Lecture 3 - Vectors and Matrices

Last time we saw that if we have n equations in n unknowns then there are n^2 coefficients (some may be zero) and n right hand side components. To efficiently study linear systems we need to write all linear systems in a generic form. To do this we need to review vectors and matrices. Once we write our linear system as a matrix problem, then we can view Gauss elimination in terms of matrices. Throughout we will assume that the entries of our vectors and matrices are *real*; the results can be easily extended to the situation where the entries are complex.

Vectors

To sketch $\mathbf{R}^2$ (Euclidean space in two dimensions) we indicate the origin and the x and y axes. Then any point can be represented as the *ordered pair* (x_1, x_2) which we can associate with a *vector* $\vec{x}$ starting at the origin (0,0) and ending at the point (x_1, x_2). In this case the vector $\vec{x}$ has a direction and a magnitude. In algebra, we calculated the length of a vector by using the standard Euclidean distance, i.e.,

$$\sqrt{x_1^2 + x_2^2}\,.$$

We call a vector which has length one a *unit vector*. We can think of $\mathbf{R}^2$ as the set of pairs (x_1, x_2) or equivalently all vectors with two components.

In $\mathbf{R}^n$ we have n dimensions so a point is represented by the ordered tuple $(x_1, x_2, x_3, \ldots, x_n)$ and we can associate a vector $\vec{x}$ as emanating from the origin and terminating at this point. $\mathbf{R}^n$ is the set of all n-tuples or equivalently all n-vectors. When we solve a system of n equations in n unknowns then there are n values for the right hand sides and n unknowns so these will be stored as vectors.

We will often use $\vec{i}$, $\vec{j}$, $\vec{k}$ as notation for unit vectors in the x, y and z directions. This means they have length one and lie along a coordinate axis.

- How do we perform standard operations with vectors such as scalar multiplication, addition/subtraction and multiplication?

Scalar multiplication means we are multiplying our vector $\vec{a}$ by a number, say k, and each component of the vector is multiplied by k. We have

$$k\vec{a} = k \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} k\ a_1 \\ k\ a_2 \end{pmatrix}$$

If we think about this in $\mathbf{R}^2$ we realize that we are just changing its Euclidean length by the magnitude of k, i.e., the length of $k\vec{a}$ is $|k|$ times the length of $\vec{a}$. To see this the length of $k\vec{a}$ is

$$\sqrt{(k\ a_1)^2 + (k\ a_2)^2} = \sqrt{k^2\left[(a_1)^2 + (a_2)^2\right]} = |k|\sqrt{a_1^2 + a_2^2}\,.$$

Multiplying a vector by -1 does not change its length but it changes its direction.

addition/subtraction of two vectors is done in the usual manner, i.e., addition/subtraction of corresponding components. We should note that addition only makes sense if the two vectors are of the same length. Because addition and scalar multiplication are performed in the standard ways, the usual properties such as commutative, etc. hold. For example,

$$\vec{x} + \vec{y} = \vec{y} + \vec{x} \qquad \alpha\vec{x} + \beta\vec{x} = (\alpha + \beta)\vec{x} \qquad \alpha(\vec{x} + \vec{y}) = \alpha\vec{x} + \alpha\vec{y}$$

Example Which of the following are defined? If defined, determine the result of the given operation. Here

$$\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} 0 \\ -4 \end{pmatrix} \qquad \vec{c} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}$$

(i) the length of 10 $\vec{c}$

(ii) $2\vec{a} - \vec{b}$

(iiii) $3(\vec{c} - \vec{b})$

For (i) the length of $10\vec{c}$ is just ten times the length of $\vec{c}$ which is $\sqrt{15+9} = 5$ so the answer is 50. For (ii) $\vec{a}$ is a vector in $\mathbf{R}^2$ and $\vec{b}$ is a vectors in $\mathbf{R}^3$ so the operation is NOT defined. For (iii) $\vec{c}, \vec{b}$ are vectors in $\mathbf{R}^2$ so the operation is defined. $\vec{c} - \vec{b} = (-4, 7)^T$ so three times this is $(-12, 21)^T$.

