# A Brief Overview of Matrix Algebra

*David Ubilava*

*August, 2017*

A *matrix* is a way of organizing information. Until 1800s matrices were known as *arrays*. The *Jiuzhang Suanshu* (Chinese: "Nine Chapters on the Mathematical Art") – composed by several generations of scholars, and completed by 200 AD – is probably the first text that gives a known example of the use of arrays to solve simultaneous equations. In 1545, Italian mathematician Girolamo Cardano brought the method to Europe when he published *Ars Magna* (Latin: "The Great Art"). Subsequently, several mathematicians applied arrays in their works. But it was not until 1850, when James Joseph Sylvester (1814-1897) coined the term "matrix" (Latin: "womb", derived from *mater* – mother).

## Notation and Definitions

A *matrix* is a rectangular array of numbers. In particular, an $n \times k$ matrix has $n$ rows and $k$ columns. The positive integer $n$ is called the *row dimension*, and the positive integer $k$ is called the *column dimension*. We can present an $n \times k$ matrix as:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,k} \end{pmatrix}$$

where $a_{i,j}$ represents the element in the $i$th row and $j$th column.

A matrix may have a single row or a single column, both referred as *vectors*. A $1 \times k$ matrix is called a *row vector*, and can be written as:

$$\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_k \end{pmatrix}$$

Similarly, an $n \times 1$ matrix is called a *column vector* and can be written as:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

A $1 \times 1$ matrix yields a single number, referred to as *scalar*. As per convention, we shall denote: scalars by lower case letters; vectors by lower case boldface letters; and matrices by upper case boldface letters.

### Square Matrix

A *square matrix* has the same number of rows and columns. That is, the dimension of a square matrix is its number of rows and columns. For example,

$$\mathbf{B} = \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix}$$

is a 2–dimensional square matrix.

## Diagonal Matrix

A *diagonal matrix* is a square matrix with all of its off-diagonal elements zero. That is, $a_{i,j} = 0$ for all $i \neq j$. We can always write an $n$–dimensional diagonal matrix as:

$$\mathbf{C} = \begin{pmatrix} c_{1,1} & 0 & \cdots & 0 \\ 0 & c_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_{n,n} \end{pmatrix}$$

## Identity Matrix

An *identity matrix*, denoted by $\mathbf{I}_n$, is the $n$–dimensional diagonal matrix with unity (one) in each diagonal position, and zero elsewhere. For example, the 2–dimensional identity matrix is:

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

## Zero Matrix

The $n \times k$ *zero matrix*, denoted by $\mathbf{0}_{n,k}$, is the $n \times k$ matrix with zero in all entries. This need not be a square matrix.

# Basic Matrix Operations

## Matrix Addition and Multiplication

Two matrices of same dimension can be added (or subtracted) element by element. That is,

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & \cdots & a_{1,k} + b_{1,k} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & \cdots & a_{2,k} + b_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} + b_{n,1} & a_{n,2} + b_{n,2} & \cdots & a_{n,k} + b_{n,k} \end{pmatrix}.$$

To multiply matrix $\mathbf{A}$ by matrix $\mathbf{B}$, the two must be *conformable*. That is, the column dimension of $\mathbf{A}$ must equal the row dimension of $\mathbf{B}$. For example, let $\mathbf{A}$ be an $n \times k$ matrix and let $\mathbf{B}$ be an $k \times m$ matrix. Then the product, $\mathbf{AB}$, is an $n \times m$ matrix defined by:

$$\mathbf{AB} = \left( \sum_{r=1}^{k} a_{i,r} b_{r,j} \right) \qquad \forall i, j.$$

That is, the $(i, j)$th element of the new matrix $\mathbf{AB}$ is obtained by multiplying each element in the $i$th row of $\mathbf{A}$ by the corresponding element in the $j$th column of $\mathbf{B}$, and adding these $k$ products together.

