Linear Statistical Models (MAST30025) Notes
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 $\mathbf{x}$.

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 $\mathbf{x}$ is a $k \times 1$ nonzero vector which satisfies the equation

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

We say that λ 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

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

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 α 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 $r(X) = r(X^T) = r(X^T X)$
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.

$$\begin{aligned} A^2\mathbf{x} &= A(Ax) && (12)\\ &= A(\lambda x), \text{since } A\mathbf{x} = \lambda\mathbf{x} && (13)\\ &= \lambda(A\mathbf{x}) && (14)\\ &= \lambda^2\mathbf{x} && (15)\\ & && (16)\\ & && (17) \end{aligned}$$

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

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

1.11 Trace:

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

$$tr(X) = \sum_{i=1}^{k} x_{ii}. \tag{18}$$

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 $\mathbf{y}$ is a $k \times 1$ vector containing variables. The quantity

$$q = \mathbf{y}^T A \mathbf{y} \tag{19}$$

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

$$q = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} y_i y_j. \tag{20}$$

Example:

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

Then

$$\mathbf{y}^T A \mathbf{y} = \Big| \, 2y_1^2 \, \Big| + \Big| \, 3y_1y_2 + y_1y_3 \, \Big| + \Big| \, y_2y_1 + 2y_2^2 + 4y_3y_1 \, \Big| + \Big| \, 0 + 6y_3y_2 \, \Big| + \Big| \, 3y_3^2 \, \Big| \tag{21}$$

$$= 2y_1^2 + 2y_2^2 + 3y_3^2 + 4y_1y_2 + 5y_1y_3 + 6y_2y_3 \tag{22}$$

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

$$z = f(\mathbf{y}). \tag{23}$$

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

$$\frac{\partial z}{\partial \mathbf{y}} = \begin{bmatrix} \frac{\partial z}{\partial y_1} \\ \frac{\partial z}{\partial y_2} \\ \vdots \\ \frac{\partial z}{\partial y_k} \end{bmatrix} \tag{24}$$

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

Example from Figure 14 and 15:

$$\frac{\partial z}{\partial \mathbf{y}} = \begin{bmatrix} 4y_1 + 4y_2 + 5y_3 \\ 4y_2 + 4y_1 + 6y_3 \\ 6y_3 + 5y_1 + 6y_2 \end{bmatrix} \tag{25}$$

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:

$$\text{If } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, \text{then } E[\mathbf{y}] = \begin{bmatrix} E[y_1] \\ E[y_2] \\ \vdots \\ E[y_k] \end{bmatrix} \tag{26}$$

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 $E[A\mathbf{y}] = AE[\mathbf{y}]$.

2.2 Variance:

Recall that the variance of a r.v Y with mean μ is defined to be $E[(Y-\mu)^2$.

We define the *variance* or *covariance matrix* of the random vector $\mathbf{y}$ to be

$$var(\mathbf{y}) = E[(\mathbf{y}-\mu)(\mathbf{y}-\mu)^T]. \tag{27}$$

The diagonal elements of the **covariance matrix** are the variances of the elements of $\mathbf{y}$:

$$[var(\mathbf{y})]_{ii} = \mathbf{y}_i, \; i = 1, 2, \ldots, k. \tag{28}$$

The off-diagonal elements are the covariances of the elements:

$$[var(\mathbf{y})]_{ij} = cov(y_i, y_j) = E[(y_1 - \mu_1)(y_1 - \mu_1)^T]. \tag{29}$$

This means that the covariance matrix is always **symmetric**.

Very Important Property!!!:
Suppose that $\mathbf{y}$ is a random vector with $var(\mathbf{y}) = V$, then,
if A is a matrix of **constants**, $var(A\mathbf{y}) = AVA^T$.

Properties that should be noted:

1. If $\mathbf{a}$ is a vector of constants, then $var(\mathbf{a}^T\mathbf{y}) = \mathbf{a}^T V \mathbf{a}$.
2. V is positive semi-definite, meaning $var(\mathbf{y}) \geq 0$.

