Math 407 Section A

MATH 308 REVIEW

In this course the notion of **linearity** plays a central role. All of the theoretical aspects of this course are based on properties of systems of linear equations and inequalities in R^n. For this reason the course prerequisite MATH 308 should be taken very seriously. Indeed, the first quiz in this course is devoted to this material. This guide has been prepared to facilitate your review of the relevant material on linear systems of equations. Reviewing this material will prepare you for the quiz.

Basic Review:

You should be able to answer the following questions:

Answer the following 6 questions for each of the four cases $n = 1$, $n = 2$, $n = 3$, and $n \in \{1, 2, 3, 4, 5, 6, \ldots\}$ beginning with the case $n = 1$.

(1) Why is R^n called a vector space?
(2) What is the *dot product* (or inner product) on R^n?
(3) What is the angle between any two vectors in R^n? (For this one simply applies the definition $a^T b = \|a\| \, \|b\| \cos\theta$.)
(4) When are two vectors in R^n said to be *orthogonal*?
(5) Describe all linear mappings from R^n to R.
(6) What is the data structure that one usually associates with a linear transformation from R^n to R^m?

You should also be able to answer the following more specific questions:

(7) Given any two points in R^2 determine (a) an equation for the line passing through these points, and (b) a vector normal to the line determined by these two points.
(8) Given any three non–collinear points in R^3 determine (a) an equation for the plane passing through these three points, and (b) a vector normal to this plane.
(9) Let $a = (a_1, a_2, \ldots, a_n)^T \in \mathrm{R}^n$ and $\alpha \in \mathrm{R}$. For $n = 1, 2, 3$ and $n > 3$ describe the set

$$\{x \in \mathrm{R}^n \ : \ a^T x = \alpha\}$$

when (a) $\alpha = 0$, and (b) $\alpha \neq 0$.

(10) Let $a_{1\cdot} = (a_{11}, a_{12}, \ldots, a_{1n})^T \in \mathrm{R}^n$, $a_{2\cdot} = (a_{21}, a_{22}, \ldots, a_{2n})^T \in \mathrm{R}^n$, and $\alpha_1, \alpha_2 \in \mathrm{R}$. For $n = 1, 2, 3$ and $n > 3$ describe the set

$$\{x \in \mathrm{R}^n \ : \ a_{1\cdot}{}^T x = \alpha_1, \ a_{2\cdot}{}^T x = \alpha_2\}.$$

(11) For $i = 1, 2, 3, \ldots, m$, let $a_{i\cdot} = (a_{i1}, a_{i2}, \ldots, a_{in})^T \in \mathrm{R}^n$ and $\alpha_i \in \mathrm{R}$.
   (a) For $n = 1, 2, 3$ and $n > 3$ with $m = n$ describe the set

$$S = \{x \in \mathrm{R}^n \ : \ a_{i\cdot}{}^T x = \alpha_i, \ i = 1, 2, \ldots, m\}.$$

   (b) How does this description change when (a) $m < n$, (b) $m > n$?
   (c) Can you express the set S in matrix notation?

(d) What does it mean to say that the underlying matrix has (a) full column rank, (b) full row rank, (c) full rank?

(12) What are the three elementary row operations, i.e. the three operations associated with row reduction?

(13) Given an $n \times n$ real matrix A, what does it mean to say that A is invertible (or, equivalently, non–singular)?

(14) Let A be an $m \times n$ real matrix, B be a $k \times n$ real matrix, and let $a \in \mathrm{R}^m$ and $b \in \mathrm{R}^k$. We say that the two systems of equations

$$Ax = a \qquad \text{and} \qquad Bx = b$$

are equivalent if they have identical solution sets.

(a) Show that if any one of the three elementary row operations is applied to the system $Ax = a$, then one obtains an equivalent linear system.

(b) If C is a non–singular $m \times m$ real matrix, show that the system $Ax = a$ is equivalent to the system $CAx = Ca$.

(c) Is it possible for the systems $Ax = a$ and $Bx = b$ to be equivalent when $n \neq m$? If it is possible, then provide an example to illustrate this possibility.

