Some Notes to Supplement the Lectures in Mathematics 700

1 The Uniqueness of the Row Reduced Echelon Form

Let A and B be two matrices of the same size and assume that they are row equivalent. Then there are elementary matrices $E_1, \dots, E_k$ so that $B = E_k E_{k-1} \cdots E_1 \cdots E_k A$. Then as all the matrices $E_1, \dots, E_k$ are invertible so is the product $P := E_k E_{k-1} \cdots E_1$. Therefore if A and B are row equivalent matrices there is an invertible matrix P so that $B = PA$. If we write $B \precapprox A$ to mean there is an invertible matrix P so that $B = PA$. Then this is an equivalence relation. (For letting $P = I$ shows $A \precapprox A$. If $B = PA$, then $A = P^{-1}B$ so $A \precapprox B$ implies $B \precapprox A$. If $B = PA$ and $C = QB$ then $C = QPA$ so $A \precapprox B$ and $B \precapprox C$ implies $A \precapprox B$. Of course we know from class work that this equivalence relation in just the same as row equivalence, but I don't want to use this fact in the following proof as then we can use the argument here to give an new proof of this. A virtue of the proof here is that it gives an example of how to use block matrices to prove things. It has several vices, such as being more computational than conceptual.

Theorem 1.1 *If R and R' are row reduced echelon matrices of the same size so that for some invertible matrix P there holds $R' = PR$, then $R = R'$. In particular if R and R' are row equivalent then they are equal.*

Proof: Assume that R and R' are both $m \times n$. We will use induction and assume that the result holds for row reduced echelon matrices with fewer than m rows and show that it holds when they have m rows. We write

$$R = \begin{bmatrix} R_1 \\ \vdots \\ R_m \end{bmatrix}, \qquad R' = \begin{bmatrix} R'_1 \\ \vdots \\ R'_m \end{bmatrix}, \qquad P = \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix}$$

where $R_1, \dots, R_m$ are the rows of R and $R'_1, \dots, R'_m$ are the rows of R. Then $R' = PR$ implies $R'_i = \sum_{j=1}^{m} p_{ij} R_j$. That is the rows of R' are linear combinations of the rows of R. By symmetry of the equivalence relation $\precapprox$ this also implies that the rows of R are linear combinations of the rows of R'. If $R = 0$ then also $R' = PR = 0$ and we are done. So we can assume $R \neq 0$. But the first row of a nonzero row reduced echelon matrix is nonzero. Assume the leading one of the first row of R_1 has occurs in the k-th place. Then the rows of R' are all linear combinations of the rows of R the leading one of the first row of R' also occurs in the k-th place. We now write R and R' in block form:

$$R = \left[\begin{array}{ccc|c} 0 & \cdots & 1 & R_{12} \\ \hline 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & R_{22} \\ 0 & \cdots & 0 & \end{array}\right], \qquad R' = \left[\begin{array}{ccc|c} 0 & \cdots & 1 & R'_{12} \\ \hline 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & R'_{22} \\ 0 & \cdots & 0 & \end{array}\right]$$

where R_{12} and R'_{12} are of size $1 \times (n-k)$ and R_{22} and R'_{22} are $(m-1) \times (n-k)$. Also write P in block from

$$P = \left[\begin{array}{c|c} p_{11} & P_{12} \\ \hline P_{21} & P_{22} \end{array}\right] = \left[\begin{array}{c|ccc} p_{11} & p_{12} & \cdots & p_{1m} \\ \hline p_{21} & & & \\ \vdots & & P_{22} & \\ p_{m1} & & & \end{array}\right] \text{ where } P_{12} = [p_{12}, \ldots, p_{1m}],\ P_{21} = \begin{bmatrix} p_{21} \\ \vdots \\ p_{m1}, \end{bmatrix}$$

