreted in the context of linear systems by saying that the linear system $Ax = b$ has a unique solution $x \in \mathbf{R}^n$ for all $b \in \mathbf{R}^n$. Also, since compositions of bijective maps are bijective, it follows that a product of invertible matrices is also invertible.

Proposition 2.2.5. If A an an invertible square matrix, then its left and right inverses coincide, and is unique. In other words, if A is an invertible $n \times n$ matrix, then there is exactly one matrix, denoted A^{-1}, such that $AA^{-1} = A^{-1}A = I_n$, and if any other square matrix satisfies either $BA = I$ or $AB = I$, then $B = A^{-1}$.

Therefore if A is invertible, the unique solution to $Ax = b$ is given by $x = A^{-1}b$. It also follows from the above uniqueness that $(A^{-1})^{-1} = A$, and if $A_1, \ldots, A_n$ are invertible square matrices of the same dimension, then $(A_1 \ldots A_n)^{-1} = A_n^{-1} \ldots A_1^{-1}$ (i.e. **the inverse of the product is the product of the inverses in the reverse order**).

Remark 2.2.6. WARNING: It is *not* true that the left inverse or right inverse is unique for (non-square) matrices. In other words, if A is some (possibly non-square) $m \times n$ matrix such that $BA = I_n$ and $CA = I_n$, we *cannot* conclude that $B = C$. The uniqueness only holds for *square* matrices.

Now, suppose we know that a matrix A is square (of dimension n), which is a *necessary* condition for it to be invertible. We ask under what conditions it is actually invertible. We know that invertibility is equivalent to bijectivity of the corresponding linear map $T : \mathbf{R}^n \to \mathbf{R}^n$, but the magic fact is that

For a linear map T from $\mathbf{R}^n$ to itself, injectivity, surjectivity, and bijectivity are all equivalent.

Remark 2.2.7. This fact is highly dependent on the fact that both the domain and codomain of T are $\mathbf{R}^n$! It is very much not true for a linear may between arbitrary vector spaces.

Writing out what this means in terms of matrices and linear systems, we get

Proposition 2.2.8. Suppose A is a *square* $n \times n$ matrix corresponding to the linear map $T : \mathbf{R}^n \to \mathbf{R}^n$. Then the following are equivalent:

1. A is an invertible matrix.
2. T is a bijective linear map. In other words, the columns of A form a basis for $\mathbf{R}^n$, and the linear system $Ax = b$ has a unique solution $x \in \mathbf{R}^n$ for every $b \in \mathbf{R}^n$, which is to say that A can be row reduced to the identity matrix I_n.
3. T is an injective linear map. In other words, the columns of A are linearly independent in $\mathbf{R}^n$, and the linear system $Ax = b$ has at least one solution $x \in \mathbf{R}^n$ for every $b \in \mathbf{R}^n$.
4. T is a surjective linear map. In other words, the columns of A span in $\mathbf{R}^n$, and the linear system $Ax = b$ has at most one solution $x \in \mathbf{R}^n$ for every $b \in \mathbf{R}^n$.

To actually find the inverse, we use the method of row reduction. By Proposition 2.2.8, we know that if a matrix A is invertible, it can be row-reduced to the $n \times n$ identity matrix I_n. To find the inverse, we can perform a series of row operations to reduce A to I_n, while performing the same row operations (in the same order) to the I_n. The result of those row operations on I_n will be A^{-1}. This is most commonly graphically represented by performing row reductions on a "partitioned matrix" $[A|I_n]$, with the matrix A on the left and the identity on the right. (**WARNING:** this is not augmentation in the same sense as in Section 1.1!).