Advanced Review:

In these notes we discuss some of the topics studied in Math 308 in greater detail. One of the intents of these notes is to discuss the *ideas* in the 308 without getting bogged down in the notational details. Hopefully, this will help you see the big picture a little better. On the other-hand, these notes are difficult mathematical reading. Reading these notes is an exercise in itself. You will be called upon to remember many of the terms and ideas defined in Math 308. These notes will exercise your knowledge of these terms and ideas and are intended to re–establish and solidify this knowledge.

Subspaces:

Recall that a subset W of R^n is a subspace if and only if it satisfies the following three conditions:

(a) The origin is an element of W.

(b) W is closed with respect to addition, i.e. $u, v \in W$ implies $u + v \in W$.

(c) W is closed with respect to scalar multiplication, i.e. $\alpha \in \mathrm{R}$ and $u \in W$ implies $\alpha u \in W$.

For example, by using these properties it is easily shown that for any set S in R^n the set

$$S^\perp = \{v \,:\, w^T v = 0 \text{ for all } w \in S\}$$

is a subspace. If S is itself a subspace, then $S^\perp$ is called the subspace orthogonal (or perpendicular) to the subspace S. Moreover, in this case we have $S = (S^\perp)^\perp$ (more generally, $(S^\perp)^\perp = \mathrm{Span}\,(S)$). If S is a subspace, it can be shown that

$$n = \dim(S) + \dim(S^\perp)\ .$$

Every subspace has what we will call *internal* and *external* representations. An *internal* representation is any representation of the subspace as a linear span of a finite set of vectors. The representation is said to be internal since the spanning vectors lie within the subspace. If the set of spanning vectors happens to be linearly independent, then it is called a *basis* of the subspace. It is known that every basis of a subspace has the same number of vectors in it. This number is called the *dimension* of the subspace.

Internal representations of a subspace can be interpreted with the aid of our notion for matrix vector multiplication. Recall that matrix vector multiplication can be viewed as taking a linear combination of the columns of the matrix. Thus, if a subspace is known to be the linear span of a finite collection of vectors, then this subspace is the same as the range of the matrix formed by taking the columns of the matrix to be the vectors that span the subspace.

An *external* representation of a subspace is any representation of the subspace as the intersection of a finite number of subspaces of the form

$$\{x \in \mathrm{R}^n \,:\, v_i^T x = 0\} \tag{1}$$

for some nonzero $v_i \neq 0$ in R^n for $i = 1, 2, \dots, k$. The representation is said to be *external* since the vectors v_i clearly cannot belong to the subspace. Another way to view an external representation of the subspace is that the vectors $\{v_1, v_2, \dots, v_k\}$ form a spanning set for the subspace *orthogonal* (or perpendicular) to the subspace we are interested in. If the dimension of the subspace is p, then the dimension of the orthogonal subspace is $n - p$. Thus, in particular, $k \geq n - p$.

External representations of a subspace can also be interpreted with the aid of our notion for matrix multiplication. Recall that matrix multiplication can also be interpreted as taking the dot product with each row of the matrix. Thus, a vector is in the null space of a matrix if it is orthogonal to every row of that matrix. Consequently, a subspace externally represented by the vectors $\{v_1, v_2, \dots, v_k\}$ is the same as the null space of the matrix formed by letting its rows be the vectors v_i in R^n for $i = 1, 2, \dots, k$.

It is important to remember that every subspace has both internal and external representations (indeed, infinitely many of them). Equivalently, every subspace can be represented either as the range of some matrix or as the null space of some matrix (indeed, there are infinitely many such matrices). Some of the computational techniques learned in Math 308 deal with passing between such representations and obtaining minimal representations, that is internal and external representations having the fewest number of elements.

Subspaces Associated with Matrices:

The discussion above tells us that matrices and subspaces are intimately connected to each other. In this regard, every matrix has associated with it four fundamental subspaces; its range and null space and the range and null space of its transpose. The relationship between these subspaces is easily understood by recalling the two different ways to think about matrix vector multiplication:

(i) linear combinations of the columns, and
(ii) dot products with the rows.

