Math 642:550 — Summer 2006
MTTh 6:00–8:30 PM   Hill 705
Prof. Bumby

Supplement 1, Prerequisites

1. Introduction

A motivation for the subject of **Linear Algebra** is to organize the techniques for solving systems of linear equations. Since you should have had a course in which this was done, it suffices to show how some simple examples fit into the theory.

1.1 One equation in three unknowns

What are **all** solutions of

$$x - 2y + 5z = 17? \qquad (Q)$$

The equation can be written as

$$x = 17 + 2y - 5z, \qquad (A)$$

which can be interpreted as: **allow** y and z to **take any values** and use (A) to find a value of x. **Every such assignment** of values to (x, y, z) satisfies (Q), and **every solution** of (Q) can be obtained in this way.

1.2 An extension

Suppose that we learn that equation (Q) was **expected to contain** a variable w. Then the solutions (w, x, y, z) are **still** given by (A), although **we must now allow** y, z, **and** w to take arbitrary values.

Although the method of solution is the same, the answers to these two questions are different. This serves as a reminder that a complete statement of the equation (Q) **must contain** a list of the variables that are allowed to appear in the equation. A solution will give values to **all** of these variables in terms of some **parameters**. The method of solution shows that these parameters can be chosen from the variables **declared** in the problem (even if they are not **seen** in any equation).

In order to have any theoretical results, **the variables must be declared in advance** and **only the declared variables will be allowed to appear in equations**.

1.3 Two equations

Suppose that our original equation (Q) is joined by the equation

$$w + 3y - 2z = -5 \qquad (Q')$$

to form a system of two equations. Solving (Q) for x, as before, and solving (Q') for w, gives expressions for all solutions w, x, y, z in terms of y and z.

This is too easy. We are more likely to meet something like

$$\begin{aligned} w + x + y + 3z &= 12 \\ 7w + 2x + 17y - 4z &= -1 \end{aligned}$$

If we just had the first equation,we would solve it for w in terms of x, y, z. This solution is still valid when we meet the second equation, so we can **substitute** $12 - x - y - 3z$ for w in the second equation to find out what the second equation says about the parameters x, y, z in the solution of the first equation. It turns out to be (Q)! This allows us to use $17 + 2y - 5z$ in place of x in our expression for w to get the same expression for w in terms of the **free variables** y and z that was found by solving (Q'). This system has the same solution as (Q) and (Q').

While the above solution is correct, **it is never used in practice**. While it retains the variables that are valuable in **interpreting the answer**, it requires the use of those variables to identify like terms, and the process of shifting terms from one side of an equation to the other introduces changes of sign. This approach to algebra is unsuitable for humans because it is **tedious** and **difficult to do accurately**, and unsuitable for computers because the terms must be repeatedly examined to find how expressions can be simplified. Fortunately, **there is a better way**.

1.4 Elimination

The substitution step is **completely equivalent** to subtracting 7 times the first equation from the second in our system. Why 7? It leads to an expression in which w does not appear. As long as we do the same operation on both sides of our equations, we get relations between our variables w, x, y, z that are consequences of the relations in the given system. When we are done, we can check that all our results satisfy the original equations. However, in some cases, this is only a partial check of the accuracy of our computation since it only shows that we haven't introduced values that fail to satisfy the given system, but it cannot detect an incorrect exclusion of some solutions.

For any system, this method of **Gaussian elimination** proceeds in two steps. First, each equation is used to express one of the variables in terms of all remaining variables and to remove that variable from the remaining equations. When all equations have been processed, any variables that remain are **free variables** that parameterize the solutions. We then work back through the list to replace references to the other variables by their expressions in terms of the free variables. This is called **back substitution**, but should be done by additional **elimination** steps. (It turns out that it is more efficient for all but very small systems to wait until the free variables have been identified before eliminating other variables from our answers.)

