Review of linear algebra

These notes discuss the most basic facts about linear algebra. All the material is covered in the Rutgers course Math 250, Introduction to Linear Algebra, and we will not go over it again in Math 252, but I hope that it will be helpful to have a brief summary here.

1. Systems of linear equations

Suppose we are given a system of m linear equations in n unknowns $x_1, \ldots, x_n$:

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

To write this in a more compact form we introduce a matrix and two vectors,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix},$$

so that (1) becomes $A\mathbf{x} = \mathbf{b}$. In the next section we turn to the problem of determining whether or not the system has any solutions and, if it does, of finding all of them. Before that, however, we make some general comments on consequences of linearity for the nature of solutions. As we develop the theory of linear systems of ODEs we will see results that are completely parallel to the results here for systems of linear equations.

The homogeneous problem. Suppose first that the system (1) is *homogeneous*, that is, that the right hand side is zero, or equivalently that $b_1 = b_2 = \cdots = b_m = 0$ or $\mathbf{b} = \mathbf{0}$:

$$A\mathbf{x} = \mathbf{0}. \tag{2}$$

Suppose further that we have found, by some method, two solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ of the equations; this means that $A\mathbf{x}_1 = \mathbf{0}$ and $A\mathbf{x}_2 = \mathbf{0}$. Then for any constants c and d, $\mathbf{x} = c\mathbf{x}_1 + d\mathbf{x}_2$ is also a solution, since

$$A\mathbf{x} = A(c\mathbf{x}_1 + d\mathbf{x}_2) = cA\mathbf{x}_1 + dA\mathbf{x}_2 = c \cdot \mathbf{0} + d \cdot \mathbf{0} = \mathbf{0}.$$

The argument extends to any number of solutions, and we have the

Principle 1: The linearity principle. If $\mathbf{x}_1$, $\mathbf{x}_2$, $\ldots$, $\mathbf{x}_k$ are all solutions of (2), and c_1, c_2, $\ldots$, c_k are constants, then

$$\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k \tag{3}$$

is also a solution of(2).

The name of this principle is related to the fact that (3) is called a *linear combination* (and sometimes, particularly by physicists studying quantum mechanics, a *linear superposition*) of the solutions $\mathbf{x}_1, \ldots, \mathbf{x}_k$. We will see later (see Principle 3 (iii) on page 8) that there is a special value of k such that (a) we can find a set of solutions $\mathbf{x}_1, \ldots, \mathbf{x}_k$ with the property that *every* solution of (1) can be built as a linear combination of these solutions, and (b) k different solutions are really needed for this to be true.

Notice also that **the homogeneous system always has at least one solution, the zero solution $\mathbf{x} = \mathbf{0}$**, since $A\mathbf{0} = \mathbf{0}$.

The inhomogeneous problem. Consider now the case in which the system (1) is *inhomogeneous*, that is, $\mathbf{b}$ is arbitrary. Suppose again that we are given two solutions, which we will now call $\mathbf{x}$ and $\mathbf{X}$, so that $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{X} = \mathbf{b}$. Then $\mathbf{x}_h = \mathbf{x} - \mathbf{X}$ is a solution of the *homogeneous* system (2), since

$$A\mathbf{x}_h = A(\mathbf{x} - \mathbf{X}) = A\mathbf{x} - A\mathbf{X} = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

This means that if we know *one* solution of our equations, X, then every other solution is obtained by $\mathbf{x} = \mathbf{X} + \mathbf{x}_h$. The converse also holds:

> **Principle 2: Extended linearity principle.** Every solution $\mathbf{x}$ of the system of inhomogeneous equations (1) is of the form $\mathbf{x} = \mathbf{X} + \mathbf{x}_h$, where $\mathbf{X}$ is some *particular* solution of the system, and $\mathbf{x}_h$ is a solution of the corresponding homogeneous system. Moreover, every vector of the form $\mathbf{x} = \mathbf{X} + \mathbf{x}_h$ is indeed a solution.