Example 2.2.9. We will find the inverse of the 3×3 matrix

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 2 & -1 & 1 \end{bmatrix}$$

(see Example 1.4.4). To do this, we set up

$$\left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

and row-reduce the left-hand side to the identity. The computation is

$$\left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & -3 & 3 & -2 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 9 & -5 & 3 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & -3 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 & 1 & 0 \\ 0 & 0 & 1 & -5/9 & 1/3 & 1/9 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1 & 0 & 1/9 & 1/3 & -2/9 \\ 0 & 0 & 1 & -5/9 & 1/3 & 1/9 \end{array}\right].$$

Therefore the desired inverse is

$$\begin{bmatrix} 1/3 & 0 & 1/3 \\ 1/9 & 1/3 & -2/9 \\ -5/9 & 1/3 & 1/9 \end{bmatrix}.$$

Of course, if the matrix on the left-hand side of our partition cannot be row-reduced to the identity matrix, then it is not invertible (Proposition 2.2.8).

Sometimes, a more abstract approach to finding the inverse is useful, as the next example shows.

Example 2.2.10. Suppose $T : \mathbf{R}^3 \to \mathbf{R}^3$ is a linear transformation such that $T(2,3,6) = (1,0,0)$, $T(1,0,1) = (0,1,0)$, and $T(-1,-1,3) = (0,1,1)$. Then by linearity, $T(-2,-1,2) = (0,1,1) - (0,1,0) = (0,0,1)$, so $\mathrm{Im}(T)$ contains the standard basis for $\mathbf{R}^3$, and since $\mathrm{Im}(T)$ is a subspace of $\mathbf{R}^3$, it follows that T is surjective (since $\mathrm{Im}(T)$ must contain $\mathrm{Span}(e_1, e_2, e_3)$, which is already all of $\mathbf{R}^3$). By Proposition 2.2.8, it follows that T is invertible; we will find the matrix corresponding to T^{-1}. Since $T^{-1} \circ T = \mathrm{id}$, it follows that T^{-1} must send $(1,0,0)$ to $(2,3,6)$, $(0,1,0)$ to $(1,0,1)$, and $(0,0,1)$ to $(-2,-1,2)$. Therefore we obtain that the matrix for T^{-1} is

$$\begin{bmatrix} 2 & 1 & -2 \\ 3 & 0 & -1 \\ 6 & 1 & 2 \end{bmatrix}$$

without ever having calculated the matrix for T!

3 Determinants

3.1 Basics of determinants

We will define determinants in a more abstract way than the course does. This is because the "Laplace expansion" definition is both quite useless for computations (beyond small cases) and does not capture the fundamental properties of the determinant (in fact, we will not discuss Laplace expansion at all).

Definition 3.1.1. For a positive integer n, the *determinant* is the unique multilinear alternating map

$$\det : \underbrace{\mathbf{R}^n \times \ldots \times \mathbf{R}^n}_{n \text{ times}} \to \mathbf{R}$$

such that $\det(e_1, e_2, \ldots, e_n) = 1$ (notice that the determinant is a function whose input is n vectors in $\mathbf{R}^n$, not just one). The determinant of a square matrix is the determinant, as defined above, applied to the n column vectors of A.

We will briefly explain what "multilinear" and "alternating" mean, but feel free to skip this; much more important are the consequences of these two properties.

- Multilinear means that if all but one of the n arguments in det are held fixed, then det is a linear function $\mathbf{R}^n \to \mathbf{R}$ with respect to the last argument. For instance, the function $v \mapsto \det(e_1, v, e_3, \ldots, e_n)$ is linear in v.

- Alternating means that if $v_i = v_j$ for some $i \neq j$, then $\det(v_1, \ldots, v_n) = 0$. Equivalently, swapping two entries in the argument changes the sign of the determinant:

$$\det(v_1, \ldots, v_n) = -\det(v_1, \ldots, \underbrace{v_j}_{i\text{th position}}, \ldots, \underbrace{v_i}_{j\text{th position}}, \ldots, v_n).$$

Let's collect some consequences that follow almost immediately from Definition 3.1.1.

Proposition 3.1.2. Suppose $v_1, \ldots, v_n$ are vectors such that some v_j is a linear combination of the other v_i. Then $\det(v_1, \ldots, v_n) = 0$. In other words, if A is a square matrix with linearly dependent columns, then $\det(A) = 0$.

