MATH 3A WEEK II
MATRICES

PAUL L. BAILEY

1. Motivation for Matrices

A linear transformation is completely described by its effect on the standard basis. Given a linear transformation, we wish to compress this information (its effect on the standard basis) into as tight a package as we can; we will call this package a matrix.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $x \in \mathbb{R}^n$ be an arbitrary vector, and call $T(v)$ the *destination* of x under T. To understand the transformation T, we wish to find a formula for $T(x)$. Since $T(x)$ is an element of $\mathbb{R}^m$, it is a linear combination of the standard basis vectors $e_1, \dots, e_m$ in $\mathbb{R}^m$; thus there exist $c_i \in \mathbb{R}$ as i runs from 1 to m such that $T(x) = \sum_{i=1}^m c_i e_i$; we seek a formula for the coefficients c_i.

Now x is a linear combination of the standard basis vectors in $\mathbb{R}^n$, so there exist $b_j \in \mathbb{R}$ as j runs from 1 to n such that $x = \sum_{j=1}^n b_j e_j$. Since T is linear, we see that $T(x) = \sum_{j=1}^n b_j T(e_j)$; thus if we know where T sends the standard basis vectors, we entirely understand T.

For each standard basis vector $e_j \in \mathbb{R}^n$, $T(e_j) \in \mathbb{R}^m$ so $T(e_j)$ is a linear combination of the standard basis vectors in $\mathbb{R}^m$. Fixing j, we see that there are real numbers a_{ij}, as i runs from 1 to m, such that $T(e_j) = \sum_{i=1}^m a_{ij} e_i$. For our arbitrary vector v we can write

$$T(x) = \sum_{j=1}^n b_j T(e_j) = \sum_{j=1}^n b_j \Big(\sum_{i=1}^m a_{ij} e_i \Big) = \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij} b_j \Big) e_i.$$

The final expression in the above equation reveals that the coefficients of $T(x)$ are given by

$$c_i = \sum_{i=1}^n a_{ij} b_j = (a_{i1}, \dots, a_{in}) \cdot (b_1, \dots b_n);$$

that is, the i^{th} coefficient of $T(v)$ is the vector whose components are the i^{th} coordinates of the destinations of the standard basis vectors dotted with v.

Thus T is completely described by the numbers a_{ij}, as i runs from 1 to m and j runs from 1 to n. These numbers form a mathematical object known as a matrix. The formula for c_i above motivates our definition of matrix multiplication.

Date: August 24, 1998.

2. Matrices

Let m, n be positive integers. An $m \times n$ *matrix* with real entries is an array of real numbers with m rows and n columns. We put brackets around the numbers; thus if A is an $m \times n$ matrix, we write

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

where a_{ij} is the real number in the i^{th} row and the j^{th} column. This can become a lot of writing; we use an abbreviated notation

$$(\text{number})_{\text{slot}}.$$

In our case,

$$A = (a_{ij})_{ij}.$$

This notation means that a_{ij} is in the ij^{th} slot. You may ask, "why repeat the ij"? The reason is, the number in the ij^{th} slot is not always indexed by ij. For example, if A is a 2×3 matrix written as $A = (2)_{ij}$ and B is a 3×2 matrix written as $B = (3j - i)_{ij}$, then then

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 5 \\ 1 & 4 \\ 0 & 3 \end{bmatrix}.$$

The *transpose* of an $m \times n$ matrix $A = (a_{ij})_{ij}$ is the $n \times m$ matrix A^t whose rows are the columns of A and whose columns are the rows of A:

$$A^t = (a_{ji})_{ij}.$$

Note that $(A^t)^t = A$. An $m \times n$ matrix is called *square* if $m = n$. A matrix A is *symmetric* if $A^t = A$; note that only square matrices can be symmetric.

A *row vector* is an $1 \times n$ matrix, and a *column vector* is a $m \times 1$ matrix. Note that if v is a column vector, then v^t is a row vector. From now on, whenever we need to consider a vector from $\mathbb{R}^n$ as a matrix, we consider to be a column vector.

Let A be an $m \times n$ matrix. Denote the i^{th} row of A by $A_{(i)}$ and the j^{th} column of A by $A^{(j)}$. Thus $A_{(i)}$ is a $1 \times n$ row vector and $A^{(j)}$ is an $m \times 1$ column vector.

Let $v_1, \dots, v_n \in \mathbb{R}^m$. We consider these to be column vectors. Let

$$A = [v_1 \mid \cdots \mid v_n]$$

denote the matrix whose j^{th} column is v_j; thus $A^{(j)} = v_j$.