In particular, if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are the special solutions of the homogeneous equation referred to above and in Principle 3 (iii), then every solution $\mathbf{x}$ of (1) is of the form

$$\mathbf{x} = \mathbf{X} + c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k. \tag{4}$$

2. Row reduction and row-echelon form

The key technique that we will use for solving linear equations, and also for investigating general properties of the solutions, is the *reduction of a matrix to row-echelon form* or to *reduced row-echelon form* by the use of *elementary row operations*, a procedure often called *row reduction* or *Gaussian elimination*. Symbolically, if A is a matrix, we have

$$A \xrightarrow[\text{operations}]{\substack{\text{elementary}\\ \text{row}}} R$$

where R is in row-echelon or reduced row-echelon form. What does this all mean?

Row-echelon form: The matrix R is in row-echelon form (REF) if it satisfies three conditions:

(i) All nonzero rows (that is, rows with at least one nonzero entry) are above any zero rows (rows with all zeros).

(ii) The first nonzero entry in any nonzero row is a 1. This entry is called a *pivot*.

(iii) Each pivot lies to the right of the pivot in the row above it.

Here is a typical matrix in row-echelon form:

$$R = \begin{pmatrix} 0 & \mathbf{1} & 3 & -2 & 3 & 5 & 0 & 12 \\ 0 & 0 & 0 & \mathbf{1} & -2 & 0 & -15 & 5 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & -7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5}$$

The pivots are the entries, all with value 1, shown in boldface.

Warning: The text by Spence, Insel and Friedberg, used in Math 250, has a different definition of row-echelon form: the pivots are the first nonzero entries in the nonzero rows, but they are not required to have value 1. Unfortunately, both definitions are in common use.

Reduced row-echelon form: It is sometimes convenient to carry the reduction further, and bring the matrix into *reduced row-echelon form* (RREF). This form satisfies conditions (i)–(iii) above, and also

(iv) All matrix entries above a pivot are zero.

When the matrix R of (5) is put into reduced row-echelon form, it becomes

$$R' = \begin{pmatrix} 0 & \mathbf{1} & 3 & 0 & -1 & 0 & 0 & 2 \\ 0 & 0 & 0 & \mathbf{1} & -2 & 0 & 0 & 65 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 0 & 28 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{1} & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

One advantage, more theoretical than practical, is that the RREF of a matrix A is unique—whatever sequence of row operations is used to go from A to R, with R in RREF, the resulting R will be the same.

Elementary row operations: There are three elementary row operations on matrices:

R1. Interchange of two rows.

R2. Multiplication of a row by a nonzero scalar.

R3. Addition of a multiple of one row to another row.

By using these operations repeatedly we can bring any matrix into row echelon form. The procedure is illustrated on the next page.

Rank: If you do the row operations in different ways you can arrive at different REF matrices R from the same starting matrix A. However, all the REF matrices you find will have the same number of nonzero rows. The number of nonzero rows in R is called the *rank* of A, and written $\text{rank}(A)$ (it is also the rank of R, since R is already in REF).

Example 1: Row reduction

Here we carry out the reduction of a 3×4 matrix first to row-echelon, and then to reduced row-echelon, form. We indicate the row operations used by a simple notation: $\mathbf{r}_i$ denotes the i^{th} row of the matrix, and the row operations are denoted by $\mathbf{r}_i \leftrightarrow \mathbf{r}_j$ (interchange rows i and j), $\mathbf{r}_i \to c\,\mathbf{r}_i$ (multiply row i by the scalar c), and $\mathbf{r}_i \to \mathbf{r}_i + c\,\mathbf{r}_j$ (add c times row j to row i). Notice that in the first step we *must* switch the first row with another: because the first column is not identically zero, the first pivot must be in the upper left corner, and we need a nonzero entry there to get started.

