

A Brief Review of Matrix Algebra

G.J. LI

This note based on Wooldridge (2008) summarizes some matrix and probability algebra, needed for the study of multiple linear regression models. It is not meant to be exhaustive. For the proofs and more details on a certain topic on linear algebra and probability, please refer to the relevant chapters in Greene (2011) and Meyer (2000).

1 Basic Definitions and Operations

Definition 1.1. A **matrix** is a rectangular array of numbers. More precisely, an $m \times n$ matrix has m rows and n columns. The positive integer m is called the row dimension, and n is called the column dimension. A typical element in a matrix A can be represented as a_{ij} , which means it is the i th row and j th column element, and A can also be written as $[a_{ij}]$. If $m = n$, the matrix is called a **square matrix**. A square matrix A is a **diagonal matrix** when all of its off-diagonal elements are zero, that is, $a_{ij} = 0$ for all $i \neq j$.

Definition 1.2. For a matrix, if $m = 1$ and $n > 1$, it is called a **row vector**; if $m > 1$ and $n = 1$, it is called a **column vector**; if $m = 1$ and $n = 1$, it is called a **scalar**. For our later discussion, we refer to a **vector** as column vector.

Definition 1.3. Matrix Addition: Two matrices A and B , each having dimension $m \times n$, can be added element by element: $A + B = [a_{ij} + b_{ij}]$. For example, $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}$.

Definition 1.4. Given a scalar λ , **scalar multiplication** is defined as $\lambda A = [\lambda a_{ij}]$. For example, $\lambda \times \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} \\ \lambda a_{21} & \lambda a_{22} \end{pmatrix}$.

Definition 1.5. Matrix and Matrix Multiplication: To multiply matrix A by matrix B to form the product AB , the column dimension of A must equal the row dimension of B . Therefore, matrix multiplication is defined

as $\begin{matrix} A & B & C \\ m \times n & n \times p & m \times p \end{matrix} = [\sum_{k=1}^n a_{ik}b_{kj}] = [c_{ij}]$. For example, $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \times \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$.

Note that $A + B = B + A$; $(A + B) + C = A + (B + C)$; $(AB)C = A(BC)$; $A(B + C) = AB + AC$; in general, $AB \neq BA$, even when both products are defined.

Definition 1.6. The **range of a matrix** $A_{m \times n}$, denoted as $range(A)$, is the set of all linear combinations of the columns of A , i.e. $\{y | y = Ax = \sum_{i=1}^n a_i x_i, x \in R^n\}$, where a_i is the i th column of A and x_i is the i th element of the vector x . It is also called the **column space** of A . If $n < m$, the column space of A is a subspace of R^m .

Definition 1.7. An **identity matrix**, denoted as I or sometimes I_n to emphasize its dimension, is a square matrix with ones on the diagonal line

and zeros elsewhere, i.e. $I_{ii} = 1$ and $I_{ij} = 0$ for $i \neq j$. For example

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Identity matrix is an extension of the concept 1 in numbers. Note that $AI = IA = A$.

Definition 1.8. Inverse Matrix: For a square matrix A , if we can find another matrix B , such that $AB = BA = I$, then B is called the inverse of A and we can write it as $B = A^{-1}$. In this case, A is said to be **invertible** or **nonsingular**. Otherwise, it is said to be **noninvertible** or **singular**.

Note that $(A')^{-1} = (A^{-1})'$. If A and C are both invertible, then $(AC)^{-1} = C^{-1}A^{-1}$.

Definition 1.9. Transpose of a Matrix: Let $A = [a_{ij}]$ be an $m \times n$ matrix. The transpose of A , denoted A' or A^T (called A prime or A transpose), is the $n \times m$ matrix obtained by interchanging the rows and columns of A . We can

write this as $A' = [a_{ji}]$. For example, $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}$. A

is called **symmetric** if $A' = A$.

Note that $(A+B)' = B' + A'$, $(AB)' = B'A'$. If x and y are both vectors, then $x'y = y'x$.

Definition 1.10. The **inner product of two vectors**, x and y , is $x'y$. If

$x'y = 0$, we say x and y are **orthogonal** to each other.

