3 Systems of Linear Equations

A *linear equation* is an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b \tag{1}$$

where $a_1, \ldots, a_n, b \in \mathbb{R}$ and $x_1, \ldots, x_n$ are *unknowns* or *variables*. We say that this equation has n unknowns.

A *system of linear equations* is a system of the form

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

with the a_{ij} and b_is all real numbers. We say this is a system of m equations in n unknowns.

Importantly, each equation sets a linear combination of the variables equal to some number; we aren't allowed to multiply variables together, or do anything else fancy with them. We will see later that this allows us to use theorems about vector spaces to solve systems of linear equations.

An element $(x_1, \ldots, x_n) \in \mathbb{R}^n$ is a *solution* to a system of linear equations if all of the equalities hold for that collection of x_i. The *solution set* of a system of linear equations is the set of all solutions, and we say two systems are equivalent if they have the same solution sets.

We say that two systems of equations are *equivalent* if they have the same set of solutions. Thus the process of solving a system of equations is mostly the process of converting a system into an equivalent system that is simpler.

There are three basic operations we can perform on a system of equations to get an equivalent system:

1. We can write the equations in a different order.
2. We can multiply any equation by a nonzero scalar.
3. We can add a multiple of one equation to another.

All three of these operations are guaranteed not to change the solution set; proving this is a reasonable exercise. Our goal now is to find an efficient way to use these rules to get a useful solution to our system.

3.1 The matrix of a system

We need an efficient way to represent the array of numbers given by the a_{ij} and the b_i. We see these numbers are naturally laid out in a rectangular grid.

Definition 3.1. A *(real) matrix* is a rectangular array of (real) numbers. A matrix with m rows and n columns is *a $m \times n$ matrix*, and we notate the set of all such matrices by $M_{m\times n}$.

A $m \times n$ matrix is *square* if $m = n$, that is, it has the same number of rows as columns. We will sometimes represent the set of $n \times n$ square matrices by M_n.

We will generally describe the elements of a matrix with the notation

$$(a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

We can then take the information from a system of linear equations and encode it in a matrix. Right now, we will just use this as a convenient notational shortcut; we will see later on in the course that this has a number of theoretical and practical advantages.

Definition 3.2. The *coefficient matrix* of a system of linear equations given by

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

and the *augmented coefficient matrix* is

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right].$$

Example 3.3. Suppose we have a system

$$\begin{aligned} a + 4b + 7c &= 0 \\ 2a + 5b + 8c &= 0 \\ 3a + 6b + 9c &= 0. \end{aligned}$$

Then the coefficient matrix is

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

and the augmented coefficient matrix is

$$\left[\begin{array}{ccc|c} 1 & 4 & 7 & 0 \\ 2 & 5 & 8 & 0 \\ 3 & 6 & 9 & 0 \end{array}\right]$$

Earlier we listed three operations we can perform on a system of equations without changing the solution set: we can reorder the equations, multiply an equation by a nonzero scalar, or add a multiple of one equation to another. We can do analogous things to the coefficient matrix.

Definition 3.4. The three *elementary row operations* on a matrix are

I Interchange two rows.

II Multiply a row by a nonzero real number.

III Replace a row by its sum with a multiple of another row.

Example 3.5. What can we do with our previous matrix? We can

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \stackrel{I}{\to} \begin{bmatrix} 2 & 5 & 8 \\ 1 & 4 & 7 \\ 3 & 6 & 9 \end{bmatrix} \stackrel{II}{\to} \begin{bmatrix} 2 & 5 & 8 \\ 2 & 8 & 14 \\ 3 & 6 & 9 \end{bmatrix} \stackrel{III}{\to} \begin{bmatrix} 2 & 5 & 8 \\ 0 & 3 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

So how do we use this to solve a system of equations? The basic idea is to remove variables from successive equations until we get one equation that contains only one variable—at which point we can substitute for that variable, and then the others. To do that with this matrix, we have

$$\left[\begin{array}{ccc|c} 1 & 4 & 7 & 0 \\ 2 & 5 & 8 & 0 \\ 3 & 6 & 9 & 0 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 4 & 7 & 0 \\ 0 & -3 & -6 & 0 \\ 3 & 6 & 9 & 0 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 4 & 7 & 0 \\ 0 & -3 & -6 & 0 \\ 0 & -6 & -12 & 0 \end{array}\right]$$
$$\stackrel{II}{\to} \left[\begin{array}{ccc|c} 1 & 4 & 7 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & -6 & -12 & 0 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 4 & 7 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \left(\stackrel{II}{\to} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \right).$$

What does this tell us? That our system of equations is equivalent to the system

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

You'll notice that this matches what we got on homework 3.

