Matrix Division

Jeff Stuart

High School Algebra Revisited – The linear system $ax = b$ where a and b are real numbers.

Why do we need division? In its earliest form, division must have arisen to answer questions such as, "If our foraging party of three gathered seven melons, what is each members's share?" Mathematically, this is captured in the problem $3x = 7$ where x is each member's share. Division arises in portioning out shares, which is to say, in solving problems of the form $ax = b$ where a and b are known numbers. Of course, we all learned early in our educations that $ax = b$ has a unique solution when $a \neq 0$, namely, $x = b/a$. Although this seems trivially familiar to us, let's think more carefully about the solution set for $ax = b$. We have observed that when $a \neq 0$, there is a unique solution for every possible choice of b, namely $x = b/a$. What happens when $a = 0$? Probably, your first thought is that there is no solution since you cannot divide by zero. Is this correct? Certainly, if $b \neq 0$, we cannot find a real number x such that $0x \neq 0$. In the language of linear systems, when $a = 0$ and $b \neq 0$, the system $ax = b$ is inconsistent. But what happens when $b = 0$? In this case, the equation becomes $0x = 0$, which holds for every real number x. That is, when $a = 0$ and $b = 0$, the system $ax = b$ has infinitely many solutions. Summarizing:

Theorem 1 *Let a and b be real numbers.*

- *If $a = 0$, then for all real numbers b with $b \neq 0$, $ax = b$ is inconsistent; and for $b = 0$, $ax = b$ has infinitely many solutions.*
- *If $a \neq 0$, then for all real numbers b, $ax = b$ is consistent and has a unique solution, $x = b/a$.*

What happens for linear systems $Ax = b$ with more equations and more variables?

From the Trichotomy Theorem, we know that exactly one of three outcomes occurs for the linear system $Ax = b$ where A and b are known: the system is inconsistent, the system is consistent with a unique solution, or the system is consistent with infinitely many solutions. Is there some analogous property to $a \neq 0$ that distinguishes the situation in which there is a unique solution to $Ax = b$ when $Ax = b$ is solvable, or that distinguishes the situation in which $Ax = b$ is solvable for all vectors b?

When do multiple solutions occur for $Ax = b$? Recall that when we form the product Ax, we are creating a vector that is a linear combination of the columns of A using the scalars that are the entries of x. One consequence of this is that for any matrix A, we can always find a vector b for which $Ax = b$ is consistent; just choose any vector b that is a linear combination of the columns of A, or even simpler, just choose b to be one of the columns of the matrix A. Also recall that for a consistent system, multiple solutions occur exactly when there is an arbitrary choice for one or more of the variables. Such arbitrary choices correspond to columns in $rref(A)$ that do not contain leading 1's. Thus to preclude arbitrary choices, we must require that every column of $rref(A)$ is a leading 1 column. If A has n columns, then this statement is equivalent to $rank(A) = n$. Note that if $rank(A) < n$, then there will be choices of b for which $Ax = b$ has multiple solutions. (Indeed, every consistent system will have infinitely many solutions.) In particular, since $rank(A) \leq \min\{m, n\}$, if the matrix A has $n > m$, there must be choices of b for which $Ax = b$ has infinitely many solutions.

When does inconsistency occur for $Ax = b$? Recall, that inconsistency can arise only when there is at least one row of zeros in $rref(A)$, the reduced row echelon form of A. When such a row of zeros exists, we can strategically choose the vector b so that when we perform row operations on $[A|b]$ to obtain $[rref(A)|c]$ there is a nonzero entry in c across from a row of zeros. Thus to guarantee that inconsistency never arises, we must require that there are no zero rows in $rref(A)$. In other words, every row of $rref(A)$ must contain a leading 1. If A has m rows, this statement is equivalent to $rank(A) = m$. Note that if $rank(A) < m$, then there will be choices of b for which the system $Ax = b$ is inconsistent. In particular, since $rank(A) \leq \min\{m, n\}$, if the matrix A has $m > n$, there must be choices of b that make $Ax = b$ inconsistent.

Apparently, if we want to guarantee that we never have multiple solutions, then we need $n = rank(A)$, and if we want to guarantee that $Ax = b$ is always consistent, then we need $m = rank(A)$. The matrices A such that $Ax = b$ has a unique solution for every choice of b are exactly those for which $m = n = rank(A)$. We call such matrices *nonsingular* matrices. Note that A is nonsingular means that every row and every column of $rref(A)$ contains a leading 1, which means that we must have $rref(A) = I_n$.

