Introduction to Linear Algebra

Thomas L. Scofield

August 31, 2014

Contents

1 Solving Linear Systems of Equations

1.1 Matrices, and Introduction to Octave

Definition 1: An m-by-n **real matrix** is a table of m rows and n columns of real numbers. We say that the matrix has **dimensions** m-by-n.

The plural of *matrix* is **matrices**.

Remarks:

1. Often we write a matrix $\mathbf{A} = (a_{ij})$, indicating that the matrix under consideration may be referred to as a single unit by the name $\mathbf{A}$, but that one may also refer to the entry in the i^{th} row, j^{th} column as a_{ij}.

2. If one of the matrix dimensions m or n is equal to 1, it is common to call the table a **vector** (or **column vector**, if $n = 1$; a **row vector** if $m = 1$). Though column vectors are just special matrices, it is common to use lowercase boldface letters for them (like $\mathbf{u}$, $\mathbf{v}$, $\mathbf{x}$, etc.), reserving uppercase boldface letters for other types of matrices. When $\mathbf{x}$ is an n-by-1 vector, we often denote its components with singly-subscripted non-bold letters—x_1 for the first component, x_2 for the 2^{nd} , and so on.

Practitioners carry out large-scale linear algebraic computations using software, and in this section we will alternate between discussions of concepts, and demonstrations of corresponding implementations in the software package Octave. To create a matrix (or vector) in Octave, you enclose elements in square brackets ([and]). Elements on the same row should be separated only by a space (or a comma). When you wish to start a new row, you indicate this with a semicolon (;). So, to enter the matrices

$$\begin{bmatrix} 1 & 5 & -2 \end{bmatrix}, \quad \begin{bmatrix} 4 \\ -1 \\ 3 \\ 7 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 3 & 0 \\ -1 & 5 \\ 2 & 1 \end{bmatrix},$$

you can type/execute

```
octave> [1 5 -2]
ans =
   1   5  -2

octave> [4; -1; 3; 7]
ans =
   4
  -1
   3
   7

octave> A = [3 0; -1 5; 2 1];
```

In all but the third of these commands, Octave echoed back to us its understanding of what we typed. It did not do so for the last command because we ended that command with a final semicolon. Also, since we preceded our final matrix with "`A =`", the resulting matrix may now be referred to by the letter (variable) `A`.

Just as writing $\mathbf{A} = (a_{ij})$ gives us license to refer to the element in the 2nd row, 1st column as a_{21}, storing a matrix in a variable in Octave gives us an immediate way to refer to its elements. The entry in the 2nd row, 1st column of the matrix `A` defined above can be obtained immediately by

```
octave> A(2,1)
ans = -1
```

That is, you can pick and choose an element from `A` by indicating its location in parentheses.

One can easily extract whole **submatrices** from `A` as well. Suppose you want the entire first row of entries. This you do by specifying the row, but using a colon (:) in place of specifying the column.

```
octave> A(1,:)
ans =
   3   0
```

Next, suppose we want to get the first and third rows of `A`. Since we want full rows here, we continue to use the colon where a column can be specified. We use a vector whose entries are 1 and 3 to specify which rows.

```
octave> A([1 3],:)
ans =
   3   0
   2   1
```

There are some shortcuts in Octave when creating matrices or vectors with particular kinds of structure. The colon may be used between numbers as a quick way to create row vectors whose entries are evenly spaced. For instance, a row vector containing the first five positive integers is produced by the command

```
octave> 1:5
  ans =
    1   2   3   4   5
```

You can also specify a "step" or "meshsize" along the way, as in

```
octave> 1:.5:3
ans =
   1.0000    1.5000    2.0000    2.5000    3.0000
```

One implication of the ability to create vectors in this fashion is that, if we wish to extract the first two rows of the matrix `A` above, either of the commands

```
octave> A(1:2, :)
```

or

```
octave> A([1 2], :)
```

will do the trick. (Here and following, I have suppressed OCTAVE's output.)

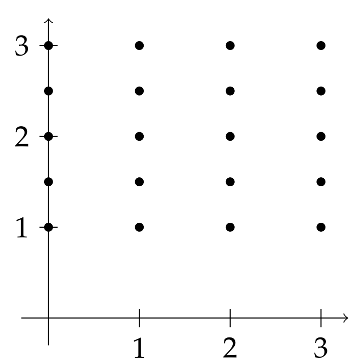
Sometimes we want matrices (a pair of them) that contain the coordinates of a mesh consisting of points in the plane. The `meshgrid()` command is useful for this purpose, as exhibited below. As you see, `meshgrid()` returns *two* matrices, not just one. In the example we graph a collection of points from a mesh in the xy-plane, along with the `meshgrid()` command generating x- and y-components for these points. Note that the contents of the matrix `Y` are flipped from what you might expect.

```
octave> [X, Y] = meshgrid(0:3, 1:.5:3)
X =

   0   1   2   3
   0   1   2   3
   0   1   2   3
   0   1   2   3
   0   1   2   3

Y =

   1.0000   1.0000   1.0000   1.0000
   1.5000   1.5000   1.5000   1.5000
   2.0000   2.0000   2.0000   2.0000
   2.5000   2.5000   2.5000   2.5000
   3.0000   3.0000   3.0000   3.0000
```



Aside: Suppose, above the region covered by our mesh, we want to view the surface given by $z = x^2/y$. You might use these commands (try them!).

```
octave> [X, Y] = meshgrid(0:3, 1:.5:3)
octave> Z = X.^2 ./ Y
octave>  mesh(X, Y, Z)
```

You may be able to place your mouse cursor over this plot, click-hold and rotate the figure around to view it from various perspectives. As you view the plot, it will occur to you that the surface would be more appealing if we used a finer mesh. As an exercise, try reproducing this surface over the same region of the xy-plane, but with grid points (in that plane) just 0.1 apart.

A special class among the **square matrices** (i.e., those having equal numbers of rows and columns) are the **diagonal** matrices. Such a matrix $\mathbf{A} = (a_{ij})$ is one whose entries a_{ij} are zero whenever $i \neq j$. The `diag()` command makes it easy to construct such a matrix in Octave, even providing the ability to place specified entries on a **super-** or **subdiagonal** (i.e., a diagonal that lies above or below the **main diagonal**). We give here two examples of the use of `diag()`. In the first case, the only argument is a vector, whose entries are then placed on the main diagonal of an appropriately-sized diagonal matrix; in the 2nd case, the additional argument of (−1) is used to request that the vector of entries be placed on the first subdiagonal.

```
octave> diag([1 3 -1])
ans =
   1   0   0
   0   3   0
   0   0  -1

octave> diag([1 3 -1], -1)
ans =
   0   0   0   0
   1   0   0   0
   0   3   0   0
   0   0  -1   0
```

Other Octave commands that are helpful in producing certain types of matrices are `zeros()`, `ones()`, `eye()`, and `rand()`. You can read the help pages to learn the purpose and required syntax of these and other Octave commands by typing

`help` <*command name*>

at the Octave prompt. It is perhaps relevant to note that numbers (scalars) themselves are considered by Octave to be 1-by-1 matrices.

The title of the next section is "Matrix Algebra." Before we can dive into that, system we must know what one means by the word *equality*.

Definition 2: Two matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are said to be *equal* if their dimensions are equal, and if the entries in every location are equal.

Example 1:

The two vectors

$$\begin{bmatrix} 3 & -1 & 2 & 5 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 3 \\ -1 \\ 2 \\ 5 \end{bmatrix}$$

cannot be considered equal, since they have different dimensions. While the entries are the same, the former is a *row* vector and the latter a *column* vector.

■

Most programming languages can interpret the concept of "equal" in several ways, and OCTAVE is no exception. A single equal sign is interpreted as *assignment*; when two equals signs appear together, it is for the purpose of *comparison*.

```
a = 4            % x has been assigned the value of 4
a == 7           % comparing value in x with 7; returns FALSE (value = 0)
A = rand(3,2)    % A is a 3-by-2 matrix with values generated randomly
a == A           % Error: Cannot compare entities of different "types"
B = A;           % B has been made to be identical to A
B(1,2) = 4       % (1,2)-entry in B has been reassigned value of 4
B == A           % Compares these two matrices entry-by-entry
A == rand(2,3)   % Guess at the result, then view the answer
```

Some tips about the OCTAVE language:

- The next examples demonstrate other relational operators.

  ```
  a ~= 7         % returns TRUE (1) if the value in a is unequal to 7
  a >= 7         % returns TRUE (1) if the value in a is at least 7
  a==4 & b==4    % & is the logical AND operator
  a==4 | b==4    % | is the logical OR operator
  ```

- When OCTAVE encounters the percent symbol (%), the rest of the line is considered a comment.

- If you see an OCTAVE command seemingly end in three dots "`...`", the command actually carries over to the next line (with the dots ignored).

- If you perform, in succession, commands like

  ```
  C = ones(4,5)       % C is now a 4-by-5 matrix of all ones
  C = [2; 1; -1]      % C is now a 3-by-1 vector with specified entries
  ```

  OCTAVE does not put up a fuss, but readily adapts from `C` being an entity holding 20 double-precision values to one holding just three. When, however, you have a matrix `B` already in existence, and you assign one of its entries (or reassign, as occurred above in the line reading "`B(1,2) = 4`"), the size and shape of `B` do not change.

The flexibility OCTAVE has over what is stored in a variable comes from the fact that it determines variable type at the time of instantiation. Some people like this feature, while others (C++ enthusiasts?) consider it an open doorway to sloppy programming. Practically speaking, most variables are double-precision numbers, or matrices/vectors containing double-precision numbers.

- OCTAVE is case-sensitive, as exhibited by the "`a == A`" line above.
- One can designate the default way that numbers are printed. Values which are clearly integers may be printed with no trailing decimal and digits. The first time you start OCTAVE, it will display non-integers with 4 digits to the right of the decimal point; very small numbers will be reported using scientific notation. Try the following commands to see how one can change this method of displaying numbers.

```
pi
format long
pi
pi / 10^6          % small number, uses scientific notation
format rat         % better rational number approx to pi than 22/7
pi
format short       % mode of display from when Octave was first started
pi
```

  There are ways, of course, to make OCTAVE print a number with 6 (or some other number) decimal places, but that seems rather unimportant for now.

In this course, the term **vector** will be synonomous with *column vector*. The set of vectors having n components, all of which are real numbers, will be called $\mathbb{R}^n$, or sometimes **Euclidean n-space**. The elements of $\mathbb{R}^n$ are n-by-1 matrices, sometimes called **n-vectors**. However, as it takes less room out of a page to list the contents of a vector horizontally rather than vertically, we will often specify an n-vector horizontally using parentheses, as in

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) .$$

1.2 Matrix Algebra

The most fundamental algebraic operations on matrices are as follows:

1. **Addition of Two Matrices**.

   Given two m-by-n matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, we define their sum $\mathbf{A} + \mathbf{B}$ to be the m-by-n matrix whose entries are $(a_{ij} + b_{ij})$. That is,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} := \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix} .$$

In order to add two matrices, they must have the same number of rows and columns (i.e., be matrices with *the same dimensions*). Note that this is not the same as saying they must be square matrices!

It is simple to add two matrices in Octave. One possibility is code like

```
octave> A = [3 1 6; 1 2 -1];
octave> A + ones(2,3)
ans =
    4    2    7
    2    3    0
```

which creates a 2-by-3 matrix **A**, and then adds to it another 2-by-3 matrix whose entries are all ones.

2. **Multiplication of a Matrix by a Scalar**.

Given an m-by-n matrix $\mathbf{A} = (a_{ij})$ and a scalar c, we define the scalar multiple $c\mathbf{A}$ to be the m-by-n matrix whose entries are (ca_{ij}). That is,

$$c\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} := \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}.$$

Our definitions for matrix addition and scalar multiplication have numerous implications. They include the following:

a) Matrix *subtraction* is merely a combination of matrix addition and scalar multiplication by (-1): $\mathbf{A} - \mathbf{B} := \mathbf{A} + (-1)\mathbf{B}$.

b) Distributive laws between matrix addition and scalar multiplication hold:

   i. $c(\mathbf{A} + \mathbf{B}) \;=\; c\mathbf{A} + c\mathbf{B}$.

   ii. $(c + d)\mathbf{A} \;=\; c\mathbf{A} + d\mathbf{A}$.

c) An appopriately-sized matrix whose entries are all zeros serves as an **additive identity** (or **zero matrix**, denoted in boldface by **0**). That is, $\mathbf{A} + \mathbf{0} = \mathbf{A}$.

d) Scalar multiplication by 0 produces the *zero matrix* **0**. That is, $(0)\mathbf{A} = \mathbf{0}$.

In the lines of code below, we generate a 3-by-2 matrix whose entries are sampled from a normal distribution with mean 0 and standard deviation 1. To exhibit scalar multiplication in Octave, we then multiply this matrix by 3.

```
octave> 3*randn(3,2)
ans =
  -4.03239    3.04860
   1.67442    2.60456
   0.33131    2.31099
```

which produces a 3-by-2 matrix whose entries are sampled from a normal distribution with mean 0 and standard deviation 1, and then multiplies it by the scalar 3.

3. **Multiplication of Two Matrices**

   When we multiply two matrices, the product is a matrix whose elements arise from **dot products**[1] between the rows of the first (matrix) factor and columns of the second. An immediate consequence of this: if **A** and **B** are matrices, the product **AB** makes sense precisely when the number of columns in **A** is equal to the number of rows in **B**. To be clearer about how such a matrix product is achieved, suppose **A** is an m-by-n matrix while **B** is an $n \times p$ matrix. If we write

$$\mathbf{A} = \begin{bmatrix} \mathbf{r_1} & \rightarrow \\ \mathbf{r_2} & \rightarrow \\ \vdots & \\ \mathbf{r_m} & \rightarrow \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} \mathbf{c_1} & \mathbf{c_2} & \cdots & \mathbf{c_p} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix},$$

   with each of the rows $\mathbf{r_i}$ of **A** having n components and likewise each of the columns $\mathbf{c_j}$ of **B**, then their product is an $m \times p$ matrix whose entry in the i^{th}-row, j^{th}-column is obtained by taking the dot product of $\mathbf{r_i}$ with $\mathbf{c_j}$. Thus if

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 0 & 3 \\ -5 & 1 \\ 7 & -4 \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} 3 & 1 & 0 \\ -2 & 4 & 10 \end{bmatrix},$$

   then the product **AB** will be the 4×3 matrix

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} (2,-1)\cdot\begin{bmatrix} 3 \\ -2 \end{bmatrix} & (2,-1)\cdot\begin{bmatrix} 1 \\ 4 \end{bmatrix} & (2,-1)\cdot\begin{bmatrix} 0 \\ 10 \end{bmatrix} \\ (0,3)\cdot\begin{bmatrix} 3 \\ -2 \end{bmatrix} & (0,3)\cdot\begin{bmatrix} 1 \\ 4 \end{bmatrix} & (0,3)\cdot\begin{bmatrix} 0 \\ 10 \end{bmatrix} \\ (-5,1)\cdot\begin{bmatrix} 3 \\ -2 \end{bmatrix} & (-5,1)\cdot\begin{bmatrix} 1 \\ 4 \end{bmatrix} & (-5,1)\cdot\begin{bmatrix} 0 \\ 10 \end{bmatrix} \\ (7,-4)\cdot\begin{bmatrix} 3 \\ -2 \end{bmatrix} & (7,-4)\cdot\begin{bmatrix} 1 \\ 4 \end{bmatrix} & (7,-4)\cdot\begin{bmatrix} 0 \\ 10 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 8 & -2 & -10 \\ -6 & 12 & 30 \\ -17 & -1 & 10 \\ 29 & -9 & -40 \end{bmatrix}. \end{aligned}$$

[1] The *dot product* of two vectors is a concept from vector calculus, studied primarily in the case where those vectors have just two components. It appears as well in elementary physics courses.

Remarks:

- When we write **AB**, where **A**, **B** are appropriately-sized matrices, we will mean the product of these two matrices using multiplication as defined above. In Octave, you must be careful to include the multiplication symbol (since `AB` is a valid variable name), as in

```
octave> A = [1 2 3; 4 5 6];
octave> B = [2; 3; 1];
octave> A*B
ans =
   11
   29
```

Vectorization

The manner in which we defined matrix multiplication is the standard (and most useful, as you will see) one. Nevertheless, there are times one has numerous pairs of numbers to multiply. If, for each pair, one of the numbers is stored in a vector **x** with its counterpart stored in the corresponding location of a vector **y**, one could use a **for** loop to achieve this; the Octave code would look something like this:

```
vectorSize = length(x)     % I assume y is of the same length
z = zeros(size(x))         % creates a vector z, same size as x
for ii = 1:vectorSize
 z(ii) = x(ii) * y(ii);
end
```

This code cycles through the elements of the vectors **x** and **y** one at a time, so the processing time of this loop increases as the length of the vectors increase. In one test conducted by the author, the processing time was 0.00554 s when the vectors were of length 500, but had jumped to 4.156 s for vectors of length 500,000.

As it is not rare for datasets to contain millions of values these days, we seek a way to speed up such computations. In Octave, the solution is to **vectorize** the calculations. Without getting into the details, this essentially means the software performs all of the calculations—products, in this case—in parallel, performing them **componentwise**. We tell Octave to carry out componentwise operations simultaneously on all elements of the vectors in question using the dot (.) symbol. You witnessed one example of this sort of vectorization in the “Aside” of the previous section, when I demonstrated a 3D plot using the `mesh()` command. We can vectorize the lines of code above (loop and all) with this single command:

```
z = x .* y;     % creates z and stores in it products from x, y
```

It really does make a difference; when the vector sizes were 500,000 in my test run, this vectorized code took 0.003328 s of processing time (as compared with 4.156 s when similar calculations were done in the **for** loop).

The reader should try creating two vectors of the same length (say, 10-by-1), perhaps using commands like

```
x = unidrnd(25, 10, 1)
```

which will fill the vector **x** with randomly-selected integers in the range 1–25. If your two vectors are called **x** and **y**, then experiment with various non-dot (unvectorized) and dot (vectorized) versions of calculations, such as

non-vectorized	vs.	vectorized
`x + y`		`x .+ y`
`x - y`		`x .- y`
`x * y`		`x .* y`
`x / y`		`x ./ y`
`x ^ 3`		`x .^ 3`

See if you can explain the differences (or lack thereof) in results, or why some commands do not work at all. (Cudos to you if you can figure out how to interpret the result of "`x / y`".) As it turns out, the same dot notation tells OCTAVE to perform operations componentwise for matrices as well as vectors—that is, if we "dot-multiply" matrices **A**, **B** (perhaps like the two below on the left side of the equals sign), we get the componentwise result (as displayed on the right side of the equals sign):

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} .* \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} a\alpha & b\beta \\ c\gamma & d\delta \end{bmatrix}$$

This explains what is happening in the code of the "Aside" (last section). Do not think of *this* as matrix multiplication (which was defined above, and yields quite different results), nor as a *dot product*, but rather as the simultaneous calculation of many products using vectorization.

Conveniently, most every mathematical function built into OCTAVE's mathematical library vectorizes its calculations. That is, you can obtain the values of the sine function simultaneously for all inputs stored in **x** by simply asking for them:

```
sin(x)        % output is vector of same size as x
exp(A)        % exponentiates every element in A
log(y)        % takes natural log of every element in y
sqrt(B)       % takes square root of every element in B
```

- Notice that, if **A** is 4-by-2 and **B** is 2-by-3, then the product **AB** is defined, but the product **BA** is not. This is because the number of columns in **B** is unequal to the number of rows in **A**. Thus, for it to be possible to multiply two matrices, one of which is m-by-n, in either order, it is necessary that the other be n-by-m. Even when both products **AB** and **BA** are possible, however, *matrix multiplication is not commutative*. That is, $\mathbf{AB} \neq \mathbf{BA}$, in general.

- We *do* have a distributive law for matrix multiplication and addition. In particular, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$, for all appropriately-sized matrices **A**, **B**, **C**.

- When an m-by-n matrix **A** is multiplied by an n-by-1 (column) vector (an n-vector, for short), the result is an m-vector. That is, for each n-vector **v**, **Av** is an m-vector. It is natural to think of left-multiplication by **A** as a *mapping* (or function) which takes n-vectors **v** as inputs and produces m-vectors **Av** as outputs. Of course, if **B** is an ℓ-by-m matrix, then one can left-multiply the product **Av** by **B** to get $\mathbf{B}(\mathbf{Av})$. The manner in which we defined matrix products ensures that things can be grouped differently with no change in the answer—that is, so
$$(\mathbf{BA})\mathbf{v} = \mathbf{B}(\mathbf{Av}) .$$

- Notice that the n-by-n matrix
$$\mathbf{I}_n := \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

  has the property that, whenever **C** is an n-by-p matrix (so that the product $\mathbf{I}_n\mathbf{C}$ makes sense), it is the case that $\mathbf{I}_n\mathbf{C} = \mathbf{C}$. Moreover, if **B** is an m-by-n matrix, then $\mathbf{B}\mathbf{I}_n = \mathbf{B}$. Since multiplication by $\mathbf{I}_n$ does not change the matrix (or vector) with which you started, $\mathbf{I}_n$ is called the **n-by-n identity matrix**. In most instances, we will write **I** instead of $\mathbf{I}_n$, as the dimensions of **I** should be clear from context.

  In Octave, the function that returns the n-by-n identity matrix is `eye(n)`. This explains the result of the commands

```
octave> A = [1 2 3; 2 3 -1]
A =
   1   2   3
   2   3  -1

octave> A*eye(3)
ans =
   1   2   3
   2   3  -1
```

- For a square (n-by-n) matrix $\mathbf{A}$, there may be a corresponding n-by-n matrix $\mathbf{B}$ having the property that
$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n .$$
If so, the matrix $\mathbf{A}$ is said to be **nonsingular** or **invertible**, with inverse matrix $\mathbf{B}$. Usually the **inverse of A**, when it exists, is denoted by $\mathbf{A}^{-1}$. This relationship is symmetric, so if $\mathbf{B}$ is the inverse of $\mathbf{A}$, then $\mathbf{A}$ is the inverse of $\mathbf{B}$ as well. If $\mathbf{A}$ is not invertible, it is said to be **singular**.

  The following fact about the product of invertible matrices is easily proved.

> **Theorem 1:** Suppose $\mathbf{A}$, $\mathbf{B}$ are both n-by-n invertible matrices. Then their product $\mathbf{AB}$ is invertible as well, having inverse $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

When $\mathbf{A}$ is invertible, it is not so easy to find $\mathbf{A}^{-1}$ as one might think. With rounding (and sometimes instability) in the calculations, one cannot, in general, get a perfect representation of the inverse using a calculator or computer, though the representation one gets is often good enough. In Octave one uses the `inv()` command.

```
octave> A = [1 2 3; 2 3 -1; 1 0 -2]
A =
   1   2   3
   2   3  -1
   1   0  -2

octave> B = inv(A)
B =
   0.66667  -0.44444   1.22222
  -0.33333   0.55556  -0.77778
   0.33333  -0.22222   0.11111

octave> B*A
ans =
   1.00000  -0.00000   0.00000
   0.00000   1.00000   0.00000
   0.00000   0.00000   1.00000
```

4. **Transposition of a Matrix**

   Look closely at the two matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 & -1 \\ -3 & -1 & 1 & -1 \\ 2 & -2 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} 1 & -3 & 2 \\ 2 & -1 & -2 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

for a connection between the two. The matrix **B** has been formed from **A** so that the first column of **A** became the first row of **B**, the second column of **A** became the 2nd row of **B**, and so on. (One might say with equal accuracy that the rows of **A** became the columns of **B**, or that the rows/columns of **B** are the columns/rows of **A**.) The operation that produces this matrix **B** from (given) matrix **A** is called **transposition**, and matrix **B** is called the **transpose of A**, denoted as $\mathbf{B} = \mathbf{A}^T$. (Note: In some texts the *prime* symbol is used in place of the ^{T}, as in $\mathbf{B} = \mathbf{A}'$.)

When you already have a matrix **A** defined in Octave, there is a simple command that produces its transpose. Strictly speaking that command is `transpose()`. However, placing an apostrophe (a *prime*) after the name of the matrix produces the tranpose as well, so long as the entries in the matrix are all *real* numbers (i.e., having zero *imaginary parts*). That is why the result of the two commands below is the same for the matrix **A** on which we use them.

```
octave> A = [1 2 3; 2 3 -1]
A =
    1    2    3
    2    3   -1

octave> transpose(A)
ans =
    1    2
    2    3
    3   -1

octave> A'
ans =
    1    2
    2    3
    3   -1
```

Remarks:

- If **A** is an m-by-n matrix, then $\mathbf{A}^T$ is n-by-m.
- Some facts which are easy to prove about matrix transposition are the following:
  - (i) For all matrices **A** it is the case that $(\mathbf{A}^T)^T = \mathbf{A}$.
  - (ii) Whenever two matrices **A** and **B** can be added, it is the case that $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
  - (iii) Whenever the product **AB** of two matrices **A** and **B** is defined, it is the case that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.
    (Compare this result to Theorem 1, a similar-looking fact about the inverse of the product of two invertible matrices.)
  - (iv) For each invertible matrix **A**, $\mathbf{A}^T$ is invertible as well, with $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

- There are some matrices $\mathbf{A}$ for which $\mathbf{A}^T = \mathbf{A}$. Such matrices are said to be **symmetric**.

1.3 Matrix Multiplication and Systems of Linear Equations

1.3.1 Several interpretations of matrix multiplication

In the previous section we saw what is required (in terms of matrix dimensions) in order to be able to produce the product $\mathbf{AB}$ of two matrices $\mathbf{A}$ and $\mathbf{B}$, and we saw how to produce this product. There are several useful ways to conceptualize this product, and in this first sub-section we will investigate them. We first make a definition.

Definition 3: Let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ be matrices all having the same dimensions. For each choice of real numbers $c_1, \ldots, c_k$, we call

$$c_1\mathbf{A}_1 + c_2\mathbf{A}_2 + \cdots + c_k\mathbf{A}_k$$

a **linear combination** of the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_k$. The set of all such linear combinations

$$S := \{c_1\mathbf{A}_1 + c_2\mathbf{A}_2 + \cdots + c_k\mathbf{A}_k \,|\, c_1, \ldots, c_k \in \mathbb{R}\}$$

is called the **linear span** (or simply **span**) of the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_k$. We sometimes write $S = \text{span}(\{\mathbf{A}_1, \ldots, \mathbf{A}_k\})$.

Here, now, are several different ways to think about product $\mathbf{AB}$ of two appropriately sized matrices $\mathbf{A}$ and $\mathbf{B}$.

1. **Block multiplication**. This is the first of four descriptions of matrix multiplication, and it is the most general. In fact, each of the three that follow is a special case of this one.

   Any matrix (table) may be separated into **blocks** (or *submatrices*) via horizontal and vertical lines. We first investigate the meaning of matrix multiplication at the block level when the left-hand factor of the matrix product $\mathbf{AB}$ has been subdivided using only vertical lines, while the right-hand factor has correspondingly been blocked using only horizontal lines.

Example 2:

Suppose

$$\mathbf{A} = \left[\begin{array}{cc|c|cc} 8 & 8 & 3 & -4 & 5 \\ 6 & -6 & 1 & -8 & 6 \\ 5 & 3 & 4 & 2 & 7 \end{array}\right] = \left[\begin{array}{c|c|c} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \end{array}\right]$$

(Note how we have named the three blocks found in **A**!), and

$$\mathbf{B} = \left[\begin{array}{cccc} -3 & 5 & -5 & -2 \\ 2 & -2 & 2 & -7 \\ \hline -6 & 6 & 0 & 3 \\ \hline -3 & 2 & -5 & 0 \\ 0 & -1 & -1 & 4 \end{array}\right] = \left[\begin{array}{c} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \\ \hline \mathbf{B}_3 \end{array}\right].$$

Then

$$\begin{aligned} \mathbf{AB} &= \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_2 + \mathbf{A}_3\mathbf{B}_3 \\ &= \begin{bmatrix} 8 & 8 \\ 6 & -6 \\ 5 & 3 \end{bmatrix}\begin{bmatrix} -3 & 5 & -5 & -2 \\ 2 & -2 & 2 & -7 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}\begin{bmatrix} -6 & 6 & 0 & 3 \end{bmatrix} + \begin{bmatrix} -4 & 5 \\ -8 & 6 \\ 2 & 7 \end{bmatrix}\begin{bmatrix} -3 & 2 & -5 & 0 \\ 0 & -1 & -1 & 4 \end{bmatrix} \\ &= \begin{bmatrix} -8 & 24 & -24 & -72 \\ -30 & 42 & -42 & 30 \\ -9 & 19 & -19 & -31 \end{bmatrix} + \begin{bmatrix} -18 & 18 & 0 & 9 \\ -6 & 6 & 0 & 3 \\ -24 & 24 & 0 & 12 \end{bmatrix} + \begin{bmatrix} 12 & -13 & 15 & 20 \\ 24 & -22 & 34 & 24 \\ -6 & -3 & -17 & 28 \end{bmatrix} \\ &= \begin{bmatrix} -14 & 29 & -9 & -43 \\ -12 & 26 & -8 & 57 \\ -39 & 40 & -36 & 9 \end{bmatrix}. \end{aligned}$$

■

While we were trying to keep things simple in the previous example by drawing only vertical lines in **A**, the number and locations of those vertical lines was somewhat arbitrary. Once we chose how to subdivide **A**, however, the horizontal lines in **B** had to be drawn to create blocks with rows as numerous as the columns in the blocks of **A**.

Now, suppose we subdivide the left factor with *both* horizontal and vertical lines. Say that

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \\ \hline \mathbf{A}_{31} & \mathbf{A}_{32} \end{array}\right].$$

Where the vertical line is drawn in **A** continues to dictate where a horizontal line must be drawn in the right-hand factor **B**. On the other hand, if we draw any vertical lines in to create blocks in the right-hand factor **B**, they can go anywhere, paying no heed to where the horizontal lines appear in **A**. Say that

$$\mathbf{B} = \left[\begin{array}{c|c|c|c} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} & \mathbf{B}_{14} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} & \mathbf{B}_{24} \end{array}\right].$$

Then

$$\begin{aligned} \mathbf{AB} &= \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \\ \hline \mathbf{A}_{31} & \mathbf{A}_{32} \end{array}\right] \left[\begin{array}{c|c|c|c} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} & \mathbf{B}_{14} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} & \mathbf{B}_{24} \end{array}\right] \\ &= \left[\begin{array}{c|c|c|c} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} & \mathbf{A}_{11}\mathbf{B}_{13} + \mathbf{A}_{12}\mathbf{B}_{23} & \mathbf{A}_{11}\mathbf{B}_{14} + \mathbf{A}_{12}\mathbf{B}_{24} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} & \mathbf{A}_{21}\mathbf{B}_{13} + \mathbf{A}_{22}\mathbf{B}_{23} & \mathbf{A}_{21}\mathbf{B}_{14} + \mathbf{A}_{22}\mathbf{B}_{24} \\ \hline \mathbf{A}_{31}\mathbf{B}_{11} + \mathbf{A}_{32}\mathbf{B}_{21} & \mathbf{A}_{31}\mathbf{B}_{12} + \mathbf{A}_{32}\mathbf{B}_{22} & \mathbf{A}_{31}\mathbf{B}_{13} + \mathbf{A}_{32}\mathbf{B}_{23} & \mathbf{A}_{31}\mathbf{B}_{14} + \mathbf{A}_{32}\mathbf{B}_{24} \end{array}\right]. \end{aligned}$$

Example 3:

Suppose **A**, **B** are the same as in Example 2. Let's subdivide **A** in the following (arbitrarily chosen) fashion:

$$\mathbf{A} = \left[\begin{array}{cccc|c} 8 & 8 & 3 & -4 & 5 \\ \hline 6 & -6 & 1 & -8 & 6 \\ 5 & 3 & 4 & 2 & 7 \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right].$$

Given the position of the vertical divider in **A**, we must place a horizontal divider in **B** as shown below. Without any requirements on where vertical dividers appear, we choose (again arbitrarily) not to have any.

$$\mathbf{B} = \left[\begin{array}{cccc} -3 & 5 & -5 & -2 \\ 2 & -2 & 2 & -7 \\ -6 & 6 & 0 & 3 \\ -3 & 2 & -5 & 0 \\ \hline 0 & -1 & -1 & 4 \end{array}\right] = \left[\begin{array}{c} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \end{array}\right].$$

Then

$$
\begin{aligned}
\mathbf{AB} &= \left[\begin{array}{c} \mathbf{A}_{11}\mathbf{B}_1 + \mathbf{A}_{12}\mathbf{B}_2 \\ \hline \mathbf{A}_{21}\mathbf{B}_1 + \mathbf{A}_{22}\mathbf{B}_2 \end{array}\right] \\
&= \left[\begin{array}{c}
\begin{bmatrix} 8 & 8 & 3 & -4 \end{bmatrix}\begin{bmatrix} -3 & 5 & -5 & -2 \\ 2 & -2 & 2 & -7 \\ -6 & 6 & 0 & 3 \\ -3 & 2 & -5 & 0 \end{bmatrix} + 5\begin{bmatrix} 0 & -1 & -1 & 4 \end{bmatrix} \\ \hline
\begin{bmatrix} 6 & -6 & 1 & -8 \\ 5 & 3 & 4 & 2 \end{bmatrix}\begin{bmatrix} -3 & 5 & -5 & -2 \\ 2 & -2 & 2 & -7 \\ -6 & 6 & 0 & 3 \\ -3 & 2 & -5 & 0 \end{bmatrix} + \begin{bmatrix} 6 \\ 7 \end{bmatrix}\begin{bmatrix} 0 & -1 & -1 & 4 \end{bmatrix}
\end{array}\right] \\
&= \left[\begin{array}{cccc} -14 & 29 & -9 & -43 \\ \hline -12 & 26 & -8 & 57 \\ -39 & 40 & -36 & 9 \end{array}\right].
\end{aligned}
$$

■

2. **Sums of rank-one matrices**. Now let us suppose that $\mathbf{A}$ has n columns and $\mathbf{B}$ has n rows. Suppose also that we block (as described allowed for in the previous case above) $\mathbf{A}$ by column—one column per block—and correspondingly $\mathbf{B}$ by row:

$$
\mathbf{A} = \left[\begin{array}{c|c|c|c} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_n \end{array}\right] \qquad \text{and} \qquad \mathbf{B} = \left[\begin{array}{c} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \\ \hline \vdots \\ \hline \mathbf{B}_n \end{array}\right].
$$

Following Example 2, we get

$$
\mathbf{AB} = \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_2 + \cdots + \mathbf{A}_n\mathbf{B}_n = \sum_{j=1}^{n} \mathbf{A}_j\mathbf{B}_j \,. \tag{1.1}
$$

The only thing new here to say concerns the individual products $\mathbf{A}_j\mathbf{B}_j$ themselves, in which the first factor $\mathbf{A}_j$ is a vector in $\mathbb{R}^m$ and the 2^{nd} $\mathbf{B}_j$ is the *transpose* of a vector in $\mathbb{R}^p$ (for some m and p).

So, take $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^p$. Since $\mathbf{u}$ is m-by-1 and $\mathbf{v}^T$ is 1-by-p, the product $\mathbf{uv}^T$, called the **outer product** of $\mathbf{u}$ and $\mathbf{v}$, makes sense, yielding an m-by-p matrix.

Example 4:

Given $\mathbf{u} = (-1, 2, 1)$ and $\mathbf{v} = (3, 1, -1, 4)$, their vector outer product is

$$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & -1 & 4 \end{bmatrix} = \begin{bmatrix} -3 & -1 & 1 & -4 \\ 6 & 2 & -2 & 8 \\ 3 & 1 & -1 & 4 \end{bmatrix}.$$

■

If you look carefully at the resulting outer product in the previous example, you will notice it has relatively simple structure—its 2^{nd} through 4^{th} columns are simply scalar multiples of the first, and the same may be said about the 2^{nd} and 3^{rd} rows in relation to the 1^{st} row. Later in these notes, we will define the concept of the **rank of a matrix**. Vector outer products are always matrices of rank 1 and thus, by (1.1), every matrix product can be broken into the sum of rank-one matrices.

3. **Linear combinations of columns of A**. Suppose **B** has p columns, and we partition it in this fashion (Notice that $\mathbf{B}_j$ represents the j^{th} *column* of **B** instead of the j^{th} row, as it meant above!):

$$\mathbf{B} = \left[\begin{array}{c|c|c|c} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_p \end{array} \right].$$

This partitioning by *vertical* lines of the right-hand factor in the matrix product **AB** does not place any constraints on how **A** is partitioned, and so we may write

$$\mathbf{AB} = \mathbf{A}\left[\begin{array}{c|c|c|c} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_p \end{array} \right] = \left[\begin{array}{c|c|c|c} \mathbf{AB}_1 & \mathbf{AB}_2 & \cdots & \mathbf{AB}_p \end{array} \right].$$

That is, for each $j = 1, 2, \ldots, p$, the j^{th} column of **AB** is obtained by left-multiplying the j^{th} column of **B** by **A**.

Having made that observation, let us consider more carefully what happens when **A**—suppose it has n columns $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$—multiplies a vector $\mathbf{v} \in \mathbb{R}^n$. (Note that each $\mathbf{B}_j$ is just such a vector.) Blocking **A** by columns, we have

$$\mathbf{Av} = \left[\begin{array}{c|c|c|c} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_n \end{array} \right] \left[\begin{array}{c} v_1 \\ \hline v_2 \\ \hline \vdots \\ \hline v_n \end{array} \right] = v_1\mathbf{A}_1 + v_2\mathbf{A}_2 + \cdots + v_n\mathbf{A}_n.$$

That is, the matrix-vector product **Av** is simply a linear combination of the columns of **A**, with the scalars multiplying these columns taken (in order, from top to bottom) from **v**. The implication for the matrix product **AB** is that each of its columns $\mathbf{AB}_j$ is a linear combination of the columns of **A**, with coefficients taken from the j^{th} column of **B**.

4. **Linear combinations of rows of B**. In the previous interpretation of matrix multiplication, we begin with a partitioning of **B** via vertical lines. If, instead, we begin with a partitioning of **A**, a matrix with m rows, via horizontal lines, we get

$$\mathbf{AB} = \left[\begin{array}{c} \mathbf{A}_1 \\ \hline \mathbf{A}_2 \\ \hline \vdots \\ \hline \mathbf{A}_m \end{array}\right] \mathbf{B} \;=\; \left[\begin{array}{c} \mathbf{A}_1\mathbf{B} \\ \hline \mathbf{A}_2\mathbf{B} \\ \hline \vdots \\ \hline \mathbf{A}_m\mathbf{B} \end{array}\right].$$

That is, the j^{th} row of the matrix product **AB** is obtained from left-multiplying the entire matrix **B** by the j^{th} *row* (considered as a submatrix) of **A**.

If **A** has n columns, then each $\mathbf{A}_j$ is a 1-by-n matrix. The effect of multiplying a 1-by-n matrix **V** by an n-by-p matrix **B**, using a blocking-by-row scheme for **B**, is

$$\mathbf{VB} \;=\; \left[\begin{array}{c|c|c|c} v_1 & v_2 & \cdots & v_n \end{array}\right] \left[\begin{array}{c} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \\ \hline \cdots \\ \hline \mathbf{B}_n \end{array}\right] \;=\; v_1\mathbf{B}_1 + v_2\mathbf{B}_2 + \cdots + v_n\mathbf{B}_n \;,$$

a linear combination of the rows of **B**. Thus, for each $j = 1, \ldots, m$, the j^{th} row $\mathbf{A}_j\mathbf{B}$ of the matrix product **AB** is a linear combination of the rows of **B**, with coefficients taken from the j^{th} row of **A**.

1.3.2 Systems of linear equations

Motivated by **Viewpoint 3** concerning matrix multiplication—in particular, that

$$\mathbf{Ax} \;=\; x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_n\mathbf{A}_n \;,$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_n$ are the columns of a matrix **A** and $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$—we make the following definition.

Definition 4: Suppose $\mathbf{A} = \left[\begin{array}{c|c|c|c} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_n \end{array}\right]$, where each submatrix $\mathbf{A}_j$ consists of a single column (so **A** has n columns in all). The set of all possible linear combinations of these columns (also known as $\text{span}(\{\mathbf{A}_1, \ldots, \mathbf{A}_n\})$)

$$\{c_1\mathbf{A}_1 + c_2\mathbf{A}_2 + \cdots + c_n\mathbf{A}_n \,|\, c_1, c_2, \ldots, c_n \in \mathbb{R}\} \;,$$

is called the **column space** of **A**. We use the symbol $\text{col}(\mathbf{A})$ to denote the column space.

The most common problem in linear algebra (and the one we seek in this course to understand most completely) is the one of solving m linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \vdots \quad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned} \tag{1.2}$$

in the n unknowns $x_1, \ldots, x_n$. If one uses the coefficients and unknowns to build a **coefficient matrix** and vectors

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

then by our definitions of matrix equality and multiplication, the system (1.2) may be expressed more concisely as the matrix equation

$$\mathbf{Ax} = \mathbf{b}\,, \tag{1.3}$$

where the vector $\mathbf{b}$ is known and $\mathbf{x}$ is to be found. Given **Viewpoint 3** for conceptualizing matrix multiplication above, problem (1.3) really presents two questions to be answered:

(I) Is $\mathbf{b}$ in the column space of $\mathbf{A}$ (i.e., is (1.3) solvable)?

(II) If it is, then what are the possible n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$ of coefficients so that the linear combination

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_n\mathbf{A}_n$$

of the columns of $\mathbf{A}$ equals $\mathbf{b}$?

When the number m of equations and the number n of unknowns in system (1.2) are equal, it is often the case that there is one unique answer for each of the variables x_i (or, equivalently, one unique vector $\mathbf{x}$ satisfying (1.3). Our main goal in the linear algebra component of this course is to understand completely when (1.3) is and is not solvable, how to characterize solutions when it is, and what to do when it is not.

One special instance of the case $m = n$ is when $\mathbf{A}$ is nonsingular. In this case, if $\mathbf{A}^{-1}$ is known, then the answer to question (I) is an immediate "yes". Moreover, one may obtain the (unique) solution of (1.3) (thus answering question (II)) via left-multiplication by $\mathbf{A}^{-1}$:

$$\begin{aligned} \mathbf{Ax} = \mathbf{b} \quad &\Rightarrow \quad \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \\ &\Rightarrow \quad \mathbf{Ix} = \mathbf{A}^{-1}\mathbf{b} \\ &\Rightarrow \quad \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}\,. \end{aligned}$$

Important Note: One should not think about the previous matrix-algebraic steps in terms of *dividing by a matrix* (and it is complete nonsense to talk about *dividing by a vector*!). One speaks, instead, of multiplying by the inverse matrix, when that exists. It is, moreover, extremely important to pay attention to which side of an expression you wish to multiply by that inverse. Often placing it on the wrong side yields a nonsensical mathematical expression!