$$\begin{pmatrix} 0 & -3 & -1 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 2 & 5 & -3 \end{pmatrix} \xrightarrow{\mathbf{r}_1 \leftrightarrow \mathbf{r}_2} \begin{pmatrix} \mathbf{1} & 2 & 3 & 0 \\ 0 & -3 & -1 & 1 \\ 2 & 2 & 5 & -3 \end{pmatrix}$$

$$\xrightarrow{\mathbf{r}_3 \to \mathbf{r}_3 - 2\,\mathbf{r}_1} \begin{pmatrix} \mathbf{1} & 2 & 3 & 0 \\ 0 & -3 & -1 & 1 \\ 0 & -2 & -1 & -3 \end{pmatrix}$$

$$\xrightarrow{\mathbf{r}_2 \to -(1/3)\,\mathbf{r}_2} \begin{pmatrix} \mathbf{1} & 2 & 3 & 0 \\ 0 & \mathbf{1} & 1/3 & -1/3 \\ 0 & -2 & -1 & -3 \end{pmatrix}$$

$$\xrightarrow{\mathbf{r3} \to \mathbf{r3} + 2\,\mathbf{r2}} \begin{pmatrix} \mathbf{1} & 2 & 3 & 0 \\ 0 & \mathbf{1} & 1/3 & -1/3 \\ 0 & 0 & -1/3 & -11/3 \end{pmatrix}$$

$$\xrightarrow{r_3 \to -3\,\mathbf{r}_3} \begin{pmatrix} \mathbf{1} & 2 & 3 & 0 \\ 0 & \mathbf{1} & 1/3 & -1/3 \\ 0 & 0 & \mathbf{1} & 11 \end{pmatrix}$$

This completes the reduction of A to row-echelon form. If we like, we can continue the process and reach reduced row-echelon form:

$$\xrightarrow[\mathbf{r}_2 \to \mathbf{r}_2 - (1/3)\,\mathbf{r}_3]{\mathbf{r}_1 \to \mathbf{r}_1 - 3\,\mathbf{r}_3} \begin{pmatrix} \mathbf{1} & 2 & 0 & -33 \\ 0 & \mathbf{1} & 0 & -4 \\ 0 & 0 & \mathbf{1} & 11 \end{pmatrix}$$

$$\xrightarrow{\mathbf{r}_1 \to \mathbf{r}_1 - 2\,\mathbf{r}_2} \begin{pmatrix} \mathbf{1} & 0 & 0 & -25 \\ 0 & \mathbf{1} & 0 & -4 \\ 0 & 0 & \mathbf{1} & 11 \end{pmatrix}$$

The extra steps for the reduction to reduced row-echelon form could also have been done at the same time as the earlier steps; for example, at the fourth step above, when we did $\mathbf{r}_3 \to \mathbf{r}_1 + 2\,\mathbf{r}_2$, we could also have done $\mathbf{r}_1 \to \mathbf{r}_1 - 2\,\mathbf{r}_2$ to leave the pivot as the only nonzero entry in column 2.

In the rest of these notes it is assumed that the reader knows what the row-echelon form and reduced row-echelon forms are and, given a matrix A, knows how to reduce it to row-echelon form and/or reduced row-echelon form. **Please review the definitions above and the procedure outlined in Example 1 to be sure that these concepts are clear.**

3. Solving systems of linear equations

Suppose now that we are given the system of linear equations (1) and want to determine whether or not it has any solutions and, if so, to find them all. The idea is to solve (1) by doing elementary operations on the equations, corresponding to the elementary row operations on matrices: interchange two equations, multiply an equation by a nonzero constant, or add a multiple of one equation to another. What is important is that these operations do not change the set of solutions of the equations, so that we can reduce the equations to simpler form, solve the simple equation, and know that we have found the all solutions of the original equations, but no extraneous ones. Moreover, instead of working with the equations, we can work with the *augmented matrix*:

$$(A\,|\,\mathbf{b}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & | & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & | & b_2 \\ \vdots & \vdots & \vdots & \vdots & | & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & | & b_m \end{pmatrix}.$$

