Matrix Algebra

William Brasic

The University of Arizona

What is Matrix Algebra?

Definition 1: Matrix Algebra

In its simplest form, matrix algebra is a convenient way to express linear equations in terms of vectors and matrices.

- It greatly simplifies the math and notation for the rest of the semester so we will cover the main topics.

Main Terms

The main terms in matrix algebra are:

- Dimensions
- Column Vector
- Row Vector
- Transpose
- Dot Product
- Matrix
- Square Matrix
- Symmetric Matrix
- Identity Matrix
- Inverse

Column Vector

Definition 2: Column Vector

A column vector $\boldsymbol{x}$ of numbers $x_1, \ldots, x_n$ is given by

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

- We say this is a $n \times 1$ dimensional column vector.
- Unless otherwise noted, a vector in this course is a column vector.

Row Vector

Definition 3: Row Vector

A row vector $\boldsymbol{x}$ of numbers $x_1, \ldots, x_n$ is given by

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}.$$

- We say this is a $1 \times n$ dimensional row vector.

Transpose

Definition 4: Transpose

The transpose of a column vector $\boldsymbol{x}$ of numbers $x_1, \ldots, x_n$ is given by

$$\boldsymbol{x}' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}' = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}.$$

- The transpose of a column vector is a row vector and vice versa.

Dot Product

Definition 5: Dot Product

The dot product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by

$$\boldsymbol{x}'\boldsymbol{y} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i.$$

- To perform a dot product, the vectors must be of the same dimension.

Dot Product

Question 1: Dot Product

$$\boldsymbol{x}'\boldsymbol{\beta} = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = ?$$

Dot Product

Question 1: Dot Product

$$\boldsymbol{x}'\boldsymbol{\beta} = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = ?$$

Answer to Question 1

$$\boldsymbol{x}'\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = \beta_0 + \sum_{i=1}^{2} \beta_i x_i.$$

Matrix

Definition 6: Matrix

A $n \times k$ matrix X is a rectangular array of numbers given by

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}.$$

- The n rows typically correspond to observations.
- The k columns typically correspond to variables.
- We say this is a $n \times k$ dimensional matrix.

Square Matrix

Definition 7: Square Matrix

A $n \times k$ matrix X of numbers is square when it has the same number of rows as columns (so $n = k$):

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{pmatrix}.$$

Matrix Addition and Scalar Multiplication

Example 1: Matrix Addition and Scalar Multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} = 2 * \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

- We can add two matrices X and Y, $X + Y$, if they have the same number of rows and columns (same dimension).

Matrix Multiplication

Example 2: Matrix Multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (1*1)+(2*3) & (1*2)+(2*4) \\ (3*1)+(4*3) & (3*2)+(4*4) \end{bmatrix}$$
$$= \begin{bmatrix} 1+6 & 2+8 \\ 3+12 & 6+16 \end{bmatrix}$$
$$= \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix}$$

- We can multiply two matrices X and Y, XY, if the number of columns of X equals the number of rows of Y.
  - If X is 4×5 and Y is 4×2, then $X'Y$ exists and is 5×2.

Matrix-Vector Multiplication

Example 3: Matrix-Vector Multiplication

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} (1*1)+(2*2) \\ (3*1)+(4*2) \end{bmatrix} = \begin{bmatrix} 1+4 \\ 3+8 \end{bmatrix} = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$$

- We can multiply a matrix X with a vector $\boldsymbol{y}$, $X\boldsymbol{y}$, if the number of columns of X equals the dimension of $\boldsymbol{y}$.
  - If X is $n \times k$ and $\boldsymbol{y}$ is $n \times 1$, then $X'\boldsymbol{y}$ is $k \times 1$.

Symmetric Matrix

Definition 8: Symmetric Matrix

A $n \times n$ square matrix X of numbers is symmetric if $X = X'$.

Example 4: Symmetric Matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 6 & 7 \\ 3 & 6 & 8 & 9 \\ 4 & 7 & 9 & 10 \end{bmatrix}$$

Transpose

Theorem 1: Matrix Multiplied by its Transpose is Symmetric

For any $n \times k$ matrix X of numbers, the resulting matrix $Y = X'X$ and its inverse is symmetric.

Theorem 2: Transpose of a Product

If X is $n \times k$ and $\boldsymbol{\beta}$ is $k \times 1$, then the transpose of $X\boldsymbol{\beta}$ is

$$(X\boldsymbol{\beta})' = \boldsymbol{\beta}'X'.$$

Identity Matrix

Definition 9: Identity Matrix

A $n \times n$ square matrix I_n of numbers is called the identity matrix if

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Identity Matrix

Property 1: Multiplication with Identity Matrices

Any matrix or vector multiplied by the identity matrix returns the original matrix or vector.

Full Rank (Non-Singular) Matrix

Property 2: Full Rank (Non-Singular) Matrix

If X is a $n \times k$ matrix with full rank, then its columns and rows are linearly independent of each other, i.e., no column can be expressed as a linear combination of other columns and no row can be expressed as a linear combination of other rows.