Multiplication of vectors is different than multiplying two numbers. We can "multiply" two vectors in two ways - in one (the *dot or scalar product*) the result is a number and in the other (*cross product*)the result is a vector. Recall that in $\mathbf{R}^2$ when we took the dot product of two vectors we multiplied corresponding components and added to get

$$\vec{x} \cdot \vec{y} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + x_2 y_2$$

The same is true in $\mathbf{R}^n$

$$\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} x_i y_i \,.$$

Thus in order for the operation of scalar product to be defined, the vectors have to have the same number of components. Note also that

$$\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$$

In vector calculus you learned an equivalent definition of dot product

$$\vec{x} \cdot \vec{y} = (\text{magnitude of } \vec{x})(\text{magnitude of } \vec{y}) \cos\theta$$

where θ is the angle between the two vectors and we use the standard Euclidean length for the magnitude. Because the $\cos\pi/2 = 0$ we immediately see that two vectors are *perpendicular or orthogonal* if their dot product is zero.

Note that if we take the scalar product of a vector with itself then the result is the square of its Euclidean length; i.e., in R^2

$$\vec{x}\cdot\vec{x} = x_1^2 + x_2^2 = \left[\sqrt{x_1^2 + x_2^2}\,\right]^2$$

So the Euclidean length of a vector in R^n can be written as

$$\sqrt{\vec{x}\cdot\vec{x}} = \Big[\sum_{i=1}^{n} x_i^2\Big]^{1/2}.$$

In general, we think of a vector $\vec{x}$ as a *column vector*, i.e.,

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Sometimes it is useful to use a *row vector*, i.e., $(x_1, x_2, \ldots, x_n)$. We write this row vector as $\vec{x}^T$ where the "T" means transpose. Because in written text it is easier to type a row vector we often write, e.g., $(x_1, x_2)^T$ to mean a column vector in R^2.

a second way to "multiply" vectors is the cross product which results in a vector. We will look at its definition in a later example.

Example Determine the following. Here

$$\vec{a} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} \qquad \vec{c} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

(i) $\vec{a}\cdot 2\vec{b}$

(ii) $3\vec{c}^T$

(iii) are $\vec{b}$, $\vec{c}$ orthogonal vectors?

all operations are defined because all are vectors in R^3. For (i) $\vec{a}\cdot 2\vec{b}$ is $\vec{a}\cdot(0,-8,2)^T$ which is $2\cdot 0 + (1)(-8) + 0(2) = 8$. For (ii) we have $(3, 0, 0)$, a row vector. For (iii) the vectors are orthogonal because $\vec{b}\cdot\vec{c} = 0$.

Matrices

Recall that we have n^2 coefficients in our system so we need to store this information. To do this, we introduce matrices which are rectangular arrays of numbers. We say A is an $m\times n$ matrix if it has m rows and n columns. If the entries of A are denoted a_{ij} where

i refers to the row and j to the column, then an $m \times n$ matrix A is written componentwise as

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix}$$

an $m \times n$ matrix has mn entries so we can store our coefficients in an $n \times n$ matrix. Note that an n-vector could be viewed as an $n \times 1$ matrix. a row of a matrix is a row vector and a column is a column vector.

Some matrices which have special structure are given individual names. The zero matrix is simply what the name implies, a matrix with all zero entries. The *diagonal* entries of a matrix are the entries a_{ii}. a *diagonal matrix* is one which $a_{ij} = 0$ for all $i \neq j$; e.g., a 3×3 diagonal matrix has the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

The *identity matrix*, usually denoted I, is a diagonal matrix whose diagonal entries are all ones. an *upper triangular* matrix is one where $a_{ij} = 0$ for $j < i$; e.g., a 3×3 upper triangular matrix has the form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}$$

Note that by this definition a diagonal matrix is also an upper triangular matrix. Similarly a *lower triangular* matrix is one where $a_{ij} = 0$ for $i < j$. Sometimes we will need a *unit* lower or upper triangular matrix; these are just special lower or upper triangular matrices which have ones as the diagonal entries; e.g., a unit 3×3 lower triangular matrix is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix}$$

