INTRODUCING MATRICES

THE WHOLE PANTHEON

ABSTRACT. This is a collection of introductions to matrices taken from mathematics texts.

Each section is the introduction to matrices from a mathematics textbook. The goal in each case is both to tell you what a matrix is *and* to explain why you ought to care. As you read them, think about why the author made these choices; about what audience is being addressed; and about what you might do yourself.

A secondary purpose of this document is to show you how to do a few things in LaTeX. One important thing to notice is that you should let LaTeX do the numbering (of sections, equations, definitions, and so on); then when you insert additional equations, definitions, and so on, the numbering will be automatically corrected. This means also that when you refer to section 4 (in this case the selection from Susan Colley's book) you should refer to the internal label rather than saying "section 4." You can also refer to a particular equation, like (1.8) (van der Waerden's definition of matrix entries).

Date: September 2, 2008.

1. van der Waerden's *Algebra*, 1966, pp. 69–70

Let $\mathfrak{M}$ and $\mathfrak{N}$ be vector spaces. A *linear transformation* is a mapping $\mathbf{A}$ from $\mathfrak{M}$ to $\mathfrak{N}$ with the following two properties:

$$\mathbf{A}(\mathbf{x}+\mathbf{y}) = \mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y} \tag{1.1}$$

$$\mathbf{A}(\mathbf{x}c) = (\mathbf{A}\mathbf{x})c \tag{1.2}$$

From (1.1) it follows as usual that

$$\mathbf{A}(\mathbf{x}-\mathbf{y}) = \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{y} \tag{1.3}$$

$$\mathbf{A}(\mathbf{x}_1 + \cdots + \mathbf{x}_r) = \mathbf{A}\mathbf{x}_1 + \cdots + \mathbf{A}\mathbf{x}_r. \tag{1.4}$$

If $\mathfrak{M}$ has finite dimension m and if $\mathbf{p}_1, \dots, \mathbf{p}_m$ form a basis, then the effect of a linear transformation $\mathbf{A}$ on an arbitrary vector is completely determined by its effect on the basis vectors. Let

$$\mathbf{x} = \mathbf{p}_1 x^1 + \cdots \mathbf{p}_m x^m.$$

Then, from (1.4) and (1.2),

$$\mathbf{y} = \mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{p}_1)x^1 + \cdots + (\mathbf{A}\mathbf{p}_m)x^m. \tag{1.5}$$

If $\mathfrak{N}$ also has finite dimension n, then on the left and right in (1.5) we can express the vectors $\mathbf{y}$ and $\mathbf{A}\mathbf{p}_k$ in terms of the basis vectors $\mathbf{q}_1, \dots, \mathbf{q}_n$ of $\mathfrak{N}$:

$$\mathbf{y} = \sum \mathbf{q}_i y^i \tag{1.6}$$

$$\mathbf{A}\mathbf{p}_k = \sum \mathbf{q}_i a_k{}^i. \tag{1.7}$$

It follows from (1.5) on comparison of coefficients that

$$y^i = \sum a_k{}^i x^k. \tag{1.8}$$

The linear transformation $\mathbf{A}$ is thus determined by a *matrix* A, that is, by a rectangular array of mn elements $a_k{}^i$ of the skew field K:

$$A = \begin{pmatrix} a_1{}^1 & a_2{}^1 & \cdots & a_m{}^1 \\ \vdots & \vdots & & \vdots \\ a_1{}^n & a_2{}^n & \cdots & a_m{}^n \end{pmatrix}.$$

If the bases $\mathbf{p}_1, \dots, \mathbf{p}_m$ and $\mathbf{q}_1, \dots, \mathbf{q}_n$ are fixed, then each linear transformation $\mathbf{A}$ determines a unique matrix A and conversely. The first index i is the *row index* and the second index k is the *column index* of a matrix element $a_k{}^i$. The elements of the kth column are by (1.7) the coordinates of the vector $\mathbf{A}\mathbf{p}_k$.