3. Matrix Addition and Scalar Multiplication

Let $A = (a_{ij})_{ij}$ and $B = (b_{ij})_{ij}$ be $m \times n$ matrices. We define the *matrix sum* $A + B$ by

$$A + B = (a_{ij} + b_{ij})_{ij}.$$

We can only add matrices of the same size. Note that if A is square, then $A + O = O + A = A$, where O is the zero matrix of the same size.

Let $A = (a_{ij})_{ij}$ be an $m \times n$ matrix and let $c \in \mathbb{R}$. We define the *scalar multiplication* cA by

$$cA = (ca_{ij})_{ij}.$$

We define $-A$ to be the scalar product of -1 and A.

Note that the sum of column vectors is a column vector, and a scalar multiple of a column vector is a column vector. Indeed, for the case of column vectors, the definitions of matrix addition and scalar multiplication agree with the definitions we previously gave for vectors in $\mathbb{R}^n$.

The *zero matrix* of size $m \times n$, denoted by $Z_{m \times n}$ or simply by Z, is the $m \times n$ matrix for which every entry is equal to zero: $Z_{m \times n} = (0)_{ij}$.

Properties of Matrix Addition and Scalar Multiplication Let A and B be $m \times n$ matrices and let $c \in \mathbb{R}$ be a scalar. Then

- **(a)** $A + B = B + A$;
- **(b)** $(A + B) + C = A + (B + C)$;
- **(c)** $A + Z = A$;
- **(d)** $A + (-A) = Z$;
- **(e)** $c(A + B) = cA + cB$.

Remark. These properties are proved directly from the definitions. $\square$

4. Matrix Multiplication

Let $A = (a_{ij})_{ij}$ be an $m \times n$ matrix and let $B = (b_{jk})_{jk}$ be an $n \times p$ matrix. We define the *matrix product* of A and B to be the $m \times p$ matrix AB given by

$$AB = (c_{ik})_{ik}, \qquad \text{where} \qquad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Viewing the i^{th} row of A and the j^{th} column of B as vectors in $\mathbb{R}^n$, we see that

$$c_{ik} = A_{(i)} \cdot B^{(j)}.$$

We have no definition for the product of an $m \times n$ matrix with a $p \times q$ matrix unless $n = p$. If $v, w \in \mathbb{R}^n$ as considered as column vectors, then $v^t w = v \cdot w$.

The *identity matrix* of dimension n, denoted by I_n or simply by I, is the $n \times n$ matrix whose entries are one along the diagonal and zero everywhere else: $I_n = (\delta_{ij})_{ij}$, where δ_{ij} is the "Kronecker delta" defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{otherwise .} \end{cases}$$

Properties of Matrix Multiplication Let A and C be $m \times n$ matrices, B and D be $n \times p$ matrices, and E be a $p \times q$ matrix. Let $c \in \mathbb{R}$ be a scalar. Then

(a) $A(BE) = (AB)E$;
(b) $I_m A = A$;
(c) $AI_n = A$;
(d) $(A + C)B = AB + CB$;
(e) $A(B + D) = AB + AD$;
(f) $c(AB) = A(cB)$;
(g) $(AB)^t = B^t A^t$
(h) $(AB)_{(i)} = A_{(i)} B$;
(i) $(AB)^{(k)} = AB^{(k)}$;
(j) $(AB)_{(i)}^{(k)} = A_{(i)} B^{(k)}$.

Remark. These properties may be proved directly from the definitions, although in some cases this could lead to a lot of notation. Of paramount importance to us are properties **(e)** and **(f)**, and we will soon examine them more closely. □

Matrix multiplication is NOT commutative.

Let $x = (x_1, \ldots, x_n)$ be a vector in $\mathbb{R}^n$. We view x as a column vector, that is, as an $n \times 1$ matrix. Thus if $A = (a_{ij})_{ij}$ is an $m \times n$ matrix, the product Ax is defined to be an $m \times 1$ matrix:

$$Ax = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

Using the distributive property of scalar multiplication over matrix addition, we see that

$$Ax = x_1 A^{(1)} + \cdots + x_n A^{(n)}.$$

This $m \times 1$ column vector is a linear combination of the columns of A.

5. Matrices and Linear Transformations

We now consider the geometric interpretation of the product of a matrix and a column vector. First we prove that this operation is linear.