In practical settings, however, $\mathbf{A}^{-1}$ must first be found (if, indeed, it exists!) before we can use it to solve the matrix problem. Despite the availability of the Octave function `inv()`, finding the inverse of a matrix is a very inefficient thing to do computationally, and quite impossible when $\mathbf{A}^{-1}$ does not exist. In the Section 1.5 we will look at *Gaussian elimination* as a procedure for solving linear systems of equations. Gaussian elimination serves as a foundation for the *LU*-factorization, which supplies us with a comprehensive method for solving $\mathbf{Ax} = \mathbf{b}$ whenever the matrix problem *can* be solved (even in cases where $\mathbf{A}^{-1}$ does not exist).

1.4 Affine transformations of $\mathbb{R}^2$

Suppose $\mathbf{A}$ is an m-by-n matrix. When we left-multiply a vector $\mathbf{v} \in \mathbb{R}^n$ by such a matrix $\mathbf{A}$, the result is a vector $\mathbf{Av} \in \mathbb{R}^m$. In this section we will focus upon functions which take inputs $\mathbf{v} \in \mathbb{R}^n$ and produce outputs $\mathbf{Av} \in \mathbb{R}^m$. A function such as this could be given a name, but we will generally avoid doing so, referring to it as "the function $\mathbf{v} \mapsto \mathbf{Av}$". When we wish to be explicit about the type of objects the input and output are, we might write "$(\mathbf{v} \mapsto \mathbf{Av})\colon \mathbb{R}^n \to \mathbb{R}^m$", which points out that the function $\mathbf{v} \mapsto \mathbf{Av}$ maps objects from $\mathbb{R}^n$ (inputs) to objects from $\mathbb{R}^m$ (outputs). But if the reader is informed that $\mathbf{A}$ is an m-by-n matrix, he should already be aware that inputs/outputs to and from the function $\mathbf{v} \mapsto \mathbf{Av}$ are in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively.

In this subsection $\mathbf{A}$ will be understood to be a 2-by-2 matrix. Assuming this, it is the case that $(\mathbf{v} \mapsto \mathbf{Av})\colon \mathbb{R}^2 \to \mathbb{R}^2$. We wish to focus our attention on the action of such a function on the entire plane of vectors for various types of 2-by-2 matrices $\mathbf{A}$.

1. **Rotations of the plane**. Our first special family of matrices are those of the form

$$\mathbf{A} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}, \tag{1.4}$$

for $\alpha \in \mathbb{R}$. We know that points in the plane may be specified using polar coordinates, so any vector $\mathbf{v} \in \mathbb{R}^2$ may be expressed as $\mathbf{v} = (r\cos\theta, r\sin\theta)$, where (r, θ) is a polar representation of the terminal point of $\mathbf{v}$. To the see the action of $\mathbf{A}$ on a typical $\mathbf{v}$, note that

$$\begin{aligned}\mathbf{A}\mathbf{v} &= \begin{bmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}r\cos\theta\\ r\sin\theta\end{bmatrix} = r\begin{bmatrix}\cos\alpha\cos\theta-\sin\alpha\sin\theta\\ \sin\alpha\cos\theta+\cos\alpha\sin\theta\end{bmatrix}\\ &= \begin{bmatrix}r\cos(\alpha+\theta)\\ r\sin(\alpha+\theta)\end{bmatrix}.\end{aligned}$$

where we have employed several **angle sum formulas**[2] in the last equality. That is, for an input vector $\mathbf{v}$ with terminal point (r, θ), the output $\mathbf{A}\mathbf{v}$ is a vector with terminal point $(r, \alpha + \theta)$. The output is the same distance r from the origin as the input, but has been rotated about the origin through an angle α. Thus, for matrices of the form (1.4), the function $\mathbf{v} \mapsto \mathbf{A}\mathbf{v}$ rotates the entire plane counterclockwise (for positive α) about the origin through an angle α. Of course, the inverse matrix would reverse this process, and hence it must be

$$\mathbf{A}^{-1} = \begin{bmatrix}\cos(-\alpha) & -\sin(-\alpha)\\ \sin(-\alpha) & \cos(-\alpha)\end{bmatrix} = \begin{bmatrix}\cos(\alpha) & \sin(\alpha)\\ -\sin(\alpha) & \cos(\alpha)\end{bmatrix}.$$

2. **Reflections across a line containing the origin**.

   First notice that, when

$$\mathbf{A} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}, \qquad \text{then} \qquad \mathbf{A}\mathbf{v} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix} = \begin{bmatrix}v_1\\ -v_2\end{bmatrix}.$$

   Thus, for this special matrix $\mathbf{A}$, $\mathbf{v} \mapsto \mathbf{A}\mathbf{v}$ maps points in the plane to their reflections through the x-axis.

   Now let $\mathbf{u} = (\cos\theta, \sin\theta)$ (i.e., $\mathbf{u}$ is a unit vector). Every line in the plane containing the origin may be expressed as a one-parameter family $L = \{t\mathbf{u} \mid t \in \mathbb{R}\}$ of multiples of $\mathbf{u}$ where θ has been chosen (fixed, hence fixing $\mathbf{u}$ as well) to be an angle the line makes with the positive x-axis. (Said another way, each line in $\mathbb{R}^2$ containing $\mathbf{0}$ is the *linear span* of some unit vector.) We can see reflections across the line L as a series of three transformations:

   i) rotation of the entire plane through an angle $(-\theta)$, so as to make the line L correspond to the x-axis,

   ii) reflection across the x-axis, and then

   iii) rotation of the plane through an angle θ, so that the x-axis is returned back to its original position as the line L.

[2]These trigonometric identities appear, for instance, in the box marked equation (4) on p. 26 of **University Calculus**, by Hass, Weir and Thomas.

These three steps may be affected through successive multiplications of matrices (the ones on the left below) which can be combined into one:

$$\begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix} = \begin{bmatrix}\cos\alpha & \sin\alpha\\ \sin\alpha & -\cos\alpha\end{bmatrix}, \tag{1.5}$$

where $\alpha = 2\theta$. That is, a matrix of the form

$$\mathbf{A} = \begin{bmatrix}\cos\alpha & \sin\alpha\\ \sin\alpha & -\cos\alpha\end{bmatrix} \tag{1.6}$$

will map points in the plane to their mirror images across a line that makes an angle $(\alpha/2)$ with the positive x-axis.

3. **Scaling relative to the origin: perpendicular lines case**. Suppose we wish to rescale vectors so that the x-coordinate of terminal points is multiplied by the quantity s, while the y-coordinates are multiplied by t. It is easy to see that multiplication by the matrix

$$\begin{bmatrix}s & 0\\ 0 & t\end{bmatrix} \tag{1.7}$$

would achieve this, since

$$\begin{bmatrix}s & 0\\ 0 & t\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix} = \begin{bmatrix}sv_1\\ tv_2\end{bmatrix}.$$

It is, in fact, only slightly more complicated to do this in directions specified by any pair of perpendicular lines, not just the x- and y-axes. It is left as an exercise to figure out how.

4. **Translations of the plane**. What we have in mind here is, for some given vector $\mathbf{w}$, to translate every vector $\mathbf{v} \in \mathbb{R}^2$ to the new location $\mathbf{v} + \mathbf{w}$. It is an easy enough mapping, described simply in symbols by $(\mathbf{v} \mapsto \mathbf{v} + \mathbf{w})$. Yet, perhaps surprisingly, it is the one type of **affine transformation** (most *affine transformations* of the plane are either of the type 1–4 described here, or combinations of these) which cannot be achieved through left-multiplication by a 2-by-2 matrix. That is, for a given $\mathbf{w} \neq \mathbf{0}$ in $\mathbb{R}^2$, there is no 2-by-2 matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{v} = \mathbf{v} + \mathbf{w}$.

When this observation became apparent to computer programmers writing routines for motion in computer graphics, mathematicians working in the area of *projective geometry* had a ready answer: **homogeneous coordinates**. The idea is to embed vectors from $\mathbb{R}^2$ into $\mathbb{R}^3$. A vector $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ is associated with the vector $\tilde{\mathbf{v}} = (v_1, v_2, 1)$ which lies on the plane $z = 1$ in $\mathbb{R}^3$. Say we want to translate all vectors $\mathbf{v} \in \mathbb{R}^2$ by the (given) vector $\mathbf{w} = (a, b)$. We can form the 3-by-3 matrix

$$\mathbf{A} = \begin{bmatrix}1 & 0 & a\\ 0 & 1 & b\\ 0 & 0 & 1\end{bmatrix}$$

and multiply $\tilde{\mathbf{v}}$ (not $\mathbf{v}$ itself) by $\mathbf{A}$. Then the translated vector $\mathbf{v} + \mathbf{w}$ is obtained from $\mathbf{A}\tilde{\mathbf{v}}$ by keeping just the first two coordinates.

We finish this section with two comments. First, we note that even though we needed homogeneous coordinates only for the translations described in 4 above, it is possible to carry out the transformations of 1–3 while in homogeneous coordinates as well. This is possible because we may achieve the appropriate analog of any of the transformations 1–3 by multiplying $\tilde{\mathbf{v}}$ by a 3-by-3 block matrix

$$\mathbf{B} = \left[\begin{array}{c|c} \mathbf{A} & 0 \\ \hline 0 & 1 \end{array}\right],$$

where $\mathbf{A}$ is a 2-by-2 block as described in 1–3 above. You should convince yourself, reviewing what we have said about multiplying matrices in blocks (**Viewpoint 1** in Section 1.3) as needed, that, if $\tilde{\mathbf{v}}$ is the homogeneous coordinates version of $\mathbf{v}$, then $\mathbf{B}\tilde{\mathbf{v}}$ is the homogeneous coordinates version of $\mathbf{A}\mathbf{v}$.

The other note to mention is that, while our discussion has been entirely about affine transformations on $\mathbb{R}^2$, all of the types we have discussed in 1–4 have counterpart transformations on $\mathbb{R}^n$, when $n > 2$. For instance, if you take a plane in $\mathbb{R}^3$ containing the origin and affix to it an axis of rotation passing perpendicularly to that plane through the origin then, for a given angle α, there is a 3-by-3 matrix $\mathbf{A}$ such that rotations of points in $\mathbb{R}^3$ through the angle α about this axis are achieved via the function $(\mathbf{v} \mapsto \mathbf{A}\mathbf{v})\colon \mathbb{R}^3 \to \mathbb{R}^3$. The 3D analogs of transformations 2–3 may be similarly achieved via multiplication by an appropriate 3-by-3 matrix. Only transformation 4 cannot, requiring, as before, that we pass into one higher dimension (homogeneous coordinates for $\mathbb{R}^3$) and multiply by an appropriate 4-by-4 matrix.

1.5 Gaussian Elimination

We have noted that linear systems of (algebraic) equations are representable in matrix form. We now investigate the solution of such systems. We begin with a definition.

Definition 5: An m-by-n matrix $\mathbf{A} = (a_{ij})$ is said to be **upper triangular** if $a_{ij} = 0$ whenever $i > j$—that is, when all entries below the main diagonal are zero. When all entries above the main diagonal are zero (i.e., $a_{ij} = 0$ whenever $i < j$), the $\mathbf{A}$ is said to be **lower triangular**. A *square* matrix that is both upper and lower triangular is called a **diagonal matrix**.

The system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$$

has two equations in two unknowns. The variables x_1 and x_2 represent *potential* **degrees of freedom**, while the equations represent **constraints**. The solution(s) are points of intersection between two lines, and as such there may be none, precisely one, or infinitely many.

In real settings, there are usually far more than two variables and equations, and the apparent number of constraints need not be the same as the number of variables. We would like a general algorithm which finds solutions to such systems of equations when solutions exist. We will develop a method called **Gaussian elimination** and, in the process, look at examples of various types of scenarios which may arise.

1.5.1 Examples of the method

Example 5:

We begin simply, with a system of 2 equations in 2 unknowns. Suppose we wish to solve

$$\begin{aligned} 7x + 3y &= 1 \\ 3y &= -6 \end{aligned} \qquad \text{or} \quad \mathbf{Ax} = \mathbf{b}\,, \quad \text{with} \qquad \mathbf{A} = \begin{bmatrix} 7 & 3 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ -6 \end{bmatrix}.$$

The problem is very easy to solve using **backward substitution**—that is, solving the equation in y alone,

$$3y = -6 \qquad \Rightarrow \qquad y = -2\,,$$

which makes the appearence of y in the other equation no problem:

$$7x + 3(-2) = 1 \qquad \Rightarrow \qquad x = \frac{1}{7}(1+6) = 1\,.$$

We have the unique solution $(1, -2)$. Notice that we can solve by backward substitution because the **coefficient matrix A** is *upper triangular*.

■

Example 6:

The system

$$\begin{aligned} 2x - 3y &= 7 \\ 3x + 5y &= 1 \end{aligned} \qquad \text{or, in matrix form} \qquad \begin{bmatrix} 2 & -3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \end{bmatrix},$$

is only mildly more difficult, though we cannot immediately resort to backward substitution as in the last example. Let us proceed by making this problem like that from the previous example. Perhaps we might leave the top equation alone, but alter the bottom one by adding (−2/3) multiples of the top equation to it. In what follows, we will employ the approach of listing the algebraic equations on the left along with a matrix form of them on the right. Instead of repeating the full matrix equation, we will abbreviate it with a matrix called an **augmented matrix** that lists only the constants in the problem. By adding (−3/2) copies of the top equation to the bottom, our original system

$$\begin{aligned} 2x - 3y &= 7 \\ 3x + 5y &= 1 \end{aligned} \qquad \text{or} \qquad \left[\begin{array}{cc|c} 2 & -3 & 7 \\ 3 & 5 & 1 \end{array}\right],$$

becomes

$$\begin{aligned} 2x - 3y &= 7 \\ (19/2)y &= -19/2 \end{aligned} \qquad \text{or} \qquad \left[\begin{array}{cc|c} 2 & -3 & 7 \\ 0 & 19/2 & -19/2 \end{array}\right].$$

Now, as the (new) coefficient matrix (the part of the matrix lying to the left of the dividing line) is upper triangular, we may finish solving our system using backward substitution:

$$\frac{19}{2}y \;=\; -\frac{19}{2} \qquad \Rightarrow \qquad y \;=\; -1\,,$$

so

$$2x - 3(-1) \;=\; 7 \qquad \Rightarrow \qquad x \;=\; 7 + 3 \;=\; 10\,.$$

Again, we have a unique solution, the point (10, −1).

■

Let's pause for some observations. Hopefully it is clear that an upper triangular system is desirable so that backward substitution may be employed to find appropriate values for the variables. When we did not immediately have that in Example 6, we added a multiple of the first equation to the second to make it so. This is listed below as number 3 of the elementary operations which are allowed when carrying out **Gaussian elimination**, the formal name given to the process of reducing a system of linear equations to a special form which is then easily solved by substitution. Your intuition about solving equations should readily confirm the validity of the other two elementary operations.

Elementary Operations of Gaussian Elimination

1. Multiply a row by a nonzero constant.
2. Exchange two rows.
3. Add a multiple of one row to another.

And what is this special form at which Gaussian elimination aims? It is an upper triangular form, yet not merely that. It is is a special form known as **echelon form** where the first nonzero entries in each row, below depicted by 'p's and asterisks, have a stair-step appearance to them:

$$\begin{bmatrix} p & * & * & * & * & * & * & * & * & \cdots \\ 0 & p & * & * & * & * & * & * & * & \cdots \\ 0 & 0 & 0 & p & * & * & * & * & * & \cdots \\ 0 & 0 & 0 & 0 & p & * & * & * & * & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p & * & \cdots \\ \vdots & & & & & & & & & \end{bmatrix}$$

In fact, there may also be zeros where the asterisks appear. The 'p's, however, called **pivots**, play a special role in the backward substitution part of the solution process, a role that requires them to be nonzero. If you look back at the pivots in our first two examples (the numbers 7 and 3 in Example 5; 2 and (19/2) in Example 6), you will see why they must be nonzero—when we get to the backward substitution stage, we divide through by these pivots. But, as the echelon form above depicts, the number of rows and columns of a matrix does not tell you just how many pivots you will have. The pivot in one row may be followed by a pivot in the next row (progressing downward) which is just one column to the right; but, that next pivot down may also skip *several* columns to the right. The final pivot may not even be in the right-most column. One thing for sure is that the pivots must progress to the right as we move down the rows; all entries below each pivot must be zero.

It is usually necessary to perform a sequence of elementary row operations on a given matrix **A** before arriving at an echelon form **R** (another m-by-n matrix). It would violate our definition of *matrix equality* to call **A** and **R** "equal". Instead, we might say that **R** is an echelon form for **A** (not "the" echelon form for **A**, as there is more than one), or that **A** and **R** are **row equivalent**.

We turn now to examples of the process for larger systems, illustrating some different scenarios in the process, and some different types of problems we might solve using it. After stating the original problem, we will carry out the steps depicting only augmented matrices. Since the various augmented matrices are not equal to their predecessors (in the sense of matrix equality), but do represent equivalent systems of equations (i.e., systems of equations which have precisely the same solutions), we will separate them with the $\sim$ symbol.

Example 7:

Find all solutions to the linear system of equations

$$\begin{aligned} 2x + y - z &= 3\,, \\ 4x + 2y + z &= 9\,. \end{aligned}$$

As these two equations both represent planes in 3-dimensional space, one imagines that there may either be no solutions, or infinitely many. We perform Gaussian elimination:

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 3 \\ 4 & 2 & 1 & 9 \end{array}\right] \quad \begin{array}{c} -2\mathbf{r}_1 + \mathbf{r}_2 \to \mathbf{r}_2 \\ \sim \end{array} \quad \left[\begin{array}{ccc|c} 2 & 1 & -1 & 3 \\ 0 & 0 & 3 & 3 \end{array}\right]$$

The latter matrix is in echelon form. It has pivots, 2 in the 1st column and 3 in the 3rd column. Unlike previous examples, these pivots are separated by a column which has no pivot. This 2nd column continues to correspond to y-terms in the system, and the absence of a pivot in this column means that y is a **free variable**. It has no special value, providing a **degree of freedom** within solutions of the system. The **pivot columns** (i.e., the ones with pivots), correspond to the x- and z-terms in the system—the **pivot variables**; their values are either fixed, or contingent on the value(s) chosen for the free variable(s). The echelon form corresponds to the system of equations (equivalent to our original system)

$$\begin{aligned} 2x + y - z &= 3 \,, \\ 3z &= 3 \,. \end{aligned}$$

Clearly, the latter of these equations implies $z = 1$. Since y is free, we do not expect to be able to solve for it. Nevertheless, if we plug in our value for z, we may solve for x in terms of the free variable y:

$$x \;=\; \frac{1}{2}(3 + 1 - y) \;=\; 2 - \frac{1}{2}\,y \,.$$

Thus, our solutions (there are infinitely many) are

$$(x, y, z) \;=\; (2 - y/2, y, 1) \;=\; (2, 0, 1) + t(-1, 2, 0) \,,$$

where $t = y/2$ may be any real number (since y may be any real number). Note that this set S of solutions traces out a line in 3D space.

■

Before the next example, we make another definition.

Definition 6: The *nullspace* of an m-by-n matrix $\mathbf{A}$ consists of those vectors $\mathbf{x} \in \mathbb{R}^n$ for which $\mathbf{A}\mathbf{x} = \mathbf{0}$. That is,

$$\text{null}\,(\mathbf{A}) \;:=\; \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{x} = \mathbf{0}\} \,.$$

A related problem to the one in the last example is the following one.

Example 8:

Find the nullspace of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & -1 \\ 4 & 2 & 1 \end{bmatrix}.$$

That is, we are asked to find those vectors $\mathbf{v} \in \mathbb{R}^3$ for which $\mathbf{Av} = \mathbf{0}$ or, to put it in a way students in a high school algebra class might understand, to solve

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

Mimicking our work above, we have

$$\left[\begin{array}{ccc|c} 2 & 1 & -1 & 0 \\ 4 & 2 & 1 & 0 \end{array}\right] \quad \begin{array}{c} -2\mathbf{r}_1 + \mathbf{r}_2 \to \mathbf{r}_2 \\ \sim \end{array} \quad \left[\begin{array}{ccc|c} 2 & 1 & -1 & 0 \\ 0 & 0 & 3 & 0 \end{array}\right],$$

which corresponds to the system of equations

$$\begin{aligned} 2x + y - z &= 0\,, \\ 3z &= 0\,. \end{aligned}$$

Now, we have $z = 0$; y is again a free variable, so $x = -\frac{1}{2}y$. Thus, our solutions (again infinitely many) are

$$(x, y, z) = (-y/2, y, 0) = t(-1, 2, 0)\,,$$

where $t = y/2$ may be any real number (since y may be any real number). Note that, like the solutions in Example 7, this set of solutions—all scalar multiples of the vector $(-1, 2, 0)$—traces out a line. This line is parallel to that of the previous example, but unlike the other, it passes through origin (or zero vector).

■

Compare the original systems of equations and corresponding solutions of Examples 7 and 8. Employing a term used earlier to describe some linear ODEs, the system of equations in Example 8 is said to be **homogeneous** as its right-hand side is the zero vector. Its solutions form a line through the origin, a line parametrized by t. Since the vector on the right-hand side of Example 7 is $(9, 3)$ (*not* the zero vector), that system is **nonhomogeneous**. Its solutions form a line as well, parallel to the line for the corresponding homogeneous system of Example 8, but translated away from the origin by the vector $(2, 0, 1)$ which, itself, is a solution of the nonhomogeneous system of Example 7. We shall see similar behavior in the solution of linear ODEs: when faced with a nonhomogeneous n^{th}-order ODE (just an ODE to solve, not an initial-value problem), one finds the general solution of the corresponding homogeneous problem, an n-parameter family of solutions, and then adds to it a particular solution of the nonhomogenous problem.

We finish with an example of describing the column space of a matrix.

Example 9:

Find the column (or range) space of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 0 & -1 \\ 1 & 0 & 3 & 1 \\ -3 & -5 & 1 & 2 \\ 1 & 0 & 3 & 1 \end{bmatrix}.$$

A quick description of the column space of **A** is to say it is

$$\text{span}(\{(2,1,-3,1),(3,0,-5,0),(0,3,1,3),(-1,1,2,1)\}) .$$

Since that is so easy, let's see if we can give a more minimal answer. After all, there may be redundancy in these columns.

Our plan of attack will be to assume that $\mathbf{b} = (b_1, b_2, b_3, b_4)$ is in col(**A**) and solve $\mathbf{Ax} = \mathbf{b}$ via elimination as before. We have

$$\left[\begin{array}{cccc|c} 2 & 3 & 0 & -1 & b_1 \\ 1 & 0 & 3 & 1 & b_2 \\ -3 & -5 & 1 & 2 & b_3 \\ 1 & 0 & 3 & 1 & b_4 \end{array}\right] \quad \begin{array}{c} \mathbf{r}_1 \leftrightarrow \mathbf{r}_2 \\ \sim \end{array} \quad \left[\begin{array}{cccc|c} 1 & 0 & 3 & 1 & b_2 \\ 2 & 3 & 0 & -1 & b_1 \\ -3 & -5 & 1 & 2 & b_3 \\ 1 & 0 & 3 & 1 & b_4 \end{array}\right]$$

$$\begin{array}{c} \mathbf{r}_2 - 2\mathbf{r}_1 \to \mathbf{r}_2 \\ 3\mathbf{r}_1 + \mathbf{r}_3 \to \mathbf{r}_3 \\ \sim \\ \mathbf{r}_4 - \mathbf{r}_1 \to \mathbf{r}_4 \end{array} \quad \left[\begin{array}{cccc|c} 1 & 0 & 3 & 1 & b_2 \\ 0 & 3 & -6 & -3 & b_1 - 2b_2 \\ 0 & -5 & 10 & 5 & 3b_2 + b_3 \\ 0 & 0 & 0 & 0 & b_4 - b_2 \end{array}\right]$$

$$\begin{array}{c} (5/3)\mathbf{r}_2 + \mathbf{r}_3 \to \mathbf{r}_3 \\ \sim \end{array} \quad \left[\begin{array}{cccc|c} 1 & 0 & 3 & 1 & b_2 \\ 0 & 3 & -6 & -3 & b_1 - 2b_2 \\ 0 & 0 & 0 & 0 & (5/3)b_1 - (1/3)b_2 + b_3 \\ 0 & 0 & 0 & 0 & b_4 - b_2 \end{array}\right].$$

For determining the range space, we focus on the last two rows which say

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = \frac{5}{3}b_1 - \frac{1}{3}b_2 + b_3 \qquad \text{or} \qquad 0 = 5b_1 - b_2 + 3b_3 ,$$

and

$$0 = b_4 - b_2 .$$

These are the constraints which must be met by the components of **b** in order to be in col(**A**). There are two constraints on four components, so two of those components are "free". We choose $b_4 = t$, which means $b_2 = t$ as well. Thus

$$b_1 = \frac{1}{5}(t - 3b_3) .$$

If we take $b_3 = -5s - 3t$, then $b_1 = 2t + 3s$. (Admittedly, this is a strange choice for b_3. However, even if t is fixed on some value, the appearance of the new parameter s makes it possible for b_3 to take on any value.) So, we have that any $\mathbf{b} \in \text{col}(\mathbf{A})$ must take the form

$$\mathbf{b} = \begin{bmatrix} 2t + 3s \\ t \\ -3t - 5s \\ t \end{bmatrix} = t \begin{bmatrix} 2 \\ 1 \\ -3 \\ 1 \end{bmatrix} + s \begin{bmatrix} 3 \\ 0 \\ -5 \\ 0 \end{bmatrix},$$

where s, t are arbitrary real numbers. That is,

$$\text{col}(\mathbf{A}) = \text{span}(\{(2, 1, -3, 1), (3, 0, -5, 0)\}) .$$

When we look back at the original matrix, these two vectors in the spanning set for col(**A**) are precisely the first two columns of **A**. Thus, while we knew before we started that col(**A**) was spanned by the columns of **A**, we now know just the first two columns suffice.

We will return to this problem of finding col(**A**) in the next chapter, in which we will see there is an easier way to determine a set of vectors that spans the columns space.

■

1.5.2 Finding an inverse matrix

What would you do if you had to solve

$$\mathbf{A}\mathbf{x} = \mathbf{b}_1 \qquad \text{and} \qquad \mathbf{A}\mathbf{x} = \mathbf{b}_2 ,$$

where the matrix **A** is the same but $\mathbf{b}_1 \neq \mathbf{b}_2$? Of course, one answer is to augment the matrix **A** with the first right-hand side vector $\mathbf{b}_1$ and solve using Gaussian elimination. Then, repeat the process with $\mathbf{b}_2$ in place of $\mathbf{b}_1$. But a close inspection of the process shows that the row operations you perform on the augmented matrix to reach row echelon form are dictated by the entries of **A**, independent of the right-hand side. Thus, one could carry out the two-step process we described more efficiently if one augmented **A** with two extra columns, $\mathbf{b}_1$ and $\mathbf{b}_2$. That is, working with the augmented matrix $[\mathbf{A}|\mathbf{b}_1\, \mathbf{b}_2]$, use elementary (row) operations 1.–3. until the part to the left of the augmentation bar is in row echelon form. This reduced augmented matrix would take the form $[\mathbf{R}|\mathbf{c}_1\, \mathbf{c}_2]$, where **R** is a row echelon form. Then we could use backward substitution separately on $[\mathbf{R}|\mathbf{c}_1]$ and $[\mathbf{R}|\mathbf{c}_2]$ to find solutions to the two matrix problems. Of course, if you have more than 2 matrix problems (with the same coefficient matrix), you tack on more than 2 columns.

This idea is key to findin the inverse of an n-by-n matrix **A**, when it exists. Let us denote the standard vectors in $\mathbb{R}^n$ by $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, \ldots, 0), \ldots, \mathbf{e}_n = (0, 0, \ldots, 1)$. These are the columns of the identity matrix. We know the inverse $\mathbf{B}$ satisfies

$$\mathbf{AB} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \cdots & \mathbf{e}_n \\ \downarrow & \downarrow & \downarrow & & \downarrow \end{bmatrix}.$$

Let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$ denote the columns of $\mathbf{B}$. Then an equivalent problem to finding the matrix $\mathbf{B}$ is to solve the n problems

$$\mathbf{Ab}_1 = \mathbf{e}_1\ , \quad \mathbf{Ab}_2 = \mathbf{e}_2\ , \quad \ldots \quad \mathbf{Ab}_n = \mathbf{e}_n\ ,$$

for the unknown vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n$. By the method above, we would augment $\mathbf{A}$ with the full n-by-n identity matrix and perform elementary operations until the part to the left of the augmentation line was in row echelon form. That is, we would reduce $[\mathbf{A}|\mathbf{I}]$ to $[\mathbf{R}|\mathbf{C}]$, where $\mathbf{C}$ is an n-by-n matrix. (Note that if $\mathbf{R}$ does not have n pivots, then $\mathbf{A}$ is singular.) We can then solve each of the problems

$$\mathbf{Rb}_1 = \mathbf{c}_1, \quad \mathbf{Rb}_2 = \mathbf{c}_2, \quad \ldots \quad \mathbf{Rb}_n = \mathbf{c}_n$$

using backward substitution, and arrive at the inverse matrix $\mathbf{B}$ putting the solutions together:

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}.$$

That's a clever method, and it's pretty much along the lines of how an inverse matrix is found when it is really desired. However, in most cases, $\mathbf{A}^{-1}$ is just an intermediate find on the way to solving a matrix problem $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{x}$. If there is a more efficient way to find $\mathbf{x}$, one requiring fewer calculations, we would employ it instead. That is the content of the next section.

1.6 *LU* Factorization of a Matrix

We have three 'legal' elementary operations when using Gaussian elimination to solve the equation $\mathbf{Ax} = \mathbf{b}$. We seek to put the matrix $\mathbf{A}$ in *echelon form* via a sequence of operations consisting of

1. multiplying a row by a nonzero constant.
2. exchanging two rows.
3. adding a multiple of one row to another.

You may have noticed that, at least in theory, reduction to echelon form may be accomplished without ever employing operation 1. Let us focus on operation 3 for the moment. In practice the multiplier is always some nonzero constant β. Moreover, in Gaussian elimination we are primarily concerned with adding a multiple of a row to some other row which is *below* it. For a fixed β, let $\mathbf{E}_{ij} = \mathbf{E}_{ij}(\beta)$ be the matrix that only differs from the m-by-m identity matrix in that its $(i, j)^{\text{th}}$ entry is β. We call $\mathbf{E}_{ij}$ an **elementary matrix**. A user-defined function written in Octave code that returns such a matrix might look like the following:

```
function emat = elementary(m, i, j, val)
  emat = eye(m,m);
  emat(i,j) = val;
endfn
```

In Exercise 1.26 you are asked to show that

$$\mathbf{E}_{ij}(\beta)\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ \vdots & \vdots & & \vdots \\ a_{i-1,1} & a_{i-1,2} & \cdots & a_{i-1,n} \\ a_{i,1} + \beta a_{j,1} & a_{i,2} + \beta a_{j,2} & \cdots & a_{i,n} + \beta a_{j,n} \\ a_{i+1,1} & a_{i+1,2} & \cdots & a_{i+1,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

In other words, pre-multiplication by $\mathbf{E}_{ij}$ performs an instance of operation 3 on the matrix $\mathbf{A}$, replacing row i with (row i) $+\beta$ (row j). Now, suppose $a_{11} \neq 0$. Then $\mathbf{E}_{21}(-a_{21}/a_{11})\mathbf{A}$ is a matrix whose entry in its $(2, 1)^{\text{th}}$ position has been made to be zero. More generally,

$$\mathbf{E}_{n1}\left(\frac{-a_{n1}}{a_{11}}\right)\cdots\mathbf{E}_{31}\left(\frac{-a_{31}}{a_{11}}\right)\mathbf{E}_{21}\left(\frac{-a_{21}}{a_{11}}\right)\mathbf{A}$$

is the matrix that results from retaining the first pivot of $\mathbf{A}$ and eliminating all entries below it. If our matrix $\mathbf{A}$ is such that no row exchanges occur during reduction to echelon form, then by a sequence of pre-multiplications by elementary matrices, we arrive at an upper-triangular matrix $\mathbf{U}$. We make the following observations:

- Each time we perform elementary operation 3 via pre-multiplication by an elementary matrix $\mathbf{E}_{ij}$, it is the case that $i > j$. Thus, the elementary matrices we use are lower-triangular.
- Each elementary matrix is invertible, and $\mathbf{E}_{ij}^{-1}$ is lower triangular when $\mathbf{E}_{ij}$ is. See Exercise 1.26.
- The product of lower triangular matrices is again lower triangular. See Exercise 1.28.

By these observations, when no row exchanges take place in the reduction of $\mathbf{A}$ to echelon form, we may amass the sequence of elementary matrices which achieve this reduction into a single matrix $\mathbf{M}$ which is lower-triangular. Let us denote the inverse of $\mathbf{M}$ by $\mathbf{L}$, also a lower-triangular matrix. Then

$$\mathbf{LM} = \mathbf{I}, \qquad \text{while} \qquad \mathbf{MA} = \mathbf{U},$$

where **U** is an upper-triangular matrix, an echelon form for **A**. Thus,

$$\mathbf{A} \;=\; (\mathbf{LM})\mathbf{A} \;=\; \mathbf{L}(\mathbf{MA}) \;=\; \mathbf{LU},$$

which is called the *LU* **factorization of A**, and

$$\mathbf{Ax} \;=\; \mathbf{b} \qquad \Leftrightarrow \qquad \mathbf{LUx} = \mathbf{b}\ .$$

Let $\mathbf{y} = \mathbf{Ux}$, so that $\mathbf{Ly} = \mathbf{b}$. Since **L** is lower-triangular, we may solve for **y** by a process known as **forward substitution**. Once we have **y**, we may solve for $\mathbf{Ux} = \mathbf{y}$ via backward substitution as in the previous section.

But let us not forget that the previous discussion was premised on the idea that no row exchanges take place in order to reduce A to echelon form. We are aware that, in some instances, row exchanges are absolutely necessary to bring a pivot into position. As it turns out, numerical considerations sometimes call for row exchanges even when a pivot would be in place without such an exchange. How does this affect the above discussion?

Suppose we can know in advance just which row exchanges will take place in reducing **A** to echelon form. With such knowledge, we can quickly write down an m-by-m matrix **P**, called a **permutation matrix**, such that **PA** is precisely the matrix **A** except that all of those row exchanges have been carried out. For instance, if we ultimately want the 1st row of **A** to wind up as row 5, we make the the 5th row of **P** be $(1, 0, 0, \ldots, 0)$. More generally, if we want the i^{th} row of **A** to wind up as the j^{th} row, we make the j^{th} row of **P** have a 1 in the i^{th} column and zeros everywhere else. To illustrate this, suppose

$$\mathbf{P} \;=\; \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Then, for any 4-by-n matrix **A**, **PA** will be another 4-by-n whose 1^{st} row is equal to the 2^{nd} row of **A**, whose 2^{nd} row equals the 4^{th} row of **A**, whose 3^{rd} row equals the 3^{rd} row of **A**, and whose 4^{th} row equals the 1^{st} row of **A**.

Now, the full story about the *LU* decomposition can be told. There is a permutation matrix **P** such that **PA** will not need any row exchanges to be put into echelon form. It is this **PA** which has an *LU* decomposition. That is, $\mathbf{PA} = \mathbf{LU}$.

Example 10:

In Octave, the following commands were entered with accompanying output:

```
octave:1> A = [0 -1 3 1; 2 -1 1 4; 1 3 1 -1];
octave:2> [L,U,P] = lu(A)
L =
   1.00000    0.00000    0.00000
   0.50000    1.00000    0.00000
   0.00000   -0.28571    1.00000
```

```
U =
     2.00000   -1.00000    1.00000    4.00000
     0.00000    3.50000    0.50000   -3.00000
     0.00000    0.00000    3.14286    0.14286

P =
   0   1   0
   0   0   1
   1   0   0
```

We will use it to solve the matrix equation $\mathbf{Ax} = \mathbf{b}$, with

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 3 & 1 \\ 2 & -1 & 1 & 4 \\ 1 & 3 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} -1 \\ 14 \\ 1 \end{bmatrix}.$$

Since we have been given the *LU* decomposition for **PA**, we will use it to solve $\mathbf{PAx} = \mathbf{Pb}$—that is, solve

$$\begin{bmatrix} 2 & -1 & 1 & 4 \\ 1 & 3 & 1 & -1 \\ 0 & -1 & 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 14 \\ 1 \\ -1 \end{bmatrix}.$$

We first solve $\mathbf{Ly} = \mathbf{Pb}$, or[3]

$$\begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 0 & -2/7 & 1 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 14 \\ 1 \\ -1 \end{bmatrix}.$$

We call our manner of solving for **y forward substitution** because we find the components of **y** in forward order, y_1 then y_2 then y_3.

$$\begin{aligned} y_1 &= 14\,, \\ \frac{1}{2}y_1 + y_2 &= 1 \qquad \Rightarrow \qquad y_2 = -6\,, \\ -\frac{2}{7}y_2 + y_3 &= -1 \qquad \Rightarrow \qquad y_3 = -\frac{19}{7}\,, \end{aligned}$$

so $\mathbf{y} = (14, -6, -19/7)$. Now we solve $\mathbf{Ux} = \mathbf{y}$, or

$$\begin{bmatrix} 2 & -1 & 1 & 4 \\ 0 & 7/2 & 1/2 & -3 \\ 0 & 0 & 22/7 & 1/7 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 14 \\ -6 \\ -19/7 \end{bmatrix},$$

[3]Asking OCTAVE to display $7 * \mathbf{L}$ shows that this is an exact representation of **L**.

via backward substitution. The result is infinitely many solutions, all with the form

$$\mathbf{x} \;=\; \begin{bmatrix} 73/11 \\ -35/22 \\ -19/22 \\ 0 \end{bmatrix} + t \begin{bmatrix} -17/11 \\ 19/22 \\ -1/22 \\ 1 \end{bmatrix}, \qquad t \in \mathbb{R}\,.$$

■

Of course, it is possible to automate the entire process—not just the part of finding the *LU*-factorization of **A**, but also the forward and backward substitution steps. And there are situations in which, for a given coefficient matrix **A**, a different kind of solution process for the matrix equation $\mathbf{Ax} = \mathbf{b}$ may, indeed, be more efficient than using the factorization $\mathbf{LU} = \mathbf{PA}$. The Octave command

```
octave> A \ b
ans =
   2.30569
   0.82917
  -0.99101
   2.80220
```

(with `A` and `b` defined as in the Example 10) is sophisticated enough to look over the matrix `A` and choose a suitable solution technique, producing a result. In fact, the solution generated by the command is one that lies along the line of solutions

$$\mathbf{x} \;=\; \left(\frac{73}{11}, -\frac{35}{22}, -\frac{19}{22}, 0\right) + t\left(-\frac{17}{11}, \frac{19}{22}, -\frac{1}{22}, 1\right), \qquad t \in \mathbb{R}\,,$$

found in Example 10, one occurring when $t \doteq 2.8022$. This, however, reveals a shortcoming of the '`A \ b`' command. It can find a particular solution, but when multiple solutions exist, it cannot find them all.

1.7 Determinants

In a previous section we saw how to use Gaussian elimination to solve linear systems of n equations in n unknowns. It is a process one can carry out without any knowledge of whether the linear algebraic system in question *has* a solution or, if it does, whether that solution is *unique* (central questions, as we have already seen, when studying initial value problems for ODEs). So long as one knows what to look for, answers to these fundamental questions do reveal themselves during the process. In this section, however, we investigate whether something about the existence and/or uniqueness of solutions may be learned *in advance* of carrying out Gaussian elimination.

1.7.1 The planar case

Consider the case of two lines in the plane

$$\begin{aligned} ax + by &= e \\ cx + dy &= f. \end{aligned} \tag{1.8}$$

In fact, we know that intersections between two lines can happen in any of three different ways:

1. the lines intersect at a unique point (i.e., solution exists and is unique),
2. the lines are coincident (that is, the equations represent the same line and there are infinitely many points of intersection; in this case a solution exists, but is not unique), or
3. the lines are parallel but not coincident (so that no solution exists).

Experience has taught us that it is quite easy to decide which of these situations we are in before ever attempting to solve a linear system of two equations in two unknowns. For instance, the system

$$\begin{aligned} 3x - 5y &= 9 \\ -5x + \frac{25}{3}y &= -15 \end{aligned}$$

obviously contains two representations of the same line (since one equation is a constant multiple of the other) and will have infinitely many solutions. In contrast, the system

$$\begin{aligned} x + 2y &= -1 \\ 2x + 4y &= 5 \end{aligned}$$

will have no solutions. This is the case because, while the left sides of each equation — the sides that contain the coefficients of x and y which determine the slopes of the lines — are in proportion to one another, the right sides are not in the same proportion. As a result, these two lines will have the same slopes but not the same y-intercepts. Finally, the system

$$\begin{aligned} 2x + 5y &= 11 \\ 7x - y &= -1. \end{aligned}$$

will have just one solution (one point of intersection), as the left sides of the equations are not at all in proportion to one another.

What is most important about the preceding discussion is that we can distinguish situation 1 (the lines intersecting at one unique point) from the others simply by looking at the coefficients a, b, c and d from equation (1.8). In particular, we can determine the ratios $a : c$ and $b : d$ and determine whether these ratios are the same or different. Equivalently, we can look at whether the quantity

$$ad - bc$$

is zero or not. If $ad - bc \neq 0$ then the system has one unique point of intersection, but if $ad - bc = 0$ then the system either has no points or infinitely many points of intersection. If we write equation (1.8) as a matrix equation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix},$$

we see that the quantity $ad - bc$ is dependent only upon the coefficient matrix. Since this quantity "determines" whether or not the system has a unique solution, it is called the *determinant* of the coefficient matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

and is sometimes abbreviated as $\det(\mathbf{A})$, $|\mathbf{A}|$ or

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

1.7.2 Calculating determinants for n-square matrices, with $n > 2$

While it is quite easy for us to determine in advance the number of solutions which arise from a system of two linear equations in two unknowns, the situation becomes a good deal more complicated if we add another variable and another equation. The solutions of such a system

$$\begin{aligned} ax + by + cz &= l \\ dx + ey + fz &= m \\ ex + hy + kz &= n \end{aligned} \tag{1.9}$$

can be thought of as points of intersection between three planes. Again, there are several possibilities:

1. the planes intersect at a unique point,
2. the planes intersect along a line,
3. the planes intersect in a plane, or
4. the planes do not intersect.

It seems reasonable to think that situation 1 can once again be distinguished from the other three simply by performing some test on the numbers a, b, c, d, e, f, g, h and k. As in the case of the system (1.8), perhaps if we write system (1.9) as the matrix equation

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} l \\ m \\ n \end{bmatrix},$$

we will be able to define an appropriate quantity $|\mathbf{A}|$ that depends only on the coefficient matrix

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$$

in such a way that, if $|\mathbf{A}| \neq 0$ then the system has a unique solution (situation 1), but if $|\mathbf{A}| = 0$ then one of the other situations (2–4) is in effect.