It turns out that the converse is also true: If $\det(A) = 0$, then A has linearly dependent columns. But because A must be square (the determinant is only defined for square matrices), Proposition 2.2.8 tells us that

Proposition 3.1.3. A square matrix A is invertible if and only if $\det(A) = 0$.

Therefore the determinant gives us another test for the invertibility of A.

The intuition for the determinant is that it represents "oriented hypervolume". The idea is that the n-dimensional box whose sides are given by vectors $v_1, \dots, v_n$ has determinant $\det(v_1, \dots, v_n)$. For instance, in the $n = 2$ case, v_1 and v_2 could be two non-parallel sides of a parallelogram inside the plane $\mathbf{R}^2$. Then the properties mentioned in the definition of the determinant carry geometric meaning. The condition that $\det(e_1, \dots, e_n)$ means that we assign the "standard n-box" to have an n-volume of 1. The alternating condition is manifested in the sense that if $v_i = v_j$ for some $i \neq j$, then our box collapses to an $n-1$-dimensional box (think of the case $n = 2$ and $n = 3$ for intuition), so its n-volume must be 0 (i.e. a flat parallelogram in 3-space has no 3-volume). Multilinearity also a geometric explanation: for instance, if I start with an n-dimensional box and scale the length of one side while keeping all other sides the same length, then the n-volume of my box should scale by the same factor.

Less obvious are these two properties of determinants:

Proposition 3.1.4. The determinant is a *multiplicative* function: **the determinant of a product is the product of the determinants**. In other words, if A and B are square matrices of the same dimension, then $\det(AB) = \det(A)\det(B)$.

From this, it follows that the determinant of the product of matrices is unchanged if we multiply the matrices in a different order—this is because mutliplication of *real numbers* is commutative, and not because multiplication of matrices is commutative (which is false!). In particular, if a product of matrices has zero determinant, then at least one of the matrices has zero determinant, so any rearrangement of the product still has zero determinant. Thinking of this in terms of invertibility/bijectivity of linear maps, the intuition is multiplying by any one non-invertible matrix/non-bijective map "ruins" the invertibility/bijectivity of the whole product of matrices/composition of linear maps, no matter which order we multiply/compose in.

Corollary 3.1.5. If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.

This is because we must have $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1$.

Proposition 3.1.6. For a square matrix, then $\det(A) = \det(A^T)$.

Therefore A is invertible if and only if A^T is, and in that case we have $(A^T)^{-1} = (A^{-1})^T$ (the transpose of the inverse is the inverse of the transpose). Moreover, since transpose switches the rows and columns, this shows that we can also define the determinant of a square matrix A as the determinant applied to its *row* vectors, and then all of the above results hold with "column of A" replaced by "row of A". In fact, we will exploit this by calculating the determinant of A via *row operations* (column operations also work, but we will do it the way the course does it).

We now demonstrate the method of calcuating determinants using row operations. By the above discussions, we can determine the effects of row operations on the determinant of a matrix A. Suppose A has rows $R_1, \ldots, R_n$. Then:

1. If A' is obtained from A by swapping two rows $R_i \leftrightarrow R_j$, then $\det(A') = \det(A)$.

2. If A' is obtained from A by scaling a row $R_i \to rR_i$, then $\det(A') = r\det(A)$. In other words, we can "factor out" a scalar r from a row if we remember to scale the new determinant by the same amount. This in particular implies that $\det(rA) = r^n \det(A)$, since each row is scaled by r.

3. If A' is obtained from A by adding a scalar multiple of R_i to R_j ($R_j \to rR_i + R_j$), then $\det(A') = \det(A)$.

We also need the fact that

Proposition 3.1.7. The determinant of a triangular matrix A (doesn't matter if upper or lower triangular) is the product of the entries on the main diagonal.

Using this, we can calculate some determinants.