(It is not really necessary to write the vertical bars here, but they remind us that the last column plays a special role.) Simplifying the original set of equations is equivalent to reducing the augmented matrix to REF or RREF. Once this is done, we can easily find the solutions explicitly, if there are any. Equally important, just by looking at the REF or RREF we can determine whether solutions exist and, if so, many of their properties. We will write this symbolically as

$$(A\,|\,\mathbf{b}) \xrightarrow[\text{operations}]{\substack{\text{elementary}\\ \text{row}}} (R\,|\,\mathbf{e})$$

The entire new augmented matrix $(R\,|\,\mathbf{e})$ is supposed to be in REF; this means that we have also reduced A to the REF matrix R.

Example 2: Suppose we want to solve the equations

$$\begin{aligned} -3x_2 - x_3 &= 1 \\ x_1 + 2x_2 + 3x_3 &= 0 \\ 2x_1 + 2x_2 + 5x_3 &= -3 \end{aligned} \tag{6}$$

The augmented matrix is the one we studied in the example in Example 1, so we already know a row-echelon form for it:

$$(A\,|\,\mathbf{b}) = \begin{pmatrix} 0 & -3 & -1 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 2 & 5 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} \mathbf{1} & 2 & 3 & 0 \\ 0 & \mathbf{1} & 1/3 & -1/3 \\ 0 & 0 & \mathbf{1} & 11 \end{pmatrix} = (R\,|\,\mathbf{e}).$$

The REF corresponds to the equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0 \\ x_2 + (1/3)x_3 &= -1/3 \\ x_3 &= 11 \end{aligned} \tag{7}$$

These may be solved by the process of "back substitution": solve first for x_3, substitute that value into the previous equation and solve for x_2, then substitute both values into the first equation to find x_1:

$$x_3 = 11, \quad x_2 = -1/2 - (1/3)x_3 = -4; \quad x_1 = -2x_2 - 3x_3 = -25, \quad \text{so} \quad \mathbf{x} = \begin{pmatrix} -25 \\ -4 \\ 11 \end{pmatrix}.$$

Notice that in this example the equations have a solution, and it is unique.

If we had used the reduced row-echelon form (which we also found in Example 1) we would have found the solution more quickly:

$$(\,A\,|\,\mathbf{b}\,) = \begin{pmatrix} 0 & -3 & -1 & 1 \\ 1 & 2 & 3 & 0 \\ 2 & 2 & 5 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} \mathbf{1} & 0 & 0 & -25 \\ 0 & \mathbf{1} & 0 & -4 \\ 0 & 0 & \mathbf{1} & 11 \end{pmatrix} = (\,R'\,|\,\mathbf{e}'\,).$$

The work we did in the first method, doing back substitution, is equivalent to the extra steps used to find the RREF in Example 1. Technically the first procedure—solving the system by finding the REF, then using back substitution—is called *Gaussian elimination*, and the second procedure is called *Gauss-Jordan elimination*, but we will not make this distinction, referring to either simply as Gaussian elimination.

In the next examples we will omit the step of row reduction and start with a matrix in reduced row-echelon form. We choose RREF because that makes the calculations somewhat simpler, but none of our conclusions would be different if we had used REF and back substitution.

Example 3: Suppose that the RREF form of the augmented matrix is

$$(\,R\,|\,\mathbf{e}\,) = \begin{pmatrix} \mathbf{1} & 2 & 0 & 1 & | & 5 \\ 0 & 0 & \mathbf{1} & 3 & | & 2 \\ 0 & 0 & 0 & 0 & | & \mathbf{1} \end{pmatrix}.$$

The last equation here is $0 = 1$, which clearly has no solutions: it expresses a contradiction. This is the signal that our original equations have no solutions. Notice that one way to say what has happened here is that the rank of R, which is 2, is less than the rank of $(\,R\,|\,\mathbf{e}\,)$, which is three. In general, we will have no solution precisely if $\text{rank}(R) < \text{rank}(\,R\,|\,\mathbf{e}\,)$.