We can also take the transpose of a matrix. A^T means to reflect the matrix around the diagonal so if $B = A^T$ then $b_{ij} = a_{ji}$. Note that the diagonal entries are unchanged. If the original matrix is not square, i.e., $m \times n$ then the transpose is $n \times m$. For example,

$$AA^T = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

a matrix which has the property that $A = A^T$ is called *symmetric*; clearly a symmetric matrix must be square. We can take the transpose of the product of two matrices for which multiplication is defined. If A is $m \times n$ and B is $n \times p$ then the product AB is $m \times p$ and its transpose is $p \times m$. We can use the following formula for computing $(AB)^T$; note that B^T is $p \times n$ and a^T is $n \times m$ so $B^T A^T$ is $p \times m$:

$$(AB)^T = B^T A^T$$

- How could you describe a matrix which is both lower and upper triangular?

addition and scalar multiplication of matrices is done in the standard way just as we did for vectors. To multiply a matrix by a scalar k we simply multiply each entry by k. To add two matrices, first they must have the same number of rows and columns, then we simply add corresponding components. Because these operations are performed in the standard way, we have the usual laws holding; e.g.,

$$A + B = B + A \qquad \alpha(A + B) = \alpha A + \alpha B$$

To define matrix multiplication, we could define it by multiplying corresponding entries. However, our goal is to use a matrix and two vectors to represent our linear system. Consequently, we need to define matrix multiplication in a meaningful way for this application.

We first look at the definition of an $m \times n$ matrix A times an $n \times p$ matrix B. For matrix multiplication to be defined, the number of columns of the first matrix must equal the number of rows of the second matrix. We have

$$A_{m \times n} B_{n \times p} = C_{m \times p} \quad \text{where} \quad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

that is, we can view this entry as taking the dot product of the ith row of A and the kth column of B.

Example Determine AB and BA, if defined, where

$$A = \begin{pmatrix} 3 & 2 \\ 1 & -1 \\ 0 & 4 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 1 \\ -2 & 1 \end{pmatrix}$$

First $C = AB$ is defined because A has two columns and B has two rows. Then $C = AB$ has three rows and two columns and is given by

$$C = \begin{pmatrix} 3 * 0 + 2 * (-2) & 3 * 1 + 2 * 1 \\ 1 * 0 + (-1)(-2) & 1 * 1 + (-1) * 1 \\ 0 * 0 + 4 * (-2) & 0 * 1 + 4 * 1 \end{pmatrix} = \begin{pmatrix} -4 & 5 \\ 2 & 0 \\ -8 & 4 \end{pmatrix}$$

Now $D = BA$ is not defined because B has two columns and A has three rows. This example shows us that, in general,

$$AB \neq BA$$

in fact, if AB is defined BA many not be. Sometimes we use the terminology *premultiply* B by A to mean AB and the terminology *post-multiply* B by A to mean BA.

Example Let I be the $n \times n$ identity matrix and A an $n \times n$ matrix. What is AI? IA? Clearly pre- or post-multiplying a matrix by the identity matrix has no effect. Consider pre-multiplying the $m \times n$ matrix A by the $m \times m$ identity matrix

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{pmatrix} = A$$

Example If A, B are square, i.e., $n \times n$ then both AB and BA are defined. If $AB = BA$ then we say that the matrices *commute*. Do all square matrices commute?

Clearly no. as a counterexample consider

$$\begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ -2 & 2 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 2 & 4 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ -1 & 3 \end{pmatrix}$$

Example What is the effect of premultiplying the given 3×3 matrix A by the matrix

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{where} \quad A = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 1 & 0 \\ -6 & 0 & 2 \end{pmatrix}?$$

Note that C is the identity matrix with the last two rows interchanged. The effect is to interchange the last two rows of A which can be seen by direct multiplication.

Example What is the effect of premultiplying the given 3×3 matrix A by the matrix

$$\mathcal{M} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \quad \text{where} \quad a = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 1 & 0 \\ -6 & 0 & 2 \end{pmatrix}?$$

The matrix $\mathcal{M}a$ is a matrix with the first row of A the same (because the first row of $\mathcal{M}$ is the same as in identity matrix) and zeros have been introduced into the $(2,1)$ and $(3,1)$ entries, i.e.,

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 3 \\ 4 & 1 & 0 \\ -6 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 3 \\ 0 & -1 & -6 \\ 0 & 3 & 11 \end{pmatrix}$$