In this method, the variables only serve to arrange their coefficients in columns and the order of columns is fixed throughout the solution. Similarly, the rows are kept in a fixed order that trace their history back to a particular equation in the original list although the other equations are used to modify the information on that row (unless it is **absolutely necessary** to change the order of the rows). The calculation can now be done with a rectangular array (or **matrix**) of coefficients. In hand computation, a new array is written summarizing a number of computations that can be done in parallel (called a **pivot step**). In machine computation, the steps are done **in place**, so it is important the matrix retain the same number if rows and columns throughout the computation.

1.5 Stopping conditions

Each step in Gaussian elimination that replaces a system of equations with one that is **simpler**, in some sense, is like the **induction step** in a proof. However, an induction proof also needs a **basis for the induction**. The current fashion is to include zero among the **natural numbers**, which leads to a search for a **vacuous case** to use for the basis of an induction. The ideal vacuous case for a system of equations is **a system of no equations**. In this case, all variables are free variables. This is what we reach if we are successful in solving each equation for one of its variables.

The other type of basic state has an equation in which all coefficients of the variables are zero. This means that the left side of the equation evaluates to zero for any choice of the variables. There are two possibilities: if the right side is also zero, the equation says that $0 = 0$, which is a tautology, so that equation is redundant, and could be omitted (although it is better to preserve the number of equations in the system by keeping it); if the right side isn't zero, the equation says that, **if the system has any solutions, then** $0 = 1$, which we interpret as saying that there are no solutions.

2. Matrices

Strictly speaking, we have described the **augmented matrix** of the system. The last column represents the right side of the equation, while the remain ones are coefficients of particular variables. This is usually indicated by separating the last column from the others by a dotted line. This dotted line corresponds to the equals sign in the equations: columns to the left of the line are coefficients of the variables; those to the right are the constants on the right sides of the equations. Typically, one is solving a single system of equations, so there is one column to the right of the dotted line. However, there could be **any number** of columns in a given problem. An important special case is the **homogeneous system** in which all right sides are zero. The operations in the solution preserve a zero column, so it is customary to omit such columns. This leads to systems with no columns in the right side of the augmented matrix. Another case that is treated as a different computation is the computation of the **inverse** that will be described below.

The steps in the solution of a system is determined entirely by the left side of its augmented matrix. If you want to solve several systems with the same left side, all right sides could be written as columns to the right of the dotted line in a large augmented matrix. A solution of one of these systems is found by looking in its column on the right side of the final augmented matrix.

Although the computations are done entirely with arrays of numbers, the variables and equations are never far away. Multiplying the entry in a column by the variable belonging to that column in the original system and adding these terms gives an expression whose equality to the value in the last column is asserted by the system being solved. This way of combining elements of matrices is the definition of **matrix multiplication**. Given a matrix M and a column v with the **same number of entries** as M has columns (equivalently, the number of entries in a row of M) Mv is found by, for each i, adding the product of the j^{th} entry of the i^{th} row of M with the j^{th} entry of v to form the i^{th} entry of the product. This allows systems of equations to be abbreviated as a matrix equation $Mx = b$.

This can be extended to give a product of matrices A and B — **in that order** — if the number of **columns** of A is **the same** as the number of **rows** of B with each column of the product being the product of A with the corresponding column of B.

Although the order of terms in a product must remain fixed since the operation is usually **non-commutative**, the **associative law** holds. That is, if either of the products $A(BC)$ or $(AB)C$ make sense, then both do, and the results are equal. One proof is to simply write expressions for the elements of these product an show that they are the same, but other approaches to matrices lead to more conceptual proofs.

An n by n matrix I like

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

with entries that are zero except 1 in every place on the **main diagonal**, has the property that multiplying by I (where defined) gives a product equal to the other factor. Such a matrix is called an **identity matrix**. If X is a matrix with $AX = I$, X is said to be a (right) **inverse** of A. One method to find an inverse (or to show that one doesn't exist) is to write an augmented matrix with A on the left and I on the right and apply Gaussian reduction.

