1 Systems of Linear Equations

Recall that a system of linear equations can be solved by the process of Gaussian elimination.

Example 1 Consider the system of equations:

\[
\begin{align*}
x + 2y + 3z &= 14 \\
x + 2y + 2z &= 11 \\
x + 3y + 4z &= 19
\end{align*}
\]

We can eliminate \(x\) from the second and third equations by subtracting the first equation from the second and third equations to get:

\[
\begin{align*}
x + 2y + 3z &= 14 \\
-2y - 2z &= -3 \\
y + z &= 5
\end{align*}
\]

We’d like to get \(y\) into the second equation, so we simply swap the second and third equations:

\[
\begin{align*}
x + 2y + 3z &= 14 \\
y + z &= 5 \\
-2y - 2z &= -3
\end{align*}
\]

Next, we can eliminate \(y\) from the first equation by subtracting two times the second equation from the first equation.

\[
\begin{align*}
x + z &= 4 \\
y + z &= 5 \\
-2y - 2z &= -3
\end{align*}
\]

Next, we multiply the third equation by \(-1\) to get an equation for \(z\).

\[
\begin{align*}
x + z &= 4 \\
y + z &= 5 \\
z &= 3
\end{align*}
\]
Finally, we eliminate $z$ from the first two equations.

\[
\begin{align*}
x &= 1 \\
y &= 2 \\
z &= 3
\end{align*}
\]

The solution to the original system of equations is $x = 1$, $y = 2$, $z = 3$. Geometrically, this system of equations describes three planes (given by the three equations) which intersect in a single point.

In solving this system of equations, we made use of three fundamental operations: adding a multiple of one equation to another equation, multiplying an equation by a nonzero constant, and swapping two equations.

This process can be extended to solve systems of equations with three, four, or more variables. Relatively small systems can be solved by hand, while systems with thousands or even millions of variables must be solved by computer.

In performing the elimination process, the actual names of the variables are insignificant— we could have renamed the variables in the above example to $a$, $b$, and $c$ without changing the solution in any significant way. Since the actual names of the variables are insignificant, we can save time by writing down the significant coefficients from the system of equations in a table called an augmented matrix. This augmented matrix form is also extremely useful in solving a system of equations by computer— the elements of the augmented matrix are simply stored in an array.

In augmented matrix form, our example system becomes:

\[
\begin{bmatrix}
1 & 2 & 3 & | & 14 \\
1 & 2 & 2 & | & 11 \\
1 & 3 & 4 & | & 19
\end{bmatrix}
\]

In this notation, our elementary row operations are adding a multiple of one row to another row, multiplying a row by a nonzero constant, and swapping two rows. The elimination process is essentially identical to the process used in the previous example. In the example, the final version of the augmented matrix is

\[
\begin{bmatrix}
1 & 0 & 0 & | & 1 \\
0 & 1 & 0 & | & 2 \\
0 & 0 & 1 & | & 3
\end{bmatrix}
\]

**Definition 1** A matrix is said to be in reduced row echelon form (RREF) if it has the following properties:

1. The first nonzero element in each row is a one. These elements of the matrix are called pivot elements. The column in which a pivot element appears is called a pivot column.

2. Except for the pivot element, all entries in pivot columns are zero.

3. If the matrix has any rows of zeros, these zero rows are at the bottom of the matrix.
In solving a system of equations in augmented matrix form, we use elementary row operations to reduce the augmented matrix to RREF and then convert back conventional notation to read off the solutions. The process of transforming a matrix into RREF can easily be automated. Most graphing calculators (such as the TI-85 and TI-92) have built in functions for computing the RREF of a matrix. In MATLAB, the \texttt{rref} command computes the RREF of a matrix. In addition to solving equations, the RREF is useful for many other calculations in linear algebra.

It can be shown that any linear system of equations has either no solutions, exactly one solution, or infinitely many solutions. For example, in two dimensions, the lines represented by the equations can fail to intersect (no solution), intersect at a point (one solution) or intersect in a line (many solutions.) In the following examples, we show how it easy is to determine the number of solutions from the RREF of the augmented matrix.