- In econometrics, we are typically concerned with X having full column rank so no regressor can be expressed as a linear combination of other regressors.
- When X has full rank, this means that $X'X$ is invertible, ensuring the OLS solution exists and is unique.

Inverse

Definition 10: Inverse

The inverse of an $n \times n$ square matrix X is denoted by X^{-1} and is defined when X has n linearly independent columns and rows. When X^{-1} exists, it satisfies $XX^{-1} = X^{-1}X = I_n$, where I_n is the $n \times n$ identity matrix.

- We think about the inverse in the same way as division between two scalars.

Inverse

Property 3: Inverse

When X and Y are invertible,

1. $\left(X^{-1}\right)^{-1} = X$.

Inverse

Property 3: Inverse

When X and Y are invertible,

1. $(X^{-1})^{-1} = X$.
2. $(cX)^{-1} = c^{-1}X^{-1}$ for any constant $c \neq 0$.

Inverse

Property 3: Inverse

When X and Y are invertible,

1. $\left(X^{-1}\right)^{-1} = X$.
2. $(cX)^{-1} = c^{-1}X^{-1}$ for any constant $c \neq 0$.
3. $(X')^{-1} = \left(X^{-1}\right)'$.

Inverse

Property 3: Inverse

When X and Y are invertible,

1. $\left(X^{-1}\right)^{-1} = X$.
2. $(cX)^{-1} = c^{-1}X^{-1}$ for any constant $c \neq 0$.
3. $(X')^{-1} = \left(X^{-1}\right)'$.
4. $(XY)^{-1} = Y^{-1}X^{-1}$.

Random Vectors

Definition 11: Random Vector

A random vector $\boldsymbol{u}$ of dimension $n \times 1$ is a vector that contains random variables $u_1, \ldots, u_n$.

Variance of a Random Vector

Property 4: Variance of a Random Vector

If $\boldsymbol{u}$ is a $n \times 1$ random vector with independent elements each having variance σ^2 and X is a $n \times k$ matrix, then

1. $\mathbb{V}[\boldsymbol{u}] = I_n\sigma^2$ is a $n \times n$ diagonal matrix with σ^2 along the diagonal.

Variance of a Random Vector

Property 4: Variance of a Random Vector

If $\boldsymbol{u}$ is a $n \times 1$ random vector with independent elements each having variance σ^2 and X is a $n \times k$ matrix, then

1. $\mathbb{V}[\boldsymbol{u}] = I_n \sigma^2$ is a $n \times n$ diagonal matrix with σ^2 along the diagonal.
2. $\mathbb{V}\left[X' \boldsymbol{u}\right] = X' \mathbb{V}[\boldsymbol{u}] X = X' I_n \sigma^2 X = \sigma^2 X' X$.

Vector and Matrix Calculus

Definition 12: Derivative of a Vector

The derivative of a function $f(\boldsymbol{x})$ where $\boldsymbol{x}$ is a $n \times 1$ vector is

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}.$$

- This is commonly called the gradient of f.

Vector and Matrix Calculus

Property 5: Matrix Differentiation

For $n \times 1$ vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and a matrix A,

1. $\frac{\partial \boldsymbol{y}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}'\boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{y}$.

Vector and Matrix Calculus

Property 5: Matrix Differentiation

For $n \times 1$ vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and a matrix A,

1. $\frac{\partial \boldsymbol{y}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}'\boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{y}$.
2. When A is not symmetric,

$$\frac{\partial \boldsymbol{x}' A \boldsymbol{x}}{\partial \boldsymbol{x}} = (A + A')\boldsymbol{x}.$$

Vector and Matrix Calculus

Property 5: Matrix Differentiation

For $n \times 1$ vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and a matrix A,

1. $\frac{\partial \boldsymbol{y}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}'\boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{y}$.
2. When A is not symmetric,
$$\frac{\partial \boldsymbol{x}'A\boldsymbol{x}}{\partial \boldsymbol{x}} = (A + A')\boldsymbol{x}.$$
3. When A is symmetric,
$$\frac{\partial \boldsymbol{x}'A\boldsymbol{x}}{\partial \boldsymbol{x}} = (A + A')\boldsymbol{x} = (A + A)\boldsymbol{x} = 2A\boldsymbol{x}.$$

Why Do We Care?

Question 2: Why Do We Care?

Why go through all the trouble of this matrix algebra stuff?

Why Do We Care?

Question 2: Why Do We Care?

Why go through all the trouble of this matrix algebra stuff?

Answer to Question 2

We can represent any OLS specification conveniently in terms of vectors and matrices.

Linear Regression Model in Vector Form

Definition 13: Vector Representation of the Linear Model

Given $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i$ for $i = 1, \ldots n$, we can write this as

$$y_i = \boldsymbol{x_i'}\boldsymbol{\beta} + u_i.$$

- y_i is agent i's outcome.
- $\boldsymbol{x_i}$ is the $(k+1) \times 1$ vector of covariates corresponding to agent i.
- $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of parameters.
- u_i is the error corresponding to agent i.

