2.6 Matrices: Basic Definitions & Review

In this section, we summarize all the Linear Algebra terms that we are going to use in this course. A complete list of matrix definitions and properties, including *random matrices* can be found here: The Matrix Cookbook.

A **matrix** is a rectangular array arranged in rows and columns. For example

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix}$$

is a 3×2 matrix. The elements of this matrix are denoted by α_{ij}, where i is the index that corresponds to the *row number* and j is the index that corresponds to the *column number*.

Square Matrix

A matrix is said to be **square** if the number of rows equals the number of columns. For example,

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix}$$

is a 3×3 matrix.

Upper /Lower Triangular Matrix

A *square* matrix is called **upper triangular** if all the entries *below* the main diagonal are zero. Similarly, a square matrix is called **lower triangular** if all the entries above the main diagonal are zero. For example, a *lower* triangular matrix is

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{bmatrix}$$

and an *upper* triangular is

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

Diagonal Matrix

A **diagonal** matrix is a matrix whose off-diagonal elements are all zero. For example,

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{bmatrix}$$

Transpose

The **transpose** of a matrix $\mathbf{A}$ is another matrix, denoted by $\mathbf{A}^T$, that is obtained by *interchanging* corresponding columns and rows of the matrix $\mathbf{A}$. As a special case, the transpose of a column vector is a row vector and vice versa.

For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 6 \\ 2 & 5 & 7 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A}^T = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}$$

Symmetric Matrix

If $\mathbf{A} = \mathbf{A}^T$ then the matrix $\mathbf{A}$ is called **symmetric**. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 5 \\ 6 & 5 & 3 \end{bmatrix} = \mathbf{A}^T$$

Identity Matrix

The **identity** or **unit matrix** is denoted by $\mathbf{I}$. It is a *diagonal* matrix whose elements on the main diagonal are all *1*. For example,

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

Equality of Matrices

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be equal if they have the *same dimension* and if all the corresponding elements are equal. Conversely, if two matrices are equal, their corresponding elements are equal. For example, if

$$\mathbf{A} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix}$$

then $\mathbf{A} = \mathbf{B}$ implies $\alpha_1 = 4$, $\alpha_2 = 3$, $\alpha_3 = 7$.

Linear Dependence & Rank of Matrix

Consider the following matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 & 1 \\ 2 & 2 & 10 & 6 \\ 3 & 4 & 15 & 1 \end{bmatrix}$$

Now, think of all the columns of this matrix as *vectors.* Then, we can think of A as being made up of 4 column vectors. Observe here that the third column is a multiple of the first one. Indeed,

$$\begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

If this is the case, we say that A is *linearly dependent*, since one column can be obtained as a linear combination of the others.

Define a set of c column vectors $\mathbf{C}_1, \ldots, \mathbf{C}_c$ in an $r \times c$ matrix to be *linearly dependent*, if one vector can be expressed as a linear combination of the others. If no vector in the set can be so expressed, we define the set of vectors to be *linearly independent*.

Rank of a Matrix

The **rank** of a matrix is defined to be the *maximum* number of linearly independent columns in the matrix. For example, the rank of matrix $\mathbf{A}$ above is 3. The rank of a matrix is **unique** and can equivalently be defined as the maximum number of linearly independent rows. It follows that the rank of an $r \times c$ matrix cannot exceed $\min(r, c)$. Also, a matrix $\mathbf{A}$ is called **full rank**, if

$$rank(\mathbf{A}) = min(r, c).$$

Inverse of a Matrix

The **inverse** of a matrix is another matrix, denoted by $\mathbf{A}^{-1}$, such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$$

The inverse of a square matrix of dimension 2 can be easily computed:

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \quad \Leftrightarrow \quad \mathbf{A}^{-1} = \frac{1}{\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}} \begin{bmatrix} \alpha_{22} & -\alpha_{12} \\ -\alpha_{21} & \alpha_{11} \end{bmatrix}$$

Note that the inverse of a matrix is defined only for *square* matrices.

Singular Matrix

A square matrix that is not invertible is called **singular** or degenerate.