Proposition 1. *Let A be an $m \times n$ matrix. Then $Ae_j = A^{(j)}$, where e_j is the j^{th} standard basis vector in $\mathbb{R}^n$.*

Proof. Since $e_j = (0, \dots, 1, \dots, 0)$, with 1 in the j^{th} slot, we have

$$Ae_j = 0 \cdot A^{(1)} + \cdots + 1 \cdot A^{(j)} + \cdots + 0 \cdot A^{(n)} = A^{(j)}.$$

□

Proposition 2. *Let A be an $m \times n$ matrix. Then*

(a) *$A(x + y) = Ax + Ay$ for all $x, y \in \mathbb{R}^n$;*
(b) *$A(ax) = a(Ax)$ for all $x \in \mathbb{R}^n$, $a \in \mathbb{R}$.*

Proof. Let $x, y \in \mathbb{R}^n$. Then $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ for some $x_i, y_i \in \mathbb{R}$. By definition of vector addition and matrix multiplication,

$$\begin{aligned} A(x+y) &= (x_1 + y_1)A^{(1)} + \cdots + (x_n + y_n)A^{(n)} \\ &= (x_1 A^{(1)} + \cdots + x_n A^{(n)}) + (y_1 A^{(1)} + \cdots + y_n A^{(n)}) \\ &= Ax + Ay. \end{aligned}$$

Now let $a \in \mathbb{R}$. Then

$$\begin{aligned} A(ax) &= ax_1 A^{(1)} + \cdots + ax_n A^{(n)} \\ &= a(x_1 A^{(1)} + \cdots + x_n A^{(n)}) \\ &= a(Ax). \end{aligned}$$

□

Proposition 3. *Let A be an $m \times n$ matrix. Define a function*

$$T_A : \mathbb{R}^n \to \mathbb{R}^m \quad \text{by} \quad T_A(x) = Ax.$$

Then T is a linear transformation.

Proof. This is immediate from the previous proposition. □

Proposition 4. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Define a matrix*

$$A_T = [T(e_1) \mid \cdots \mid T(e_n)].$$

Then

(a) *$T(x) = A_T x$ for all $x \in \mathbb{R}^n$;*
(b) *$T_{A_T} = T$;*
(c) *$A_{T_A} = A$.*

Proof. A linear transformation is completely determined by its effect on the standard basis. The effect of A_T on the standard basis is the same as that of T; but A_T induces a linear transformation, so it must be the transformation T. □

Thus $m \times n$ matrices correspond to linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. The zero matrix corresponds to the zero transformation (that transformation which sends every element to the origin), and the identity matrix corresponds to the identity transformation (that transformation which sends every element to itself).

We emphasize that the columns of a matrix A are the destinations of the standard basis vectors.

Example 1. Find the matrix R_θ of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates the plane by an angle of θ radians.

Solution. We only need to discover what T does to the standard basis vectors. We see that $T(e_1) = (\cos\theta, \sin\theta)$ and $T(e_2) = (-\sin\theta, \cos\theta)$. Then

$$R_\theta = A_T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Select a point on the unit circle to test this: Then

$$\begin{aligned} R_\theta \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} &= \begin{bmatrix} \cos\theta\cos\alpha - \sin\theta\sin\alpha \\ \sin\theta\cos\alpha + \cos\theta\sin\alpha \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta+\alpha) \\ \sin(\theta+\alpha) \end{bmatrix}; \end{aligned}$$

this is what we would expect. □

Example 2. Find the matrix F_θ of the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which reflects the plane across a line whose angle with the x-axis is θ.

Solution. We see that

$$T(e_1) = (\cos 2\theta, \sin 2\theta)$$

and that

$$T(e_2) = -(\cos(2\theta + \frac{\pi}{2}), \sin(2\theta + \frac{\pi}{2}) = (\sin 2\theta, -\cos 2\theta).$$

Thus

$$F_\theta = \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}.$$

□

6. Matrices and Compositions of Linear Transformations

We now consider the geometric interpretation of matrix multiplication. Recall that if $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}_p \to \mathbb{R}_n$ are linear transformations, then the composition $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ given by $T \circ S(x) = T(S(x))$ is a linear transformation.

Proposition 5. *Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Then*

$$T_{AB} = T_A \circ T_B.$$

Proof. This means that the transformation associated to a product of matrices is the composition of the associated transformations. To show this, we only need to show that these transformations have the same effect on an arbitrary basis vector $e_k \in \mathbb{R}^p$.

