Review of Matrix Algebra for Regression

A.Colin Cameron
Department of Economics
University of California - Davis
accameron@ucdavis.edu

May 8, 2008

Abstract

This provides a review of key matrix algebra / linear algebra results. The most essential results are given first. More complete results are given in e.g. Greene Appendix A.

Contents

1 Matrices and Vectors

Matrix: $\mathbf{A}$ is an $m \times n$ matrix with m rows and n columns

$$\underset{(m\times n)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{mn} & \cdots & \cdots & a_{mn} \end{bmatrix}$$

The typical element is a_{ij} in the i^{th} row and j^{th} column.
A 2×3 example is

$$\underset{(2\times 3)}{\mathbf{A}} = \begin{bmatrix} 1 & 2 & 0 \\ 4 & -1 & 2 \end{bmatrix}.$$

Column vector: Matrix with one column ($n = 1$)

$$\underset{(m\times 1)}{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

Row vector: Matrix with one row ($m = 1$)

$$\underset{(1\times n)}{\mathbf{a}} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

Vectors are often defined to be column vectors in econometrics.
In particular the parameter vector

$$\underset{(k\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

2 Types of matrices

Addition: Can add matrices that are of the same dimension. i.e. both are $m \times n$.
Then ij^{th} element of $\mathbf{A} + \mathbf{B}$ equals $\mathbf{A}_{ij} + \mathbf{B}_{ij}$.

$$\begin{bmatrix} 1 & 2 & 0 \\ 4 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 3 & 1 & 0 \\ 4 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 0 \\ 8 & 0 & 4 \end{bmatrix}.$$

Subtraction: Add minus the matrix.
Multiplication: Can multiply $\mathbf{A} \times \mathbf{B}$ if

$$\text{Number of columns in } \mathbf{A} = \text{ Number of rows in } \mathbf{B}.$$

If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$ then $\mathbf{A} \times \mathbf{B}$ is $m \times p$.
The ij^{th} element of $\mathbf{A} \times \mathbf{B}$ is the inner product of the i^{th} row of $\mathbf{A}$ and the j^{th} column of $\mathbf{B}$.

$$\{\mathbf{A} \times \mathbf{B}\}_{ij} = \sum_{k=1}^{p} a_{ik} b_{ik}.$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} 1 \times 7 + 2 \times 8 + 3 \times 9 & 1 \times 10 + 2 \times 11 + 3 \times 12 \\ 4 \times 7 + 5 \times 8 + 6 \times 9 & 4 \times 10 + 5 \times 11 + 6 \times 12 \end{bmatrix}$$

$$= \begin{bmatrix} 7 + 16 + 27 & 10 + 22 + 36 \\ 28 + 40 + 54 & 40 + 55 + 72 \end{bmatrix} = \begin{bmatrix} 50 & 68 \\ 122 & 167 \end{bmatrix}.$$

Division: Does not exist. Instead multiply by the inverse of the matrix.
Transpose: Converts rows of matrix into columns. Denoted $\mathbf{A}^T$ or $\mathbf{A}'$.
If $\mathbf{A}$ is $m \times n$ with ij entry a_{ij} then $\mathbf{A}'$ is $n \times m$ with ij entry a_{ji}.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

If $\mathbf{AB}$ exists then

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'.$$

3 Operators

Square matrix: $\mathbf{A}$ is $n \times n$ (same number of rows as columns).
Diagonal matrix: Square matrix with all off-diagonal terms equal to 0.
Block-diagonal matrix: Square matrix with off-diagonal blocks equal to $\mathbf{0}$. e.g. $\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$.
Identity matrix: Diagonal matrix with diagonal terms equal to 1.

$$\underset{(n \times n)}{\mathbf{I}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix}.$$

Then $\mathbf{A} \times \mathbf{I} = \mathbf{A}$ and $\mathbf{I} \times \mathbf{B} = \mathbf{B}$ for conformable matrices $\mathbf{A}$ and $\mathbf{B}$.
Orthogonal matrix: Square matrix such that $\mathbf{A}'\mathbf{A} = \mathbf{I}$.
Idempotent matrix: Square matrix such that $\mathbf{A} \times \mathbf{A} = \mathbf{A}$.
Positive definite matrix: Square matrix $\mathbf{A}$ such that $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all vector $\mathbf{x} \neq \mathbf{0}$.
Nonsingular matrix: Square matrix $\mathbf{A}$ with inverse that exists (also called full rank matrix).

