Matrix Notation - A Quick Introduction

Andre C. Andersen

Originally published June 28, 2012

Abstract

Vector and matrix notation is a convenient and succinct way to represent dense data structures and complex operations. Its application is called matrix algebra, but it is best viewed as an extension of ordinary algebra rather than an alternative to it. Matrix notation is useful across linear algebra, multivariate analysis, economics, optimization, and machine learning because it lets us reason about structured collections of variables without drowning in indices. This short note gives a brief introduction to some of the most useful parts of the notation.

Data Structures

A vector $\boldsymbol{v} \in \mathbb{R}^n$ represents an ordered list of scalar variables $v_i \in \mathbb{R}$. The index $i \in \{1, \dots, n\}$ denotes the position of each variable within the vector. By default, vectors are written vertically:

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \tag{1}$$

A vector written this way is called a column vector. Orientation matters when performing operations. The transpose operation turns a column vector into a row vector:

$$\boldsymbol{v}^{\mathrm{T}} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \tag{2}$$

A compact way to represent a column vector is therefore

$$\boldsymbol{v} = [v_1 \dots v_n]^{\mathrm{T}} \tag{3}$$

A matrix $A \in \mathbb{R}^{m \times n}$ is a rectangular array of scalar entries $a_{ij} \in \mathbb{R}$. The first index i denotes the row and the second index j denotes the column. A matrix may be viewed as a collection of row vectors:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_m^{\mathrm{T}} \end{bmatrix}, \qquad \boldsymbol{a}_i^{\mathrm{T}} = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \tag{4}$$

Alternatively, a matrix may be viewed as a collection of column vectors:

$$D = \begin{bmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \ddots & \vdots \\ d_{m1} & \cdots & d_{mn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}_1 & \cdots & \boldsymbol{d}_n \end{bmatrix}, \qquad \boldsymbol{d}_j = \begin{bmatrix} d_{1j} \\ \vdots \\ d_{mj} \end{bmatrix} \tag{5}$$

Conversely, a vector may be interpreted as a matrix with only one column or one row, hence the terms column vector and row vector.

Arithmetic Operations

There are several ways to multiply vectors and matrices. The simplest is the dot product:

$$\boldsymbol{v} \cdot \boldsymbol{u} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = v_1 u_1 + \cdots + v_n u_n \tag{6}$$

The vectors must have the same dimension. Corresponding entries are multiplied and then summed.

General matrix multiplication extends this idea. Each row vector of $A \in \mathbb{R}^{m \times n}$ is dotted with each column vector of $B \in \mathbb{R}^{n \times k}$:

$$AB = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_1 & \cdots & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_k \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{b}_1 & \cdots & \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{b}_k \end{bmatrix} \tag{7}$$

The inner dimensions must agree. The resulting matrix AB has dimension $m \times k$, and each entry is

$$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{b}_j = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{bmatrix} = a_{i1}b_{1j} + \cdots + a_{in}b_{nj}. \tag{8}$$

A special case is matrix-vector multiplication:

$$A\boldsymbol{v} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{v} \\ \vdots \\ \boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{v} \end{bmatrix} = \begin{bmatrix} a_{11}v_1 + \cdots + a_{1n}v_n \\ \vdots \\ a_{m1}v_1 + \cdots + a_{mn}v_n \end{bmatrix} \tag{9}$$

In computer science and numerical work, the entrywise product (also called the Hadamard product) is also common. It multiplies corresponding entries of two objects with the same dimensions:

$$A \circ D = \begin{bmatrix} a_{11}d_{11} & \cdots & a_{1n}d_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}d_{m1} & \cdots & a_{mn}d_{mn} \end{bmatrix} \tag{10}$$

Multiplication by a scalar simply scales every entry:

$$c\boldsymbol{v} = c \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} cv_1 \\ \vdots \\ cv_m \end{bmatrix} \tag{11}$$

Addition and subtraction are also entrywise:

$$\boldsymbol{x} \pm \boldsymbol{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \pm \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_1 \pm y_1 \\ \vdots \\ x_m \pm y_m \end{bmatrix} \tag{12}$$

Here too, the dimensions must agree.

Vector Norms

An important class of vector operations is the norm, often denoted by $\|\cdot\|$. Norms measure the size or length of a vector. The most familiar norm is the Euclidean norm:

$$\|\boldsymbol{v}\|_2 = \sqrt{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \tag{13}$$

This is closely related to Euclidean distance:

$$d(\boldsymbol{v}, \boldsymbol{u}) = \|\boldsymbol{v} - \boldsymbol{u}\|_2 = \sqrt{(\boldsymbol{v} - \boldsymbol{u})^{\mathrm{T}}(\boldsymbol{v} - \boldsymbol{u})} = \sqrt{\sum_{i=1}^{n}(v_i - u_i)^2} \tag{14}$$

A broader family is the p-norm:

$$\|\boldsymbol{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{1/p}, \qquad p \geq 1. \tag{15}$$

Two especially common special cases are the Manhattan norm and the infinity norm:

$$\|\boldsymbol{v}\|_1 = \sum_{i=1}^{n} |v_i| \tag{16}$$

$$\|\boldsymbol{v}\|_\infty = \max\{|v_1|, \ldots, |v_n|\} \tag{17}$$

Linear and Quadratic Equations

Matrix notation gives a compact way to express linear and quadratic systems. A system of linear equations may be written as

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0} \tag{18}$$

Likewise, a quadratic form may be written as

$$y = \boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} + b = 0 \tag{19}$$

Miscellaneous Structures and Operations

It is often useful to represent vectors with repeated entries. The null vector $\mathbf{0}$ and the one vector $\mathbf{1}$ contain only zeros and ones, respectively. Their dimensions are usually implied by context. One useful identity is that $\mathbf{1}^{\mathrm{T}}\boldsymbol{x}$ sums the entries of $\boldsymbol{x}$.

$$\mathbf{0} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{20}$$

The $\operatorname{diag}(\cdot)$ operator transforms a vector $\boldsymbol{v} \in \mathbb{R}^n$ into a diagonal matrix $V \in \mathbb{R}^{n \times n}$:

$$\operatorname{diag}(\boldsymbol{v}) = \operatorname{diag}\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & v_n \end{bmatrix} = V. \tag{21}$$

A closely related object is the identity matrix,

$$I_n = \mathrm{diag}(\mathbf{1}) \tag{22}$$

which satisfies $AI_n = A$ and plays the same role in matrix multiplication that 1 plays in ordinary arithmetic.

Finally, vectors and matrices can be augmented by concatenating objects of compatible dimensions. For example, a matrix $A \in \mathbb{R}^{m\times n}$ may be augmented by another matrix $E \in \mathbb{R}^{m\times k}$:

$$\begin{bmatrix} A & E \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & e_{11} & \cdots & e_{1k} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & e_{m1} & \cdots & e_{mk} \end{bmatrix} \tag{23}$$

Likewise, a matrix may be augmented by a row vector:

$$\begin{bmatrix} A \\ \boldsymbol{v}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \\ v_1 & \cdots & v_n \end{bmatrix} \tag{24}$$