The k^{th} column of AB is equal to A times the k^{th} column of B, and we have seen that multiplying a matrix by e_k picks out the k^{th} column, so

$$T_{AB}(e_k) = (AB)^{(k)} = AB^{(k)}.$$

On the other hand,

$$T_A \circ T_B(e_k) = T_A(B^{(k)}) = AB^{(k)}.$$

Thus these transformations have the same effect on e_k, and we conclude that they are the same transformation. □

Proposition 6. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}_p \to \mathbb{R}_n$ be linear transformations. Then*

$$A_{T \circ S} = A_T A_S.$$

Proof. This means that the matrix associated to a composition of transformations is the product of the associated matrices. It suffices to show that the k^{th} column of $A_{T \circ S}$ is the same as the k^{th} column of $A_T A_S$. But

$$A^{(k)}_{T \circ S} = T \circ S(e_k) = T(S(e_k))$$

and

$$A_T A^{(k)}_S = A_T A_S e_k = A_T S(e_k) = T(S(e_k)).$$

□

Example 3. Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which stretches the plane horizontally by a factor of 2, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which rotates the plane by 90 degrees (all rotations are counterclockwise). Then

$$A = A_S = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = A_T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Note that intuitively we see that $T \circ S$ and $S \circ T$ have different effects on the plane. Indeed,

$$BA = \begin{bmatrix} 0 & -1 \\ 2 & 0 \end{bmatrix} \quad \text{but} \quad AB = \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}.$$

We see that matrix multiplication is NOT commutative.

Example 4. Show that the composition of rotations is a rotation whose angle is the sum of the original angles.

Solution. We compute with matrices:

$$\begin{aligned} R_\alpha R_\beta &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix} \\ &= \begin{bmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\cos\alpha\sin\beta - \sin\alpha\cos\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & \sin\alpha\sin\beta + \cos\alpha\cos\beta \end{bmatrix} \\ &= \begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix}. \end{aligned}$$

□

7. Matrices and Invertible Linear Transformations

Recall that the identity transformation $J_n : \mathbb{R}^n \to \mathbb{R}^n$ is the function that has no effect on $\mathbb{R}^n$; it is given by $J_n(v) = v$. Since the identity matrix I_n has no effect on the standard basis (viewed as column vectors), we see that

$$A_{J_n} = I_n \quad \text{and} \quad T_{I_n} = J_n.$$

Recall that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is *invertible* if it is bijective, in which case there is an inverse function $T^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ such that $T^{-1} \circ T = J_n$.

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is an invertible linear transformation. We will see that this implies $m = n$; for now, just assume $m = n$. Then the matrix corresponding to T is an $n \times n$ matrix; that is, it is square.

Suppose that T is invertible; then $T \circ T^{-1} = T^{-1} \circ T = \mathrm{id}_{R^n}$. Then $A_T A_{T^{-1}} = A_{T^{-1}} A_T = A_{\mathrm{id}} = I$.

A matrix A is called *invertible* if there exists a matrix B such that $AB = BA = I$. We see that two matrices are invertible if and only if the corresponding linear transformations are bijective. The matrix B is called the *inverse* of A, and is denoted by A^{-1}. Note that A^{-1} is invertible (with inverse A).

Properties of Matrix Inverses

Let A, B, C, and D be square matrices of the same size.

- **(a)** Inverses are unique.
- **(b)** If A and B are invertible, then so is AB, with $(AB)^{-1} = B^{-1}A^{-1}$.
- **(c)** If $AC = DA = I$, then $C = D$.
- **(d)** If $AB = I$, then $BA = I$, so A and B are invertible.

Proof of **(c)** is in FB §1.5 Theorem 1.9. Proof of **(d)** is postponed for now.

Now if $T : \mathbb{R}^n \to \mathbb{R}^n$ is bijective, then for every $b \in \mathbb{R}^n$ there exists a unique $x \in \mathbb{R}^n$ such that $T(x) = b$; indeed, we have $x = T^{-1}(b)$. In matrix form, this says that the matrix equation $Ax = b$ has a unique solution, given by $x = A^{-1}b$.

We would like a method to find A^{-1}. The idea is to "dissolve" A by multiplying both sides of the equation $AX = I$ with invertible matrices: $E_1AX = E_1I = E_1$, then $E_2E_1AX = E_2E_1$, et cetera, at each step getting closer to the identity (e.g. E_2E_1A looks more like the identity than E_1A), until finally we obtain $E_n \cdots E_1AX = E_n \cdots E_1$, where $E_n \cdots E_1A = I$, so $X = E_n \cdots E_1$. Now X is the product of invertible matrices, so it is invertible, and it is the inverse of A since $AX = I$.

8. Elementary Row Operations and Elementary Invertible Matrices

The invertible matrices E_i mentioned above are called "elementary"; they correspond to *elementary row operations*. A row operation is a way of modifying a row of a matrix to change it into a different matrix. Tradition demands that we list three *elementary row operations*:

$\mathsf{R}_i + c\mathsf{R}_j$	Type **E**	Multiply j^{th} row by c and add to i^{th} row
$c\mathsf{R}_i$	Type **D**	Multiply i^{th} row by c
$\mathsf{R}_i \leftrightarrow \mathsf{R}_j$	Type **P**	Swap the i^{th} row and the j^{th} row

For each of these three row operations, there is an invertible matrix E such that EA is the result of the row operation applied to A. To find E, just perform the row operation on the identity matrix.