正交矩阵 $Q^{-1}=Q^t$

Definition 1.11. The **euclidean norm of a vector** x is defined to be $\|x\| = \sqrt{x'x}$. The **distance** between two vectors x and y is $\|x - y\|$.

Definition 1.12. Let $\{a_1, a_2, \dots, a_r\}$ be a set of $n \times 1$ vectors. These are **linearly independent** vectors if and only if

$$\alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_r a_r = 0 \quad (1)$$

implies that $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$. If (1) holds for a set of scalars that are not all zero, then $\{a_1, a_2, \dots, a_r\}$ is **linearly dependent**.

Definition 1.13. Let A be an $n \times m$ matrix. The **rank of a matrix** A , denoted $rank(A)$, is the maximum number of linearly independent columns of A . If A is $n \times m$ and $rank(A) = m$, then A has **full column rank**.

Note that $rank(A') = rank(A)$; if A is $n \times k$, then $rank(A) \leq \min(n, k)$; if A is $k \times k$ and $rank(A) = k$, then A is **nonsingular**.

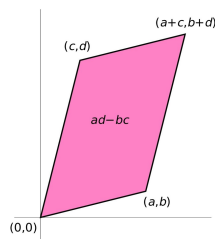
非奇异矩阵

Definition 1.14. The **trace of a square matrix** $A_{n \times n}$, denoted $tr(A) = \sum_{i=1}^n a_{ii}$, is the sum of its diagonal elements.

Note that $tr(A') = tr(A)$, $tr(AB) = tr(BA)$ and $tr(A + B) = tr(A) + tr(B)$.

Definition 1.15. The **determinant** is a value associated with a square matrix, (denoted as $det(\cdot)$ or $|\cdot|$) which measures the volume spanned by

the vectors in the matrix. For example, $\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc$.



We usually use computer program to calculate matrix determinant. Note that the determinant of a triangular matrix is the product of all of its diagonal elements. A singular square matrix has 0 determinant while the determinant of a nonsingular matrix is nonzero.

Definition 1.16. A **symmetric matrix** $A = A^t$ is said to be **positive definite (p.d.)** if $x'Ax > 0$ for all $n \times 1$ vectors x except $x = 0$. A symmetric matrix A is **positive semi-definite (p.s.d.)** if $x'Ax \geq 0$ for all $n \times 1$ vectors.

If a matrix is positive definite or positive semi-definite, it is automatically assumed to be symmetric. A positive definite matrix has diagonal elements that are strictly positive, while a p.s.d. matrix has nonnegative diagonal elements. If A is p.d., then A^{-1} exists and is p.d.. If X is $n \times k$, then $X'X$ and XX' are p.s.d. If X is $n \times k$ and $\text{rank}(X) = k$, then $X'X$ is p.d. (and therefore nonsingular).

自己乘自己的于自己

Definition 1.17. A is said to be an **idempotent matrix** if and only if $AA = A$. If A also happens to be a symmetric matrix, then A is called an **orthogonal projection matrix**.

If A is idempotent, then $\text{rank}(A) = \text{tr}(A)$ and A is positive semi-definite. If $\text{rank}(X_{n \times k}) = k$, then $P_X = X(X'X)^{-1}X'$ is the orthogonal projection

matrix, which projects any vector in R^n orthogonally onto the column space of X , i.e. $P_X y \in \text{range}(X)$ for any $y \in R^n$ and the distance between y and $P_X y$ is the smallest among the vectors in the column space of X . The matrix $M_X = I - P_X$ is also an orthogonal projection matrix, which projects any vector in R^n onto the space orthogonal to the column space of X , i.e. $(M_X y)' X = 0$. (*Q: What is the rank of P_X and M_X respectively?*)

2 Some Vector Calculus

For a given $n \times 1$ vector a , consider the linear function defined by $f(x) = a'x$, for all $n \times 1$ vectors x . The derivative of f with respect to x is the $1 \times n$ vector of partial derivatives¹, which is simply $\frac{\partial f(x)}{\partial x'} = a'$. For a given matrix A and the vector function $h(x) = Ax$, we have $\frac{\partial h(x)}{\partial x'} = A$. For a square matrix B , define the quadratic form $g(x) = x' B x$. Then $\frac{\partial g(x)}{\partial x'} = x'(B + B')$.