Orthogonal Matrix

An **orthogonal** matrix is a real square matrix whose columns and rows are orthonormal vectors. For an orthogonal matrix

$$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$$

and as a consequence $\mathbf{A}^T = \mathbf{A}^{-1}$.

Trace of a Matrix

The **trace** of a square matrix $\mathbf{A}$, denoted $tr(\mathbf{A})$ is defined to be the sum of the elements on the main diagonal of $\mathbf{A}$.

Idempotent

An **idempotent** matrix is a matrix which, when multiplied by itself, yields itself; that is

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}$$

2.6.1 Basic Results on Matrices

Some basic properties on matrices are summarized below:

$$\begin{aligned}
\mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \\
(\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \\
(\mathbf{A}\mathbf{B})\mathbf{C} &= \mathbf{A}(\mathbf{B}\mathbf{C}) \\
\mathbf{C}(\mathbf{A} + \mathbf{B}) &= \mathbf{C}\mathbf{A} + \mathbf{C}\mathbf{B} \\
k(\mathbf{A} + \mathbf{B}) &= k\mathbf{A} + k\mathbf{B} \\
(\mathbf{A}^T)^T &= \mathbf{A} \\
(\mathbf{A} + \mathbf{B})^T &= \mathbf{A}^T + \mathbf{B}^T \\
(\mathbf{A}\mathbf{B})^T &= \mathbf{B}^T\mathbf{A}^T \\
(\mathbf{A}\mathbf{B}\mathbf{C})^T &= \mathbf{C}^T\mathbf{B}^T\mathbf{A}^T \\
(\mathbf{A}\mathbf{B})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{A}\mathbf{B}\mathbf{C})^{-1} &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{A}^{-1})^{-1} &= \mathbf{A} \\
(\mathbf{A}^T)^{-1} &= (\mathbf{A}^{-1})^T
\end{aligned}$$

2.6.2 Random Vectors & Matrices

Let $\mathbf{Z}$ be a random vector of size $m \times 1$, with components $Z_1, Z_2, \ldots, Z_m$.

- The **mean** of $\mathbf{Z}$ is equal to vector μ defined as:

$$\mu = \mathbb{E}(\mathbf{Z}) = \begin{pmatrix} \mathbb{E}(Z_1) \\ \mathbb{E}(Z_2) \\ \dots \\ \mathbb{E}(Z_m) \end{pmatrix}$$

- The **variance** of a random vector $\mathbf{Z}$ is a matrix – the Variance-Covariance matrix. This matrix is *symmetric* (why?) of size $m \times m$ with component (i, j) equal to the $Cov(Z_i, Z_j)$. Specifically,

$$\begin{aligned} \Sigma_{m \times m} &= Cov(\mathbf{Z}) \\ &= \mathbb{E}\left((\mathbf{Z} - \mu)(\mathbf{Z} - \mu)^T\right) \\ &= \begin{pmatrix} Var(Z_1) & \dots & Cov(Z_1, Z_m) \\ \dots & \dots & \dots \\ Cov(Z_m, Z_1) & \dots & Var(Z_m) \end{pmatrix} \end{aligned}$$

Affine transformations of $\mathbf{Z}$

A matrix $\mathbf{W}$ is an affine transformation of a random vector $Z_{m \times 1}$:

$$\mathbf{W} = \mathbf{a_{n \times 1}} + \mathbf{B_{n \times m} Z_{m \times 1}}$$

- The *mean* and *covariance* matrix of $\mathbf{W}$

$$\mathbb{E}(\mathbf{W}) = \mathbf{a} + \mathbf{B}\mu,$$

$$Cov(\mathbf{W}) = \mathbf{B \Sigma B}^T$$

Another transformation of $\mathbf{Z}$ is the following

$$\mathbf{U} = \mathbf{v^T Z} = v_1 Z_1 + v_2 Z_2 + \dots + v_m Z_m$$

- The *mean* and *covariance* matrix of $\mathbf{U}$ is

$$\mathbb{E}(W) = \mathbf{v}^T\mu = \sum_{i=1}^{m} v_i\mu_i$$

$$Var(W) = \mathbf{v}^T\Sigma\mathbf{v} = \sum_{i=1}^{m} v_i^2 Var(Z_i) + 2\sum_{i<j} v_i v_j Cov(Z_i, Z_j)$$