2. Strang's *Introduction to Linear Algebra*, 2003, pp. 24–26

The three unknowns are x, y, and z. The linear equations $A\mathbf{x} = \mathbf{b}$ are

$$\begin{array}{rcrcrcl} x & + & 2y & + & 3z & = & 6 \\ 2x & + & 5y & + & 2z & = & 4 \\ 6x & - & 3y & + & z & = & 2 \end{array} \tag{2.1}$$

We look for numbers x, y, z that solve all three equations at once.

$$\vdots$$

We have three rows in the row picture and three columns in the column picture (plus the right side). The three rows and three columns contain nine numbers. *These nine numbers fill a* 3 *by* 3 *matrix.* The "coefficient matrix" has the rows and columns that have so far been kept separate:

$$\textbf{\textit{The coefficient matrix is}}\ A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix}.$$

The capital letter A stands for all nine coefficients (in this square array). The letter $\mathbf{b}$ denotes the column vector with components 6, 4, 2. The unknown $\mathbf{x}$ is also a column vector, with components x, y, z. (We use boldface because it is a vector, $\mathbf{x}$ because it is unknown.) By rows the equations were…and now by matrices they are (2.2). The shorthand is $A\mathbf{x} = \mathbf{b}$:

$$\textbf{\textit{Matrix equation}}\quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 2 \\ 6 & -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}. \tag{2.2}$$

We multiply the matrix A times the unknown vector $\mathbf{x}$ to get the right side $\mathbf{b}$.

Basic question: What does it mean to "multiply A times $\mathbf{x}$"? We can multiply by rows or by columns. Either way, $A\mathbf{x} = \mathbf{b}$ must be a correct representation of the three equations. You do the same nine multiplications either way.

3. Herstein's *Topics in Algebra*, 1964 pp. 228–229

Although we have been discussing linear transformations for some time, it has always been in a detached and impersonal way; to us a linear transformation has been a symbol (very often, T) which acts in a certain way on a vector space. When one gets right down to it, outside of the few concrete examples encountered in the problems, we have really never come face to face with specific linear transformations. At the same time it is clear that if one were to pursue the subject further there would often arise the need of making a thorough and detailed study of a given linear transformation. To mention one precise problem, presented with a linear transformation (and suppose, for the moment, that we have a means of recognizing it), how does one go about, in a "practical" and computable way, finding its characteristic roots?

What we seek first is a simple notation, or, perhaps more accurately, representation, for linear transformations. We shall accomplish this by use of a particular basis of the vector space and by use of the action of a linear transformation on his basis. Once this much is achieved, by means of the operations in $A(V)$ [the linear transformations on a vector space V] we can induce operations for the symbols created, making of them an algebra. This new object, infused with an algebraic life of its own, can be studied as a mathematical entity having an interest by itself. This study is what comprises the subject of *matrix theory.*

However, to ignore the sources of these matrices, that is to investigate the set of symbols independently of what they represent, can be costly, for we would be throwing away a great deal of useful information. Instead we shall always use the interplay between the abstract, $A(V)$, and the concrete, the matrix algebra, to obtain information one about the other.

Let V be an n-dimensional vector space over a field F and let $v_1, \ldots, v_n$ be a basis of V over F. If $T \in A(V)$ then T is determined on any vector as soon as we know its action on a basis of V. Since T maps V into V, v_1T, v_2T,$\ldots$,v_nT must all be in V. As elements of V each of these is realizable in a *unique* way as a linear combination of $v_1, \ldots, v_n$ over F. Thus:

$$
\begin{aligned}
v_1T &= \alpha_{11}v_1 + \alpha_{12}v_2 + \cdots + \alpha_{1n}v_n \\
v_2T &= \alpha_{21}v_1 + \alpha_{22}v_2 + \cdots + \alpha_{2n}v_n \\
v_iT &= \alpha_{i1}v_1 + \alpha_{i2}v_2 + \cdots + \alpha_{in}v_n \\
&\vdots \\
v_nT &= \alpha_{n1}v_1 + \alpha_{n2}v_2 + \cdots + \alpha_{nn}v_n,
\end{aligned}
$$

where each $\alpha_{ij} \in F$. This system of equations can be written more compactly as

$$v_iT = \sum_{j=1}^{n} \alpha_{ij}v_j, \text{ for } \; i = 1, 2, \ldots, n.$$

The ordered set of n^2 numbers α_{ij} in F completely describes T. They will serve as the means of representing T.

