

A quick overview of linear algebra for econometrics

Rob Letzler

January 19, 2007

Thanks to Professor Jill Pipher for teaching me linear algebra and to John Leen for providing useful input on this talk. This document is prepared in \LaTeX which I learned from Leslie Lamport's *LaTeX: A Document Preparation System* book.

1 Background: Linear Algebra

Linear Algebra:

- Is Useful. It was developed to solve a practical problems.
- Provides natural, convenient tools to write down and characterize operations involving adding or multiplying lots of numbers.
- Is not difficult. There are no deep new ideas here.
- Involves a considerable amount of mechanics, jargon, and notation.

This document provides a quick overview of linear algebra. There are plenty of linear algebra textbooks and appendices in econometrics books that provide more details; show you how to do proofs involving linear algebra; provide plenty of practice problems; and present a lot of topics that you do not really need to do the kind of statistics that social science grad students typically deal with. (I can give you the lay of the land in 90 minutes; but you have to practice before it will get into your bloodstream.) Howard Anton's *Elementary Linear Algebra* books are particularly good. William Greene's *Econometric Analysis* textbook has an appendix that covers linear algebra. It is very terse and provides few examples and little context about why things are important.

2 Three and a Half Motivating Questions

2.1 Motivation: Regression Analysis

Suppose we want to use a dataset to analyze how public action affects outcomes.

For example, we might be wondering whether to give unemployed people more training and want to estimate the relationship between the number of job offers the person gets (y_i) and the person's training and age:

$$y_i = \alpha + \beta_1 \times (\text{number of hours of training taken}_i) + \beta_2 \times (\text{age}_i) + \epsilon_i$$

Similarly, we might wonder whether to require safety equipment on cars and want to estimate the relationship between the probability of driver injury y_i and a characteristic of the car:

$$y_i = \alpha + \beta_1 \times (\text{number of airbags on car}_i) + \epsilon_i$$

In other words, this is an relationship between

- a list of outcomes, y_i , with one entry for each person i ,
- data, $x_{i,j}$, listing j variables for each person i and
- the list of coefficients, β_j , for each of the j variables that we are trying to solve for,

Linear algebra lets us express this relationship as: $\vec{y} = \mathbf{X}\vec{\beta}$ where \vec{y} and $\vec{\beta}$ are **vectors** (1-dimensional lists of numbers) and \mathbf{X} is a **matrix** (a two-dimensional array of numbers). In linear algebra, we call a single number, like 5, a **scalar**.

(Throughout this document, I will write vectors in lower case with arrows above them, and matrices in uppercase bold, and scalars in lower case. Often I use an (optional) dot (\cdot) to indicate multiplication. Books vary slightly in the notation they use for vectors and matrices.)

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix}$$

Notice that we write an entry in matrix \mathbf{X} as $x_{\text{column}\#, \text{row}\#}$ with the first subscript being the column.

2.2 Motivation: Solving Systems of Equations

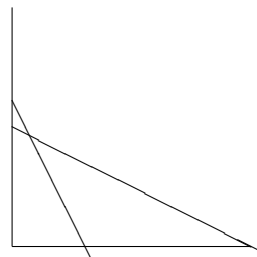
Suppose we want to solve the system of equations:

$$x + 2y = 4$$

$$2x + y = 5$$

(How would you solve this? The matrix approach to solving it – called “Gaussian Elimination” that yields a “triangular matrix” – is almost exactly the same as the grade school algebra approach of adding and subtracting the equations from each other in order to eliminate all but one variable from one equation.)