To write our linear system as a matrix equation we will have a matrix times a vector. Because we can view an n-vector as simply an $n \times 1$ matrix we already know how to do this. If A is an $n \times n$ which contains our n^2 coefficients, $\vec{x}$ is an n-vector of the unknowns and $\vec{b}$ is the right hand side terms, then we have $A\vec{x} = \vec{b}$. The ith row of A contains the coefficients (in order) in the ith equation because the ith component of the vector $a\vec{x}$ is

$$(A\vec{x})_i = \sum_{j=1}^{n} a_{ij} x_j$$

We can also view our system using the columns of A which we denote by $\vec{a}_1, \vec{a}_2, \ldots \vec{a}_n$. Then solving the linear system is equivalent to finding $\vec{x}$ such that

$$x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3 + \cdots + x_n\vec{a}_n = \vec{b}$$

Example Form $A\vec{x}$ where

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Using matrix multiplication we have $A\vec{x}$ is the vector

$$A\vec{x} = \begin{pmatrix} 2x_1 + x_2 + x_3 \\ 4x_1 - 6x_2 \\ -2x_1 + 7x_2 + 2x_3 \end{pmatrix}$$

Example Use your result in the previous example to write the linear system

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 1 \\ 4x_1 - 6x_2 &= 2 \\ -2x_1 + 7x_2 + 2x_3 &= 3 \end{aligned}$$

as a matrix equation $a\vec{x} = \vec{b}$. Then write the equation in column form.
Clearly we have

$$\begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

So now we can see the reason that matrix multiplication was defined in this way. We can also view the linear system as finding $\vec{x}$ such that

$$x_1 \begin{pmatrix} 2 \\ 4 \\ -2 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ -6 \\ 7 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

that is, we seek a combination of the columns of A (which themselves are vectors) which forms the right hand side. This viewpoint is especially useful when we talk about the solvability of $A\vec{x} = \vec{b}$.

For an $n \times n$ matrix A we may be able to define an $n \times n$ matrix B which has the property that

$$AB = BA = I$$

where I is the $n \times n$ identity matrix. If this holds, then B is called the *inverse* of A. If such a matrix exists, then we denote it by A^{-1}. Please realize that this is notation, it does not mean $1/a$ because this doesn't make sense. If A^{-1} exists then

$$AA^{-1} = A^{-1}A = I$$

Why is A^{-1} important? Because

$$A\vec{x} = \vec{b} \Rightarrow A^{-1}A\vec{x} = A^{-1}\vec{b} \Rightarrow I\vec{x} = A^{-1}\vec{b} \Rightarrow \vec{x} = A^{-1}\vec{b}$$

So analytically, if we have A^{-1} then we can obtain the solution by performing a matrix times vector operation to get $A^{-1}\vec{b}$. It turns out that computationally this is NOT an efficient way to find the solution; we will address this later.

A matrix whose inverse is its transpose is called *orthogonal*; that is $A^{-1} = A^T$

$$\text{if } AA^T = A^TA = I \text{ then } A \text{ is orthogonal.}$$

This tells us that if A is orthogonal then the linear system $A\vec{x} = \vec{b}$ is readily solved by forming $\vec{x} = A^T\vec{b}$; i.e., we only have to perform a matrix times vector operation.

Example Let

$$A = \begin{pmatrix} 2 & 4 \\ -1 & -3 \end{pmatrix}, \quad A^{-1} = \frac{-1}{2}\begin{pmatrix} -3 & -4 \\ 1 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \quad B^{-1} = \frac{1}{2}\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

Demonstrate that $AA^{-1} = I$ and $BB^{-1} = I$, then form AB, its inverse and the product $B^{-1}A^{-1}$.

Now AB and its inverse are given by

$$AB = \begin{pmatrix} 2 & 6 \\ -1 & -5 \end{pmatrix} \qquad (AB)^{-1} = \frac{-1}{4}\begin{pmatrix} -5 & -6 \\ 1 & 2 \end{pmatrix}$$

and $B^{-1}A^{-1}$ is

$$B^{-1}A^{-1} = \frac{-1}{4}\begin{pmatrix} -5 & -6 \\ 1 & 2 \end{pmatrix} = (AB)^{-1}$$

In the last example we saw that the inverse of the product of two specific matrices is the product of their inverses with the order of multiplication reversed. This is true in general.