**Example 2** Consider the system of equations:

\[
\begin{align*}
    x + y &= 1 \\
    x + y &= 2
\end{align*}
\]

This system of equations describes two parallel lines. The augmented matrix is:

\[
\begin{bmatrix}
    1 & 1 & | & 1 \\
    1 & 1 & | & 2
\end{bmatrix}
\]

The RREF of this augmented matrix is:

\[
\begin{bmatrix}
    1 & 1 & | & 0 \\
    0 & 0 & | & 1
\end{bmatrix}
\]

In equation form, this is

\[
\begin{align*}
    x + y &= 0 \\
    0 &= 1
\end{align*}
\]

There are obviously no solutions to $0 = 1$, so the original system of equations has no solutions.

**Example 3** In this example, we’ll consider a system of two equations in three variables which has many solutions. Our system of equations is:

\[
\begin{align*}
    x_1 + x_2 + x_3 &= 0 \\
    x_1 + 2x_2 + 2x_3 &= 0
\end{align*}
\]

We put this system of equations into augmented matrix form and then find the RREF

\[
\begin{bmatrix}
    1 & 0 & 0 & | & 0 \\
    0 & 1 & 1 & | & 0
\end{bmatrix}
\]

We can translate this back into equation form as

\[
\begin{align*}
    x &= 0 \\
    y + z &= 0
\end{align*}
\]
Clearly, $x$ must be 0 in any solution to the system of equations. However, $y$ and $z$ are not fixed. We can treat $z$ as a free variable and allow it to take on any value. However, whatever value $z$ takes on, $y$ must be equal to $-z$. Geometrically, this system of equations describes the intersection of two planes. The intersection of the two planes consists of the points on the line $y = -z$ in the $x = 0$ plane.

A linear system of equations may have more equations than variables, in which case the system of equations is over determined. Although over determined systems often have no solutions, it is possible for an over determined system of equations to have many solutions or exactly one solution. Can you construct examples of this?

A system of equations with fewer equations than variables is under determined. Although in many cases, under determined systems of equations have infinitely many solutions, it is possible for an under determined system of equations to have no solutions. Can you construct an example? Is it possible for an under determined system of equations to have exactly one solution?

A system of equations with all zeros on the right hand side is homogeneous. Every homogeneous system of equations has at least one solution- you can set all of the variables to 0 and get the trivial solution. A system of equations with a nonzero right hand side is non homogeneous.

2 Matrix and Vector Algebra

As we have seen in the previous section, a matrix is a table of numbers laid out in rows and columns. A vector is a matrix that consists of a single column of numbers.

There are several important notational conventions used with matrices and vectors. Capital letters such as $A$, $B$, ... are used to denote matrices. Lower case letters such as $x$, $y$, ... are used to denote vectors. Lower case letters or Greek letters such as $\alpha$, $\beta$, ... will be used to denote scalars. At times we will need to refer to specific parts of a matrix. The notation $A_{ij}$ denotes the element of the matrix $A$ in row $i$ and column $j$. We denote the $j$th element of the vector $x$ by $x_j$. The notation $A_j$ is used to refer to column $j$ of the matrix $A$. We can also build up larger matrices from smaller matrices. The notation $A = [B \ C]$ means that the matrix $A$ is composed of the matrices $B$ and $C$, with matrix $C$ beside matrix $B$. The notation $A = [B; \ C]$ means that matrix $A$ is composed of the matrices $B$ and $C$, with $C$ below matrix $B$.

It’s also possible to perform arithmetic on matrices and vectors. If $A$ and $B$ are two matrices of the same size, we can add them together by simply adding the corresponding elements of the two matrices. Similarly, we can subtract $B$ from $A$ by subtracting the elements of $B$ from the elements of $A$. We can multiply a scalar times a vector by multiplying the scalar times each element of the vector. Since vectors are just $n$ by 1 matrices, we can perform the same arithmetic operations on vectors.