Graphically, we want to know where the lines intersect:



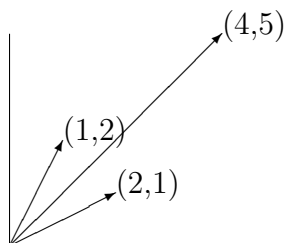
We can write the system of equations as matrices and vectors:

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

2.3 Motivation: combining two vectors into a third vector

We can express the same question as above as the challenge of finding scaling factors x and y that stretch or shrink two vectors so that their sum is equal to a third vector:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} 2 \\ 1 \end{bmatrix} y = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$



Thinking of matrix multiplication as an operation that can stretch, rotate, and potentially flatten vectors will be a very natural way to understand whether a matrix is “well behaved” – which I talk about below.

2.4 Motivation: Can we do basic math on matrices and vectors?

- Can we “add” vectors? Can we “add” matrices? Answer: yes, if two matrices (vectors) have are of the same size, addition works by adding the corresponding items in the “obvious way. If the vectors or matrices are different sizes, they cannot be added. Here is an example:

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

- Can we multiply matrices and vectors by scalars. Answer: Yes, it happens in the “obvious” way: multiply every entry in the vector or matrix by the scalar value.

$$2 \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 6 \\ 10 \end{bmatrix}$$

- Can we “multiply” matrices and vectors? What are the properties of matrix multiplication? Answer: Yes. The rest of this document discusses matrix multiplication and its pitfalls.

3 Matrix Multiplication

Here is a simple example of matrix multiplication:

- Notice that we multiply row times column – in other words, we multiply rows in the first matrix times the columns in the second matrix.
- Each multiplication of a row times a column maps to a scalar in the result. Specifically, we multiply the first (nth) entry in the row times the first (nth) entry in the column, which yields n pairs of products – which we then add up. The example below illustrates this:

$$\begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \times 1 + 4 \times 2 \\ 5 \times 1 + 6 \times 2 \end{bmatrix} = \begin{bmatrix} 11 \\ 17 \end{bmatrix}$$

As we move right (down) in the columns of the second (first) matrix, we move right (down) in the results matrix. Here is a simple, abstract example involving multiplying a matrix consisting of two rows – r_1 and r_2 – by a matrix containing two columns c_1 and c_2 .

$$\begin{bmatrix} - & r_1 & - \\ - & r_2 & - \end{bmatrix} \cdot \begin{bmatrix} | & | \\ c_1 & c_2 \\ | & | \end{bmatrix} = \begin{bmatrix} r_1 \cdot c_1 & r_1 \cdot c_2 \\ r_2 \cdot c_1 & r_2 \cdot c_2 \end{bmatrix}$$

We can only multiply two matrices if the number of columns in the first matrix matches the number of rows in the second matrix

3.1 Sometimes “transposing” the matrix, which makes rows into columns, enables us to multiply

An OLS regression seeks to minimize the the sum of the squared errors in its prediction. It is easy to write down the errors as a vector:

$$\vec{e} = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

And we’d like to be able to multiply this vector by itself to get the sum of the squared errors ($\sum e_i^2$), but just multiplying the vector times itself does not work because the vectors are of the wrong sizes – we cannot match the single entry in each row to n entries in each column:

$$\vec{e} \cdot \vec{e} = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

Solution: Taking the transpose of the error vector, denoted either \vec{e} or \vec{e}^T , solves the problem and yields:

$$\vec{e}^T \vec{e} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix} = \sum e_i^2$$

Taking the transpose:

- exchanges the row and column index, so entry x_{ij} in matrix \mathbf{X} is entry x_{ji} in \mathbf{X}'
- reflects items across the diagonal in a square matrix. (The diagonal that runs from the top left to the bottom right is very important in linear algebra. When we talk about a diagonal, we mean that one – we generally ignore the other diagonal.)
- converts column vectors into row vectors – and row vectors into column vectors.

If $\mathbf{A} = \mathbf{A}'$ we say that \mathbf{A} is symmetric.

3.2 Order Matters in Matrix Multiplication: $\mathbf{AB} \neq \mathbf{BA}$

Multiplying matrices is different from multiplying individual numbers because order matters.

Consider the following multiplications that take the same two matrices, \mathbf{A} and \mathbf{B} , but get different results depending on their order.