Indeed it is possible to define $|\mathbf{A}|$ for a square matrix $\mathbf{A}$ of arbitrary dimension. For our purposes, we do not so much wish to give a rigorous definition of such a determinant as we do wish to be able to find it. As of now, we do know how to find it for a 2×2 matrix:

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

For square matrices of dimension larger than 2 we will find the determinant using *cofactor expansion*.

Let $\mathbf{A} = (a_{ij})$ be an arbitrary $n \times n$ matrix; that is,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

We define the (i, j)**-minor** of $\mathbf{A}$, M_{ij}, to be the determinant of the matrix resulting from crossing out the i^{th} row and the j^{th} column of $\mathbf{A}$. Thus, if

$$\mathbf{B} = \begin{bmatrix} 1 & -4 & 3 \\ -3 & 2 & 5 \\ 4 & 0 & -1 \end{bmatrix},$$

we have nine possible minors M_{ij} of $\mathbf{B}$, two of which are

$$M_{21} = \begin{vmatrix} -4 & 3 \\ 0 & -1 \end{vmatrix} = 4 \qquad \text{and} \qquad M_{33} = \begin{vmatrix} 1 & -4 \\ -3 & 2 \end{vmatrix} = -10.$$

A concept that is related to the (i, j)-minor is the **(i, j)-cofactor**, C_{ij}, which is defined to be

$$C_{ij} := (-1)^{i+j} M_{ij}.$$

Thus, the matrix $\mathbf{B}$ above has 9 cofactors C_{ij}, two of which are

$$C_{21} = (-1)^{2+1} M_{21} = -4 \qquad \text{and} \qquad C_{33} = (-1)^{3+3} M_{33} = -10.$$

Armed with the concept of cofactors, we are prepared to say how the determinant of an arbitrary square matrix $\mathbf{A} = (a_{ij})$ is found. It may be found by **expanding in cofactors** along the i^{th} row:

$$\det(\mathbf{A}) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{k=1}^{n} a_{ik}C_{ik}.$$

Or, alternatively, it may be found by expanding in cofactors along the j^{th} column:

$$\det(\mathbf{A}) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} = \sum_{k=1}^{n} a_{kj}C_{kj}.$$

It is perhaps surprising that expansion in cofactors gives you the same final result regardless of the row or column you choose. In practice, then, one usually chooses a row or column that contains the most zeros. Expanding in cofactors along the third row, we calculate the determinant of the 3×3 matrix $\mathbf{B}$ above as follows:

$$\begin{aligned}\det(\mathbf{B}) = \begin{vmatrix} 1 & -4 & 3 \\ -3 & 2 & 5 \\ 4 & 0 & -1 \end{vmatrix} &= 4(-1)^{3+1}\begin{vmatrix} -4 & 3 \\ 2 & 5 \end{vmatrix} + (0)(-1)^{3+2}\begin{vmatrix} 1 & 3 \\ -3 & 5 \end{vmatrix} + (-1)(-1)^{3+3}\begin{vmatrix} 1 & -4 \\ -3 & 2 \end{vmatrix} \\ &= 4(-20-6) + 0 - (2-12) \quad = \quad -94.\end{aligned}$$

If, instead, we expand in cofactors along the 2^{nd} column, we get

$$\begin{aligned}\det(\mathbf{B}) &= (-4)(-1)^{1+2}\begin{vmatrix} -3 & 5 \\ 4 & -1 \end{vmatrix} + (2)(-1)^{2+2}\begin{vmatrix} 1 & 3 \\ 4 & -1 \end{vmatrix} + (0)(-1)^{3+2}\begin{vmatrix} 1 & 3 \\ -3 & 5 \end{vmatrix} \\ &= 4(3-20) + 2(-1-12) + 0 \quad = \quad -94.\end{aligned}$$

You should verify that expansions along other rows or columns of $\mathbf{B}$ yield the same result.

This process can be used iteratively on larger square matrices. For instance

$$\begin{aligned}\begin{vmatrix} 3 & 1 & 2 & 0 \\ -1 & 0 & 5 & -4 \\ 1 & 1 & 0 & -1 \\ 0 & 0 & -3 & 1 \end{vmatrix} &= (0)(-1)^{4+1}\begin{vmatrix} 1 & 2 & 0 \\ 0 & 5 & -4 \\ 1 & 0 & -1 \end{vmatrix} + (0)(-1)^{4+2}\begin{vmatrix} 3 & 2 & 0 \\ -1 & 5 & -4 \\ 1 & 0 & -1 \end{vmatrix} \\ &\quad + (-3)(-1)^{4+3}\begin{vmatrix} 3 & 1 & 0 \\ -1 & 0 & -4 \\ 1 & 1 & -1 \end{vmatrix} + (1)(-1)^{4+4}\begin{vmatrix} 3 & 1 & 2 \\ -1 & 0 & 5 \\ 1 & 1 & 0 \end{vmatrix},\end{aligned}$$

where our cofactor expansion of the original determinant of a 4×4 matrix along its fourth row expresses it in terms of the determinants of several 3×3 matrices. We may proceed to find these latter determinants using cofactor expansions as well:

$$\begin{aligned}\begin{vmatrix} 3 & 1 & 0 \\ -1 & 0 & -4 \\ 1 & 1 & -1 \end{vmatrix} &= (3)(-1)^{1+1}\begin{vmatrix} 0 & -4 \\ 1 & -1 \end{vmatrix} + (1)(-1)^{1+2}\begin{vmatrix} -1 & -4 \\ 1 & -1 \end{vmatrix} + (0)(-1)^{1+3}\begin{vmatrix} -1 & 0 \\ 1 & 1 \end{vmatrix} \\ &= 3(0+4) - (1+4) + 0 \\ &= 7,\end{aligned}$$

where we expanded in cofactors along the first row, and

$$\begin{aligned}\begin{vmatrix} 3 & 1 & 2 \\ -1 & 0 & 5 \\ 1 & 1 & 0 \end{vmatrix} &= (1)(-1)^{1+2}\begin{vmatrix} -1 & 5 \\ 1 & 0 \end{vmatrix} + (0)(-1)^{2+2}\begin{vmatrix} 3 & 2 \\ 1 & 0 \end{vmatrix} + (1)(-1)^{3+2}\begin{vmatrix} 3 & 2 \\ -1 & 5 \end{vmatrix} \\ &= -(0-5) + 0 - (15+2) \\ &= -12,\end{aligned}$$

where this cofactor expansion was carried out along the second column. Thus

$$\begin{vmatrix} 3 & 1 & 2 & 0 \\ -1 & 0 & 5 & -4 \\ 1 & 1 & 0 & -1 \\ 0 & 0 & -3 & 1 \end{vmatrix} = 0 + 0 + (3)(7) + (1)(-12) = 9.$$

Now consider the linear algebraic systems

$$\begin{aligned} x_1 - 4x_2 + 3x_3 &= b_1 \\ -3x_1 + 2x_2 + 5x_3 &= b_2 \\ 4x_1 - x_3 &= b_3 \end{aligned} \qquad \text{whose coefficient matrix is} \qquad \begin{bmatrix} 1 & -4 & 3 \\ -3 & 2 & 5 \\ 4 & 0 & -1 \end{bmatrix}$$

(i.e., the matrix **B** above), and

$$\begin{aligned} 3x_1 + x_2 + 2x_3 &= b_1 \\ -x_1 + 5x_3 - 4x_4 &= b_2 \\ x_1 + x_2 - x_4 &= b_3 \\ -3x_3 + x_4 &= b_4 \end{aligned} \qquad \text{whose coefficient matrix is} \qquad \begin{bmatrix} 3 & 1 & 2 & 0 \\ -1 & 0 & 5 & -4 \\ 1 & 1 & 0 & -1 \\ 0 & 0 & -3 & 1 \end{bmatrix}.$$

We have computed the determinants of both of these coefficient matrices and found them to be nonzero. As in the case of planar linear systems (1.8), this tells us that both of these systems have unique solutions. That is, if we write each of these systems in matrix form $\mathbf{Ax} = \mathbf{b}$, with

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix},$$

then no matter what values are used for b_1, b_2, b_3 (and b_4 in the latter system), there is exactly one solution for the unknown vector $\mathbf{x} = (x_1, x_2, x_3, x_4)$.

You should endeavor to develop proficiency in use of cofactor expansion to find determinants. When you are not practicing this skill, however, you may rely on software or a calculator to find determinants for you, particularly in the case of a square matrix with $n \geq 4$ columns. In Octave the command that calculates the determinant of a matrix `A` is `det(A)`.

1.7.3 Some facts about determinants

Next, we provide some important facts about determinants. This is not a summary of things demonstrated in the section, though perhaps some of them may seem intuitively correct from what you have learned about computing determinants. The list is not arranged in a sequence that indicates a progression in level of importance though, for our purposes, letter F is probably the most important.

A. The idea of the determinant of a matrix does not extend to matrices which are non-square. We only talk about the determinant of square matrices.

B. If $\mathbf{A}, \mathbf{B}$ are square matrices having the same dimensions, then $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.

C. The determinant of an upper- (or lower-) triangular matrix $\mathbf{A} = (a_{ij})$ is the product of its diagonal elements. That is, $\det(\mathbf{A}) = a_{11}a_{22}\cdots a_{nn}$.

D. Suppose $\mathbf{A}$, $\mathbf{B}$ are square matrices having the same dimensions and, in addition, $\mathbf{B}$ has been obtained from $\mathbf{A}$ via one of the elementary row operations described in Section 1.5. If $\mathbf{B}$ was obtained from $\mathbf{A}$ via

- **row op 1** (multiplication of a row by a constant c), then $\det(\mathbf{B}) = c\det(\mathbf{A})$.
- **row op 2** (exchanging two rows), then $\det(\mathbf{B}) = -\det(\mathbf{A})$.
- **row op 3** (adding a multiple of one row to another), then $\det(\mathbf{B}) = \det(\mathbf{A})$.

E. If any of the rows or columns of $\mathbf{A}$ contain all zeros, then $|\mathbf{A}| = 0$.

F. The matrix $\mathbf{A}$ is nonsingular (i.e., $\mathbf{A}^{-1}$ exists) if and only if $|\mathbf{A}| \neq 0$. As a consequence, the matrix problem $\mathbf{Ax} = \mathbf{b}$ has a unique solution for all right-hand side vectors $\mathbf{b}$ if and only if $\det(\mathbf{A}) \neq 0$. When $\det(\mathbf{A}) = 0$, there is *not a single vector* $\mathbf{b}$ for which a unique solution $\mathbf{x}$ exists. As to which of the alternatives that leaves us—infinitely many solutions $\mathbf{x}$, or no solution $\mathbf{x}$ at all—that depends on the choice of $\mathbf{b}$ (i.e., whether $\mathbf{b}$ is in $\text{col}(\mathbf{A})$ or not).

1.7.4 Cramer's Rule

Cramer's rule provides a method for solving a system of linear algebraic equations for which the associated matrix problem $\mathbf{Ax} = \mathbf{b}$ has a coefficient matrix which is *nonsingular*. It is of no use if this criterion is not met and, considering the effectiveness of algorithms we have learned already for solving such a system (inversion of the matrix $\mathbf{A}$, and Gaussian elimination, specifically), it is not clear why we need yet another method. Nevertheless, it is a tool (some) people use, and should be recognized/understood by you when you run across it. We will describe the method, but not explain why it works, as this would require a better understanding of determinants than our time affords.

So, let us assume the n-by-n matrix $\mathbf{A}$ is nonsingular, that $\mathbf{b}$ is a known vector in $\mathbb{R}^n$, and that we wish to solve the equation $\mathbf{Ax} = \mathbf{b}$ for an unknown (unique) vector $\mathbf{x} \in \mathbb{R}^n$. Cramer's rule requires the construction of matrices $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$, where each $\mathbf{A}_j$, $1 \le j \le n$ is built from the original $\mathbf{A}$ and $\mathbf{b}$. These are constructed as follows: the j^{th} column of $\mathbf{A}$ is replaced by $\mathbf{b}$ to form $\mathbf{A}_j$.

Example 11: Construction of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ when $\mathbf{A}$ is 3-by-3

Suppose $\mathbf{A} = (a_{ij})$ is a 3-by-3 matrix, and $\mathbf{b} = (b_i)$, then

$$\mathbf{A}_1 = \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix}, \quad \text{and} \quad \mathbf{A}_3 = \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{pmatrix}.$$

■

Armed with these $\mathbf{A}_j$, $1 \le j \le n$, the solution vector $\mathbf{x} = (x_1, \dots, x_n)$ has its j^{th} component given by

$$x_j = \frac{|\mathbf{A}_j|}{|\mathbf{A}|}, \qquad j = 1, 2, \dots, n. \tag{1.10}$$

It should be clear from this formula why it is necessary that $\mathbf{A}$ be nonsingular.

Example 12:

Use Cramer's rule to solve the system of equations

$$\begin{aligned} x + 3y + z - w &= -9 \\ 2x + y - 3z + 2w &= 51 \\ x + 4y + 2w &= 31 \\ -x + y + z - 3w &= -43 \end{aligned}$$

Here, $\mathbf{A}$ and $\mathbf{b}$ are given by

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 1 & -1 \\ 2 & 1 & -3 & 2 \\ 1 & 4 & 0 & 2 \\ -1 & 1 & 1 & -3 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} -9 \\ 51 \\ 31 \\ -43 \end{pmatrix}, \qquad \text{so} \qquad |\mathbf{A}| = \begin{vmatrix} 1 & 3 & 1 & -1 \\ 2 & 1 & -3 & 2 \\ 1 & 4 & 0 & 2 \\ -1 & 1 & 1 & -3 \end{vmatrix} = -46.$$

Thus,

$$x = \frac{|\mathbf{A}_1|}{|\mathbf{A}|} = \frac{1}{|\mathbf{A}|}\begin{vmatrix} -9 & 3 & 1 & -1 \\ 51 & 1 & -3 & 2 \\ 31 & 4 & 0 & 2 \\ -43 & 1 & 1 & -3 \end{vmatrix} = \frac{-230}{-46} = 5,$$

$$y = \frac{|\mathbf{A}_2|}{|\mathbf{A}|} = \frac{1}{|\mathbf{A}|}\begin{vmatrix} 1 & -9 & 1 & -1 \\ 2 & 51 & -3 & 2 \\ 1 & 31 & 0 & 2 \\ -1 & -43 & 1 & -3 \end{vmatrix} = \frac{-46}{-46} = 1,$$

$$z = \frac{|\mathbf{A}_3|}{|\mathbf{A}|} = \frac{1}{|\mathbf{A}|}\begin{vmatrix} 1 & 3 & -9 & -1 \\ 2 & 1 & 51 & 2 \\ 1 & 4 & 31 & 2 \\ -1 & 1 & -43 & -3 \end{vmatrix} = \frac{276}{-46} = -6,$$

$$w = \frac{|\mathbf{A}_4|}{|\mathbf{A}|} = \frac{1}{|\mathbf{A}|}\begin{vmatrix} 1 & 3 & 1 & -9 \\ 2 & 1 & -3 & 51 \\ 1 & 4 & 0 & 31 \\ -1 & 1 & 1 & -43 \end{vmatrix} = \frac{-506}{-46} = 11,$$

yielding the solution $\mathbf{x} = (x, y, z, w) = (5, 1, -6, 11)$.

■

1.8 Linear Independence and Matrix Rank

We have defined the nullspace of an m-by-n matrix $\mathbf{A}$ as the set of vectors $\mathbf{v} \in \mathbb{R}^n$ satisfying the equation $\mathbf{A}\mathbf{v} = \mathbf{0}$. In light of the discussion in Section 1.3, the components of any $\mathbf{v} \in \text{null}(\mathbf{A})$ offer up a way to write $\mathbf{0}$ as a linear combination of the columns of $\mathbf{A}$:

$$\mathbf{0} = \mathbf{A}\mathbf{v} = \left[\begin{array}{c|c|c|c} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_n \end{array}\right]\left[\begin{array}{c} v_1 \\ \hline v_2 \\ \hline \vdots \\ \hline v_n \end{array}\right] = v_1\mathbf{A}_1 + v_2\mathbf{A}_2 + \cdots + v_n\mathbf{A}_n .$$

For some matrices $\mathbf{A}$, the nullspace consists of just one vector, the zero vector $\mathbf{0}$. We make a definition that helps us characterize this situation.

Definition 7: Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^m$. If the zero vector $\mathbf{0} \in \mathbb{R}^n$ can be written as a linear combination

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k \;=\; \mathbf{0}\,,$$

with at least one of the coefficients $c_1, \ldots, c_k$ nonzero, then the set S of vectors is said to be **linearly dependent**. If, however, the *only* linear combination of the vectors in S that yields $\mathbf{0}$ is the one with $c_1 = c_2 = \cdots = c_k = 0$, then set S is **linearly independent**.

Employing this terminology, when $\text{null}\,(\mathbf{A}) = \{\mathbf{0}\}$ the set of columns of $\mathbf{A}$ are linearly independent. Otherwise, this set is linearly dependent.

Suppose, as in Example 8, we set out to find the nullspace of $\mathbf{A}$ using Gaussian elimination. The result of elementary row operations is the row equivalence of augmented matrices

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{0} \end{array}\right] \sim \left[\begin{array}{c|c} \mathbf{R} & \mathbf{0} \end{array}\right],$$

where $\mathbf{R}$ is an echelon form for $\mathbf{A}$. We know that $\mathbf{v} \in \text{null}\,(\mathbf{A})$ if and only if $\mathbf{v} \in \text{null}\,(\mathbf{R})$. Let's look at several possible cases:

1. Case: $\mathbf{R}$ has no free columns.
   Several possible appearances of $\mathbf{R}$ are

$$\mathbf{R} \;=\; \begin{bmatrix} p & * & * & \cdots & * \\ 0 & p & * & \cdots & * \\ 0 & 0 & p & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p \end{bmatrix}, \tag{1.11}$$

and

$$\mathbf{R} \;=\; \left[\begin{array}{ccccc} p & * & * & \cdots & * \\ 0 & p & * & \cdots & * \\ 0 & 0 & p & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p \\ \hline 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{array}\right]. \tag{1.12}$$

Regardless of whether $\mathbf{R}$ has form (1.11) or form (1.12), the elements of $\mathbf{v}$ are uniquely determined—there is no other solution to $\mathbf{R}\mathbf{v} = \mathbf{0}$ but the one with each component

$v_j = 0$ for $j = 1, \ldots, n$. This means that null $(\mathbf{R})$ = null $(\mathbf{A}) = \{\mathbf{0}\}$ and, correspondingly, that the columns of **A** are linearly independent.

2. Case: **R** has free columns.
   A possible appearance of **R** is

$$\mathbf{R} \;=\; \left[\begin{array}{cccccccc} p & * & * & * & * & \cdots & * & * \\ 0 & 0 & 0 & p & * & \cdots & * & * \\ 0 & 0 & 0 & 0 & p & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & p & * \\ \hline 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{array}\right] \tag{1.13}$$

   The matrix **R** pictured here (as a 'for instance') has at least 3 free columns (the 2^{nd}, 3^{rd} and last ones), each providing a degree of freedom to the solution of $\mathbf{Av} = \mathbf{0}$. If the solution of $\mathbf{Av} = \mathbf{0}$ has even one degree of freedom (one free column in an echelon form of **A**), then the columns of **A** are linearly dependent.

   It should be evident that the set of *pivot columns* of **R** are linearly independent. That is, if we throw out the free columns to get a smaller matrix $\tilde{\mathbf{R}}$ of form (1.11) or (1.12), then the columns of $\tilde{\mathbf{R}}$ (and correspondingly, those of **A** from which these pivot columns originated) are linearly independent.

The number of linearly independent columns in **A** is a quantity that deserves a name.

Definition 8: The **rank** of an m-by-n matrix **A**, denoted by rank(**A**), is the *number* of pivots (equivalently, the number of pivot columns) in an echelon form **R** for **A**.
The number of free columns in **R** is called the **nullity** of **A**, denoted by nullity (**A**).

Note that, for an m-by-n matrix, rank(**A**) + nullity (**A**) = n.

Now, suppose some vector $\mathbf{b} \in \text{span}(S)$, where $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is some collection of vectors. That is,

$$\mathbf{b} \;=\; a_1\mathbf{u}_1 + \cdots + a_k\mathbf{u}_k \ ,$$

for some choice of coefficients $a_1, \ldots a_k$. If the vectors in S are linearly dependent, then there is a choice of coefficients $c_1, \ldots c_k$, not all of which are zero, such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_k + \cdots + c_k\mathbf{u}_k \;=\; \mathbf{0} \ .$$

Let us assume that $c_k \neq 0$. Solving this equation for $\mathbf{u}_k$, we get

$$\mathbf{u}_k \;=\; -\frac{1}{c_k}(c_1\mathbf{u}_1 + \cdots + c_{k-1}\mathbf{u}_{k-1})\;,$$

which we may then plug back into our equation for $\mathbf{b}$:

$$\begin{aligned}
\mathbf{b} &= a_1\mathbf{u}_1 + \cdots + a_{k-1}\mathbf{u}_{k-1} + a_k\mathbf{u}_k \\
&= a_1\mathbf{u}_1 + \cdots + a_{k-1}\mathbf{u}_{k-1} - \frac{a_k}{c_k}(c_1\mathbf{u}_1 + \cdots + c_{k-1}\mathbf{u}_{k-1}) \\
&= \left(a_1 - \frac{a_k c_1}{c_k}\right)\mathbf{u}_1 + \left(a_2 - \frac{a_k c_2}{c_k}\right)\mathbf{u}_2 + \cdots + \left(a_{k-1} - \frac{a_k c_{k-1}}{c_k}\right)\mathbf{u}_{k-1}\;,
\end{aligned}$$

which shows that, by taking $d_j = a_j - a_k c_j/c_k$ for $j = 1, \ldots, k-1$, this $\mathbf{b}$ which was already known to be a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_k$ may be rewritten as a linear combination $d_1\mathbf{u}_1 + \cdots + d_{k-1}\mathbf{u}_{k-1}$ of the reduced collection $\{\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}\}$. Of course, if this reduced set of vectors is linearly dependent, we may remove another vector—let us assume $\mathbf{u}_{k-1}$ would suit our purposes—to arrive at an even smaller set $\{\mathbf{u}_1, \ldots, \mathbf{u}_{k-2}\}$ which has the same span as the original set S, and continue in this fashion until we arrive at a subcollection of S which is linearly independent. We have demonstrated the truth of the following result.

Theorem 2: Suppose S is a collection of vectors in $\mathbb{R}^n$. Then some **subset** B of S (that is, every vector in B comes from S, but there *may* be vectors in S excluded from B) has the property that $\text{span}(B) = \text{span}(S)$ and B is linearly independent.

The collection **B** is called a **basis** (a term we will define more carefully in a later section) for span(S).

The previous theorem is an "existence" theorem, akin to the Existence/Uniqueness theorems of Sections 2.4, 2.8 and 7.1 in the Boyce and DiPrima text; it tells you that something exists without necessarily describing how to find it. But the text leading to the theorem suggests a way—namely, we may build a matrix whose columns are the vectors in S

$$\mathbf{A} \;=\; \left[\; \mathbf{u}_1 \;\middle|\; \mathbf{u}_2 \;\middle|\; \cdots \;\middle|\; \mathbf{u}_k \;\right].$$

We may then reduce $\mathbf{A}$ to echelon form $\mathbf{R}$ (another matrix whose dimensions are the same as $\mathbf{A}$—there is no need to augment $\mathbf{A}$ with an extra column for this task), and take B to be the set of columns of $\mathbf{A}$—it may be all of them—which correspond to *pivot* columns in $\mathbf{R}$. The number of elements in B will be rank($\mathbf{A}$).

There is an important relationship between the value of nullity ($\mathbf{A}$) and the number of solutions one finds when solving $\mathbf{Ax} = \mathbf{b}$. You may already have suspected this, given the similarity of results in Examples 7 and 8, both of which involved the same matrix $\mathbf{A}$. In Example 8, we found null ($\mathbf{A}$) to be the line of vectors passing through the origin in $\mathbb{R}^3$

$$t(-1, 2, 0), \qquad t \in \mathbb{R} .$$

In Example 7, we solved $\mathbf{Ax} = (3, 9)$, getting solution

$$(2, 0, 1) + t(-1, 2, 0), \qquad t \in \mathbb{R} ,$$

another line of vectors in $\mathbb{R}^3$, parallel to the first line, but offset from the origin by the vector $(2, 0, 1)$. One could describe this latter solution as being the sum of the nullspace of $\mathbf{A}$ and a *particular solution* of $\mathbf{Ax} = \mathbf{b}$.[4] Observe that, if $\mathbf{x}_p$ satisfies the equation $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x}_n \in$ null ($\mathbf{A}$), then for $\mathbf{v} = \mathbf{x}_p + \mathbf{x}_n$,

$$\mathbf{Av} \;=\; \mathbf{A}(\mathbf{x}_p + \mathbf{x}_n) \;=\; \mathbf{Ax}_p + \mathbf{Ax}_n \;=\; \mathbf{b} + \mathbf{0} \;=\; \mathbf{b} .$$

Thus, when $\mathbf{Ax} = \mathbf{b}$ has a solution, the number of solutions is at least as numerous as the number of vectors in null ($\mathbf{A}$). In fact, they are precisely as numerous, as stated in the next theorem.

Theorem 3: Suppose the m-by-n matrix $\mathbf{A}$ and vector $\mathbf{b} \in \mathbb{R}^m$ (both fixed) are such that the matrix equation $\mathbf{Ax} = \mathbf{b}$ is **consistent** (i.e., the equation *has* a solution). Then the solutions are in one-to-one correspondence with the elements in null ($\mathbf{A}$). Said another way, if null ($\mathbf{A}$) has just the zero vector, then $\mathbf{Ax} = \mathbf{b}$ has just one solution. If null ($\mathbf{A}$) is a line (plane, etc.) of vectors, then so is the set of solutions to $\mathbf{Ax} = \mathbf{b}$.

If you review Examples 7 and 8 you will see that the appearance of the free variable t is due to a free column in the echelon form we got for $\mathbf{A}$. The rank of $\mathbf{A}$—its number of linearly independent columns it has, is 2, not 3.

We finish this section with an important theorem. Some of these results have been stated (in some form or other) elsewhere, but the theorem provides a nice overview of facts about *square* matrices.

[4] This will be a theme in the course: whenever we have a nonhomogenous linear problem that is *consistent* (i.e., has at least one solution), the collection of all solutions may be characterized as the sum of the nullspace and any particular solution.

Theorem 4: Suppose $\mathbf{A}$ is an n-by-n matrix. The following are equivalent (that is, if you know one of them is true, then you know all of them are).

(i) The matrix $\mathbf{A}$ is nonsingular.

(ii) The matrix equation $\mathbf{Ax} = \mathbf{b}$ has a unique solution for each possible n-vector $\mathbf{b}$.

(iii) The determinant $\det(\mathbf{A}) \neq 0$.

(iv) The nullspace $\text{null}\,(\mathbf{A}) = \{\mathbf{0}\}$.

(v) The columns of $\mathbf{A}$ are linearly independent

(vi) $\text{rank}(\mathbf{A}) = n$.

(vii) $\text{nullity}\,(\mathbf{A}) = 0$.

1.9 Eigenvalues and Eigenvectors

The product $\mathbf{Ax}$ of a n-by-n real matrix (i.e., having real number entries) $\mathbf{A}$ and an n-vector $\mathbf{x}$ is itself an n-vector. Of particular interest in many settings (of which differential equations is one) is the following question:

> For a given matrix $\mathbf{A}$, what are the vectors $\mathbf{x}$ for which the product $\mathbf{Ax}$ is a scalar multiple of $\mathbf{x}$? That is, what vectors $\mathbf{x}$ satisfy the equation
>
> $$\mathbf{Ax} = \lambda\mathbf{x}$$
>
> for some scalar λ?

It should immediately be clear that, no matter what $\mathbf{A}$ and λ are, the vector $\mathbf{x} = \mathbf{0}$ (that is, the vector whose elements are all zero) satisfies this equation. With such a trivial answer, we might ask the question again in another way:

> For a given matrix $\mathbf{A}$, what are the *nonzero* vectors $\mathbf{x}$ that satisfy the equation
>
> $$\mathbf{Ax} = \lambda\mathbf{x}$$
>
> for some scalar λ?

To answer this question, we first perform some algebraic manipulations upon the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. We note first that, if $\mathbf{I} = \mathbf{I_n}$ (the $n \times n$ multiplicative identity matrix), then we can write

$$\begin{aligned} \mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad &\Leftrightarrow \quad \mathbf{A}\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \quad \mathbf{A}\mathbf{x} - \lambda\mathbf{I}\mathbf{x} = \mathbf{0} \\ &\Leftrightarrow \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}. \end{aligned}$$

Remember that we are looking for nonzero $\mathbf{x}$ that satisfy this last equation. But $\mathbf{A} - \lambda\mathbf{I}$ is an $n \times n$ matrix and, should its determinant be nonzero, this last equation will have exactly one solution, namely $\mathbf{x} = \mathbf{0}$. Thus our question above has the following answer:

> The equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ has nonzero solutions for the vector $\mathbf{x}$ if and only if the matrix $\mathbf{A} - \lambda\mathbf{I}$ has zero determinant.

As we will see in the examples below, for a given matrix $\mathbf{A}$ there are only a few special values of the scalar λ for which $\mathbf{A} - \lambda\mathbf{I}$ will have zero determinant, and these special values are called the **eigenvalues** of the matrix $\mathbf{A}$. Based upon the answer to our question, it seems we must first be able to find the eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_n$ of $\mathbf{A}$ and then see about solving the individual equations $\mathbf{A}\mathbf{x} = \lambda_i\mathbf{x}$ for each $i = 1, \ldots, n$.

Example 13:

Find the eigenvalues of the matrix $\mathbf{A} = \begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix}$.

The eigenvalues are those λ for which $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Now

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= \det\left(\begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= \begin{vmatrix} 2-\lambda & 2 \\ 5 & -1-\lambda \end{vmatrix} = (2-\lambda)(-1-\lambda) - 10 = \lambda^2 - \lambda - 12. \end{aligned}$$

The eigenvalues of $\mathbf{A}$ are the solutions of the quadratic equation $\lambda^2 - \lambda - 12 = 0$, namely $\lambda_1 = -3$ and $\lambda_2 = 4$.

■

As we have discussed, if $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$ then the equation $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{b}$ has either no solutions or infinitely many. When we take $\mathbf{b} = \mathbf{0}$ however, it is clear by the existence of the solution $\mathbf{x} = \mathbf{0}$ that there are infinitely many solutions (i.e., we may rule out the "no solution" case). If we continue using the matrix $\mathbf{A}$ from the example above, we can expect nonzero solutions $\mathbf{x}$ (infinitely many of them, in fact) of the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ precisely when $\lambda = -3$ or $\lambda = 4$. Let us procede to characterize such solutions.

First, we work with $\lambda = -3$. The equation $\mathbf{Ax} = \lambda\mathbf{x}$ becomes $\mathbf{Ax} = -3\mathbf{x}$. Writing

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and using the matrix $\mathbf{A}$ from above, we have

$$\mathbf{Ax} = \begin{bmatrix} 2 & 2 \\ 5 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 5x_1 - x_2 \end{bmatrix},$$

while

$$-3\mathbf{x} = \begin{bmatrix} -3x_1 \\ -3x_2 \end{bmatrix}.$$

Setting these equal, we get

$$\begin{aligned} \begin{bmatrix} 2x_1 + 2x_2 \\ 5x_1 - x_2 \end{bmatrix} = \begin{bmatrix} -3x_1 \\ -3x_2 \end{bmatrix} \quad &\Rightarrow \quad 2x_1 + 2x_2 = -3x_1 \qquad \text{and} \qquad 5x_1 - x_2 = -3x_2 \\ &\Rightarrow \quad 5x_1 = -2x_2 \\ &\Rightarrow \quad x_1 = -\frac{2}{5}x_2. \end{aligned}$$

This means that, while there are infinitely many nonzero solutions (solution vectors) of the equation $\mathbf{Ax} = -3\mathbf{x}$, they all satisfy the condition that the first entry x_1 is $-2/5$ times the second entry x_2. Thus all solutions of this equation can be characterized by

$$\begin{bmatrix} 2t \\ -5t \end{bmatrix} = t\begin{bmatrix} 2 \\ -5 \end{bmatrix},$$

where t is any real number. The nonzero vectors $\mathbf{x}$ that satisfy $\mathbf{Ax} = -3\mathbf{x}$ are called **eigenvectors** associated with the eigenvalue $\lambda = -3$. One such eigenvector is

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$$

and all other eigenvectors corresponding to the eigenvalue (-3) are simply scalar multiples of $\mathbf{u}_1$ — that is, $\mathbf{u}_1$ spans this set of eigenvectors.

Similarly, we can find eigenvectors associated with the eigenvalue $\lambda = 4$ by solving $\mathbf{Ax} = 4\mathbf{x}$:

$$\begin{aligned} \begin{bmatrix} 2x_1 + 2x_2 \\ 5x_1 - x_2 \end{bmatrix} = \begin{bmatrix} 4x_1 \\ 4x_2 \end{bmatrix} \quad &\Rightarrow \quad 2x_1 + 2x_2 = 4x_1 \qquad \text{and} \qquad 5x_1 - x_2 = 4x_2 \\ &\Rightarrow \quad x_1 = x_2. \end{aligned}$$

Hence the set of eigenvectors associated with $\lambda = 4$ is spanned by

$$\mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Example 14:

Find the eigenvalues and associated eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 7 & 0 & -3 \\ -9 & -2 & 3 \\ 18 & 0 & -8 \end{bmatrix}.$$

We first find the eigenvalues, doing so by getting an expression for $\det(\mathbf{A} - \lambda\mathbf{I})$, setting it equal to zero and solving:

$$\begin{aligned} \begin{vmatrix} 7-\lambda & 0 & -3 \\ -9 & -2-\lambda & 3 \\ 18 & 0 & -8-\lambda \end{vmatrix} &= (-2-\lambda)(-1)^4 \begin{vmatrix} 7-\lambda & -3 \\ 18 & -8-\lambda \end{vmatrix} \\ &= -(2+\lambda)[(7-\lambda)(-8-\lambda)+54] \\ &= -(\lambda+2)(\lambda^2+\lambda-2) \;=\; -(\lambda+2)^2(\lambda-1)\,. \end{aligned}$$

Thus $\mathbf{A}$ has two distinct eigenvalues, $\lambda_1 = -2$ (its **algebraic multiplicity**, as a *zero* of $\det(\mathbf{A} - \lambda\mathbf{I})$ is 2), and $\lambda_3 = 1$ (algebraic multiplicity 1).

To find eigenvectors associated with $\lambda_3 = 1$, we solve the matrix equation $(\mathbf{A}-\mathbf{I})\mathbf{v} = \mathbf{0}$ (that is, we find the nullspace of $(\mathbf{A} - \mathbf{I})$). Our augmented matrix appears on the left, and an equivalent echelon form on the right:

$$\left[\begin{array}{ccc|c} 6 & 0 & -3 & 0 \\ -9 & -3 & 3 & 0 \\ 18 & 0 & -9 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 2 & 0 & -1 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Since the algebraic multiplicity of λ_3 is 1, the final bullet point on the previous page indicates we should expect precisely one free column in the echelon form and, indeed, the 3rd column is the free one. Writing $x_3 = 2t$, we have $x_1 = t$ and $x_2 = -t$, giving that

$$\text{null}\,(\mathbf{A} - \mathbf{I}) \;=\; \{t(1,-1,2) \,|\, t \in \mathbb{R}\} \;=\; \text{span}(\{(1,-1,2)\})\,.$$

That is, the eigenvectors associated with $\lambda_3 = 1$ form a line in $\mathbb{R}^3$ characterized by $(1,-1,2)$.

Now, to find eigenvectors associated with $\lambda_1 = -2$ we solve $(\mathbf{A}+2\mathbf{I})\mathbf{v} = \mathbf{0}$. We know going in that λ_1 has algebraic multiplicity 2, so we should arrive at an echelon form with either 1 or 2 free columns. We find that the augmented matrix

$$\left[\begin{array}{ccc|c} 9 & 0 & -3 & 0 \\ -9 & 0 & 3 & 0 \\ 18 & 0 & -6 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 3 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Columns 2 and 3 are free, and we set $x_2 = s$, $x_3 = 3t$. This means $x_1 = t$, and hence

$$\text{null}\,(\mathbf{A} + 2\mathbf{I}) \;=\; \{s(0,1,0) + t(1,0,3) \,|\, s,t \in \mathbb{R}\} \;=\; \text{span}(\{(0,1,0),(1,0,3)\})\,.$$

So, the eigenvectors associated with $\lambda_1 = -2$ form a plane in $\mathbb{R}^3$, with each of these eigenvectors obtainable as a linear combination of $(0, 1, 0)$ and $(1, 0, 3)$.

It should not be surprising that commands are available to us in OCTAVE for finding the **eigenpairs** (collective way to refer to eigenvalue-eigenvector pairs) of a square matrix. The relevant OCTAVE code looks like this:

```
octave> A = [7 0 -3; -9 -2 3; 18 0 -8]
A =
     7     0    -3
    -9    -2     3
    18     0    -8

octave> [V, lam] = eig(A)
V =
    0.00000    0.40825    0.31623
    1.00000   -0.40825    0.00000
    0.00000    0.81650    0.94868

lam =
   -2    0    0
    0    1    0
    0    0   -2
```

Compare the results with our analysis above.

■

Example 15:

Find the eigenvalues and associated eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}.$$

First, we have

$$\det(\mathbf{A} - \lambda \mathbf{I}) \;=\; \begin{vmatrix} -1-\lambda & 2 \\ 0 & -1-\lambda \end{vmatrix} \;=\; (\lambda+1)^2 \,,$$

showing $\lambda = -1$ is an eigenvalue (the only one) with algebraic multiplicity 2. Reducing an augmented matrix for $(\mathbf{A} - (-1)\mathbf{I})$, we should have either one or two free columns. In fact, the augmented matrix is

$$\left[\begin{array}{cc|c} 0 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right],$$

and does not need to be reduced, as it is already an echelon form. Only its first column is free, so we set $x_1 = t$. This augmented matrix also tells us that $x_2 = 0$, so

$$\text{null}\,(\mathbf{A} + \mathbf{I}) \;=\; \{t(1,0) \,|\, t \in \mathbb{R}\} \;=\; \text{span}(\{(1,0)\}) \,.$$

Note that, though the eigenvalue has algebraic multiplicity 2, the set of eigenvectors consists of just a line in $\mathbb{R}^2$ (a one-dimensional object), the scalar multiples of $(1, 0)$.

Compare this work with software output:

```
octave> [V, lam] = eig([-1 2; 0 -1])
V =
    1.00000   -1.00000
    0.00000    0.00000

lam =
   -1    0
    0   -1
```

■

What is new in this last example is that we have an eigenvalue whose **geometric multiplicity**, the term used for the corresponding number of linearly independent eigenvectors (also known as nullity $(\mathbf{A} - \lambda\mathbf{I})$), is smaller than its algebraic multiplicity.

Example 16:

Find the eigenvalues and associated eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$

We compute

$$\det(\mathbf{A} - \lambda\mathbf{I}) \;=\; \begin{vmatrix} 2-\lambda & -1 \\ 1 & 2-\lambda \end{vmatrix} \;=\; (\lambda-2)^2 + 1 \;=\; \lambda^2 - 4\lambda + 5\,.$$

The roots of this polynomial (found using the quadratic formula) are $\lambda_1 = 2 + i$ and $\lambda_2 = 2 - i$; that is, the eigenvalues are not real numbers. This is a common occurrence, and we can press on to find the eigenvectors just as we have in the past with real eigenvalues. To find eigenvectors associated with $\lambda_1 = 2 + i$, we look for $\mathbf{x}$ satisfying

$$(\mathbf{A} - (2+i)\mathbf{I})\mathbf{x} = \mathbf{0} \quad\Rightarrow\quad \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad\Rightarrow\quad \begin{bmatrix} -ix_1 - x_2 \\ x_1 - ix_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad\Rightarrow\quad x_1 = ix_2\,.$$

Thus all eigenvectors associated with $\lambda_1 = 2 + i$ are scalar multiples of $\mathbf{u}_1 = (i, 1)$. Proceeding with $\lambda_2 = 2 - i$, we have

$$(\mathbf{A} - (2-i)\mathbf{I})\mathbf{x} = \mathbf{0} \quad\Rightarrow\quad \begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad\Rightarrow\quad \begin{bmatrix} ix_1 - x_2 \\ x_1 + ix_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad\Rightarrow\quad x_1 = -ix_2\,,$$

which shows all eigenvectors associated with $\lambda_2 = 2 - i$ to be scalar multiples of $\mathbf{u}_2 = (-i, 1)$.

Notice that $\mathbf{u}_2$, the eigenvector associated with the eigenvalue $\lambda_2 = 2 - i$ in the last example, is the complex conjugate of $\mathbf{u}_1$, the eigenvector associated with the eigenvalue $\lambda_1 = 2 + i$. It is indeed a fact that, if an m-by-n real matrix $\mathbf{A}$ has a nonreal eigenvalue $\lambda_1 = \lambda + i\mu$ with corresponding eigenvector ξ_1, then it also has eigenvalue $\lambda_2 = \lambda - i\mu$ with corresponding eigenvector $\xi_2 = \overline{\xi_1}$.

The relevant software commands that parallel our work look like this:

```
octave> [V, lam] = eig([2 -1; 1 2])
V =
   0.70711 + 0.00000i   0.70711 - 0.00000i
   0.00000 - 0.70711i   0.00000 + 0.70711i

lam =
   2 + 1i   0 + 0i
   0 + 0i   2 - 1i
```

■

To sum up, our search for eigenpairs of $\mathbf{A}$ consists of

1. finding the eigenvalues, all roots of the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$, then

2. for each eigenvalue λ, finding a linearly independent set of corresponding eigenvectors that spans the nullspace of $\mathbf{A} - \lambda\mathbf{I}$. For a given λ, the number of vectors in this minimal spanning set is equal to nullity $(\mathbf{A} - \lambda\mathbf{I})$.

Some remarks:

- If n is the number of rows/columns in $\mathbf{A}$, then the quantity $\det(\mathbf{A} - \lambda\mathbf{I})$ is (always) an n^{th}-degree polynomial. Hence it has, counting multiplicities, exactly n roots which are the eigenvalues of $\mathbf{A}$.

- If $\mathbf{A}$ is upper or lower triangular, its eigenvalues are precisely the elements found on its main diagonal.

- Even if the problems we consider have corresponding matrices $\mathbf{A}$ with real-number entries, the eigenvalues of $\mathbf{A}$ may be non-real (complex). However, such eigenvalues always come in conjugate pairs—if $(a + bi)$ is an eigenvalue of $\mathbf{A}$, then so is $(a - bi)$.