4 Inverse of 2x2 matrix

Matrix inverse: Inverse $\mathbf{A}^{-1}$ of the square matrix $\mathbf{A}$ is a matrix such that

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

For a 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Example:

$$\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}^{-1} = \frac{1}{4 \times 3 - 2 \times 1} \begin{bmatrix} 3 & -2 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 0.3 & -0.2 \\ -0.1 & 0.4 \end{bmatrix}.$$

Check:

$$\begin{aligned} \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0.3 & -0.2 \\ -0.1 & 0.4 \end{bmatrix} &= \begin{bmatrix} 4 \times 0.3 + 2 \times (-0.1) & 4 \times (-0.2) + 2 \times 0.4 \\ 1 \times 0.3 + 2 \times (-0.1) & 1 \times (-0.2) + 3 \times 0.4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

5 Determinant

For inversion of larger matrices we first introduce determinant.
The determinant is important as inverse of matrix exists only if $|\mathbf{A}| \neq 0$.
Determinant: $|\mathbf{A}|$ or $\det \mathbf{A}$ is a scalar measure of a square $n \times n$ matrix $\mathbf{A}$ that can be computed in the following recursive way.

$$|\mathbf{A}| = a_{i1}c_{i1} + a_{i2}c_{i2} + \cdots + a_{in}c_{in} \text{ (for any choice of row } i\text{)}$$

where c_{ij} are the **cofactors:**

$$\begin{aligned} a_{ij} &= ij^{th} \text{ element of } \mathbf{A} \\ c_{ij} &= ij^{th} \textbf{ cofactor} \text{ of } \mathbf{A} \\ &= (-1)^{i+j}|\mathbf{A}_{ij}| \\ |\mathbf{A}_{ij}| &= \text{ minor of } \mathbf{A} \end{aligned}$$

Minor: The minor of **A** is the determinant of $(n-1) \times (n-1)$ matrix formed by deleting the i^{th} row and j^{th} column of **A**.
Determinant of 2×2 matrix example:

$$\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} = 5 \times (-1)^{1+1} \times 10 + 6 \times (-1)^{1+2} \times 8 = 50 - 48 = 2.$$

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a \times (-1)^{1+1} \times d + b \times (-1)^{1+2} \times c = ad - bc.$$

Determinant of 3×3 matrix example:

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} &= 1 \times (-1)^{1+1} \times \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 2 \times (-1)^{1+2} \times \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} \\ &\quad +3 \times (-1)^{1+3} \times \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 1 \times (50 - 48) - 2 \times (40 - 42) + 3 \times (32 - 35) \\ &= 2 + 4 - 9 \\ &= -3. \end{aligned}$$

6 Inverse

Matrix inverse: Inverse $\mathbf{A}^{-1}$ of the square matrix **A** is a matrix such that

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

The inverse is the transpose of the matrix of cofactors divided by the determinant.

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix} \text{ where } c_{ij} \text{ are the cofactors.}$$

Inverse of $n \times n$ matrix $\mathbf{A}$ exists if and only if any of the following

$$\text{rank}(\mathbf{A}) = n$$
$$\mathbf{A} \text{ is nonsingular}$$
$$|\mathbf{A}| \neq \mathbf{0}$$

7 Rank of a matrix

Rank: Consider $m \times n$ matrix $\mathbf{A}$ that is not necessarily square.

$$\begin{aligned} rank(\mathbf{A}) &= \text{maximum number of linearly independent rows} \\ &= \text{maximum number of linearly independent columns} \\ &\leq \min(m, n) \end{aligned}$$

Let

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \cdots \mathbf{a}_n] \text{ where } \mathbf{a}_i \text{ is the } i^{th} \text{ column of } \mathbf{A}$$

Then if the only solution to

$$\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + \cdots + \lambda_n \mathbf{a}_n = \mathbf{0}$$

is $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$ then $\mathbf{a}_1$, $\mathbf{a}_2, ..., \mathbf{a}_n$ are linearly independent.
If at least one λ is nonzero then they are linearly dependent.

This is important because if rank$(\mathbf{A}) = n$ and $\mathbf{x}$ is $n \times 1$ then
(1) the system of equations $\mathbf{Ax} = \mathbf{b}$ has a unique nonzero solution for $\mathbf{x}$.
(2) the system of equations $\mathbf{Ax} = \mathbf{0}$ has no solution for $\mathbf{x}$ other than $\mathbf{x} = \mathbf{0}$.