Example 4: Suppose that the RREF form of the augmented matrix is

$$(\,R\,|\,\mathbf{e}\,) = \begin{pmatrix} 0 & \mathbf{1} & 2 & 0 & 1 & | & 5 \\ 0 & 0 & 0 & \mathbf{1} & 3 & | & 2 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

Now the idea is to solve for the variables x_2 and x_4, the variables for the columns containing pivots, in terms of the other variables, which are treated as parameters. To remind us that we are treating these variables as parameters, we will give them new names: $\alpha = x_1$, $\beta = x_3$, and $\gamma = x_5$. then our solution is

$$x_1 = \alpha, \quad x_2 = 5 - 2\beta - \gamma, \quad x_3 = \beta, \quad x_4 = 2 - 3\gamma, \quad x_5 = \gamma.$$

In vector form,

$$\mathbf{x} = \begin{pmatrix} \alpha \\ 5 - 2\beta - \gamma \\ \beta \\ 2 - 3\gamma \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \\ 0 \\ 2 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ -1 \\ 0 \\ -3 \\ 1 \end{pmatrix}. \tag{8}$$

Here we have three parameters, one for each column of R which does not contain a pivot. There are $n = 5$ unknowns and $r = \text{rank}(R) = 2$ pivots, and subtracting these numbers indeed gives $n - r = 3$ free parameters.

The pattern here is quite general. A solution will exist if $\text{rank}(R) = \text{rank}(\,R\,|\,\mathbf{e}\,)$, and it will have the general form

$$\mathbf{x} = \mathbf{X} + c_1\mathbf{x}_1 + c_2\mathbf{x}_x + \cdots + c_k\mathbf{x}_k.$$

The free parameters $c_1, \ldots c_k$ are just the original unknowns corresponding to the columns without pivots. Since there are $r = \text{rank}(R) = \text{rank}(A)$ pivots there will be $n - r$ free parameters in the solution (that is, $k = n - r$). Since we can choose the parameters freely, we can take $c_1 = c_2 = \cdots = c_k = 0$ and we thus find that $\mathbf{X}$ itself a solution. This is the *particular* solution we discussed in Section 1. If we consider now the homogeneous problem—the same equations, but with $\mathbf{b} = 0$—then we will also have $\mathbf{e} = 0$, and by looking at (8) we can see that we will have $\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_x + \cdots + c_k\mathbf{x}_k$ with the *same* vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$; this means that we have recovered (4).

We summarize in Principle 3 on page 8.

4. The case of n equations in n unknowns

Probably the most common systems of linear equations have the same number of equations as unknowns—say n equations in n unknowns. The coefficient matrix A is then *square*, with n rows and n columns. In this case there is a connection between the questions of whether a solution *exists*, and whether a solution which does exist is *unique*. As we shall see, one of two things may happen. Suppose that the augmented matrix has been reduced to RREF $(\,R\,|\,\mathbf{e}\,)$.

Case 1: $\text{rank}(A) = n$**.** Since R is an $n \times n$ matrix in RREF with no zero rows, it must be the identity matrix, so that $(\,R\,|\,\mathbf{e}\,) = (\,I\,|\,\mathbf{e}\,)$. The corresponding equations $x_1 = e_1$, $x_2 = e_2$, ..., $x_n = e_n$ will have a solution $\mathbf{x} = \mathbf{e}$ no matter what $\mathbf{e}$ is, and hence no matter what the original $\mathbf{b}$ was; moreover, the solution is clearly always unique.

Principle 3: Solving linear equations. Suppose that the augmented matrix $(A\,|\,\mathbf{b})$ is reduced to the REF or RREF $(R\,|\,\mathbf{e})$. Then:

(i) If $\text{rank}(R) < \text{rank}(R\,|\,\mathbf{e})$, so that the last nonzero equation is $0 = 1$, then the equations have no solutions. This cannot happen if the system is homogeneous.