2.1 Row operations

Another role of matrices is to express the steps of Gaussian elimination as row operations that can be expressed by multiplying the augmented matrix of the system **on the left** by a suitable matrix. In all cases, this matrix is obtained by applying the intended operation to an identity matrix.

The elementary row operations of Gaussian reduction (including the rarely used row interchange) are represented by **square** matrices (same number of rows and columns) that have unique inverses giving row operations that **undo** that operation. Insisting that the number of rows and columns remain fixed throughout the computation allows the work to be described in terms of matrix multiplication..

In the first part of Gaussian elimination, M is reduced to a triangular (or **echelon**) form. The role of b is to accept these row operations so that each step in the process gives the augmented matrix of a system with the same solutions as $Mx = b$.

The left multiplications expressing these steps involve matrices in which nonzero elements can only appear on or below the main diagonal. The product of such matrices also has the same property.

2.2 The LU factorization

The reduction of an m by n matrix M to echelon form, when it is possible, can then be summarized by a factorization $M = LU$ where L is a m by m **lower** triangular matrix whose inverse gives the product of the operations and U is the m by n **upper** triangular matrix that is the result of the operations. The matrix L can also be found in the record of the computation (this is described in detail in the textbook — see index entries for "pivot", "LU factorization", and "LDU factorization"). If the original matrix is augmented by a column b, the echelon form U will be augmented by a column c with $Lc = b$. Note that we have required L to be a square matrix with 1 on its main diagonal. The matrix U has the same shape as M. These are the conventions of the text — other conventions may be used elsewhere.

It is easy to see from the echelon form whether the system has solutions, and to determine **qualitative** information about the nature of the set of all solutions.

If there are rows of zeros at the bottom of U, then a test for $Mx = b$ to have a solution, is that c also have zero in these positions, since such a row represents an equation whose left side is zero for all values of the variables, and $Mx = b$ is **equivalent** to $Ux = c$.

2.3 Uniqueness

If no row of U consists entirely of zero, then this is the only LU factorization of M. We prove this by induction. The first **column** of an **upper** triangular matrix can only have a nonzero entry in its first position. Similarly, the first **row** of a **lower**triangular matrix can only have a nonzero entry in its first position. The first row of a product is found by multiplying the first row of the first factor with the whole second factor. Thus, the first row of U is a scalar multiple of the the first row of M. Furthermore, the LU factorization has been normalized so that the entry in the (1, 1) position of L is 1, so the first rows of U and M are identical. Likewise, the first column of a product is the product of the whole first factor with the first column of the second factor. Here, the first column of U has zero in all places but the first, so the first column of M is a multiple of the first column of L.

After this pivot step, both M and L have been reduced to matrices in which all rows except the first have 0 in the first position. If we now remove the first row and column from L, U, and M, we have an LU factorization of the replacement for M. If we can continue until the rows of U are exhausted, every element of the factorization will have been characterized by the calculation that we used.

On the other hand, if the last row of U consists entirely of zeros, the elements in the last column of L are multiplied by zero when L and U are multiplied to give M. Since we normalized L to have 1 on its diagonal, the result is still unique. However, if **two or more** rows of U are zero, there will be entries of L that do not contribute to the product LU. The factorization will not be unique in this case.

3. Row interchanges

Gaussian elimination doesn't always work. Induction shows that the case in which the first variable doesn't appear in the first equation is typical. To get around this, one selects a different equation to solve for the first variable. The simplest way to do this is to exchange two equations in our list. At the matrix level, this means interchanging two rows in the matrix.

This difficulty only arises rarely, so its role in solving systems can be treated in an *ad hoc* manner. In numerical work, there are often reasons to perform row interchanges to assure that the calculation can be performed accurately. This will be treated later in the course.

For now, we assume that we can find an LU factorization.