$E(i,j;c)$ is I except $a_{ij} = c$; $E(i,j;c)^{-1} = E(i,j;-c)$.

$D(i;c)$ is I except $a_{ii} = c$; $D(i;c)^{-1} = D(i;c^{-1})$.

$P(i,j)$ is I except $a_{ii} = a_{jj} = 0$ and $a_{ij} = a_{ji} = 1$; $P(i,j)^{-1} = P(i,j)$.

We give an organized algorithm for applying row operations to attempt to find the inverse of a matrix.

Algorithm for Row Reduction of a Square Matrix to Find an Inverse

- Make all entries below the diagonal into zero, starting with the second entry in the first column, proceeding downward, then doing the third column, etc.
- Make all diagonal entries equal to one.
- Make all entries above the diagonal zero, starting with the lowest entry in the last column, working upward in that column, then starting on the next to last column, etc.

Step one is always possible; it may be necessary to swap some rows to do this. Use only type **E** and **P** row operations.

Step two is possible if all diagonal entries are nonzero via use of type **D** row operations; otherwise, the matrix is not invertible. To see this, let Q be the matrix obtained after step one, and suppose that $Q^{(i)}_{(i)} = 0$ is the first zero diagonal entry. Then all entries in column i below the diagonal are also zero, so $Q^{(i)}$ is a linear combination of the previous columns (to see this may take some effort, but it is true); say $Q^{(i)} = a_1 Q^{(1)} + \cdots + a_{i-1} Q^{(i-1)}$. Then $a_1 e_1 + \cdots + a_{i-1} Q^{(i-1)} - e_i$ is in the kernel of T_A, so T_A is not injective, and A is not invertible.

Step three is possible whenever step two is possible. Use only type **E** row operations.

Thus every invertible matrix is a product of elementary invertible matrices. To see this, let A be invertible and suppose that X is its inverse. Then $AX = I$. Following the above algorithm, we obtain elementary invertible matrices $E_1, \ldots, E_r$ such that

$$X = E_r \cdots E_1 AX = E_r \cdots E_1 I = E_r \cdots E_1.$$

9. Introduction to Linear Equations

Consider the system of linear equations

$$\begin{aligned} 3x_1 - 4x_2 &= 11; \\ x_1 + 2x_2 &= 7. \end{aligned}$$

Solving this system means finding x_1 and x_2 which make the equations true.

The loci of the equations $3x_1 - 4x_2 = 11$ and $x_1 + 2x_2 = 7$ are lines in $\mathbb{R}^2$ (we have replaced the standard x and y by x_1 and x_2 because we want to use the variables x and y to indicate vectors).

A second geometric interpretation of the problem comes from forming the matrix of coefficients and the column vectors

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

and considering the matrix equation $Ax = b$. Since A corresponds to a linear transformation, solving the system of equations is equivalent to finding the preimage of the point b under this linear transformation.

To solve this system, we can multiply the second equation by 2 and add it to the first to get $5x_1 = 25$, so $x_1 = 5$; then plug this into the second equation to get $5 + 2x_2 = 7$, so $x_2 = 1$.

Generalizing this solution technique to many equations in many unknowns could lead to a lot of confusion and difficulty without a more organized approach. We now search for a failsafe algorithm for finding the solution.

10. Linear Equations

A *linear equation* in n variables $x_1, \dots, x_n$ is an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b_1,$$

where $a_1, \dots, a_n, b_1 \in \mathbb{R}$ are fixed constants.

Let $a = (a_1, \dots, a_n)$, $x = (x_1, \dots, x_n)$, and $q = (\frac{b_1}{a_1}, 0, \dots, 0)$. The above equation becomes $a \cdot x = a \cdot q$, or

$$(x - q) \cdot a = 0.$$

We recognize this as the equation of a hyperplane in $\mathbb{R}^n$ through the point q with normal vector a.

Consider an arbitrary system of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \vdots &= \vdots \\ a_{i1}x_1 + \cdots + a_{in}x_n &= b_i \\ \vdots &= \vdots \\ a_{1n}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where $a_{ij}, b_i \in \mathbb{R}$ are constants and x_i are indeterminants.

Our goal is to use invertible matrices to help us solve such systems of linear equation; that is, we wish to find all vectors $x \in \mathbb{R}^n$ such that when we plug their coordinates into the equations, all of the resulting equations are true.

