Linear Statistical Models - MAST30025
Akira Wang

# 1 Linear Algebra Review

## 1.1 Transposition:

The *transpose* of a matrix results when the rows and columns are interchanged.

A matrix $X$ is **symmetric** if and only if (iff) $X^T = X$.

Properties that should be noted:

1. $(X^T)X = X$.
2. $(XY)^T = Y^T X^T$.

## 1.2 Identity:

The matrix *identity* $I$ is a square matrix of arbitrary size with 1's on the diagonals (diags).

A $k \times k$ identity matrix is denoted as $I_k$.

## 1.3 Inverse:

If $X$ is a **square matrix**, its *inverse* is the matrix $X^{-1}$ of the same size which satisfies the following:

$$XX^{-1} = X^{-1}X = I \tag{1}$$

Properties that should be noted:

1. $(X^{-1})^{-1} = X$.
2. $(XY)^{-1} = Y^{-1}X^{-1}$.
3. $(X^T)^{-1} = (X^{-1})^T$

## 1.4 Orthogonal Vectors and Matrices:

Two $n \times 1$ vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* iff:

$$\mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i y_i = 0. \tag{2}$$

A square matrix $X$ is *orthogonal* iff:

$$X^T X = I. \tag{3}$$

If X is orthogonal, then

$$X^{-1} = X^T. \tag{4}$$

## 1.5 Orthogonality:

$$\text{If } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ then } \mathbf{x}^T\mathbf{x} = ||\mathbf{x}||^2 = \sum_{i=1}^{n} x_i^2 \tag{5}$$

The square root of $||\mathbf{x}^2||$, denoted by $||\mathbf{x}||$, is called the **norm** or **length** of x.

A matrix $X$ is an orthogonal matrix iff the columns of $X$ form an orthonormal set.

## 1.6 Eigenvalues and Eigenvectors:

Suppose $A$ is a $k \times k$ matrix and x is a $k \times 1$ nonzero vector which satisfies the equation

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{6}$$

We say that $\lambda$ is an *eigenvalue* of $A$, with associated *eigenvector* $\mathbf{x}$.

To find the eigenvalues, solve the following equation:

$$|A - \lambda I| = 0. \tag{7}$$

To then find the eigenvector(s) of $A$ associated with the eigenvalue(s), solve the linear system of equations:

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{8}$$

The linear system should have an infinite number of solutions, which will always happen for an eigenvector system.

> Note:
>
> If $A$ is **symmetric**, then all its eigenvalues are **real**, and its eigenvectors all **orthogonal**.

## 1.7 Diagonalization:

Let $A$ be a **symmetric** $k \times k$ ,matrix. Then an **orthogonal** matrix $P$ exists such that

$$P^T AP = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix} \text{where } \lambda_i, i = 1, 2, \dots, k, \text{are the eigenvalues of } A. \tag{9}$$

We say that $P$ *diagonalizes* $A$, where $A$ is the *diagonalizable* matrix, and $P^T AP$ is the *diagonalized* matrix.

## 1.8 Linear Independence:

A set of vectors is *linearly depending* iff there exists some numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$, which are not all zero, such that

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_k\mathbf{x}_k = \mathbf{0}. \tag{10}$$

If all $\alpha$ are zero, then they are are *linearly independent*.

If a set of vectors is linearly dependent, then at least one of the vectors can be written as a **linear combination** of some or all the other vectors.

## 1.9 Rank:

Some definitions (Yao-Ban):

1. Tall Matrix has more rows than columns ($m > n$) and is of ***full rank***.
2. Short Matrix has more columns than rows ($n > m$).

The *rank* of $X$, denoted by $r(X)$, is the greatest number of **linearly independent vectors** in the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$.

Properties that should be noted:

1. For any matrix $X$, we have $r(X) = r(X^T) = r(X^TX)$
2. The rank of a diagonal matrix is equal to the number of nonzero diagonal entries in the matrix.

## 1.10 Idempotence:

A square matrix $A$ is *idempotent* iff

$$A^2 = A \tag{11}$$

It should be noted that if $A$ is diagonizable, then $A$ is idempotent.

Properties that should be noted:

1. The eigenvalues of any idempotent matrix is always either 0 or 1
   **Proof.**

   $\mathrm{A}^2\mathbf{x} = A(Ax)$
   $= A(\lambda x), \text{since } A\mathbf{x} = \lambda\mathbf{x}$
   $= \lambda(A\mathbf{x})$
   $= \lambda^2\mathbf{x}$

   By definition, $\mathbf{x} \neq \mathbf{0}$, so $\lambda = \lambda^2$. Therefore $\lambda = 0, 1$.