4. Vector spaces

The collection of all columns of n real numbers is denoted $\mathbb{R}^n$. It has an operation of **addition** defined by putting the sum of corresponding entries in that place of the sum. It is also possible to multiply by a number (called a **scalar** in this context) by multiplying each entry by that scalar. These operations have the expected algebraic properties.

Any collection of object with operations of addition and multiplication by scalars satisfying these properties is called a **vector space**. An introductory course on Linear Algebra explores the consequences of this definition in order to apply it to a large number of examples.

We have seen that $v \in \mathbb{R}^n$ can be multiplied by an m by n (the first number always counts the rows and the second counts the columns) matrix M to give $Mv \in \mathbb{R}^m$. Another interpretation of this product is that each column of M is multiplied by a scalar taken from the corresponding position in v and the results are added together. This gives a **linear combination** of the columns of M. The collection of all such linear combinations is a vector space called the **column space** of M.

In this language, $Mx = b$ has a solution if and only if b lies in column space of M.

Every **subspace** of $\mathbb{R}^m$, i.e., a sub**set** of $\mathbb{R}^m$ that is a vector space with the operations of addition and multiplication by scalars inherited from $\mathbb{R}^m$, can be realized as the column space of some matrix with m rows.

4.1 Intersection of subspaces

Having a record of the reduction of M to the echelon form U, one can solve a system $Mx = b$ by augmenting the matrices appearing in the reduction with those determined by b. The process is the same even if the record contains matrices that were previously augmented since only the part to the left of the dotted line is used to find the TU factorization. Having provided the details of the computation for two columns b_0 and b_1, to obtain U augmented by columns c_0 and c_1, there is a shortcut to the c corresponding to a linear combination $b = \alpha_0 b_0 + \alpha_1 b_1$. At each stage, one would get the same linear combination of columns, so it is not necessary to write them all — one can skip to the last step and write $c = \alpha_0 c_0 + \alpha_1 c_1$. This is useful if neither $Mx = b_0$ nor $Mx = b_1$ has a solution because c_0 and c_1 have nonzero entries where U has a zero row, since we can immediately describe those c which have zero in this position and use the coefficient α to identify the corresponding b. This is much simpler than multiplying by L.

This leads to a method for finding the intersection of subspaces V and W. Begin by using a **basis** for V as the columns in the left side of an augmented matrix M and a **basis** for W as the columns in the right side. Then begin Gaussian elimination to try to express each vector in the given basis of W in terms of the given basis of V. This leads to $M = LU$. Here L is a square matrix and U has the same shape as M and may be considered as an augmented matrix in the same way. If the left side of U has a zero row, any column c in the right side not having zero in this position corresponds to a vector in the basis for W that does **not** belong to V. Typically, none of the vectors in the basis will belong to V, but there may be other **combinations** of these vectors that give elements of the intersection. Specifically, a combination belongs to the intersection if it has zero in all positions corresponding to the zero rows on the left side. The coefficients in this combination can be found by solving the homogeneous system whose left side consists of the right sides of all rows of M whose left sides are zero. This is equivalent to continuing the LU factorization of M using columns from the right side of M.

However, the LU factorization is only **halfway** to the solution of a system. One must also do the **back substitution** steps. However, when the back substitution has been for the rows of U with zero left side, the expressions for vectors of the intersection in terms of the basis of W has been revealed. This allows a basis of the intersection to be written. Further row operations will give expressions in terms of the basis of V, but this only serves as a check of the answer.

The above description may not be clear, so here is a simple example.

4.2 Example

Let

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 9 & -1 & 11 \\ 6 & -4 & 2 \\ -3 & 5 & 20 \end{bmatrix} \qquad B = \begin{bmatrix} 4 & -3 \\ 19 & -10 \\ 0 & -3 \\ 8 & 15 \end{bmatrix}$$

Find the intersection of the space V spanned by the columns of A and the space W spanned by the columns of B.