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix}$$

The discussion above looked at: $\mathbf{X}\vec{b}$ which is very well behaved. We cannot calculate the product of the matrices in the opposite order, $\vec{b}\mathbf{X}$, since the two items are the wrong shape to be multiplied together:

$$\begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

It is, however, always true that $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

4 (Multiplicative) Inverses

Econometrics is the story of the valiant search for the correct (vector of) coefficients, $\vec{\beta}$, that show how X affects Y. There's a dramatic moment in a matrix-based presentation of

econometrics where we have almost got the vector $\vec{\beta}$ that we want. The story so far shows that getting the best estimate implies that:¹

$$\mathbf{A}\vec{\beta} = \mathbf{X}'\vec{y}$$

So we can discover the $\vec{\beta}$ vector if we can just get rid of that pesky \mathbf{A} matrix.

Your algebra 1 instincts are (almost) right here about what we need to do. If \mathbf{A} were a single number, we would divide both sides by \mathbf{A} – or equivalently – we would multiply both sides of the equation by \mathbf{A} 's multiplicative inverse, $\frac{1}{\mathbf{A}}$. With a matrix, we will do the same thing – we will multiply both sides of the equation by the multiplicative inverse of matrix \mathbf{A} , which we denote as \mathbf{A}^{-1} and call “A-inverse” when we talk.

Notice that multiplying scalar A by its multiplicative inverse, $\frac{1}{A}$, yields the multiplicative identity, 1: $\frac{1}{A} \times A = 1$. Multiplying a matrix by its inverse yields the identity matrix², \mathbf{I} : $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. While in general, order matters in matrix multiplication, it turns out that the left and right inverses are the same, so $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

5 Singular matrices destroy information and are not invertible

Consider the following regression:³

$$\text{earnings} = \beta_1 \times \text{years of education} + \beta_2 \times \text{semesters of education} + \epsilon$$

Suppose that each year of education is exactly two semesters of education. This yields an example data matrix \mathbf{X} and outcome vector \vec{y} as follows:

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 1 & 2 \\ 5 & 10 \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} 6 \\ 2 \\ 10 \end{bmatrix}$$

¹For purposes of keeping the presentation simple, I gloss over the fact that the \mathbf{A} matrix is actually the matrix we get by transposing the data matrix and multiplying by itself: $(\mathbf{X}'\mathbf{X}) = \mathbf{A}$

²The identity matrix has ones on the diagonal that runs from the top left to the bottom right. All its other entries are zero. Here is a two dimensional identity matrix:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

³Notice that this regression asks an impossible-to-answer question: it demands that we separate the effects of education from the effects of education measured in exactly the same way. If you were to feed it to Stata, Stata would drop one of the two education variables before attempting to answer.

We want to solve for the $\vec{\beta}$ in the equations of the form $\mathbf{X}\vec{\beta} = \vec{y}$ below.⁴ The following calculations show that there are two $\vec{\beta}$ vectors that solve $\mathbf{X}\vec{\beta} = \vec{y}$ equally well.⁵

$$\begin{bmatrix} 3 & 6 \\ 1 & 2 \\ 5 & 10 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 6 \\ 1 & 2 \\ 5 & 10 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 10 \end{bmatrix}$$

This \mathbf{X} matrix is singular. Multiplying by a singular matrix destroys information (since we can multiply any two dimensional vector – which can represent any point in two dimensional space and will get points on a single plane of the form $y = \frac{1}{3}x = \frac{1}{5}Z$). It maps many vectors to the same \vec{y} vector. Thus, we cannot tell which of the many ways to get to this result we actually took so this matrix has no inverse.

If one column of the matrix is a linear combination (weighted sum) of other columns in the matrix, it is not “linearly independent” of them. (In our singular matrix in this section, the right column is exactly twice the left column.) That means we can capture an equal amount of information in fewer columns and that the matrix is singular. The **rank** of the matrix is the number of linearly independent rows. A matrix is said to have **full rank** if all of its columns are linearly independent. The matrix in the example above is not of full rank. A square matrix of full rank is invertible. All other matrices are singular.

