Chapter 9

Monday, July 25, 2011

9.1 Linear Algebra: The Idea of Linearity

Linear Algebra is the study of real n-dimensional spaces and maps between them that preserve the linear structure of $\mathbb{R}^n$. This linearity actually makes many computations in Linear Algebra very manageable. Furthermore, these linear maps have far-reaching applications in many fields of Mathematics.

9.2 Understanding $\mathbb{R}^n$

9.2.1 Vectors

We will see that $\mathbb{R}^n$ will essentially be n copies of $\mathbb{R}$. An *n-vector* is an n-tuple of real numbers

$$\mathbf{x} = (x_1, x_2, x_3, \ldots, x_n)$$

with $x_i \in \mathbb{R}$. Thus, any *ordered* collection of n real numbers will specify an n-vector. The space of all n-vectors is known as $\mathbb{R}^n$:

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \,|x_i \in \mathbb{R}\}.$$

Notice that it is important they are ordered as we need to know which real number to put in which slot. As a *vector space*, $\mathbb{R}^n$ also has some additional structure. There is a notion of *scalar multiplication*. For any $\alpha \in \mathbb{R}$ (called a *scalar*), we can multiply α with any n-vector by

$$\alpha \cdot \mathbf{x} = \alpha \cdot (x_1, x_2, \ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)\,.$$

Intuitively, this stretches our vector $\mathbf{x}$ by a factor of α. We also have a notion of *vector addition*; given any two n-vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, we can form their sum:

$$\mathbf{x} + \mathbf{y} = (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) =$$

$$(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)\,.$$

Of course, it only makes sense to add two vectors of the same dimension. Notice that these two operations of scalar multiplication and vector addition behave nicely with respect to each other; we have the following distributive property:

$$\alpha \cdot (\mathbf{x} + \mathbf{y}) = \alpha \cdot \mathbf{x} + \alpha \cdot \mathbf{y}.$$

The notion of the vector space $\mathbb{R}^n$ is a generalization of the additive and multiplicative structure of $\mathbb{R}$. Clearly, $\mathbb{R}^1 = \mathbb{R}$ as vector spaces since scalar multiplication here is just regular multiplication and vector addition is regular addition.

9.2.2 Visualizing $\mathbb{R}^n$

We will be focusing on $\mathbb{R}^1, \mathbb{R}^2$, and $\mathbb{R}^3$ since they are the real vector spaces that we can visualize geometrically. We already have a good notion of $\mathbb{R}$ as a flat infinitely long line. This one-dimensional object is the object of the study of single-variable calculus. Now, if we take one horizontal copy of $\mathbb{R}$ and one vertical copy of $\mathbb{R}$ and intersect them at their zero points, this gives us a coordinate description of $\mathbb{R}^2$. Given any 2-vector (x_1, x_2), we find this on $\mathbb{R}^2$ by finding x_1 on the horizontal copy of $\mathbb{R}$ and finding x_2 on the vertical copy of $\mathbb{R}$. Then, we find the vector (x_1, x_2) by going over to x_1 on the vertical $\mathbb{R}$ and up to x_2 on $\mathbb{R}^2$. Thus, $\mathbb{R}^2$ is geometrically a flat plane and any point on this plane can be represented by a 2-vector.

Visualizing $\mathbb{R}^3$ is just a generalization of the above geometric construction of $\mathbb{R}^2$. Since our universe is (at least on a visual level) 3-dimensional, many physical applications require a concrete understanding of $\mathbb{R}^3$. To construct this, we consider three copies of $\mathbb{R}$, but arrange them so that the first and second copies are flat (like on a table top) and meet perpendicularly and the third one is perpendicular to both copies of $\mathbb{R}$ (coming up out of the table top). Again, any 3-vector (x_1, x_2, x_3), we find it on $\mathbb{R}^3$ by finding x_1 on the first copy of $\mathbb{R}$, go over to x_2 on the second copy of $\mathbb{R}$, and then go up to x_3 on the last copy. Once again, any 3-vector can be represented uniquely in this manner.