We call matrices for which $m = n > rank(A)$ *singular* matrices. Note that if A is singular, then the number of leading 1's in $rref(A)$ must be less than the number of rows, hence $rref(A)$ has at least one zero row; and hence, for some choices of b, $Ax = b$ must be inconsistent. Since the number of leading 1's in $rref(A)$ must be less than the number of columns in A, there must be nonleading variables associated with columns of $rref(A)$ that do not contain leading 1's. Thus, for all choices of b for which $Ax = b$ is consistent, there must be infinitely many solutions.

We can answer our question about what property for $Ax = b$ is analogous to $a \neq 0$ for $ax = b$.

Theorem 2 ***(Refined Trichotomy Theorem)*** *Let A be an $m \times n$ real matrix. Let b be a vector in $\mathbf{R}^m$.*

- *If $m < n$, then there always exist vectors b such that $Ax = b$ has infinitely many solutions.*

- *If $m > n$, then there always exist vectors b such that $Ax = b$ is inconsistent.*
- *If $m = n$, and if A is singular, then for some choices of the vector b, $Ax = b$ is inconsistent; and for all choices of the vector b such that $Ax = b$ is consistent, $Ax = b$ has infinitely many solutions.*
- *If $m = n$, and if A is nonsingular, then $Ax = b$ has a unique solution for every choice of the vector x.*

One very important area of matrix theory is the study of what statements about a square matrix are equivalent to the statement that the matrix is nonsingular. From the discussion above, we can start this study.

Theorem 3 ***(Equivalence to Nonsingularity Theorem)*** *Let A be an $n \times n$ real matrix The following are equivalent.*

1. *The matrix A is nonsingular.*
2. *The matrix A has $rank(A) = n$.*
3. *The matrix A has $rref(A) = I_n$.*
4. *The unique solution to $Ax = \mathbf{0}_{n\times 1}$ is $x = \mathbf{0}_{n\times 1}$.*
5. *For some vector b in $\mathbf{R}^n$, $Ax = b$ has a unique solution.*
6. *For every vector b in $\mathbf{R}^n$, $Ax = b$ has a solution.*
7. *For every vector b in $\mathbf{R}^n$, $Ax = b$ has a unique solution.*

Why don't we just write $x = b/A$?

If we compare Theorem 1 and Theorem 2 above, we should notice that the result for $ax = b$ with $a \neq 0$, contains a formula for the unique solution: $x = b/a$. Is there an analogous manner for expressing the unique solution to $Ax = b$ when A is nonsingular? In particular, since we know there is a unique solution to $Ax = b$ when A is nonsingular, why don't we just write $x = b/A$?

When $a \neq 0$ and we solve $ax = b$ by dividing both sides by a to obtain $x = b/a$, what we are really doing is following the following steps:

$$\begin{aligned} ax &= b \\ \frac{1}{a}(ax) &= \frac{1}{a}(b) \\ \left(\frac{1}{a}\cdot a\right)x &= \frac{1\cdot b}{a} \\ 1x &= \frac{b}{a} \\ x &= \frac{b}{a} \end{aligned}$$

In particular, what we are really doing is introducing a real number c such that $c \cdot a = 1$, and then using the fact that $1x = x$. Of course, because we already understand fractions, we do not write c, but rather we write $\frac{1}{a}$. This suggests that when A is a nonsingular $n \times n$ real matrix, we want to find a real matrix C such that $CA = I_n$, so that we can follow the analogous steps:

$$\begin{aligned} Ax &= b \\ C(Ax) &= Cb \\ (CA)x &= Cb \\ I_n x &= Cb \\ x &= Cb \end{aligned}$$

Notice that this formulation allows us to avoid defining matrix division, and in particular, it enables us to avoid the fact that matrix division would have to reflect the noncommutativity of matrix multiplication. Given an $n \times n$ nonsingular matrix A, can we find a matrix C such that $CA = I_n$?

Recall that when we perform row operations on a matrix M to obtain a matrix N, we can achieve same result by matrix multiplication. That is, $M \to N$ via row operations means that there is a square matrix P such that $PM = N$, and further, P is a product of elementary row operation matrices. In particular, since $A \to rref(A)$ via row operations, we can find a square matrix C such that $CA = rref(A)$. When A is nonsingular, $rref(A) = I_n$. Thus when A is nonsingular, there must be at least one $n \times n$ matrix C such that $CA = I_n$!!!

