Matrix Notation

In order to discuss the problems of linear algebra it will be convenient to use vector/matrix notation.

Vectors

The term "vector" will mean an array of numbers arranged in a column and denoted by $\vec{x}$; the notation $\vec{x}^\top$ will denote the same array of numbers arranged in a row. That is,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \vec{x}^\top = [x_1, \ldots, x_n]$$

The purpose of distinguishing between column and row vectors will become apparent when we discuss matrix multiplication.

Matrices

The term, "$m \times n$ matrix" refers to a rectangular array of numbers consisting of m rows and n columns. For example,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \cdots & \vdots \\ b_{n1} & \cdots & b_{np} \end{bmatrix}$$

denote, respectively, an m by n matrix and an n by p matrix. Here $a_{ij} =$ the entry in the i-th row and j-th column of the matrix A. Note that a column vector is an n by 1 matrix and a row vector is a 1 by n matrix.

Scalars

The term "scalar" refers to ordinary numbers, real or complex, not arranged in any sort of array and subject to the usual rules of arithmetic.

Products

It is useful to define various products involving scalars, vectors and matrices.

Scalar Multiplication

The operation of multiplying a scalar, α, times a vector $\vec{x}$ or a matrix A is called *scalar multiplication* and is defined as follows:

$$\alpha\,\vec{x}^{\top} = [\alpha x_1, \ldots, \alpha x_n] \qquad \alpha A = \begin{bmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \cdots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{bmatrix}$$

Vector times Vector (**Inner Product**)

A scalar valued product of two vectors of length n is referred to variously as the "dot product", "inner product" or "scalar product" of two vectors. It is defined as follows:

$$\vec{x} \bullet \vec{y} = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

Note that this product is only defined between two vector of equal length. Other notations for this product include $\vec{x}^{\top}\,\vec{y}$ and $(\vec{x},\vec{y})$. This product satisfies:

i) $(\vec{x},\vec{y}) = (\vec{y},\vec{x})$
ii) $\alpha(\vec{x}+\vec{y}) = \alpha\vec{x}+\alpha\vec{y}$
iii) $(\vec{x},\vec{x}) \geq 0$ and $(\vec{x},\vec{x}) = 0$ if and only if $\vec{x}^{\top} = [0,\cdots,0]$

Vectors $\vec{x}$ and $\vec{y}$ which satisfy $\vec{x} \bullet \vec{y} = 0$ are said to be *orthogonal*. We will often denote this by writing $\vec{x} \perp \vec{y}$. In two and three dimensions, vectors which are orthogonal are in fact perpendicular in the geomentric sense. In higher dimensions there is no such interpretation other than to say that orthogonal vectors are perpendicular in a generalized sense.

Matrix times Vector

The product of an m by n matrix, $A,$ times an n –vector $\vec{x}$ is defined as follows

$$A\vec{x} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix}$$

This product, $A\vec{x},$ can only be formed if the number of columns of A is the same as the number of entries in $\vec{x}.$ In that case, the result is a vector with m entries, where m is the number of rows of A. Note that if we denote the rows of A by $R_1^{\top} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \end{bmatrix}, \ldots, R_m^{\top} = \begin{bmatrix} a_{m1} & \cdots & a_{mn} \end{bmatrix}$ then

$$A\vec{x} = \begin{bmatrix} R_1^{\top}\vec{x} \\ \vdots \\ R_m^{\top}\vec{x} \end{bmatrix}$$

Similarly, if we denote the columns of A by $C_1, \cdots, C_n$ then

$$A\vec{x} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 \\ \vdots \\ a_{m1}x_1 \end{bmatrix} + \cdots + \begin{bmatrix} a_{1n}x_n \\ \vdots \\ a_{mn}x_n \end{bmatrix}$$

$$= x_1\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n\begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = x_1C_1 + \cdots + x_nC_n$$