One geometric interpretation of this problem is to find the intersection of the hyperplanes in $\mathbb{R}^n$ which are the loci of the given equations.

A second geometric interpretation comes from forming the matrix $A = (a_{ij})_{ij}$. Then setting $x = (x_1, \dots, x_n)$ and $b = (b_1, \dots, b_m)$, we see that the solution set of the matrix equation $Ax = b$ is exactly the solution set of the system of equations. This matrix equation, stated in terms of linear transformations, is $T_A(x) = b$; solving means finding $T_A^{-1}(b)$, the preimage of the point b under the linear transformation T_A.

Our approach to the problem uses matrices; we seek column vectors x such that $Ax = b$.

A *general solution* is the set of all such column vectors x.

A *particular solution* is a specific such column vector x.

The system is called *homogeneous* if $b_i = 0$ for $i = 1, \dots, m$. In this case, solving the system of equations means finding the kernel of T_A. Otherwise, the system is *nonhomogeneous*.

We have seen that if b is in the image of T_A, say $T_A(v) = b$ for some $v \in \mathbb{R}^n$, then $T^{-1}(b) = v + \ker(T_A)$. If T_A is injective, then $\ker(T_A)$ consists of a single point (the origin), so $T^{-1}(b) = \{v\}$. Otherwise, $\ker(T_A)$ is a nontrivial subspace, so $v + \ker(T_A)$ at least one line, and possibly a plane or more. Thus there are three possibilities:

(1) there are no solutions (b is not in the image of T_A);
(2) there is exacly one solution (T_A is injective);
(3) there are infinitely many solutions (T_A has a nontrivial kernel).

If we have infinitely many solutions, they are of the form

$$v_0 + c_1 v_1 + \cdots + c_k v_k,$$

where $v_0, \dots, v_k$ are vectors which span the kernel of T_A, $c_1, \dots, c_k$ are free scalars, v_0 is a particular solution to $Ax = b$, and $c_1 v_1 + \cdots + c_k v_k$ is the general solution to the homogeneous equation $Ax = 0$ (the kernel of T_A).

Suppose that there exists an invertible matrix E such that the matrix EA has a particularly nice form. Then $Ax = b \Rightarrow EAx = Eb$; since E is invertible, we have $EAx = Eb \Rightarrow E^{-1}EAx = E^{-1}Eb \Rightarrow Ax = b$. Thus the solution set of $Ax = b$ is exactly the solution set of $EAx = Eb$, so is suffices to find the solution set of $EAx = Eb$.

The nice form we refer to here is known as reduced row echelon form.

11. Reduced Row Echelon Form

A matrix is said to be in *row echelon form* if

i. All zero rows lie below nonzero rows;
ii. The first nonzero entry in any row appears in a column to the right of the first nonzero entry in any preceding row.

The first nonzero entry in a row is called a *pivot*.

Given a matrix A, there is a sequence of row operations which brings A into row echelon form. The final product is not unique.

A matrix is said to be in *reduced row echelon form* if

i. It is in row echelon form;
ii. All the pivots equal 1;
iii. All nonpivot entries in a column containing a pivot are equal to 0.

Given a matrix A, there is a sequence of row operations which brings A into row echelon form. Although the sequence of row operations is not unique, the final product is unique.

Gaussian elimination is an algorithm for using elementary row operations to bring a matrix into reduced row echelon form. There are two stages: *forward elimination* brings the matrix into row echelon form, and *backward elimination* brings the row echelon matrix into reduced row echelon form.

Forward elimination:

(1) Start with the first column, and proceed through all columns in order.
(2) If the diagonal entry in the column is zero, permute with the first available lower row so that the diagonal entry is nonzero (use P).
(3) Eliminate all entries below this one in order (use E).

Note that forward elimination does not use D. Also note that this algorithm is so specific, the sequence of elementary matrices and the row echelon form obtained is unique.

Backward elimination:

(1) Make all pivots equal to one (use D).
(2) Starting from the right, working upward then leftward, make all entries above a pivot equal to zero (use E).

To solve a system of linear equations $Ax = b$, form the augmented matrix $[A \mid b]$ and work A and b simultaneously. Perform forward elimination and backward elimination on A, and then read off the solution. We describe this last step momentarily.

Once the matrix is in reduced row echelon form, it is easy to read off the general solution. We give an example, then list the exact steps to take.