$$\mathbf{AB} = \begin{pmatrix} \sum_{r=1}^{k} a_{1,r} b_{r,1} & \sum_{r=1}^{k} a_{1,r} b_{r,2} & \cdots & \sum_{r=1}^{k} a_{1,r} b_{r,m} \\ \sum_{r=1}^{k} a_{2,r} b_{r,1} & \sum_{r=1}^{k} a_{2,r} b_{r,2} & \cdots & \sum_{r=1}^{k} a_{2,r} b_{r,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{r=1}^{k} a_{n,r} b_{r,1} & \sum_{r=1}^{k} a_{n,r} b_{r,2} & \cdots & \sum_{r=1}^{k} a_{n,r} b_{r,m} \end{pmatrix}$$

For example, if $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 4 & 3 \\ 2 & 1 & 0 \end{pmatrix}$, then $\mathbf{AB} = \begin{pmatrix} 4 & 9 & 6 \\ 8 & 4 & 0 \end{pmatrix}$

We can also multiply a matrix and a vector (that are conformable). If $\mathbf{A}$ is an $n \times k$ matrix and $\mathbf{y}$ is an $k \times 1$ vector, then $\mathbf{Ay}$ is an $n \times 1$ vector. If $\mathbf{x}$ is a $1 \times n$ vector, then $\mathbf{xA}$ is a $1 \times k$ vector.

For any given real number $c$, scalar multiplication, that is, the multiplication of a matrix by a scalar, is an element by element operation. Specifically,

$$c\mathbf{A} = \begin{pmatrix} ca_{1,1} & ca_{1,2} & \cdots & ca_{1,k} \\ ca_{2,1} & ca_{2,2} & \cdots & ca_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n,1} & ca_{n,2} & \cdots & ca_{n,k} \end{pmatrix}.$$

To reiterate, if $\mathbf{A}$ is $n \times k$ and $\mathbf{B}$ is $k \times m$ then $\mathbf{AB}$ is defined, but $\mathbf{BA}$ is defined only if $n = m$. So, if $\mathbf{A}$ is $n \times k$ and $\mathbf{B}$ is $k \times n$, then $\mathbf{AB}$ and $\mathbf{BA}$ are both defined, but they are not usually the same. In fact, $\mathbf{AB}$ and $\mathbf{BA}$ have different dimensions, unless $\mathbf{A}$ and $\mathbf{B}$ are both square matrices. But even when $\mathbf{A}$ and $\mathbf{B}$ are both square matrices, $\mathbf{AB} \neq \mathbf{BA}$, except under special circumstances.

### Transpose

The *transpose* of an $n \times k$ matrix $\mathbf{A}$ is the $k \times n$ matrix $\mathbf{A}'$, sometimes also denoted as $\mathbf{A}^t$ or $\mathbf{A}^T$, which is obtained by interchanging the rows and columns of $\mathbf{A}$. For example, if $\mathbf{A} = \begin{pmatrix} 2 & 1 & 7 \\ 4 & 5 & 0 \end{pmatrix}$, then $\mathbf{A}' = \begin{pmatrix} 2 & 4 \\ 1 & 5 \\ 7 & 0 \end{pmatrix}$.

### Symmetric Matrix

A square matrix $\mathbf{A}$ is a *symmetric matrix* iff $\mathbf{A}' = \mathbf{A}$. For any $n \times k$ matrix $\mathbf{X}$, $\mathbf{X}'\mathbf{X}$ is always defined and is a symmetric matrix.

### Trace

The trace of a matrix is defined only for square matrices. For any $n \times n$ matrix $\mathbf{A}$, the *trace* of $\mathbf{A}$, $\text{tr}(\mathbf{A})$, is the sum of its diagonal elements. Mathematically, $\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{i,i}$.

### Matrix Inverse

The notion of a *matrix inverse* applies only to square matrices. An $n \times n$ matrix $\mathbf{A}$ has an inverse, denoted by $\mathbf{A}^{-1}$, provided that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$. When the above equalities hold, $\mathbf{A}$ is said to be invertible or *nonsingular*. Otherwise, it is said to be noninvertible of *singular*.