Given $A \in \mathrm{R}^{m\times n}$ recall that the null space of A is given by

$$\mathrm{Null}\,(A) = \{x \mid Ax = 0\}\ ,$$

that is, $x \in \mathrm{Null}\,(A)$ if and only if x is orthogonal to every row of A since Ax is just the dot product of x with every row of A. Therefore, $x \in \mathrm{Null}\,(A)$ if and only if x is perpendicular to the linear span of the rows of A. But the linear span of the rows of A is precisely the range of A^T! Therefore, $\mathrm{Null}\,(A) = \mathrm{Ran}\,(A^T)^\perp$. By replacing A by A^T in this expressiong, we also have $\mathrm{Null}\,(A^T) = \mathrm{Ran}\,(A)^\perp$. Putting this all together we obtain the following relationsships:

$$\begin{aligned}
\mathrm{Ran}\,(A) &= \mathrm{Null}\,(A^T)^\perp \\
\mathrm{Null}\,(A) &= \mathrm{Ran}\,(A^T)^\perp \\
m &= \dim(\mathrm{Ran}\,(A^T)) + \dim(\mathrm{Null}\,(A^T)) \\
n &= \dim(\mathrm{Ran}\,(A)) + \dim(\mathrm{Null}\,(A)).
\end{aligned}$$

The dimension of the range of a matrix is called the *rank* of the matrix and the dimension of the null space of a matrix is called the *nullity* of the matrix. An important fact in this regard is that the rank of a matrix equals the rank of its transpose. However, if the matrix is not square, then the nullity of a matrix may differ from the nullity of its transpose.

Echelon Form:

The key computational tool of Math 308 and Math 407 is Gaussian elimination. The purpose of Gaussian elimination is to put a matrix into *echelon form.* Let us review this process in the light of our knowledge of subspaces.

Recall that in Gaussian elimination one employs the three elementary row operations to put a matrix into upper triangular form (or echelon form). These operations can be viewed as operations on the vectors that form the rows of the matrix. That is, they are operations in the row space of the matrix. In Math 308 it was shown that the three elementary row operations do not change the row space of the matrix. Let us review this fact by considering the row operations one at a time.

Row Interchange: In this operation we simply change the order in which we write the rows. Clearly this does not change the row space of the matrix.

Multiply a Row by a Nonzero Scalar: Again the row space remains unchanged since the row space contains all linear combinations of the rows and scalar multiplication is simply a special case of this.

Replace a Row by the Sum of Itself and a Multiple of Another Row: This just represents a linear combination of two rows. Thus the new row remains in the row space. Furthermore, *the new collection of rows have the same span as the rows you started with.* In order to see this we just need to show that the only row that was changed can be obtained as some linear combination of the new rows. This is also obvious. The old row is simply the new row minus what was added to it, namely a scalar multiple of one of the unchanged rows.

Thus, the elementary row operations do not change the row space of the matrix. In particular, we have that two matrices are *row equivalent* if and only if they have the same row space.

To say that we have tranformed a matrix into echelon form implies that we have made as many rows of the matrix zero as we possibly can. Thus, none of the remaining nonzero rows can be represented as a linear combination of the other remaining nonzero rows. That is, the remaining nonzero rows are linearly independent and have the same span as the row space of the original matrix, or equivalently, the nonzero rows form a basis for the row space of the original matrix. This is a remarkable fact! Our primary tool for solving equations is also an efficient way to obtain a basis for a subspace from a spanning set for that subspace!

Echelon form is a very powerful tool. It can be used to solve a wide variety of problems. We take a moment to review a few of the problems it is used to solve in Math 308.

<u>Equation Solving:</u> Given a system of linear equations, $Ax = b$, where $A \in \mathrm{R}^{m\times n}$ and $b \in \mathrm{R}^m$, we can characterize the solution set to this system by reducing the associated augmented matrix $[A|b]$ to echelon form.