Solution. Begin by forming the augmented matrix AB, then reduce A to echelon form (pivots will be shown in bold type).

$$\left[\begin{array}{ccc|cc} \mathbf{3} & -1 & 2 & 4 & -3 \\ 9 & -1 & 11 & 19 & -10 \\ 6 & -4 & 2 & 0 & -3 \\ -3 & 5 & 20 & 8 & 15 \end{array}\right]$$

$$\left[\begin{array}{ccc|cc} 3 & -1 & 2 & 4 & -3 \\ 0 & \mathbf{2} & 5 & 7 & -1 \\ 0 & -2 & -2 & -8 & 3 \\ 0 & 4 & 22 & 12 & 12 \end{array}\right]$$

$$\left[\begin{array}{ccc|cc} 3 & -1 & 2 & 4 & -3 \\ 0 & 2 & 5 & 7 & -1 \\ 0 & 0 & \mathbf{3} & -1 & 2 \\ 0 & 0 & 12 & -2 & 14 \end{array}\right]$$

$$\left[\begin{array}{ccc|cc} 3 & -1 & 2 & 4 & -3 \\ 0 & 2 & 5 & 7 & -1 \\ 0 & 0 & 3 & -1 & 2 \\ 0 & 0 & 0 & 2 & 6 \end{array}\right]$$

The elements of right side of the bottom row are not zero. This means that no column of B belongs to V. However, if 3 times the first column is subtracted from the second column, the result has has a zero in the last position. Performing this operation on the columns of B gives

$$\begin{bmatrix} -15 \\ -67 \\ -3 \\ -9 \end{bmatrix}.$$

This is a basis for the intersection, i.e., the multiples of this vector give all elements of $V \cap W$.

Since the vector is a linear combination of columns of B, it belongs to W. Properties of the echelon form show that it also belongs to V.

When you have a basis for $V \cap W$, finding the descriptions of those vectors in terms of **both** the given basis of V and the given basis of W is equivalent to finding the nullspace of a matrix whose set of columns is the union of those bases (this observation was given in a previous edition of the textbook, but is one of the few things removed in the transition to the current edition). The method of these notes, while based on row operations on the same matrix, involves less computation since it only finds expressions with respect to the basis of one of the subspaces after formulating the property that a combination of the vectors in a basis for W belongs to V.

A variation on the method for finding both representations is to do the back substitution only with columns on the right side that come from the intersection. These are recognized in the echelon form by having zero in the positions where there is a row of zeroes on the left side of the augmented matrix. Thus, if you continue the solution with U augmented by

$$\begin{bmatrix} -3 \\ -1 \\ 2 \\ 6 \end{bmatrix} - 3 \begin{bmatrix} 4 \\ 7 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -15 \\ -22 \\ 5 \\ 0 \end{bmatrix}$$

you will get the solution $\left(-67/6, -91/6, 5/3\right)$ giving the coefficients in the expression of this vector in terms of the columns of A.

In general, the intersection can have any dimension, so many combination of columns may need to be found in order to arrive at a basis.

5. Exercise A

Let

$$A = \begin{bmatrix} 1 & 3 & -2 \\ -2 & -5 & 3 \\ 3 & 9 & -5 \\ -2 & -4 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & -3 \\ -12 & 7 \\ 15 & -6 \\ -11 & 13 \end{bmatrix}$$

(a) Find the intersection of the space V spanned by the columns of A and the space W spanned by the columns of B by forming an augmented matrix with A on the left and B on the right.

(b) Repeat with a matrix that has B on the left and A on the right. (This has **not** been constructed to avoid fractions in the computation. You may use of a calculator or computer giving decimal values for the matrix entries.)

(c) Interpret these results. Although the details are different, both claim to find the intersection of V and W. You should see that the spaces are the same.

Note. In simple cases, you can **see** that the spaces are the same. However, if the dimension of the intersection is greater than 1, different computations may give **very** different bases for the space, so another test must be used to recognize that the same space has been found.

End of Supplement