- Once we know the eigenvalues, the search for eigenvectors is essentially the same as Example 8. For each eigenvalue λ, we find the nullspace of a certain matrix, namely $(\mathbf{A} - \lambda\mathbf{I})$. In each instance, when you reduce $(\mathbf{A} - \lambda\mathbf{I})$ to echelon form, there will be at least one free column, and there can be no more free columns than the multiplicity of λ as a zero of $\det(\mathbf{A} - \lambda\mathbf{I})$. Thus, the **geometric multiplicity** of λ, also known as nullity $(\mathbf{A} - \lambda\mathbf{I})$, must lie between 1 and the **algebraic multiplicity** of λ.

Though these examples, and the observations taken from them, leave open a number of possibilities involving nonreal eigenvalues and geometric multiplicities that are strictly less than algebraic ones, there is one special type of square matrix we mention here for which the outcome is quite definite. We give the result as a theorem.

Theorem 5 (Spectral Theorem for Symmetric Matrices): Suppose $\mathbf{A}$ is a *symmetric* matrix (i.e., $\mathbf{A}^T = \mathbf{A}$) with real-number entries. Then

- each eigenvalue of $\mathbf{A}$ is a *real* number, and
- each eigenvalue has geometric multiplicity equal to its algebraic multiplicity.

We describe this latter situation by saying that $\mathbf{A}$ has a **full set of eigenvectors**.

In Section 1.4, we investigated the underlying geometry associated with matrix multiplication. We saw that certain kinds of 2-by-2 matrices transformed the plane $\mathbb{R}^2$ by rotating it about the origin; others produced reflections across a line. Of particular interest here is Case 3 from that section, where the matrices involved caused rescalings that were (possibly) different along two perpendicular axes. Now, using our knowledge of eigenpairs, we can discuss the general case where these axes may not be perpendicular.

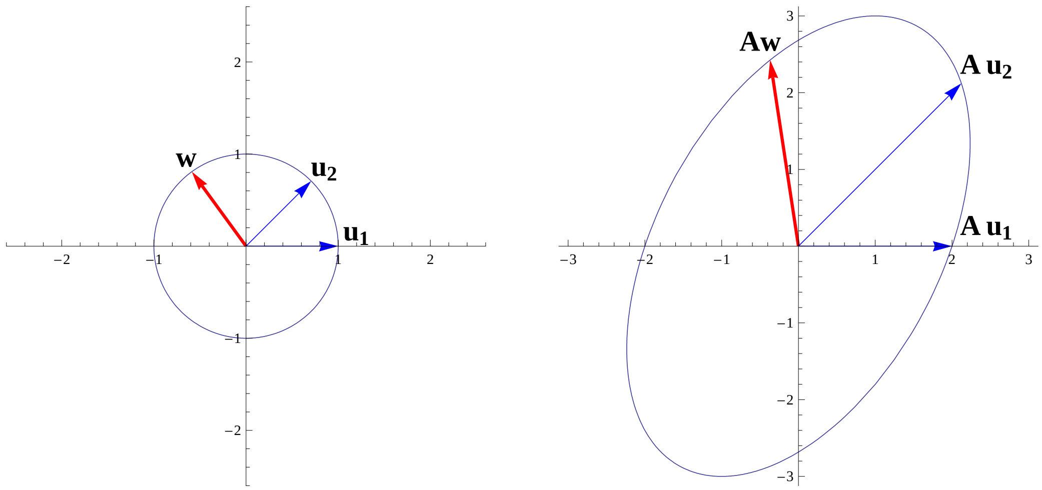
Recall that an eigenpair $(\lambda, \mathbf{v})$ of $\mathbf{A}$ satisfies the relationship $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. This says that the output $\mathbf{A}\mathbf{v}$ (from the function $(\mathbf{x} \mapsto \mathbf{A}\mathbf{x})$) corresponding to input $\mathbf{v}$ is a vector that lies in the "same direction" as $\mathbf{v}$ itself and, in fact, is a predictable rescaling of $\mathbf{v}$ (i.e., it is λ times $\mathbf{v}$).

Example 17:

Suppose $\mathbf{A}$ is a 2-by-2 matrix that has eigenvalues $\lambda_1 = 2$, $\lambda_2 = 3$ with corresponding eigenvectors $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (1/\sqrt{2}, 1/\sqrt{2})$. The matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

is just such a matrix, and the associated function $(\mathbf{x} \mapsto \mathbf{A}\mathbf{x})$ rescales vectors in the direction of $(1, 0)$ by a factor of 2 relative to the origin, while vectors in the direction of $(1, 1)$ will be similarly rescaled but by a factor of 3. (See the figure below.) The affect of multiplication by $\mathbf{A}$ on all other vectors in the plane is more complicated to describe, but will nevertheless conform to these two facts. The figure shows (on the left) the unit circle and eigenvectors $\mathbf{u}_1$, $\mathbf{u}_2$ of $\mathbf{A}$. On the right is displayed how this circle is transformed via multiplication by $\mathbf{A}$. Notice that $\mathbf{A}\mathbf{u}_1$ faces the same direction as $\mathbf{u}_1$, but is twice as long; the same is true of $\mathbf{A}\mathbf{u}_2$ in relation to $\mathbf{u}_2$, except it is 3 times as long. The figure displays one more unit vector $\mathbf{w}$ along with its image $\mathbf{Aw}$ under matrix multiplication by $\mathbf{A}$.



■

We leave it to the exercises to discover what may be said about the eigenvalues of a 2-by-2 matrix $\mathbf{A}$ when the associated function ($\mathbf{x} \mapsto \mathbf{Ax}$) rotates the plane about the origin. We also investigate similar ideas when $\mathbf{A}$ is a 3-by-3 matrix.

Exercises

1.1 Give a particularly simple command in OCTAVE (one which does not require you to type in every entry) which will produce the matrix

a) $\begin{bmatrix} 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

b) $\begin{bmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 \\ 0 & 7 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 3 & -2 \end{bmatrix}$

1.2 Suppose **A** is a 5-by-3 matrix.

a) If **B** is another matrix and the matrix product **AB** makes sense, what must be true about the dimensions of **B**?

b) If the matrix product **BA** makes sense, what must be true about the dimensions of **B**?

1.3 Suppose **A**, **B** are matrices for which the products **AB** and **BA** are both possible (both defined).

a) For there to be any chance that **AB** = **BA**, what must be true about the dimensions of **A**? Explain.

b) When we say that **AB** ≠ **BA** in general, we do not mean that it never happens, but rather that you cannot count on their equality. Write a function in OCTAVE which, when called, generates two random 3-by-3 matrices **A** and **B**, finds the products **AB** and **BA**, and checks whether they are equal. Run this code 20 times, and record how many of those times it happens that **AB** = **BA**. Hand in a printout of your function.

c) Of course, when both **A** and **B** are square matrices with one of them equal to the identity matrix, it will be the case that **AB** = **BA**. What other instances can you think of in which **AB** = **BA** is guaranteed to hold?

1.4 Verify, via direct calculation, Theorem 1. That is, use the knowledge that $\mathbf{A}$, $\mathbf{B}$ are n-by-n nonsingular matrices to show that $\mathbf{AB}$ is nonsingular as well, having inverse $\mathbf{B}^{-1}\mathbf{A}^{-1}$.

1.5 We have learned several properties of the operations of inversion and transposition of a matrix. The table below summarizes these, with counterparts appearing on the same row.

	matrix transposition	matrix inversion
i.	$(\mathbf{A}^T)^T = \mathbf{A}$	$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
ii.	$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$	$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
iii.	$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$	

Show that property iii. has no counterpart in the "matrix inversion" column. That is, in general it is not the case that $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} + \mathbf{B}^{-1}$.

1.6 The previous two problems asked you to "prove" or "show" (basically synonymous words in mathematics) something. Yet there is something fundamentally different about what is required in the two problems. In one problem, all you need to do is come up with a specific instance—matrices $\mathbf{A}$, $\mathbf{B}$ whose entries are concrete numbers—to prove the assertion. In the other problem, if you resort to specific matrices then all you succeed in doing is showing the assertion is true in one particular instance. In which problem is it that you cannot get specific about the entries in $\mathbf{A}$, $\mathbf{B}$? What is it in the wording of these problems that helps you determine the level of generality required?

1.7

a) Explain why it is necessary that a symmetric matrix be square.

b) Suppose $\mathbf{A} = (a_{ij})$ is an n-by-n matrix. Prove that $\mathbf{A}$ is symmetric if and only if $a_{ij} = a_{ji}$ for each $1 \leq i, j \leq n$.

1.8 Suppose there is a town which perenially follows these rules:

- The number of households always stays fixed at 10000.
- Every year 30 percent of households currently subscribing to the local newspaper cancel their subscriptions.
- Every year 20 percent of households not receiving the local newspaper subscribe to it.

a) Suppose one year, there are 8000 households taking the paper. According to the data above, these numbers will change the next year. The total of subscribers will be

$$(0.7)(8000) + (0.2)(2000) = 6000 ,$$

and the total of nonsubscribers will be

$$(0.3)(8000) + (0.8)(2000) = 4000 .$$

If we create a 2-vector whose first component is the number of subscribers and whose 2nd component is the number of nonsubscribers, then the initial vector is (8000, 2000), and the vector one year later is

$$\begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \begin{bmatrix} 8000 \\ 2000 \end{bmatrix} = \begin{bmatrix} 6000 \\ 4000 \end{bmatrix}.$$

What is the long-term outlook for newspaper subscription numbers?

b) Does your answer above change if the initial subscription numbers are changed to 9000 subscribing households? Explain.

1.9 In OCTAVE, generate 50 random 4-by-4 matrices. Determine how many of these matrices are singular. (You may find the command `det()` helpful. It's a simple command to use, and like most commands in OCTAVE, you can find out about its use by typing `help det`. You may also wish to surround the work you do on one matrix with the commands `for i = 1:50` and `end`.) Based upon your counts, how prevalent among all 4-by-4 matrices would you say that singular matrices are? What if you conduct the same experiment on 5-by-5 matrices? 10-by-10? (Along with your answers to the questions, hand in the code you used to conduct one of these experiments.)

1.10 Consider a matrix $\mathbf{A}$ that has been blocked in the following manner:

$$\mathbf{A} = \left[\begin{array}{c|c|c} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \end{array}\right],$$

where $\mathbf{A}_{11}$ is 2-by-3, $\mathbf{A}_{23}$ is 4-by-2, and the original matrix $\mathbf{A}$ has 7 columns.

a) How many rows does $\mathbf{A}$ have?

b) Determine the dimensions of the submatrices $\mathbf{A}_{12}$, $\mathbf{A}_{13}$, $\mathbf{A}_{21}$, and $\mathbf{A}_{22}$.

c) Give at least three different ways to partition a matrix $\mathbf{B}$ that has 5 columns so that a block-multiplication of the matrix product $\mathbf{AB}$ makes sense. For each of your answers, specify the block structure of $\mathbf{B}$ using $\mathbf{B}_{ij}$ notation just as we originally gave the block structure of $\mathbf{A}$, and indicate the dimensions of each block.

d) For each of your answers to part (c), write out the corresponding block structure of the product $\mathbf{AB}$, indicating how the individual blocks are computed from the blocks of $\mathbf{A}$ and $\mathbf{B}$ (as was done in the notes immediately preceding Example 3).

1.11

a) Suppose $\mathbf{A}$ is a 4-by-n matrix. Find a matrix $\mathbf{P}$ (you should determine appropriate dimensions for $\mathbf{P}$, as well as specify its entries) so that $\mathbf{PA}$ has the same entries as $\mathbf{A}$ but the 1^{st}, 2^{nd}, 3^{rd} and 4^{th} rows of $\mathbf{PA}$ are the 2^{nd}, 4^{th}, 3^{rd} and 1^{st} rows of $\mathbf{A}$ respectively. Such a matrix $\mathbf{P}$ is called a **permutation matrix**.

b) Suppose $\mathbf{A}$ is an m-by-4 matrix. Find a matrix $\mathbf{P}$ so that $\mathbf{AP}$ has the same entries as $\mathbf{A}$ but the 1^{st}, 2^{nd}, 3^{rd} and 4^{th} *columns* of $\mathbf{AP}$ are the 2^{nd}, 4^{th}, 3^{rd} and 1^{st} columns of $\mathbf{A}$ respectively.

c) Suppose $\mathbf{A}$ is an m-by-3 matrix. Find a matrix $\mathbf{B}$ so that $\mathbf{AB}$ again has 3 columns, the first of which is the sum of all three columns of $\mathbf{A}$, the 2^{nd} is the difference of the 1^{st} and 3^{rd} columns of $\mathbf{A}$ (column 1 - column 3), and the 3^{rd} column is 3 times the 1^{st} column of $\mathbf{A}$.

1.12 We have given two alternate ways of achieving translations of the plane by a vector $\mathbf{w} = (a, b)$:

(i) $(\mathbf{v} \mapsto \mathbf{v} + \mathbf{w})$, and

(ii) $(\mathbf{v} \mapsto \tilde{\mathbf{v}} \mapsto \mathbf{A}\tilde{\mathbf{v}})$, where $\mathbf{A} = \left[\begin{array}{cc|c} 1 & 0 & a \\ 0 & 1 & b \\ \hline 0 & 0 & 1 \end{array}\right]$.

If $\mathbf{v} \in \mathbb{R}^2$ has homogeneous coordinates $\tilde{\mathbf{v}} \in \mathbb{R}^3$, use the indicated blocking on $\mathbf{A}$ in (ii) and what you know about block multiplication to show that the upper block of $\mathbf{A}\tilde{\mathbf{v}}$ gives the same result as the mapping in (i).

1.13

a) Multiply out the matrices on the left-hand side of (1.5) to show that, indeed, they are equal to the matrix on the right-hand side for $\alpha = 2\theta$.

b) Show that a matrix in the form (1.6) may be expressed in an alternate form

$$\begin{bmatrix} a^2 - b^2 & 2ab \\ 2ab & b^2 - a^2 \end{bmatrix},$$

for some choice of constants a, b such that $a^2 + b^2 = 1$.

1.14 Determine which of the following is an echelon form. For those that are, indicate what are the pivot columns and the pivots.

a) $\begin{bmatrix} 0 & 2 & 1 & 6 & 5 & -1 \\ 0 & 0 & 0 & 3 & 2 & 7 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$

b) $\begin{bmatrix} 2 & 7 & 3 & -1 & -5 \\ -1 & 1 & 1 & 4 & 2 \\ 0 & 2 & 3 & 5 & 1 \\ 0 & 0 & -1 & -1 & 7 \\ 0 & 0 & 0 & -3 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}$

d) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 4 & 2 & 8 \end{bmatrix}$

f) $\begin{bmatrix} 1 & 0 & -2 & 3 \\ 0 & 2 & 3 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & -1 & 0 & 5 \end{bmatrix}$

1.15 Use backward substitution to solve the following systems of equations.

a) $$\begin{aligned} x_1 - 3x_2 &= 2 \\ 2x_2 &= 6 \end{aligned}$$

b) $$\begin{aligned} x_1 + x_2 + x_3 &= 8 \\ 2x_2 + x_3 &= 5 \\ 3x_3 &= 9 \end{aligned}$$

c) $$\begin{aligned} x_1 + 2x_2 + 2x_3 + x_4 &= 5 \\ 3x_2 + x_3 - 2x_4 &= 1 \\ -x_3 + 2x_4 &= -1 \\ 4x_4 &= 4 \end{aligned}$$

1.16 Write out the system of equations that corresponds to each of the following augmented matrices.

a) $\left[\begin{array}{cc|c} 3 & 2 & 8 \\ 1 & 5 & 7 \end{array}\right]$

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

c) $\left[\begin{array}{ccc|c} 2 & 1 & 4 & -1 \\ 4 & -2 & 3 & 4 \\ 5 & 2 & 6 & -1 \end{array}\right]$

1.17 Suppose we wish to perform elementary operation 3 on some matrix **A**. That is, we wish to produce a matrix **B** which has the same dimensions as **A**, and in most respects is identical to **A** except that

$$(\text{row } i \text{ of } \mathbf{B}) = (\text{row } i \text{ of } \mathbf{A}) + \beta(\text{row } j \text{ of } \mathbf{A}) .$$

If **A** is m-by-n, then $\mathbf{B} = \mathbf{E}_{ij}\mathbf{A}$, where $\mathbf{E}_{ij}$ is the m-by-m **elementary matrix**

$$\mathbf{E}_{ij} := \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & \beta & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

which looks just like the m-square identity matrix $\mathbf{I}_m$ except for the entry β appearing in its i^{th} row, j^{th} column. Octave code for producing such a matrix might look like

```
function elementaryMatrix = emat(m, i, j, val)
  elementaryMatrix = eye(m,m);
  elementaryMatrix(i,j) = val;
end
```

a) Create a text file containing the code above and called `emat.m`. Place this file in your working directory, the one you are in when running Octave (or Matlab).

b) Another list of commands to put into a file, call it `simpleGE.m`, is the following:

```
B = A;
numRows = size(B)(1);
numCols = size(B)(2);
currRow = 1;
currCol = 1;
while ((currRow < numRows) && (currCol < numCols))
  while ((abs(B(currRow, currCol)) < 10^(-10)) && (currCol < numCols))
    B(currRow, currCol) = 0;
    currCol = currCol + 1;
```

```
    end

    if (currCol < numCols)
      pivot = B(currRow, currCol);
      for ii = (currRow + 1):numRows
        B = emat(numRows, ii, currRow, -B(ii,currCol)/pivot) * B;
        % Remove the final semicolon in the previous line
        % if you would like to see the progression of matrices
        % from the original one (A) to the final one in echelon form.
      end
    end
    currRow = currRow + 1;
    currCol = currCol + 1;
  end
  B
```

One would run `simpleGE` after first storing the appropriate coefficients of the linear system in an augmented matrix `A`.

Save the commands above under the filename `simpleGE.m` in your working directory. Then test it out on the matrix (assumed to be already augmented)

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 2 & 1 & 5 \\ 3 & 2 & 6 & 3 & 1 \\ 6 & 2 & 12 & 4 & 3 \end{bmatrix}.$$

If you have arranged and entered everything correctly, the result will be the matrix in echelon form

$$\begin{bmatrix} 1 & 3 & 2 & 1 & 5 \\ 0 & -7 & 0 & 0 & -14 \\ 0 & 0 & 0 & -2 & 5 \end{bmatrix}.$$

1.18 Using `simpleGE` (see Exercise 1.17) as appropriate, find all solutions to the following linear systems of equations:

a)
$$\begin{aligned} 2x - z &= -4 \\ -4x - 2y + z &= 11 \\ 2x + 2y + 5z &= 3 \end{aligned}$$

b)
$$\begin{aligned} x_1 + 3x_2 + 2x_3 + x_4 &= 5 \\ 3x_1 + 2x_2 + 6x_3 + 3x_4 &= 1 \\ 6x_1 + 2x_2 + 12x_3 + 4x_4 &= 3 \end{aligned}$$

c)
$$\begin{aligned} x + 3y &= 1 \\ -x - y + z &= 5 \\ 2x + 4y - z &= -7 \end{aligned}$$

1.19 Using `simpleGE` (see Exercise 1.17) as appropriate, find all solutions to the following linear systems of equations:

$$\begin{aligned} x + y &= 5 \\ x - 7y - 12z &= 1 \\ 3x - y - 5z &= 15 \\ 2x + 4y + 3z &= 11 \\ x - y - 3z &= 4\,. \end{aligned}$$

1.20 Find choices of constants c_1, c_2 and c_3 such that $\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$. That is, write $\mathbf{b}$ as a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. If there is only one such linear combination, state how you know this is so. Otherwise, your answer should include all possible choices of the constants c_1, c_2 and c_3.

a) $\mathbf{v}_1 = (1, 3, -4)$, $\mathbf{v}_2 = (0, 1, 2)$, $\mathbf{v}_3 = (-1, -5, 1)$, $\mathbf{b} = (0, -5, -6)$.

b) $\mathbf{v}_1 = (1, 2, 0)$, $\mathbf{v}_2 = (2, 3, 3)$, $\mathbf{v}_3 = (-1, 1, -8)$, $\mathbf{b} = (5, 9, 4)$.

c) $\mathbf{v}_1 = (1, 0, 3, -2)$, $\mathbf{v}_2 = (0, 1, 2, -1)$, $\mathbf{v}_3 = (3, -4, 1, -2)$, $\mathbf{b} = (1, -5, -7, 3)$.

1.21

a) For each given set of matrices, show that they commute (i.e., can be multiplied in any order and give the same answer; find an easy way if you can), and find the product of all matrices in the set. (A missing entry should be interpreted as a zero.)

(i) $\begin{bmatrix} 1 & & \\ b_{21} & 1 & \\ & & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & \\ & 1 & \\ b_{31} & & 1 \end{bmatrix}$

(ii) $\begin{bmatrix} 1 & & & \\ b_{21} & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & & \\ & 1 & & \\ b_{31} & & 1 & \\ & & & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ b_{41} & & & 1 \end{bmatrix}$

(iii) $\begin{bmatrix} 1 & & & \\ & 1 & & \\ & b_{32} & 1 & \\ & & & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & b_{42} & & 1 \end{bmatrix}$

b) Describe as precisely as you can what characterizes the sets of matrices in (i)–(iii) of part (a). (Each is the set of all matrices which . . .)

c) State and prove a general result for n-by-n matrices, of which (i)–(iii) above are special cases.

1.22 Find the nullspace of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 & 3 \\ 3 & 6 & 18 & 9 & 9 \\ 2 & 4 & 6 & 2 & 6 \\ 4 & 8 & 12 & 10 & 12 \\ 5 & 10 & 24 & 11 & 15 \end{bmatrix}.$$

1.23 Consider the system of linear equations

$$\begin{aligned} x_1 + 3x_2 + 2x_3 - x_4 &= 4 \\ -x_1 - x_2 - 3x_3 + 2x_4 &= -1 \\ 2x_1 + 8x_2 + 3x_3 + 2x_4 &= 16 \\ x_1 + x_2 + 4x_3 + x_4 &= 8\,. \end{aligned}$$

a) Determine the associated augmented matrix for this system. Run `simpleGE` on this matrix to see that the algorithm fails to put this augmented matrix into echelon form. Explain why the algorithm fails to do so.

b) Though this would not normally be the case, the output from `simpleGE` for this system may be used to find all solutions to the system anyway. Do so.

1.24 Solve the two linear systems

$$\begin{aligned} x_1 + 2x_2 - 2x_3 &= 1 \\ 2x_1 + 5x_2 + x_3 &= 9 \\ x_1 + 3x_2 + 4x_3 &= 9 \end{aligned} \qquad \text{and} \qquad \begin{aligned} x_1 + 2x_2 - 2x_3 &= 9 \\ 2x_1 + 5x_2 + x_3 &= 9 \\ x_1 + 3x_2 + 4x_3 &= -2 \end{aligned}$$

by doing elimination on a 3-by-5 augmented matrix and then performing two back substitutions.

1.25 A well-known formula for the inverse of a 2-by-2 matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{is} \qquad \mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Use Gaussian elimination (do it *by hand*) on the matrix $\mathbf{A}$ above to derive this formula for the inverse matrix $\mathbf{A}^{-1}$. Handle separately the following cases: I) $a \neq 0$, II) $a = 0$ but $c \neq 0$, and III) both $a, c = 0$. What does a nonzero determinant for $\mathbf{A}$ have to do with nonsingularity in this case?

1.26 Show that the elementary matrix $\mathbf{E}_{ij}$ of Exercise 1.17 is invertible, and find the form of its inverse. You may assume, as is always the case when such elementary matrices are used in Gaussian elimination, that $i > j$.

1.27

a) Suppose

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & a_{32} & 1 \end{bmatrix}.$$

Find a general expression (give the entries) for each of

(i) $\mathbf{A}_1\mathbf{A}_2$ (iii) $\mathbf{A}_2\mathbf{A}_1$

(ii) $(\mathbf{A}_1\mathbf{A}_2)^{-1}$ (iv) $(\mathbf{A}_2\mathbf{A}_1)^{-1}$

b) Suppose

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ a_{31} & 0 & 1 & 0 \\ a_{41} & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & a_{32} & 1 & 0 \\ 0 & a_{42} & 0 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{A}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & a_{43} & 1 \end{bmatrix}.$$

Find a general expression (give the entries) for each of

(i) $\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3$ (iii) $\mathbf{A}_3\mathbf{A}_2\mathbf{A}_1$

(ii) $(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3)^{-1}$ (iv) $(\mathbf{A}_3\mathbf{A}_2\mathbf{A}_1)^{-1}$

c) What special feature does the calculation of $\mathbf{A}_1\mathbf{A}_2$ and $(\mathbf{A}_2\mathbf{A}_1)^{-1}$ (in part (a)) and $\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3$ and $(\mathbf{A}_3\mathbf{A}_2\mathbf{A}_1)^{-1}$ (in part (b)) have? State the corresponding result for arbitrary $n \geq 2$.

1.28 Prove that the product of lower triangular matrices is again lower triangular.

1.29

a) We know that two points in a plane determine a unique line. When those points are not located at the same x-coordinate, the line will take the form

$$p(x) = mx + b\,,$$

a polynomial of degree at most one. Under what conditions on the points would this really be a polynomial of degree zero?

b) If you remember anything about your study of Simpson's Rule in calculus, you may suspect that, when three points no two of which share the same x-coordinate are specified in the plane, there is a unique polynomial

$$p(x) \;=\; ax^2 + bx + c\,,$$

having degree at most 2, that passes through the three points. This statement is, indeed, true. One might say the polynomial p **interpolates** the given points, in that it passes through them filling in the gaps between.

i. Write the similar statement that applies to a set of n points in the plane, no two of which share the same x-coordinate.

ii. Consider the problem of finding the smallest degree polynomial that interpolates the n points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ in the plane. Once the coefficients $a_0, a_1, \ldots, a_{n-1}$ of

$$p(x) \;=\; a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$$

are found, we are done. The information we have at our disposal to find these coefficients is that

$$p(x_1) = y_1\,, \quad p(x_2) = y_2\,, \quad \ldots\,, \quad p(x_n) = y_n\,.$$

That is, we have n equations to determine the n unknowns. Find the matrix $\mathbf{B}$ so that the problem of finding the coefficients of p is equivalent to solving the matrix problem

$$\mathbf{B}\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

c) Use Octave and your answer to the previous part to find the coefficients of the polynomial that interpolates the six points $(-2, -63)$, $(-1, 3)$, $(0, 1)$, $(1, -3)$, $(2, 33)$, and $(3, 367)$.

1.30 Below we have the output from Octave's `lu()` command for a particular matrix.

```
octave> A = [6 -4 5; -4 3 1; 2 -1 1];
octave> [L, U, P] = lu(A)
L =
    1.00000    0.00000    0.00000
   -0.66667    1.00000    0.00000
    0.33333    1.00000    1.00000

U =
    6.00000   -4.00000    5.00000
    0.00000    0.33333    4.33333
    0.00000    0.00000   -5.00000

P =
    1   0   0
    0   1   0
    0   0   1
```

Use it (and *not* some other means) to find all solutions to the linear system of equations

$$\begin{aligned} 6x - 4y + 5z &= -10 \\ -4x + 3y + z &= -1 \\ 2x - y + z &= -1\,. \end{aligned}$$

1.31 Below we have the output from Octave's `lu()` command for a particular matrix.

```
octave> [L, U, P] = lu([1 -2 3; 1 -4 -7; 2 -5 1])
L =
    1.00000    0.00000    0.00000
    0.50000    1.00000    0.00000
    0.50000   -0.33333    1.00000

U =
    2.00000   -5.00000    1.00000
    0.00000   -1.50000   -7.50000
    0.00000    0.00000    0.00000

P =
    0    0    1
    0    1    0
    1    0    0
```

Use it (and *not* some other means) to find all solutions to the linear system of equations

$$\begin{aligned} x - 2y + 3z &= -13 \\ x - 4y - 7z &= 1 \\ 2x - 5y + z &= -19\,. \end{aligned}$$

1.32 Suppose **A** is an m-by-n matrix. Explain why it is not possible for rank(**A**) to exceed m. Deduce that rank(**A**) cannot exceed the minimum value of m and n.

1.33 Give an example of an m-by-n matrix **A** for which you can tell at a glance $\mathbf{Ax} = \mathbf{b}$ is not always consistent—that is, there are right-hand side vectors $\mathbf{b} \in \mathbb{R}^m$ for which no solution exists.

1.34 Let

$$\mathbf{A} := \begin{bmatrix} 1 & -2 & 1 & 1 & 2 \\ -1 & 3 & 0 & 2 & -2 \\ 0 & 1 & 1 & 3 & 4 \\ 1 & 2 & 5 & 13 & 5 \end{bmatrix}.$$

a) Use Gaussian elimination (you may use `simpleGE`, or the (better) alternative called `rref()`) to find the rank and nullity of **A**.

b) Find a basis for the column space of **A**.

c) State another way to phrase the question of part (b) that employs the words "linear independent" and "span".

1.35 Suppose $\mathbf{A}$ is an m-by-n matrix, with $m > n$ and $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$.

a) Are the column vectors of $\mathbf{A}$ linearly independent? How do you know?

b) How many solutions are there to the matrix equation $\mathbf{Ax} = \mathbf{b}$ if $\mathbf{b} \in \text{col}(\mathbf{A})$?

c) How many solutions are there to the matrix equation $\mathbf{Ax} = \mathbf{b}$ if $\mathbf{b} \notin \text{col}(\mathbf{A})$?

1.36 Can a nonzero matrix (i.e., one not completely full of zero entries) be of rank 0? Explain.

1.37 We know that, for an m-by-1 vector $\mathbf{u}$ and 1-by-n matrix (row vector) $\mathbf{v}$, the matrix product $\mathbf{uv}$ is defined, yielding an m-by-n matrix sometimes referred to as the **outer product** of $\mathbf{u}$ and $\mathbf{v}$. In Section 1.3 we called this product a **rank-one matrix**. Explain why this term is appropriate.

1.38 Determine whether the given set of vectors is linearly independent.

a) $S = \{(3, 2, 5, 1, -2), (5, 5, -2, 0, 1), (2, 2, 6, -1, -1), (0, 1, 4, 1, 2)\}$

b) $S = \{(3, 6, 4, 1), (-1, -1, 2, 5), (2, 1, 3, 0), (6, 13, 0, -8)\}$

1.39 For the matrix $\mathbf{A}$, find its nullspace: $\mathbf{A} = \begin{bmatrix} 3 & 6 & 4 & 1 \\ -1 & -1 & 2 & 5 \\ 2 & 1 & 3 & 0 \\ 6 & 13 & 0 & -8 \end{bmatrix}$

1.40 Octave has a command `rank()` which returns a number it thinks equals the rank of a matrix. (Type `help rank` to see how to use it.) The command can be used on square and non-square matrices alike. Use Octave commands to find both the rank and determinant of the following *square* matrices:

(i) $\begin{bmatrix} 2 & 5 & -5 \\ 7 & 0 & 7 \\ -4 & 7 & 0 \end{bmatrix}$

(ii) $\begin{bmatrix} -10 & 0 & 5 & 7 \\ -5 & 3 & -3 & 9 \\ 7 & 7 & -1 & 6 \\ 0 & -9 & -3 & 1 \end{bmatrix}$

(iii) $\begin{bmatrix} 3 & 3 & 0 & -1 & 3 \\ -4 & -2 & 1 & 1 & 5 \\ 1 & -1 & -4 & -5 & -2 \\ 4 & 3 & -3 & 1 & 0 \\ 1 & 5 & 5 & 4 & 1 \end{bmatrix}$

(iv) $\begin{bmatrix} -6 & 5 & 6 & -7 & -2 \\ 4 & -7 & -2 & 5 & 5 \\ -2 & -2 & 4 & -2 & 3 \\ -10 & 12 & 8 & -12 & -7 \\ -8 & 3 & 10 & -9 & 1 \end{bmatrix}$

Using these results, write a statement that describes, for *square* matrices, what knowledge of one of these numbers (the rank or determinant) tells you about the other.

1.41 Theorem 4 tells of many things one can know about a square matrix when is has **full rank**—that is, rank(**A**) = n for a matrix **A** with n columns. Look back through that theorem, and determine which of the conditions (i)–(vii) still hold true when **A** is non-square but has full rank.

1.42 Here is a quick tutorial of how one might use Octave to produce the circle and oval in the figure from Example 17. First, to get the circle, we create a row vector of parameter values (angles, in radians) running from 0 to 2π in small increments, like 0.05. Then we create a 2-row matrix whose 1st row holds the x-coordinates around the unit circle (corresponding to the parameter values) and whose 2nd row contains the corresponding y-coordinates. We then plot the list of x-coordinates against the list of y-coordinates.

```
octave> t = 0:.05:2*pi;
octave> inPts = [cos(t); sin(t)];
octave> plot(inPts(1,:),inPts(2,:))
octave> axis(''square'')
```

The last command in the group above makes sure that spacing between points on the x- and y-axes look the same. (Try the same set of commands omitting the last one.) At this stage, if you wish to draw in some lines connecting the origin to individual points on this circle, you can do so. For instance, given that I chose spacing 0.05 between my parameter values, the "circle" drawn above really consists of 126 individual points (pixels), as evidenced by the commands

```
octave> length(t)
ans = 126

octave> size(inPts)
ans =
    2    126
```

So, choosing (in a somewhat haphazard fashion) to draw in vectors from the origin to the 32nd (green), 55th (red) and 111th (black) of these points, we can use the following commands (assuming that you have not closed the window containing the plot of the circle):

```
octave> hold on
octave> plot([0 inPts(1,32)], [0 inPts(2,32)], 'g')
octave> plot([0 inPts(1,55)], [0 inPts(2,55)], 'r')
octave> plot([0 inPts(1,111)], [0 inPts(2,111)], 'k')
octave> hold off
```

To get the corresponding oval, we need to multiply the vectors that correspond to the points on the circle (drawn using the commands above) by the **A** in Example 17.

```
octave> A = [2 1; 0 3];
```

```
octave> outPts = A*inPts;
octave> plot(outPts(1,:),outPts(2,:))
octave> axis("square")
```

Of course, if you want to know what point corresponds to any individual vector, you can explicitly ask for it. For instance, you can get the point **Av** on the oval corresponding to $\mathbf{v} = (-1/\sqrt{2}, 1/\sqrt{2})$ quite easily using the commands

```
octave> v = [-1/sqrt(2); 1/sqrt(2)]
v =
   -0.70711
    0.70711

octave> A*v
ans =
   -0.70711
    2.12132
```

To see **Av** for the three (colored) vectors we added to our circle's plot, you can use the commands (assuming the window containing the oval is the last plot your produced)

```
octave> subInPts = inPts(:,[32 55 111]);
octave> subOutPts = A*subInPts;
octave> hold on
octave> plot([0 subOutPts(1,1)], [0 subOutPts(2,1)], 'g')
octave> plot([0 subOutPts(1,2)], [0 subOutPts(2,2)], 'r')
octave> plot([0 subOutPts(1,3)], [0 subOutPts(2,3)], 'k')
octave> hold off
```

Use commands like these to help you answer the following questions.

a) Choose an angle $\alpha \in [0, 2\pi)$ and form the corresponding matrix **A** of the form (1.4). In Item 1 of Section 1.4 we established that multiplication by **A** achieves a rotation of the plane. Find the eigenvalues of **A**.

b) Consider the matrix **A** from Example 16. What are its eigenvalues? Describe as accurately as you can the way the plane $\mathbb{R}^2$ is transformed when vectors are multiplied by this **A**.

c) Still working with the matrix **A** from Example 16, write it as a product $\mathbf{A} = \mathbf{BC}$, where both matrices on the right side are 2-by-2, one of which has the form (1.4) and the other has the form (1.7). (Hint: If $(a + bi)$ is one of the eigenvalues of **A**, then the quantity $\sqrt{a^2 + b^2}$ should come into play somewhere.)

d) Make your best effort to accurately finish this statement:

> If **A** is a 2-by-2 matrix with complex eigenvalues $(a + bi)$ and $(a - bi)$ (with $b \neq 0$), then multiplication by **A** transforms the plane $\mathbb{R}^2$ by . . . .

1.43 Suppose **A** is a 2-by-2 matrix with real-number entries and having at least one eigenvalue that is real.

a) Explain how you know **A** has at most one other eigenvalue.

b) Can **A** have a non-real eigenvalue along with the real one? Explain.

c) Consider the mapping $(\mathbf{x} \mapsto \mathbf{A}\mathbf{x})\colon \mathbb{R}^2 \to \mathbb{R}^2$. Is it possible that, given the matrix **A**, this function brings about a (rigid) rotation of the plane? Explain.

1.44 Write a matrix **A** such that, for each $\mathbf{v} \in \mathbb{R}^2$, $\mathbf{A}\mathbf{v}$ is the reflection of **v**

a) across the y-axis. Then use Octave to find the eigenpairs of **A**.

b) across the line $y = x$. Use Octave to find the eigenpairs of **A**.

c) across the line $y = (-3/4)x$. Use Octave to find the eigenpairs of **A**.

d) across the line $y = (a/b)x$, where a, b are arbitrary real numbers with $b \neq 0$.

1.45

a) Write a 3-by-3 matrix **A** whose action on $\mathbb{R}^3$ is to reflect across the plane $x = 0$. That is, for each $\mathbf{v} \in \mathbb{R}^3$, $\mathbf{A}\mathbf{v}$ is the reflection of **v** across $x = 0$. Use Octave to find the eigenpairs of **A**.

b) Write a 3-by-3 matrix **A** whose action on $\mathbb{R}^3$ is to reflect across the plane $y = x$. (Hint: Your answer should be somehow related to your answer to part (b) of Exercise 1.44.) Use Octave to find the eigenpairs of **A**.

c) Suppose P is a plane in 3D space containing the origin, and **n** is a normal vector to P. What, in general, can you say about the eigenpairs of a matrix **A** whose action on $\mathbb{R}^3$ is to reflect points across the plane P?

1.46

a) Consider a coordinate axes system whose origin is always fixed at the Earth's center, and whose positive z-axis always passes through the North Pole. While the positive x- and y- axes always pass through the Equator, the rotation of the Earth causes the points of intersection to change, cycling back every 24 hours. Determine a 3-by-3 matrix **A** so that, given any $\mathbf{v} \in \mathbb{R}^3$ that specifies the current location of a point on (or in) the Earth relative to this coordinate system, $\mathbf{A}\mathbf{v}$ is the location of this same point in 3 hours.

b) Repeat the exerise, but now assuming that, in every 3-hour period, the poles are 1% farther from the origin than they were before.

1.47 When connected in the order given, the points $(0,0)$, $(0.5,0)$, $(0.5,4.5)$, $(4,4.5)$, $(4,5)$, $(0.5,5)$, $(0.5,7.5)$, $(5.5,7.5)$, $(5.5,8)$, $(0,8)$ and $(0,0)$ form the letter 'F', lying in Quadrant I with the bottom of the stem located at the origin.

a) Give Octave commands that produce a plot of the letter with the proper aspect. (Include among them the command you use to store the points, doing so not storing the points themselves, but their corresponding homogeneous coordinates, storing them as `hPts`.)

b) What 3-by-3 matrix would suffice, via matrix multiplication, to translate the letter to Quadrant III, with its top rightmost point at the origin? Give Octave commands that carry out this transformation on `hPts` and produce the plot of the letter in its new position.

c) What 3-by-3 matrix would suffice, via matrix multiplication, to rotate the letter about its effective center (the point $(2.75,4)$), so that it still lies entirely in Quadrant I, but is now upside down? Give Octave commands that carry out this transformation on `hPts` and produce the plot of the letter in its new position.

d) Extract the original points from their homogeneous coordinates with the command

```
octave> pts = hPts(1:2,:);
```

Now consider 2-by-2 matrices of the form

$$\mathbf{A} = \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}.$$

Choose several different values of c, run the command

```
octave> chpts = A*pts;
```

and observe the effect by plotting the altered points (found in `chpts`). These matrices are called **shear matrices**. For each $\mathbf{A}$ you try, find the eigenpairs of $\mathbf{A}$. Summarize your observations about the effect of shear matrices on the letter, and what you note about the eigenpairs.

1.48

a) Suppose $\mathbf{D}$ is a diagonal matrix with entries along the main diagonal $d_1, \ldots, d_n$. Suppose also that $\mathbf{A}$, $\mathbf{S}$ are n-by-n matrices with $\mathbf{S}$ nonsingular, such that the equation $\mathbf{AS} = \mathbf{SD}$ is satisfied. If $\mathbf{S}_j$ denotes the j^{th} column (a vector in $\mathbb{R}^n$) of $\mathbf{S}$, show that each $(d_j, \mathbf{S}_j)$ is an eigenpair of $\mathbf{A}$.

b) Find a matrix **A** for which $(4, 1, 0, -1)$ is an eigenvector corresponding to eigenvalue (-1), $(1, 2, 1, 1)$ is an eigenvector corresponding to eigenvalue 2, and both $(1, -1, 3, 3)$ and $(2, -1, 1, 2)$ are eigenvectors corresponding to eigenvalue 1. (You may use OCTAVE for this part, supplying your code and using commentary in identifying the result.)

c) Show that, under the conditions of part (a), $\det(A) = \prod_{j=1}^{n} d_j$. That is, $\det(A)$ is equal to the product of the eigenvalues of **A**. (This result is, in fact, true even for square matrices **A** which do not have this form.)

d) Two square matrices **A**, **B** are said to be **similar** if there is an invertible matrix **P** for which $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. Show that, if λ is an eigenvalue of **B**, then it is also an eigenvalue of **A**.

2 Vector Spaces

2.1 Properties and Examples of Vector Spaces

In Chapter 1 we often worked with individual vectors. We now shift our focus to entire collections, or **spaces**, of vectors all of the same type.

2.1.1 Properties of $\mathbb{R}^n$

The set of all 2-by-1 matrices with real-number entries is called **Euclidean 2-space**, or $\mathbb{R}^2$. The 3-by-1 matrices with real entries are collectively referred to as $\mathbb{R}^3$, Euclidean 3-space. The sets $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, ..., $\mathbb{R}^n$ and so on, are all examples of **vector spaces**. The fundamental properties shared by these Euclidean spaces are summed up in the following theorem.

Theorem 6: Let $n > 0$ be an integer. The following properties hold in $\mathbb{R}^n$.

(i) For each pair of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, $\mathbf{u} + \mathbf{v} \in \mathbb{R}^n$ as well. This property is usually summed up by saying $\mathbb{R}^n$ is **closed under addition**.

(ii) For each $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, $c\mathbf{u} \in \mathbb{R}^n$. We summarize this property by saying $\mathbb{R}^n$ is **closed under scalar multiplication**.

(iii) Addition in $\mathbb{R}^n$ is **associative**: Given any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$.

(iv) Addition in $\mathbb{R}^n$ is **commutative**: Given any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

(v) Existence of an **additive identity**: There is an element, call it $\mathbf{0}$, found in $\mathbb{R}^n$ having the property that, for each $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{0} + \mathbf{v} = \mathbf{v}$.

(vi) Existence of **additive inverses**: Still denoting the additive identity of $\mathbb{R}^n$ by $\mathbf{0}$, for each $\mathbf{v} \in \mathbb{R}^n$ there is a corresponding element $\tilde{\mathbf{v}}$ such that $\mathbf{v} + \tilde{\mathbf{v}} = \mathbf{0}$.