<u>Passing between internal and external representations of a subspace:</u>

<u>Internal to external:</u> This is the same as expressing the range of one matrix as the null space of another. Let A be the matrix in question and assume y is in the range of A. Since $y \in \mathrm{Ran}\,(A)$, the system $Ax = y$ is consistent. We then formally solve this system by reducing the augmented matrix $[A|y]$ to echelon form. Once in echelon form some of the rows will be all zero except for the right–hand side which will be a algebraic expression in the components of y. The consistency of the system implies that each of these expressions must be zero. By writing these expressions in matrix form and setting them equal to zero we get a matrix equation of the form $By = 0$ for some matrix B. That is, we have expressed the range of A as the null space of B.

<u>External to internal:</u> This is the same as expressing the null space of one matrix as the range of another. Again, letting A be the matrix in question, we reduce A to *reduced echelon form*. In general, this will give us a row equivalent matrix with block structure

$$C = \begin{bmatrix} I & B \\ 0 & 0 \end{bmatrix} .$$

The the columns of the matrix $D = \begin{bmatrix} -B \\ I \end{bmatrix}$ form a basis for the null space of A. Indeed, the null space of A is the subspace orthogonal to the row space of A. Since C is row equivalent to A, the null space of A is the subspace orthogonal to the row space of C. Next observe that $CD = 0$, that is, the columns of D are orthogonal to the rows of C. Since the dimension counts work out, the columns of D must be a basis for the null space of C.

Generating a basis for the span of a finite collection of vectors: This has already been discussed above as one of the primary consequences of the echelon form. Just write the vectors as the row vectors of a matrix and then reduce this matrix to echelon form. The rows of the reduced matrix will be the desired basis.

Generating a basis for the null space of a matrix: This also has been discussed above in another guise. It is really the same as expressing the null space of the matrix as the range of another matrix where the other matrix must have the fewest number of columns possible since then the columns of this other matrix will form a basis for the null space. This was addresses above in our discussion of how to go from an external representation of a subspace to an internal one.

Generating a basis for the range of a matrix: Since the range of a matrix is just the linear span of its columns (or the column space), this is the same as generating a basis for the column space. This also was discussed above in our description of how to obtain a basis for the linear span of a finite collection of vectors. In short, reduce the transpose of the matrix to echelon form. The nonzero rows of the reduced matrix will be a basis for the range.

We give a concrete illustrations of how to use echelon, and reduced echelon form to compute bases for the 4 elementary subspaces associated with a matrix.

Computing Bases for the Four Fundamental Subspaces of a Matrix

Let $A \in \mathrm{R}^{m\times n}$. An efficient procedure for computing bases for the four fundamental subspaces associated with A begins by reducing the augmented system

$$[A \,|\, I_{m\times m}] \in \mathrm{R}^{m\times(n+m)} \tag{2}$$

to echelon form. Recall that this is equivalent to multiplying the augmented system on the left by some nonsingular $m\times m$ matrix M, yielding

$$[MA \,|\, M] \in \mathrm{R}^{m\times(n+m)} \ . \tag{3}$$

Decomposing this augmented matrix conformally with respect to the nonzero and zero rows of MA yields the block matrix

$$\left[\begin{array}{c|c} T_1 & T_2 \\ 0 & T_3 \end{array}\right] \ , \tag{4}$$

where, for some k, the matrix $T_1 \in \mathrm{R}^{k\times n}$ is in echelon form and has no zero rows and $M = \begin{bmatrix} T_2 \\ T_3 \end{bmatrix}$ is block conformal with $\begin{bmatrix} T_1 \\ 0 \end{bmatrix}$. From Math 308 we know that

- $T_1 \in \mathrm{R}^{k\times n}$ where $k := \operatorname{rank} A \le \min\{m, n\}$ is the rank of A,
- the rows of the matrix T_1 form a basis for $\operatorname{Ran}(A^T)$, and
- the rows of the matrix $T_3 \in \mathrm{R}^{(m-k)\times m}$ form a basis for $\operatorname{Null}(A^T)$.