2. If $A$ is a **symmetric** and **idempotent** matrx, $r(A) = tr(A)$.

## 1.11 Trace:

The *trace* of a square matrix $X$, denoted by $tr(X)$, is the sum of its diagonal entries

$$tr(X) = \sum_{i=1}^{k} x_{ii}. \tag{12}$$

Properties that should be noted:

1. If $c$ is a scalar, $tr(cX) = ctr(X)$
2. If $XY$ and $YX$ both exist, $tr(XY) = tr(YX)$

## 1.12 Quadratic Forms:

Suppose $A$ is a square matrix, and y is a $k \times 1$ vector containing variables. The quantity

$$q = \mathbf{y}^T A\mathbf{y} \tag{13}$$

is called a *quadratic form* in $\mathbf{y}$, and $A$ is called the matrix of the *quadratic form*.

However, since $q$ is a scalar, it can be re-expressed as

$$q = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} y_i y_j. \tag{14}$$

**Example:**

$$\text{Let } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \text{and } A = \begin{pmatrix} \boxed{2} & 3 & 1 \\ 1 & 2 & 0 \\ 4 & 6 & \boxed{3} \end{pmatrix}$$

Then

$$\begin{aligned}
\mathbf{y}^T A \mathbf{y} &= \Big|\; 2y_1^2 \;\Big| + \Big|\; 3y_1y_2 + y_1y_3 \;\Big| + \Big|\; y_2y_1 + 2y_2^2 + 4y_3y_1 \;\Big| + \Big|\; 0 + 6y_3y_2 \;\Big| + \Big|\; 3y_3^2 \;\Big| \qquad (15) \\
&= 2y_1^2 + 2y_2^2 + 3y_3^2 + 4y_1y_2 + 5y_1y_3 + 6y_2y_3 \qquad (16)
\end{aligned}$$

This can be found from either the summation formula (14) or by multiplying out the matrices.

Positive Definiteness:

1. If $\mathbf{y}^T A \mathbf{y} > 0$, then it is positive definite.
2. If $\mathbf{y}^T A \mathbf{y} \geq 0$, then it is positive semi-definite.

Properties that should be noted:

1. The quadratic form will **never** be negative.
2. A symmetric matrix $A$ is **positive definite** iff all its eigenvalues are all (strictly) positive.
3. A symmetric matrix $A$ is **positive semi-definite** iff all its eigenvalues are all non-negative.

## 1.13 Differentiation of Quadratic Forms:

Suppose we have a vector of variables $\mathbf{y} = (y_1, y_2, \ldots, y_k)^T$, and some scalar function of them

$$z = f(\mathbf{y}). \tag{17}$$

We can then define the derivative of $z$ with respect to $\mathbf{y}$ as follows:

$$\frac{\partial z}{\partial \mathbf{y}} = \begin{bmatrix} \frac{\partial z}{\partial y_1} \\ \frac{\partial z}{\partial y_2} \\ \vdots \\ \frac{\partial z}{\partial y_k} \end{bmatrix} \tag{18}$$

To do so, you can take the **quadratic form** of $z$ and **partial derive** with respect to $\mathbf{y}$.

**Example from Figure 14:**

$$\frac{\partial z}{\partial \mathbf{y}} = \begin{bmatrix} 4y_1 + 4y_2 + 5y_3 \\ 4y_2 + 4y_1 + 6y_3 \\ 6y_3 + 5y_1 + 6y_2 \end{bmatrix} \tag{19}$$

# 2 Random Vectors

## 2.1 Expectation:

**Note that we will denote random variables (r.v's) as lower-cases, according to linear algebra notation.**

We define the *expectation* of a random vector $\mathbf{y}$ as follows:

$$\text{If } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, \text{then } E[\mathbf{y}] = \begin{bmatrix} E[y_1] \\ E[y_2] \\ \vdots \\ E[y_k] \end{bmatrix} \tag{20}$$

Properties that should be noted:

1. If $\mathbf{a}$ is a vector of constants, then:
   (a) $E[\mathbf{a}] = \mathbf{a}$.
   (b) $E[\mathbf{a}^T\mathbf{y}] = \mathbf{a}^T e[\mathbf{a}]$.
2. If $A$ is a matrix of constants, then $E[A\mathbf{y}] = AE[\text{y}]$.

## 2.2 Variance: