W4912
Spring 2003

Section Handout #1 - Linear Algebra Review

Notation

$A_{n,m}$ is an n×m matrix, i.e. a matrix with *n* rows and *m* columns, such that:

$$A_{n,m} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & & & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

x_n is an n-dimensional ***column*** vector, i.e. a matrix with one column, such that:

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

Note that we'll usually use capital letters to denote matrices, and lower case letters to denote vectors.

Matrix Operations

- Transpose: $B_{m,n}=(A_{n,m})'$ such that $b_{ij} = a_{ji}$
- Adding: $C_{n,m} = A_{n,m} + B_{n,m}$ such that $c_{ij} = a_{ij} + b_{ij}$

  Note that we can add two matrices only if they have the same dimensions.

- Multiplying by scalar: $B_{n,m} = cA_{n,m}$ such that $b_{ij} = ca_{ij}$
- Multiplying: $C_{n,m} = A_{n,k}B_{k,m}$ such that $c_{ij} = \sum_{l=1}^{k} a_{il} b_{lj}$

  Note that: (1) We can multiply two matrices only if the number of columns in the first is the same as the number of rows in the second, (2) The order is important, so even if AB is a valid operation, it does not necessarily mean that BA is valid, and (3) If x is an n×1 vector then xx' is an n×n matrix while x'x is a scalar.

- Comparison: A = B if and only if $a_{ij} = b_{ij}$ for each i and j

Some Properties to Remember

$$(A+B)' = A' + B'$$

$$(A+B)+C = A+(B+C)$$

$$(AB)' = B'A'$$

$$(AB)C = A(BC)$$

$$A(B+C) = AB + AC \quad \text{or} \quad (A+B)C = AC + BC$$

Special Matrices

- Square matrix: a matrix with the number of rows equals the number of columns
- Symmetric matrix: the matrix A is symmetric iff A=A'. Note that for a matrix to be symmetric it must be a square matrix.
- Lower triangular matrix: a square matrix for which all the elements above the main diagonal equal to zero.
- Upper triangular matrix: a square matrix for which all the elements below the main diagonal equal to zero.
- Diagonal matrix: a matrix which is upper and lower triangular, so all off-diagonal elements are zeros.
- The identity matrix: a diagonal matrix with all elements on the main diagonal equal to one. We denote the matrix by I. Note that this matrix is the matrix analogue for the scalar 1, in the sense that for any matrix A, we have that: AI = IA = A.
- The zero matrix: a matrix with all its elements equal to zero. Note that this matrix is the matrix analogue for the scalar 0, in the sense that for any matrix A, we have that: $A\underline{0} = \underline{0}A = \underline{0}$.
- Idenpotent matrix: a matrix A that satisfies $A^2 = AA = A$, which also means that $A^3 = AAA = (AA)A = AA = A$ and that $A^n = A$ for any n. Note that for a matrix to be idenpotent it must be a square matrix.

Rank

An n×k matrix A can be thought of as comprising k vectors: $A = [v_1,v_2,...,v_k]$. The number of these k vectors which are linearly independent is the *column rank* or simply the *rank* of a matrix. The row and column rank of a matrix are always equal.

A matrix with rank equal to its number of columns is of *full rank.*

If C=AB then each of the columns of C will be a linear combination of the columns of A. Each of the rows of C will be a linear combination of the rows of B. So:

$$\text{rank}(AB) \le \min\{\text{rank}(A), \text{rank}(B)\}$$

A useful corollary is that: rank(X) = rank(X'X) = rank (XX').

Inverting Matrices and Determinants

The matrix A is invertible (or non-singular) if there exists another matrix B such that we have AB = BA = I (where I is the identity matrix). A matrix is invertible only if it is a square matrix, but not all square matrices are invertibles. A square matrix is invertible if and only if it is of full rank.

We denote the inverse of A as A^{-1}, so if A^{-1} exits, it is always the case that:

$$AA^{-1} = A^{-1}A = I$$

For example:

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 3 & 3 & 2 \\ 1 & 3 & 4 \end{bmatrix} \Rightarrow A^{-1} = \begin{bmatrix} -3 & 1 & 1 \\ 5 & -1 & -2 \\ -3 & 0.5 & 1.5 \end{bmatrix}$$

We always have that: $(A')^{-1} = (A^{-1})'$

If A and B are both invertible then $(AB)^{-1} = B^{-1}A^{-1}$, but note that $(AB)^{-1}$ can exist even if A and B are not invertible (A and B can even be non-square matrices).