(vii) **Distributivity I**: For each pair of real numbers c, d and each element $\mathbf{v} \in \mathbb{R}^n$, $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$.

(viii) **Distributivity II**: For each real number c and each pair of elements $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.

(ix) Given any pair of real numbers c, d and any element $\mathbf{v} \in \mathbb{R}^n$, $(cd)\mathbf{v} = c(d\mathbf{v}) = d(c\mathbf{v})$.

(x) Given any $\mathbf{v} \in \mathbb{R}^n$, $(1)\mathbf{v} = \mathbf{v}$. (That is, scalar multiplication by 1 leaves $\mathbf{v}$ unchanged.)

Indeed, there are other sets besides the Euclidean spaces $\mathbb{R}^n$ $(n = 1, 2, \ldots)$ which have these same ten properties. Let $k \geq 0$ be an integer. Then the set $\mathscr{C}^k(a, b)$ consisting of functions whose

- domains include the interval $a < x < b$, and
- are k-times continuously differentiable on that interval,

also have these ten properties. (That is, one can replace every instance of "$\mathbb{R}^n$" in Theorem 6 with "$\mathscr{C}^k(a, b)$" and the theorem still holds.) Since these ten properties are prevalent outside the Euclidean spaces, they have come to be the defining criteria for a **real vector space**, the criteria any set of objects called a *(real) vector space* must have. We will state this as a definition.

Definition 9: Suppose $\mathcal{V}$ is a collection of objects that comes equipped with definitions for addition (i.e., the sum of any two objects in $\mathcal{V}$ is defined) and scalar multiplication (i.e., any object in $\mathcal{V}$ may be multiplied by any real number c). If Theorem 6 still holds when every instance of "$\mathbb{R}^n$" appearing in the theorem is replaced by "$\mathcal{V}$", then $\mathcal{V}$ is a **real vector space** (or **vector space over the reals**).

Example 18:

Let $\mathbf{v} \in \mathbb{R}^3$ be fixed and nonzero, and consider the set $\mathcal{U} = \{t\mathbf{v} \,|\, t \in \mathbb{R}\}$ consisting of all scalar multiples of $\mathbf{v}$. Notice that another way to describe $\mathcal{U}$ is that it is span($\{\mathbf{v}\}$), and that it represents a line in $\mathbb{R}^3$ that passes through the origin. This $\mathcal{U}$ is a vector space. ■

Example 19:

Let $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^3$ be fixed, nonzero, non-parallel vectors. Let $\mathcal{U} = \text{span}(\{\mathbf{u}, \mathbf{v}\})$. This $\mathcal{U}$ is a vector space as well. Speaking geometrically, $\mathcal{U}$ is a plane in $\mathbb{R}^3$, once again containing the origin.

■

The last two examples have followed a theme: Start with a collection of vectors, and look at their span. In the next section, we will show that the span of a collection S of vectors in $\mathbb{R}^n$ is always a vector space. To verify this, it seems one should have to do as described in Definition 9—that is, run through all ten items in Theorem 6, replacing $\mathbb{R}^n$ with $\mathcal{U} = \text{span}(S)$, and check that each holds. As we will learn in the next section, since the items in $\mathcal{U}$ are all from $\mathbb{R}^n$, already known to be a vector space, there is a much shorter list of things to check.

2.1.2 Some non-examples

Theorem 6 tells us $\mathbb{R}^2$ is a (Euclidean) vector space. One may equate this space with the 2-dimensional plane—the set of points $\{(x, y) \,|\, x, y \text{ are real nos.}\}$. Our next example shows why, when you take incomplete portions (or **proper subsets**) of the plane, you often wind up with a set that is not a vector space.

Example 20:

Within $\mathbb{R}^2$, consider the set of vectors S which, when placed in standard position (i.e., each having initial point at the origin), have terminal point on or above the x-axis. That is, $S := \{(x, y) \,|\, y \geq 0\}$. S is not a vector space for several reasons. One is that it is not closed under scalar multiplication (property (ii)). As an example, even though the vector $(3, 1)$ is in S and (-1) is a real number, $(-1)(3, 1) = (-3, -1)$ is *not* found in S.

A related reason is this. Though S has an additive identity $(0, 0)$, it does not, in general contain additive inverses (property (vi)). There is simply no element *found in* S that you can add to $(3, 1)$ to get $(0, 0)$.

■

Example 21:

In Example 7 we found the solution set of the matrix equation

$$\begin{bmatrix} 2 & 1 & -1 \\ 4 & 2 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

to be

$$S = \{(2, 0, 1) + t(-1, 2, 0) \,|\, t \in \mathbb{R}\} \ .$$

This set is not a vector space, since it is not closed under addition (property (i)). For instance, both the vectors $\mathbf{u} = (2, 0, 1)$ and $\mathbf{v} = (1, 2, 1)$ are found in S, the first arising when $t = 0$ and the 2^{nd} when $t = 1$. Were S closed under addition, then $\mathbf{u} + \mathbf{v} = (3, 2, 2)$ would be in S as well. However

$$\begin{bmatrix} 2 & 1 & -1 \\ 4 & 2 & 1 \end{bmatrix} (\mathbf{u} + \mathbf{v}) \;=\; \begin{bmatrix} 2 & 1 & -1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix} \;=\; \begin{bmatrix} 6 \\ 18 \end{bmatrix},$$

not $(3,9)$, so $\mathbf{u}+\mathbf{v}$ is not in the solution set S.

■

In the same vein as the last one, we next consider an example where the set consists of solutions to a linear nonhomogeneous differential equation.

Example 22:

Consider the linear differential equation

$$y''(x) + y(x) = x .$$

Using standard techniques, one can show that the general solution of this ODE is

$$y(x) = x + c_1 \cos(x) + c_2 \sin(x) .$$

That is, the general solution is a set of functions

$$S = \{x + c_1 \cos x + c_2 \sin x \,|\, c_1, c_2 \in \mathbb{R}\} .$$

Both

$$y_1(x) = x + \cos x \qquad \text{and} \qquad y_2(x) = x + \sin x$$

are in the set S, the first arising when $c_1 = 1$, $c_2 = 0$ and the 2^{nd} when $c_1 = 0$, $c_2 = 1$. If S were closed under addition, then

$$y_1(x) + y_2(x) = 2x + \cos(x) + \sin x$$

would also be in S (i.e., it would solve the differential equation). We leave it to the reader to show this is not the case.

■

2.2 Vector Subspaces

In the previous section we suggested (without proof) that, given any collection S of vectors from a vector space $\mathcal{V}$, span(S) is a vector space. This idea that inside any given vector space $\mathcal{V}$ there lies many other vector spaces is key to a complete understanding of the functions $(\mathbf{v} \mapsto \mathbf{A}\mathbf{v})$ (where $\mathbf{A}$ is some given m-by-n matrix) we studied in Chapter 1.

Definition 10: Let $\mathcal{V}$ be a vector space, equipped with addition + and scalar multiplication $\cdot$, and $\mathcal{W}$ be a subset of $\mathcal{V}$. If $\mathcal{W}$, when equipped with the same addition and scalar multiplication as $\mathcal{V}$, is itself a vector space, then we call $\mathcal{W}$ a **vector subspace** (or just **subspace**) of $\mathcal{V}$.

We intimated in the last section that there is a much easier process for checking if a collection of vectors is a subspace than verifying the properties in Theorem 6. We give this test now.

Theorem 7 (Subspace Test): A subset $\mathcal{W}$ of a vector space $\mathcal{V}$ is a subspace of $\mathcal{V}$ if and only if it is the case that, for each $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, $\text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$ is a subset of $\mathcal{W}$. That is, for each $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$ and each scalar c_1, c_2, the linear combination $c_1\mathbf{w}_1 + c_2\mathbf{w}_2$ is in $\mathcal{W}$ as well (i.e., $\mathcal{W}$ is **closed under taking linear combinations**).

Following are the most important examples of vector subspaces. A couple of these have been asserted (in the previous section) to be vector spaces, but we put off demonstrating how we knew this because we did not yet have the Subspace Test.

Example 23:

1. **span(S) is a subspace**.
   Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a subset of the vector space $\mathbb{R}^n$. Then $\mathcal{W} = \text{span}(S)$ is a subspace of $\mathbb{R}^n$. To see this, let $c_1, c_2 \in \mathbb{R}$ and $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, the latter meaning that there are coefficients $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ such that

   $$\mathbf{w}_1 = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k \qquad \text{and} \qquad \mathbf{w}_2 = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k .$$

   Thus, the linear combination

   $$\begin{aligned} c_1\mathbf{w}_1 + c_2\mathbf{w}_2 &= c_1(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k) + c_2(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k) \\ &= (c_1a_1 + c_2b_1)\mathbf{v}_1 + (c_1a_2 + c_2b_2)\mathbf{v}_2 + \cdots + (c_1a_k + c_2b_k)\mathbf{v}_k \end{aligned}$$

   is, once again, a linear combination of vectors in S.
   Note that, in one fell swoop, this establishes that
   - All lines in $\mathbb{R}^n$ that pass through the origin (zero element) are subspaces of $\mathbb{R}^n$. That is, for any such line one may select a representative vector $\mathbf{v}$ (one whose direction matches that of the line) so that the line is just $\text{span}(\{\mathbf{v}\})$.
   - All planes in $\mathbb{R}^n$ that pass through the origin are subspaces of $\mathbb{R}^n$. That is, for any such plane one may select two *linearly independent* vectors $\mathbf{u}, \mathbf{v}$ lying in the plane so that the plane is simply $\text{span}(\{\mathbf{u}, \mathbf{v}\})$.
   - Similar to the above, all **hyperplanes** (higher-dimensional planes) in $\mathbb{R}^n$ that pass through the origin are subspaces of $\mathbb{R}^n$.
   - The column space of an m-by-n matrix $\mathbf{A}$ is a subspace of $\mathbb{R}^m$.
2. **The trivial subspace**.
   Let $\mathcal{V}$ be a vector space, and $\mathbf{0}$ its additive identity. We verify that $\mathcal{W} = \{\mathbf{0}\}$ is a subspace by assuming that $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, and $c_1, c_2 \in \mathbb{R}$. Though these elements were arbitrarily chosen, since $\mathcal{W}$ has only $\mathbf{0}$ in it, both $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{0}$. Thus,

   $$c_1\mathbf{w}_1 + c_2\mathbf{w}_2 = (c_1)\mathbf{0} + (c_2)\mathbf{0} = \mathbf{0} ,$$

showing that $c_1\mathbf{w}_1 + c_2\mathbf{w}_2 \in \mathcal{W}$; that is, $\mathcal{W}$ is closed under taking linear combinations.

3. **Nullspace is a subspace**.
More specifically, if $\mathbf{A}$ is an m-by-n matrix, then $\text{null}(\mathbf{A})$ is a subspace of $\mathbb{R}^n$. To see this, let $c_1, c_2 \in \mathbb{R}$ and $\mathbf{v}_1, \mathbf{v}_2 \in \text{null}(\mathbf{A})$. Then
$$\mathbf{A}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) \;=\; c_1\mathbf{A}\mathbf{v}_1 + c_2\mathbf{A}\mathbf{v}_2 \;=\; (c_1)\mathbf{0} + (c_2)\mathbf{0} \;=\; \mathbf{0}\,,$$
showing that the linear combination $(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)$ is in $\text{null}(\mathbf{A})$ as well.

4. **Eigenspace is a subspace**.
Let $\mathbf{A}$ be an n-by-n matrix with real eigenvalue λ. The set
$$E_\lambda \;:=\; \{\mathbf{v} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$$
is known as the **λ-eigenspace** (or the **eigenspace associated with eigenvalue** λ) of $\mathbf{A}$. To see that E_λ is a subspace of $\mathbb{R}^n$, first notice that it contains only vectors from $\mathbb{R}^n$. Moreover, for any pair of vectors $\mathbf{v}_1, \mathbf{v}_2 \in E_\lambda$, and any real numbers c_1, $c_2 \in \mathbb{R}$, we have
$$\mathbf{A}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) \;=\; c_1\mathbf{A}\mathbf{v}_1 + c_2\mathbf{A}\mathbf{v}_2 \;=\; c_1(\lambda\mathbf{v}_1) + c_2(\lambda\mathbf{v}_2) \;=\; \lambda(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)\,,$$
which shows that the linear combination $(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)$ satisfies the equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ it is required to satisfy in order to be in E_λ.

Since it is the case, sometimes, that an eigenvalue λ is nonreal, a slight modification of the above is needed in such cases. In this case, the λ-eigenspace of $\mathbf{A}$ is
$$E_\lambda \;:=\; \{\mathbf{v} \in \mathbb{C}^n \,|\, \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}\,.$$
It is still a vector space in its own right, but is now a vector subspace of $\mathbb{C}^n$.

■

2.3 Bases and Dimension

In Section 1.8 we introduced the term **basis**. The context in which we used this word was that we had a collection $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ of vectors—perhaps the columns of a matrix—and we used a process (Gaussian elimination) to eliminate whatever vectors necessary from this collection to arrive at a linear independent (sub-)collection B. We called B a *basis* for $\text{span}(S)$, not just because B is a linearly independent collection of vectors, but also because its *span* (the collection of vectors which may be written as a linear combination of vectors in B) is the same as $\text{span}(S)$. We summarize this in the following definition.

Definition 11: A collection B of vectors from a vector space $\mathcal{V}$ is called a **basis** for $\mathcal{V}$ if

- B **spans** $\mathcal{V}$ (i.e., $\mathcal{V} = \text{span}(B)$), and
- B is linearly independent.

There are some situations in which a basis is readily apparent.

Example 24:

The collection of vectors $B = \{\mathbf{i}, \mathbf{j}\}$ is a basis for $\mathbb{R}^2$. To verify this, we must check several things. First, the vectors in B must come from $\mathbb{R}^2$; they do, so long as we understand $\mathbf{i}$ to be $(1, 0)$ and $\mathbf{j} = (0, 1)$. Next, it must be the case that every vector $\mathbf{v} \in \mathbb{R}^2$ is expressible as a linear combination of the vectors in B; this is so, since

$$\mathbf{v} \;=\; (v_1, v_2) \;=\; v_1\mathbf{i} + v_2\mathbf{j}\,.$$

Finally, it must be the case that the set B is linearly independent; we can see this is so by building a matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ whose columns are the vectors in B and noting that it is already an echelon form having no free columns. (There are, of course, other ways to see that the vectors in B are linearly independent.)

We may also think of $\mathbf{i}, \mathbf{j}$ as vectors in $\mathbb{R}^3$ (that is, $\mathbf{i} = (1, 0, 0)$ and $\mathbf{j} = (0, 1, 0)$). If we join to these the vector $\mathbf{k} = (0, 0, 1)$, then the set $B = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is a basis for $\mathbb{R}^3$.

More generally, for any fixed positive integer n we may take $\mathbf{e}_j \in \mathbb{R}^n$, for $j = 1, 2, \ldots, n$, to be the vector all of whose elements are zero except for the i^{th} element, which is 1. That is, $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, etc. Then the collection $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$. The claims of this and the previous paragraph may be verified in precisely the same way as was our claim of a basis for $\mathbb{R}^2$ in the first paragraph.

■

While the **standard bases** of the previous example are important bases for the Euclidean vector spaces, it is sometimes more convenient to use an alternate basis. There are, in fact, infinitely many bases for $\mathbb{R}^n$, $n \geq 1$. The next example gives an alternate basis for $\mathbb{R}^3$.

Example 25:

Show that the set $B = \{(1,0,0),(1,1,0),(1,1,1)\}$ is a basis for $\mathbb{R}^3$.

Obviously all of the vectors in B come from $\mathbb{R}^3$. Let us form a matrix whose columns are the vectors in B:

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

For a different set of vectors B, we might have had to use Gaussian elimination to obtain an echelon form. But here our matrix is already an echelon form. Since it has no free columns, B is a linear independent set.

The final criterion (for B to be a basis for $\mathbb{R}^3$) is that each vector $\mathbf{b} \in \mathbb{R}^3$ needs to be expressible as a linear combination of the vectors in B. That is, given any $\mathbf{b} \in \mathbb{R}^3$, there must be scalars c_1, c_2, c_3 that satisfy the criterion

$$c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{b}\,, \qquad \text{or} \qquad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Since the (echelon) matrix $\mathbf{A}$ has no row of zeros at the bottom, this problem (i.e., "Given $\mathbf{b}$, find $\mathbf{c} = (c_1, c_2, c_3)$ so that $\mathbf{Ac} = \mathbf{b}$") is consistent. (We need not solve for the coefficients c_1, c_2, c_3 to know that the equation *can* be solved.)

■

Let's consider the previous example more carefully. We started with a collection of vectors. We formed a matrix using these vectors as columns, and found a row equivalent echelon form (in the example, this did not require any work). Since our echelon form has no free columns, we conclude that the vectors are linearly independent; since the echelon form has no row consisting solely of zeros, we conclude they span our space (in this case, span $\mathbb{R}^3$).

When will a set $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of vectors in $\mathbb{R}^n$ *not* be a basis for $\mathbb{R}^n$? Assembling the vectors of S into the matrix $\mathbf{A} = \left[\; \mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_k \;\right]$, we see the set S *cannot be a basis* if

- $k > n$. In this case, it is a certainty that the matrix $\mathbf{A}$ has an echelon form with free columns. (Recall, by Exercise 1.32, that the rank of a matrix is never larger than the smaller of its two dimensions. In our case, the matrix is n-by-k with $k > n$, so $\text{rank}(\mathbf{A}) < k$, or $\text{nullity}(\mathbf{A}) > 0$.)

- $k < n$. To see this, suppose $\text{rank}(\mathbf{A}) = k$ (i.e., every column of an echelon form $\mathbf{R}$ for $\mathbf{A}$ contains a pivot). In this case, there are $n - k$ rows of zeros at the bottom of $\mathbf{R}$, and there are certainly vectors $\mathbf{b} \in \mathbb{R}^n$ for which

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \end{array}\right] \sim \left[\begin{array}{cccccc|c} p & * & * & \cdots & * & & * \\ 0 & p & * & \cdots & * & & * \\ 0 & 0 & p & \cdots & * & & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & p & & * \\ 0 & 0 & 0 & \cdots & 0 & & q_1 \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & & q_{n-k} \end{array}\right],$$

where one or more of the values $q_1, \ldots, q_{n-k}$ in the augmented column is nonzero. Thus, the system is inconsistent for some vectors $\mathbf{b} \in \mathbb{R}^n$, and hence S does not span $\mathbb{R}^n$. In the case $\operatorname{rank}(\mathbf{A}) < k$, this only adds more rows of zeros at the bottom, compounding the problem.

- it is a linearly *dependent* set of vectors. Equivalently, it cannot be a basis when the matrix $\mathbf{A} = \left[\begin{array}{c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{array}\right]$ has an echelon form with free columns.

These observations suggest the following theorem which, though we do not prove it here, is nonetheless true.

Theorem 8: Every basis of a (fixed) vector space $\mathcal{V}$ contains the same number of vectors. Moreover, if the number of elements comprising a basis for $\mathcal{V}$ is n (finite), then a collection B containing n vectors from $\mathcal{V}$ is a basis for $\mathcal{V}$ if and only if B is linearly independent.

We have seen that there are vector spaces which lie *inside* of $\mathbb{R}^n$, the *subspaces* of $\mathbb{R}^n$.

1. Each line through the origin of $\mathbb{R}^n$ is a subspace $\mathcal{W}$ of $\mathbb{R}^n$. If we select any nonzero vector $\mathbf{v}$ that is parallel to the line, then $\{\mathbf{v}\}$ is a basis for $\mathcal{W}$. A basis for such a subspace $\mathcal{W}$ must contain exactly one vector, even though there is considerable freedom in the choice of that vector.

2. Let $\mathcal{W}$ be a plane in $\mathbb{R}^n$ passing through the origin. If we select nonzero, nonparallel vectors (i.e., linearly independent) $\mathbf{u}, \mathbf{v} \in \mathcal{W}$, then $\{\mathbf{u}, \mathbf{v}\}$ is a basis for $\mathcal{W}$. While there are many bases (many different choices of $\mathbf{u}, \mathbf{v}$) for such a subspace $\mathcal{W}$, a basis must contain exactly two vectors.

3. For a given m-by-n matrix $\mathbf{A}$, col($\mathbf{A}$) is a subspace of $\mathbb{R}^m$. Our method for finding a basis for col($\mathbf{A}$) is to use Gaussian elimination to find an echelon form $\mathbf{R}$ for $\mathbf{A}$. We then take as our basis those columns of $\mathbf{A}$ which correspond to pivot columns in $\mathbf{R}$. While there are other bases for col($\mathbf{A}$), all such bases have precisely rank($\mathbf{A}$) vectors.

4. For a given m-by-n matrix $\mathbf{A}$, null ($\mathbf{A}$) is a subspace of $\mathbb{R}^n$. Our method for finding null ($\mathbf{A}$) is to use Gaussian elimination to find an echelon form $\mathbf{R}$ for $\mathbf{A}$. We then write out solutions to the equation $\mathbf{Rx} = \mathbf{0}$. Except in the case null ($\mathbf{R}$) = $\{\mathbf{0}\}$, these solutions take the form

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k, \qquad t_1, t_2, \ldots, t_k \in \mathbb{R} ,$$

   where k = nullity ($\mathbf{A}$) is the number of free columns in $\mathbf{R}$. The result is $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for null ($\mathbf{A}$). Again, there are other bases for null ($\mathbf{A}$), but all such bases contain precisely nullity ($\mathbf{A}$) vectors.

All of these examples reflect the conclusions of Theorem 8—i.e., that for any (but the trivial) real vector space $\mathcal{V}$, there are many bases, but the *number* of elements in a basis for $\mathcal{V}$ is fixed. That number is known as the **dimension** of $\mathcal{V}$, or dim($\mathcal{V}$). The lines which are subspaces of $\mathbb{R}^n$ have dimension 1; the planes which are subspaces have dimension 2. By convention, the trivial vector (sub)space $\{\mathbf{0}\}$ is said to have dimension 0. By Example 24, the dimension of the entire space $\mathbb{R}^n$ is n.

Along with col($\mathbf{A}$) and null ($\mathbf{A}$), there are two other vector spaces commonly associated with a given m-by-n matrix $\mathbf{A}$. The first is the **row space** of $\mathbf{A}$, denoted by col($\mathbf{A}^T$). As the notation suggests, the row space of $\mathbf{A}$ is just the column space of $\mathbf{A}^T$, and hence is a vector space. More specifically, it is a subspace of $\mathbb{R}^n$.

Were we to want to know the dimension of col($\mathbf{A}^T$), or to find a basis for it, we could carry out this now-familiar process: (i) reduce $\mathbf{A}^T$ to echelon form, (ii) determine the quantity (which equals dim(col($\mathbf{A}^T$))) and placement of free columns, and keep just the pivot columns of $\mathbf{A}^T$ for our basis. At the heart of this method is the understanding that, using $\mathbf{r}_1, \ldots, \mathbf{r}_m$ to denote the rows (1-by-n matrices) of $\mathbf{A}$, col($\mathbf{A}^T$) = span($\{\mathbf{r}_1^T, \ldots, \mathbf{r}_m^T\}$). Thus, $S = \{\mathbf{r}_1^T, \ldots, \mathbf{r}_m^T\}$ serves as a starter set *en route* to a basis for col($\mathbf{A}^T$). To get a basis, we eliminate any of the rows $\mathbf{r}_j$ which are *redundant* in the sense that $\mathbf{r}_j$ is, itself, a linear combination of the other rows.

But there is no need to work with $\mathbf{A}^T$ to figure out which rows are redundant. Gaussian elimination *on* $\mathbf{A}$ *itself* involves instances of **Row Operation 3** (adding a multiple of one row to another) in sufficient quantity to zero out those rows which are linear combinations of previous rows. Those rows which, in an echelon matrix for $\mathbf{A}$, do not contain *only* zeros—rows we might call **pivot rows** because they each contain a pivot—are *not* zeroed out precisely because they are *not* linear combinations of the previous ones. And, the term **row equivalence** we use when $\mathbf{A} \sim \mathbf{B}$ (i.e., when $\mathbf{B}$ may be obtained from $\mathbf{A}$ through a sequence of *elementary row operations*) is suggestive of the truth, which is that both matrices have the same row space. We thus arrive at the following alternate procedure for finding a basis for and the dimension of $\mathrm{col}(\mathbf{A}^T)$:

Finding the dimension of, and a basis for, the row space of A:

- Reduce $\mathbf{A}$ to an echelon form $\mathbf{R}$.
- Count the number of pivots (or the number of pivot rows/columns) in $\mathbf{R}$. This number is $\dim(\mathrm{col}(\mathbf{A}^T))$. (Note that this also means $\dim(\mathrm{col}(\mathbf{A}^T)) = \mathrm{rank}(\mathbf{A})$.)
- Determine the pivot (not-fully-zero) rows in $\mathbf{R}$. (These rows will conveniently appear together above any zeroed-out rows.) Take the tranposes of these pivot rows; let's call the resulting vectors (elements of $\mathbb{R}^n$) $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Then $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis for $\mathrm{col}(\mathbf{A}^T)$.

The other subspace associated with an m-by-n matrix $\mathbf{A}$ is called the **left nullspace** of $\mathbf{A}$. Once again, the symbol we use to denote the left nullspace, $\mathrm{null}\left(\mathbf{A}^T\right)$, is sufficient to suggest to the reader the criteria vectors must satisfy to be in it. We relegate investigation of $\mathrm{null}\left(\mathbf{A}^T\right)$ to the exercises.

Example 26:

The matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 6 & 1 & 1 \\ 1 & 0 & -1 & 3 \\ 1 & -6 & -4 & 8 \end{bmatrix} \qquad \text{has echelon form} \qquad \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & 1 & 1/2 & -5/6 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Use this to find a basis for the row, column and null spaces of $\mathbf{A}$.

A basis for the row space is $\{(1, 0, -1, 3), (0, 1, 1/2, -5/6)\}$. (Note there are many bases; another basis for $\mathrm{col}(\mathbf{A}^T)$, one that avoids fractions, is $\{(1, 0, -1, 3), (0, 6, 3, -5)\}$.) A basis for the column space is $\{(2, 1, 1), (6, 0, -6)\}$.

There is still some work required to find a basis for $\mathrm{null}\,(\mathbf{A})$. We have two free columns, so we get from the echelon matrix that vectors $\mathbf{x} = (x_1, x_2, x_3, x_4)$ in the nullspace satisfy

$$\begin{aligned} x_1 - x_3 + 3x_4 &= 0\,, \\ x_2 + \tfrac{1}{2}x_3 - \tfrac{5}{6}x_4 &= 0\,, \\ x_3 &= s\,, \qquad s \in \mathbb{R}\,, \\ x_4 &= t\,, \qquad t \in \mathbb{R}\,. \end{aligned}$$

which means

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} s - 3t \\ (5/6)t - (1/2)s \\ s \\ t \end{bmatrix} = s \begin{bmatrix} 1 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 5/6 \\ 0 \\ 1 \end{bmatrix}.$$

So, a basis for $\mathrm{null}\,(\mathbf{A})$ is $\{(1, -1/2, 1, 0), (-3, 5/6, 0, 1)\}$.

■

Exercises

2.1 In Subsection 2.1.1 it is claimed that $\mathscr{C}^k(a,b)$, the set of real-valued functions defined on the open interval (a,b) of the real line which have continuous derivatives up to the k^{th} order, is a real vector space. That is, a theorem like Theorem 6 (with every instance of "$\mathbb{R}^k$" replaced by "$\mathscr{C}^k(a,b)$") is, in fact, true. If so, then what, precisely, is the *additive identity* in $\mathscr{C}^k(a,b)$? Describe, using notation like $(x \mapsto f(x))$, exactly what this additive identity function does. State how you know the function you give as an answer is a member of $\mathscr{C}^k(a,b)$.

2.2 In Example 19 we indicate that, when two nonzero, non-parallel vectors $\mathbf{u}$, $\mathbf{v}$ are selected in $\mathbb{R}^3$, $\mathcal{U} = \text{span}(\{\mathbf{u},\mathbf{v}\})$ is a vector space, described geometrically as a plane through the origin.

a) Show that the requirements "nonzero" and "non-parallel" guarantee that the set $\{\mathbf{u},\mathbf{v}\}$ are linearly independent. Is the opposite (i.e., linear independence of $\{\mathbf{u},\mathbf{v}\}$ guarantees $\mathbf{u}$ and $\mathbf{v}$ are nonzero and non-parallel) true?

b) For $\mathbf{u} = (1,2,-1)$ and $\mathbf{v} = (0,1,1)$, write the equation (write it in the form $Ax+By+Cz = D$) of the plane $\text{span}(\{\mathbf{u},\mathbf{v}\})$. (Hint: This involves material covered in Section 10.5 of **University Calculus**. It involves taking a *cross product*.)

c) You should have found in part (b) that the value of D is zero. Does this always happen with planes through the origin?

2.3 The subset of $\mathbb{R}^3$ described by

$$\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \qquad s,\, t \in \mathbb{R}\,,$$

is a plane in $\mathbb{R}^3$.

a) Show that this plane is not a vector space.

b) Write a similar description of the plane parallel to the given one which passes through the origin.

c) Write, in $Ax + By + Cz = D$ form, the equations of both planes.

2.4 **Error-Correcting Codes: The Hamming (7,4) Code**. In this problem, we wish to look at a method for transmitting the 16 possible 4-bit **binary words**

0000	0001	0010	0011	0100	0101	0110	0111
1000	1001	1010	1011	1100	1101	1110	1111

in such a way that if, for whatever reason (perhaps electrostatic interference), some digit is reversed in transmission (a 0 becomes a 1 or vice versa), then the error is *both* detected and corrected.

First, consider the set $\mathbb{Z}_2^n$. The objects in this set are n-by-1 matrices (in that respect they are like the objects in $\mathbb{R}^n$), with entries that are *all zeros or ones*. We wish to define what it means to *add* objects in $\mathbb{Z}_2^n$, and how to multiply these objects by a reduced list of scalars—namely 0 and 1. When we add vectors from $\mathbb{Z}_2^n$, we do so componentwise (as in $\mathbb{R}^n$), but with each sum calculated mod 2.[1] Scalar multiplication is done mod 2 as well. For instance, in $\mathbb{Z}_2^3$ we have

$$\begin{bmatrix}1\\0\\1\end{bmatrix} + \begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0\\1\end{bmatrix}, \qquad 0\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} \qquad \text{and} \qquad 1\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}1\\0\\1\end{bmatrix}.$$

a) Show that, with this definition of vector addition and scalar multiplication, $\mathbb{Z}_2^n$ is a vector space (over the **field** of scalars $Z_2 = \{0, 1\}$).

Note that, when operations are performed mod 2, an m-by-n matrix times a vector in $\mathbb{Z}_2^n$ produces another vector in $\mathbb{Z}_2^m$. For instance

$$\begin{bmatrix}3 & 1 & 2\\1 & 6 & 0\\2 & 0 & 1\\7 & 2 & 1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\\1\\0\end{bmatrix}, \quad \text{and is equivalent to} \quad \begin{bmatrix}1 & 1 & 0\\1 & 0 & 0\\0 & 0 & 1\\1 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}0\\1\\1\\0\end{bmatrix}.$$

Consider the matrix

$$\mathbf{H} := \begin{bmatrix}0 & 0 & 0 & 1 & 1 & 1 & 1\\0 & 1 & 1 & 0 & 0 & 1 & 1\\1 & 0 & 1 & 0 & 1 & 0 & 1\end{bmatrix}.$$

An easy way to remember this matrix, known as the **Hamming Matrix**, is through noting that beginning from its left column you have, in sequence, the 3-bit binary representations of the integers 1 through 7.

b) Find a basis for null $(\mathbf{H})$, where the matrix product $\mathbf{Hx}$ is to be interpreted mod 2 as described above. (Hint: When you add a scalar multiple of one row to another during row reduction, both the scalar multiplication and the row addition is done componentwise mod 2.)

[1]Modular arithmetic is the type of *integer* arithmetic we use with clocks. For a standard clock, the *modulus* is 12, resulting in statements like "It is now 8 o'clock; in 7 hours it will be 3 o'clock" (i.e., "8 + 7 = 3"). In mod 2 arithmetic, the modulus is 2, and we thereby act as if the only numbers on our "clock" are 0 and 1.

One basis (probably different than the one you found in part (b)) of null$(\mathbf{H})$ is $\mathbf{u}_1 = (1,0,0,0,0,1,1)$, $\mathbf{u}_2 = (0,1,0,0,1,0,1)$, $\mathbf{u}_3 = (0,0,1,0,1,1,0)$, and $\mathbf{u}_4 = (0,0,0,1,1,1,1)$.

c) Show that the collection $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is, indeed, a basis for null$(\mathbf{H})$.

Transmitting a 4-bit Word

Let (c_1, c_2, c_3, c_4) be a 4-bit word (i.e., each c_i is 0 or 1), one we wish to transmit. We use the values $c_1, \ldots, c_4$ to generate a 7-bit word via a linear combination (mod 2) of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. To be precise, instead of the original 4-bit word, we transmit the 7-bit word

$$\mathbf{v} := c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4 = (c_1, c_2, c_3, c_4, c_2 + c_3 + c_4, c_1 + c_3 + c_4, c_1 + c_2 + c_4).$$

This $\mathbf{v}$ is in both $\mathbb{Z}_2^7$ and null$(\mathbf{H})$. (Do you see why it is an element of the latter?)

Error Detection and Correction

Suppose a 7-bit word $\tilde{\mathbf{v}}$ is received. It may be the same as the transmitted $\mathbf{v}$, or it may be a corrupted version of $\mathbf{v}$. Suppose that at most one binary digit of $\tilde{\mathbf{v}}$ is in error. Then the matrix product $\mathbf{H}\tilde{\mathbf{v}}$ tells us what we need to know. To see this, consider two cases:

- There are no errors (that is, $\tilde{\mathbf{v}} = \mathbf{v}$).
  In this case, $\mathbf{H}\tilde{\mathbf{v}} = \mathbf{H}\mathbf{v} = \mathbf{0}$, and the receiver can take the first 4 bits (entries) of $\mathbf{v}$ as the original 4-bit word intended.

- There is an error in position i (so $\tilde{\mathbf{v}} = \mathbf{v} + \mathbf{e}_i$, where $\mathbf{e}_i$ is a vector of zeros except in its i^{th} position, where it has a 1).
  In this case, $\mathbf{H}\tilde{\mathbf{v}} = \mathbf{H}(\mathbf{v} + \mathbf{e}_i) = \mathbf{H}\mathbf{v} + \mathbf{H}\mathbf{e}_i = \mathbf{0} + \mathbf{H}\mathbf{e}_i = \mathbf{H}\mathbf{e}_i = i$th column of $\mathbf{H}$. Thus, $\mathbf{H}\tilde{\mathbf{v}} \neq \mathbf{0}$ in this case. Moreover, by inspecting which column of $\mathbf{H}$ is equal to $\mathbf{H}\tilde{\mathbf{v}}$, we learn which of $\tilde{\mathbf{v}}$'s digits is different from those of $\mathbf{v}$. The receiver may correct that bit in $\tilde{\mathbf{v}}$, and once again take the first 4 bits of this (newly-corrected) $\tilde{\mathbf{v}}$ as the intended word.

d) Suppose that the 7-bit word $(1,0,1,1,1,0,0)$ is received. Assuming that this was originally a 4-bit word that was sent using the Hamming (7,4) error-correcting code, and assuming at most one binary digit becomes corrupted during transmission, what was the original 4-bit word?

e) Investigate what happens if more than one binary digit becomes corrupted in the transmission process (i.e., more than one of the 7 bits changes from 0 to 1 or 1 to 0) and report on your findings.

2.5 Let n be a positive integer. Give an appropriate definition for the vector space $\mathbb{R}^n$.

2.6 Show that the set of vectors

$$S := \left\{\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_1 = x_2\right\},$$

is a subspace of $\mathbb{R}^3$.

2.7 Determine whether the set of vectors

$$S := \left\{ \mathbf{x} = (x_1, 1) \in \mathbb{R}^2 \,|\, x_1 \in \mathbb{R} \right\} ,$$

is a subspace of $\mathbb{R}^2$. (Indicate how you know.)

2.8 Consider the matrices $\mathbf{A} = \begin{bmatrix} 5 & -15 & -25 \\ -5 & 15 & 25 \\ 5 & -15 & -25 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 8 & -15 & -25 \\ -5 & 18 & 25 \\ 5 & -15 & -22 \end{bmatrix}$.

a) Find a basis for null $(\mathbf{A})$.

b) Use Octave to find the eigenpairs of $\mathbf{B}$.

c) Write a sentence describing why part (a) may be viewed as a sub-problem of part (b) (i.e., one of the tasks performed along the way to getting the eigenpairs of $\mathbf{B}$).

d) In light of part (c), you should be able to write down an answer to part (a) using only the Octave output you obtained in part (b). What basis do you get? This basis (i.e., the one you extract from part (b)) probably looks quite different than the answer you worked out in part (a). Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ denote the basis you obtained in part (a), and let $\{\mathbf{w}_1, \mathbf{w}_2\}$ be the basis obtained from Octave output. Show that $\mathbf{v}_1$ and $\mathbf{v}_2$ are in span$(\{\mathbf{w}_1, \mathbf{w}_2\})$—that is, show that there are coefficients (scalars) c_1, c_2, d_1 and d_2 such that

$$\mathbf{v}_1 = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 \qquad \text{and} \qquad \mathbf{v}_2 = d_1\mathbf{w}_1 + d_2\mathbf{w}_2.$$

Is the reverse true as well—i.e., are $\mathbf{w}_1$, $\mathbf{w}_2$ in span$(\{\mathbf{v}_1, \mathbf{v}_2\})$?

e) **Continuation of part (d)**. Explain why it is not *automatically a point of concern* if the bases $\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{w}_1, \mathbf{w}_2\}$ do not agree.

f) **Continuation of part (e)**. Of course, there is the possibility that the bases $\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{w}_1, \mathbf{w}_2\}$ disagree *because of calculation error*, which *would* be a point of concern. One should not be so comfortable with the likelihood of a discrepancy between the two bases that a *faulty* answer obtained by hand goes unnoticed. Describe a method one might use to check a "by-hand" answer.

2.9 Find a basis for span(S).

a) $S = \{(2, 1, 6), (3, 2, -1), (1, 1, -7), (1, 0, 13)\}$

b) $S = \{(1, 3, 2, 1), (-1, -1, 1, 2), (0, 2, 2, 1), (1, 1, 1, -1)\}$

2.10 Suppose $\mathcal{V}$ is a vector space over the reals whose dimension is n, and that S is a spanning set for $\mathcal{V}$. If S contains precisely n vectors from $\mathcal{V}$, must S be a basis? Explain.

2.11 Suppose $\mathbf{A}$ is an m-by-n matrix whose nullspace $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$. Show that, for each $\mathbf{b} \in \mathbb{R}^m$, the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has *at most one* solution. That is, if $\mathbf{x}_1$ and $\mathbf{x}_2$ both satisfy the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ for a given $\mathbf{b}$, then it must be the case that $\mathbf{x}_1 = \mathbf{x}_2$.

2.12 Determine whether the set

a) $\{(2,1),(3,2)\}$ is a basis for $\mathbb{R}^2$, and explain how you know.

b) $\{(1,-1),(1,3),(4,-1)\}$ is a basis for $\mathbb{R}^2$, and explain how you know.

c) $\{(1,0,1),(1,1,0),(0,1,-1)\}$ is a basis for $\mathbb{R}^3$, and explain how you know.

d) $\{(1,0,0),(1,1,0),(1,1,1)\}$ is a basis for $\mathbb{R}^3$, and explain how you know.

2.13 Show that the set $B = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is, indeed, a basis for $\mathbb{R}^3$.

2.14 Consider the vector space $\mathbb{Z}_2^n$ (a *vector space over* $\mathbb{Z}_2$ instead of being *over the reals*) of Exercise 2.4. What, would you guess, is $\dim(\mathbb{Z}_2^n)$? Find a basis for $\mathbb{Z}_2^n$ with the appropriate number of elements to verify your answer.

2.15 Suppose $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a subset of a vector space $\mathcal{V}$, and that $\mathbf{u} \in \text{span}(S)$. Show that

$$\text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\}) = \text{span}(S) .$$

Notice that you are being asked to show the equality of two sets. Two sets A and B are often shown to be equal by showing first that $A \subset B$ and then that $B \subset A$.

2.16 Suppose $\mathbf{u} \in \text{span}(S)$, with $\mathbf{u}$ a nonzero vector. Then, by definition, there exist vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in S$ and corresponding real numbers $a_1, \ldots, a_k$ such that

$$\mathbf{u} \;=\; a_1\mathbf{v}_1 + \ldots + a_k\mathbf{v}_k \, .$$

(Note that you may assume, without any loss of generality, that the $\mathbf{v}_i$ are distinct, and each $a_i \neq 0$.) Show that, if S is linearly independent, then there is no other way to write $\mathbf{u}$ as a linear combination of vectors in S. That is, show that if

$$\mathbf{u} \;=\; b_1\mathbf{w}_1 + \ldots + b_m\mathbf{w}_m \, ,$$

where the $\mathbf{w}_j$ are distinct vectors in S, and each $b_j \neq 0$, then

- $m = k$,
- for each $1 \leq j \leq m$, there is some i with $1 \leq i \leq k$ for which $\mathbf{v}_i = \mathbf{w}_j$, and

- whenever $\mathbf{v}_i = \mathbf{w}_j$, it is the case that $a_i = b_j$.

2.17 Suppose that S is a subset of a vector space $\mathcal{V}$ and that the number of elements in S is finite. If S is linearly independent, show that every nonempty subset of S is linearly independent as well. More specifically, suppose that $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and show that, for $T = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ with $1 \leq m \leq k$, T is linearly independent.

2.18 Suppose S and T are subsets of a vector space $\mathcal{V}$, each containing finitely many vectors. Suppose that S spans $\mathcal{V}$, while T is linearly independent. Denoting the number of elements in T by $|T|$, what can you say about $|T|$ in relation to $|S|$? Explain.

2.19 Suppose $\mathbf{A}$ is an m-by-n matrix.

a) We express the nullspace of $\mathbf{A}$ as

$$\text{null}\,(\mathbf{A}) \;=\; \{\mathbf{v} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{v} = \mathbf{0}\}\ .$$

Write a similar expression for the left nullspace $\text{null}\left(\mathbf{A}^T\right)$ of $\mathbf{A}$.

b) Find a basis for the left nullspace in the specific instance where

$$\mathbf{A} \;=\; \begin{bmatrix} 2 & 6 & 1 & 1 \\ 1 & 0 & -1 & 3 \\ 1 & -6 & -4 & 8 \end{bmatrix}.$$

Also, find $\dim(\text{null}\left(\mathbf{A}^T\right))$.

c) Give an expression involving the number $\text{rank}(\mathbf{A})$ for the dimension of the left nullspace of $\mathbf{A}$. (Make this a general statement, not reliant on any particular knowledge of the entries of $\mathbf{A}$, only using that its dimensions are m-by-n.)