P_{22} is $(m-1) \times (m-1)$ and $P_{11} = p_{11}$ is 1×1 (that is just a scalar). Now writing out $PR = R'$ in block form

$$\begin{aligned} PR &= \left[\begin{array}{c|c} p_{11} & P_{12} \\ \hline P_{21} & P_{22} \end{array}\right] \left[\begin{array}{ccc|c} 0 & \cdots & 1 & R_{12} \\ \hline 0 & \cdots & 0 & \\ \vdots & \ddots & \vdots & R_{22} \\ 0 & \cdots & 0 & \end{array}\right] \\ &= \left[\begin{array}{cccc|c} 0 & \cdots & 0 & p_{11} & p_{11}R_{12} + P_{12}R_{22} \\ \hline 0 & \cdots & 0 & p_{21} & \\ \vdots & \ddots & \vdots & \vdots & P_{21}R_{12} + P_{22}R_{22} \\ 0 & \cdots & 0 & p_{m1} & \end{array}\right] \\ &= R' = \left[\begin{array}{cccc|c} 0 & \cdots & 0 & 1 & R'_{12} \\ \hline 0 & \cdots & 0 & 0 & \\ \vdots & \ddots & \vdots & \vdots & R'_{22} \\ 0 & \cdots & 0 & 0 & \end{array}\right] \end{aligned}$$

This implies $p_{11} = 1$ and $p_{21} = p_{31} = \cdots = p_{m1} = 0$. That is $P_{21} = 0$. Putting this information back into the equation gives

$$PR = \left[\begin{array}{cccc|c} 0 & \cdots & 0 & 1 & R_{12} + P_{12}R_{22} \\ \hline 0 & \cdots & 0 & 0 & \\ \vdots & \ddots & \vdots & \vdots & P_{22}R_{22} \\ 0 & \cdots & 0 & 0 & \end{array}\right] = \left[\begin{array}{cccc|c} 0 & \cdots & 0 & 1 & R'_{12} \\ \hline 0 & \cdots & 0 & 0 & \\ \vdots & \ddots & \vdots & \vdots & R'_{22} \\ 0 & \cdots & 0 & 0 & \end{array}\right] = R'.$$

In particular this implies $R'_{22} = P_{22}R_{22}$ and that

$$P = \left[\begin{array}{c|ccc} 1 & p_{12} & \cdots & p_{1m} \\ \hline 0 & & & \\ \vdots & & P_{22} & \\ 0 & & & \end{array}\right].$$

But then P invertible implies P_{22} invertible. Also R_{22} and R'_{22} are both in row reduced echelon form and $R'_{22} = P_{22}R_{22}$ along with the induction hypothesis implies $R'_{22} = R_{22}$.

But this in turn implies that R and R have the same number r of nonzero rows and that the leading ones in each of these rows is in the same column. Let $R_1, \dots, R_r$ be the nonzero rows of R and $R'_1, \dots, R'_r$ the nonzero rows of R'. Let k_i be the position of the leading one in R_i (which is the same as the position of the leading one in R'_i). Then as every row of R' is a linear combination of the rows of R for each i there are scalars $c_1, \dots, c_r$ so that

$$R'_i = c_1R_1 + c_2R_2 + \cdots + c_rR_r.$$

If we look at the k_i position in R'_i then it is a 1 (as k_i was where the leading 1 in R'_i occurs) and the same is true of R_i. But all of the other vectors R_j have a 0 in this place as R is row reduced. This means that the element in the k_i position of $c_1R_1 + c_2R_2 + \cdots + c_rR_r$ is just c_i and that his must be equal to 1 (the element in the k_i of R_i). For $j \neq i$ we look a similar argument shows that the element in the k_j position of $c_1R_1 + c_2R_2 + \cdots + c_rR_r$ is c_j is c_j and this must be equal to the element in the k_j position of R'_i which is 0. In summary: $c_i = 1$ and $c_j = 0$ for $j \neq i$. Therefore

$$R'_i = c_1R_1 + c_2R_2 + \cdots + c_rR_r = R_i.$$

As this works for any i this shows $R'_i = R_i$ for $i = 1, \dots, r$. Therefore R and R' have all their rows equal and thus are equal. This completes the proof. ❑

2 Summary of Chapter 1

This chapter deals with solving systems of linear equations over a field. A ***field*** is a collection of elements $\mathbf{F}$ together with two binary operations addition and multiplication with satisfy the usual rules of high school algebra in the sense that it is possible to add, subtract, multiple and divide as usual. In the context of linear algebra elements of the field will often be called ***scalars***.