Asume that X is a matrix of full rank (more rows than columns), which implies X^TX is nonsingular (invertible). Let

$$\mathbf{z} = (X^TX)^{-1}X^T\mathbf{y} = A\mathbf{y}. \tag{30}$$

Then (Remember that $V = var(\mathbf{y}),\ A = (X^TX)^{-1}X^T$),

$$\begin{aligned}
var(\mathbf{z}) &= AVA^T \\
&= [(X^TX)^{-1}X^T]\sigma^2 I[(X^TX)^{-1}X^T]^T \\
&= (X^TX)^{-1}X^T(X^T)^T[(X^TX)^{-1}]^T\sigma^2,\ X^T(X^T)^T = X^TX \\
&= (X^TX)^{-1}[(X^TX)^{-1}]^T(X^TX)\sigma^2 \\
&= (X^TX)^{-1}\sigma^2.
\end{aligned}$$

2.3 Matrix Square Root:

The square root of a matrix A is a matrix B such that $B^2 = A$. If A is **symmetric** and **positive semi-definite**, there is a unique symmetric positive semi-definite square root, called the *principle root*, denoted $A^{\frac{1}{2}} = (P\Lambda^{\frac{1}{2}}P^T)$

Proof.

$$\begin{aligned} A &= P\Lambda P^T \\ &= (P\Lambda^{\frac{1}{2}}P^T)(P\Lambda^{\frac{1}{2}}P^T) \end{aligned}$$

2.4 Multivariate Normal Distribution:

We say that

$$\mathbf{x} = A\mathbf{z} + \mathbf{b} \tag{31}$$

follows a multivariate normal distribution, with just $\mu = \mathbf{b}$ and covariance matrix $\sum = AA^T$, and write $\mathbf{x} \sim MVN(\mu, \sum)$.

Any linear combination of multivariate normals results in another multivariate normal. For example:

$$\text{If } \mathbf{x} \sim MVN(\mu, \sum), \text{then } \mathbf{y} = A\mathbf{x} + \mathbf{b} \sim MVN(A\mu + \mathbf{b}, A\sum A^T). \tag{32}$$

2.5 Random Quadratic Forms:

We have seen that a matrix induces a quadratic form (multivariate function) that looks like $\mathbf{y}^T A\mathbf{y}$. The form becomes a scalar function of r.v's, so it itself is a r.v.

Theorem 3.2:
Let $\mathbf{y}$ be a random vector with $E[\mathbf{y}] = \mu$, $V = var(\mathbf{y})$, and A be a matrix of constants. Then

$$E[\mathbf{y}^T A\mathbf{y}] = tr(AV) + \mu^T A\mu \tag{33}$$

Example:
Let $\mathbf{y}$ be a 2×1 random vector with

$$E[\mathbf{y}] = \mu = \begin{bmatrix} 1 \\ 3 \end{bmatrix},\ V = var(\mathbf{y}) = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}. \tag{34}$$

$$\text{And let } A = \begin{pmatrix} \boxed{4} & \boxed{1} \\ \boxed{1} & \boxed{2} \end{pmatrix}$$

The quadratic form is given as

$$\mathbf{y}^T A\mathbf{y} = 4y_1^2 + 2y_1y_2 + 2y_2^2. \tag{35}$$

The expecation of this form is

$$E[\mathbf{y}^T A\mathbf{y}] = 4E[y_1^2] + 2E[y_1y_2] + 2E[y_2^2]. \tag{36}$$

From the given covariance matrix V,

1. $V_{11} = 2 = vary_1 = E[y_1^2] - E[y_1]^2 = E[y_1^2] - 1^2$
2. $V_{22} = 5 = vary_2 = E[y_2^2] - E[y_2]^2 = E[y_2^2] - 3^2$

Solving both equations gives $E[y_1^2] = 3$ and $E[y_2^2] = 14$. Finally,

$$V_{12} = V_{21} = 1 = cov(y_1, y_2) = E[y_1y_2] - E[y_1]E[y_2] = E[y_1y_2] - (1 \times 3) \tag{37}$$

Solving the equation gives $E[y_1y_2] = 4$, which gives

$$E[\mathbf{y}^T A\mathbf{y}] = 4 \times 3 + 2 \times 4 + 2 \times 14 = 48. \tag{38}$$

