Linear Equations, Matrices, and Vectors

Here is a quick review of what we have learned about linear equations in the last couple of weeks. Let F be any infinite field. Let us start with the system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n &= b_m \end{aligned}$$

where the coefficients a_{ij} and b_i and the variables x_i are all in F. In fact, all scalars will be in the same field F, so I will stop saying this.

The $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

is called the *coefficient matrix*, while the $m \times (n+1)$ matrix

$$(A|b) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_n \end{pmatrix}$$

is called the *augmented matrix* of the equation. The notation $(A|b)$ comes from writing the scalars b_i in a column vector

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

We have learned that we can solve such a linear equation by row reduction (aka Gaussian elimination), which is just the matrix version of solving a system of linear equations by elimininination. We do this by performing the three kinds of elementary row operations

- multiplying a row by a nonzero scalar,
- adding a scalar multiple of a row to another row,
- switching two rows

on the augmented matrix until it is in *reduced row-echelon form* (see handout from Section 1.2 of Anton for a complete description of what reduced row-echelon form means).

We have learned to recognize when a system of linear equations has a unique solution. If it does, the reduced row-echelon form of the augmented matrix has an $n \times n$ identity matrix in the upper left part, possibly followed by rows of 0s if $m > n$. When the linear equation does not have a unique solution, the reduced row-echelon form of the augmented matrix has a more general shape: the leading 1s in the rows may not line up along the diagonal and there may be entries that are neither 0 nor 1 in the upper right part of the matrix. In this case, there could be infinitely many solutions or no solutions at all, which will depend on what the vector b is. If there is a row whose first n entries are 0 and the entry in the last column is nonzero, then there is no solution because such a row would correspond to the equation

$$0x_1 + 0x_2 + \cdots + 0x_n = b_i$$

for some $b_i \neq 0$, which clearly has no solution. Otherwise, one or more of the variables $x_1, \ldots, x_n$ is a *free variable*, i.e. its value could be anything in F. This means there are infinitely many solutions.

We have also learned to write the linear equation above in matrix-vector form as

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

or

$$Ax = b,$$

where x is the column vector

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

We saw that each elementary row operation corresponds to multiplying the augmented matrix by an elementary matrix. If $E_1, E_2, \ldots, E_k$ are the elementary matrices that correspond to the consecutive steps in the row-reduction, and $C = E_k E_{k-1} \cdots E_1$, then multiplying the augmented matrix $(A|b)$ by C does the same thing as the whole row-reduction. So $C(A|b)$ is the reduced row-echelon form of the augmented matrix, and CA is the reduced row-echelon form of the coefficient matrix A.

Now, if A is an $n \times n$ matrix, then we know that

- If $CA = I_n$, then

$$C(A|b) = \begin{pmatrix} 1 & 0 & \cdots & 0 & d_1 \\ 0 & 1 & \cdots & 0 & d_2 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & d_n \end{pmatrix},$$

  where d_i is the i-th entry of the column vector $d = Cb$. This corresponds to the equations

$$\begin{aligned} x_1 &= d_1 \\ x_2 &= d_2 \\ &\vdots \\ x_n &= d_n. \end{aligned}$$

  That is the equation has a unique solution in this case. Notice that C only depends on the coefficient matrix A and the matrix multiplication Cb can be done no matter what b is. So the equation would have a unique solution for any choice of the vector $b \in F^n$.
- Conversely, if the linear equation has a unique solution, then solving it by elimination should give some result of the form

$$\begin{aligned} x_1 &= d_1 \\ x_2 &= d_2 \\ &\vdots \\ x_n &= d_n. \end{aligned}$$

This means the reduced-row echelon form of the augmented matrix must look like

$$C(A|b) = \begin{pmatrix} 1 & 0 & \cdots & 0 & d_1 \\ 0 & 1 & \cdots & 0 & d_2 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & d_n \end{pmatrix}.$$

We can conclude CA must be the identity matrix I_n in this case.

Hence a system of n linear equations in n variables has a unique solution if and only if its coefficient matrix can be row-reduced to the identity I_n.