Example 3.6. Solve the system of equations

$$\begin{aligned} a + 2b + c &= 3 \\ 3a - b - 3c &= -1 \\ 2a + 3b + c &= 4. \end{aligned}$$

This system has augmented coefficient matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 2 & 3 & 1 & 4 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & -1 & -1 & -2 \end{array}\right]$$
$$\stackrel{II}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & 1 & 1 & 2 \end{array}\right] \stackrel{I}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & -7 & -6 & -10 \end{array}\right] \stackrel{III}{\to} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

which gives us the system

$$\begin{aligned} a + 2b + c &= 3 \\ b + c &= 2 \\ c &= 4. \end{aligned}$$

The last equation tells us $c = 4$, which then gives $b = -2$ and $a = 3$. We can check that this solves the system.

3.2 Row Echelon Form

We want to solve systems of linear equations, using these matrix operations. We want to be somewhat more concrete about our goals: what exactly would it look like for a system to be solved?

Definition 3.7. A matrix is in *row echelon form* if

- Every row containing nonzero elements is above every row containing only zeroes; and
- The first (leftmost) nonzero entry of each row is to the right of the first nonzero entry of the above row.

Remark 3.8. Some people require the first nonzero entry in each nonzero row to be 1. This is really a matter of taste and doesn't matter much, but you should do it to be safe; it's an easy extra step to take by simply dividing each row by its leading coefficient.

Example 3.9. The following matrices are all in Row Echelon Form:

$$\begin{bmatrix} 1 & 3 & 2 & 5 \\ 0 & 3 & -1 & 4 \\ 0 & 0 & -2 & 3 \end{bmatrix} \quad \begin{bmatrix} 5 & 1 & 3 & 2 & 8 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -7 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 5 \\ 0 & -2 & 3 \\ 0 & 0 & 7 \end{bmatrix}.$$

The following matrices are not in Row Echelon Form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 2 & 5 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 5 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

Definition 3.10. The process of using elementary row operations to transform a system into row echelon form is *Gaussian elimination*.

A system of equations sometimes has a solution, but does not always. We say a system is *inconsistent* if there is no solution; we say a system is *consistent* if there is at least one solution.

Example 3.11. Consider the system of equations given by

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\ -1x_1 + -1x_2 + x_5 &= -1 \\ -2x_1 + -2x_2 + 3x_5 &= 1 \\ x_3 + x_4 + 3x_5 &= -1 \\ x_1 + x_2 + 2x_3 + 2x_4 + 4x_5 &= 1. \end{aligned}$$

This translates into the augmented matrix

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

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

We see that the final two equations are now $0 = -4$ and $0 = -3$,s othe system is inconsistent.

Example 3.12. Let's look at another system that is almost the same.

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\ -1x_1 + -1x_2 + x_5 &= -1 \\ -2x_1 + -2x_2 + 3x_5 &= 1 \\ x_3 + x_4 + 3x_5 &= 3 \\ x_1 + x_2 + 2x_3 + 2x_4 + 4x_5 &= 4. \end{aligned}$$

This translates into the augmented matrix

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

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

We see this system is now consistent. Our three equations are

$$x_1 + x_2 + x_3 + x_4 + x_5 = 1 \qquad x_3 + x_4 + 2x_5 = 0 \qquad x_5 = 3.$$

Via back-substitution we see that we have

$$x_5 = 3 \qquad x_3 + x_4 = -6 \qquad x_1 + x_2 = 4.$$

Thus we could say the set of solutions is $\{(\alpha, 4-\alpha, \beta, -6-\beta, 3)\} \subseteq \mathbb{R}^5$.

What we were just doing definitely worked, but even after we finished transforming the matrix we still needed to do some more work. So we'd like to reduce the matrix even further until we can just read the answer off from it.

Definition 3.13. A matrix is in *reduced row echelon form* if it is in row echelon form, and the first nonzero entry in each row is the only entry in its column.

This means that we will have some number of columns that each have a bunch of zeroes and one 1. Other than that we may or may not have more columns, which can contain basically anything; we've used up all our degrees of freedom to fix those columns that contain the leading term of some row.

Note that the columns we have fixed are not necessarily the first columns, as the next example shows.

Example 3.14. The following matrices are all in reduced Row Echelon Form:

$$\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix} \quad \begin{bmatrix} 1 & 17 & 0 & 2 & 8 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The following matrices are not in reduced Row Echelon Form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 3 & 0 & 3 \\ 0 & 0 & 2 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 15 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example 3.15. Let's solve the following system by putting the matrix in reduced row echelon form.

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 2 \\ x_1 + x_2 + x_3 + 2x_4 + 2x_5 &= 3 \\ x_1 + x_2 + x_3 + 2x_4 + 3x_5 &= 2 \end{aligned}$$

We have

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

From this we can read off the solution $x_1 + x_2 + x_3 = 1, x_4 = 2, x_5 = -1$. Thus the set of solutions is $\{(1 - \alpha - \beta, \alpha, \beta, 2, -1)\}$.

We say some systems of equations are "overdetermined", which means that there are more equations than varaibles. Overdetermined equations are "usually" inconsistent, but not always—they can be consistent when some of the equations are redundant.