BUT, from Theorem 3.2,

$$\begin{aligned} E[\mathbf{y}^T A\mathbf{y}] &= tr(AV) + \mu^T A\mu \\ &= 9 + 11 + 7 + 21 \\ &= 48. \end{aligned}$$

2.6 Noncentral χ^2 Distribution:

Let $\mathbf{y} \sim MVN(\mu, I)$, then

$$x = \mathbf{y}^T\mathbf{y} = \sum_{i=1}^{k} y_i^2 \sim \chi^2_{k,\lambda} \tag{39}$$

has a *noncentral* χ^2 *distribution* with k degrees of freedom (d.o.f) and a noncentrality parameter $\lambda = \frac{1}{2}\mu^T\mu$.

WARNING:
R defines λ to be $\mu^T\mu$, so **double** it before putting it in!!!

Properties that should be noted:

1. $var(x) = 2k + 8\lambda$.
2. The noncentrality parameter λ is 0 iff $\mu = \mathbf{0}$.

Theorem 3.4:
Let $X^2_{k_1,\lambda_1}, \ldots, X^2_{k_k,\lambda_k}$ be a collection of n independent noncentral χ^2 random variables. Then

$$\sum_{i=1}^{n} = X^2_{k_i,\lambda_i} \tag{40}$$

has a noncentral χ^2 distribution with $k = \sum_{i=1}^{n}$ d.o.f and noncentrality parameter $\lambda = \sum_{i=1}^{n}\lambda_i$.
Setting $\lambda_i = 0$ results in a sum of independent χ^2 distributions.

2.7 Distribution of Quadratic Forms:

Theorem 3.5:
Let $\mathbf{y} \sim MVN(\mu, I)$ be a $n \times 1$ random vector and let A be a $n \times n$ **symmetric** matrix. Then $\mathbf{y}^T A\mathbf{y}$ has a **noncentral** χ^2 **distribution** with k d.o.f and noncentrality parameter $\lambda = \mu^T A\mu$ iff A **is idempotent and has rank** k.

Some definitions (Yao-Ban):
Let J_n be an $n \times n$ matrix filled with 1's. Example:

$$J_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{41}$$

Example:
Let y_1 and y_2 be independent normal r.v's with means 3 and -2 respectively, and variance 1. Let

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2}J_2 \tag{42}$$

Since A is **symmetric** and **idempotent**, it has rank 1. Therefore

$$\mathbf{y}^T A\mathbf{y} = \frac{1}{2}\begin{bmatrix} y_1 & y_2 \end{bmatrix} J_2 \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{2}y_1^2 + y_1y_2 + \frac{1}{2}y_2^2 \tag{43}$$

has a noncentral χ^2 distribution with 1 d.o.f and noncentrality parameter

$$\lambda = \frac{1}{2}\begin{bmatrix} 3 & -2 \end{bmatrix} J_2 \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \frac{1}{4} \tag{44}$$

2.8 Independence of Quadratic Forms:

Theorem 3.11:
Let $\mathbf{y} \sim MVN(\mu, V)$ be a $n \times 1$ random vector with nonsingular (invertible) variance V, and let A and B be square matrices. Then $\mathbf{y}^T A\mathbf{y}$ and $\mathbf{y}^T B\mathbf{y}$ are independent iff

$$AVB = 0 \tag{45}$$

Example:
Let y_1 and y_2 follow a MVN distribution with covariance matrix

$$V = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{46}$$

Consider the symmetric matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{47}$$

Then

$$\mathbf{y}^T A \mathbf{y} = y_1^2, \; \mathbf{y}^T B \mathbf{y} = y_2^2. \tag{48}$$

To find if they are independent, we solve $AVB = 0$ (the zero matrix)

$$\begin{aligned} AVB &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} \end{aligned}$$

This means that $b = 0$ and implies that $cov(y_1, y_2) = 0$. Therefore, the quadratic forms are independent.

$\boxed{\text{As a result, we can see that the quadratic forms are independent iff } AB = 0.}$