Example 3.1.8. We calculate the determinant of

$$A = \begin{bmatrix} 10 & 11 & 12 & 13 & 426 \\ 2000 & 2001 & 2002 & 2003 & 421 \\ 2 & 2 & 1 & 0 & 419 \\ 100 & 101 & 101 & 102 & 2000 \\ 2003 & 2004 & 2005 & 2006 & 421 \end{bmatrix}.$$

(see next page)

Using row reductions,

$$\det\begin{bmatrix} 10 & 11 & 12 & 13 & 426 \\ 2000 & 2001 & 2002 & 2003 & 421 \\ 2 & 2 & 1 & 0 & 419 \\ 100 & 101 & 101 & 102 & 2000 \\ 2003 & 2004 & 2005 & 2006 & 421 \end{bmatrix} = \det\begin{bmatrix} 10 & 11 & 12 & 13 & 426 \\ 2000 & 2001 & 2002 & 2003 & 421 \\ 2 & 2 & 1 & 0 & 419 \\ 100 & 101 & 101 & 102 & 2000 \\ 3 & 3 & 3 & 3 & 0 \end{bmatrix}$$
$$= 3\det\begin{bmatrix} 10 & 11 & 12 & 13 & 426 \\ 2000 & 2001 & 2002 & 2003 & 421 \\ 2 & 2 & 1 & 0 & 419 \\ 100 & 101 & 101 & 102 & 2000 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$
$$= 3\det\begin{bmatrix} 0 & 1 & 2 & 3 & 426 \\ 0 & 1 & 2 & 3 & 421 \\ 0 & 0 & -1 & -2 & 419 \\ 0 & 1 & 1 & 2 & 2000 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$
$$= -3\det\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 421 \\ 0 & 0 & -1 & -2 & 419 \\ 0 & 1 & 1 & 2 & 2000 \\ 0 & 1 & 2 & 3 & 426 \end{bmatrix}$$
$$= -3\det\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 421 \\ 0 & 0 & -1 & -2 & 419 \\ 0 & 0 & -1 & -1 & 1579 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$
$$= -3\det\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 421 \\ 0 & 0 & -1 & -2 & 419 \\ 0 & 0 & 0 & 1 & 1160 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}.$$

This last matrix is upper-triangular, so the determinant is the product of the diagonal entries, which is -5. Therefore $\det(A) = (-3)(-5) = 15$.

Example 3.1.9. Let A be a $n\times n$ matrix whose diagonal entries are a, and whose off-diagonal entries are b. So

$$A = \begin{bmatrix} a & b & \dots & b \\ b & a & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & a \end{bmatrix}.$$

We will calculate $\det(A)$ in terms of a, b, and n. First, upon replacing the last row by $R_1 + R_2 + \ldots + R_n$, we have

$$\det(A) = \det \begin{bmatrix} a & b & \dots & b \\ b & a & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ a+(n-1)b & a+(n-1)b & \dots & a+(n-1)b. \end{bmatrix} = (a+(n-1)b) \det \begin{bmatrix} a & b & \dots & b \\ b & a & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1. \end{bmatrix}.$$

We then subtract bR_j from each of the other rows. This doesn't change the determinant, so that

$$\det(A) = (a+(n-1)b) \det \begin{bmatrix} a-b & 0 & \dots & 0 \\ 0 & a-b & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}.$$

This last matrix is lower triangular with diagonal entries $a-b, a-b, \ldots, a-b, 1$ (there are $n-1$ occurrences of $a-b$). It follows that $\det(A) = (a+(n-1)b)(a-b)^{n-1}$.

Example 3.1.10. We will show that if $n \geq 2$, the determinant of an $n \times n$ matrix with odd entries is divisible by 2^{n-1}. Indeed, replace each row $R_2, R_3, \ldots, R_n$ by $R_1 + R_2, R_1 + R_3, \ldots, R_1 + R_n$ respectively; this does not change the determinant. Then each of the rows from the 2nd to the nth row now has all even entries, so we can factor out a 2 from each of those $n-1$ rows, so that the determinant of this new matrix is divisible by 2^{n-1}. Therefore our original determinant must also be divisible by 2^{n-1}.