A matrix whose elements are all zero plays the same role in matrix algebra
as the scalar 0.

\[ A + 0 = 0 + A = A. \]

**Example 4** Let
\[
A = \begin{bmatrix} 2 & 0 & 4 \\ 1 & 5 & 3 \end{bmatrix}
\]
and
\[
B = \begin{bmatrix} 1 & 5 & 3 \\ 2 & 5 & 3 \end{bmatrix}.
\]

Then
\[
B_2 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}
\]
and
\[ A_{12} = 0. \]

We can add \( A \) and \( B \) to get
\[
A + B = \begin{bmatrix} 3 & 5 & 7 \\ 3 & 10 & 6 \end{bmatrix}
\]

Using the vector notation, we can write a linear system of equations in **vector form**.

**Example 5** Recall the system of equations
\[
\begin{align*}
    x_1 + x_2 + x_3 &= 0 \\
    x_1 + 2x_2 + 2x_3 &= 0
\end{align*}
\]
from example 3. We can write this in vector form as
\[
\begin{bmatrix} 1 \\ 1 \end{bmatrix} x_1 + \begin{bmatrix} 1 \\ 2 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 2 \end{bmatrix} x_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

The expression on the left hand side of the equation in which scalars are multiplied by vectors and then added up is called a **linear combination**. Although this form is not convenient for solving the system of equations, it can be useful in setting up a system of equations.

If \( A \) is a \( m \) by \( n \) vector, and \( x \) is an \( n \) element vector, we can multiply \( A \) time \( x \). The product is defined by
\[
Ax = x_1 A_1 + x_2 A_2 + \ldots + x_n A_n.
\]

**Example 6** Let
\[
A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}
\]
and
\[
x = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}
\]
Then
\[ Ax = 1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 0 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 16 \end{bmatrix} \]

Notice that the formula for \( Ax \) involves a linear combination much like the one that occurred in the vector form of a system of equations. It is possible to take any linear system of equations and rewrite the system of equations as \( Ax = b \), where \( A \) is a matrix containing the coefficients of the variables in the equations, \( b \) is a vector containing the coefficients on the right hand sides of the equations, and \( x \) is a vector containing the variables.

**Example 7** The system of equations
\[
\begin{align*}
  x_1 + x_2 + x_3 &= 0 \\
  x_1 + 2x_2 + 2x_3 &= 0
\end{align*}
\]
can be written as \( Ax = b \), where
\[
A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

If \( A \) is a matrix of size \( m \) by \( n \), and \( B \) is a matrix of size \( n \) by \( r \), then we can define the matrix–matrix product \( C = AB \). To obtain the product, we multiply \( A \) times each of the columns of \( B \) and assemble the matrix vector products into an array. That is,
\[
C = [AB_1 \ AB_2 \ \ldots \ AB_r].
\]

Note that the product is only possible if the two matrices are of compatible sizes—\( B \) must have as many rows as \( A \) has columns. When the matrices are compatible, the product matrix is of size \( m \) by \( r \). In some cases, it is possible to multiply \( AB \) but not \( BA \). Also, it turns out that even when both \( AB \) and \( BA \) exist, \( AB \) is not always equal to \( BA \)!

There is an alternate way to compute the matrix–matrix product. In the row–column expansion method, we obtain the entry in row \( i \) and column \( j \) of \( C \) by computing the matrix product of row \( i \) of \( A \) and column \( j \) of \( B \).

**Example 8** Let
\[
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 5 & 2 \\ 3 & 7 \end{bmatrix}
\]
and let \( C = AB \). The matrix \( C \) will be of size 3 by 2. We compute the product using both of the methods. First, using the matrix–vector approach:

\[
C = [AB_1 \ AB_2]
\]

\[
C = \begin{bmatrix}
5 & 3 \\
3 & 5 \\
5 & 6
\end{bmatrix} + 3 \begin{bmatrix}
2 & 1 \\
4 & 3 \\
5 & 6
\end{bmatrix} + 7 \begin{bmatrix}
2 & 1 \\
4 & 3 \\
5 & 6
\end{bmatrix}
\]

\[
C = \begin{bmatrix}
11 & 16 \\
27 & 34 \\
43 & 52
\end{bmatrix}
\]