$$\text{Let } AA^{-1} = I,\ BB^{-1} = I \text{ then } \quad \left(AB\right)^{-1} = B^{-1}A^{-1}$$

The Matrix Form of Gauss Elimination (GE)

Now that we have written our linear system of equations as a matrix problem $A\vec{x} = \vec{b}$, we want to describe our GE algorithm in terms of matrix operations. First, recall an example that we did in the previous lecture.

$$\begin{aligned} 2x + y + z &= 5 \\ 4x - 6y &= -2 \\ -2x + 7y + 2z &= 9 \end{aligned} \quad \Rightarrow \quad \begin{aligned} 2x + y + z &= 5 \\ 0 - 8y - 2z &= -12 \\ 0 + 8y + 3z &= 14 \end{aligned} \quad \Rightarrow \quad \begin{aligned} 2x + y + z &= 5 \\ 0 - 8y - 2z &= -12 \\ z &= 2 \end{aligned}$$

Lets concentrate on the left hand side for now and write the coefficient matrices for the three systems above.

$$\begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

We first recognize that the last matrix is upper triangular so the goal in GE is to convert the original system to an *equivalent* upper triangular system because we saw that this type of system is easy to solve. What we want to do now is take the process we used to eliminate x from the second and third equations in the first step of GE above and write it as a matrix times A; e.g., find $\mathcal{M}^1$ such that

$$\mathcal{M}^1 \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{pmatrix}$$

Now we want $\mathcal{M}^1$ as simple as possible. If we return to a previous example, we can guess what $\mathcal{M}^1$ should be. We have

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{pmatrix}$$

Our $\mathcal{M}^1$ is just a modified identity matrix where in the first column we have included the factors we multiplied the first equation by so that when we added it to another equation it removed the first variable. Note that $\mathcal{M}^1$ is unit lower triangular.

Our next step is to find $\mathcal{M}^2$ such that

$$\mathcal{M}^2 \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

Now $\mathcal{M}^2$ must have the property that the first two rows remain unchanged but we know that multiplication by the identity matrix doesn't change anything. Clearly the first two rows of $\mathcal{M}^2$ must be the identity matrix. Now we modify the (3,2) entry of $\mathcal{M}^2$ so that we introduce a zero into the (3,2) entry . We have

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

because our multiplier in the equation was one. We have converted the original matrix A to

$$\mathcal{M}^2\mathcal{M}^1 A = U$$

where U is an upper triangular matrix and each $\mathcal{M}^i$ is unit lower triangular.

We multiplied the left hand side of our equation first by $\mathcal{M}^1$ and then by $\mathcal{M}^2$ so we have to do the same thing to both sides of the equations:

$$\mathcal{M}^2\mathcal{M}^1 A\vec{x} = \mathcal{M}^2\mathcal{M}^1\vec{b}$$

Notice that the order here is important because, in general, matrix multiplication is not commutative. Now in our example if we multiply our vector $\vec{b}$ by $\mathcal{M}^1$ and then by $\mathcal{M}^2$ we should get $(5, -12, 2)^T$. You should verify this.

If we have an $n \times n$ system then $\mathcal{M}^1$ has the form

$$\mathcal{M}^1 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ m_{21}^1 & 1 & 0 & \cdots & 0 \\ m_{31}^2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{n1} & 0 & 0 & \cdots & 1 \end{pmatrix}$$

where

$$m_{21}^1 = -a_{21}/a_{11} \qquad m_{31}^1 = -a_{31}/a_{11} \qquad m_{i1}^1 = -a_{i1}/a_{11}$$

Recall that a_{11} is called our first pivot and must be nonzero for the algorithm to work. $\mathcal{M}^1$ is a unit lower triangular matrix but moreover, it is the identity matrix modified to have nonzero entries only in the first column below the diagonal. $\mathcal{M}^2$ is a unit lower triangular matrix but moreover, it is the identity matrix modified to have nonzero entries only in the second column below the diagonal. Note that the entries of $\mathcal{M}^2$ in the second column below the diagonal are determined by the pivot in the modified matrix $A^1 = \mathcal{M}^1 A$.

- If A is the 10×10 coefficient matrix for a linear system (with ten unknowns) that has a unique solution, what is the maximum number of matrices $\mathcal{M}^i$ that we must determine to transform A into an upper triangular matrix?

