Review of Matrix Algebra and related topics
Econ 837[1]

Outline:

1. Matrices: addition, multiplication, transposition and rank.
2. Square matrices: inversion, determinant, and trace.
3. Orthogonal projections.
4. Moments of random variables.
5. Matrix differentiation.
6. Some other results.

[1] If you find any error or typo, please email me at bertille_antoine@sfu.ca

1 Matrices: addition, multiplication, transposition and rank.

Basic operations on numbers, like addition and multiplication, can usefully be extended to arrays of numbers.

A matrix is a rectangular array of real numbers. For example,

$$A = \begin{pmatrix} 4 & -1.5 & 0 \\ -0.5 & 1 & 5 \end{pmatrix}$$

is a matrix of order (or dimension) 2×3, meaning that is has 2 rows and 3 columns. We write A_{ij} to denote the element on the $i-th$ row and $j-th$ column of A, e.g. $A_{23} = 5$.

A row vector is a matrix with only one row. A column vector, or vector, is a matrix with only one column.

1.1 Addition and Scalar Multiplication

If A and B are matrices of the same order, the sum $A + B$ is defined by:

$$(A+B)_{ij} = A_{ij} + B_{ij}$$

Matrix addition is commutative and associative:

$$\begin{aligned} A+B &= B+A \\ (A+B)+C &= A+(B+C) \end{aligned}$$

If p is a scalar (i.e. a real number) and A a matrix, the product pA is defined by:

$$(pA)_{ij} = pA_{ij}$$

Some properties:

$$\begin{aligned} (p+q)A &= pA+qA \\ p(A+B) &= pA+pB \\ p(qA) &= (pq)A \\ A(pB) &= (pA)B = p(AB) \end{aligned}$$

1.2 Matrix Multiplication

If A is a $k \times l$ matrix and B an $l \times m$ matrix, the product AB is the $k \times m$ matrix defined by:

$$(AB)_{ij} = \sum_{s=1}^{l} A_{is} B_{sj}$$

For example:

$$\begin{pmatrix} 4 & 0 & 0 \\ -2 & 1 & 5 \end{pmatrix} \begin{pmatrix} 2 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 & -4 & 4 & 0 \\ 1 & 8 & -2 & 5 \end{pmatrix}$$

Note that AB exists only when A has the same number of columns as B has rows.

Matrix multiplication satisfies:

$$\begin{aligned} (AB)C &= A(BC) \\ A(B+C) &= AB + AC \\ (A+B)C &= AC + BC \end{aligned}$$

Unlike multiplication of real number, matrix multiplication is not commutative. That is in general $AB \neq BA$, and BA need not even exist.

1.3 Matrix Transpose

The *transpose* of a $k \times l$ matrix A, denoted $A^{'}$, is the $l \times k$ matrix defined by:

$$(A')_{ij} = A_{ji}$$

Thus, transposition amounts to exchanging rows with columns. Properties:

$$\begin{aligned} (A')' &= A \\ (A+B)' &= A' + B' \\ (AB)' &= B'A' \end{aligned}$$

1.4 Linear Independence

The (row or column) vectors $a_1, ..., a_l$ are linearly independent if any non-zero linear combination of $a_1, ..., a_l$ is non-zero. That is, if:

$$\sum_{j=1}^{l} \lambda_j a_j \neq 0$$

for any set of scalars $\lambda_1, ..., \lambda_l$ not all equal to 0. Otherwise, $a_1, ..., a_l$ are linearly dependent. (We write $A = 0$ if all elements of A are 0, and $A \neq 0$ otherwise.) For example,

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

are linearly independent, while

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

are linearly dependent.

1.5 Rank

The rank of a matrix is the maximal number of linearly independent columns (or of linearly independent rows). Equivalently, the column rank of A is the dimension of the column space of A, while the row rank of A is the dimension of the row space of A.

The column rank and the row rank are always equal. This number (i.e. the number of linearly independent rows or columns) is simply called the rank of A. For example,

$$rank \begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} = 2$$

$$rank \begin{pmatrix} 2 & -1 & 1 \\ -4 & 2 & -2 \end{pmatrix} = 1$$

For any matrix A,

$$rankA = rank(AA') = rank(A'A)$$

and

$$rank(AB) \leq \min(rankA, rankB).$$

A $k \times l$ matrix A has full column rank if $rankA = l$, and full row rank if $rankA = k$. Notice that if the matrix A is not squared, then it cannot have both full column rank and full row rank, although we still have that rank = row rank = column rank.

2 Square matrices: inversion, determinant, and trace.

2.1 Inverse

Non-zero numbers can be inverted, e.g. the inverse of 3 is $\frac{1}{3}$, because $3 \times \frac{1}{3} = 1$. This is extended to square, nonsingular matrices.

A matrix is *square* if it has as many rows as columns. A square matrix A is *symmetric* if $A = A'$, and *diagonal* if $A_{ij} = 0$ for all $i \neq j$. The *identity matrix* (of any particular order) is the diagonal matrix which has all diagonal elements equal to 1. It is denoted by I or I_n, where n is its order. For example,

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

The identity matrix is the neutral element of matrix multiplication. That is, for any $l \times m$ matrix B,

$$BI_m = B \qquad \text{and} \qquad I_l B = B$$

An $n \times n$ matrix A is *nonsingular* if $rankA = n$, and *singular* if $rankA < n$. Nonsingular matrices can be inverted. The *inverse* of a nonsingular $n \times n$ matrix A is the $n \times n$ matrix, denoted A^{-1}, that satisfies

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

For example,

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

provided $ad - bc \neq 0$ (the condition for this matrix to be nonsingular). The inverse is always unique. Singular matrices cannot be inverted. Useful properties are

$$\begin{aligned} (A^{-1})' &= (A')^{-1} \\ (AB)^{-1} &= B^{-1}A^{-1} \end{aligned}$$

for nonsingular matrices A and B.