If no row interchanges were required for the reduction to echelon form (i.e only k Gaussian elimination steps), then T_2 and T_3 have the form

$$T_2 = [T_{21}\ 0_{k\times(m-k)}] \quad \text{and} \quad T_3 = \left[T_{31}\ I_{(m-k)\times(m-k)}\right] \ , \tag{5}$$

where $T_{21} \in \mathrm{R}^{k\times k}$ is invertible and $T_{31} \in \mathrm{R}^{(m-k)\times k}$. In this case, the columns of the matrix

$$\begin{bmatrix} I \\ -T_{31} \end{bmatrix} \tag{6}$$

form a basis for Ran (A). Finally, to get a basis for Null (A), we again use Gaussian elimination to transform the matrix T_1 to reduced echelon form, yielding a matrix which typically has block structure

$$[I_{k\times k}\ T_{12}] \in \mathrm{R}^{m\times k}\ . \tag{7}$$

In this case, the columns of the matrix

$$\begin{bmatrix} -T_{12} \\ I \end{bmatrix} \tag{8}$$

form a basis for Null (A).

We now illustrate this process on the matrix

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 0 \\ -1 & 1 & -4 \\ 3 & 5 & 4 \end{bmatrix}.$$

In this case, the augmented system (2) has the form

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

After reduction to echelon form, we obtain the matrix

$$\left[\begin{array}{ccc|cccc} 1 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 4 & -3 & 1 & 0 \\ 0 & 0 & 0 & -4 & 1 & 0 & 1 \end{array}\right].$$

In this case, the matrices T_1, T_2, T_3 and T_{31} appearing in (4) and (5) are

$$T_1 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix}, \quad T_3 = \begin{bmatrix} 4 & -3 & 1 & 0 \\ -4 & 1 & 0 & 1 \end{bmatrix},$$

and $T_{13} = \begin{bmatrix} 4 & -3 \\ -4 & 1 \end{bmatrix}$. Therefore, as indicated above,

$$\mathrm{Ran}\,(A^T) = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}, \quad \mathrm{Null}\,(A^T) = \mathrm{Span}\left\{ \begin{pmatrix} 4 \\ -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right\},$$

and

$$\mathrm{Ran}\,(A) = \mathrm{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -4 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \\ -1 \end{pmatrix}\right\}.$$

Finally, transforming T_1 to reduced echelon form yields the matrix

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}.$$

Therefore, the matrix T_{12} appearing in (7) is $T_{12} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$, and so

$$\mathrm{Null}\,(A) = \mathrm{Span}\left\{\begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix}\right\}.$$

Recapping, we see that beginning with the augmented matrix (2), one can reduce to echelon form to obtain the matrix (4). At this point one can immediately read off bases for both $\mathrm{Ran}\,(A^T)$ and $\mathrm{Null}\,(A^T)$. This leads to an alternative method for computing bases for the subspaces $\mathrm{Ran}\,(A)$ and $\mathrm{Null}\,(A)$. That is to simply repeat the process described above with A replaced by A^T. In this approach one begins with the augmented matrix

$$[A^T \mid I]$$

and reduces to the echelon form

$$\left[\begin{array}{c|c} S_1 & S_2 \\ 0 & S_3 \end{array}\right], \tag{9}$$

where S_1 is in echelon form and has no zero rows. Then the rows of S_1 form a basis for $\mathrm{Ran}\,(A)$, while the rows of S_3 form a basis for $\mathrm{Null}\,(A)$. Applying this approach to the example given above, we begin by reducing the augmented matrix

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

to the echelon form

$$\left[\begin{array}{cccc|ccc} 1 & 1 & -1 & 3 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & -3 & 1 & 1 \end{array}\right].$$

Consequently,

$$S_1 = \begin{bmatrix} 1 & 1 & -1 & 3 \\ 0 & 1 & 3 & -1 \end{bmatrix} \quad \text{and} \quad S_3 = \begin{bmatrix} -3 & 1 & 1 \end{bmatrix}.$$

Therefore,

$$\mathrm{Ran}\,(A) = \mathrm{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \\ -1 \end{pmatrix}\right\} \quad \text{and} \quad \mathrm{Null}\,(A) = \mathrm{Span}\left\{\begin{pmatrix} -3 \\ 1 \\ 1 \end{pmatrix}\right\}.$$

Working with Block Structured Matrices

Numerical linear algebra lies at the heart of modern scientific computing and computational science. Today it is not uncommon to perform numerical computations with matrices having millions of components. The key to understanding how to implement such algorithms is to exploit underlying structure within the matrices. In these notes we touch on a few ideas and tools for dissecting matrix structure. Specifically we are concerned with block matrix structures.