5.1 A synopsis of some major results about singular and invertible matrices

A linear algebra book spends a great deal of time exploring the implications of being a good or bad matrix. Here are some of the punchlines. Some but not all of these results have direct implications for econometrics.

⁴I have simplified this example by presenting an example where we can find $\vec{\beta}$ so that the regression line passes through every data point exactly and all the residuals are zero. If this were a real econometrics problem where there is no perfect regression line, the fact that the data matrix is not of full rank, so $\mathbf{X}'\mathbf{X}$ is singular, means that there would be more than one regression line that minimized the residuals.

⁵There are in fact an infinite family of $\vec{\beta}$'s that solve this set of equations; which is a real problem when econometrics aspires to provide a single, unique solution.

Invertible (Good) Matrices	Singular (Bad) Matrices
Multiplying a vector by an invertible matrix rotates and stretches the vector	Multiplying a vector by an singular matrix rotates, stretches, and flattens the vector. In other words, the multiplication destroys information.
Multiplication maps each input vector (matrix) to a unique output vector (matrix)	Multiplication maps many different inputs to each output
A unique inverse \mathbf{A}^{-1} exists. Multiplying by \mathbf{A}^{-1} undoes the effects of multiplying by \mathbf{A} .	Has no inverse.
is square (full rank rectangular matrices are somewhat well behaved; and if \mathbf{A} is a full rank rectangular matrix, then $\mathbf{A}'\mathbf{A}$ is a well behaved invertible square matrix).	square or rectangular
Maps only the zero vector ($\vec{0}$) to $\vec{0}$	Maps an infinite family of non-zero vectors to $\vec{0}$. This family is called its null space.
is a basis for R^N and spans R^N (i.e. for every $\vec{Y} \in R^N$ there exists an $\vec{x} \in R^N$ such that $\mathbf{A}\vec{x} = \vec{y}$)	is NOT a basis for R^N ; spans some space, but it's smaller than R^N .
is of full rank; i.e. there is no way to write down a matrix that contains fewer vectors in its narrowest dimension but contains as much information	is not of full rank; at least one of the vectors in the matrix can be expressed as a linear combination of the others.
has N (potentially imaginary) nonzero eigenvalues	at least one eigenvalue is zero. the corresponding eigenvector is in its null space.

6 Natural Connections Between Econometrics and Linear Algebra

- A database of right hand side variables with one entry per observation can be a matrix. We typically call it the \mathbf{X} matrix.
- Vectors are natural ways to write down the outcomes \vec{Y} and the coefficients we are solving for, $\vec{\beta}$
- The sum of squares is just the dot product of two vectors: $\sum x_i y_i = \vec{x}' \cdot \vec{y}$ and $\sum x_i x_i = \vec{x}' \cdot \vec{x}$

- Further, we can compute the OLS $\vec{\beta}$ through matrix multiplication: $\vec{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{y}$

6.1 Symmetric idempotent matrices have nice properties and are useful statistically

A matrix \mathbf{A} is **idempotent** if multiplying it by itself does not affect its value, i.e. $\mathbf{A}\mathbf{A} = \mathbf{A}^2 = \mathbf{A}$

We can calculate the standard deviation of a vector, \vec{x} , using the following matrix multiplication:

$$\vec{x}' \begin{bmatrix} \frac{n-1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots \\ \frac{1}{n} & \frac{n-1}{n} & \frac{1}{n} & \cdots \\ \frac{1}{n} & \frac{1}{n} & \frac{n-1}{n} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \vec{x}$$

We can get a vector containing the average value in \vec{x} through the following multiplication:

$$\vec{x}' \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \vec{x}$$

- These idempotent matrices are singular.
- Their rank is the sum of the diagonal