It is also helpful to see how the algebraic operations of scalar multiplication and vector addition are reflected geometrically. We will focus on $\mathbb{R}^3$. Recall that scalar multiplication of a scalar α with a 3-vector $\mathbf{x} = (x_1, x_2, x_3)$ is given by

$$\alpha \cdot \mathbf{x} = \alpha \cdot (x_1, x_2, x_3) = (\alpha x_1, \alpha x_2, \alpha x_3)\,.$$

Thus, every single entry x_i is scaled by α. The overall effect on the vector $\mathbf{x}$ is that it is stretched to be α-times as long in the direction of $\mathbf{x}$. Note that if $\alpha = 0$, then $\alpha \cdot \mathbf{x}$ will give the zero vector $(0, 0, 0)$; further, if $\alpha < 0$, then $\mathbf{x}$ will be stretched by $|\alpha|$ in the opposite direction (i.e., in the direction of $-\mathbf{x}$). If we consider two 3-vectors $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$, we recall how to add these vectors:

$$\mathbf{x} + \mathbf{y} = (x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3).$$

Now, we consider a parallelogram with one vertex at $(0, 0, 0)$, one at (x_1, x_2, x_3) and the third (y_1, y_2, y_3). The only way to complete this parallelogram is by a fourth vector; this vector is realized geometrically as the sum $\mathbf{x} + \mathbf{y}$.

9.3 The World of Matrices

Notice that a vector is essentially a convenient way to keep track of n independent real variables. In a similar way, we can define a matrix as an arrangement of real numbers in an array. Formally, an $n \times m$-matrix is an array of n rows and m columns, with each of the nm slots being filled by one real number. The number that we put in the i-th row and the j-th column will be denoted a_{ij}.

Many times, we will denoted an $n \times m$-matrix by $A = [a_{ij}]$ or can be written in its rows and columns:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}.$$

We can see an n-vector as an $n \times 1$-matrix, which is also called a column vector. Thus, the vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ can be written as

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

As we placed scalar multiplication and vector addition on $\mathbb{R}^n$, we will place these structures as well as a matrix multiplication structure on the space of matrices. First, given any $n \times m$ matrix A and a scalar α, the scalar product $\alpha \cdot A$ scales each entry of A by α:

$$\alpha \cdot A = \alpha \cdot \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} = \begin{bmatrix} \alpha \cdot a_{11} & \alpha \cdot a_{12} & \dots & \alpha \cdot a_{1m} \\ \alpha \cdot a_{21} & \alpha \cdot a_{22} & \dots & \alpha \cdot a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha \cdot a_{n1} & \alpha \cdot a_{n2} & \dots & \alpha \cdot a_{nm} \end{bmatrix}.$$

Given any two $n \times m$-matrices A and B, we can form the matrix addition of A and B by adding the matrices term-wise:

$$A + B = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1m} + b_{1m} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2m} + b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \dots & a_{nm} + b_{nm} \end{bmatrix}.$$

Notice that addition only makes sense when the matrices have the same m and n.

Now, we will be able to place a multiplication structure on matrices. If we are given an $n \times m$ matrix A and an $m \times p$ matrix B, the matrix product of A and B will be an $n \times p$-matrix C. We obtain the ij entry of the product matrix by

$$c_{ij} = \sum_{k=1}^{m} a_{ik} \cdot b_{kj}.$$

Notice that there are exactly m terms in this sum since the i-th row of A has m entries and the j-th column of B also has m entries. Essentially, this sum formula means that to find the ij-th entry of the product $A \cdot B$, we take the i-th row of A and the j-th column of B and add the product of the first terms and second terms and so on, until we reach the m-th (last) terms. So, multiplication by matrices can be a bit tedious to compute, but for small matrices (e.g., when $m, n \leq 3$), this computation is manageable.

9.3.1 Computation Examples

Let's take the 2×3-matrix

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

and the 3×2-matrix

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

Consider the following scalar multiplications:

$$-1 \cdot A = -1 \cdot \begin{bmatrix} 1 & 2 & 0 \\ -2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} -1 & -2 & 0 \\ 2 & -3 & 1 \end{bmatrix};$$

$$3 \cdot B = 3 \cdot \begin{bmatrix} 0 & 12 \\ 15 & -3 \\ 6 & 6 \end{bmatrix}.$$

Now, given these A and B that we have, the multiplication $A \cdot B$ makes sense because A is an 2×3 matrix and B is an 3×2; furthermore, this product matrix $A \cdot B$ will be a 2×2 matrix. If we look at our matrices,