(ii) If $\text{rank}(R) = \text{rank}(R\,|\,\mathbf{e})$ then the equations have at least one solution. Write $r = \text{rank}(R) = \text{rank}(A)$; then the solution is unique if $n = r$, i.e., if every column in R has a pivot. Otherwise, the equations have a family of solutions with $k = n - r$ free parameters. The general solution may be written in the form

$$\mathbf{x} = \mathbf{X} + c_1\mathbf{x}_1 + c_2\mathbf{x}_x + \cdots + c_k\mathbf{x}_k, \tag{9}$$

where $\mathbf{X}$ is a particular solution, $c_1, \ldots, c_k$ are the parameters, and $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are solutions of the homogeneous equations $A\mathbf{x} = \mathbf{0}$. The specific solutions are found by solving the reduced equations for the variables corresponding to the columns with pivots in terms of the other variables, which become the parameters.

(iii) The *homogeneous* system always has at least one solution: $\mathbf{x} = \mathbf{0}$. This is the *trivial* solution. The system has nontrivial solutions if and only if there are columns in R which do not contain pivots, that is, if and only if $r < n$. The general solution of the homogeneous equation is of the form

$$\mathbf{x} = c_1\mathbf{x}_1 + c_2\mathbf{x}_x + \cdots + c_k\mathbf{x}_k, \tag{10}$$

with $k = r - n$.

Case 2: $\text{rank}(A) < n$**.** In this case, the last row of R is a zero row. This means that for some choices of $\mathbf{b}$, the right hand side of the original equations, the vector $\mathbf{e}$ can have one more nonzero component than there are nonzero rows in R, i.e., that the equations will have no solution for some $\mathbf{b}$. On the other hand, if a solution does exist, then because there is a column without a pivot, our solution method will lead to a solution with at least one free parameter—that is, any solution that does exist will not be unique. We have:

Principle 4: n **equations in** n **unknowns.** If A is a square matrix then the system of equations $A\mathbf{x} = \mathbf{b}$ **either** has a unique solution for every $\mathbf{b}$ (Case 1), **or** fails to have a solution for some $\mathbf{b}$, and never has a unique solution (Case 2).

Note, for example, that if we know that for some $\mathbf{b}$ the system $A\mathbf{x} = \mathbf{b}$ has a unique solution, then we must be in Case 1 and we immediately know that it has a solution, and in fact a unique solution, for *every* $\mathbf{b}$. Note also that the homogeneous system $A\mathbf{x} = \mathbf{0}$ can have a nontrivial solution only in Case 2, that is, if and only if $\text{rank}(A) = 0$.

There is another way to distinguish between Case 1 and Case 2 which we will use but not prove: **we are in Case 1, that is,** $\text{rank}(A) = n$**, only if the determinant of** A**,** $\det(A)$**, is not zero**.

Much more can be said in Case 1. Suppose that we are in this case, i.e., that $\mathrm{rank}(A) = n$. Let us define the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ to be the columns of the $n \times n$ identity matrix:

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \mathbf{u}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

We know that the system $A\mathbf{x} = \mathbf{u}_i$ has a unique solution, which we will call $\mathbf{v}_i$, that is, $A\mathbf{v}_i = \mathbf{u}_i$. Now consider a matrix B with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$: $B = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \ \mathbf{v}_n)$. Because of the definition of matrix multiplication, if we compute AB we just multiply each column of B by the matrix A: thus

$$AB = (A\mathbf{v}_1 \ A\mathbf{v}_2 \ A\mathbf{v}_3 \ \cdots \ A\mathbf{v}_m) = (\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \ \cdots \ \mathbf{u}_n) = I.$$