Example 3.16. The system

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 1 \\ 2x_1 - x_2 + x_3 &= 2 \\ 4x_1 + 3x_2 + 3x_3 &= 4 \\ 2x_1 - x_2 + 3x_3 &= 5 \end{aligned}$$

gives the matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 2 & -1 & 1 & 2 \\ 4 & 3 & 3 & 4 \\ 2 & -1 & 3 & 5 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -5 & -1 & 0 \\ 0 & -5 & -1 & 0 \\ 0 & -5 & 1 & 3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -5 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 3 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & 1 & 1/5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3/2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 3/5 & 1 \\ 0 & 1 & 1/5 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3/2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1/10 \\ 0 & 1 & 0 & -3/10 \\ 0 & 0 & 1 & 3/2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This gives us the solution $x_1 = 1/10, x_2 = -3/10, x_3 = 3/2$, which you can go back and check solves the original system.

This overdetermined system does have a solution, but only because two of the equations were redundant, as we could see in the second matrix where two lines are identical. In fact we can go back to the original set of equations, and see that if we add two times the first equation to the second equation, we get the third—which is the redundancy.

Other systems of equations are "underdetermined", which means there are more variables than equations. These systems are usually but not always consistent.

Example 3.17. Let's consider the system

$$\begin{aligned} -x_1 + x_2 - x_3 + 3x_4 &= 0 \\ 3x_1 + x_2 - x_3 - x_4 &= 0 \\ 2x_1 + x_2 - 2x_3 - x_4 &= 0. \end{aligned}$$

This gives us the matrix

$$\left[\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 3 & 1 & -1 & -1 & 0 \\ 2 & 1 & -2 & -1 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & -1 & 1 & -3 & 0 \\ 0 & 4 & -4 & 8 & 0 \\ 0 & 3 & -4 & 5 & 0 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & -1 & 1 & -3 & 0 \\ 0 & 1 & -1 & 2 & 0 \\ 0 & 3 & -4 & 5 & 0 \end{array}\right]$$

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

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

We see that we can't "simplify" the fourth column in any way; we don't have any degrees of freedom after we fix the first three columns. This means that we can pick x_4 to be anything we want, and the other variables are given by $x_1 - x_4 = 0, x_2 - 3x_4 = 0, x_3 + x_4 = 0$. Thus the set of solutions is $\{(\alpha, 3\alpha, -\alpha, \alpha)\}$.

Remark 3.18. A system of any size can be either consistent or inconsistent. $0 = 1$ is an inconsistent system with one equation, and

$$\begin{aligned} x_1 + \cdots + x_{100} &= 0 \\ x_1 + \cdots + x_{100} &= 1 \end{aligned}$$

is an inconsistent system with a hundred variables and only two equations. In contrast,

$$\begin{aligned} x_1 &= 1 \\ x_1 &= 1 \\ &\vdots \\ x_1 &= 1 \end{aligned}$$

has only one variable, and many equations, and is still consistent.

3.3 Matrix Algebra

3.3.1 The vector space $M_{m\times n}$

Definition 3.19. If $A = (a_{ij})$ and $B = (b_{ij})$ are $m \times n$ matrices, and $r \in \mathbb{R}$, then we define matrix scalar multiplication by

$$rA = (ra_{ij}) = \begin{bmatrix} ra_{11} & ra_{12} & \dots & ra_{1n} \\ ra_{21} & ra_{22} & \dots & ra_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ra_{m1} & ra_{m2} & \dots & ra_{mn} \end{bmatrix}.$$

We define matrix addition by

$$A + B = (a_{ij} + b_{ij}) = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix}.$$

Fact 3.20. *Under these operations of addition and scalar multiplication, the set $M_{m\times n}$ is a vector space, with zero vector given by*

$$\mathbf{0} = (0) = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

I'm not going to prove this, but you can see that it should be true for the same reason $\mathbb{R}^{mn}$ is a vector space: they're both just lists of real numbers, but one is arranged in a column and the other in a rectangle. The operations are the same.

Definition 3.21. We define the standard basis matrices E_{ij} to be the matrices with a 1 in the ij space and zeroes in every other space.

In $M_{4\times 3}$ we have

$$E_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad E_{41} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

The set $\{E_{ij} : 1 \leq i \leq m, j \leq 1 \leq n\}$ is the standard basis for $M_{m\times n}$.

All of this is fairly standard, but also fairly boring; as a vector space, $M_{m\times n}$ really is just $\mathbb{R}^{mn}$ written in a different order. The interesting aspect of matrices comes from the ability to multiply them.

3.3.2 Matrix Multiplication

Definition 3.22. If $A \in M_{\ell\times m}$ and $B \in M_{m\times n}$, then there is a matrix $AB \in M_{\ell\times n}$ whose ij element is

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.$$

If you're familiar with the dot product, you can think that the ij element of AB is the dot product of the ith row of A with the jth column of b.

Note that A and B don't have to have the same dimension! Instead, A has the same number of columns that B has rows. The new matrix will have the same number of rows as A and the same number of columns as B.

Example 3.23.