Example 5. Consider the matrix equation

$$\begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Computing the matrix product on the left gives

$$\begin{bmatrix} x_1 + 2x_2 \\ x_3 \\ x_4 + 5x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

The solution set of this equation is a subset of $\mathbb{R}^6$, so we actually seek six dimensional vectors. Insert the free variables into the equation in an appropriate fashion to arrive at

$$\begin{bmatrix} x_1 + 2x_2 \\ x_2 \\ x_3 \\ x_4 + 5x_5 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ x_2 \\ 2 \\ 3 \\ x_5 \\ 4 \end{bmatrix}$$

By the definition of vector addition, this is the same as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 0 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Subtract the free columns from both sides and use the distributive property to obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \\ 0 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 0 \\ 0 \\ -5 \\ 1 \\ 0 \end{bmatrix}$$

12. Solution Method

Let $A = [a_{ij}]_{ij}$ be an $m \times n$ matrix, $x = [x_1, \ldots, x_n]^t$ an n-dimensional variable column vector, and $b = [b_1, \ldots, b_m]$ an m-dimensional constant column vector. Then the solution set of the matrix equation $Ax = b$ is the solution set of a system of linear equations.

Let O be the product of all the elementary matrices whose corresponding row operations bring the matrix A into row echelon form, that is, those used in forward elimination. Set $Q = OA$, where Q is in row echelon form. Let $c = Ob$. Then the solution set of $Ax = b$ is equal to the solution set of $OAx = Ob$, i.e., $Qx = c$.

At this point, we can tell if there is no solution: this happens when the a row of the nonaugmented matrix contains only zeros, but the corresponding entry of the augmentation column is nonzero. We can also tell if the solution is unique: this happens when the number of nonzero rows equals the number of columns.

Let U be the product of all the elementary matrices whose corresponding row operations bring the matrix A into reduced row echelon form; that is, $R = UA$, where R is in reduced row echelon form. Let $d = Ub$. Then the solution set of $Ax = b$ is equal to the solution set of $UAx = Ub$, i.e., $Rx = d$. We describe how to read off the general solution from the matrix equation $Rx = d$.

We say that $R^{(j)}$ is a *basic column* if $R^{(j)}$ (or $Q^{(j)}$) contains a pivot; otherwise $R^{(j)}$ is a *free column*.

We say that x_j is a *basic variable* if $A^{(j)}$ contains a pivot; otherwise x_j is a *free variable*.

The general solution will be of the form

$$v_0 + c_1 v_1 + \cdots + c_k v_k,$$

where k is the number of free variables; we have $k = n - r$, where r is the number of nonzero rows.

The vector v_0 is the particular solution obtained by setting the free variables equal to 0 and solving for the basic variables.

The vectors v_i are found by replacing d by the zero vector, setting the i^{th} free variable equal to 1 and the other free variables equal to 0, and solving for the basic variables.

We can read off the general solution from the reduced matrix as follows:

(1) eliminate any zero rows at the bottom of the reduced matrix;
(2) insert a zero row at row i for every free variable x_i;
(3) multiply each free column by -1;
(4) add e_i to each free column for every free variable x_i;
(5) the particular solution is now the augmentation column;
(6) the homogeneous solution is now the span of the adjusted free columns.

13. Geometric Interpretation of Systems of Linear Equations

We have two geometric interpretations for solving a system of linear equations: as the intersection of the loci of the equations, and as the preimage of a linear transformation. How do these geometric interpretations correspond?

Reconsider the system of linear equations

$$3x_1 - 4x_2 = 11;$$
$$x_1 + 2x_2 = 7.$$

There are two ways of viewing this problem geometrically:

We want to find a point (x_1, x_2) which satisfies both equations, that is, which lies on both lines. This is an AND condition, and AND corresponds to the set operation of intersection (just as OR corresponds to the set operation of union); so we intersect the lines (which are, after all, subsets of $\mathbb{R}^2$) and find that the only point of intersection is $(11, 7)$.

The second geometric interpretation comes from putting the coefficients on the left hand side of the system of equations into a matrix A, the indeterminates into a column vector x and the values on the left hand side into a column vector b:

$$A = \begin{bmatrix} 3 & -4 \\ 1 & 2 \end{bmatrix}; \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad \text{and } b = \begin{bmatrix} 11 \\ 7 \end{bmatrix}.$$

We see that solving the system of equations is equivalent to solving the matrix equation

$$Ax = b.$$

But A corresponds to a linear transformation T_A; thus we seek the preimage of b under the linear transformation T_A.

How do these two geometric interpretations coincide?

14. Component Functions

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. For each $i \in \{1, \dots, m\}$, define a function

$$f_i : \mathbb{R}^n \to \mathbb{R} \quad \text{by } f_i(v) = \operatorname{proj}_{e_i} f(v);$$

this is called the i^{th} *component function* of f.

Example 6. Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f(t) = (\cos t, \sin t)$. Then $f_1 = \cos$ and $f_2 = \sin$. Note that the image of the function f is a circle in $\mathbb{R}^2$.