2.20 Consider the matrix $\mathbf{A} \;=\; \begin{bmatrix} 2 & 3 & 0 & -1 \\ 1 & 0 & 3 & 1 \\ -3 & -5 & 1 & 2 \\ 1 & 0 & 3 & 1 \end{bmatrix}$.

a) Find bases for its **four fundamental subspaces**: $\text{col}(\mathbf{A})$, $\text{null}\,(\mathbf{A})$, $\text{col}(\mathbf{A}^T)$ and $\text{null}\left(\mathbf{A}^T\right)$. (Note: You may find it interesting to compare your basis for $\text{col}(\mathbf{A})$ with the results of Example 9.)

b) Select a vector $\mathbf{u}$ from your basis for $\text{col}(\mathbf{A})$ and a vector $\mathbf{v}$ from your basis for $\text{null}\left(\mathbf{A}^T\right)$. Compute their dot product (i.e., compute $\mathbf{u}^T\mathbf{v}$). What do you get? Try selecting different combinations of basis elements from these two spaces. Try it again, now taking $\mathbf{u}$ from $\text{col}(\mathbf{A}^T)$ and $\mathbf{v}$ from $\text{null}\,(\mathbf{A})$. Formulate a general statement encapsulating your findings.

3 Orthogonality and Least-Squares Solutions

3.1 Inner Products, Norms, and Orthogonality

In MATH 162 (and elsewhere), we encounter the **inner product** (also known as a **scalar** or **dot product**) of vectors in $\mathbb{R}^3$, and use it to find things like the length of a vector and the angle between two vectors. In this first part of Section 3.1, we extend these ideas to $\mathbb{R}^n$. In the second part, we use it to define *orthogonality* in $\mathbb{R}^n$, particularly orthogonality of sets. Finally, we discuss briefly the idea of an inner product in a more general setting (outside $\mathbb{R}^n$), and vector spaces that have an inner product.

3.1.1 Inner products

We will denote the **inner product** between vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ by $\langle \mathbf{u}, \mathbf{v} \rangle$, given by

$$\langle \mathbf{u}, \mathbf{v} \rangle := \sum_{j=1}^{n} u_j v_j. \tag{3.1}$$

Example 27:

1. In $\mathbb{R}^3$, the inner product of $\mathbf{u} = (1, 5, -2)$ and $\mathbf{v} = (3, -1, 1)$ is

$$\langle \mathbf{u}, \mathbf{v} \rangle = (1)(3) + (5)(-1) + (-2)(1) = -4.$$

2. For the vectors $\mathbf{u} = (-1, 2, 0, 5)$ and $\mathbf{v} = (-4, 3, 8, -2)$ from $\mathbb{R}^4$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = (-1)(-4) + (2)(3) + (0)(8) + (5)(-2) = 0 .$$

■

We make a few preliminary observations about the inner product defined in (3.1):

- For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} .$$

  That is, the inner product of $\mathbf{u}$ and $\mathbf{v}$ is equal to the matrix product of $\mathbf{u}^T$ and $\mathbf{v}$. Recall that this product *can* be performed in the other order, $\mathbf{v}\mathbf{u}^T$, in which case it is called an **outer product**, and yields an n-by-n **rank-one** matrix.

- One might consider (by holding $\mathbf{u}$ fixed) the function $(\mathbf{v} \mapsto \langle \mathbf{u}, \mathbf{v} \rangle)\colon \mathbb{R}^n \to \mathbb{R}$. This map has the property that, for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and scalars $a, b \in \mathbb{R}$,

$$\langle \mathbf{u}, a\mathbf{v} + b\mathbf{w} \rangle \;=\; a \langle \mathbf{u}, \mathbf{v} \rangle + b \langle \mathbf{u}, \mathbf{w} \rangle \;. \tag{3.2}$$

- There is no provision in our definition of $\langle \mathbf{u}, \mathbf{v} \rangle$ for vectors coming from different vector spaces $\mathbb{R}^n$ and $\mathbb{R}^k$ with $k \neq n$. We need our vectors to have the same number of components.

We take the **length of a vector v** in $\mathbb{R}^n$ to be

$$\|\mathbf{v}\| \;:=\; \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \;=\; \left(\sum_{j=1}^{n} v_j^2 \right)^{1/2} \;.$$

Notice that, according to this definition, $\|\mathbf{v}\|$ is a nonnegative number for all $\mathbf{v} \in \mathbb{R}^n$. In fact, the only way for $\|\mathbf{v}\|$ to be zero is for all components of $\mathbf{v}$ to be zero. A vector whose length is 1 is called a **unit vector**.

Armed with this concept of length, we may define the notion of **distance between two vectors**. In particular, given $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$, we take the distance between $\mathbf{u}$ and $\mathbf{v}$ to be the length of their difference $\|\mathbf{u} - \mathbf{v}\|$.

Given two nonzero vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$, we define the **angle between u and v** to be the number $\theta \in [0, \pi]$ which satisfies

$$\cos\theta \;=\; \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \|\mathbf{v}\|} \;. \tag{3.3}$$

This definition of an angle cannot be visualized when the dimension n of the space exceeds 3, but coincides precisely with our geometric intuition of *angle* when $n = 2$ or 3.

3.1.2 Orthogonality

When the angle θ between two nonzero vectors in $\mathbb{R}^2$ is $\pi/2$ (=90°), we call them **perpendicular**. In expression (3.3) this corresponds to the numerator $\langle \mathbf{u}, \mathbf{v} \rangle$ being zero. Orthogonality, as defined below, is a generalization of this idea of perpendicularity.

Definition 12: Two vectors $\mathbf{u}$ and $\mathbf{v}$ from $\mathbb{R}^n$ are said to be **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. We often write $\mathbf{u} \perp \mathbf{v}$ to indicate $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Note that the two vectors in part 2 of Example 27 are orthogonal. The zero vector in $\mathbb{R}^n$ is orthogonal to every vector of $\mathbb{R}^n$.

Orthogonality of vectors leads to an important result about the lengths of the legs of a triangle with sides $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}+\mathbf{v}$.

Theorem 9 (Pythagorean Theorem): Suppose $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are such that $\mathbf{u} \perp \mathbf{v}$. Then $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof: We have

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\ &= \ldots \quad \text{(You fill in the missing details—see Exercise 3.2)} \quad \ldots \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 . \end{aligned}$$

□

We want to be able to talk not only about orthogonality of vectors, but also of sets.

Definition 13: Two subsets S and T of $\mathbb{R}^n$ are said to be orthogonal if $\mathbf{s} \perp \mathbf{t}$ for every $\mathbf{s} \in S$ and every $\mathbf{t} \in T$. In this case, we write $S \perp T$.

It may be necessary to develop a little intuition about orthogonal sets. For instance, one might think of the floor and a wall of a room as being orthogonal. Translating this intuition to $\mathbb{R}^3$, one might think the sets $\mathcal{U} = \{\mathbf{x} = (x_1, x_2, x_3) \,|\, x_3 = 0\}$ and $\mathcal{W} = \{\mathbf{x} = (x_1, x_2, x_3) \,|\, x_2 = 0\}$, corresponding to the xy- and xz-planes, respectively, were orthogonal. Indeed, if MATH 162 students were asked to find the *angle* between these planes, they would be correct to say it is $\pi/2$. But this is not what we need, according to Definition 13, for two *sets of vectors* to be orthogonal. In fact, $\mathcal{U}$ and $\mathcal{W}$ are *not orthogonal*, as the vector $\mathbf{i} = (1, 0, 0)$ is in both $\mathcal{U}$ and $\mathcal{W}$ but is not orthogonal to itself. The vectors that *are* orthogonal to every vector in $\mathcal{U}$ are those which are parallel to (i.e., scalar multiples of) $\mathbf{k} = (0, 0, 1)$—that is, those in span($\{\mathbf{k}\}$). Span($\{\mathbf{k}\}$) is, in fact, the *orthogonal complement* of $\mathcal{U}$, a concept defined for any subset S of $\mathbb{R}^n$ as follows:

Definition 14: Let S be a subset of $\mathbb{R}^n$. The **orthogonal complement of** S, denoted by $S^\perp$ (pronounced "S perp"), is the set of vectors $\mathbf{u} \in \mathbb{R}^n$ which are orthogonal to every vector in S. That is,

$$S^\perp := \{\mathbf{u} \in \mathbb{R}^n \,|\, \langle \mathbf{u}, \mathbf{s} \rangle = 0 \text{ for every } \mathbf{s} \in S\}.$$

Let $\mathcal{U}$ be a plane in $\mathbb{R}^3$ passing through the zero vector, a 2-dimensional subspace of $\mathbb{R}^3$. If we select a pair of nonzero, non-parallel vectors $\mathbf{u}_1$, $\mathbf{u}_2 \in \mathcal{U}$, we know from MATH 162 that $\mathbf{w} = \mathbf{u}_1 \times \mathbf{u}_2$ is normal to $\mathcal{U}$. The set of vectors which are orthogonal to all vectors in $\mathcal{U}$ is given by

$$\mathcal{U}^\perp = \{t\mathbf{w} \,|\, t \in \mathbb{R}\} ,$$

a one-dimensional subspace of $\mathbb{R}^3$. In fact, it is also the case that, for this same choice of $\mathbf{u}_1$, $\mathbf{u}_2 \in \mathcal{U}$,

$$\{\mathbf{u}_1, \mathbf{u}_2\}^\perp = \{t\mathbf{w} \,|\, t \in \mathbb{R}\} ,$$

which is a special case of the following theorem.

Theorem 10: Let S be a subset of $\mathbb{R}^n$. For each set A containing S as a subset, and itself contained in span(S) (i.e., $S \subset A \subset \text{span}(S)$), $A^\perp = S^\perp$.

Example 28:

Let $\mathbf{A}$ be an m-by-n matrix, and let $\mathbf{v} \in \text{null}(\mathbf{A})$ so that $\mathbf{A}\mathbf{v} = \mathbf{0}$. Thus, for each row vector $\mathbf{r}_i$ of $\mathbf{A}$, $\mathbf{v} \perp \mathbf{r}_i^T$—that is, $\mathbf{v} \in \{\mathbf{r}_1^T, \mathbf{r}_2^T, \ldots, \mathbf{r}_m^T\}^\perp$. But by Theorem 10, this means that $\mathbf{v} \in \text{span}(\{\mathbf{r}_1^T, \mathbf{r}_2^T, \ldots, \mathbf{r}_m^T\})^\perp = \text{col}(\mathbf{A}^T)^\perp$. So, $\text{null}(\mathbf{A}) \subset \text{col}(\mathbf{A}^T)^\perp$.

Moreover, if $\mathbf{v} \in \text{col}(\mathbf{A}^T)^\perp$, then $\mathbf{v}$ is perpendicular to each column of $\mathbf{A}^T$, which allows us to retrace our steps and conclude that $\mathbf{v} \in \text{null}(\mathbf{A})$. Thus, $\text{col}(\mathbf{A}^T)^\perp \subset \text{null}(\mathbf{A})$ as well, giving us that the two sets $\text{col}(\mathbf{A}^T)^\perp$ and $\text{null}(\mathbf{A})$ are, in fact, equal. ■

The content of this example is very important. We state it as part (i) in the next theorem. Part (ii) follows from (i) in that every matrix is the transpose of another matrix.

Theorem 11: Let $\mathbf{A}$ be a matrix with real number entries. We have

(i) $\text{null}(\mathbf{A}) = \text{col}(\mathbf{A}^T)^\perp$.

(ii) $\text{null}\left(\mathbf{A}^T\right) = \text{col}(\mathbf{A})^\perp$.

We have now seen several examples of orthogonal complements. The orthogonal complement of two nonzero, non-parallel vectors $\mathbf{u}_1$, $\mathbf{u}_2$ in $\mathbb{R}^3$ is $\text{span}(\{\mathbf{u}_1 \times \mathbf{u}_2\})$, a subspace of $\mathbb{R}^3$. For an arbitrary m-by-n matrix $\mathbf{A}$ we know that $\text{col}(\mathbf{A}^T)$ is a subset (subspace, in fact) of $\mathbb{R}^n$ and, by the previous example, $\text{col}(\mathbf{A}^T)^\perp = \text{null}(\mathbf{A})$, also a subspace of $\mathbb{R}^n$. These two examples might lead us to suspect the truth of the following theorem.

Theorem 12: For any subset S of $\mathbb{R}^n$, $S^\perp$ is a subspace of $\mathbb{R}^n$.

Proof: We will use the Subspace Test (Theorem 7). Let $\mathbf{u}, \mathbf{v} \in S^\perp$, and a, b be real numbers (scalars). For any vector $\mathbf{s} \in S$, we know $\mathbf{u} \perp \mathbf{s}$ and $\mathbf{v} \perp \mathbf{s}$, and so

$$\begin{aligned} \langle \mathbf{s}, a\mathbf{u} + b\mathbf{v} \rangle &= a \langle \mathbf{s}, \mathbf{u} \rangle + b \langle \mathbf{s}, \mathbf{v} \rangle \qquad \text{(this is equation (3.2)} \\ &= a \cdot 0 + b \cdot 0 \qquad (\text{since } \mathbf{s} \in S,\ \mathbf{u}, \mathbf{v} \in S^\perp) \\ &= 0\,. \end{aligned}$$

Thus, an arbitrary linear combination $a\mathbf{u} + b\mathbf{v}$ of $\mathbf{u}$ and $\mathbf{v}$ is still orthogonal to $\mathbf{s}$. The choice of $\mathbf{s} \in S$ was arbitrary, so $a\mathbf{u} + b\mathbf{v}$ is orthogonal to *every* vector in S, which is just the requirement for $a\mathbf{u} + b\mathbf{v}$ to be in $S^\perp$. Thus, $S^\perp$ is closed under linear combinations. □

3.1.3 Inner product spaces

The inner product (3.1) provides a mapping whose inputs are two vectors from $\mathbb{R}^n$ and whose output is a real number. This inner product has some nice properties, including the *linearity in the second argument* noted in (3.2). There are other ways besides (3.1) to define mappings from $\mathbb{R}^n \times \mathbb{R}^n$ (or $\mathcal{V} \times \mathcal{V}$, where $\mathcal{V}$ is any vector space) into $\mathbb{R}$ that possess the same "nice" properties. The properties we require are as follows:

Definition 15: Let $\mathcal{V}$ be a vector space over the reals. An **inner product on $\mathcal{V}$** must satisfy the following properties:

(i) $\langle u, v\rangle$ is a real number for all vectors $u, v \in \mathcal{V}$.

(ii) $\langle v, v\rangle \geq 0$ for all $v \in \mathcal{V}$, with equality if and only if v is the zero vector of $\mathcal{V}$.

(iii) $\langle u, v\rangle = \langle v, u\rangle$ for all vectors $u, v \in \mathcal{V}$.

(iv) $\langle au, v\rangle = a\langle u, v\rangle$ for all vectors $u, v \in \mathcal{V}$ and all real numbers a.

(v) $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all vectors $u, v, w \in \mathcal{V}$.

It is not difficult to verify that the inner product 3.1 satisfies all of these properties. Here is another example of an inner product, this time on the vector space $\mathscr{C}([a, b])$.

Example 29:

In $\mathscr{C}([a, b])$ (where the 'vectors' are continuous functions on the closed interval $[a, b]$) it is common to define the inner product between $f, g \in \mathscr{C}([a, b])$ as

$$\langle f, g\rangle := \int_a^b f(x)g(x)\,dx\,. \tag{3.4}$$

Thus, if $f(x) = x^2$ and $g(x) = \sin x$ are considered as functions in $\mathscr{C}([-\pi, \pi])$, then they are orthogonal under this inner product, since

$$\langle f, g\rangle = \int_{-\pi}^{\pi} x^2 \sin x\,dx = 0.$$

(This integral equals zero because the product of an even and odd function is odd.)

The inner product (3.4) on $\mathscr{C}([a, b])$ is instrumental in the theory of Fourier series, being used in the computation of Fourier coefficients.

■

Whenever we have a valid inner product on a vector space $\mathcal{V}$, we can emphasize this fact by calling $\mathcal{V}$ an **inner product space**. If $\mathcal{V}$ is an inner product space, we again use the notation $\langle u, v\rangle$ to indicate the inner product of two vectors $u, v \in \mathcal{V}$.

Because of property (ii) in Definition 15, we can once again define the notion of length and distance in an inner product space. The usual way to do this is through the **norm derived from the inner product**. That is, for each v in an inner product space $\mathcal{V}$, we may define the (derived) norm of v to be

$$\|v\| := \sqrt{\langle v, v\rangle}\,. \tag{3.5}$$

By **length** of v, we mean precisely its norm. The **distance between two vectors** u and v is once again the norm of their difference, or $\|u - v\|$. Orthogonality, as well, is a concept associated with *any* inner product as we saw in Example 29.

3.2 The Fundamental Subspaces

3.2.1 Direct Sums

Recall from the last section that, if we fix a plane $\mathcal{U}$ through the origin (a 2-dimensional subspace) in $\mathbb{R}^3$, then $\mathcal{U}^\perp$ is a line of vectors through the origin all parallel to a vector normal to the plane. It is perhaps believable that if you had some destination vector $\mathbf{v} \in \mathbb{R}^3$ to reach, and if you had to resolve the trip to $\mathbf{v}$ into two steps $\mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in \mathcal{U}$ and $\mathbf{w} \in \mathcal{U}^\perp$, that there would be only one choice for $\mathbf{u}$ and one choice for $\mathbf{w}$. This is the situation we wish to investigate in this section.

Definition 16: Suppose $\mathcal{U}$ and $\mathcal{W}$ are subspaces of a vector space $\mathcal{V}$. Suppose also that each $v \in \mathcal{V}$ can be written *uniquely* as a sum $u + w$, with $u \in \mathcal{U}$ and $w \in \mathcal{W}$. Then we say that $\mathcal{V}$ is a **direct sum** of $\mathcal{U}$ and $\mathcal{W}$, denoting this by writing $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$.

From the motivating example that precedes this definition, one might guess the following result.

Theorem 13: If $\mathcal{U}$ is a subspace of $\mathbb{R}^n$, then $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$.

The rest of Subsection 3.2.1 is devoted to proving this theorem and, for those readers in a rush to get to the point, may be skipped for now.

Before we do so, we state and prove the following intermediate result.

Lemma 1: If $\mathcal{U}$ is a subspace of $\mathbb{R}^n$, then $\dim(\mathcal{U}) + \dim(\mathcal{U}^\perp) = n$. Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is a basis for $\mathcal{U}$ and $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ is a basis for $\mathcal{U}^\perp$, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

Proof: If $\mathcal{U}$ is the trivial subspace $\{\mathbf{0}\}$ of $\mathbb{R}^n$, then $\mathcal{U}^\perp = \mathbb{R}^n$ (see Exercise 3.4), and all of the claims of this lemma are obviously true. Let us assume, then, that $\dim(\mathcal{U}) = r > 0$, so that a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ of $\mathcal{U}$ is nonempty. Let $\mathbf{A}$ be an r-by-n matrix whose i^{th} row is $\mathbf{u}_i^T$, for $1 \leq i \leq r$. Note that the row space of $\mathbf{A}$ is all of $\mathcal{U}$, and that $\text{rank}(\mathbf{A}) = r$. By Theorem 11,

$$\mathcal{U}^\perp = \text{col}(\mathbf{A}^T)^\perp = \text{null}(\mathbf{A}) ,$$

so

$$\dim(\mathcal{U}^\perp) = \text{nullity}(\mathbf{A}) = n - r ,$$

proving the first claim of the lemma.

To prove the remaining claim, we must simply show that the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ is linearly independent. So, let $c_1, \ldots, c_n$ be real numbers such that

$$c_1\mathbf{u}_1 + \ldots + c_r\mathbf{u}_r + c_{r+1}\mathbf{u}_{r+1} + \ldots + c_n\mathbf{u}_n = \mathbf{0} .$$

Let

$$\mathbf{u} := c_1\mathbf{u}_1 + \ldots + c_r\mathbf{u}_r \qquad \text{and} \qquad \mathbf{w} := c_{r+1}\mathbf{u}_{r+1} + \ldots + c_n\mathbf{u}_n ,$$

so that $\mathbf{u} + \mathbf{w} = \mathbf{0}$, or $\mathbf{u} = -\mathbf{w}$. Since $\mathcal{U}^\perp$ is a subspace of $\mathbb{R}^n$ containing $\mathbf{w}$, it follows that $\mathbf{u} \in \mathcal{U}^\perp$. By Exercise 3.12, $\mathbf{u} \in \mathcal{U} \cap \mathcal{U}^\perp$ implies that $\mathbf{u} = \mathbf{0}$. Thus,

$$c_1\mathbf{u}_1 + \ldots + c_r\mathbf{u}_r = \mathbf{0}$$

and, by the linear independence of $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, we have $c_1 = \cdots = c_r = 0$. Similarly, $\mathbf{w} \in \mathcal{U} \cap \mathcal{U}^\perp$, which leads to

$$c_{r+1}\mathbf{u}_{r+1} + \ldots + c_n\mathbf{u}_n = \mathbf{0} ,$$

and hence $c_{r+1} = \cdots = c_n = 0$. □

Proof: [Proof of Theorem 13] We must show that, given any $\mathbf{v} \in \mathbb{R}^n$, there are vectors $\mathbf{u} \in \mathcal{U}$ and $\mathbf{w} \in \mathcal{U}^\perp$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$ and, moreover, there is no second pair of vectors $\tilde{\mathbf{u}} \in \mathcal{U}$ and $\tilde{\mathbf{w}} \in \mathcal{U}^\perp$ that somehow deviates from $\mathbf{u}$ and $\mathbf{w}$ (i.e., either $\tilde{\mathbf{u}} \neq \mathbf{u}$, $\tilde{\mathbf{w}} \neq \mathbf{w}$, or both) for which $\mathbf{v} = \tilde{\mathbf{u}} + \tilde{\mathbf{w}}$. Note that, should it be the case that $\mathcal{U} = \{\mathbf{0}\}$ or $\mathcal{U} = \mathbb{R}^n$, then these claims are quite straightforward. Thus, we focus on the situation where $\mathcal{U}$ is neither all of $\mathbb{R}^n$ nor the trivial subspace, but has dimension r with $0 < r < n$.

Let $\mathbf{v} \in \mathbb{R}^n$. By Lemma 1 there is a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots \mathbf{u}_n\}$ of $\mathbb{R}^n$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is a basis for $\mathcal{U}$ and $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$ is a basis for $\mathcal{U}^\perp$. Any basis of $\mathbb{R}^n$ spans $\mathbb{R}^n$, so there exist scalars $c_1, \ldots, c_n$ such that

$$\mathbf{v} \;=\; c_1\mathbf{u}_1 + \ldots + c_r\mathbf{u}_r + c_{r+1}\mathbf{u}_{r+1} + c_n\mathbf{u}_n \;.$$

Now take $\mathbf{u} := c_1\mathbf{u}_1 + \ldots + c_r\mathbf{u}_r$ and $\mathbf{w} := c_{r+1}\mathbf{u}_{r+1} + c_n\mathbf{u}_n$ to see that there are vectors $\mathbf{u} \in \mathcal{U}$, $\mathbf{w} \in \mathcal{U}^\perp$ for which $\mathbf{v} = \mathbf{u} + \mathbf{w}$.

It remains to show that such a choice of $\mathbf{u}$, $\mathbf{w}$ is unique. So, suppose there are $\tilde{\mathbf{u}} \in \mathcal{U}$ and $\tilde{\mathbf{w}} \in \mathcal{U}^\perp$ with $\mathbf{v} = \tilde{\mathbf{u}} + \tilde{\mathbf{w}}$. Then

$$\mathbf{u} + \mathbf{w} \;=\; \tilde{\mathbf{u}} + \tilde{\mathbf{w}}\,, \qquad \text{or} \qquad \mathbf{u} - \tilde{\mathbf{u}} = \tilde{\mathbf{w}} - \mathbf{w}\,.$$

This shows $\mathbf{u} - \tilde{\mathbf{u}} \in \mathcal{U} \cap \mathcal{U}^\perp = \{\mathbf{0}\}$ and, likewise, $\tilde{\mathbf{w}} - \mathbf{w} \in \mathcal{U} \cap \mathcal{U}^\perp$. Thus, $\mathbf{u} = \tilde{\mathbf{u}}$ and $\mathbf{w} = \tilde{\mathbf{w}}$. □

3.2.2 Fundamental subspaces, the normal equations, and least-squares solutions

Recall that if we start with a plane $\mathcal{U}$ through the origin in $\mathbb{R}^3$, then $\mathcal{U}^\perp$ is a line through the origin. We could very well have begun with the line through the origin, calling this $\mathcal{W}$, and found that $\mathcal{W}^\perp$ was the corresponding plane. This might lead you to suspect that, in general, $(\mathcal{U}^\perp)^\perp = \mathcal{U}$. The next theorem shows this is almost correct.

Theorem 14: Let S be a subset of $\mathbb{R}^n$. Then $(S^\perp)^\perp = \text{span}(S)$. In particular, if S is a *subspace* of $\mathbb{R}^n$, then $(S^\perp)^\perp = S$.

A corollary to this theorem, then, is that the relationship in Theorem 11 is reciprocated. That is, for any matrix $\mathbf{A}$, along with the relationships expressed in Theorem 11 we also have

$$\text{null}\,(\mathbf{A})^\perp = \text{col}(\mathbf{A}^T) \qquad \text{and} \qquad \text{null}\left(\mathbf{A}^T\right)^\perp = \text{col}(\mathbf{A})\,. \tag{3.6}$$

If $\mathbf{A}$ is m-by-n, then we know the function $(\mathbf{x} \mapsto \mathbf{A}\mathbf{x})\colon \mathbb{R}^n \to \mathbb{R}^m$.

Because of the relationships (3.6) and Theorem 13, we know

$$\mathbb{R}^n = \operatorname{col}(\mathbf{A}^T) \oplus \operatorname{null}(\mathbf{A}) \,, \tag{3.7}$$

$$\mathbb{R}^m = \operatorname{col}(\mathbf{A}) \oplus \operatorname{null}\left(\mathbf{A}^T\right) . \tag{3.8}$$

The four vector spaces $\operatorname{col}(\mathbf{A}^T)$ (the row space), $\operatorname{null}(\mathbf{A})$ (the nullspace), $\operatorname{col}(\mathbf{A})$ (the column space), and $\operatorname{col}(\mathbf{A}^T)$ (the left nullspace) are, as a group, referred to as the **fundamental subspaces of A**. The direct sums in (3.7)–(3.8) show how these fundamental subspaces provide an important decomposition to the domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$ of the mapping $(\mathbf{x} \mapsto \mathbf{A}\mathbf{x})$.

These decompositions gives rise to the situation depicted in Figure 3.1. By (3.7), every $\mathbf{x} \in \mathbb{R}^n$ is uniquely written as a sum

$$\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n \,,$$

where $\mathbf{x}_r \in \operatorname{col}(\mathbf{A}^T)$ and $\mathbf{x}_n \in \operatorname{null}(\mathbf{A})$. If $\mathbf{b} \in \operatorname{col}(\mathbf{A})$, then $\mathbf{b} = \mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. But then

$$\mathbf{b} = \mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{x}_r + \mathbf{x}_n) = \mathbf{A}\mathbf{x}_r + \mathbf{A}\mathbf{x}_n = \mathbf{A}\mathbf{x}_r \,, \tag{3.9}$$

which shows that it is the part of $\mathbf{x}$ lying in $\operatorname{col}(\mathbf{A}^T)$ that entirely accounts for where the image $\mathbf{b} = \mathbf{A}\mathbf{x}$ of $\mathbf{x}$ lies in $\mathbb{R}^m$.

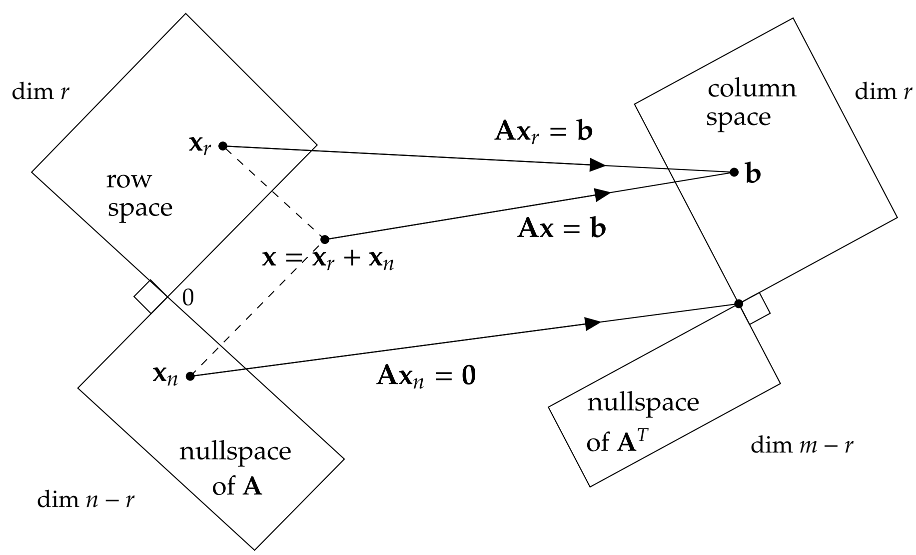


Figure 3.1: The Fundamental Subspaces of **A**

Now, suppose $\mathbf{b} \notin \operatorname{col}(\mathbf{A})$, so that the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ is **inconsistent**, having no solution. On many occasions we still want some kind of *approximate solution*. By (3.8), $\mathbf{b}$ is uniquely written as a sum

$$\mathbf{b} = \mathbf{p} + \mathbf{e} \,,$$

where $\mathbf{p} \in \text{col}(\mathbf{A})$ and $\mathbf{e} \in \text{null}\left(\mathbf{A}^T\right)$, the situation depicted in Figure 3.2. Since $\mathbf{p} \in \text{col}(\mathbf{A})$, there is some element $\bar{\mathbf{x}} \in \text{col}(\mathbf{A}^T)$ for which $\mathbf{A}\bar{\mathbf{x}} = \mathbf{p}$. It is this $\bar{\mathbf{x}}$ which we choose to take as our approximate solution because, over all elements $\mathbf{Ax}$ in the column space, $\mathbf{p} = \mathbf{A}\bar{\mathbf{x}}$ is the closest to $\mathbf{b}$. That is, the quantity

$$\|\mathbf{b} - \mathbf{Ax}\| \qquad \text{or, equivalently,} \qquad \|\mathbf{b} - \mathbf{Ax}\|^2 ,$$

is minimized when $\mathbf{x} = \bar{\mathbf{x}}$. (Note, however, that when nullity $(\mathbf{A}) > 0$, equation (3.9) shows there are many $\mathbf{x} \in \mathbb{R}^n$ that minimize this quantity; each has the same component $\bar{\mathbf{x}}$ in $\text{col}(\mathbf{A}^T)$.) The vector

$$\mathbf{r}(\mathbf{x}) \; := \; \mathbf{b} - \mathbf{Ax} \tag{3.10}$$

is called the **residual**. Any $\mathbf{x} \in \mathbb{R}^n$ that minimizes the length (or length-squared) of the residual $\mathbf{r}(\mathbf{x})$ is called a **least-squares solution of $\mathbf{Ax} = \mathbf{b}$**.

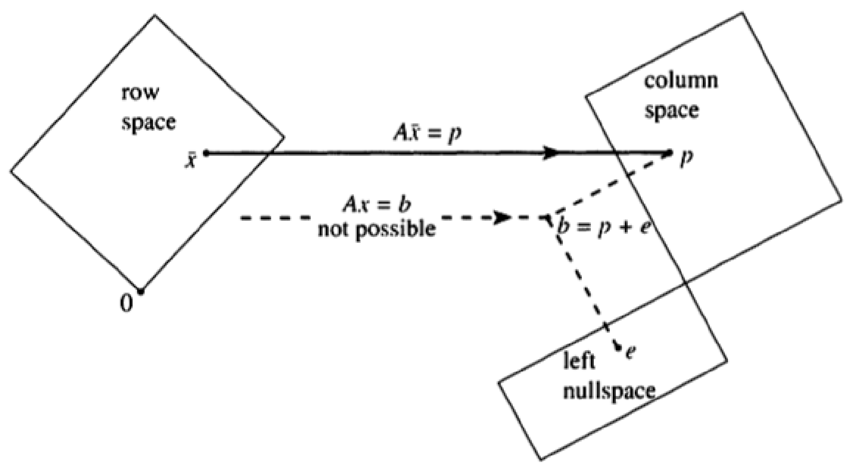


Figure 3.2: The action of $\mathbf{A}$ (isolated to its row space) into $\mathbb{R}^m$

Note that when $\mathbf{x}$ is a least-squares solution, then $\mathbf{r}(\mathbf{x}) = \mathbf{e}$, the part of $\mathbf{b}$ that lies in the left nullspace of $\mathbf{A}$ (see Figure 3.2). This is key to finding least-squares solutions for, assuming that $\mathbf{x}$ is a least-squares solution, we have

$$\mathbf{A}^T\mathbf{e} \; = \; \mathbf{A}^T(\mathbf{b} - \mathbf{Ax}) \; = \; \mathbf{0} ,$$

or

$$(\mathbf{A}^T\mathbf{A})\mathbf{x} \; = \; \mathbf{A}^T\mathbf{b} . \tag{3.11}$$

The name given to (3.11) is the **normal equations**. Any solution of them is a least-squares solution of $\mathbf{Ax} = \mathbf{b}$. It can be shown that (3.11) always has a solution (see Exercise 3.15), so the methods of Chapter 1 may be used to find it/them. Moreover, when the columns of $\mathbf{A}$ are linearly independent (i.e., $\text{rank}(\mathbf{A}) = n$, or $\text{null}(\mathbf{A}) = \{\mathbf{0}\}$), then

$$\begin{aligned}(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{0} \quad &\Rightarrow \quad \|\mathbf{A}\mathbf{x}\|^2 = 0 \qquad \text{(see Exercise 3.14)} \\ &\Rightarrow \quad \mathbf{A}\mathbf{x} = \mathbf{0} \\ &\Rightarrow \quad \mathbf{x} = \mathbf{0}\,.\end{aligned}$$

That is, when $\mathbf{A}$ has linearly independent columns, then $\text{null}\left(\mathbf{A}^T\mathbf{A}\right) = \{\mathbf{0}\}$ and $\mathbf{A}^T\mathbf{A}$, being square, is nonsingular. Thus, $\text{nullity}\,(\mathbf{A}) = 0$ means that the normal equations (3.11) have precisely one solution, which must be $\bar{\mathbf{x}}$.

Finally, we mention that, in the case where $\mathbf{b} \in \text{col}(\mathbf{A})$ (so that $\mathbf{e} = \mathbf{0}$, and the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent), least-squares solutions coincide with solutions in the usual sense. This is because such solutions make the residual equal to zero in that case.

Exercises

3.1 Show that the inner product (3.1) on $\mathbb{R}^n$ indeed has the property asserted in (3.2). This property is called **linearity in the 2nd argument**.

3.2

a) Observe that, for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$. Use this and equation (3.2) to show that, given any real numbers a and b, and any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$,

$$\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle \;=\; a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$$

(called **linearity in the 1st argument**).

b) Show that, for any pair of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and any real numbers a, b,

$$\langle a\mathbf{u} + b\mathbf{v}, a\mathbf{u} + b\mathbf{v} \rangle \;=\; a^2\|\mathbf{u}\|^2 + 2ab\langle \mathbf{u}, \mathbf{v} \rangle + b^2\|\mathbf{v}\|^2 \,.$$

c) Fill in the missing details of the proof of the Pythagorean Theorem (Theorem 9).

3.3 Show that the zero vector $\mathbf{0} \in \mathbb{R}^n$ is the only vector in $\mathbb{R}^n$ which is orthogonal to itself.

3.4

a) Suppose $\mathcal{U}$ is the trivial subspace of $\mathbb{R}^n$, the one consisting of just the zero vector. Show that $\mathcal{U}^\perp = \mathbb{R}^n$.

b) Show that for $\mathcal{U} = \mathbb{R}^n$, the orthogonal complement is the trivial subspace of $\mathbb{R}^n$.

3.5 The fundamental subspaces associated with an arbitrary m-by-n matrix $\mathbf{A}$ are the row space $\text{col}(\mathbf{A}^T)$, the column space $\text{col}(\mathbf{A})$, the nullspace $\text{null}(\mathbf{A})$, and the left nullspace $\text{null}\left(\mathbf{A}^T\right)$. For each $\mathbf{A}$ given below, determine a basis for each of these fundamental subspaces.

a) $\mathbf{A} \;=\; \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}$

b) $\mathbf{A} \;=\; \begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 0 \end{bmatrix}$

c) $\mathbf{A} \;=\; \begin{bmatrix} 4 & -2 \\ 1 & 3 \\ 2 & 1 \\ 3 & 4 \end{bmatrix}$

d) $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 2 & 2 \end{bmatrix}$

In light of Theorem 11, what orthogonality relationships should you see between the various basis vectors? Do you?

3.6

a) Let $\mathcal{U}$ be the subspace of $\mathbb{R}^3$ spanned by $(1, -1, 1)$. Find a basis for $\mathcal{U}^\perp$. (Hint: Exploit the fact that $\text{col}(\mathbf{A}^T)^\perp = \text{null}\,(\mathbf{A})$.)

b) If $\mathcal{V} = \text{span}(\{(1, 2, 1), (1, -1, 2)\})$, then find $\mathcal{V}^\perp$.

3.7 Is it possible for a matrix to have $(1, 5, -1)$ in its row space and $(2, 1, 6)$ in its nullspace? Explain.

3.8 Let $\mathbf{A}$ be an m-by-n matrix with $\text{rank}(\mathbf{A}) = r$. What are the dimensions of $\text{null}\,(\mathbf{A})$ and $\text{null}\left(\mathbf{A}^T\right)$? Explain.

3.9 Suppose $\mathbf{A}$ is an m-by-n matrix. Show that

a) if $\mathbf{x} \in \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$, then $\mathbf{A}\mathbf{x}$ is in both $\text{col}(\mathbf{A})$ and $\text{null}\left(\mathbf{A}^T\right)$.

b) $\text{null}\left(\mathbf{A}^T\mathbf{A}\right) = \text{null}\,(\mathbf{A})$. (Hint: $\text{col}(\mathbf{A}) \cap \text{null}\left(\mathbf{A}^T\right) = \{\mathbf{0}\}$, a result proved in part a) of Exercise 3.12.)

c) $\mathbf{A}$ and $\mathbf{A}^T\mathbf{A}$ have the same rank.

d) if $\mathbf{A}$ has (all) linearly independent columns, then $\mathbf{A}^T\mathbf{A}$ is nonsingular.

3.10 Using only the fact that $\mathcal{V}$ is an inner product space (whose inner product satisfies Definition 15), prove that the inner product of any vector $\mathbf{v} \in \mathcal{V}$ with the zero vector of $\mathcal{V}$ is zero.

3.11 Suppose $\mathcal{V}$ is an inner product space. Show that, under the derived norm (3.5), it is impossible that $\|\mathbf{v}\| = 0$ for $\mathbf{v} \in \mathcal{V}$ unless $\mathbf{v} = \mathbf{0}$ (the zero vector).

3.12

a) Suppose $\mathcal{U}$ is a subspace of $\mathbb{R}^n$. Show that $\mathcal{U} \cap \mathcal{U}^\perp = \{\mathbf{0}\}$. Here the symbol "$\cap$" denotes the "intersection"—that is, given two sets A and B, $A \cap B$ stands for the set of elements that are in both A and B.

b) Suppose U is only a subset of $\mathbb{R}^n$. What may be said about $U \cap U^\perp$ in this case? Give an example that helps illustrate your answer.

3.13 Use Theorems 11 and 14 to show that

$$\text{null}\,(\mathbf{A})^\perp = \text{col}(\mathbf{A}^T) \qquad \text{and} \qquad \text{null}\left(\mathbf{A}^T\right)^\perp = \text{col}(\mathbf{A})\,.$$

3.14 In the string of implications following equation (3.11), we have the statement that $(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{0}$ implies $\|\mathbf{A}\mathbf{x}\|^2 = 0$. Show this. (Hint: Recall that an inner product between two vectors in $\mathbb{R}^m$ has an equivalent formulation as a matrix product; see the first bullet following Example 27.)

3.15 In this exercise, we show that equation (3.11) is consistent. To do so, we show $\mathbf{A}^T\mathbf{b} \in \text{col}(\mathbf{A}^T\mathbf{A})$, doing so via the following "steps":

a) Show that $\mathbb{R}^n = \text{col}(\mathbf{A}^T\mathbf{A}) \oplus \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$.

b) Use (3.8) and the result of part b), Exercise 3.9 to deduce that $\text{col}(\mathbf{A}^T) = \text{col}(\mathbf{A}^T\mathbf{A})$.

c) Explain why the equality of spaces deduced in the previous part implies a solution to (3.11) exists.

3.16 Let $\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 4 & -2 \\ -2 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 0 \\ -2 \\ 2 \end{bmatrix}$.

a) Write out the normal equations associated with the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. That is, write out just what the matrix $\mathbf{A}^T\mathbf{A}$ and right-hand side vector $\mathbf{A}^T\mathbf{b}$ look like.

b) Find all least-squares solutions of $\mathbf{A}\mathbf{x} = \mathbf{b}$ (i.e., all solutions of the normal equations). These solutions form a subset of $\mathbb{R}^2$; plot this subset on the coordinate plane.

c) What is the nullspace of $\mathbf{A}$? Plot this on the same coordinate plane with your plot from part (b).

d) What is the row space of $\mathbf{A}$? Plot this on the same coordinate plane with your plot from part (b). Find the vector which is both in the row space of $\mathbf{A}$ and is a least-squares solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$.

e) What is the column space of $\mathbf{A}$? On a coordinate grid (necessarily different from the one you have been using), sketch both this column space and the vector $\mathbf{b}$. What vector $\mathbf{p} \in \text{col}(\mathbf{A})$ is closest to $\mathbf{b}$?

3.17 Let $\mathbf{A} = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 6 & 2 \\ 1 & 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 4 \\ 6 \\ 3 \end{bmatrix}$.

a) Show that, for this $\mathbf{A}$ and $\mathbf{b}$, the matrix problem $\mathbf{A}\mathbf{x} = \mathbf{b}$ is inconsistent.

b) Solve (i.e., find *all* solutions) of the normal equations $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$. You are welcome to use Octave along the way—for instance, in reducing an augmented matrix to echelon form.

c) Employ the "backslash" command in Octave to solve $\mathbf{A}\mathbf{x} = \mathbf{b}$. That is, enter the given matrices into variables named `A` and `b`, and then run the command `A \ b`. How does this "answer" relate to your answers from parts (a) and (b)?