Linear Regression Model in Matrix Form

Definition 14: Matrix Representation of the Linear Model

We can write $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i$ for $i = 1, \ldots n$, as $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u}$,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

- $\boldsymbol{y}$ is the $n \times 1$ vector of outcomes for each agent.
- X is the $n \times (k+1)$ vector of covariates for each agent.
- $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of parameters.
- $\boldsymbol{u}$ is the $n \times 1$ vector of errors for each agent.

SSR in Vector Form

Definition 15: SSR in Vector Form

The sum of squared residuals (SSR) in vector notation is

$$\sum_{i=1}^{n} \widehat{u}_i^2 = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}} = \left(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}\right)$$

SSR in Vector Form

Definition 15: SSR in Vector Form

The sum of squared residuals (SSR) in vector notation is

$$\sum_{i=1}^{n} \widehat{u}_i^2 = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}} = \left(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}\right)$$
$$= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'X\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'X'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}}$$

SSR in Vector Form

Definition 15: SSR in Vector Form

The sum of squared residuals (SSR) in vector notation is

$$\begin{aligned}\sum_{i=1}^{n}\widehat{u}_i^2 = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}} &= \left(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}\right) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'X\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'X'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\widehat{\boldsymbol{\beta}}'X'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}}.\end{aligned}$$

- We can now derive the OLS estimator without using summations! So, lets do it!

The OLS Estimator in Matrix Form

Theorem 3: OLS Solution

The solution, $\widehat{\boldsymbol{\beta}}$, to the OLS problem is given by

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}.$$

- $X'\boldsymbol{y}$ is analogous to $\text{Cov}(x, y)$.
- $X'X$ is analogous to $\mathbb{V}(x)$.
- The inverse operator is analogous to division.

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 1

First, find the derivative of the SSR:

$$\begin{aligned}\frac{\partial SSR\left(\widehat{\boldsymbol{\beta}}\right)}{\partial \widehat{\boldsymbol{\beta}}} &= \frac{\partial \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{\partial \widehat{\boldsymbol{\beta}}} \\ &= \frac{\partial}{\widehat{\boldsymbol{\beta}}}\left(\boldsymbol{y}'\boldsymbol{y} - 2\widehat{\boldsymbol{\beta}}'X'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'X'X\widehat{\boldsymbol{\beta}}\right) \\ &= -2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}}.\end{aligned}$$

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 2

Second, use the first order condition for minimization:

$$\frac{\partial SSR\left(\widehat{\boldsymbol{\beta}}\right)}{\partial \widehat{\boldsymbol{\beta}}} = -2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}} = 0$$

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 3

Lastly, solve the equation for $\widehat{\boldsymbol{\beta}}$:

$$-2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}} = 0 \iff 2X'X\widehat{\boldsymbol{\beta}} = 2X'\boldsymbol{y}$$

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 3

Lastly, solve the equation for $\widehat{\boldsymbol{\beta}}$:

$$
\begin{aligned}
-2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}} = 0 &\iff 2X'X\widehat{\boldsymbol{\beta}} = 2X'\boldsymbol{y} \\
&\iff X'X\widehat{\boldsymbol{\beta}} = X'\boldsymbol{y}
\end{aligned}
$$

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 3

Lastly, solve the equation for $\widehat{\boldsymbol{\beta}}$:

$$\begin{aligned} -2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}} = 0 &\iff 2X'X\widehat{\boldsymbol{\beta}} = 2X'\boldsymbol{y} \\ &\iff X'X\widehat{\boldsymbol{\beta}} = X'\boldsymbol{y} \\ &\iff (X'X)^{-1}X'X\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \end{aligned}$$

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 3

Lastly, solve the equation for $\widehat{\boldsymbol{\beta}}$:

$$\begin{aligned} -2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}} = 0 &\iff 2X'X\widehat{\boldsymbol{\beta}} = 2X'\boldsymbol{y} \\ &\iff X'X\widehat{\boldsymbol{\beta}} = X'\boldsymbol{y} \\ &\iff (X'X)^{-1}X'X\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \\ &\iff I_{k+1}\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \end{aligned}$$

Proof of the OLS Solution

Proof 1: Proof of OLS Solution Part 3

Lastly, solve the equation for $\widehat{\boldsymbol{\beta}}$:

$$\begin{aligned} -2X'\boldsymbol{y} + 2X'X\widehat{\boldsymbol{\beta}} = 0 &\iff 2X'X\widehat{\boldsymbol{\beta}} = 2X'\boldsymbol{y} \\ &\iff X'X\widehat{\boldsymbol{\beta}} = X'\boldsymbol{y} \\ &\iff (X'X)^{-1}X'X\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \\ &\iff I_{k+1}\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \\ &\iff \widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}. \quad \square \end{aligned}$$

Hooray!

Thank You!