We will give explicit formulas for determinants of 2×2 and 3×3 matrices. These are useful to remember since the matrices are so small, but *do not* try to extend them to matrices of higher dimension.

Proposition 3.1.11. The determinant of a 2×2 matrix is

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

The determinant of a 3×3 matrix is

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

You can remember these rules by the following heuristic: add products of entries on diagonals going right-down and subtract products of entries on diagonals going left-down, "wrapping diagonals around" the matrix if necessary for the 3×3 case.

Finally, we can use the determinant to give an explicit formula for the inverse of a 2×2 matrix, which is sometimes useful to remember. As above, do not try to extend this rule to matrices of higher dimension!

Proposition 3.1.12. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible, then

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Here, the necessary $\det(A) \neq 0$ condition appears as we need to divide by precisely that quantity.

Remark 3.1.13. I recommend against using the Laplace expansion method of calculating the determinant, since it really isn't useful unless $n = 4$. When $n = 2$ or $n = 3$ you have explicit formulas, and when $n \geq 5$ it is slow and error-prone in hand calculations due to the recursive nature. It is much better to use row reduction when calculating all determinants of dimension ≥ 4.

3.2 Cramer's Rule and the Adjugate

Cramer's rule is an explicit formula of solving the linear system $Ax = b$ when A is invertible. For each $1 \leq i \leq n$, define A_i to be the same as the matrix A, but with the ith column replaced by the column vector b. Then Cramer's rule says that

Proposition 3.2.1. The ith component of the unique solution x to the linear system $Ax = b$ is given by

$$x_i = \frac{\det(A_i)}{\det(A)}.$$

Since Cramer's rule is cumbersome to apply when $n \geq 4$ (due to the repeated determinants that need to be calculated), it is *not* recommended as an alternative to the usual row reduction method for concrete calculations. Usually Cramer's rule is invoked in an abstract context.

Finally, we will briefly discuss the adjugate matrix of a square matrix A. This is a mysterious construction whose utility, like Cramer's rule, is usually in a theoretical context.

Definition 3.2.2. Let A be $n \times n$ matrix. The *(i, j)-th minor* M_{ij} is the determinant of the $n - 1 \times n - 1$ matrix given by deleting the ith row and jth column from A. The *(i, j)-th cofactor* C_{ij} is defined as $(-1)^{i+j} M_{ij}$. Let C be the $n \times n$ matrix whose (i, j)-th entry is C_{ij}. We define the *adjugate* matrix of A to be $\operatorname{adj}(A) = C^T$.

From this definition, one can show that

Proposition 3.2.3. For any square matrix A, $A\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I$. In particular, if A is invertible, then $A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$.

Just like with Cramer's rule, the adjugate is pretty useless for actually computing anything (for instance, the inverse of an invertible matrix). But to show its theoretical utility, we will finish with a pretty result.

Proposition 3.2.4. Suppose an invertible square matrix A has integer entries. Then A^{-1} has integer entries if and only if $\det(A) = \pm 1$.

Proof. It can be proved from the Definition 3.1.1 of the determinant of a matrix with integer entries is also an integer. One can see this via induction on dimension with Laplace expansion, or using the permutation expression of the determinant—the point is that the determinant can be written as sums of products of the entries of the matrix, and sums and products of integers are still integers.

So, if A^{-1} has integer entries, then both $\det(A)$ and $\det(A^{-1}) = \det(A)^{-1}$ are integers, which is only possible if $\det(A) = \pm 1$. Conversely, if $\det(A) = \pm 1$, then $A^{-1} = \det(A)\mathrm{adj}(A) = \pm\mathrm{adj}(A)$. But the entries of $\mathrm{adj}(A)$ are cofactors of A, which are, up to sign, determinants of $n-1 \times n-1$ matrices with integer entries. Therefore $\mathrm{adj}(A)$ has integer entries, so A^{-1} does as well. $\square$