Now we say that an $n \times n$ matrix A is *invertible* if it has an *inverse*: a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$ (A^{-1} must necessarily also be $n \times n$). We want to show that if A falls under Case 1 then it is invertible and the matrix B found above is A^{-1}. To do so we observe that B also falls under Case 1, since if $\mathbf{x}$ is a vector with $B\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = I\mathbf{x} = AB\mathbf{x} = A\mathbf{0} = \mathbf{0}$, so that the equations $B\mathbf{x} = 0$ have a unique solution. But then by the argument above there is a matrix C with $BC = I$, and then we have $A = AI = ABC = IC = C$, so $BA = BC = I$ and with $AB = I$ this shows that $B = A^{-1}$. It is also clear that if A is invertible then it must fall under Case 1, since the equations $A\mathbf{x} = \mathbf{b}$ have a solution $\mathbf{x} = A^{-1}\mathbf{b}$ for any $\mathbf{b}$.

These ideas also tell us how to compute A^{-1}. First, how do we find $\mathbf{v}_i$? We do Gaussian elimination on the augmented matrix $(\,A \,|\, \mathbf{u}_i\,)$, and $\mathbf{v}_i$, the solution, will just be the last column of the result, that is, the row reduction will be $(\,A \,|\, \mathbf{u}_i\,) \to (\,I \,|\, \mathbf{v}_i\,)$. Doing all these different problems to find all the $\mathbf{v}_i$ is a terrible duplication of effort, however, so we do them all at once:

$$(A|\mathbf{u}_1\,\mathbf{u}_2\,\cdots\mathbf{u}_n) \to (I|\mathbf{v}_1\,\mathbf{v}_2\,\cdots\mathbf{v}_n) \quad \text{or equivalently} \quad (\,A\,|\,I\,) \to (\,I\,|\,A^{-1}\,).$$

In general this is the simplest method of computing A^{-1} when one is given a specific numerical matrix A.

We can conclude that if A is a square matrix then any one of the following conditions is enough to guarantee that we are in Case 1, and hence that in fact all the conditions hold:

C1: The system $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$.

C2: Whenever the system $A\mathbf{x} = \mathbf{b}$ has a solution, the solution is unique.

C3. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

C4: $\mathrm{rank}(A) = n$

C5: A has an inverse matrix A^{-1} satisfying $AA^{-1} = A^{-1}A = I$.

C6: The reduced row-echelon form of A is the identity matrix I.

C7: The determinant of A is not zero.

5. Linear independence of vectors

Suppose we are given k vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_k$, these might be either row or column vectors, but they are all one or the other, and they all have the same number of components. To be specific we will think of them as column vectors belonging to the space $\mathbb{R}^n$ of all column vectors with n components. We then ask the question: can any one of these vectors be expressed as a linear combination of the remaining ones? If so, the vectors are *linearly dependent*, if not, they are *linearly independent*.

Example 5: (a) The vectors $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and $\mathbf{x}_3 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$, all of which belong to $\mathbb{R}^2$, are linearly *dependent*, since $\mathbf{x}_3$ can be expressed as a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$: $\mathbf{x}_3 = 3\mathbf{x}_1 - 2\mathbf{x}_2$,

(b) The vectors $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $\mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, from $\mathbb{R}^3$, are linearly *independent*. For example, we cannot write $\mathbf{x}_1 = a\mathbf{x}_2 + b\mathbf{x}_3$ no matter how we choose a and b, since $a\mathbf{x}_2 + b\mathbf{x}_3 = \begin{pmatrix} 0 \\ a \\ b \end{pmatrix}$ has first component 0, and $\mathbf{x}_1$ has first component 1.

(c) The vectors $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 5 \\ -3 \\ 2 \end{pmatrix}$, $\mathbf{x}_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, and $\mathbf{x}_3 = \begin{pmatrix} 7 \\ -1 \\ 2 \\ 0 \end{pmatrix}$, from $\mathbb{R}^4$, are linearly *dependent*, since $\mathbf{x}_2 = 0\mathbf{x}_1 + 0\mathbf{x}_3$. Clearly, any set of vectors in which one vector is $\mathbf{0}$ must be linearly dependent, by the same reasoning.