In the 308 Review above, we have already made heavy use of block structures. Indeed, the augmented matrix associated with the linear system $Ax = b$ given by $[A|b]$ is the first example of a block structured matrix. But we also made use of block structures associated with echelon and reduced echelon form to compute bases associated with the four fundamental subspaces of a matrix. In math 407, we will make even greater use of such structures so it is very important to understand them well.

To illustrate the general idea of block structures consider the following matrix.

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

Visual inspection tells us that this matrix has structure. But what is it, and how can it be represented? We re-write the the matrix given above *blocking* out some key structures:

$$A = \left[\begin{array}{ccc|ccc} 3 & -4 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 & 3 \end{array}\right] = \left[\begin{array}{c|c} B & I_{3\times 3} \\ \hline 0_{2\times 3} & C \end{array}\right],$$

where

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

$I_{3\times 3}$ is the 3×3 identity matrix, and $0_{2\times 3}$ is the 2×3 zero matrix. Having established this structure for the matrix A, it can now be exploited in various ways. As a simple example, we consider how it can be used in matrix multiplication.

Consider the matrix

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

The matrix product AM is well defined since A is 5×6 and M is 6×2. We show how to compute this matrix product using the structure of A. To do this we must first *block* *decompose M conformally with the block decomposition of A.* Another way to say this is that we must give M a block structure that allows us to do block matrix multiplication with the blocks of A:

$$M = \begin{bmatrix} X \\ Y \end{bmatrix},$$

where

$$X = \begin{bmatrix} 1 & 2 \\ 0 & 4 \\ -1 & -1 \end{bmatrix}, \quad \text{and} \quad Y = \begin{bmatrix} 2 & -1 \\ 4 & 3 \\ -2 & 0 \end{bmatrix}.$$

Then

$$\begin{aligned} AM &= \begin{bmatrix} B & I_{3\times 3} \\ 0_{2\times 3} & C \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} BX + Y \\ CY \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 2 & -9 \\ 2 & 12 \\ -1 & -2 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 4 & 3 \\ -2 & 0 \end{bmatrix} \\ \\ \begin{bmatrix} 4 & 1 \\ -4 & -1 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 4 & -10 \\ 6 & 15 \\ 1 & -2 \\ 4 & 1 \\ -4 & -1 \end{bmatrix}. \end{aligned}$$

The example given above is a numerical example of how block matrix multiplication works and why it might be useful. One of the most powerful uses of block structures is in understanding and implementing standard matrix factorizations. We consider one such factorization: the LU Factorization.

The LU Factorization

Recall from linear algebra that Gaussian elimination is a method for solving linear systems of the form

$$Ax = b,$$

where $A \in \mathrm{R}^{m\times n}$ and $b \in \mathrm{Ran}(A)$. In this method, the augmented system

$$[A\,|b]$$

is formed and then the three elementary row operations are used to put this system into row echelon form (or upper triangular form). A solution x is obtained by back substitution, or back solving, starting with the component x_n. We show how the process of bringing a matrix to upper triangular form can be performed by left matrix multiplication.

The key step in Gaussian elimination is to transform a vector of the form

$$\begin{bmatrix} a \\ \alpha \\ b \end{bmatrix},$$

where $a \in \mathrm{R}^k$, $0 \neq \alpha \in \mathrm{R}$, and $b \in \mathrm{R}^{n-k-1}$ (here the possible values for k are $k = 0, 1, \ldots, n-1$), into one of the form

$$\begin{bmatrix} a \\ \alpha \\ 0 \end{bmatrix}.$$

This can be accomplished by left matrix multiplication as follows:

$$\begin{bmatrix} I_{k\times k} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\alpha^{-1}b & I_{(n-k-1)\times(n-k-1)} \end{bmatrix} \begin{bmatrix} a \\ \alpha \\ b \end{bmatrix} = \begin{bmatrix} a \\ \alpha \\ 0 \end{bmatrix}.$$

The matrix

$$\begin{bmatrix} I_{k\times k} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\alpha^{-1}b & I_{(n-k-1)\times(n-k-1)} \end{bmatrix}$$

is called a Gaussian elimination matrix. This matrix is invertible with inverse

$$\begin{bmatrix} I_{k\times k} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha^{-1}b & I_{(n-k-1)\times(n-k-1)} \end{bmatrix}.$$

We now use this basic idea to show how a matrix can be put into upper triangular form.