Conversely, suppose that A is an $n \times n$ matrix, possibly singular and possibly nonsingular, and suppose that there were a matrix C such that $CA = I_n$. Clearly the requirement $CA = I_n$ forces C to be $n \times n$. Suppose that b in $\mathbf{R}^n$ is one of the vectors such that $Ax = b$ is consistent (for example, $b = \mathbf{0}$). Suppose also that v and w are two solutions of $Ax = b$. Then $v = I_n v = (CA)v = C(Av) = Cb$ and $w = I_n w = (CA)w = C(Aw) = Cb$, which forces $v = Cb = w$. That is, when $Ax = b$ is consistent, the solution must be unique. From the Equivalence to Nonsingularity Theorem, if there is a C such that $CA = I_n$, then A must be nonsingular.

Observe that for the matrices C and A where

$$C = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 3 \end{bmatrix} \quad \textit{and} \quad A = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \\ 0 & 0 \end{bmatrix},$$

that $CA = I_2$ holds but A is clearly not nonsingular. Thus we need to be careful when we use results based on C and A such that $CA = I_n$ that A is $n \times n$.

What can we determine about a matrix C such that $CA = I_n$ when A is $n \times n$ and nonsingular? Consider the problem $Cx = b$ where b is any vector in $\mathbf{R}^n$. If we let $x = Ab$, then $Cx = C(Ab) = (CA)b = I_n b = b$, which means that $Cx = b$ is always consistent. That is, C must be nonsingular! Since C is itself nonsingular, we can find a matrix D such that $DC = I_n$. This means that

$$\begin{aligned} (DC)A &= I_nA \\ D(CA) &= A \\ DI_n &= A \\ D &= A \end{aligned}$$

Notice that this says that A is the only choice for a matrix D such that $DC = I_n$. Summarizing:

Theorem 4 *Let A be an $n \times n$ matrix. Then A is nonsingular if and only if there is a matrix C such that $AC = CA = I_n$. Further such a matrix C is unique and nonsingular.*

We say that a pair of $n \times n$ matrices A and B are *inverses* of each other if $AB = BA = I_n$. When A has an inverse, we say that A is *invertible.* Since the inverse is unique when it exists, we write A^{-1} for the inverse of an invertible matrix A. Notice that this captures the idea of division since for a nonzero real number a, we can write $\frac{1}{a}$ as a^{-1}.

Summarizing, we can add two additional conditions to the Equivalence to Nonsingularity Theorem:

8. A is an invertible matrix (A^{-1} exists).
9. There is a matrix C such that at least one of $AC = I_n$ and $CA = I_n$ holds.

Finally, we can write the analog to $x = b/a$ when $a \neq 0$:

Lemma 5 *Let A be an $n \times n$ nonsingular matrix. Let b be any vector in $\mathbf{R}^n$. Then the unique solution to $Ax = b$ is $x = A^{-1}b$.*

Example 6 *Let A be the matrix*

$$A = \begin{bmatrix} 1 & 5 \\ 2 & 9 \end{bmatrix}.$$

Direct computation shows that

$$C = \begin{bmatrix} -9 & 5 \\ 2 & -1 \end{bmatrix}$$

satisfies $AC = CA = I_2$, so $C = A^{-1}$. If b is the vector

$$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

then the unique solution to $Ax = b$ is

$$x = A^{-1}b = \begin{bmatrix} -9 & 5 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -9b_1 + 5b_2 \\ 2b_1 - 1b_2 \end{bmatrix}.$$

One-sided inverses for nonsquare matrices

We have already seen that if A is square and there is a matrix B such that either $AB = I$ or $BA = I$, then $B = A^{-1}$ and $AB = BA = I$. What can we say about when A is $m \times n$ and $m \neq n$? If there exists an $n \times m$ matrix B such that $BA = I_n$ then we call B a *left inverse* of A. Similarly, if there exists an $m \times n$ matrix C such that $AC = I_m$, then we call C a *right inverse* of A.

Suppose that there is a matrix B such that $BA = I_n$. Now consider the linear system $Ax = \mathbf{0}$. Observe that

$$x = I_n x = (BA)x = B(Ax) = B\mathbf{0} = \mathbf{0}.$$

Apparently, $x = \mathbf{0}$ is the unique solution to the homogeneous system.