As this example shows, we may turn our definition around; that is, given m functions $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$, we construct a function $f : \mathbb{R}^n \to \mathbb{R}^m$ by defining $f(v) = (f_1(v), \dots, f_m(v))$.

Now let $f_1 : \mathbb{R}^2 \to \mathbb{R}$ be given by $f_1(x, y) = 3x - 4y$ and let $f_2 : \mathbb{R}^2 \to \mathbb{R}$ be given by $f_2(x, y) = x + 2y$. The line in $\mathbb{R}^2$ which is the locus the equation $3x_1 - 4x_2 = 11$ is the preimage of 11 under the function f_1; the second line is the preimage of 7 under f_2. A solution (x_1, x_2) for the system of equations is an element of the set $f_1^{-1}(11) \cap f_2^{-1}(7)$.

Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f(x) = (f_1(x), f_2(x))$; that is, $f(x_1, x_2) = (3x_1 - 4x_2, x_1 + 2x_2)$. Then the solution to the system of linear equations we started out with is the preimage of the point $(11, 7)$ under this new function; that is, we wish to find v such that $f(v) = (11, 7)$, which is the same as saying that we wish to discover the set $f^{-1}(11, 7) = f_1^{-1}(11) \cap f_2^{-1}(7)$.

By a previous proposition, we see that the function f is a linear transformation; let us relabel it by T.

Solving the system of equations is equivalent to finding the preimage of the point $(11, 7)$ under the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(x_1, x_2) = (3x_1 - 4x_2, x_1 + 2x_2)$. What is the effect of T on the standard basis, and what is the matrix associated to T?

We have $T(e_1) = T(1, 0) = (3, 1)$ and $T(e_2) = T(0, 1) = (-4, 2)$. Thus the matrix which corresponds to T is

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

and finding the preimage of $(11, 7)$ under T is equivalent to solving the matrix equation

$$Ax = b, \quad \text{where } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } b = \begin{bmatrix} 11 \\ 7 \end{bmatrix}.$$

In general, we have m equations in n unknowns. We obtain an $m \times n$ matrix A of coefficients, an $n \times 1$ column vector b of indeterminates, and an $m \times 1$ column vector of values. Solving the system is equivalent to solving the matrix equation $Ax = b$. The associated transformation $T = T_A$ is obtained by creating m linear functions $T_i : \mathbb{R}^n \to \mathbb{R}$ given by the left hand sides of our equations; these become the component functions of $T : \mathbb{R}^n \to \mathbb{R}^m$. The preimage of each T_i at b_i is a hyperplane in $\mathbb{R}^n$. The solution set is the intersection of the hyperplanes, which is the same as the preimage of the point b under the linear transformation T.

We may also view this as follows. Let $f : A \to B$ be any function, and let $C, D \subset B$. Then $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.

Let H_i be the hyperplane in $\mathbb{R}^m$ (the range of T) given by

$$H_i = \{(y_1, \ldots, y_m) \in \mathbb{R}^m \mid y_i = b_i\}.$$

Then

$$\{b\} = \cap_{i=1}^m H_i.$$

Let L_i be the hyperplane in R^n (the domain of T) which is the locus of the equation

$$a_{i1}x_i + \cdots + a_{in}x_n = b_i.$$

Then if X is the solution set to our system of linear equations, we have

$$X = \cap_{i=1}^m L_i.$$

But $L_i = T^{-1}(H_i)$, and

$$X = \cap_{i=1}^m T^{-1}(H_i) = T^{-1}(\cap_{i=1}^m H_i) = T^{-1}(b).$$

15. Geometric Interpretation of the Solution Process

We have the matrix equation $Ax = b$, where A is an $m \times n$ matrix. We know that A corresponds to a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. The columns of A are the destinations of the standard basis vectors in $\mathbb{R}^n$ under the transformation T. We ask if b is a linear combination of these destinations, in which case there is a solution to the equation.

When row reducing the augmented matrix $[A \mid b]$, we are in theory multiplying both sides of the equation $Ax = b$ by elementary invertible $m \times m$ matrices. Each such multiplication corresponds to an invertible linear transformation of $\mathbb{R}^m$, which is the range space of the linear transformation T. What in fact we are doing is transmuting $\mathbb{R}^m$ so that the *labeling* of the destinations of the standard basis vectors is more to our liking; in the process, b is also moved to a new location. That is, we are relabeling the points in $\mathbb{R}^m$ so that we can see more clearly the manner in which b is a linear combination of the destinations of the standard basis vectors.

Department of Mathematics, University of California, Irvine
E-mail address: pbailey@math.uci.edu