We have pointed out that elementary row operations on a matrix take linear combinations of the rows of that matrix. In fact, if C is as above, CA is a matrix that consists of linear combinations of the row vectors of A. For example, the first row of CA would be the vector

$$c_{11}a_{1*} + c_{12}a_{2*} + \cdots + c_{1n}a_{n*}$$

where

$$a_{i*} = (a_{i1}, a_{i2}, \ldots, a_{in})$$

is the i-th row of A. So if an $n \times n$ matrix A can be row-reduced to the identity matrix, then the row vectors of the identity matrix

$$\begin{aligned} e_1 &= (1, 0, \cdots, 0) \\ e_2 &= (0, 1, \cdots, 0) \\ &\vdots \\ e_n &= (0, 0, \cdots, 1) \end{aligned}$$

can be expressed as linear combinations of the rows $a_{1*}, \ldots, a_{n*}$:

$$\begin{aligned} e_1 &= c_{11}a_{1*} + c_{12}a_{2*} + \cdots + c_{1n}a_{n*} \\ e_2 &= c_{21}a_{1*} + c_{22}a_{2*} + \cdots + c_{2n}a_{n*} \\ &\vdots \\ e_n &= c_{n1}a_{1*} + c_{n2}a_{2*} + \cdots + c_{nn}a_{n*} \end{aligned}$$

Now, since every vector $v = (v_1, v_2, \ldots, v_n) \in F^n$ can be expressed as a linear combination of $e_1, \ldots, e_n$, namely

$$v = v_1e_1 + v_2e_2 + \cdots + v_ne_n,$$

it can also be expressed as a linear combination of $a_{1*}, \ldots, a_{n*}$ by replacing each e_i by $e_i = c_{i1}a_{1*} + c_{i2}a_{2*} + \cdots + c_{in}a_{n*}$ above. That is the row vectors of A span F^n. Now, if the row vectors $a_{1*}, \ldots, a_{n*}$ span F^n then they also form a basis of F^n since there are exactly $\dim(F^n) = n$ of them. This means they must be linearly independent. We can reverse this argument: if the rows of A are linearly independent then they form a basis of F^n, hence they span F^n, hence each element e_i of the standard basis is a linear combination of the rows of A:

$$e_i = c_{i1}a_{1*} + c_{i2}a_{2*} + \cdots + c_{in}a_{n*}$$

for some coefficients c_{ij}. Now if C is the matrix (c_{ij}) then $CA = I_n$. So we have just proved

Theorem 1. *If $A \in M_{n\times n}(F)$ then the following are equivalent:*
(a) A has a left inverse $C \in M_{n\times n}(F)$ such that $CA = I_n$.
(b) The rows of A span F^n.
(c) The rows of A are linearly independent.
(d) The rows of A form a basis of F^n.

Note that if A can be row-reduced to I_n then it satisfies all of the statements in the theorem above. We will show soon that the converse is also true: if A satisfies the statements in the theorem above, then it can be row-reduced to I_n.

Another way to view the original linear equation is as

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

That is we are looking for coefficients $x_1, \ldots, x_n$ to express the vector b as a linear combination of the column vectors $a_{*1}, a_{*2}, \ldots, a_{*n}$ of A. We noted earlier that if A is an $n \times n$ matrix that can be row-reduced to I_n, then the equation $Ax = b$ has a unique solution for any choice of b. This means that every vector in F^n is a linear combination of the column vectors of A, that is the column vectors of A span F^n. If so, the column vectors of A form a basis of F^n since $\dim(F^n) = n$. Hence they are also linearly independent. Also, since every vector in F^n is in the span of the column vectors of A, in particular, the elements of the standard basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

can all be expressed as linear combinations of the columns of A as

$$e_j = d_{1j}a_{*1} + d_{2j}a_{*2} + \cdots + d_{nj}a_{*n}$$

for some coefficients $d_{1j}, \ldots, d_{nj}$. If D is the matrix (d_{ij}) then $AD = I_n$, that is D is a right inverse of A. Just like with the rows and the left inverse, this argument also works in reverse. So we now have

Theorem 2. *If $A \in M_{n\times n}(F)$ then the following are equivalent:*
(a) A has a right inverse $D \in M_{n\times n}(F)$ such that $AD = I_n$.
(b) The columns of A span F^n.
(c) The columns of A are linearly independent.
(d) The columns of A form a basis of F^n.