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 5 & -1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1\cdot 5 + 3\cdot 3 & 1\cdot(-1) + 3\cdot 2 \\ 2\cdot 5 + 4\cdot 3 & 2\cdot(-1) + 4\cdot 2 \end{bmatrix} = \begin{bmatrix} 14 & 5 \\ 22 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 6 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 5 \\ 4 & 1 & 6 \end{bmatrix} = \begin{bmatrix} 4\cdot 3 + 6\cdot 4 & 4\cdot 1 + 6\cdot 1 & 4\cdot 5 + 6\cdot 6 \\ 2\cdot 3 + 1\cdot 4 & 2\cdot 1 + 1\cdot 1 & 2\cdot 5 + 1\cdot 6 \end{bmatrix} = \begin{bmatrix} 36 & 10 & 56 \\ 10 & 3 & 16 \end{bmatrix}.$$

Matrix multiplication is *associative*, by which we mean that $(AB)C = A(BC)$.

Matrix multiplication is not commutative: in general, it's not even the case that AB and BA both make sense. If $A \in M_{3\times 4}$ and $B \in M_{4\times 2}$ then AB is a 3×2 matrix, but BA isn't a thing we can compute. But even if AB and BA are both well-defined, they are not equal.

Example 3.24.

$$\begin{bmatrix} 3 & 5 & 1 \\ -2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 5 \cdot 1 + 1 \cdot 4 & 3 \cdot 1 + 5 \cdot 3 + 1 \cdot 1 \\ -2 \cdot 2 + 0 \cdot 1 + 2 \cdot 4 & -2 \cdot 1 + 0 \cdot 3 + 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 15 & 19 \\ 4 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 3 & 5 & 1 \\ -2 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 3 + 1 \cdot (-2) & 2 \cdot 5 + 1 \cdot 0 & 2 \cdot 1 + 1 \cdot 2 \\ 1 \cdot 3 + 3 \cdot (-2) & 1 \cdot 5 + 3 \cdot 0 & 1 \cdot 1 + 3 \cdot 2 \\ 4 \cdot 3 + 1 \cdot (-2) & 4 \cdot 5 + 1 \cdot 0 & 4 \cdot 1 + 1 \cdot 2 \end{bmatrix} = \begin{bmatrix} 4 & 10 & 4 \\ -3 & 5 & 7 \\ 10 & 20 & 6 \end{bmatrix}.$$

Particularly nice things happen when our matrices are square. Any time we have two $n \times n$ matrices we can multiply them by each other in either order (though we will still get different things each way!).

Example 3.25.

$$\begin{bmatrix} 4 & 1 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ 8 & -13 \end{bmatrix}$$

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

3.3.3 Transposes

Definition 3.26. If A is a $m \times n$ matrix, then we can form a $n \times m$ matrix B by flipping A across its diagonal, so that $b_{ij} = a_{ji}$. We say that B is the *transpose* of A, and write $B = A^T$.

If $A = A^T$ we say that A is *symmetric*. (Symmetric matrices must always be square).

Example 3.27.

$$\text{If } A = \begin{bmatrix} 1 & 3 & 5 \\ -1 & 4 & 2 \end{bmatrix} \text{ then } A^T = \begin{bmatrix} 1 & -1 \\ 3 & 4 \\ 5 & 2 \end{bmatrix}.$$

$$\text{If } B = \begin{bmatrix} 5 & 3 \\ 3 & -2 \end{bmatrix} \text{ then } B^T = \begin{bmatrix} 5 & 3 \\ 3 & -2 \end{bmatrix}$$

and thus B is symmetric.

Fact 3.28.

- $(A^T)^T = A$.
- $(A + B)^T = A^T + B^T$.
- $(rA)^T = rA^T$.
- *If* $A \in M_{\ell \times m}$ *and* $B \in M_{m \times n}$ *then* $(AB)^T = B^T A^T$.

3.3.4 Matrices and Systems of Equations

We will do a lot with matrices in the future (a linear algebra class that doesn't cover general vector spaces is often called a matrix algebra class). In the current context we mostly want it to make it easier to talk about systems of equations.

Let

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

be a system of linear equations. Then $A = (a_{ij})$ is its coefficient matrix, and $\mathbf{b} = (b_1, \dots, b_m)$ is a vector in $\mathbb{R}^m$. If we take $\mathbf{x} = (x_1, \dots, x_n)$ to be a variable vector in $\mathbb{R}^n$, we can rewrite our linear system as the equation

$$A\mathbf{x} = \mathbf{b}$$

which is certainly much easier to write down.

3.4 Homogeneous systems and subspaces

Definition 3.29. A system of linear equations $A\mathbf{x} = \mathbf{b}$ is called *homogeneous* if $\mathbf{b} = \mathbf{0}$, that is, if all of the constant terms are zero. Otherwise, it is *non-homogeneous*.

It's pretty clear that every homogeneous system has at least one solution: the solution where every variable is equal to zero. You can also see, by playing around with the algebra, that if you add two solutions together you get another solution. And if you multiply a solution by a scalar, you get another solution. This list of properties should seem familiar.