3.18 In class we discussed the particulars of fitting a polynomial of degree d to n prescribed data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. In this problem, we will investigate the case where $d = 1$ (i.e., fitting a line); in particular, our polynomial is $p(x) = a_0 + a_1 x$.

a) Write out the form of $\mathbf{A}$, $\mathbf{b}$ (what we did in class for general d) so that least-squares solutions of

$$\mathbf{A}\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \mathbf{b}$$

yield the "best-fit" line.

b) Describe conditions under which the rank of your matrix $\mathbf{A}$ (from part (a)) would be less than 2?

c) One website lists the following numbers (perhaps they are fictitious) for different kinds of sandwiches offered at a restaurant:

Sandwich	Total Fat (grams)	Total Calories
Hamburger	9	260
Cheeseburger	13	320
Quarter Pounder	21	420
Quarter Pounder w/ Cheese	30	530
Big Mac	31	560
Arch Sandwich Special	31	550
Arch Special w/ Bacon	34	590
Crispy Chicken	25	500
Fish Fillet	28	560
Grilled Chicken	20	440
Grilled Chicken Light	5	300

Treating the "Total Fat" variable as the x-coordinate, determine the least-squares best-fit line to these data points.

d) For a given n-by-p matrix $\mathbf{A}$, $\mathbf{b} \in \mathbb{R}^n$, a least squares solution $\mathbf{x} \in \mathbb{R}^p$ of the matrix equation $\mathbf{Ax} = \mathbf{b}$ minimizes the quantity $\|\mathbf{b} - \mathbf{Ax}\|^2$. In the case $p = 2$, if we write $\mathbf{x} = (a_0, a_1)$, then

$$\|\mathbf{b} - \mathbf{Ax}\|^2 = \left\| \begin{bmatrix} y_1 - (a_0 + a_1 x_1) \\ y_2 - (a_0 + a_1 x_2) \\ \vdots \\ y_n - (a_0 + a_1 x_n) \end{bmatrix} \right\|^2 = \sum_{j=1}^n (y_j - a_0 - a_1 x_j)^2 .$$

Let us define $f(a_0, a_1) = \sum_{j=1}^n (y_j - a_0 - a_1 x_j)^2$, a function of two variables. (Note that the *variables* in f are a_0 and a_1; the x_js and y_js are given numbers.) Recall, from MATH 162, that a critical point of a such a function (a location at which f might reach a local extremum) is a point (a_0^*, a_1^*) which simultaneously satisfies the system of equations

$$\frac{\partial f}{\partial a_0} = 0 , \qquad \frac{\partial f}{\partial a_1} = 0 .$$

Show that there is a solution (a_0^*, a_1^*) of this system of equations, and that its values are

$$a_1^* = \frac{n \sum_{j=1}^n x_j y_j - \left(\sum_{j=1}^n x_j\right)\left(\sum_{j=1}^n y_j\right)}{n \sum_{j=1}^n x_j^2 - \left(\sum_{j=1}^n x_j\right)^2} \quad \text{and} \quad a_0^* = \frac{1}{n}\left(\sum_{j=1}^n y_j - a_1 \sum_{j=1}^n x_j\right).$$

e) Use the formulas from part (d) to see that they yield the same "best fit" line for the data in part(c) as you obtained solving the normal equations.

3.19 Write down an m-by-n matrix $\mathbf{A}$ for which

a) the least-squares solutions of $\mathbf{Ax} = \mathbf{b}$ are always (no matter the choice of $\mathbf{b} \in \mathbb{R}^m$) the same as the actual solutions of $\mathbf{Ax} = \mathbf{b}$. (In particular, there are no vectors $\mathbf{b} \in \mathbb{R}^m$ for which $\mathbf{Ax} = \mathbf{b}$ is inconsistent.)

b) there are vectors $\mathbf{b} \in \mathbb{R}^m$ for which $\mathbf{Ax} = \mathbf{b}$ is inconsistent, but for all $\mathbf{b} \in \mathbb{R}^m$, the normal equations have a unique solution.

c) there are vectors $\mathbf{b} \in \mathbb{R}^m$ for which $\mathbf{Ax} = \mathbf{b}$ is inconsistent, but for all $\mathbf{b} \in \mathbb{R}^m$, the normal equations have infinitely many solutions.

4 Detailed Solutions to Exercises

1.1

a) `diag([3 5], 2)`

b) `diag([2 7 1], -2) + diag([-1 2 1 -4], 1)`

c) `[ones(3, 2); [3 -2]]`

1.2

a) It must be the case that **B** is 3-by-p, for some positive integer p.

b) It must be the case that **B** is ℓ-by-5, for some positive integer ℓ.

1.3

a) Let us suppose **A** is m-by-n. For **AB** to be defined, it must be the case that **B** is n-by-p (i.e., the number of rows of **B** equals the number of rows of **A**). But, for **BA** to make sense, we must have $p = n$, which means, just on the grounds that **AB** and **BA** make sense, we now have that **B** is n-by-m. In that case, **AB** is an m-by-m matrix, while **BA** is an n-by-n matrix. Now, if we add the constraint that **AB** = **BA**, we get that $n = m$—that is, **A** must be a square matrix.

b) The Octave code I give below is not technically an Octave *function*, in that there is no return value. I have instead just placed these commands in an `.m` file and executed it.

```
numEqual = 0;

for k = 1:20
  A = rand(3);
  B = rand(3);
  if (sum(sum(A*B ~= B*A)) == 0)
    numEqual = numEqual + 1;
  end
end

numEqual
```

Each time I run this code, it comes up with zero instances of **AB** equalling **BA**.

c) We are guaranteed to have $\mathbf{AB} = \mathbf{BA}$ whenever **B** is a power of **A**—that is, whenever $\mathbf{B} = \mathbf{A}^n$.

1.4 First, we have

$$(\mathbf{B}^{-1}\mathbf{A}^{-1})\mathbf{AB} = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{IB} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}\,.$$

Next,

$$\mathbf{AB}(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}\,.$$

Thus, $\mathbf{B}^{-1}\mathbf{A}^{-1}$ is the matrix which, when multiplied on either the left or right-hand side of **AB**, produces the identity matrix.

1.5 Perhaps the easiest place to do this is in Octave. The following code illustrates that it need not be true when **A**, **B** are 2-by-2 matrices.

```
octave> A = [1 3; 2 1];
octave> B = [2 7; 9 6];
octave> inv(A) + inv(B)
ans =
   -0.31765    0.73725
    0.57647   -0.23922

octave> inv(A+B)
ans =
   -0.078652    0.112360
    0.123596   -0.033708
```

1.6 The former problem asks you to show something is true without regard for the entries in the matrices involved. Thus, if you resort to giving these matrices specific entries and show that the inverse matrix of **AB** is $\mathbf{B}^{-1}\mathbf{A}^{-1}$, you will have shown the result for this pair of matrices only.

On the other hand, the latter problem gives a another universal-sounding proposition and asks you to show it is not true—i.e., does not hold universally. So you only have to come up with one example (with specific entries) that serves as a counterexample.

1.7

a) In transposition, columns are made into rows and vice versa. So, the number of columns of a matrix **A** becomes the number of rows in $\mathbf{A}^T$, and the number of rows of **A** becomes the number of columns of $\mathbf{A}^T$. A symmetric matrix is one for which $\mathbf{A} = \mathbf{A}^T$ and, in particular, the number of rows of **A** equals the number of rows of $\mathbf{A}^T$ (but that equals the number of columns of **A**).

b) Let $\mathbf{A}$ be an m-by-n matrix with $\mathbf{A} = (a_{ij})$. Then

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ a_{13} & a_{23} & \cdots & a_{m3} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix}.$$

By the definition of matrix equality, it is clear that $\mathbf{A} = \mathbf{A}^T$ implies each $a_{ij} = a_{ji}$, and vice versa.

1.8

a) To find out how many subscribers after 2 years, you could left-multiply the vector $(6000, 4000)$ by the same matrix as before. Alternatively, you could left-multiply $(8000, 2000)$ by $\mathbf{A}^2$, where `A = [0.7 0.2; 0.3 0.8]`. In the code below, I have chosen 25 years as long enough for the long-term behavior to appear.

```
octave> A = [0.7  0.2;  0.3  0.8];
octave> A^25*[6000;  4000]
ans =
    4000.0
    6000.0
```

So, after 25 years, there are 4000 subscribing households and 6000 non-subscribers. To see that this is, in fact, a steady state, you can left-multiply by $\mathbf{A}$ again, and the figure does not change.

b) Repeating the above with an initial vector of $(9000, 1000)$ still yields a long-term outlook of 4000 subscribing households. This is not surprising, since just one year gets the initial number of subscribing households from 9000 down to 6500.

1.9 My code looks something like the following, and is stored in a file named `singularCount.m`.

```
function numSing = singularCount(n)
  tol = 0.0000000001;
  numSing = 0;
  for i = 1:50
    A = randn(n);
    if (abs(det(A)) < tol)
      numSing = numSing + 1;
    end
  end
end
```

You might not have thought to put in a tolerance level, in which case your code may be more like this:

```
function numSing = singularCount(n)
  numSing = 0;
  for i = 1:50
    A = randn(n);
    if (det(A) == 0)
      numSing = numSing + 1;
    end
  end
end
```

But numerical calculations are rarely done perfectly, free from roundoff error. So, a matrix whose determinant is truly zero might still get computed to be some nonzero (but small in magnitude) number. The code I first presented might count a NS matrix among the singular ones, whereas the code as it appears the 2nd time might have a singular matrix which doesn't get classified as one.

Either way, the function is called with a command like `nsCount(4)`, in which case it returns the number of 4-by-4 matrices which were singular. I tried it with matrices of size 4-by-4, 5-by-5 and 10-by-10, and in no case did any of my randomly-generated matrices come up singular.

This leads one to believe that singular matrices, even among n-by-n matrices, are fairly rare.

1.10

a) 6

b) 2-by-2, 2-by-2, 4-by-3, and 4-by-2 respectively

c) Every possibility requires that **B** be partitioned with horizontal dividers appearing as in

$$\mathbf{B} = \left[\begin{array}{ccccc} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ \hline * & * & * & * & * \\ * & * & * & * & * \\ \hline * & * & * & * & * \\ * & * & * & * & * \end{array}\right].$$

It is permissible to use vertical dividers anywhere in **B** to create more submatrices, or to not have any at all. One possibility might be the above (which gives blocks $\mathbf{B}_1$,

$\mathbf{B}_2$, $\mathbf{B}_3$ running from top to bottom). Another is

$$\mathbf{B} \;=\; \left[\begin{array}{ccc|cc} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ \hline * & * & * & * & * \\ * & * & * & * & * \\ \hline * & * & * & * & * \\ * & * & * & * & * \end{array}\right] \;=\; \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \\ \hline \mathbf{B}_{31} & \mathbf{B}_{32} \end{array}\right].$$

The students can give the dimensions of the various blocks as I have above (via picture), or do so explicitly.

d) Just by way of examples, using the the first of my partitions of $\mathbf{B}$ above,

$$\mathbf{AB} \;=\; \left[\begin{array}{c} \mathbf{A}_{11}\mathbf{B}_1 + \mathbf{A}_{12}\mathbf{B}_2 + \mathbf{A}_{13}\mathbf{B}_3 \\ \hline \mathbf{A}_{21}\mathbf{B}_1 + \mathbf{A}_{22}\mathbf{B}_2 + \mathbf{A}_{23}\mathbf{B}_3 \end{array}\right].$$

Using the 2nd of my partitions of $\mathbf{B}$, we get

$$\mathbf{AB} \;=\; \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} + \mathbf{A}_{13}\mathbf{B}_{31} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} + \mathbf{A}_{13}\mathbf{B}_{32} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} + \mathbf{A}_{23}\mathbf{B}_{31} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} + \mathbf{A}_{23}\mathbf{B}_{32} \end{array}\right].$$

1.11

a) $\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$

b) For this one $\mathbf{P}$ should be the transpose of the $\mathbf{P}$ from part (a).

c) $\mathbf{B} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}.$

1.12 Method (i), which does not involve matrix multiplication, yields output (v_1+a, v_2+b) for input (v_1, v_2). Using approach (ii) and the knowledge that $\tilde{\mathbf{v}} = (v_1, v_2, 1)$, we have

$$\mathbf{A}\tilde{\mathbf{v}} \;=\; \left[\begin{array}{ccc} 1 & 0 & a \\ 0 & 1 & b \\ \hline 0 & 0 & 1 \end{array}\right]\begin{bmatrix} v_1 \\ v_2 \\ 1 \end{bmatrix} \;=\; \left[\begin{array}{c} v_1 + a \\ v_2 + b \\ \hline 1 \end{array}\right].$$

The upper block of the result is identical to the result from method (i).

1.13

a) We have

$$\begin{aligned}
\begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix} &= \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ \sin(\theta) & -\cos(\theta)\end{bmatrix}\\
&= \begin{bmatrix}\cos^2\theta-\sin^2\theta & 2\cos\theta\sin\theta\\ 2\cos\theta\sin\theta & \sin^2\theta-\cos^2\theta\end{bmatrix}\\
&= \begin{bmatrix}\cos(2\theta) & \sin(2\theta)\\ \sin(2\theta) & -\cos(2\theta)\end{bmatrix}.
\end{aligned}$$

b) From part (a)—the second-to-last matrix, particularly—we see that we may take $a = \cos\theta$ and $b = \sin\theta$. The fact that $a^2 + b^2 = 1$, for this particular choice of a and b, is probably the most well-known trigonometric identity.

1.14 The matrix in letter (a) is in echelon form. Its 2nd, 4th, 5th and 6th columns are pivot columns with pivots 2, 3, 1 and 2 respectively.

The matrix in letter (c) is in echelon form. Its 1st and 3rd columns are pivot columns with pivots 1 and 3 respectively.

The matrices in letters (d) and (e) are in echelon form, both having pivot 1 in the only pivot column, column 1.

The matrices in letters (b) and (f) are not in echelon form.

1.15

a) $(11, 3)$

b) $(4, 1, 3)$

c) $(-2, 0, 3, 1)$

1.16

a) $$\begin{aligned} 3x_1 + 2x_2 &= 8\\ x_1 + 5x_2 &= 7 \end{aligned}$$

b) $$\begin{aligned} 5x_1 - 2x_2 + x_3 &= 3\\ 2x_1 + 3x_2 - 4x_3 &= 0 \end{aligned}$$

c) $$\begin{aligned} 2x_1 + x_2 + 4x_3 &= -1\\ 4x_1 - 2x_2 + 3x_3 &= 4\\ 5x_1 + 2x_2 + 6x_3 &= -1 \end{aligned}$$

1.17 There is plenty to *do* here, but nothing to hand in.

1.18

a) The initial steps in OCTAVE are

```
octave> A = [2 0 -1 -4; -4 -2 1 11; 2 2 5 3];
octave> simpleGE
B =

     2     0    -1    -4
     0    -2    -1     3
     0     0     5    10
```

The last row says $5z = 10$, which yields $z = 2$. The first row says

$$y = -\frac{1}{2}(3+2) = -\frac{5}{2}.$$

Finally,

$$x = \frac{1}{2}(-4+2) = -1.$$

Thus, there is just one solution: $\mathbf{x} = (-1, -5/2, 2)$.

b) The initial steps in OCTAVE are

```
octave> A = [1 3 2 1 5; 3 2 6 3 1; 6 2 12 4 3];
octave> simpleGE
B =

     1     3     2     1     5
     0    -7     0     0   -14
     0     0     0    -2     5
```

The third column is free, so we will take x_3 to be free. Solving this system via backward substitution leads to solutions of the form $(3/2-2t, 2, t, -5/2) = (3/2, 2, 0, -5/2) + t(-2, 0, 1, 0)$, $t \in \mathbb{R}$.

c) The initial steps in OCTAVE are

```
octave> A = [1 3 0 1; -1 -1 1 5; 2 4 -1 -7];
octave> simpleGE
B =
    1    3    0    1
    0    2    1    6
    0    0    0   -3
```

The last of these rows says "0 = 3". Such a statement is nonsense, and so this system of equations has no solution.

1.19 There is only one solution, namely $(7.5, -2.5, 2)$.

1.20

a) There is only one choice: $c_1 = 4$, $c_2 = 3$ and $c_3 = 4$.

b) There is only one choice: $c_1 = -2$, $c_2 = 4$ and $c_3 = 1$.

c) There are infinitely many choices of the constants. They all take the form

$$(c_1, c_2, c_3) = t(-3, 4, 1) + (1, -5, 0)\,, \qquad t \in \mathbb{R}.$$

1.21

a) The product, regardless of order of the factors, is

(i) $\begin{bmatrix} 1 & & \\ b_{21} & 1 & \\ b_{31} & & 1 \end{bmatrix}$

(ii) $\begin{bmatrix} 1 & & & \\ b_{21} & 1 & & \\ b_{31} & & 1 & \\ b_{41} & & & 1 \end{bmatrix}$

(iii) $\begin{bmatrix} 1 & & & \\ & 1 & & \\ & b_{32} & 1 & \\ & b_{42} & & 1 \end{bmatrix}$

b) Each is the set of square matrices which are lower triangular, have 1's along the main diagonal, and just one nonzero element off the main diagonal. These nonzero elements all appear on the same column, and there is one matrix for each eligible position below the main diagonal.

c) The general statement: The product of a family of matrices as described in part (b) is obtainable by multiplying them in any order, and the result is a single lower-triangular matrix whose entries below the diagonal are the same as the sum of those entries from the matrices in the family.

1.22 When finding the nullspace of a matrix, the column we would use to augment **A** is full of zeros. Since nothing in this extra column changes as a result of elementary operations 1–3, we need not really add it. In the code below, we just do row reduction on **A** itself.

```
octave> A = [1 2 3 4 3; 3 6 18 9 9; 2 4 6 2 6; 4 8 12 10 12; 5 10 24 11 15]
A =
      1     2     3     4     3
      3     6    18     9     9
      2     4     6     2     6
      4     8    12    10    12
      5    10    24    11    15

octave> simpleGE
B =
    1    2    3    4    3
    0    0    9   -3    0
    0    0    0   -6    0
    0    0    0    0    0
    0    0    0    0    0
```

Columns 1, 3 and 4 are the pivot columns, while columns 2 and 5 are free. We now solve

$$\left.\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 + 3x_5 &= 0 \\ 9x_3 - 3x_4 &= 0 \\ -6x_4 &= 0 \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} x_4 &= 0 \\ x_3 &= \tfrac{1}{9}(3)(0) \;=\; 0 \\ x_1 &= -2(x_2) - 3(0) - 4(0) - 3x_5 \\ &= -2s - 3t\ , \end{aligned}\right.$$

where we have set $x_2 = s$ and $x_5 = t$. Thus the nullspace consists of vectors of the form

$$(-2s - 3t, s, 0, 0, t) \;=\; s(-2, 1, 0, 0, 0) + t(-3, 0, 0, 0, 1)\ ,$$

or $\text{null}(\mathbf{A}) = \text{span}(\{(-2, 1, 0, 0, 0), (-3, 0, 0, 0, 1)\})$.

1.23

a) Here is the result of entering the augmented matrix and running `simpleGE`:

```
octave> A = [1 3 2 -1 4; -1 -1 -3 2 -1; 2 8 3 2 16; 1 1 4 1 8]
A =
      1     3     2    -1     4
     -1    -1    -3     2    -1
      2     8     3     2    16
      1     1     4     1     8

octave> simpleGE
B =
    1    3    2   -1    4
    0    2   -1    1    3
    0    0    0    3    5
    0    0    1    0    2
```

The result is not in echelon form, though it would be easy enough to put it in echelon form by swapping the last two rows. A close inspection of the algorithm in `simpleGE` reveals that, after zeroing below a particular pivot in position (i, j), it moves on to position $(i+1, j+1)$ to search for the next pivot. If the entry in that position is too near zero (i.e., less than 10^{-10} in absolute value), it is just set equal to zero and we move on to look for a pivot in position $(i+1, j+2)$. No attempt is made in this algorithm to look in positions *underneath* $(i+1, j+1)$ for a nonzero entry, in which case we a pivot is placed into position $(i+1, j+1)$ via a row exchange. That is, `simpleGE` carries out Gaussian elimination successfully only when echelon form is obtainable without any row exchanges.

b) The third row says $3x_4 = 5$, which we solve to get $x_4 = 5/3$. The fourth row yields $x_3 = 2$. The second row says

$$2x_2 - x_3 + x_4 = 3 \qquad \Rightarrow \qquad x_2 = \frac{1}{2}\left(3 + 2 - \frac{5}{3}\right) = \frac{5}{3}.$$

Finally,

$$x_1 + 3x_2 + 2x_3 - x_4 = 4 \qquad \Rightarrow \qquad x_1 = 4 - 3\left(\frac{5}{3}\right) - 2(2) + \frac{5}{3} = -\frac{10}{3},$$

so the solution is $(-10/3, 5/3, 2, 5/3)$.

1.24 We let Octave do the reduction to echelon form:

```
octave> A = [1 2 -2 1 9; 2 5 1 9 9; 1 3 4 9 -2]
A =
    1    2   -2    1    9
    2    5    1    9    9
    1    3    4    9   -2

octave> simpleGE
B =
    1    2   -2    1    9
    0    1    5    7   -9
    0    0    1    1   -2
```

So, the solution of the first system is found by backward substitution

$$\left.\begin{aligned} x_1 + 2x_2 - 2x_3 &= 1 \\ x_2 + 5x_3 &= 7 \\ x_3 &= 1 \end{aligned}\right\} \qquad \Rightarrow \qquad \left\{\begin{aligned} x_2 &= 7 - 5(1) = 2 \\ x_1 &= 1 - 2(2) + 2(1) = -1. \end{aligned}\right.,$$

so there is a unique solution $\mathbf{x}_1 = (-1, 2, 1)$.

The second system is solved similarly, and is the unique solution $\mathbf{x}_2 = (3, 1, -2)$.

1.25 We augment the original matrix $\mathbf{A}$ with the 2-by-2 identity matrix, and then perform Gaussian elimination. In what follows, we go farther than just echelon form so that, still employing the three elementary operations, we make every pivot equal to 1 and zero out entries above pivots as well as below. Assuming $a \neq 0$, we have

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \begin{array}{c} \mathbf{r}_2 + (-c/a)\mathbf{r}_1 \to \mathbf{r}_2 \\ \sim \end{array} \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & (ad-bc)/a & -c/a & 1 \end{array}\right]$$

$$\begin{array}{c} a/(ad-bc)\mathbf{r}_2 \to \mathbf{r}_2 \\ \sim \end{array} \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right]$$

$$\begin{array}{c} \mathbf{r}_1 - b\mathbf{r}_2 \to \mathbf{r}_1 \\ \sim \end{array} \left[\begin{array}{cc|cc} a & 0 & ad/(ad-bc) & -ab/(ad-bc) \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right]$$

$$\begin{array}{c} (1/a)\mathbf{r}_1 \to \mathbf{r}_1 \\ \sim \end{array} \left[\begin{array}{cc|cc} 1 & 0 & d/(ad-bc) & -b/(ad-bc) \\ 0 & 1 & -c/(ad-bc) & a/(ad-bc) \end{array}\right].$$

Separating this into two linear systems, solving, and putting the two solutions together as the columns of the inverse matrix gives the result.

1.26 The key to this one is the following observation. If you get a new matrix **B** from an old one **A** by setting

$$(\text{row } i \text{ of } \mathbf{B}) \; = \; (\text{row } i \text{ of } \mathbf{A}) + \beta(\text{row } j \text{ of } \mathbf{A}) \,,$$

while transmitting all other rows of **A** intact into **B**, then this process may simply be reversed. That is, starting with B, take

$$(\text{row } i \text{ of } \mathbf{A}) \; = \; (\text{row } i \text{ of } \mathbf{B}) - \beta(\text{row } j \text{ of } \mathbf{B}) \,.$$

The resulting matrix $\mathbf{E}_{ij}^{-1}$ looks just like $\mathbf{E}_{ij}$ except β is replaced by $-\beta$.

1.27

a) (i) $\mathbf{A}_1\mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix}$ (iii) $\mathbf{A}_2\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} + a_{32}a_{21} & a_{32} & 1 \end{bmatrix}$

(ii) $(\mathbf{A}_1\mathbf{A}_2)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ a_{32}a_{21} - a_{31} & -a_{32} & 1 \end{bmatrix}$ (iv) $(\mathbf{A}_2\mathbf{A}_1)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a_{21} & 1 & 0 \\ -a_{31} & -a_{32} & 1 \end{bmatrix}$

b) (i) $\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 \\ a_{41} & a_{42} & a_{43} & 1 \end{bmatrix}$ (iv) $(\mathbf{A}_3\mathbf{A}_2\mathbf{A}_1)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a_{21} & 1 & 0 & 0 \\ -a_{31} & -a_{32} & 1 & 0 \\ -a_{41} & -a_{42} & -a_{43} & 1 \end{bmatrix}$

(ii) $(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a_{21} & 1 & 0 & 0 \\ a_{32}a_{21} - a_{31} & -a_{32} & 1 & 0 \\ a_{21}a_{42} + a_{31}a_{43} - a_{41} - a_{21}a_{32}a_{43} & a_{32}a_{43} - a_{42} & -a_{43} & 1 \end{bmatrix}$

(iii) $\mathbf{A}_3\mathbf{A}_2\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 \\ a_{31} + a_{32}a_{21} & a_{32} & 1 & 0 \\ a_{41} + a_{31}a_{43} + a_{21}a_{42} + a_{21}a_{32}a_{43} & a_{42} + a_{32}a_{43} & a_{43} & 1 \end{bmatrix}$

c) Suppose for each $j = 1, \dots, n$, $\mathbf{A}_j$ is an n-by-n lower triangular matrix with ones along the main diagonal and nonzero entries only in column j. Then the product $\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_n$ is another lower triangular matrix with ones along the main diagonal, and whose subdiagonal entries are equal to the sum of the subdiagonal entries of $\mathbf{A}_1$ through $\mathbf{A}_n$. The matrix $(\mathbf{A}_n\mathbf{A}_{n-1} \cdots \mathbf{A}_1)^{-1}$ is almost identical to $\mathbf{A}_1\mathbf{A}_2 \cdots \mathbf{A}_n$, except that their subdiagonal elements are opposite in sign.

1.28 We use a block interpretation of $\mathbf{AB}$. We know that

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_1 \\ \hline \mathbf{A}_2 \\ \hline \vdots \\ \hline \mathbf{A}_m \end{bmatrix} \mathbf{B} = \begin{bmatrix} \mathbf{A}_1\mathbf{B} \\ \hline \mathbf{A}_2\mathbf{B} \\ \hline \vdots \\ \hline \mathbf{A}_m\mathbf{B} \end{bmatrix},$$

where the blocks $\mathbf{A}_1, \dots, \mathbf{A}_m$ are the rows of $\mathbf{A}$. To show $\mathbf{AB}$ is lower triangular when $\mathbf{A}$ and $\mathbf{B}$ are, it suffices to show that all the elements to the right of the j^{th} one on the j^{th} row $\mathbf{A}_j\mathbf{B}$ are zero. We know that $\mathbf{A}_j\mathbf{B}$ is a linear combination of the rows of $\mathbf{B}$. Since all the elements to the right of the j^{th} one on row $\mathbf{A}_j$ of $\mathbf{A}$ are zero, the rows of $\mathbf{B}$ included in this linear combination are only rows $\mathbf{B}_1, \dots, \mathbf{B}_j$ from $\mathbf{B}$. But all of these rows contain zeros to the right of the j^{th} element. Thus $\mathbf{A}_j\mathbf{B}$ does too.

1.29

a) This interpolating linear polynomial would be of degree zero (i.e., be a constant function, one with zero slope) precisely when the two interpolation points have the same y-coordinate.

b)
i. Given n points in the plane, no two of which share the same x-coordinate, there is a unique polynomial having degree at most $n-1$ that passes through these n points.

ii. An equation $p(x_j) = y_j$ translates into

$$a_0 + x_j a_1 + x_j^2 a_2 + \cdots + x_j^{n-1} a_{n-1} \;=\; y_j \,.$$

Since there is one such equation in the unknowns $a_0, \ldots, a_{n-1}$ for each value of $j = 1, 2, \ldots, n$, we get our matrix $\mathbf{B}$ to be

$$\mathbf{B} \;=\; \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & x_2^3 & \cdots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & x_3^3 & \cdots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & \cdots & x_n^{n-1} \end{bmatrix}.$$

c) In Octave:

```
octave> x = [-2; -1; 0; 1; 2; 3];
octave> y = [-63; 3; 1; -3; 33; 367];
octave> B = [x.^0 x x.^2 x.^3 x.^4 x.^5];
octave> B \ y
ans =
   1.0000e+00
  -4.0000e+00
   9.4739e-15
  -1.0000e+00
  -1.0000e+00
   2.0000e+00
```

The command "`A \ b`" is a new one to us. It is the one used in Octave/Matlab to solve the linear system $\mathbf{Ax} = \mathbf{b}$. It shows us none of the intermediate steps—reduction to echelon form, backward substitution, etc.—but just gives the answer. It also fails to show us multiple solutions when the problem has them, providing only one representative solution in such cases. It reports an answer even in cases, like Exercise 1.18, where no solution exists, which means it ought to be used with some wariness. Nevertheless, it reports coefficients yielding the polynomial

$$p(x) = 1 - 4x - x^3 - x^4 + 2x^5 \,.$$

1.30 In this case our permutation matrix is just the identity. Thus, solving $\mathbf{Ly} = \mathbf{Pb}$ is simply a matter of solving $\mathbf{Ly} = \mathbf{b}$ using forward substitution. First, we solve

$$\mathbf{Ly} = \mathbf{Pb} = \begin{bmatrix} -10 \\ -1 \\ -1 \end{bmatrix}$$

via backward substitution to get $\mathbf{y} = (-10, -23/3, 10)$. Then we solve $\mathbf{Ux} = \mathbf{y}$, which looks like

$$\begin{aligned} 6x_1 - 4x_2 + 5x_3 &= -10 \\ \frac{1}{3}x_2 + \frac{13}{3}x_3 &= -\frac{23}{3} \\ -5x_3 &= 10. \end{aligned}$$

This leads to the solution $(2, 3, -2)$.

1.31 First, we solve

$$\mathbf{Ly} = \mathbf{Pb} = \begin{bmatrix} -19 \\ 1 \\ -13 \end{bmatrix}$$

via forward substitution to get $\mathbf{y} = (-19, 21/2, 0)$. Then we solve $\mathbf{Ux} = \mathbf{y}$, which looks like

$$\begin{aligned} 2x - 5y + z &= -19 \\ -\frac{3}{2}y - \frac{15}{2}z &= \frac{21}{2} \\ 0 &= 0. \end{aligned}$$

The third equation is vacuous, so we really have two equations in the three unknowns; z is a free variable. The 2nd equation yields $y = -7 - 5z$, while the first yields $x = -27 - 13z$. Thus, solutions have the form $(-27, -7, 0) + z(-13, -5, 1)$, $z \in \mathbb{R}$.

1.32 No row can have more than one pivot (and some have none). Since you move both downward and to the right in going from one pivot to the next, this $\min\{m, n\}$ is an upper bound on the number of pivots one can have.

1.33 Any matrix with a row whose entries are all zero works.

1.34

a) $\text{rank}(\mathbf{A}) = 3$, nullity $(\mathbf{A}) = 2$

b) A basis of $\text{col}(\mathbf{A})$ is $B = \{(1, -1, 0, 1), (-2, 3, 1, 2), (2, -2, 4, 5)$

c) One way: "Find a linear independent collection of vectors whose span is col($\mathbf{A}$)."

1.35

a) The columns of $\mathbf{A}$ are linearly independent, since the lack of any nonzero element in null ($\mathbf{A}$) indicates that the only linear combination of columns $\mathbf{A}_1, \ldots, \mathbf{A}_n$ of $\mathbf{A}$ that yields the zero vector is
$$(0)\mathbf{A}_1 + (0)\mathbf{A}_2 + \cdots + (0)\mathbf{A}_n.$$

b) There is precisely one solution.

c) There is no solution.

1.36 No. You can look through the first column for a nonzero entry. If none is found, you can proceed to the 2^{nd} (and beyond) column until a nonzero entry is found (guaranteed, since it is a *nonzero* matrix). The row in which this first nonzero entry is found may be brought to the top (via **Elementary Operation 2**), and the entry is a pivot in the echelon form of the matrix. With (at least) one pivot column, the original matrix is (at least) of rank 1.

1.37 Using a block interpretation of the product, we have

$$\mathbf{uv} = \begin{bmatrix} u_1 \\ \hline u_2 \\ \hline \vdots \\ \hline u_m \end{bmatrix} \mathbf{v} = \begin{bmatrix} u_1\mathbf{v} \\ \hline u_2\mathbf{v} \\ \hline \vdots \\ \hline u_m\mathbf{v} \end{bmatrix}.$$

This illustrates that each row of $\mathbf{uv}$ is a scalar multiple of $\mathbf{v}$. Thus, when we find the lead nonzero entry (the *pivot*) in row 1 of $\mathbf{uv}$ and then use **Elementary Operation 3** to zero out the entries in the column below this pivot, the result will be zero in every position (i, j) of the matrix with $i > 1$.

1.39 null ($\mathbf{A}$) = $\{\mathbf{0}\}$

1.40 An n-by-n matrix has nonzero determinant if and only if its rank equals n.

1.41 Assuming (vi), then (iv), (v) and (vii) hold as well. The problem with (ii) is that, for non-square matrices of full rank, some $\mathbf{b} \in \mathbb{R}^m$ will make the system inconsistent (i.e., not have a solution). Nevertheless, for those $\mathbf{b} \in \mathbb{R}^m$ which yield a consistent system, solutions are unique.

1.42

a) Here are sample results, taking $\alpha = 0.75$ (radians) as a "for instance".

```
octave> alph = .75;
octave> A = [cos(alph) -sin(alph); sin(alph) cos(alph)];
octave> eig(A)
ans =
   0.73169 + 0.68164i
   0.73169 - 0.68164i
```

b) The eigenvalues are $(2 \pm i)$. The effect multiplication by $\mathbf{A}$ has on the plane is to rotate it (0.4637 radians) and expand points so that their (new) distance from the origin is $\sqrt{5}$ times as far as their original distance from it.

c) We have $\mathbf{A} = \mathbf{BC}$ where

$$\mathbf{B} = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{5} \end{bmatrix} \qquad \text{and} \qquad \mathbf{C} = \begin{bmatrix} \cos(0.4637) & -\sin(0.4637) \\ \sin(0.4637) & \cos(0.4637) \end{bmatrix}.$$

d) Under these conditions, multiplication by $\mathbf{A}$ transforms the plane by rotating it and rescaling it so that the new location of points is $\sqrt{a^2 + b^2}$ times as far from the origin as their original locations.

1.43

a) For an n-by-n matrix, the eigenvalues are the roots of an n^{th}-degree polynomial. Such a polynomial has n roots, though some may repeated, which means it has at most n distinct roots. If $\mathbf{A}$ is 2-by-2, this means it can have at most one other root.

b) So long as $\mathbf{A}$ is a *real* matrix (i.e., has entries that are real nos.), then the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$ has real coefficients. The non-real roots of a polynomial with real coefficients come in complex-conjugate pairs. Since we already know $\mathbf{A}$ has one real eigenvalue, its other cannot be complex.

c) The quick answer is "No", and it's almost the right one. The matrices bringing about rotations in Exercise 1.42 were ones that had complex (nonreal) eigenvalues. Moreover, if $\mathbf{A}$ has an eigenpair $(\lambda, \mathbf{v})$, then good intuition into this fact tells us that $\mathbf{v}$ and $\mathbf{Av}$ have the same "direction". But how can any vector $\mathbf{v}$ have the same "direction" as $\mathbf{Av}$ if multiplication by $\mathbf{A}$ is a rotation of the plane? Strictly speaking, it cannot happen. But eigenpairs satisfy $\mathbf{Av} = \lambda\mathbf{v}$, and when $\lambda < 0$ the direction is "reversed". Thus, if $\lambda = -1$ with multiplicity (as an eigenvalue) 2, then $\mathbf{Av} = -\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^2$. That's the same as a rotation of the plane through an angle of π (= 180°). So, there is precisely one matrix $\mathbf{A}$, namely

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

which has real eigenvalues yet brings about a rigid rotation of $\mathbb{R}^2$.

1.44

a) $\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, eigenpairs: $(-1,(1,0)), (1,(0,1))$

b) $\mathbf{A} = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\pi/4) & \sin(\pi/4) \\ -\sin(\pi/4) & \cos(\pi/4) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(Note: Students can get this same matrix appealing to equation (1.6) with $\alpha = 2\theta = \pi/2$, or perhaps by appealing to part (b) of Exercise 1.13; these same observations are true of parts (c) and (d) below.), with eigenpairs $(-1,(-1,1)), (1,(1,1))$ (Any rescaling of the *eigenvectors* is permissible.)

c) $\mathbf{A} = \begin{bmatrix} \cos(0.6435) & \sin(0.6435) \\ -\sin(0.6435) & \cos(0.6435) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(0.6435) & -\sin(0.6435) \\ \sin(0.6435) & \cos(0.6435) \end{bmatrix} = \frac{1}{25} \begin{bmatrix} 7 & -24 \\ -24 & -7 \end{bmatrix}$,

$(-1,(3,4)), (1,(-4,3))$

d) $\mathbf{A} = \begin{bmatrix} \cos(\arctan(a/b)) & -\sin(\arctan(a/b)) \\ \sin(\arctan(a/b)) & \cos(\arctan(a/b)) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\arctan(a/b)) & \sin(\arctan(a/b)) \\ -\sin(\arctan(a/b)) & \cos(\arctan(a/b)) \end{bmatrix}$

$= \begin{bmatrix} \frac{b}{\sqrt{a^2+b^2}} & -\frac{a}{\sqrt{a^2+b^2}} \\ \frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{b}{\sqrt{a^2+b^2}} & \frac{a}{\sqrt{a^2+b^2}} \\ -\frac{a}{\sqrt{a^2+b^2}} & \frac{b}{\sqrt{a^2+b^2}} \end{bmatrix} = \frac{1}{a^2+b^2} \begin{bmatrix} b^2-a^2 & 2ab \\ 2ab & a^2-b^2 \end{bmatrix}$

1.45

a) $\mathbf{A} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, eigenpairs: $(-1,\mathbf{i}), (1,\mathbf{j}), (1,\mathbf{k})$

b) $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, eigenpairs: $(-1,(-1,1,0)), (1,(1,1,0)), (1,(0,0,1))$

c) The eigenvalues will be (-1) and 1 (the latter has multiplicity 2). The eigenvalue (-1) will have $\mathbf{n}$ as an associated eigenvector. The eigenvectors associated with eigenvalue 1 are those in the plane P.

1.46

a) Three hours corresponds to a rotation through an angle of $\pi/4$. Thus,

$$\mathbf{A} = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) & 0 \\ \sin(\pi/4) & \cos(\pi/4) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

b) To get the desired expansion of the z-axis, we have

$$\mathbf{A} = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) & 0 \\ \sin(\pi/4) & \cos(\pi/4) & 0 \\ 0 & 0 & 1.01 \end{bmatrix}.$$

1.47

a)

```
octave> hPts = [0 0.5 0.5 4 4 0.5 0.5 5.5 5.5 0 0];
octave> hPts = [hPts; 0 0 4.5 4.5 5 5 7.5 7.5 8 8 0];
octave> hPts = [hPts; ones(1,11)];
octave> plot(hPts(1,:), hPts(2,:))
octave> axis("square")
octave> axis([-1 6 -1 9])
```

b) The matrix which will achieve this is $\mathbf{A} = \begin{bmatrix} 1 & 0 & -5.5 \\ 0 & 1 & -8 \\ 0 & 0 & 1 \end{bmatrix}$.

```
octave> A = [1 0 -5.5; 0 1 -8; 0 0 1];
octave> newPts = A*hPts;
octave> plot(newPts(1,:), newPts(2,:))
octave> axis("square")
octave> axis([-6 1 -9 1])
```

c) We carry out this transformation in several steps. We must first translate the effective center to the origin. We then rotate the xy-plane (leaving the z-axis fixed) 180°, then translate back. That is,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2.75 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -2.75 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 5.5 \\ 0 & -1 & 8 \\ 0 & 0 & 1 \end{bmatrix}.$$

```
octave> H = [1 0 -2.75; 0 1 -4; 0 0 1];
octave> A = inv(H)*[-1 0 0; 0 -1 0; 0 0 1]*H;
octave> newPts = A*hPts;
octave> plot(newPts(1,:), newPts(2,:))
octave> axis("square")
octave> axis([-1 6 -1 9])
```

d) The effect of multiplication by such **A** is to leave the lower left corner of the "F" in its original position, and to leave horizontal line segments unaltered. Vertical lines, however, are transformed into slanted lines, slanting as with 'positive slope' when $c > 0$ and 'negative slope' when $c < 0$.

In every case where $c \neq 0$, these shear matrices have repeated eigenvalue (1), but only one line of eigenvectors. The eigenvectors associated with this eigenvalue all lie along the x-axis. Much of the time a matrix with real eigenvalues will have (at least) two different lines of eigenvectors.

1.48

a) One way to prove this is through blocking. Blocking **S** by column (so that the j^{th} column is called $\mathbf{S}_j$), the left-hand side

$$\mathbf{AS} = \mathbf{A}\left[\begin{array}{c|c|c|c} \mathbf{S}_1 & \mathbf{S}_2 & \cdots & \mathbf{S}_n \end{array}\right] = \left[\begin{array}{c|c|c|c} \mathbf{AS}_1 & \mathbf{AS}_2 & \cdots & \mathbf{AS}_n \end{array}\right],$$

while the right hand side

$$\mathbf{SD} = \left[\begin{array}{c|c|c|c} \mathbf{S}_1 & \mathbf{S}_2 & \cdots & \mathbf{S}_n \end{array}\right]\left[\begin{array}{c|c|c|c|c} d_1 & 0 & 0 & \cdots & 0 \\ \hline 0 & d_2 & 0 & \cdots & 0 \\ \hline 0 & 0 & d_3 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & 0 & \cdots & d_n \end{array}\right] = \left[\begin{array}{c|c|c|c} d_1\mathbf{S}_1 & d_2\mathbf{S}_2 & \cdots & d_n\mathbf{S}_n \end{array}\right].$$

Since $\mathbf{AS} = \mathbf{SD}$, their columns must be equal—that is, $\mathbf{AS}_j = d_j\mathbf{S}_j$ for $j = 1, 2, \ldots, n$. This, indeed, is what it means for $(d_j, \mathbf{S}_j)$ to be an eigenpair, so long as we know $\mathbf{S}_j$ is a nonzero vector, which is the case since **S** is invertible (so has nonzero determinant).

b) Using Octave, such a matrix **A** is produced below:

```
octave> S = transpose([4 1 0 -1; 1 2 1 1; 1 -1 3 3; 2 -1 1 2]);
octave> D = diag([-1 2 1 1]);
octave> A = S*D*inv(S)
A =

   -0.461538    0.846154   -2.230769    3.000000
   -0.410256    1.974359   -0.871795    1.333333
   -0.025641    0.435897    0.820513    0.333333
    0.333333    0.333333    0.333333    0.666667
```

c) Note first that for any square matrix **B**, $\det(\mathbf{B}) = \det(\mathbf{BI}) = \det(\mathbf{B})\det(\mathbf{I})$. This equation holds whether **B** has nonzero determinant or not, so when it is nonzero we may divide to get $\det(\mathbf{I}) = 1$. Thus, $\det(\mathbf{A}) = \det(\mathbf{SDS}^{-1}) = \det(\mathbf{S})\det(\mathbf{D})\det(\mathbf{S}^{-1}) = \det(\mathbf{SS}^{-1})\det(\mathbf{D}) = \det(\mathbf{I})\det(\mathbf{D}) = \det(\mathbf{D}) = d_1d_2d_3\cdots d_n$, this last equality coming from Fact B in Section 1.7.

d) Since λ is an eigenvalue of $\mathbf{B}$ there must be a nonzero vector $\mathbf{v}$ for which

$$\lambda\mathbf{v} \;=\; \mathbf{B}\mathbf{v} \;=\; (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})\mathbf{v} \;=\; \mathbf{P}^{-1}\mathbf{A}(\mathbf{P}\mathbf{v})\,.$$

Multiplying the two ends of this equality by $\mathbf{P}$, we get $\lambda(\mathbf{P}\mathbf{v}) = \mathbf{A}(\mathbf{P}\mathbf{v})$, which is the same as saying that $(\lambda, \mathbf{P}\mathbf{v})$ is an eigenpair of $\mathbf{A}$.