The main subject of this chapter is systems of m linear equations in n unknowns with coefficients from the field $\mathbf{F}$. That is a system of the form

$$\begin{aligned} A_{11}x_n + \cdots + A_{1n}x_n &= y_1 \\ \vdots \qquad\qquad \vdots \;&\;\vdots\; \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_m &= y_m. \end{aligned} \tag{1}$$

The equation is ***homogeneous*** iff all the y_i's vanish. That is the system is of the form

$$\begin{aligned} A_{11}x_n + \cdots + A_{1n}x_n &= 0 \\ \vdots \qquad\qquad \vdots \;&\;\vdots\; \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_m &= 0. \end{aligned}$$

The more general system (1) is the ***inhomogeneous system***. One basic difference between the homogeneous and inhomogeneous system is that the homogeneous system always has at least the solution $x_1 = x_2 = \cdots x_n = 0$ (called the ***trivial solution***) while the inhomogeneous system need not have any solutions in which case it is called ***inconsistent***. A exmaple of an inconsistent system would be $x_1 + x_2 = 1$ and $2x_2 + 2x_2 = 1$.

Given such a system of equations some other equation is a ***linear combination*** of elements in this system if it is formed by multiplying some of the equations in the system by scalars and adding the results together. Two systems of linear equations are ***equivalent*** if and only if every equation one system is a linear combination of equations in the other system and vice versa. What makes this notation of equivalence interesting is

Theorem 2.1 *Equivalent systems of equations have the same set of solutions.*

In doing calculations with systems such as taking linear combinations it soon becomes clear that all the actual work of the calculations is done with the coefficients A_{ij} of the system and that the unknowns x_i are just along for the ride. This has lead to the notation of a ***matrix*** which is just a rectangular array $A = [A_{ij}]$ of elements. To be more precise a $m \times n$ matrix has m rows, n columns and the element A_{ij} is in the i th row and j th column of A. Associated to the inhomogeneous system (1) there are four matrices

$$A := \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \qquad X := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad Y := \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

The system (1) can then be written as $AX = Y$ and the homogeneous system can be written as $AX = 0$. Before going to give the relation between matrices and solving systems of linear equations it we note that matrices have a good deal of algebra associated with them. If particular it is possible to multiply a matrix by a scalar so that cA is the matrix obtained form A by multiplying all of its elements by c. The sum $A + B$ of two matrices of the same size is matrix whose (i, j)th entry is the sum of the corresponding elements of A and B and the product of an $m \times n$ matrix A with an $m \times p$ matrix B is the $m \times p$ matrix AB with elements

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}.$$

This product is associative (*i.e.* $(AB)C = A(BC)$) but not commutative (sometimes $AB \neq BA$). Note that when we rewrite the system (1) as $AX = Y$ that the product between A and X is matrix multiplication. The $n \times n$ ***identity matrix*** is the matrix $I_{n\times n}$ with entries $\delta_{ij} = 1$ if $i = j$ and $= 0$ if $i \neq j$. Or what is the same thing $I_{n\times n}$ is the $n \times n$ matrix with 1's down the main diagonal and all other elements zero. This matrix is special in that for any $m \times n$ matrix A we have

$$I_{m\times m} A = A I_{n\times n} = A.$$

For a square matrix A one to define a matrix B to be the ***inverse*** of A iff $AB = BA = I$ where I is the identity matrix. (While at first it would seem that one might have to define left inverses, right inverse and two sided inverses we have shown that if A has a one sided inverse then this one sided inverse is the unique two sided inverse to A, thus at least for square matrices we do not have to worry about these distinctions). A square matrix with an inverse will be called a ***invertible*** or ***nonsingular*** iff it has in inverse and it is called ***singular*** iff it has no inverse. The inverse of a nonsingular matrix A will be denoted by A^{-1}. Note if $A_1, \ldots, A_k$

are all invertible matrices of the same size then the product $A_1A_2\cdots A_k$ is also invertible and $(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}$.

After this digression on matrix algebra we return to systems of equations. The ***elementary row operations*** on a $m \times n$ matrix A are

1. Multiplying a row of A by a nonzero scalar.
2. Adding a scalar multiple of one row to anther row.
3. Interchanging two rows of A.