Matrix times Matrix

For any m by n matrix A and any n by p matrix $B,$ the product AB is defined to be an m by p matrix G whose entries G_{ij} are given by

$$G_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = R_i^{\top}C_j \qquad \text{for} \qquad 1 \le i \le m, \qquad 1 \le j \le p$$

i.e., $$AB = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \cdots & \vdots \\ b_{n1} & \cdots & b_{np} \end{bmatrix} = \begin{bmatrix} R_1^{\top}C_1 & \cdots & R_1^{\top}C_p \\ \vdots & \cdots & \vdots \\ R_m^{\top}C_1 & \cdots & R_m^{\top}C_p \end{bmatrix}$$

where $R_i^{\top}\ \ 1 \le i \le m$ denote the rows of A and $C_j\ \ 1 \le j \le p$ denote the columns of B

If A is m by n and B is n by $m,$ then the products AB and BA are both defined but are not equal. In fact, AB is m by m while BA is n by $n.$ Even in the case $m = n$ AB is not necessarily equal to $BA.$ The matrix product is not commutative.

Identity Matrix

Let I denote the n by n matrix whose entries I_{jk} equal 1 if $j = k$ and equal 0 otherwise. Then I is called the n by n *identity matrix* since for any n by n matrix $A,$ we have $IA = AI = A.$ Thus I plays the role of the identity in matrix multiplication.

Matrix Inverse

For n by n matrices A and $B,$ both the products AB and BA are defined and produce n by n matrices as a result of the multiplication. In the special case that $AB = BA = I,$ we say that B is the *matrix inverse* for A and write $B = A^{-1}.$ Note that only a square matrix can have an inverse but not every square matrix has an inverse. A matrix which has an inverse is said to be invertible or non-singular while a matrix with no inverse is said to be singular or non-invertible.

Matrix Identities

If A is an m by n matrix, then $A^\top$ denotes the n by m matrix whose rows are the columns of A; i.e.,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \qquad A^\top = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \cdots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix}$$

For vectors $\vec{x}$ and $\vec{y}$, and matrices A and B, we have the following identities

$$\vec{x} \bullet \vec{y} = \vec{x}^\top \vec{y} = y^\top \vec{x}$$

$$(A^\top)^\top = A$$

$$(A\vec{x})^\top = \vec{x}^\top A^\top \qquad \text{and} \qquad (AB)^\top = B^\top A^\top$$

$$(A\vec{x}) \bullet \vec{y} = (A\vec{x})^\top \vec{y} = \vec{x}^\top A^\top \vec{y} = \vec{x}^\top (A^\top \vec{y}) = \vec{x} \bullet A^\top \vec{y}$$

Matrix Notation for Systems of LInear Equations

Consider the system of m linear algebraic equations in the n unknowns $x_1, \ldots, x_n$

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$\vdots$$

$$a_{m1}x_1 + a_{12}x_2 + \cdots + a_{mn}x_n = b_m$$

This can be written in matrix notation as, $A\vec{x} = \vec{b}$. The use of matrix notation is of considerable value in studying problems of this type.

The First Problem of Linear Algebra

Let A denote an n by n matrix, and let $\vec{b}$ denote an arbitrary vector in R^n. The first fundamental problem of linear algebra is the problem of finding an $\vec{x}$ in R^n such that $A\vec{x} = \vec{b}$. That is, $\vec{x}$ solves a system of n linear algebraic equations in n unknowns. There are two things we want to know about this problem: first, does it have a solution, and second, is the solution unique?

In order to answer these questions, it will be helpful to define two ***subspaces*** associated with the matrix A.

N_A = the ***null space*** of A = all vectors $\vec{x} \in R^n$ satisfying $A\vec{x} = \vec{0}$.

R_A = the ***range*** of A = all vectors $\vec{b} \in R^n$ satisfying $A\vec{x} = \vec{b}$. for some $\vec{x} \in R^n$.

Note that the null space always contains the zero vector (at least) but there may be other vectors in the null space as well. Similarly the range of A also contains the zero vector (since $A\vec{0} = \vec{0}$) but may contain vectors besides the zero vector. Note that the system $A\vec{x} = \vec{b}$ has a solution if and only if $\vec{b}$ belongs to the range of A (this is just the definition of the range, it is the set of $\vec{b}'s$ for which the system has a solution). The system has at most one solution (i.e., any solution is unique) if and only if the null space of A contains only the zero vector. To see this, suppose there are two solutions for the system, e.g., $A\vec{x} = \vec{b}$ and $A\vec{z} = \vec{b}$. Then

$$A\vec{x} - A\vec{z} = \vec{b} - \vec{b} = \vec{0}.$$

That is, $$A\vec{x} - A\vec{z} = A(\vec{x} - \vec{z}) = \vec{0}, \qquad \text{or} \qquad (\vec{x} - \vec{z}) \in N_A$$

But if the null space of A contains only the zero vector, then $\vec{x} - \vec{z} = \vec{0}$, which is to say, $\vec{x} = \vec{z}$, and the two solutions are, in fact, equal (so there is really only one solution).