## Linear Independence and Matrix Rank

### Linear Independence

For a set of vectors of the same dimension, it is important to know whether one vector can be expressed as a linear combination of the remaining vectors. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ be a set of $n \times 1$ vectors. These are *linearly independent vectors* iff

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_r\mathbf{x}_r = \mathbf{0}$$

holds for $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$. Alternatively, if the aforementioned equality holds for a set of scalars that are not all zero, then $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ are *linearly dependent*.

### Rank

Let $\mathbf{A}$ be an $n \times k$ matrix, where $k \leq n$. The *rank* of a matrix $\mathbf{A}$, denoted as rank($\mathbf{A}$), is the maximum number of linearly independent columns of $\mathbf{A}$. So, the rank of $\mathbf{A}$ can be at most $m$. If rank($\mathbf{A}$) $= k$, then $\mathbf{A}$ has *full rank*. That is, a matrix has full rank if its columns form a linearly independent set. For example, the rank of the following $3 \times 2$ matrix

$$\begin{pmatrix} 1 & 3 \\ 2 & 6 \\ 0 & 0 \end{pmatrix}$$

can be at most two. But its rank, in fact, is only one because the second column is three times the first column.

## Quadratic Forms and Positive Definiteness

Let $\mathbf{A}$ be an $n \times n$ symmetric matrix. The quadratic form associated with the matrix $\mathbf{A}$ is the real-valued function defined for all $n \times 1$ vectors $\mathbf{x}$:

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n} a_{i,i} x_i^2 + 2 \sum_{i=1}^{n} \sum_{j>1} a_{i,j} x_i x_j.$$

A symmetric matrix $\mathbf{A}$ is said to be *positive definite* (p.d.), if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $n \times 1$ vectors $\mathbf{x}$ except $\mathbf{x} = \mathbf{0}$. A symmetric matrix $\mathbf{A}$ is said to be *positive semi-definite* (p.s.d.), if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $n \times 1$ vectors. If a matrix is positive definite or positive semi-definite, it is automatically assumed to be symmetric.

## Matrix Differentiation

For an $n$–dimensional vectors $\mathbf{a}$ and $\mathbf{x}$, consider a linear function given by

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x}.$$

Then,

$$\partial f(\mathbf{x})/\partial \mathbf{x} = \mathbf{a}',$$

which is a $1 \times n$ vector of partial derivatives.

For an $n \times n$ symmetric matrix $\mathbf{A}$, define a quadratic function

$$g(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}.$$

Then,

$$\partial g(\mathbf{x})/\partial \mathbf{x} = 2\mathbf{x}'\mathbf{A},$$

which is, again, a $1 \times n$ vector of partial derivatives.

## An Application: The Ordinary Least Squares

Consider a linear model:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

where $\mathbf{y}$ is an $n$–dimensional vector of the dependent variable; $\mathbf{X}$ is an $n \times k$ matrix of the independent variables; $\beta$ is a $k$–dimensional vector of the parameters; and $\varepsilon$ is an $n$–dimensional vector of error terms.

Let $\mathbf{b}$ be an estimator of the $\beta$, so that

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e},$$

where $\mathbf{e}$ is an $n$–dimensional vector of residuals. We can obtain the least squares estimator $\hat{\beta}$ by minimizing the sum of squared residuals with respect to $\mathbf{b}$:

$$\min_{\mathbf{b}} \mathbf{S} = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}.$$

Differentiating $\mathbf{S}$ with respect to $\mathbf{b}$ yields the vector of first order conditions:

$$\frac{\partial \mathbf{S}}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{0}.$$

After rearranging terms,

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

Given that the rank of the matrix $\mathbf{X}$ is $k$, the symmetric matrix $\mathbf{X}'\mathbf{X}$ will be of full rank, and its inverse, $(\mathbf{X}'\mathbf{X})^{-1}$, will exist. Premultiplying the above equation by this inverse, will yield the expression for $\hat{\beta}$:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$