Proposition 3.30. *If $A\mathbf{x} = \mathbf{0}$ is a homogeneous system of linear equations, and U is the set of solutions to this system, then U is a subspace of $\mathbb{R}^n$.*

Proof. See Homework 5. □

Definition 3.31. For a given matrix A, the subspace of solutions to the equation $A\mathbf{x} = \mathbf{0}$ is called the *nullspace* $N(A)$ or the *kernel* $\ker(A)$ of the matrix A. The dimension of the nullspace is the *nullity* of A.

Example 3.32. Find a basis for the null space of $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix}$.

We row reduce the matrix

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

We see that x_3 and x_4 are fixed variables, and x_1, x_2 are determined by x_3 and x_4. (You could of course do this the other way around). Then we have $x_1 = x_3 - x_4$ and $x_2 = x_4 - 2x_3$.

Thus $N(A) = \{(\alpha - \beta, \beta - 2\alpha, \alpha, \beta)\} = \{\alpha(1, -2, 1, 0) + \beta(-1, 1, 0, 1)\}$, so a basis for the space is $B = \{(1, -2, 1, 0), (-1, 1, 0, 1)\}$.

(We can check that these two elements are both in the nullspace; we will see soon how to predict the dimension of the nullspace).

In contrast, the set of solutions to a non-homogeneous system is never a subspace. The easy way to see this is that if all the variables are zero, all the constants must be zero as well; thus the set of solutions to a non-homogeneous system never includes the zero vector.

But the solution set to a non-homogeneous system is almost a subspace in a very specific way.

Proposition 3.33. *Suppose $A\mathbf{x} = \mathbf{b}$ is a non-homogeneous linear system.*

If $U = N(A))$ and $\mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{b}$, then the set of solutions to the system $A\mathbf{x} = \mathbf{b}$ is the set

$$N(A) + \mathbf{x}_0 = \{\mathbf{y} + \mathbf{x}_0 : \mathbf{y} \in N(A)\}.$$

Proof. We want to show that two sets are equal, so we show that each is a subset of the other.

First, suppose that $\mathbf{x}_1$ is a solution to $A\mathbf{x}_1 = \mathbf{b}$. Then we have

$$\begin{aligned} b &= A\mathbf{x}_0 \\ b &= A\mathbf{x}_1 \\ b - b &= A\mathbf{x}_1 - A\mathbf{x}_0 = A(\mathbf{x}_1 - \mathbf{x}_0) \\ \mathbf{0} &= A(\mathbf{x}_1 - \mathbf{x}_0). \end{aligned}$$

Thus $\mathbf{y} = \mathbf{x}_1 - \mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{0}$, and then $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{y}$ for some $\mathbf{y} \in U$.

Conversely, suppose $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{y}$ for some $\mathbf{y} \in U$. Then

$$A\mathbf{x}_1 = A(\mathbf{x}_0 + \mathbf{y}) = A\mathbf{x}_0 + A\mathbf{y} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Thus x_1 is a solution to $A\mathbf{x} = \mathbf{b}$. □

Remark 3.34. Notice this did not depend on the matrices specifically; it only depends on the ability to distribute the matrix across sums of vectors. Operations with this property are called "linear" and we will discuss them in much more detail in section 4.

Example 3.35. Let's find a set of solutions to the system

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 + 2x_2 + 3x_3 &= 6 \\ 2x_1 + 3x_2 + 4x_2 &= 9. \end{aligned}$$

Gaussian elimination gives

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

Taking $x_3 = \alpha$ as a free variable, our solution set is $\{(\alpha, 3-2\alpha, \alpha)\} = \{(0,3,0)+\alpha(1,-2,1)\}$. Indeed, we see that this set corresponds to elements of the vector space spanned by $\{(1,-2,1)\}$, plus a specific solution $(0,3,0)$.

Alternatively, we could have solved the homogeneous system first, and seen that the solution was $x_1 - x_3 = 0, x_2 + 2x_3 = 0$ giving us the subspace spanned by $\{(1,-2,1)\}$. Then we just need to find a solution; to my eyes the obvious solution is $(1,1,1)$. So our theorem tells us that the solution set is $\{(1,1,1) + \alpha(1,-2,1)\}$. This may not *look* like the solution we got before, but it is in fact the same set, since $(1,1,1) = (0,3,0) + (1,-2,1)$.

Example 3.36. Now consider the system

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 + 2x_2 + 3x_3 &= 3 \\ 2x_1 + 3x_2 + 4x_2 &= 3. \end{aligned}$$

It's easy enough to see that this system has no solutions, since the sum of the first two equations should be the third.

The empty set isn't a vector space plus a vector, since every vector space contains the zero vector. But this doesn't violate our theorem, since our theorem assumes that a solution exists; no matter what the homogeneous system looks like, it's always possible for the non-homogeneous system to have no solutions at all if it contradicts itself. Our theorem only tells us that *if* any solution exists, then the homogeneous system tells us how many solutions exist.