In order to determine whether N_A and R_A contain vectors besides the zero vector, we will need to define the notion of ***dimension*** for these subspaces (the precise definition for the term subspace has not yet been given but will be given later). In order to define dimension, we will first have to define the term ***linearly independent***.

a collection of vectors $\left\{\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_N\right\}$ is said to be ***linearly independent***.
if the following two statements are equivalent:

$$1) \quad C_1\vec{X}_1 + C_2\vec{X}_2 + \ldots + C_N\vec{X}_N = \vec{0} \qquad\qquad 2) \; C_1 = C_2 = \cdots = C_N = 0$$

A set of vectors is linearly independent if none of them can be written as a linear combination of the others. For example, the vectors

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \vec{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

are easily seen to be a linearly independent set in R^4. These vectors are also mutually orthogonal since $\vec{e}_j \bullet \vec{e}_k = 0$ for $j \neq k$ from which one gets the impression that mutual orthogonality is in some sense "maximal linear independence".

Now we can define the ***dimension*** of a subspace M to be the maximum number of linearly independent vectors in the subspace. It is clear that the dimension of a subspace of vectors from R^n cannot exceed n. With regard to the questions of existence and uniqueness of solutions for $A\vec{x} = \vec{b}$, the range of A is all of R^n (so that the system is solvable for any choice of $\vec{b}$) if the dimension of R_A equals $n,$ and N_A contains only the zero vector (so the solution is unique, if it exists) if the dimension of N_A equals zero. That is,

$A\vec{x} = \vec{b}$ has at least one solution for every $\vec{b} \in R^n$ if $\dim(R_A) = n$

$A\vec{x} = \vec{b}$ has at most one solution for every $\vec{b} \in R^n$ if $\dim(N_A) = 0$

In order to determine the dimensions of R_A and $N_A,$ we have to now define two more subspaces associated with A.

the ***row space*** of A $\quad RS[A] = span\left[\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_n\right]$

and the ***column space*** of A $\quad CS[A] = span\left[\vec{C}_1, \vec{C}_2, \ldots, \vec{C}_n\right]$

where $\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_n =$ the n rows of A $\quad$ *and* $\quad$ $\vec{C}_1, \vec{C}_2, \ldots, \vec{C}_n =$ the n columns of A

$span\left[\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n\right] =$ all possible linear combinations of $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$

Note the following facts about the row space and column space of $A.$ Since

$$A\vec{x} = \begin{bmatrix} \vec{R}_1 \bullet \vec{x} \\ \vec{R}_2 \bullet \vec{x} \\ \vdots \\ \vec{R}_n \bullet \vec{x} \end{bmatrix} \quad \text{and} \quad A\vec{x} = x_1\vec{C}_1 + x_2\vec{C}_2 + \cdots + x_n\vec{C}_n$$

it follows that:

$\vec{x} \in N_A$ if and only if $\vec{R}_1 \bullet \vec{x} = \vec{R}_2 \bullet \vec{x} = \cdots = \vec{R}_n \bullet \vec{x} = 0,$
i.e., if and only if $\vec{x}$ is orthogonal to all the rows of A

and $A\vec{x} = \vec{b}$ if and only if $\vec{b} = x_1\vec{C}_1 + x_2\vec{C}_2 + \cdots + x_n\vec{C}_n$
i.e., if and only if $\vec{b} \in CS[A] = span\left[\vec{C}_1, \vec{C}_2, \ldots, \vec{C}_n\right]$

These two observations can be stated more concisely as:

$N_A =$ the orthogonal complement of $RS[A] =: (RS[A])^{\perp}$
$R_A = CS[A]$

Here we define, $M^{\perp},$ the orthogonal complement of a subspace M to be the set of all vectors $\vec{z}$ which are orthogonal to every vector in $M;$ i.e., $M^{\perp} = \left\{\vec{z} : \vec{z} \bullet \vec{x} = 0 \text{ for every } \vec{x} \text{ in } M\right\}$. It is a fundamental result from linear algebra (a result not proved here) that the sum of the dimensions of a subspace and its orthogonal complement equals the dimension of the surrounding space. That is, if M is a subspace in $\mathbb{R}^n,$ then $M^{\perp}$ is also a subspace in $\mathbb{R}^n$ and the sum of the dimesions of M and $M^{\perp}$ equals n. Since $(RS[A])^{\perp} = N_A$, it follows that

$$\dim(N_A) + \dim(RS[A]) = n$$

The second observation, $R_A = CS[A]$, implies that

$$\dim(R_A) = \dim(CS[A])$$

A second fundamental fact (proved in any linear algebra course but not proved here) asserts that

$$\dim(RS[A]) = \dim(CS[A]) = r$$

Combining these facts leads to the following fundamental result for solving systems of linear equations

$$\dim(R_A) = n - \dim(N_A), \qquad \text{or} \qquad \dim(N_A) = n - r$$

We will now show how to compute the number r. In order to do so, we introduce the notion of the ***rank*** of the matrix A. Beginning with the matrix A, we perform ***row operations*** on A to reduce it to ***echelon form***; e.g.;

$$A = \begin{bmatrix} 2 & -3 & -3 \\ -3 & -2 & 0 \\ 3 & -1 & 2 \end{bmatrix}, \quad \Rightarrow \text{ row operations } \Rightarrow \quad B = \underbrace{\begin{bmatrix} 2 & -3 & -3 \\ 0 & \frac{7}{2} & \frac{13}{2} \\ 0 & 0 & \frac{53}{7} \end{bmatrix}}_{\text{echelon form}}$$

The ***row operations*** consist of:

1) multiply any row by a nonzero scalar
2) multiply any row by a nonzero scalar and add it to another row
3) interchange two rows

Two matrices are said to be ***row equivalent*** if one is obtained from the other by a sequence of row operations. For example, the matrices A and B above are row equivalent. The matrix B above is in ***echelon form***; this means that in each row, the first nonzero entry lies to the right of the first nonzero entry in the row above and in each column, all entries below the diagonal entry are zeroes Now we can define the ***rank*** of A to be the number of nontrivial rows (not all zero entries) in a row equivalent matrix that is in echelon form. Thus the rank of the matrix A in the example above is 3 since the row equivalent matrix B has 3 nontrivial rows. In the following example,

$$A = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 3 & -1 & 2 \end{bmatrix}, , \quad \Rightarrow \text{row operations} \Rightarrow: B = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \text{echelon form}$$

we see that A has rank equal to 2. Evidently, the rank of an n by n matrix can equal any number between zero (all entries of A would have to be zeroes) and n (the rows of A are a linearly independent set of vectors).

The previously stated (but not proved) results, .

$$\dim[RS[A]] + \dim[N_A] = n$$

$$\dim[RS[A]] = \dim[CS[A]] = \dim[R_A]$$

now can be combined to provide the assertions:

1. $r = rank\, A = \dim[R_A]$

2. $\dim[N_A] = n - r$

It is evident from 1) that if the matrix A has rank equal to n, then $\dim[R_A] = n$ and $R_A = R^n$. In this case $A\vec{x} = \vec{b}$ is solvable for any $\vec{b} \in R^n$.

It is also evident from 2) that if $rank[A] = n$, then $\dim[N_A] = 0$, which is to say N_A contains only the zero vector and the solution to $A\vec{x} = \vec{b}$ is unique.

An alternative means for determining whether the matrix A has rank equal to n is to compute the determinant of A. Computing the determinant of an n by n matrix A when n is large is impractical so this approach is more theoretical than practical. At any rate, it can be shown that the determinant of A is different from zero if and only if the rank of A equals n. Then, at least in cases of small n, the determinant computation can be used to decide the rank of A.

These observations lead to the fundamental result regarding the first problem of linear algebra:

Theorem- The following are equivalent statements

1) $A\vec{x} = \vec{b}$ is uniquely solvable for every $\vec{b} \in R^n$
2) $rank[A] = n$
3) $\det(A) \neq 0$