Next, we use the row–column approach:

\[
C = \begin{bmatrix}
1 \times 5 + 2 \times 3 & 1 \times 2 + 2 \times 7 \\
3 \times 5 + 4 \times 3 & 3 \times 2 + 4 \times 7 \\
5 \times 5 + 6 \times 3 & 5 \times 2 + 6 \times 7
\end{bmatrix}
\]

\[
C = \begin{bmatrix}
11 & 16 \\
27 & 34 \\
43 & 52
\end{bmatrix}
\]

The \( n \) by \( n \) identity matrix consists of 1’s on the diagonal and 0’s on the off diagonal. For example, the 3 by 3 identity matrix is

\[
I_3 = \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

We often write \( I \) without specifying the size of the matrix in situations where the size of matrix is obvious from context. You can easily show that if \( A \) is an \( m \) by \( n \) matrix, then

\[
AI_n = A
\]

and

\[
I_mA = A.
\]

Thus multiplying by \( I \) in matrix algebra is similar to multiplying by 1 in conventional scalar algebra.

We have not defined division of matrices, but it is possible to define the matrix algebra equivalent of the reciprocal.

**Definition 2** If \( A \) is an \( n \) by \( n \) matrix, and there is a matrix \( B \) such that

\[
AB = BA = I,
\]

then \( B \) is the **inverse** of \( A \). We write \( B = A^{-1} \).

Unfortunately, not all \( n \) by \( n \) matrices have inverses!

How do we compute the inverse of a matrix? If \( AB = I \), then

\[
[AB_1 \ AB_2 \ldots \ AB_n] = I
\]
Since the columns of the identity matrix are known, and $A$ is known, we can solve

$$AB_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

To obtain $B_1$. In the same way, we can find the remaining columns of the inverse. If any of these systems of equations are inconsistent, then $A^{-1}$ does not exist.

**Example 9** Let

$$A = \begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}$$

To find the first column of $A^{-1}$, we solve the system of equations with the augmented matrix

$$\begin{bmatrix} 2 & 1 & | & 1 \\ 5 & 2 & | & 0 \end{bmatrix}$$

The RREF is

$$\begin{bmatrix} 1 & 0 & | & -2 \\ 0 & 1 & | & -5 \end{bmatrix}$$

Thus the first column of $A^{-1}$ is

$$\begin{bmatrix} -2 \\ 5 \end{bmatrix}$$

Similarly, the second column of $A^{-1}$ is

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

Thus

$$A^{-1} = \begin{bmatrix} -2 & 1 \\ 5 & -2 \end{bmatrix}.$$

The inverse matrix can be used to solve an $n$ by $n$ system of equations. Given the system of equations $Ax = b$, and $A^{-1}$, we can calculate

$$Ax = b$$

$$A^{-1}Ax = A^{-1}b$$

$$Ix = A^{-1}b$$

$$x = A^{-1}b$$

This argument shows that if $A^{-1}$ exists, then for any right hand side $b$, the system of equations $Ax = b$ has a unique solution.

**Example 10** Consider the system of equations $Ax = b$, where

$$A = \begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}$$
and 
\[ b = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \]

From the previous example, we know \( A^{-1} \), so
\[ x = A^{-1}b \]
\[ x = \begin{bmatrix} -2 & 1 \\ 5 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \]

**Definition 3** When \( A \) is an \( n \) by \( n \) matrix, \( A^k \) is the product of \( k \) copies of \( A \). By convention, we define \( A^0 = I \).

The **transpose** of a matrix \( A \) is obtained by taking the columns of \( A \) and writing them as the rows of the transpose. The transpose of \( A \) is denoted by \( A^T \).