Now, note that we can solve any of the linear equations

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} d_{1j} \\ d_{2j} \\ \vdots \\ d_{nj} \end{pmatrix} = e_j$$

by row reducing the augmented matrix $(A|e_j)$. In fact, we can solve all of them at the same time by row-reducing the augmented matrix

$$(A|I_n) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & 1 & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & a_{2n} & 0 & 1 & \ldots & 0 \\ \vdots & & & \vdots & \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

If A can be row-reduced to I_n, then the result of row-reducing $(A|I_n)$ is a matrix $(I_n|D)$ such that $AD = I_n$. Now we know that if A can be row-reduced to I_n then it satisfies the statements in both Theorems 1 and 2. Let us call a square matrix that can row-reduced to I_n *nonsingular* and a square matrix that cannot be row-reduced to I_n *singular*.

We have just learned that a nonsingular square matrix A has both a left inverse C and a right inverse D. In fact, these must be equal because

$$C = CI_n = C(AD) = (CA)D = I_nD = D,$$

where we used the fact that matrix multiplication is associative. That is a nonsingular square matrix has an inverse. In fact, the proof above also shows that the inverse is unique.

The same way we can row-reduce a matrix, we can also column-reduce it by doing elementary column operations:

- multiplying a column by a nonzero scalar,
- adding a scalar multiple of a column to another column,
- switching two columns.

Such column operations correspond to multiplying A by corresponding elementary matrices on the right.

We can of course make the analogous arguments about column-reduction as about row-reduction. That is a square matrix that can be column-reduced to I_n also has an inverse, and both its rows and columns are bases of F^n. We could find the left inverse (and hence the inverse) of such a matrix A by column-reducing the augmented matrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Now, let $A \in M_{n\times n}(F)$ such that A has a right inverse D. Then the column vectors of A span F^n. Let $T \in \mathcal{L}(F^n, F^n)$ be the linear map

$$T(x) = Ax.$$

As we noted above Ax is a linear combination of the columns of A. Then

$$\mathrm{range}(T) = \{T(x) \mid x \in F^n\} = \{Ax \mid x \in F^n\} = \mathrm{span}(a_{*1}, a_{*2}, \dots, a_{*n}) = F^n.$$

By the Fundamental Theorem of Linear Maps

$$\dim(F^n) = \dim(\mathrm{null}(T)) + \dim(\mathrm{range}(T)).$$

Since $\dim(F^n) = \dim(\mathrm{range}(T)) = n$, we find that $\dim(\mathrm{null}(T)) = 0$. That is $\mathrm{null}(T) = \{0\}$. This means that the only solution of the homogeneous linear equation $Ax = 0$ is $x = 0$. As we already noted above, this means that solving this equation by row-reducing the augmented matrix $(A|0)$ must result in $(I_n|0)$. That is A is a nonsingular matrix. By Theorem 1, A has a left inverse as well. That is a square matrix that has a right inverse also has a left inverse and hence has an inverse. By symmetry, a square matrix that has a left inverse must also be nonsingular and must therefore have a right inverse and a two-sided inverse as well. This nicely completes what we know about linear equations, matrices, and row and column-reduction. We can summarize it as

Theorem 3. *Let $A \in M_{n\times n}(F)$. Then the following are equivalent:*

(a) A can be row-reduced to I_n.

(b) A can be column-reduced to I_n.

(c) A has a left inverse $C \in M_{n\times n}(F)$ *such that* $CA = I_n$.
(d) A has a right inverse $D \in M_{n\times n}(F)$ *such that* $AD = I_n$.
(e) A has an inverse $A^{-1} \in M_{n\times n}(F)$ *such that* $AA^{-1} = A^{-1}A = I_n$.
(f) The rows of A span F^n.
(g) The columns of A span F^n.
(h) The rows of A are linearly independent.
(i) The columns of A are linearly independent.
(j) The rows of A form a basis of F^n.
(k) The columns of A form a basis of F^n.

A square matrix that has an inverse is called *invertible*. We have learned that square matrices that are invertible are exactly those that are nonsingular.

Let me note for the sake of completeness that all of the above works the same if F is a finite field, such as Z_n, but instead of infinitely many solutions, some linear equations would have more than one but finitely many solutions, as a free variable would have finitely many different values.