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

we see that to compute the $1,1$ entry c_{11}. According to our formula, this will be

$$c_{11} = \sum_{k=1}^{3} a_{1k} b_{k1} = a_{11} b_{11} + a_{12} b_{21} + a_{13} b_{31} = 1 \cdot 0 + 2 \cdot 5 + 0 \cdot 2 = 10.$$

Notice that this is just multiplying the first row and the first column termwise and them adding up. Doing the other computations, we have

$$c_{12} = \sum_{k=1}^{3} a_{1k} b_{k2} = 1 \cdot 4 + 2 \cdot -1 + 0 \cdot 2 = 2;$$

$$c_{21} = \sum_{k=1}^{3} a_{2k} b_{k1} = -2 \cdot 0 + 3 \cdot 5 + -1 \cdot 2 = 13;$$

$$c_{22} = \sum_{k=1}^{3} a_{2k} b_{k2} = -2 \cdot 4 + 3 \cdot -1 + -1 \cdot 2 = -13.$$

Thus, we have that

$$A \cdot B = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 3 & -1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 4 \\ 5 & -1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 13 & -13 \end{bmatrix}.$$

Notice that we can also form the product $B \cdot A$ since B is a 3×2 matrix and A is a 2×3 matrix; thus, the product is a 3×3 matrix. Doing the matrix multiplication as above, we have that

$$B \cdot A = \begin{bmatrix} 0 & 4 \\ 5 & -1 \\ 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 0 \\ -2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} -8 & 12 & -4 \\ 3 & 7 & 1 \\ -2 & 10 & -2 \end{bmatrix}.$$

Notice, of course, that $A \cdot B \neq B \cdot A$ since they are not even of the same dimensions. Even if they were of the same dimension, it is (generally) true that $A \cdot B \neq B \cdot A$.

Notice that the product structure on matrices and the additivity structure on matrices behave distributively. Given an $n \times m$-matrix A and two $m \times p$-matrix B and C, then

$$A \cdot (B + C) = A \cdot B + A \cdot C.$$

Example - The Identity Matrix. The most important and basic matrix is an $n \times n$-matrix I_n, which has 1's along the diagonal and 0's everywhere else. Thus, it is given by

$$I_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

The most important property of the identity matrix is that it leaves any matrix with which it is multiplies unchanged. Thus, if A is any $n \times p$ matrix and B is any $p \times n$-matrix, then I_n is a multiplicative identity:

$$I_n \cdot A = A$$

$$B \cdot I_n = B.$$

9.3.2 Matrices as Maps on Vector Spaces

Recall that we can view an n-vector $\mathbf{x} = (x_1, x_2, \dots, x_m)$ as an $n \times 1$-matrix

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

If we take any $n \times m$-matrix A and form the multiplication $A \cdot \mathbf{x}$ will produce an $n \times 1$-matrix (i.e., an n-vector). Thus, we can see left multiplication L_A by an $n \times m$-matrix A is a function from $\mathbb{R}^m$ to $\mathbb{R}^n$. Formally, this map is given by

$$L_A : \mathbb{R}^m \to \mathbb{R}^n$$

$$\mathbf{x} \mapsto A \cdot \mathbf{x}.$$

One of the major reasons that these maps given by left-multiplication by A is that it preserves the *linear structure* of $\mathbb{R}^n$; we say that such a map is a *linear transformation.* Precisely, if A is an $n \times m$-matrix, then map $L_A : \mathbb{R}^m \to \mathbb{R}^n$ (given by left-multiplication of an m-vector by A) preserves the scalar structure:

$$L_A(\alpha \mathbf{x}) = A \cdot (\alpha \mathbf{x}) = \alpha A \cdot \mathbf{x} = \alpha L_A \mathbf{x}.$$

Further, this map L_A preserves the additive structure:

$$L_A(\mathbf{x} + \mathbf{y}) = A \cdot (\mathbf{x} + \mathbf{y}) = A \cdot \mathbf{x} + A \cdot \mathbf{y} = L_A \mathbf{x} + L_A \mathbf{y}.$$

Thus, this map L_A of a scaled vector is L_A of the vector scaled; also, the map L_A takes the sum of two vectors $\mathbf{x} + \mathbf{y}$ to the sum of $L_A \mathbf{x}$ and $L_A \mathbf{y}$.