Example 3.37. Let's find the set of solutions to

$$\begin{aligned} x + y + z &= 0 \\ x - 2y + 2z &= 4 \\ x + 2y - z &= 2. \end{aligned}$$

We form the matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 1 & -2 & 2 & 4 \\ 1 & 2 & -1 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -3 & 1 & 4 \\ 0 & 1 & -2 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 2 \\ 0 & -3 & 1 & 4 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & -5 & 10 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -2 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|c} 1 & 0 & 3 & -2 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

giving us the sole solution $x_1 = 4, x_2 = -2, x_3 = -2$.

If we look at the corresponding homogeneous system, we see that we can reduce the matrix to $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and thus the sole solution to the homogeneous system of equations is $x_1 = x_2 = x_3 = 0$. This is indeed a vector space; in fact, it is the trivial vector space. Then every solution to our non-homogeneous system is a solution to our homogeneous system plus some element of the trivial vector space; since there is only one vector in the trivial vector space, there is only one solution to our system.

3.5 Row space and column space

We've seen that our theorem is only helpful if a solution exists, so we'd like to know when solutions to our system exist at all. The concepts of the row space and column space allow us to determine this neatly.

Definition 3.38. If $A = (a_{ij})$ is a $m \times n$ matrix, then each row can be viewed as a vector in $\mathbb{R}^n$; we call these vectors the *row vectors* of A. We may notate them as $\mathbf{r}_i = (a_{i1}, a_{i2}, \ldots, a_{in})$.

Similarly, we can view each column as vector in $\mathbb{R}^m$, and we call these the *column vectors* of A. We may notate them as $\mathbf{c}_j = (a_{1j}, a_{2j}, \ldots, a_{mj})$

Thus each matrix gives us two sets of vectors. We can look at these vectors and see which vector spaces they span.

Definition 3.39. If A is a $m \times n$ matrix, we say that the span of the row vectors of A is the *row space* of A, which we will sometimes denote $row(A)$. It is a subspace of $\mathbb{R}^n$. The dimension of the row space is the *rank* of A, denoted $\mathrm{rk}(A)$.

The span of the column vectors of A is the *column space* of A, sometimes denoted $col(A)$.

The concept of column space allows us to answer the question we just asked ourselves.

Proposition 3.40. *Let A be a $m \times n$ matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$. Then $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in $col(A)$.*

Proof. The system $A\mathbf{x} = \mathbf{b}$ is the same as the system

$$\begin{aligned}
a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \qquad\qquad & \quad \vdots \\
a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m
\end{aligned}$$

which we can rewrite as

$$x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \cdots + x_n\mathbf{c}_n = \mathbf{b}.$$

Thus the equation has a solution precisely when $\mathbf{b}$ is in the span of the $\mathbf{c}_i$, which is the column space of A by definition. □

Corollary 3.41. *the system $A\mathbf{x} = \mathbf{b}$ has a solution for every $b \in \mathbb{R}^m$ if and only if $col(A) = \mathbb{R}^m$, that is, the column vectors span $\mathbb{R}^m$.*

The system has a unique solution if and only if the column vectors are linearly independent.

Proof. $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^m$ if and only if every $\mathbf{b} \in \mathbb{R}^m$ is in the column space, that is, if the column vectors span $\mathbb{R}^m$.

The column vectors are linearly independent if and only if every vector in their span can be represented uniquely as a linear combination of the column vectors. □

This doesn't help that much though, since it doesn't actually give us a way to figure out anything about the column space. To learn about that, we shift to looking at the row space, which is somewhat easier to understand.

Proposition 3.42. *Two row-equivalent matrices have the same row space.*

Proof. We need to check that each elementary row operation doesn't change the span of the set of vectors.

I. (Switch two rows) Switching the order of two vectors does not affect the span at all.

II. (Multiply a row by a nonzero scalar) Multiplying a vector by a non-zero scalar won't change the span of the set of vectors, since in any linear combination we can always just multiply the relevant coefficient by the inverse of our non-zero scalar.

III. (Add a multiple of one row to another) This won't add anything to the span, since a linear combination of the new vectors will still be a linear combination of the old vectors.

   This won't lose anything from the span, since we can undo the row operation, and so every old vector is a linear combination of new vectors.

□

Corollary 3.43. *Suppose A is a $m \times n$ matrix and A_R is the matrix obtained by using Gauss-Jordan elimination to reduce it to reduced row echelon form. Then the non-zero rows of A_R form a basis for the row space of A.*

Proof. The non-zero rows of A_R are clearly linearly independent, since each one has a 1 in a column where every other row has a zero. Thus the non-zero rows of A_R form a basis for the space they span, which is the rowspace of A_R. But we just saw that A_R and A have the same rowspace, so they form a basis for the rowspace of A. □

Example 3.44. Find a basis for the rowspace of $\begin{bmatrix} 1 & 5 & -9 & 11 \\ -2 & -9 & 15 & -21 \\ 3 & 17 & -30 & 36 \\ -1 & 2 & -3 & -1 \end{bmatrix}$

$$\begin{bmatrix} 1 & 5 & -9 & 11 \\ -2 & -9 & 15 & -21 \\ 3 & 17 & -30 & 36 \\ -1 & 2 & -3 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 5 & -9 & 11 \\ 0 & 1 & -3 & 1 \\ 0 & 2 & -3 & 3 \\ 0 & 7 & -12 & 10 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 6 & 6 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 9 & 3 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 6 & 6 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 9 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So a basis for $row(A)$ is $\{(1,0,0,4),(0,1,0,2),(0,0,1,1/3)\}$. The matrix has rank 3.