**Example 11** Let
\[ A = \begin{bmatrix} 2 & 1 \\ 5 & 2 \end{bmatrix}. \]
Then
\[ A^T = \begin{bmatrix} 2 & 5 \\ 1 & 2 \end{bmatrix}. \]

**Definition 4** A matrix is **symmetric** if \( A = A^T \).

### 3 Laws and non-Laws of Matrix Algebra

The following statements are true for any scalars \( s \) and \( t \) and any matrices \( A, B, \) and \( C \). It is assumed that the matrices are of the appropriate size for the operations involved and that whenever an inverse occurs, the matrix is invertible.

1. \( A + 0 = 0 + A = A \).
2. \( A + B = B + A \).
3. \( (A + B) + C = A + (B + C) \).
4. \( A(BC) = (AB)C \).
5. \( A(B + C) = AB + AC \).
6. \( (A + B)C = AC + BC \).
7. \( (st)A = s(tA) \).
8. \( s(AB) = (sA)B = A(sB) \).
9. \( (s + t)A = sA + tA \).
10. \( s(A + B) = sA + sB \).
11. $(A^T)^T = A$.
12. $(sA)^T = s(A^T)$.
15. $(AB)^{-1} = B^{-1} A^{-1}$.
16. $(A^{-1})^{-1} = A$.
17. $(A^T)^{-1} = (A^{-1})^T$.
18. If $A$ and $B$ are $n$ by $n$ matrices, and $AB = I$, then $A^{-1} = B$ and $B^{-1} = A$.

The first ten rules in this list are identical to rules of conventional algebra, and you should have little trouble in applying them. The rules involving transposes and inverses are new, but they can be mastered without too much trouble.

Many students have difficulty with the following statements, which would appear to be true on the surface, but which are in fact false for at least some matrices.

1. $AB = BA$.
2. If $AB = 0$, then $A = 0$ or $B = 0$.
3. If $AB = AC$, then $B = C$.

It is a worthwhile exercise to construct examples of 2 by 2 matrices for which these statements are false.

We conclude with some examples that demonstrate matrix algebra.

**Example 12** Suppose that $A$, $B$, and $C$ are square matrices. Solve

$$C^{-1}(A + X)B^{-1} = I$$

for $X$ in terms of $A$, $B$, and $C$.

$$C^{-1}(A + X)B^{-1} = I.$$  
$$CC^{-1}(A + X)B^{-1} = CI.$$  
$$I(A + X)B^{-1} = C.$$  
$$(A + X)B^{-1}B = CB.$$  
$$(A + X)I = CB.$$  
$$(A + X) = CB.$$  
$$X = CB - A.$$
Example 13 Let $A$ be an $m$ by $n$ matrix, and let

$$P = I - A(A^T A)^{-1}A^T.$$ 

We will assume that $(A^T A)^{-1}$ exists, but not that $A^{-1}$ exists. Show that $P$ is symmetric.

$$P^T = (I - A(A^T A)^{-1}A^T)^T.$$ 
$$P^T = I^T - (A(A^T A)^{-1}A^T)^T.$$ 
$$P^T = I - (A(A^T A)^{-1}A^T)^T.$$ 
$$P^T = I - (A^T)^T ((A^T A)^{-1})^T A.$$ 
$$P^T = I - A((A^T A)^{-1})^T A.$$ 
$$P^T = I - A((A^T A)^{-1})^T A.$$ 
$$P^T = I - A(A^T A)^{-1}A^T.$$ 
$$P^T = P.$$ 

Show that $P^2 = P.$

$$P^2 = (I - A(A^T A)^{-1}A^T)^2.$$ 
$$P^2 = (I - A(A^T A)^{-1}A^T)(I - A(A^T A)^{-1}A^T).$$ 
$$P^2 = I^2 - A(A^T A)^{-1}A^T - A(A^T A)^{-1}A^T + A(A^T A)^{-1}A^T A(A^T A)^{-1}A^T.$$ 
$$P^2 = I - 2A(A^T A)^{-1}A^T + A(A^T A)^{-1}(A^T A)(A^T A)^{-1}A^T.$$ 
$$P^2 = I - 2A(A^T A)^{-1}A^T + A(A^T A)^{-1}A^T.$$ 
$$P^2 = I - A(A^T A)^{-1}A^T.$$ 
$$P^2 = P.$$