In particular for ordinary least squares the estiamting equations are

$$\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}.$$

To solve for $k \times 1$ vector $\widehat{\boldsymbol{\beta}}$ need rank$(\mathbf{X}'\mathbf{X}) = k$ which in turn requires rank$(\mathbf{X}) = k$.

8 Positive definite matrices

Quadratic form: The scalar $\mathbf{x}'\mathbf{A}\mathbf{x}$ based on a symmetric matrix.

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n \\ &\quad + a_{22}x_1^2 + 2a_{23}x_2x_3 + \cdots + 2a_{2n}x_2x_n \\ &\quad + \cdots + a_{nn}x_n^2.\end{aligned}$$

Positive definite matrix: $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
Positive semidefinite matrix: $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$.

Variance matrix: The variance matrix of a vector random variable is always positive semidefinite, and is positive definite if there is no linear dependence among the components of $\mathbf{x}$.

A useful propery is that if a matrix $\mathbf{A}$ is symmetric and positive definite a nonsingular matrix $\mathbf{P}$ exists such that

$$\mathbf{A} = \mathbf{P}\mathbf{P}'.$$

9 Matrix differentiation

This is initially for reference only. Use when obtain estimating equaitons.
There are rules for ways to store e.g. derivative of a vector with respect to a vector.
The starting point is that the derivative of a scalar with respect to a column vector is a column vector, and the derivative of a scalar with respect to a row vector is a row vector.

Let $\mathbf{b}$ be an $n \times 1$ column vector.

1. Differentiation of scalar wrt column vector.
Let $f(\mathbf{b})$ be a scalar function of $\mathbf{b}$.

$$\underset{(n\times 1)}{\frac{\partial \mathbf{f}(\mathbf{b})}{\partial \mathbf{b}}} = \begin{bmatrix} \frac{\partial \mathbf{f}(\mathbf{b})}{\partial b_1} \\ \vdots \\ \frac{\partial \mathbf{f}(\mathbf{b})}{\partial b_n} \end{bmatrix}$$

2. Differentiation of row vector wrt column vector.
Let $\mathbf{f}(\mathbf{b})$ be a $m \times 1$ row function of $\mathbf{b}$.

$$\mathbf{f}(\mathbf{b}) = \begin{bmatrix} f_1(\mathbf{b}) & \cdots & f_m(\mathbf{b}) \end{bmatrix}$$

Then

$$\underset{(n\times m)}{\frac{\partial \mathbf{f}(\mathbf{b})}{\partial \mathbf{b}}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{b})}{\partial \mathbf{b}} & \cdots & \frac{\partial f_m(\mathbf{b})}{\partial \mathbf{b}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\mathbf{b})}{\partial b_1} & \cdots & \frac{\partial f_m(\mathbf{b})}{\partial b_1} \\ \vdots & & \vdots \\ \frac{\partial f_1(\mathbf{b})}{\partial b_n} & & \frac{\partial f_m(\mathbf{b})}{\partial b_n} \end{bmatrix}$$

3. Second derivative of scalar function $f(\mathbf{b})$ with respect to column vector.

$$\begin{aligned} \underset{(n\times n)}{\frac{\partial^2 f(\mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}'}} &= \frac{\partial}{\partial \mathbf{b}} \left(\frac{\partial f(\mathbf{b})}{\partial \mathbf{b}'} \right) \\ &= \frac{\partial}{\partial \mathbf{b}} \begin{bmatrix} \frac{\partial f(\mathbf{b})}{\partial b_1} & \cdots & \frac{\partial f(\mathbf{b})}{\partial b_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial \mathbf{b}} \left(\frac{\partial f(\mathbf{b})}{\partial b_1} \right) & \cdots & \frac{\partial}{\partial \mathbf{b}} \left(\frac{\partial f(\mathbf{b})}{\partial b_n} \right) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial^2 f(\mathbf{b})}{\partial b_1 \partial b_1} & \cdots & \frac{\partial f(\mathbf{b})}{\partial b_1 \partial b_n} \\ \vdots & & \vdots \\ \frac{\partial f(\mathbf{b})}{\partial b_n \partial b_1} & & \frac{\partial f(\mathbf{b})}{\partial b_n \partial b_n} \end{bmatrix} \end{aligned}$$

4. Chain rule for $f()$ scalar and $\mathbf{x}$ and $\mathbf{z}$ column vectors.
Suppose $f(\mathbf{x}) = f(g(\mathbf{x})) = f(\mathbf{z})$ where $\mathbf{z} = g(\mathbf{x})$. Then

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{z}'}{\partial \mathbf{x}} \times \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}}.$$