Remark 3.45. We can use this to find a "simple" basis for any vector space we have a spanning set for: write a matrix with our spanning set as rows, and row-reduce it until we have a basis.

Theorem 3.46 (Rank-Nullity). *If $A \in M_{m\times n}$ then rank of A plus nullity of A equals n.*

Proof. If U is the reduced row echelon form of A, then $A\mathbf{x} = \mathbf{0}$ is equivalent to $U\mathbf{x} = \mathbf{0}$. Since the matrix has rank r, the matrix U will have r nonzero rows and $n - r$ zero rows; thus it will have $n - r$ free variables and r lead variables.

The dimension of $N(A)$ is equal to the number of free variables, and thus to $n - r$. □

We have managed to relate the rank and the nullity, but we still want to know about the column space. But the column space is tied to the row space in a fundamental way.

Proposition 3.47. *If A is a $m \times n$ matrix, the dimension of the row space of A equals the dimension of the column space of A.*

Proof. We will use a trick with the transpose matrix, since the rows of A are the columns of A^T and vice versa. We will prove that the dimension of the column space of a matrix is at least as great as the dimension of the row space. But since this result will also hold for the transpose matrix, this gives us our answer.

Suppose A has rank r, and let U be the row echelon form of A. It will have r leading 1s, and the columns containing the leading 1s will be linearly independent. (They do not form a basis for the column space, since we have no reason to believe that the row operations preserve the span of the *columns*).

Let U_L be the matrix obtained by deleting the columns of U corresponding to free variables, leaving only the columns that contain a leading 1. Delete the same columns from A, and call the resulting matrix A_L.

The matrices U_L and A_L are row-equivalent, so $A_L\mathbf{x} = \mathbf{0}$ if and only if $U_L\mathbf{x} = \mathbf{0}$, and since the columns of U_L are linearly independent, this happens if and only if $\mathbf{x} = \mathbf{0}$. Thus we see that the columns of A_L are linearly independent. We know that A_L will have exactly r columns, so the column space contains at least r linearly independent vectors, and so the dimension of the column space is at least r. Thus $\dim(col(A)) \geq \dim(row(A)) = r$.

Now consider the matrix A^T. By the previous result, $\dim(col(A^T)) \geq \dim(row(A^T))$. But we know that $col(A^T) = row(A)$ and $row(A^T) = col(A)$, so this tells us that $\dim(row(A)) \geq \dim(col(A))$, which combined with the previous result gives us that $\dim(row(A)) = \dim(col(A))$. □

Corollary 3.48. *Let A be a $m \times n$ matrix, and let U be the reduced row echelon form of A. Then the columns of A corresponding to columns of U that contain a leading "1" form a basis for the columnspace of A.*

Proof. We just showed that these columns are linearly independent, and there are r of them. Thus they are a basis. □

Remark 3.49. Note that the columns of U do not (usually) span the column space of A! But looking at U tells us which columns we should take to find a basis for the column space.

Note that we could also find a basis for the column space by simply taking A^T, row reducing it, and finding a basis for the rowspace of A^T.

Example 3.50. Find a basis for the column space of $\begin{bmatrix} 1 & 5 & -9 & 11 \\ -2 & -9 & 15 & -21 \\ 3 & 17 & -30 & 36 \\ -1 & 2 & -3 & -1 \end{bmatrix}$

We saw that the reduced row echelon form of this matrix has leading ones in the first three columns. So the first three columns form a basis for the columnspace, and thus a basis is $\{(1, -2, 3, -1), (5, -9, 17, 2), (-9, 15, -30, -3)\}$.

Example 3.51. Find bases for the row, column, and nullspace of $\begin{bmatrix} 1 & -2 & 1 & 1 & 2 \\ -1 & 3 & 0 & 2 & -2 \\ 0 & 1 & 1 & 3 & 4 \\ 1 & 2 & 5 & 13 & 5 \end{bmatrix}$

We first row reduce the matrix.

$$\begin{bmatrix} 1 & -2 & 1 & 1 & 2 \\ -1 & 3 & 0 & 2 & -2 \\ 0 & 1 & 1 & 3 & 4 \\ 1 & 2 & 5 & 13 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 & 1 & 2 \\ 0 & 1 & 1 & 3 & 0 \\ 0 & 1 & 1 & 3 & 4 \\ 0 & 4 & 4 & 12 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 3 & 7 & 2 \\ 0 & 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$
$$\to \begin{bmatrix} 1 & 0 & 3 & 7 & 2 \\ 0 & 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 3 & 7 & 0 \\ 0 & 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

To find the rowspace, we just take these rows; so a basis for the rowspace is $\{(1,0,3,7,0),(0,1,1,3,0),(0,0,0,0,1)\}$. Thus the rank of the matrix ix 3.