3 Moments of Random Vectors

Definition 3.1. If y is an $n \times 1$ random vector, the **expectation** of y , denoted $E(y)$, is the vector of expected values: $E(y) = [E(y_1), E(y_2), \dots, E(y_n)]'$. If Z is an $n \times m$ random matrix, $E(Z)$ is the $n \times m$ matrix of expected values: $E(Z) = [E(z_{ij})]$.

If A is an $m \times n$ matrix and b is an $n \times 1$ vector, where both are non-random, then $E(Ay + b) = AE(y) + b$. If A is $p \times n$ and B is $m \times k$, where both are nonrandom, then $E(AZB) = AE(Z)B$ and $E(\text{tr}(Z)) = \text{tr}(E(Z))$.

¹To see why it is defined in this way, please refer to Magnus and Neudecker (2007).

Definition 3.2. If y is an $n \times 1$ random vector, its variance-covariance ma-

trix, denoted $Var(y)$, is defined as $E((y - E(y))(y - E(y))')$, i.e.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix}$$

where $\sigma_j^2 = Var(y_j)$ and $\sigma_{ij} = Cov(y_i, y_j)$.

In other words, the variance-covariance matrix has the variances of each element of y down its diagonal, with covariance terms in the off diagonals. Because $Cov(y_i, y_j) = Cov(y_j, y_i)$, it immediately follows that a variance-covariance matrix is symmetric.

If a is an $n \times 1$ nonrandom vector, then $Var(a'y) = a'[Var(y)]a \geq 0$. If $Var(a'y) > 0$ for all $a \neq 0$, $Var(y)$ is positive definite. If the elements of y are uncorrelated, $Var(y)$ is a diagonal matrix. If, in addition, $Var(y_j) = \sigma^2$ for $j = 1, 2, \dots, n$, then $Var(y) = \sigma^2 I_n$. If A is an $m \times n$ nonrandom matrix and b is an $n \times 1$ nonrandom vector, then $Var(Ay + b) = A[Var(y)]A'$.

4 Distributions of Random Vectors

Definition 4.1. If y is an $n \times 1$ **multivariate normal random vector** with mean μ and variance-covariance matrix Σ , we write $y \sim Normal(\mu, \Sigma)$.

Ay and By are independent if and only if $A\Sigma B' = 0$. (Why?) If $y \sim Normal(0, \sigma^2 I_n)$, A is a $k \times n$ nonrandom matrix, and B is an $n \times n$ symmetric, idempotent matrix, then Ay and $y'By$ are independent if and only if $AB = 0$; if A and B are nonrandom symmetric, idempotent matrices, then $y'Ay$ and $y'By$ are independent if and only if $AB = 0$.

If $u \sim \text{Normal}(0, I_n)$, then $u'u \sim \chi^2(n)$ (**chi-squared distribution with degrees of freedom n**); if A is an $n \times n$ symmetric, idempotent matrix with $\text{rank}(A) = q$, then $u'Au \sim \chi^2(q)$.

If $u \sim \text{Normal}(0, I_n)$, c is an $n \times 1$ nonrandom vector, A is a nonrandom $n \times n$ symmetric, idempotent matrix with rank q , and $Ac = 0$, then $\frac{c'u/\sqrt{(c'c)}}{\sqrt{u'Au/q}} \sim t_q$ (**student's t distribution with degrees of freedom q**).

If $u \sim \text{Normal}(0, I_n)$ and A and B are $n \times n$ nonrandom symmetric, idempotent matrices with $\text{rank}(A) = k_1$, $\text{rank}(B) = k_2$, and $AB = 0$, then $\frac{u'Au/k_1}{u'Bu/k_2} \sim F_{k_1, k_2}$ (**F distribution with degrees of freedom k_1 and k_2**).

References

- [1] W.H. Greene, *Econometric Analysis*, 7th ed., Pearson Education, 2011.
- [2] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3 ed., John Wiley & Sons Ltd., 2007.
- [3] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics, 2000.
- [4] J.M. Wooldridge, *Introductory Econometrics (a modern approach)*, 4th ed., South Western College, 2008.