The matrices $\mathcal{M}^i$ are called *elementary matrices* or *Gauss transformation matrices*. Because they differ from the identity matrix in only one column their inverse can be easily determined. From our example above

$$\mathcal{M}^1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \qquad \left[\mathcal{M}^1\right]^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

and

$$\mathcal{M}^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \qquad \left[\mathcal{M}^2\right]^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

You should verify this by multiplication. So we can immediately obtain the inverse of $\mathcal{M}^k$ by multiplying our entries in the kth column below the diagonal by -1. So the inverse of an elementary matrix is also a unit lower triangular matrix and it differs from the identity by entries in one column below the diagonal.

Example Find the Gauss transformation matrices $\mathcal{M}^1$ and $\mathcal{M}^2$ which converts the given matrix to an upper triangular matrix; give this upper triangular matrix. If the right hand side of the corresponding linear system is $\vec{b} = (-1, 0, -2)^T$, what is the resulting upper triangular system that needs to be solved? Here

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 3 & 4 & 7 \\ -2 & 0 & 1 \end{pmatrix}$$

We have

$$\mathcal{M}^1 A = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 3 & 4 & 7 \\ -2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 7 & 7 \\ 0 & -2 & 1 \end{pmatrix}$$

and

$$\mathcal{M}^2(\mathcal{M}^1 A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2/7 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 7 & 7 \\ 0 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 7 & 7 \\ 0 & 0 & 3 \end{pmatrix} = U$$

The resulting system is then

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 7 & 7 \\ 0 & 0 & 3 \end{pmatrix} \vec{x} = \mathcal{M}^2 \mathcal{M}^1 \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & 2/7 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -4 \end{pmatrix}$$

So far, we have constructed Gauss transformation matrices $\mathcal{M}^k$ such that

$$\mathcal{M}^q \mathcal{M}^{q-1} \cdots \mathcal{M}^2 \mathcal{M}^1 A = U$$

where U is an upper triangular matrix. Our system $A\vec{x} = \vec{b}$ becomes

$$\mathcal{M}^q \mathcal{M}^{q-1} \cdots \mathcal{M}^2 \mathcal{M}^1 A\vec{x} = \mathcal{M}^q \mathcal{M}^{q-1} \cdots \mathcal{M}^2 \mathcal{M}^1 \vec{b} \Rightarrow U\vec{x} = \mathcal{M}^q \mathcal{M}^{q-1} \cdots \mathcal{M}^2 \mathcal{M}^1 \vec{b}.$$

Because we can immediately write down the inverse of each $\mathcal{M}^k$ we can write

$$A = \left[\left(\mathcal{M}^1\right)^{-1} \left(\mathcal{M}^2\right)^{-1} \cdots \left(\mathcal{M}^{q-1}\right)^{-1} \left(\mathcal{M}^q\right)^{-1} \right] U$$

which says that A can be written as the product of unit lower triangular matrices times an upper triangular matrix. If we can show that the product of unit lower triangular matrices is itself a unit lower triangular matrix, then we have shown that GE is equivalent to writing

$$A = LU$$

where L is *unit lower triangular* matrix and U is upper triangular. We will see later how this interpretation of GE can lead to an algorithm which is equivalent to GE on paper but its implementation can be much more efficient in some situations.

Example Perform the product $\left[\mathcal{M}^1\right]^{-1}\left[\mathcal{M}^2\right]^{-1}$ where these are the matrices above.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}$$

We notice that after multiplying these two matrices we see a definite pattern to the product. Is this true in general? Let's do a product $\left[\mathcal{M}^1\right]^{-1}\left[\mathcal{M}^2\right]^{-1}$ for general entries

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \nu_3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ \mu_2 & 1 & 0 \\ \mu_3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \mu_2 & 1 & 0 \\ \mu_3 & \nu_3 & 1 \end{pmatrix}$$

Of course this doesn't prove that it is true in general but you can convince yourselves that *a product of lower triangular matrices is lower triangular* and a *product of unit lower triangular matrices is unit lower triangular.*

Of course if in the process of constructing $\mathcal{M}^k$ we find a zero pivot, i.e., $a_{kk}^k = 0$ then the method fails because we are dividing by zero in our formula for the entries of $\mathcal{M}^k$. Previously, we found that we could interchange the equations to eliminate a zero (or small) pivot. We now want to determine how interchanging rows of our matrix can be done by premultiplying by some other matrix. If you recall we had an example where we premultiplied a matrix by a matrix which looked like the identity matrix but the last two rows were interchanged. Premultiplying by this matrix had the effect of interchanging the last two rows.