To find the columnspace, we look at the columns corresponding to those with leading 1s, which are the first, second, and fifth. Thus a basis for the columnspace is $\{(1,-1,0,1),(-2,3,1,2),(2,-2,4,5)\}$.

To find the nullspace, we see there are two free variables, which we set to be parameters $x_3 = \alpha, x_4 = \beta$. Then the nullspace is

$$\begin{aligned} \{(-3\alpha - 7\beta, -\alpha - 3\beta, \alpha, \beta, 0)\} &= \{(-3\alpha, -\alpha, \alpha, 0, 0) + (-7\beta, -3\beta, 0, \beta, 0)\} \\ &= \{\alpha(-3,-1,1,0,0) + \beta(-7,-3,0,1,0)\} \end{aligned}$$

so a basis for the nullspace is $\{(-3,-1,1,0,0),(-7,-3,0,1,0)\}$. The nullity is 2, which is what we expected from the rank-nullity theorem.

3.6 The identity matrix and matrix inverses

There is one more fundamental object of matrix algebra we would like to study.

Definition 3.52. For any n we define the *identity matrix* to be $I_n \in M_n$ to have a 1 on every diagonal entry, and a zero everywhere else. For example,

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If $A \in M_n$ then $I_n A = A = A I_n$. Thus it is a *multiplicative identity* in the ring of $n \times n$ matrices.

The identity matrix is symmetric (that is, $I_n^T = I_n$). It is already in reduced row echelon form, so it is easy to see that it has rank n and nullity 0.

Since we have a multiplicative inverse, which is the equivalent of "1" in our matrix algebra, we would like to define multiplicative inverses, the equivalent of reciprocals. The definition is not difficult to invent:

Definition 3.53. Let A and B be $n \times n$ matrices, such that $AB = I_n = BA$. Then we say that B is the *inverse* (or *multiplicative inverse*) of A, and write $B = A^{-1}$.

If such a matrix exists, we say that A is *invertible* or *nonsingular*. If no such matrix exists, we say that A is *singular*.

Example 3.54. The identity matrix I_n is its own inverse, and thus invertible.

The matrices

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1/10 & 2/5 \\ 3/10 & -1/5 \end{bmatrix}$$

are inverses to each other, as you can check.

Example 3.55. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ has no inverse, since

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$$

won't be the identity for any a, b, c, d. Thus this matrix is singular.

Remark 3.56. If $AB = I_n$ then $BA = I_n$. This isn't really trivial but we won't prove it.

As the last example shows, finding the inverse to a matrix is a matter of solving a big pile of linear equations at the same time (one for each coefficient of the inverse matrix). Fortunately, we just got good at solving linear equations. Even more fortunately, there's an easy way to organize the work for these problems.

Proposition 3.57. *Let A be a $n \times n$ matrix. Then if we form the augmented matrix $\begin{bmatrix} A & I_n \end{bmatrix}$, then A is invertible if and only if the reduced row echelon form of this augmented matrix is $\begin{bmatrix} I_n & B \end{bmatrix}$ for some matrix B, and furthermore $B = A^{-1}$.*

Proof. Let X be a $n \times n$ matrix of unknowns, and set up the system of equations implied by $AX = I_n$. This will be the same set of equations we are solving with this row reduction, and thus a matrix X exists if and only if this system has a solution, which happens if and only if the reduced row echelon form of $\begin{bmatrix} A & I_n \end{bmatrix}$ has no all-zero rows on the A side. $\square$

Corollary 3.58. *A $n \times n$ matrix A is invertible if and only if the rank of A is n, if and only if the nullity of A is 0.*

Thus A is invertible if and only if the rows are linearly independent, if and only if the columns are linearly independent, if and only if the rows span $\mathbb{R}^n$, if and only if the columns span $\mathbb{R}^n$.

Corollary 3.59. *A matrix $A \in M_n$ is invertible if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b} \in \mathbb{R}^n$.*

Example 3.60. Let's find an inverse for $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$.

We form and reduce the augmented matrix

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

Thus $A^{-1} = \begin{bmatrix} 1 & -2 & 5 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}$. We can check this by multiplying the matrices back together.

Example 3.61. Find the inverse of $B = \begin{bmatrix} 1 & 0 & 4 \\ 1 & 1 & 6 \\ -3 & 0 & -10 \end{bmatrix}$.

We form and reduce the augmented matrix

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

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

Thus $B^{-1} = \begin{bmatrix} -5 & 0 & -2 \\ -4 & 1 & -1 \\ 3/2 & 0 & 1/2 \end{bmatrix}$.

There are some more properties of inverse matrices we'd like to prove, but it turns out they are much easier to prove from the perspective of *functions*. We will discuss these functions in the next section.