2.1 The additive identity (function) in $\mathscr{C}^k(a,b)$ is the function that is constantly zero—that is, $(x \mapsto 0)$. This function is differentiable (it is its own derivative) up to an order you like (i.e., it has a k^{th} derivative for any choice of nonnegative integer k), and all of its derivatives are continuous.

2.2

a) Let us assume that $\mathbf{u}$, $\mathbf{v}$ are nonzero and non-parallel in $\mathbb{R}^3$. Suppose that we have scalars c_1, c_2 such that

$$c_1\mathbf{u} + c_2\mathbf{v} \;=\; \mathbf{0}\,.$$

If $\{\mathbf{u}, \mathbf{v}\}$ were linearly *dependent* then it would be possible to choose at least one of these scalars, say c_2, to be something other than 0. But that would mean

$$\mathbf{v} \;=\; -\frac{c_1}{c_2}\mathbf{u}\,.$$

Looking carefully at this equation, we see that the requirement "both $\mathbf{u}$, $\mathbf{v}$ are nonzero" means it would be impossible for c_1 to equal zero (if $c_2 \neq 0$). But that would mean $\mathbf{v}$ was a scalar multiple of $\mathbf{u}$ (i.e., parallel vectors).

b) The $\mathbf{u} \times \mathbf{v} = (3, -1, 1)$, and so the equation is $3x - y + z = 0$.

c) Yes, this always happens. That is because, "containing the *origin*" means that $(0, 0, 0)$ must satisfy the equation $Ax + By + Cz = D$ or, equivalently, $0 + 0 + 0 = D$.

2.3

a) We will show this plane is not closed under addition, as would be required if it were a vector space. Let $\mathbf{v}_1 = (3, -1, 1)$ (the vector we get by taking both $s = t = 0$) and $\mathbf{v}_2 = (4, -1, 2)$ (the vector we get by taking $s = 1$, $t = 0$). Then $\mathbf{w} = \mathbf{v}_1 + \mathbf{v}_2 = (7, -2, 3)$. To show that $\mathbf{w}$ is not in our plane, we must demonstrate that there is *no choice* of real numbers s and t such that $\mathbf{w} = (3, -1, 1) + s(1, 0, 1) + t(0, 1, -1)$. But this is equivalent to showing that no s, t exist such that $\mathbf{w} - (3, -1, 1) = s(1, 0, 1) + t(0, 1, -1)$, or that the matrix equation

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix}$$

has no solution. The latter problem is probably the easiest one to work with, as we need only reduce the augmented matrix to echelon form:

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

The presence of a nonzero element (-3) in the augmented column alongside a row of zeros indicates the system has no solution.

b) $s(1,0,1) + t(0,1,-1)\,, \qquad s,\ t \in \mathbb{R}$.

c) Since the cross product (yielding a normal vector) of $(0,1,-1)$ and $(1,0,1)$ is $(1,-1,-1)$, both planes are going to have equations of the form $x - y - z = D$. For the plane in part (b), the equation is simply $x - y - z = 0$. Since the original plane needs to pass through the point $(3,-1,1)$, we note that $D = 3 - (-1) - 1 = 3$, and so its equation is $x - y - z = 3$.

2.4

b) By reducing $\mathbf{H}$ to an echelon form $\mathbf{R}$ and then solving for $\text{null}(\mathbf{R}) = \text{null}(\mathbf{H})$, we get (one possible) basis

$$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\} = \{(1,1,1,0,0,0,0), (1,0,0,1,1,0,0), (0,1,0,1,0,1,0), (1,1,0,1,0,0,1)\}.$$

(One can check that this is, indeed, a linearly independent set of vectors in $\mathbb{Z}_2^7$ by making them the columns of a matrix, reducing *that* matrix to echelon form, and noting that there are no free columns.)

c) Perhaps the easiest way to do this is in two steps. First, one shows that each $\mathbf{u}_j$ is in $\text{null}(\mathbf{H})$—that is, show that $\mathbf{H}\mathbf{u}_j = \mathbf{0}$ for $j = 1, \ldots, 4$. In doing so, we demonstrate that the set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\}$ is a collection of vectors all of which are from $\text{null}(\mathbf{H})$. (Do you see why $\text{span}(S) = \text{span}(\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\})$, and thereby that $\text{span}(S) = \text{null}(\mathbf{H})$?) Next, we can form the matrix whose columns are the vectors in S—that is, $\mathbf{A} = \left[\begin{array}{c|c|c|c|c|c|c|c} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 & \mathbf{w}_4 \end{array}\right]$. Then by reducing $\mathbf{A}$ to echelon form, we find that the first four columns are pivot columns while the last four are free, thereby demonstrating that only the first four are needed as a basis for $\text{span}(S)$.

d) For this (received) $\tilde{\mathbf{v}}$ we have

$$\mathbf{H}\tilde{\mathbf{v}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},$$

so there is an error in transmission (since $\tilde{\mathbf{v}} \notin \text{null}(\mathbf{H})$). If we presume that only *one* bit of the transmitted word is corrupt, we conclude it is the 3rd bit (since $\mathbf{H}\tilde{\mathbf{v}}$ is equal to the 3rd column of $\mathbf{H}$). Extracting the first four elements from $\tilde{\mathbf{v}}$ and altering the 3rd of these yields $(1, 0, 0, 1)$ (or 1001).

e) Sadly, the use of the Hamming (7,4) scheme for detecting and correcting errors breaks down if two (or more) bits from a 7-bit transmitted word are corrupt. To see this, notice that if the 7-bit word $\mathbf{v} = (1, 0, 0, 1, 1, 0, 0)$ is corrupted to $\tilde{\mathbf{v}} = (1, 1, 0, 1, 0, 0, 0)$ (*two* altered bits), then we will, indeed, *detect* an error (it is still the case, with this $\tilde{\mathbf{v}}$, that $\tilde{\mathbf{v}} \notin \text{null}(\mathbf{H})$), but that our process for correction would make us think that the 7th bit alone was faulty (not the pair of 2nd and 5th bits). With *three* altered bits we might not even detect the error!

2.5 $\mathbb{R}^n$ is the set of all n-tuples whose elements are real numbers. That is,

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \,|\, \text{each } x_i \in \mathbb{R}\} .$$

2.6 Let $\mathbf{u}, \mathbf{v} \in S$, and $a, b \in \mathbb{R}$. We have $\mathbf{u} = (u_1, u_1, u_3)$ and $\mathbf{v} = (v_1, v_1, v_3)$, so

$$a\mathbf{u} + b\mathbf{v} = (au_1 + bv_1, au_1 + bv_1, au_3 + bv_3) ,$$

which shows a linear combination of vectors from S is once again in S. By the subspace test, S is a subspace.

2.7 This S cannot be a subspace, because it does not contain the zero element $(0, 0)$ of $\mathbb{R}^2$.

2.8

a) The most likely correct answer is $\{(3, 1, 0), (5, 0, 1)\}$. There are other correct answers (see part (f) below).

b)
```
B =
    8.0000   -15.0000   -25.0000
   -5.0000    18.0000    25.0000
    5.0000   -15.0000   -22.0000

octave-3.0.0:13> [V, lam] = eig(B)
V =
   0.81650    0.57735   -0.26243
  -0.40825   -0.57735    0.80338
   0.40825    0.57735   -0.53452

lam =
   3.00000    0.00000    0.00000
   0.00000   -2.00000    0.00000
   0.00000    0.00000    3.00000
```

From this output, we get eigenpairs $(3, (0.8165, -0.40825, 0.40825))$, $(3, (-0.26243, 0.80338, -0.53452))$, and $(-2, (0.57735, -0.57735, 0.57735))$.

c) The first step to obtaining the eigenpairs is to find the eigenvalues, in this case $\lambda_{1,2} = 3$ and $\lambda_3 = -2$. Next we turn to finding the corresponding eigenvectors; for each $j = 1, 2, 3$, we find null$\left(\mathbf{B} - \lambda_j \mathbf{I}\right)$. In the case of $\lambda_{1,2} = 3$, this means finding the nullspace of $(\mathbf{B} - 3\mathbf{I})$, which is precisely our matrix $\mathbf{A}$.

d) The basis extracted from Octave output is

$$\{\mathbf{w}_1, \mathbf{w}_2\} = \{(0.8165, -0.40825, 0.40825), (-0.26243, 0.80338, -0.53452)\}.$$

Taking $\mathbf{v}_1 = (3, 1, 0)$ and $\mathbf{v}_2 = (5, 0, 1)$, we have

$$\mathbf{v}_1 = 4.8697\mathbf{w}_1 + 3.7193\mathbf{w}_2 \qquad \text{and} \qquad \mathbf{v}_2 = 7.3192\mathbf{w}_1 + 3.7193\mathbf{w}_2 .$$

I used Octave to obtain these values for the coefficients, employing the commands

```
octave-3.0.0:16> V(:,[1 3]) \ [3; 1; 0]
ans =
   4.8697
   3.7193

octave-3.0.0:17> V(:,[1 3]) \ [5; 0; 1]
ans =
   7.3192
   3.7193
```

e) There are (nearly) always infinitely many different possible bases for a given vector (sub)space, and we have no reason to suspect that Octave (or Matlab) finds a basis in exactly the same fashion as we would. What is important here is that, whatever basis you provide, its span should represent the same *plane* of vectors that comprise null$(\mathbf{B})$.

f) One could always do what was done in part (d)—make sure the number of vectors in each answer is the same, and that the basis vectors obtained by hand are all in the span of the basis obtained via software. An easier approach, however, is to simply check that $\mathbf{Av}_i = \mathbf{0}$ (i.e., that each of the basis vectors you obtained by hand is, indeed, in the nullspace of $\mathbf{A}$).

2.9

a) We can form a matrix whose columns are the vectors from S and reduce to echelon form:

```
A =
    2    3    1    1
    1    2    1    0
    6   -1   -7   13

octave-3.0.0:31> rref(A)
ans =
   1.00000   0.00000  -1.00000   2.00000
   0.00000   1.00000   1.00000  -1.00000
   0.00000   0.00000   0.00000   0.00000
```

This shows that the 3rd and 4th columns may be removed from S without reducing its span (because those columns are free). Thus a basis for span(S) is $\{(2, 1, 6), (3, 2, -1)\}$.

b) As in the previous part, one might find an echelon form for $\mathbf{A}$ and, in so doing, discover that it has no free columns. Another way to discover the same thing is to compute det($\mathbf{A}$).

```
octave-2.9.13:42> A = [1 -1 0 1; 3 -1 2 1; 2 1 2 1; 1 2 1 -1]
A =
   1  -1   0   1
   3  -1   2   1
   2   1   2   1
   1   2   1  -1

octave-2.9.13:43> det(A)
ans =  6
```

Since the determinant of the matrix $\mathbf{A}$ (whose columns are precisely the given vectors) is nonzero, these vectors are linearly independent. A basis might therefore consist of the four vectors themselves. An alternate basis (which is just as valid) might consist of any four linearly independent vectors in $\mathbb{R}^4$, such as $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$, the standard basis for $\mathbb{R}^4$ (see Example 24).

2.10 Yes, S *must* be a basis for $\mathcal{V}$. Since span(S) = $\mathcal{V}$, the only way S could *not* be a basis for $\mathcal{V}$ is if S were linearly dependent. But if that were so, we could *extract* vectors (at least one, and perhaps more) from S until we had paired down to a sub-collection B which was a basis for span(S) = $\mathcal{V}$. But then B would have fewer than n vectors, which would contradict the assumption that dim($\mathcal{V}$) = n.

2.11 Let $\mathbf{b} \in \mathbb{R}^m$, and suppose $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ both satisfy the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. Then

$$\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) \;=\; \mathbf{A}\mathbf{x}_1 - A\mathbf{x}_2 \;=\; \mathbf{b} - \mathbf{b} \;=\; \mathbf{0}.$$

This shows $(\mathbf{x}_1 - \mathbf{x}_2) \in \text{null}\,(\mathbf{A})$. But since $\mathbf{0}$ is the only element of null ($\mathbf{A}$), it must be that

$$\mathbf{x}_1 - \mathbf{x}_2 \;=\; \mathbf{0}\,, \qquad \text{or} \qquad \mathbf{x}_1 \;=\; \mathbf{x}_2\,.$$

2.12

a) We know dim($\mathbb{R}^2$) = 2, and there are two vectors here. Since they are both nonzero vectors, and are not parallel, these vectors are linearly independent. By Theorem 8, this is a basis for $\mathbb{R}^2$.

b) A basis for $\mathbb{R}^2$ must have two vectors, no more or less (Theorem 8 again). Thus, our set here cannot be a basis.

c) Since dim($\mathbb{R}^3$) = 3, this set at least has the right number of vectors. If we can show it is a linearly independent set, then Theorem 8 tells us it is a basis. So, suppose

$$(0,0,0) \;=\; c_1(1,0,1) + c_2(1,1,0) + c_3(0,1,-1)\,.$$

We can write this same equation in matrix form:

$$\mathbf{A}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \;=\; \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \text{where} \qquad \mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

If det($\mathbf{A}$) $\neq 0$, then the equation above would have just one solution, which would be $c_1 = c_2 = c_3 = 0$. However, it is the case here that det($\mathbf{A}$) = 0, so there are other choices of the c_i's (ones where not all of them are zero) that solve this equation. Thus, this set is linearly dependent, and not a basis.

d) As in part c), we can look at the determinant of the matrix

$$\mathbf{A} \;=\; \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

This time, det($\mathbf{A}$) = 1, which shows that this set of vectors is linearly independent. Since there are three of them (and dim($\mathbb{R}^3$) = 3), they form a basis for $\mathbb{R}^3$.

2.13 That $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ spans $\mathbb{R}^3$ is a fact we have used since MATH 162, easily going back and forth between component notation $\mathbf{x} = (x_1, x_2, x_3)$ and writing $\mathbf{x}$ as a linear combination of vectors in S: $\mathbf{x} = x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$. To see that S is linearly independent, we note that if

$$\mathbf{0} = c_1\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k} ,$$

then $(0, 0, 0) = (c_1, c_2, c_3)$. Equality of vectors means that $c_1 = c_2 = c_3 = 0$.

2.14 One basis for $\mathbb{Z}_2^n$ is the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$ (see Example 24). (The demonstration of this is very much like Exercise 2.13.) Thus, $\dim(\mathbb{Z}_2^n) = n$.

2.15 By supposition, $\mathbf{u} \in \text{span}(S)$, so there exist real numbers $a_1, \ldots, a_m$ for which

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m .$$

We first show that $\text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\})$ is a subset of $\text{span}(S)$. Let $\mathbf{w} \in \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\})$. Then for some choice of real numbers $b_1, \ldots, b_{m+1}$,

$$\begin{aligned} \mathbf{w} &= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_m\mathbf{v}_m + b_{m+1}\mathbf{u} \\ &= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_m\mathbf{v}_m + b_{m+1}(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m) \\ &= (b_1 + b_{m+1}a_1)\mathbf{v}_1 + (b_2 + b_{m+1}a_2)\mathbf{v}_2 + \cdots + (b_m + b_{m+1}a_m)\mathbf{v}_m , \end{aligned}$$

which shows $\mathbf{w} \in \text{span}(S)$. Since $\mathbf{w}$ is an arbitrary element of $\text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\})$, we have that $\text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\})$ is a subset of $\text{span}(S)$.

Now we must show $\text{span}(S)$ is a subset of $\text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\})$. But this is much easier, because for any element $\mathbf{w} \in \text{span}(S)$, there are real numbers $b_1, \ldots, b_m$ for which

$$\mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_m\mathbf{v}_m ,$$

which is already a linear combination of vectors in $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}\}$.

2.17 Suppose such a subset T were linearly *dependent*. Then there would exist scalars c_1, $\ldots$, c_m, not all of which were zero, such that

$$c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \mathbf{0} .$$

Now take $a_1 = c_1, a_2 = c_2, \ldots, a_m = c_m$, and $a_{m+1} = a_{m+2} = \cdots = a_k = 0$. Then

$$a_1\mathbf{u}_1 + \cdots + a_m\mathbf{u}_m + a_{m+1}\mathbf{u}_{m+1} + \cdots + a_k\mathbf{u}_k = c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \mathbf{0} ,$$

and not all of the a_j's are zero. This would mean that the original set S is linearly dependent as well.

2.18 The set S may already be linearly independent, in which case $\dim(\mathcal{V}) = |S|$. If S is linearly dependent, then we can remove elements from S in such quantity as needed until the resulting (new) set B is linearly independent. Since S spans $\mathcal{V}$, B spans as well (recall that the process we use to remove elements from S does not result in a smaller span), and so is a basis for $\mathcal{V}$. Either way, (whether B and S are the same, or S has elements not found in B) $\dim(\mathcal{V}) = |B| \leq |S|$.

On the other hand, the fact that T is linearly independent means that every subset of T is also linearly independent (see Exercise 2.17). Thus, if $|T| > \dim(\mathcal{V})$, then you could take any subset of T that contained $\dim(\mathcal{V})$ elements and, by Theorem 8, this subset would be a basis for $\mathcal{V}$. But then T, which supposedly has even more elements than B, would be a basis as well, which violates Theorem 8. Thus, $|T| \leq |B| \leq |S|$.

2.19

a) In words, the left nullspace of $\mathbf{A}$ is the set of vectors $\mathbf{v} \in \mathbb{R}^m$ satisfying $\mathbf{A}^T\mathbf{v} = \mathbf{0}$ or, equivalently, satisfying $\mathbf{v}^T\mathbf{A} = \mathbf{0}^T$. Using set notation, this is

$$\text{null}\left(\mathbf{A}^T\right) \;=\; \left\{\mathbf{v} \in \mathbb{R}^m \mid \mathbf{A}^T\mathbf{v} = \mathbf{0}\right\} \;.$$

b) $\text{null}\left(\mathbf{A}^T\right)$ has basis $\{(1, -3, 1)\}$, and so $\dim(\text{null}\left(\mathbf{A}^T\right)) = 1$.

c) $\dim(\text{null}\left(\mathbf{A}^T\right)) = m - \text{rank}(\mathbf{A})$

2.20

a) $\text{col}(\mathbf{A})$ has basis $\{(2, 1, -3, 1), (3, 0, -5, 0)\}$.
$\text{col}(\mathbf{A}^T)$ has basis $\{(1, 0, 3, 1), (0, 1, -2, -1)\}$.
$\text{null}\,(\mathbf{A})$ has basis $\{(-3, 2, 1, 0), (-1, 1, 0, 1)\}$.
$\text{null}\left(\mathbf{A}^T\right)$ has basis $\{5, -1, 3, 0), (0, -1, 0, 1)\}$.
For this particular matrix $\mathbf{A}$, the dimension of each subspace is 2.

b) Every vector in a basis for $\text{col}(\mathbf{A})$ is orthogonal to (or has zero dot product with) every vector in a basis for $\text{null}\left(\mathbf{A}^T\right)$. (Actually, it is not only true for basis vectors: Every vector in $\text{col}(\mathbf{A})$ is orthogonal to every vector in $\text{null}\left(\mathbf{A}^T\right)$.) Naturally, this also means that vectors in $\text{col}(\mathbf{A}^T)$ are orthogonal to vectors in $\text{null}\,(\mathbf{A})$.

3.1 For $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$, we have

$$\begin{aligned}\langle \mathbf{u}, a\mathbf{v} + b\mathbf{w} \rangle &= \sum_{j=1}^{n} u_j(av_j + bw_j) = \sum_{j=1}^{n}(au_jv_j + bu_jw_j) \\ &= a\sum_{j=1}^{n} u_jv_j + b\sum_{j=1}^{n} u_jw_j = a\langle \mathbf{u}, \mathbf{v} \rangle + b\langle \mathbf{u}, \mathbf{w} \rangle \;,\end{aligned}$$

which establishes property (3.2).

3.2

a) We have

$$\begin{aligned} \langle a\mathbf{u} + b\mathbf{v}, \mathbf{w}\rangle &= \langle \mathbf{w}, a\mathbf{u} + b\mathbf{v}\rangle = a\langle \mathbf{w}, \mathbf{u}\rangle + b\langle \mathbf{w}, \mathbf{v}\rangle \qquad \text{(using linearity in } 2^{\text{nd}} \text{ argument)} \\ &= a\langle \mathbf{u}, \mathbf{w}\rangle + b\langle \mathbf{v}, \mathbf{w}\rangle \ . \end{aligned}$$

b) We have

$$\begin{aligned} \langle a\mathbf{u} + b\mathbf{v}, a\mathbf{u} + b\mathbf{v}\rangle &= a\langle a\mathbf{u} + b\mathbf{v}, \mathbf{u}\rangle + b\langle a\mathbf{u} + b\mathbf{v}, \mathbf{v}\rangle \qquad \text{(using linearity in the } 2^{\text{nd}} \text{ argument)} \\ &= a^2\langle \mathbf{u}, \mathbf{u}\rangle + ab\langle \mathbf{v}, \mathbf{u}\rangle + ab\langle \mathbf{u}, \mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle \qquad \text{(using linearity in the } 1^{\text{st}} \text{ argume} \\ &= a^2\|\mathbf{u}\|^2 + 2ab\langle \mathbf{u}, \mathbf{v}\rangle + b^2\|\mathbf{v}\|^2 \ . \end{aligned}$$

c) We start with the expression

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}\rangle = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \ , \end{aligned}$$

where the $\langle \mathbf{u}, \mathbf{v}\rangle$-term disappears since $\mathbf{u} \perp \mathbf{v}$.

3.3 Suppose $\mathbf{v} \in \mathbb{R}^n$ satisfies $\langle \mathbf{v}, \mathbf{v}\rangle = 0$. Then

$$0 = \langle \mathbf{v}, \mathbf{v}\rangle = \|\mathbf{v}\|^2 = v_1^2 + v_2^2 + \cdots + v_n^2 \ .$$

With the expression on the far right containing only perfect squares of real numbers, it follows that each $v_j = 0$. Since all of the components of $\mathbf{v}$ are zero, $\mathbf{v}$ is the zero vector.

3.4

a) Since $\mathbf{0}$ is orthogonal to every vector in $\mathbb{R}^n$, the orthogonal complement of $\mathcal{U}$ contains every vector.

b) Let $\mathbf{v} \in \mathcal{U}^\perp$. Assuming $\mathcal{U} = \mathbb{R}^n$, we have that $\mathbf{v} \in \mathcal{U}$ as well. Thus, $\langle \mathbf{v}, \mathbf{v}\rangle = 0$. But, by Exercise 3.3 Exercise 3.2, the only vector which is orthogonal to itself is $\mathbf{v} = \mathbf{0}$.

3.5

a) A basis for:
$\text{col}(\mathbf{A}^T)$: $\{(3,4)\}$
$\text{col}(\mathbf{A})$: $\{(1,2)\}$
$\text{null}\,(\mathbf{A})$: $\{(4,-3)\}$
$\text{null}\left(\mathbf{A}^T\right)$: $\{(2,-1)\}$

b) A basis for:
$\text{col}(\mathbf{A}^T)$: $\{(1,0,-2),(0,1,1)\}$
$\text{col}(\mathbf{A})$: $\{(1,2),(3,4)\}$
(the standard basis is also appropriate here) $\text{null}\,(\mathbf{A})$: $\{(2,-1,1)\}$
$\text{null}\left(\mathbf{A}^T\right) = \{\mathbf{0}\}$: By convention a basis for the trivial subspace is the empty set $\{\}$.

c) A basis for:
$\text{col}(\mathbf{A}^T)$: $\{(1,0),(0,1\}$
$\text{col}(\mathbf{A})$: $\{(4,1,2,3),(-2,3,1,4)\}$
$\text{null}\,(\mathbf{A})$: $\{\}$
$\text{null}\left(\mathbf{A}^T\right)$: $\{(-5/14,-4/7,1,0),(-5/14,-11/7,0,1)\}$

d) A basis for:
$\text{col}(\mathbf{A}^T)$: $\{(1,0,0,0),(0,1,1,1),(0,0,1,1)\}$
$\text{col}(\mathbf{A})$: $\{(1,0,0,1),(0,1,0,1),(0,1,1,2)\}$
$\text{null}\,(\mathbf{A})$: $\{(0,0,-1,1)\}$
$\text{null}\left(\mathbf{A}^T\right)$: $\{(1,1,1,-1)\}$

The vectors in $\text{col}(\mathbf{A})$ should be orthogonal to those in $\text{null}\left(\mathbf{A}^T\right)$; those in $\text{col}(\mathbf{A}^T)$ should be orthogonal to those in $\text{null}\,(\mathbf{A})$. Both of these relationships are realized.

3.6

a) The subspace $\mathcal{U}$ of $\mathbb{R}^3$ is precisely the row space of $\mathbf{A} = \begin{bmatrix}1 & -1 & 1\end{bmatrix}$. We know $\mathcal{U}^\perp = \text{col}(\mathbf{A}^T)^\perp = \text{null}\,(\mathbf{A})$. Finding a basis for the latter, we get $\{(-1,0,1),(1,1,0)\}$.

b) We may simply take the cross product to find a normal vector:

$$\mathbf{n} = \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 1 & 2 & 1\\ 1 & -1 & 2\end{vmatrix} = \begin{bmatrix}5\\-1\\-3\end{bmatrix}.$$

Then $\{\mathbf{n}\}$ is a basis for $\mathcal{V}^\perp$.

3.7 From Theorem 11 we have that $\text{col}(\mathbf{A}^T)^\perp = \text{null}\,(\mathbf{A})$. Thus, if these vectors were in these respective spaces, they would be orthogonal. However,

$$\langle(1,5,-1),(2,1,6)\rangle = 1 \neq 0\,,$$

so it is *not* possible.

3.8 We know rank($\mathbf{A}$) equals the number of pivot columns in an echelon form for $\mathbf{A}$, while $\dim(\text{null}(\mathbf{A})) = \text{nullity}(\mathbf{A})$ equals the number of free columns. Since the total number of columns is n, $\text{nullity}(\mathbf{A}) = n - r$. Since rank($\mathbf{A}$) also equals the dimension of the row space of $\mathbf{A}$, we may apply similar reasoning to reveal that $\text{nullity}\left(\mathbf{A}^T\right) = m - r$.

3.9

a) That $\mathbf{Ax}$ is in col($\mathbf{A}$) is automatic, since the column space of $\mathbf{A}$ is exactly the same thing as the range of the map $(\mathbf{x} \mapsto \mathbf{Ax})\colon \mathbb{R}^n \to \mathbb{R}^m$.

Now suppose $\mathbf{x} \in \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$. Then

$$(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{0} \qquad \Rightarrow \qquad \mathbf{A}^T(\mathbf{Ax}) = \mathbf{0} \qquad \Rightarrow \qquad \mathbf{Ax} \in \text{null}\left(\mathbf{A}^T\right) .$$

b) First let $\mathbf{x} \in \text{null}(\mathbf{A})$. Then

$$(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{A}^T(\mathbf{Ax}) = \mathbf{A}^T\mathbf{0} = \mathbf{0} .$$

Thus, $\mathbf{x} \in \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$. This shows $\text{null}(\mathbf{A}) \subset \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$.

Now let $\mathbf{x} \in \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$. By part a), this means $\mathbf{Ax} \in \text{col}(\mathbf{A}) \cap \text{null}\left(\mathbf{A}^T\right) = \{\mathbf{0}\}$. That is, $\mathbf{Ax} = \mathbf{0}$, showing $\mathbf{x} \in \text{null}(\mathbf{A})$. Thus, $\text{null}\left(\mathbf{A}^T\mathbf{A}\right) \subset \text{null}(\mathbf{A})$. Together with the previous inclusion, we get the two sets are equal.

c) Since their nullspaces are equal, we have $\text{nullity}\left(\mathbf{A}^T\mathbf{A}\right) = \text{nullity}(\mathbf{A})$. Both $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}$ have n columns. The rank of a matrix is equal to the number of its columns minus its nullity. So

$$\text{rank}(\mathbf{A}^T\mathbf{A}) = n - \text{nullity}\left(\mathbf{A}^T\mathbf{A}\right) = n - \text{nullity}(\mathbf{A}) = \text{rank}(\mathbf{A}) .$$

d) This follows immediately from Theorem 4, since $\mathbf{A}^T\mathbf{A}$ is a square matrix.

3.10 Let $\mathbf{v} \in \mathcal{V}$ be selected arbitrarily, and let $\mathbf{0}$ denote the additive identity in $\mathcal{V}$. Then

$$\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0} + \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle + \langle \mathbf{0}, \mathbf{v} \rangle .$$

Subtracting $\langle \mathbf{0}, \mathbf{v} \rangle$ from both sides, we get $0 = \langle \mathbf{0}, \mathbf{v} \rangle$.

3.11 If $\|\mathbf{v}\| = \langle \mathbf{v}, \mathbf{v} \rangle = 0$, then point (ii) of Definition 15 declares that $\mathbf{v}$ is the zero vector.

3.12

a) For any vector $\mathbf{v} \in \mathcal{U} \cap \mathcal{U}^{\perp}$, $\mathbf{v}$ must be orthogonal to itself. In Exercise 3.3 we showed the only vector with this property is $\mathbf{0}$. Moreover, we know that $\mathbf{0}$ is in every subspace of $\mathbb{R}^n$, so it is in both $\mathcal{U}$ and $\mathcal{U}^{\perp}$.

b) If $S = \{\mathbf{i}, \mathbf{j}\}$ we know $S^{\perp} = \{t\mathbf{k} \mid t \in \mathbb{R}\}$. Neither $\mathbf{i}$ nor $\mathbf{j}$ are in $S^{\perp}$, and hence $S \cap S^{\perp}$ is empty.

3.13 Suppose $\mathbf{A}$ is an m-by-n matrix. Theorem 11 says $\text{null}\,(\mathbf{A}) = \text{col}(\mathbf{A}^T)^{\perp}$. Taking the orthogonal complement of both sides, we get

$$\text{null}\,(\mathbf{A})^{\perp} \;=\; (\text{col}(\mathbf{A}^T)^{\perp})^{\perp} \;=\; \text{col}(\mathbf{A}^T)\,,$$

with this last equality holding because of Theorem 14 and the fact that $\text{col}(\mathbf{A}^T)$ is a subspace of $\mathbb{R}^n$.

The other half of this exercise is carried out in the same fashion.

3.14 We start at the point of supposing $(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{0}$, where this is the zero (column) vector in $\mathbb{R}^n$, an n-by-1 matrix. We can multiply it by the 1-by-n matrix $\mathbf{x}^T$, resulting in the value (scalar) zero:

$$\begin{aligned} 0 &= \mathbf{x}^T\mathbf{0} \;=\; \mathbf{x}^T(\mathbf{A}^T\mathbf{A}\mathbf{x}) \\ &= (\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x} \;=\; \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}\rangle \\ &= \|\mathbf{A}\mathbf{x}\|^2\,. \end{aligned}$$

3.15

a) We note that $\mathbf{A}^T\mathbf{A}$ is n-by-n (assuming $\mathbf{A}$ is m-by-n), so $\text{col}(\mathbf{A}^T\mathbf{A})$ is a subspace of $\mathbf{R}^n$. Moreover, by Theorem 11 we have $\text{col}(\mathbf{A}^T\mathbf{A})^{\perp} = \text{null}\left((\mathbf{A}^T\mathbf{A})^T\right) = \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$. Thus, the claim holds by Theorem 13.

b) We have

$$\text{col}(\mathbf{A}^T) \;=\; \text{null}\,(\mathbf{A})^{\perp} \;=\; \text{null}\left(\mathbf{A}^T\mathbf{A}\right)^{\perp} \;=\; \text{col}((\mathbf{A}^T\mathbf{A})^T) \;=\; \text{col}(\mathbf{A}^T\mathbf{A})\,.$$

c) Equation (3.11) has a right-hand side $\mathbf{A}^T\mathbf{b}$ which is automatically in $\text{col}(\mathbf{A}^T)$ for all $\mathbf{b} \in \mathbb{R}^m$. So, by part b), this right-hand side is in $\text{col}(\mathbf{A}^T\mathbf{A})$, which guarantees the existence of some $\mathbf{x} \in \mathbb{R}^n$ maps to $\mathbf{A}^T\mathbf{b}$ under the function $(\mathbf{x} \mapsto \mathbf{A}^T\mathbf{A}\mathbf{x})\colon \mathbb{R}^n \to \mathbb{R}^n$.

3.16

a) The normal equations are $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, where

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 24 & -12 \\ -12 & 6 \end{bmatrix}, \qquad \text{and} \qquad \mathbf{A}^T\mathbf{b} = \begin{bmatrix} -12 \\ 6 \end{bmatrix}.$$

b) The augmented matrix

$$\left[\begin{array}{c|c} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{b} \end{array}\right] = \left[\begin{array}{cc|c} 24 & -12 & -12 \\ -12 & 6 & 6 \end{array}\right] \sim \left[\begin{array}{cc|c} 2 & -1 & -1 \\ 0 & 0 & 0 \end{array}\right].$$

So, we may take the 2nd coordinate of solutions $\mathbf{x} = (x_1, x_2)$ to be free, say $x_2 = t$. Then

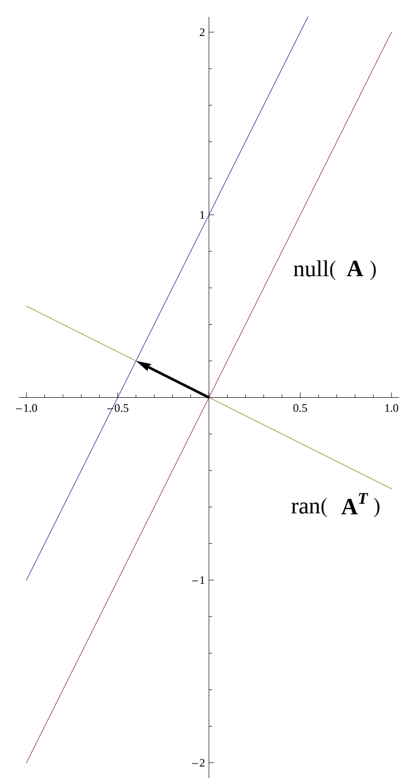
$$2x_1 - t = -1 \quad \Rightarrow \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1/2(t-1) \\ t \end{bmatrix} = = t\begin{bmatrix} 1/2 \\ 1 \end{bmatrix} + \begin{bmatrix} -1/2 \\ 0 \end{bmatrix}.$$

Replacing x_2 with y, x_1 with x, the relationship between x and y is

$$x = \frac{1}{2}(y-1), \qquad \text{or} \qquad y = 2x + 1.$$

c) We know $\text{null}(\mathbf{A}) = \text{null}\left(\mathbf{A}^T\mathbf{A}\right)$. We can read off from the solutions to part (b) just what $\text{null}\left(\mathbf{A}^T\mathbf{A}\right)$ is—we get $\text{null}(\mathbf{A}) = \text{span}(\{(1/2, 1)\})$.

d) We know $\text{col}(\mathbf{A}^T)$ is the orthogonal complement of $\text{null}(\mathbf{A})$. Thus, it is a line through the origin that is perpendicular to the line representing $\text{null}(\mathbf{A})$. So, $\text{col}(\mathbf{A}^T) = \text{span}(\{(2, -1)\})$. The equation for the line in the plane which represents $\text{col}(\mathbf{A}^T)$ has equation $y = (-1/2)x$. So, to find the vector which is both a least-squares solution and lies in $\text{col}(\mathbf{A}^T)$, we find the point of intersection between the lines $y = 2x + 1$ and $y = (-1/2)x$. This yields the vector $(-2/5, 1/5)$.



e) It is pretty clear looking at $\mathbf{A}$ that its two columns are parallel vectors, and so $\text{col}(\mathbf{A}) = \text{span}(\{(1, 2, -1)\})$. Any least-squares solution $\overline{\mathbf{x}}$ from part (b) may be used to find $\mathbf{p}$. In particular,

$$\mathbf{p} \;=\; \mathbf{A}\begin{bmatrix} -1/2 \\ 0 \end{bmatrix} \;=\; \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix}.$$

3.17

a) Perhaps the easiest way is to reduce the augmented matrix to echelon form. We get

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{b} \end{array}\right] \;=\; \left[\begin{array}{ccc|c} 1 & 3 & 1 & 4 \\ 2 & 6 & 2 & 6 \\ 1 & 1 & 1 & 3 \end{array}\right] \;\sim\; \left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

The last row makes it clear the associated system of equations has no solution, since it says "0 = 1".

b) We have

$$\left[\begin{array}{c|c} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{b} \end{array}\right] \;=\; \left[\begin{array}{ccc|c} 6 & 16 & 6 & 19 \\ 16 & 46 & 16 & 51 \\ 6 & 16 & 6 & 19 \end{array}\right] \;\sim\; \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2.9 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The least-squares solutions (of $\mathbf{Ax} = \mathbf{b}$) are

$$(2.9 - t, 0.1, t) \;=\; t(-1, 0, 1) + (2.9, 0.1, 0)\,, \qquad t \in \mathbb{R}\,.$$

c)

```
A =
   1   3   1
   2   6   2
   1   1   1

octave-3.0.0:73> b
b =
   4
   6
   3

octave-3.0.0:74> A \ b
warning: matrix singular to machine precision, rcond = 0
warning: attempting to find minimum norm solution
warning: dgelsd: rank deficient 3x3 matrix, rank = 2
ans =
   1.45000
   0.10000
   1.45000
```

It is easy to see that this answer is among those we described in part (b) (i.e., it is a least-squares solution of $\mathbf{Ax} = \mathbf{b}$). Moreover, one can see, using commands like

```
octave-3.0.0:77> rref([A' [1.45;.1;1.45]])
ans =

    1.00000    2.00000    0.00000   -0.67500
    0.00000    0.00000    1.00000    2.12500
    0.00000    0.00000    0.00000    0.00000
```

that this solution is in the row space of **A**. It is the one least-squares solution that is also in col($\mathbf{A}^T$).

3.18

a) We have $\mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$.

b) If rank($\mathbf{A}$) < 2 then it has linearly dependent columns. This only happens if every one of the x_j's (for the given points) are equal.

c) Using commands like these

```
octave-3.0.0:84> A = [ones(1,11); 9 13 21 30 31 31 34 25 28 20 5]';
octave-3.0.0:85> b = [260 320 420 530 560 550 590 500 560 440 300]';
octave-3.0.0:86> rref([A'*A A'*b])
ans =
      1.00000       0.00000     193.85215
      0.00000       1.00000      11.73128
```

we get that $a_0 \doteq 193.85$ and $a_1 \doteq 11.73$. Thus, our least-squares best-fit line is

$$(\text{Total Calories}) = 193.85 + 11.73\,(\text{Total Fat}) .$$

d) We have

$$\frac{\partial f}{\partial a_0} = \sum_{j=1}^{n} \frac{\partial}{\partial a_0}(y_j - a_0 - a_1 x_j)^2 = -2\sum_{j=1}^{n}(y_j - a_0 - a_1 x_j)$$
$$= 2\left(n a_0 + a_1 \sum_{j=1}^{n} x_j - \sum_{j=1}^{n} y_j\right), \qquad \text{and}$$
$$\frac{\partial f}{\partial a_1} = -2\sum_{j=1}^{n} x_j(y_j - a_0 - a_1 x_j)$$
$$= 2\left(a_0 \sum_{j=1}^{n} x_j + a_1 \sum_{j=1}^{n} x_j^2 - \sum_{j=1}^{n} x_j y_j\right).$$

Thus, the system of equations in a_0, a_1 which arises from setting these partial derivatives equal to zero is

$$\left.\begin{array}{l} n a_0 + a_1(\sum_j x_j) - \sum_j y_j = 0 \\ a_0(\sum_j x_j) + a_1(\sum_j x_j^2) - \sum_j x_j y_j = 0 \end{array}\right\} \quad \text{or} \quad \left\{\begin{array}{l} n a_0 + \alpha_x a_1 = \alpha_y \\ \alpha_x a_0 + \beta_x a_1 = \beta_{xy}, \end{array}\right.$$

where we have defined $\alpha_x = \sum_j x_j$, $\alpha_y = \sum_j y_j$, $\beta_x = \sum_j x_j^2$, and $\beta_{xy} = \sum_j x_j y_j$. Employing Cramer's rule for a_1, we get

$$a_1 = \frac{\begin{vmatrix} n & \alpha_y \\ \alpha_x & \beta_{xy} \end{vmatrix}}{\begin{vmatrix} n & \alpha_x \\ \alpha_x & \beta_x \end{vmatrix}} = \frac{n\beta_{xy} - \alpha_x \alpha_y}{n\beta_x - \alpha_x^2} = \frac{n\left(\sum_j x_j y_j\right) - \left(\sum_j x_j\right)\left(\sum_j y_j\right)}{n\left(\sum_j x_j^2\right) - \left(\sum_j x_j\right)^2}.$$

Substituting this answer into the top equation, we get

$$a_0 = \frac{1}{n}(\alpha_y - a_1 \alpha_x) = \frac{1}{n}\left(\sum_j y_j - a_1 \sum_j x_j\right).$$

e) Assuming we still have our matrix `A` and vector `b` from the Octave code above, we may use commands like

```
octave-3.0.0:87> x = A(:,2);
octave-3.0.0:88> y = b;
octave-3.0.0:89> n = length(x);
octave-3.0.0:92> a1 = (n*sum(x.*y)-sum(x)*sum(y)) / (n*sum(x.*x)-sum(x)^2)
a1 =  11.731

octave-3.0.0:93> a0 = (sum(y) - a1*sum(x)) / n
a0 =  193.85
```

3.19

a) Any m-by-n matrix **A** satisfying rank(**A**) $\geq m$ should be considered correct.

b) Any m-by-n matrix **A** satisfying $m > n =$ rank(**A**) should be considered correct.

c) Any m-by-n matrix **A** whose rank is less than both m and n should be considered correct.

Author Notes