Suppose

$$A = \begin{bmatrix} \alpha_1 & v_1^T \\ u_1 & \widetilde{A}_1 \end{bmatrix} \in \mathbb{C}^{m\times n},$$

with $0 \neq \alpha_1 \in \mathbb{C}$, $u_1 \in \mathbb{C}^{m-1}$, $v_1 \in \mathbb{C}^{n-1}$, and $\widetilde{A}_1 \in \mathbb{C}^{(m-1)\times(n-1)}$. Then using the first row to zero out u_1 amounts to left multiplication of the matrix A by the matrix

$$\begin{bmatrix} 1 & 0 \\ -\alpha_1^{-1}u_1 & I \end{bmatrix}$$

to get

$$(*) \qquad \begin{bmatrix} 1 & 0 \\ -\alpha_1^{-1}u_1 & I \end{bmatrix} \begin{bmatrix} \alpha_1 & v_1^T \\ u_1 & \widetilde{A}_1 \end{bmatrix} \in \mathbb{C}^{m\times n} = \begin{bmatrix} \alpha_1 & v_1^T \\ 0 & A_1 \end{bmatrix},$$

where

$$A_1 = \widetilde{A}_1 - u_1 v_1^T/\alpha_1 .$$

Define

$$L_1 = \begin{bmatrix} 1 & 0 \\ \alpha_1 & I \end{bmatrix} \in \mathbb{C}^{m\times m} \quad \text{and} \quad U_1 = \begin{bmatrix} \alpha_1^{-1}u_1 & v_1^T \\ 0 & A_1 \end{bmatrix} \in \mathbb{C}^{m\times n} .$$

and observe that

$$L_1^{-1} = \begin{bmatrix} 1 & 0 \\ -\alpha_1^{-1}u_1 & I \end{bmatrix} .$$

Hence (*) becomes

$$L_1^{-1}A = U_1, \text{ or equivalently, } \quad A = L_1 U_1 \ .$$

Note that L_1 is *unit* lower triangular (ones on the mail diagonal) and U_1 is block upper-triangular with one 1×1 block and one $(m-1) \times (n-1)$ block on the block diagonal. The multipliers are usually denoted

$$u/\alpha = [\mu_{21}, \ \mu_{31}, \ \ldots, \ \mu_{m1}]^T \ .$$

If the $(1, 1)$ entry of A_1 is not 0, we can apply the same procedure to A_1: if

$$A_1 = \begin{bmatrix} \alpha_2 & v_2^T \\ u_2 & \widetilde{A}_2 \end{bmatrix} \in \mathbb{C}^{(m-1)\times(n-1)}$$

with $\alpha_2 \neq 0$, letting

$$\widetilde{L}_2 = \begin{bmatrix} 1 & 0 \\ \alpha_2^{-1} u_2 & I \end{bmatrix} \in \mathbb{C}^{(m-1)\times(m-1)},$$

and forming

$$\widetilde{L}_2^{-1} A_1 = \begin{bmatrix} 1 & 0 \\ -\alpha_2^{-1} u_2 & I \end{bmatrix} \begin{bmatrix} \alpha_2 & v_2^T \\ u_2 & \widetilde{A}_1 \end{bmatrix} = \begin{bmatrix} \alpha_2 & v_2^T \\ 0 & A_2 \end{bmatrix} \equiv \widetilde{U}_2 \in \mathbb{C}^{(m-1)\times(n-1)},$$

where $A_2 \in \mathbb{C}^{(m-2)\times(n-2)}$. This process amounts to using the second row to zero out elements of the second column below the diagonal. Setting

$$L_2 = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{L}_2 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} \alpha_1 & v^T \\ 0 & \widetilde{U}_2 \end{bmatrix},$$