Definition 3.1. *Let V be an n-dimensional vector space over F and let $v_1, \ldots, v_n$ be a basis of V over F. If $T \in A(V)$ then the* matrix of T in the basis $v_1, \ldots, v_n$*, written $m(T)$, is*

$$m(T) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix}$$

where $v_iT = \sum_{j=1}^{n} \alpha_{ij}v_j$.

A matrix, then is an ordered, square array of elements of F, with, as yet, no further properties, which represents the effect of a linear transformation on a given basis.

4. Colley's *Vector Calculus*, 2006 pp. 29, 50

A **matrix** is a rectangular array of numbers. Examples of matrices are

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \begin{bmatrix} 1 & 3 \\ 2 & 7 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If a matrix has m rows and n columns, we call it "$m \times n$" (read "m by n"). Thus, the three matrices just mentioned are, respectively, 2×3, 3×2, and 4×4. To some extent, matrices behave algebraically like vectors. We discuss some elementary matrix algebra in §1.6. For now, we are mainly interested in the notion of a **determinant**, which is a real number associated to an $n \times n$ (square) matrix. (There is no such thing as the determinant of a nonsquare matrix.) In fact, for the purposes of understanding the cross product, we need only study 2×2 and 3×3 determinants.

Definition 4.1. *Let A by a 2×2 or 3×3 matrix. Then the **determinant** of A, denoted $\det A$ or $|A|$, is the real number computed from the individual entries of A as follows:*

- 2×2 *case*

  If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, *then* $|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$

- 3×3 *case*

  If $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, *then*

$$|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + ccdh - ceg - afh - bdi$$

$$= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

  in terms of 2×2 determinants.

⋮

We had a brief glance at matrices and determinants in §1.4 in connection with computation of cross products. Now it's time for another look.

A matrix is defined in §1.4 as a rectangular array of numbers. To extend the discussion, we need a good notation for matrices and their individual entries. We used the upper case Latin alphabet to denote entire matrices and will continue to do so. We shall also adopt the standard convention and two sets of indices (one set for rows, the other for columns) to identify matrix entries. Thus, the generaal $m \times n$ matrix can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = (\text{shorthand})(a_{ij}).$$

The first index will *always* represent the row position and the second index the column position.

5. Artin's *Algebra*, 1991, pp. 1-2

Matrices play a central role in this book. They form an important part of the theory, and many concrete examples are based on them. Therefore it is essential to develop facility in matrix manipulation. Since matrices pervade much of mathematics, the techniques needed here are sure to be useful elsewhere.

The concepts which require practice to handle are *matrix multiplication* and *determinants*.

Let m and n be positive integers. An $m \times n$ matrix is a collection of mn numbers arranged in a rectangular array:

(5.1)
$$\begin{array}{c} n \text{ columns} \\ m \text{ rows} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \end{array}$$

For example, $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 5 \end{bmatrix}$ is a 2×3 matrix.

The numbers in a matrix are called the *matrix entries* and are denoted by a_{ij}, where i, j are indices (integers) with $1 \le i \le m$ and $1 \le j \le n$. The index i is called the *row index*, and j is the *column index*. So a_{ij} is the entry which appears in the ith row and jth column of the matrix:

$$\begin{array}{c} \qquad j \\ i \begin{bmatrix} & & & \cdot & & & & & & \\ \cdot & \cdot & \cdot & a_{ij} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & \cdot & & & & & & \\ & & & \cdot & & & & & & \\ & & & \cdot & & & & & & \\ & & & \cdot & & & & & & \\ & & & \cdot & & & & & & \end{bmatrix} \end{array}$$

In the example above, $a_{11} = 2$, $a_{13} = 0$, and $a_{23} = 5$.

We usually introduce a symbol such as A to denote a matrix, or we may write it as (a_{ij}).

A $1 \times n$ matrix is called an n-dimensional *row vector*. We will drop the index i when $m = 1$ and write the row vector as

(5.2) $$A = [a_1 \cdots a_n], \quad \text{or as} \quad A = (a_1, \ldots, a_n).$$