There is another way to describe linear dependence: the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_k$ are linearly dependent if there exist scalars $c_1, \ldots, c_k$, **not all zero**, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k = \mathbf{0}. \tag{11}$$

The restriction that not all the c_i be zero is important, since we could always make (11) true by taking $c_1 = c_2 = \cdots = c_k = 0$. This new definition of linear dependence is the same as our original definition. For if the vectors are linearly dependent according to our first definition then one of them, say $\mathbf{x}_1$, can be expressed as a linear combination of the others: $\mathbf{x}_1 = d_2\mathbf{x}_2 + d_3\mathbf{x}_3 + \cdots + d_k\mathbf{x}_k$; but then

$$\mathbf{x}_1 - d_2\mathbf{x}_2 - d_3\mathbf{x}_3 - \cdots - d_k\mathbf{x}_k = \mathbf{0},$$

which shows that (11) holds with the coefficients c_i not all zero (since $c_1 = 1$). Conversely, if (11) holds with some coefficient not zero—say, $c_1 \neq 0$—then we can solve the equation for $\mathbf{x}_1$, expressing it as a linear combination of the others:

$$\mathbf{x}_1 = -\left(\frac{c_2}{c_1}\right)\mathbf{x}_2 - \cdots - \left(\frac{c_k}{c_1}\right)\mathbf{x}_k,$$

so that the vectors are linearly dependent by our first definition.

How can we determine if the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent or linearly independent? Here is one way. Suppose that these are column vectors with n components, and build a matrix A with these vectors as columns: $A = (\mathbf{x}_1\, \mathbf{x}_2\, \ldots\, \mathbf{x}_k)$. The matrix A is $n \times k$. To say that (11) holds is just to say that $A\mathbf{c} = \mathbf{0}$, where $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$. This means that $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent—i.e., that (11) holds with the c_i not all zero—if the system of equations $A\mathbf{c} = \mathbf{0}$ has a nontrivial solution for $\mathbf{c}$. One can determine whether or not it does by reducing A to REF or RREF.

Finally, suppose we have n vectors, each with n components, and want to know if they are linearly independent. Then the matrix A is an $n \times n$ square matrix, and we can study it via the ideas of the previous section. The system $A\mathbf{c} = \mathbf{0}$ has no nontrivial solution if and only if we are in Case 1 (this is condition **C3** for being in case 1), i.e., if the matrix A satisfies any of the conditions **C1–C7**. Note that this means that we could add another condition to the list **C1–C7**, equivalent to all the rest:

C8. The columns of A are linearly independent vectors.

We summarize:

Principle 5: Linear independence. The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent if the system of equations $A\mathbf{c} = \mathbf{0}$, where $A = (\mathbf{x}_1\, \mathbf{x}_2\, \ldots\, \mathbf{x}_k)$, has a nontrivial solution. The vectors are linearly independent if the system $A\mathbf{c} = \mathbf{0}$ has only the trivial solution $\mathbf{c} = \mathbf{0}$. If $k = n$ and the vectors are column vectors with n components, then they are linearly independent if and only if the matrix A satisfies any of the conditions **C1–C7** of Section 4.

Remark 1: In Principle 3 on page 8 we found that $k = n - \operatorname{rank}(A) = n - r$ vectors are needed to express every solution of the equations $A\mathbf{x} = \mathbf{b}$, and observed that row reduction produced the needed vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$. We want to observe here that these k vectors are **linearly independent**. To see this, consider Example 4 and form a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ produced there:

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ -1 \\ 0 \\ -3 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 \\ -2c_2 - c_3 \\ c_2 \\ -3c_3 \\ c_3 \end{pmatrix}.$$

By looking at the first, third, and fifth components of the final form of this vector we see that if $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = 0$ then necessarily $c_1 = c_2 = c_3 = 0$, and this is precisely linear independence of $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$. The pattern is the same for any system $A\mathbf{x} = \mathbf{b}$.