Let A_{ij} be the (n-1)×(n-1) submatrix obtained by deleting row i and column j from A. Then $M_{ij}/\det(A_{ij})$ is the (i,j)th *minor* of A and $C_{ij}/(-1)^{i+j}M_{ij}$ is the (i,j)th *cofactor* of A.

The determinant of an n×n matrix A is $\det(A) = a_{11}C_{11} + a_{12}C_{12} + ... + a_{1n}C_{1n}$, where a_{ij} is the (i,j)th element of A. The *adjoint* of A, adj(A) is the matrix whose (i,j)th element is Cji, the (j,i)th cofactor of A.

The inverse A^{-1} of a matrix A is defined as adj(A)/det(A). The determinant of a matrix is nonzero if and only if it has full rank. Consequently, a matrix with zero determinant is non-invertible or *singular*.

Econometric application: the OLS estimates of the parameters β are b = $(X'X)^{-1}X'Y$. From the above, these can only be computed if X is of full rank.

Taking Derivatives

Taking derivative with respect to a matrix:

$$\frac{\partial f(A)}{\partial A_{n,m}} = \begin{bmatrix} \frac{\partial f(A)}{\partial a_{11}} & \frac{\partial f(A)}{\partial a_{12}} & \cdots & \frac{\partial f(A)}{\partial a_{1m}} \\ \frac{\partial f(A)}{\partial a_{21}} & & & \frac{\partial f(A)}{\partial a_{2m}} \\ \vdots & & & \vdots \\ \frac{\partial f(A)}{\partial a_{n1}} & \frac{\partial f(A)}{\partial a_{n2}} & \cdots & \frac{\partial f(A)}{\partial a_{nm}} \end{bmatrix}$$

Taking derivative with respect to a vector:

$$\frac{\partial f(x)}{\partial x_n} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Then we have:

1. For a, b vectors: $\frac{\partial (a'b)}{\partial b} = \frac{\partial (b'a)}{\partial b} = a$

2. For a square matrix A: $\frac{\partial (b'Ab)}{\partial b} = 2Ab$, $\frac{\partial (b'Ab)}{\partial A} = bb'$

Block Matrices

Sometimes we find it more convenient to separate the matrix into blocks, and not into elements, and then do all the operations block by block instead of element by element (provided that the dimensions match).

For example:

$$A = \begin{bmatrix} 1 & 9 & 7 & | 7 & 4 \\ 4 & 5 & 6 & | 7 & 8 \\ 1 & 3 & 5 & | 4 & 5 \\ 3 & 5 & 5 & | 3 & 0 \\ \overline{4} & \overline{5} & \overline{0} & | \overline{8} & \overline{8} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \end{bmatrix}$$

where: $A_{11} = \begin{bmatrix} 1 & 9 & 7 \\ 4 & 5 & 6 \\ 1 & 3 & 5 \\ 3 & 5 & 5 \end{bmatrix}$ $\quad A_{12} = [4 \quad 5 \quad 0]$ $\quad A_{21} = \begin{bmatrix} 7 & 4 \\ 7 & 8 \\ 4 & 5 \\ 3 & 0 \end{bmatrix}$ $\quad A_{22} = [8 \quad 8]$

Eigenvectors and Eigenvalues

Given a square matrix A, it is often useful to think of a transformation which would make it diagonal. Eigenvectors and eigenvalues allow us to turn A into a diagonal or nearly diagonal matrix.

The number λ is an *eigenvalue* of A if and only if det(A-λI) = 0. This is the *characteristic equation* for A, and is an n'th degree polynomial in λ, with exactly n (possibly repeated and complex) solutions. The eigenvalues of a matrix are sometimes referred to as the *characteristic values*.

If det(A-λI)=0, there must be a nonzero vector (an *eigenvector*) v such that (A-λI)v = 0, so Av = λv. Each eigenvalue thus defines a corresponding eigenvector. Note that we could multiply v by any scalar and it would still satisfy this condition, so the eigenvectors are defined only up to normalization.