Next, suppose that there is a matrix C such that $AC = I_m$. Now consider the linear system $Ax = b$ where b is an arbitrary vector in $\mathbf{R}^m$. If we let $x = Cb$, then

$$Ax = A(Cb) = (AC)b = I_m b = b.$$

That is, $Ax = b$ always has at least one solution, $x = Cb$. Summarizing,

Lemma 7 *Let A be $m \times n$.*

1. *Let B be a left inverse of A. Then the only solution to the linear system $Ax = \mathbf{0}$ is $x = \mathbf{0}$.*
2. *Let C be a right inverse of A. Then for all b in $\mathbf{R}^m$, the system $Ax = b$ always has at least one solution, $x = Cb$.*

We have already encountered a connection between $rank(A)$ and the conclusions in the previous lemma. Specifically, we saw that a linear system with $m \times n$ coefficient matrix A has a unique solution exactly when $rank(A) = n$, and that a linear system with coefficient matrix A is always consistent exactly when $rank(A) = m$. We have seen that for square matrices, right inverses are left inverses are inverses, that inverses are unique, and that inverses only exist when the rank of the matrix is equal to the number of rows (equivalently, number of columns).. In this section, we will focus on nonsquare matrices.

Suppose that A is $m \times n$ with $m > n$. From the Refined Trichotomy Theorem, there must be choices of b such that $Ax = b$ is inconsistent, and hence, A cannot have a right inverse. What about a left inverse? Since $Ax = \mathbf{0}$ has multiple solutions unless every column of A corresponds to a leading 1 in the matrix $rref(A)$, $rank(A) < n$ implies that A cannot have a left inverse. Now suppose that $rank(A) = n$. Then since $m > n$ and A is $m \times n$,

$$rref(A) = \begin{bmatrix} I_n \\ \mathbf{0}_{(m-n)\times n} \end{bmatrix}.$$

There is an $m \times m$ invertible matrix P such that $PA = rref(A)$. Partition P as

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

where P_1 is $n \times m$ and where P_2 is $(m-n) \times m$. Then

$$PA = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} A = \begin{bmatrix} P_1 A \\ P_2 A \end{bmatrix}$$

so $P_1 A = I_n$. That is, A has a left inverse. Also notice that $P_2 A = \mathbf{0}_{(m-n)\times n}$, so if M is any $n \times (m-n)$ matrix,

$$\begin{aligned} (P_1 + MP_2)\, A &= P_1 A + (MP_2)A \\ &= I_n + M(P_2 A) \\ &= I_n + M\mathbf{0}_{(m-n)\times n} \\ &= I_n \end{aligned}$$

That is, for every choice of M, $P_1 + MP_2$ is a left inverse of A. Since P is invertible, P_2 cannot be the zero matrix, and hence, there are choices for M such that MP_2 is not the zero matrix. If M is any such matrix, then cM also works for all choices of the scalar c. This means that there are infinitely many different left inverses for A. Summarizing:

Lemma 8 *Let A be $m \times n$ with $m > n$. If rank$(A) < n$, then A does not have a left inverse. If rank$(A) = n$, then A has infinitely many left inverses, and row operations can be used to create a family of left inverses for A.*

Example 9 *Let P be the 3×3 nonsingular matrix and let A be the 3×2 matrix given by*

$$P = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix} \quad \textit{and} \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Then

$$PA = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_2 \\ \mathbf{0}_{1\times 2} \end{bmatrix},$$

so rank$(A) = 2$. Then

$$P_1 = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \end{bmatrix}$$

is a left inverse of A. Further, if $M = \begin{bmatrix} a \\ b \end{bmatrix}$, then

$$\begin{aligned} P_1 + MP_2 &= \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1-a & a & -1 \\ -1-b & 1+b & 1 \end{bmatrix} \end{aligned}$$

is a left inverse of A for all choices of real numbers a and b.

What happens when $m < n$? If $rank(A) < m$, then as argued in the paragraphs preceding the Refined Trichotomy Theorem, there must be inconsistent systems whose coefficient matrix is A. Thus, A cannot have a right inverse when $rank(A) < m$. What happens when $rank(A) = m$? When $rank(A) = m$, $Ax = b$ can be solved for every b in $\mathbf{R}^m$. In particular, if b is obtained from the zero vector by replacing a single 0 with a 1, we can solve the resulting linear system. Specifically, let the vector $e^{(j)}$ denote the j^{th} column of I_m. Then we can find at least one solution to $Ax = e^{(j)}$. Let $c^{(j)}$ in $\mathbf{R}^n$ be one such solution. Let C be the $n \times m$ matrix whose j^{th} column is $c^{(j)}$ for each j. Then since the j^{th} column of AC is exactly $Ac^{(j)} = e^{(j)}$, we have $AC = I_m$. Also note that since $m < n$, each of the linear systems $Ax = e^{(j)}$ has infinitely many solutions, hence there are infinitely many choices for C.