we have

$$L_2^{-1} L_1^{-1} A = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{L}_2^{-1} \end{bmatrix} \begin{bmatrix} \alpha_1 & v^T \\ 0 & A_1 \end{bmatrix} = U_2,$$

or equivalently,

$$A = L_2 L_1 U_2.$$

Here U_2 is block upper triangular with two 1×1 blocks and one $(m-2) \times (n-2)$ block on the diagonal, and again L_2 is unit lower triangular. We can continue in this fashion at most $\tilde{m} - 1$ times, where

$$\tilde{m} = \min\{m, n\}.$$

If we *can* proceed $\tilde{m} - 1$ times, then

$$L_{\tilde{m}-1}^{-1} \cdots L_2^{-1} L_1^{-1} A = U_{\tilde{m}-1} = U$$

is upper triangular provided that along the way that the $(1, 1)$ entries of

$$A, \ A_1, \ A_2, \ \ldots, \ A_{\tilde{m}-2}$$

are nonzero so the process can continue. Define

$$L = (L_{\tilde{m}-1}^{-1} \cdots L_1^{-1})^{-1} = L_1 L_2 \cdots L_{\tilde{m}-1}.$$

The matrix L is square unit lower triangular, and so is invertable. Moreover, $A = LU$, where the matrix U is the so called *row echelon form* of A. In general, a matrix $T \in \mathbb{C}^{m\times n}$ is said to be in row echelon form if for each $i = 1, \ldots, m-1$ the first non-zero entry in the $(i+1)^{\text{st}}$ row lies to the right of the first non-zero row in the i^{th} row.

Let us now suppose that $m = n$ and $A \in \mathbb{C}^{n\times n}$ is invertible. Writing $A = LU$ as a product of a unit lower triangular matrix $L \in \mathbb{C}^{n\times n}$ (necessarily invertible) and an upper triangular matrix $U \in \mathbb{C}^{n\times n}$ (also nessecarily invertible in this case) is called the *LU factorization* of A.

Remarks

(1) If $A \in \mathbb{C}^{n\times n}$ is invertible and has an LU factorization, it is unique.

(2) One can show that $A \in \mathbb{C}^{n\times n}$ has an LU factorization iff for $1 \le j \le n$, the upper left $j \times j$ principal submatrix

$$\begin{bmatrix} a_{11} & \cdots & a_{ij} \\ \vdots & & \\ a_{j1} & \cdots & a_{jj} \end{bmatrix}$$

is invertible.

(3) Not every invertible $A \in \mathbb{C}^{n\times n}$ has an LU-factorization.

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

Typically, one must permute the rows of A to move nonzero entries to the appropriate spot for the elimination to proceed. Recall that a permutation matrix $P \in \mathbb{C}^{n\times n}$ is the identity I with its rows (or columns) permuted: so

$$P \in \mathrm{R}^{n\times n} \text{ is orthogonal, and } P^{-1} = P^T.$$

Permuting the rows of A amounts to left multiplication by a permutation matrix P^T; then $P^T A$ has an LU factorization, so $A = PLU$ (called the PLU factorization of A).

(4) Fact: Every invertible $A \in \mathbb{C}^{n\times n}$ has a (not necessarily unique) PLU factorization.

(5) The LU factorization can be used to solve linear systems $Ax = b$ (where $A = LU \in \mathbb{C}^{n\times n}$ is invertible). The system $Ly = b$ can be solved by forward substitution (1^{st} equation gives x_1, etc.), and $Ux = y$ can be solved by back-substitution (n^{th} equation gives x_n, etc.), giving the solution to? $Ax = LUx = b$.

Example: We now use the procedure outlined above to compute the LU factorization of the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 2 \\ -1 & 1 & 3 \end{bmatrix}.$$

$$\begin{aligned} L_1^{-1} A &= \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 2 \\ -1 & 1 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 2 & 5 \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
L_2^{-1}L_1^{-1}A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 2 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix}
\end{aligned}
$$

We now have

$$
U = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix},
$$

and

$$
L = L_1 L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}.
$$