Let P be the matrix whose columns are the eigenvectors of A: P = $[v_1, v_2, \ldots, v_n]$. A useful result is that if A has n distinct eigenvalues, the corresponding n eigenvectors are linearly independent.

$$AP = A[v_1, \ldots, v_n] = [Av_1, \ldots, Av_n] = [\lambda_1 v_1, \ldots, \lambda_n v_n] = [v_1, \ldots, v_n] \begin{bmatrix} \lambda_1 & \ldots & 0 \\ : & & : \\ 0 & \ldots & \lambda_n \end{bmatrix} = P\Gamma$$

where Γ is a diagonal matrix with the eigenvectors of A on its diagonal. Then, if P is invertible, $P^{-1}AP = \Gamma$ and we have *diagonalized* A. Alternately, $A = P\Gamma P^{-1}$.

Symmetric Matrices

In econometrics we will use a lot of symmetric matrices. In particular, given the regression model Y = Xb + e, the variance-covariance matrix E = ee' is a symmetric n×n matrix.

If an n×n matrix A is symmetric, it has n distinct **real** eigenvalues and its eigenvectors are mutually orthogonal, i.e. $v_i' v_j = v_j' v_i = 0$ for any i≠j.

Proof: start with $Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2$, $A = A'$, $\lambda_1 \neq \lambda_2$. Then $\lambda_1 v_1' v_2 = (\lambda_1 v_1)' v_2 = (Av_1)' v_2 = v_1' A' v_2 = v_1' A v_2 = v_1' (\lambda_2 v_2) = \lambda_2 v_1' v_2$.
So $\lambda_1 v1' v_2 = \lambda_2 v_1' v_2$, which, by $\lambda_1 \neq \lambda_2$, implies $v_1' v_2 = 0$.

Remember that we can scale eigenvectors by any factor we like. So let's scale all the eigenvectors in the matrix P to be of length one. Now if the eigenvectors are mutually orthogonal we have that P'P = PP' = I (where I is the identity matrix), so $P' = P^{-1}$.

This result turns out to be very useful in finding powers of matrices: consider $AA = A^2$. If A is symmetric, $AA = (P\Gamma P')(P\Gamma P') = P\Gamma P' P\Gamma P' = P\Gamma^2 P$, since P'P = I. This is useful to know, since Γ^2 is simply the matrix Γ with each diagonal element squared.

Consider $\Gamma^{1/2}$. This will only be defined if the eigenvalues of A are all non-negative.

$P\Gamma^{1/2} P' P\Gamma^{1/2} P' = P\Gamma P' = A$, so $P\Gamma^{1/2} P'$ is the square root of the matrix A. This is the *Cholesky factorization* of A.
Sometimes we will want to construct $A^{-1/2}$. This will be $P\Gamma^{-1/2} P'$ and will be defined if all the eigenvalues of A are positive.

Definiteness of Quadratic Forms

A quadratic form is a polynomial in $x_1 \ldots x_n$ where the highest power is two. Thus matrix, for $Q(x_1,\ldots x_n) = \Sigma a_{ij} x_i x_j$ or, if n=2, $Q(x_1, x_2) = a_{11} x_1^2 + a_{12} x_1 x_2 + a_{22} x_2^2$
We can always write a quadratic form in matrix notation as x'Ax where A is a symmetric matrix, for example in the case n=2:

$$Q(x_1, x_2) = (x_1, x_2) \begin{pmatrix} a_{11} & \frac{1}{2} a_{12} \\ \frac{1}{2} a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

- A matrix A is positive definite if x'Ax > 0 for any vector x≠0
- A matrix A is positive semidefinite if x'Ax ≥ 0 for any vector x

- A matrix A is negative definite if x'Ax < 0 for any vector $x \neq 0$

- A matrix A is negative semidefinite if x'Ax ≤ 0 for any vector x

- A matrix A is indefinite if it is neither positive semidefinite nor negative semidefinite

The definiteness of a quadratic form is determined by its eigenvalues. Suppose we choose x to be an eigenvector v_1 of A. Then $x'Ax = \lambda_1 \|v_1\| \|v_1\|$. Although this is only a one-way proof, it is true that

A is positive definite ⇔ the eigenvalues of A are all positive,
A is positive semidefinite ⇔ the eigenvalues of A are all non-negative, and so on.

We suggest that positive-definite matrices are thought of as being like positive numbers, in that one can find their square roots.

More useful facts about eigenvalues are that:

$$\det(A) = \prod_{i=1}^{n} \lambda_i \qquad \text{and} \qquad trace(A) = \sum_{i=1}^{n} \lambda_i$$