Permutation matrices are $n \times n$ matrices which are formed by interchanging rows (or columns) of the identity matrix. For example,

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are permutation matrices. Premultiplying a 3×3 matrix by P interchanges the first and third rows; premultiplying a 3×3 matrix by Q interchanges the first and second rows.

- What does post-multiplying a matrix by a permutation matrix do?

If we include permutation matrices our GE process can be written as

$$\mathcal{M}^q P^q \mathcal{M}^{q-1} P^{q-1} \cdots \mathcal{M}^2 P^2 \mathcal{M}^1 P^1 A = U$$

Of course if row interchanges are not needed then P^k is just the identity matrix. One can demonstrate that

$$\left[(\mathcal{M}^q)^{-1} (P^q)^{-1} (\mathcal{M}^{q-1})^{-1} (P^{q-1})^{-1} \cdots (\mathcal{M}^2)^{-1} (P^2)^{-1} (\mathcal{M}^1)^{-1} (P^1)^{-1} \right]$$

is also unit lower triangular so that we still have $A = LU$.

Back Solving

We have seen that GE can be viewed as converting our original problem $A\vec{x} = \vec{b}$ into an *equivalent* upper triangular system $U\vec{x} = \vec{c}$. In our examples, we saw that once we have an upper triangular system we can easily solve it by starting with the last variable and solving until we get to the first. This process is called a *back solve.* We can easily describe the algorithm for a back solve. First, lets do a specific example to remind ourselves.

Example Solve the upper triangular system

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -12 \\ 2 \end{pmatrix}$$

Clearly $x_3 = 2$; because $-8x_2 - 2x_3 = -12$ we have $-8x_2 = 8$ which implies $x_2 = 1$. Lastly, substituting our values for x_2, x_3 into $2x_1 + x_2 + x_3 = 5$ implies $2x_1 = 5 - 1 - 2 = 2$ or $x_1 = 1$.

Now consider the general upper triangular system

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & u_{24} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & u_{34} & \cdots & u_{3n} \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & u_{n-1,n-1} & u_{n-1,n} \\ 0 & 0 & 0 & 0 & \cdots & u_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{n-1} \\ b_n \end{pmatrix}$$

To write the equations we just equate entries on the vectors of the right and left sides of the equation; recall that $U\vec{x}$ is itself a vector. We have

$$x_n = \frac{b_n}{u_{nn}}$$

Then to obtain x_{n-1} we equate the $n-1$ component

$$u_{n-1,n-1}x_{n-1} + u_{n-1,n}x_n = b_{n-1} \Rightarrow x_{n-1} = \frac{b_{n-1} - u_{n-1,n}x_n}{u_{n-1,n-1}}$$

For x_{n-2} we equate the $n-2$ component

$$u_{n-2,n-2}x_{n-2}+u_{n-2,n-1}x_{n-1}+u_{n-2,n}x_n = b_{n-2} \Rightarrow x_{n-2} = \frac{b_{n-2} - u_{n-2,n-1}x_{n-1} - u_{n-2,n}x_n}{u_{n-2,n-2}}$$

In general, we can find the ith component of $\vec{x}$ for $i < n$ by

$$x_i = \frac{b_i - \sum_{j=i+1}^{n} u_{i,j}x_j}{u_{ii}}$$

We can write the general algorithm in the following manner using *pseudo-code*. Note that we want to write it in such a way that another person could take the algorithm and code it in the language of their choice.

Given an $n \times n$ upper triangular matrix U with entries u_{ij} and an n-vector $\vec{b}$ with components b_i then the solution of $U\vec{x} = \vec{b}$ is given by the following algorithm.

Set $x_n = \dfrac{b_n}{u_{nn}}$

For $i = n-1, n-2, \ldots, 1$

$$x_i = \frac{b_i - \sum_{j=i+1}^{n} u_{i,j}x_j}{u_{ii}}$$