Summarizing:

Lemma 10 *Let A be $m \times n$ with $m < n$. If $rank(A) < m$, then A does not have a right inverse. If $rank(A) = m$, then A has infinitely many right inverses, and row operations can be used to create a family of right inverses for A.*

Exercise 11 *Let A be the 2×3 matrix*

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

Clearly $rank(A) = 2 = m$. Let $e^{(1)}$ and $e^{(2)}$ be the vectors

$$e^{(1)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } e^{(2)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

so $\begin{bmatrix} e^{(1)} & e^{(2)} \end{bmatrix} = I_2$. Solving $Ax = e^{(1)}$ gives

$$x = \begin{bmatrix} -1 \\ 2-\alpha \\ \alpha \end{bmatrix}$$

where α is an arbitrary scalar. Let $c^{(1)} = x$. Solving $Ax = e^{(2)}$ gives

$$x = \begin{bmatrix} 1 \\ -1-\beta \\ \beta \end{bmatrix}$$

where β is an arbitrary scalar. Let $c^{(2)} = x$. Then

$$C = \begin{bmatrix} -1 & 1 \\ 2-\alpha & -1-\beta \\ \alpha & \beta \end{bmatrix}$$

satisfies $AC = I_2$ for all scalars α and β. Thus there are infinitely many right inverses C for A.

If you carefully examine the arguments that we have used, you will observe that ultimately, we have used row operations on linear systems to construct inverses, right inverses and left inverses. Can every right inverse and every left inverse be obtained by row operations? Given a known $m \times n$ matrix A with $rank(A) = n$, finding an unknown $n \times m$ matrix B such that $BA = I_n$ is equivalent to solving n^2 linear equations in the unknown entries of B with coefficients taken from A whose right hand side values are the n^2 entries in I_n. For example, consider again the rank 2 matrix A given by

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

We seek all 2×3 matrices B such that $BA = I_2$. If

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix},$$

then $BA = I_2$ is equivalent to the four linear equations:

$$\begin{array}{ll} 1b_{11} + 1b_{12} + 0b_{13} = 1 & \qquad 1b_{11} + 1b_{12} + 1b_{13} = 0 \\ 1b_{21} + 1b_{22} + 0b_{23} = 0 & \qquad 1b_{21} + 1b_{22} + 1b_{23} = 1 \end{array}$$

That is, we are solving

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{12} \\ b_{13} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_{21} \\ b_{22} \\ b_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Clearly, the left inverse B must be constructible via applying row operations to a matrix. (In fact, the matrix is $[A^T \mid I_n]$.)

Given a known $m \times n$ matrix A with $rank(A) = m$, finding an unknown $m \times n$ matrix C such that $AC = I_m$ is equivalent to solving m^2 linear equations in the unknown entries of C with coefficients taken from A whose right hand side values are the m^2 entries in I_m. For example, consider again the rank 2 matrix A given by

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

We seek all 3×2 matrices C such that $AC = I_2$. If

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{23} \end{bmatrix},$$

then solving $AC = I_2$ can be accomplished by performing row operations on $[A \mid I_2]$.

In summary,

Theorem 12 *Let A be $m \times n$ with $m \neq n$. If $rank(A) < \min\{m, n\}$, then A has neither a left inverse nor a right inverse. If $rank(A) = m$, then A has infinitely many right inverses and no left inverse. If $rank(A) = n$, then A has infinitely many left inverses and no right inverse. Finally, all left inverses and all right inverses are obtained by applying row operations to some augmented matrix obtained from A and the identity matrix.*

A natural question to ask is what happens when $rank(A) < \min\{m, n\}$? For example, if A is the 2×2 matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

then $rank(A) = 1$, and hence, A has neither a right nor a left inverse. What kinds of inverse-like matrices exist for this A, and what properties do they satisfy? Such matrices are called *generalized inverses*, and have been an important area of research for almost one hundred years. For the given matrix A, every generalized inverse is a matrix G of the form

$$G = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where $a+b+c+d = 1$. Depending on which additional properties are required for G, other conditions are imposed in its entries. On widely used generalized inverse is called the Moore-Penrose inverse, denoted A^+. For the given matrix A,

$$A^+ = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

It is easy to check that $rank(A^+) = rank(A)$, that $AA^+A = A$, that $A^+AA^+ = A^+$, and that A^+A and AA^+ are symmetric, among the many other special relationships between A and A^+.