In terms of the corresponding system of equations these operations are the same as multiplying one of the equations by a nonzero scalar, adding a scalar multiple of one equation to anther and final interchanging two of the equations. Note that any of these operations can be reversed by anther elementary row operation of the same type. We also note that each elementary row operation e can be accomplished by multiplying A on the left by a $m \times m$ matrix E. That is $a(A) = EA$. These matrices are called ***elementary matrices***. Two matrices A and B are ***row equivalent*** iff there is a finite sequence of row operations that transform A into B. Or what is the same thing if there is a finite number of elementary matrices $E_1, \ldots, E_k$ is that $B = E_k \cdots E_1 A$. This is very closely related to matrix multiplication because of

Theorem 2.2 *Two $m \times n$ matrices A and B are row equivalent iff there is a nonsingular $m \times m$ matrix P so that $B = PA$.*

To study the solutions of the homogeneous system $AX = 0$ note if A and B are row equivalent then the systems $AX = 0$ and $BX = 0$ are equivalent in the sense of Theorem 2.1 and thus have the same solutions. This makes it interesting to try to row reduce the matrix to as simple a form as possible. The first step in this is a matrix is ***row reduced*** iff the first nonzero element in any row is a 1 and any column that contains a leading row has all its other elements zero. A matrix is in ***row reduced echelon form*** iff if is row reduced, all its nonzero rows appear above all its zero rows and the leading one in any row appears to the right of all the leading ones in the rows above it. This turns out to be our simple form under row equivalence:

Theorem 2.3 *Every matrix is row equivalent to a unique row reduced echelon matrix.*

This allows us to define an important invariant of a matrix, its ***rank*** which is the number of nonzero rows in its row reduced echelon form. This invariant is used in understanding when the inhomogeneous system $AX = 0$ has non-trivial solutions. In particular we have

Theorem 2.4 *Let R be the row reduced echelon form of the matrix A. Then the general solution to $AX = 0$ is obtained by letting the variables that do* not *correspond to leading ones in R have arbitrary values and then solving for the variables that correspond to leading ones in the system $RX = 0$. Thus if the rank of A is equal to the number of rows (that is the number of equations) then all variables correspond to leading ones and the system has only the trivial solution $X = 0$. However if the rank is less than the number of rows then there will be non-trivial solutions. In particular this leads to the very important result than any homogeneous system $AX = 0$ where the number of unknowns is larger than the number of equations will have a non-trivial solution.*

This tells close to the full story about the homogeneous system $AX = 0$. To understand the inhomogeneous equation $AX = Y$ form the ***augmented matrix***

$$A' := \begin{bmatrix} A_{11} & \cdots & A_{1n} & y_1 \\ \vdots & \ddots & \vdots & \\ A_{m1} & \cdots & A_{mn} & y_m \end{bmatrix}$$

Obtained from A by adding Y as an extra column. Let r be the rank of A which is the number of nonzero rows in R. We then row reduce A (not the whole matrix A') to row reduced echelon form R. The result is

$$\begin{bmatrix} & R & & Z_1 \\ 0 & \cdots & 0 & z_{r+1} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & z_m \end{bmatrix} \qquad \text{where} \qquad Z_1 := \begin{bmatrix} z_1 \\ \vdots \\ z_r \end{bmatrix}$$

and each of the z_i's is a linear combination of the variables $y_1, \ldots, y_m$. For any given values of $y_1, \ldots, y_m$ if any of the linear expressions $z_{r+1}, \ldots, z_m$ is nonzero, then the system will be inconsistent (these conditions are often called ***compatibility conditions***). If $z_{r+1} = z_{r+2} = \cdots z_m = 0$ then the system is consistent and the general solution is found as in the homogeneous case by letting the variable not corresponding to leading 1's in R have arbitrary values and then solving for the variables that do correspond to leading ones by use of the system $RX = Z_1$. Note if the rank of A is equal to the number of rows, there are no compatibility conditions and we have

Theorem 2.5 *If the rank of A is equal to the number of rows of A (which is the number of equations in the system $AX = Y$) then the inhomogeneous system $AX = Y$ has a solution for any choice of Y. More generally the number of compatibility conditions is difference between the number of rows of A and the rank of A.*

We specialize this to the case of "square" systems. That is where the number of equations is equal to the number of unknowns.

Theorem 2.6 *For a square $n \times n$ matrix the following are equivalent:*

1. *A is invertible.*
2. *The homogeneous system $AX = 0$ has only the trivial solution.*
3. *The inhomogeneous $AX = Y$ has a solution for every choice of Y.*
4. *A is row equivalent to the identity matrix I.*
5. *The rank of A is n.*

Finally there are some computational issues that I have not mentioned in the above. These are things such as finding the inverse of a invertible square matrix A (which is done by augmenting A by the identity matrix to get $[A, I]$ and row reducing this to $[I, B]$. Then $B = A^{-1}$. When appropriate the best way to deal with this type problem is to use some computer package such as Maple or Matlab. ❑