A nonsingular matrix A is *orthogonal* if $A^{-1} = A'$. If A is orthogonal, then $A'A = AA' = I$.

A square matrix A is *idempotent* if $A^2 = A$.

2.2 Determinant

The determinant of an $n \times n$ matrix A is defined as:

$$|A| = \sum (-1)^{r(j_1,\ldots,j_n)} \prod_{i=1}^{n} A_{ij}$$

where the summation is over all permutations $(j_1, ..., j_n)$ of $(1, ..., n)$, and $r(j_1, ..., j_n)$ is the number of pairwise interchanges to transform $(j_1, ..., j_n)$ to $(1, ..., n)$. For example:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Notice that:

$$\begin{vmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ 0 & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_{nn} \end{vmatrix} = \prod_{i=1}^{n} A_{ii},$$

so

$$|I_n| = 1.$$

A very useful property is:

$$A \text{ is nonsingular} \iff |A| \neq 0.$$

Therefore, A is invertible iff $|A| \neq 0$.

Furthermore for $n \times n$ matrices A and B and a scalar a,

$$\begin{aligned} |AB| &= |A|\,|B| \\ |A'| &= |A| \\ |aA| &= a^n\,|A| \end{aligned}$$

and, if A^{-1} exists,

$$\left|A^{-1}\right| = \frac{1}{|A|}.$$

2.3 Trace

The *trace* of an $n \times n$ matrix A is defined as

$$trA = \sum_{i=1}^{n} A_{ii}.$$

Useful properties are

$$\begin{aligned} tr(A+B) &= trA + trB \\ tr(aA) &= atrA \\ trA' &= trA \\ tr(AB) &= tr(BA) \end{aligned}$$

Remark: while AB and BA must be square, they are not of the same order if A and B are not square.

If A is idempotent,

$$trA = rankA.$$

3 Orthogonal Projections

Vectors can be viewed as points in a Euclidean space. A few such points generate a subspace of that space. Any point can be orthogonally projected onto that subspace.

The column space of a $k \times l$ matrix , written $M(A)$, is the set of linear combinations of the columns of A,

$$M(A) = \{y|y = Ax \quad \text{for some vector } x\}$$

Thus, when equipped with vector addition and scalar multiplication, $M(A)$ is the vector space (over the field of real numbers) generated (or spanned) by the columns of A, and has dimension $rank(A)$. If A has full column rank, l, its columns are basis vectors of $M(A)$ (although not orthogonal and not normalized). Note that $M(A)$ is an l-dimensional subspace of the k- dimensional Euclidean space, $\mathbb{R}^k$.

Let the $k \times l$ matrix have full column rank, l. Note that $k \geq l$ and that the $l \times l$ matrix $A'A$ is nonsingular. Consider a $k \times l$ vector y. The orthogonal projection of y onto $M(A)$, denoted $p_A(y)$, is the $k \times 1$ vector in $M(A)$ that is closest to y in the Euclidean metric:

$$d(a, b) = \|a - b\| = \sqrt{(a - b)'(a - b)}.$$

Thus $p_A(y)$ solves

$$\min_{p \in M(A)} \|p - y\| .$$

Because $p \in M(A)$,we can write $p = Ax$ for some vector x, and solve instead the unconstrained problem,

$$\min_{x} \|Ax - y\| .$$

This is least-squares problem (as you have seen in class), whose solution for x is $(AA')^{-1}A'y$, and so

$$p_A(y) = A(A'A)^{-1}A'y = P_A \; y,$$

where the $k \times k$ matrix

$$P_A = A(A'A)^{-1}A'$$

is called the *projection matrix onto* $M(A)$.

P_A is symmetric and has rank l. Furthermore, P_A is idempotent. Hence,

$$p_A(p_A(y)) = P_A(P_A(y)) = P_A y = p_A(y).$$

Consider now the set of vectors in $\mathbb{R}^k$ that are orthogonal to $M(A)$. Denote this set as $M(A)^{\perp}$. Note that this has dimension $k - l$ and that

$$a'b = 0, \text{ for any } a \in M(A) \text{ and } b \in M(A)^{\perp}.$$

Furthermore, we can decompose any vector $y \in \mathbb{R}^k$ into orthogonal components

$$y = P_A \; y + M_A y,$$

where

$$M_A = I - P_A = I - A(A'A)^{-1}A'$$

is the projection matrix onto $M(A)^{\perp}$. Note that, indeed, $P_A M_A = M_A P_A = 0$.

Remark: The columns of a matrix generate a subspace of a Euclidean space. The projection of any vector onto that subspace is found by pre-multiplying that vector by the appropriate projection matrix.

4 Moments of random variables

The following properties are extremely useful. For any constants a, B, c, and D, with a, B, c, D, X, Y of dimensions $q \times 1$, $q \times p$, $p \times 1$, $p \times q$, $p \times 1$, $q \times 1$ respectively,

$$\begin{aligned}
E(E(X)) &= E(X) \\
E(a + BX) &= a + BE(X) \\
Var(a + BX) &= BVar(X)B' \\
Cov(X, Y) &= [Cov(Y, X)]' \\
Cov(a + BX, c + DY) &= BCov(X, Y)D' \\
Var(X + Y) &= Var(X) + Var(Y) + Cov(X, Y) + Cov(Y, X) \quad \text{(if } p = q\text{)} \\
Cov(X + Y, Z) &= Cov(X, Z) + Cov(Y, Z) \;\; \text{(if } p = q\text{, and } Z \text{ is } l \times 1\text{)}
\end{aligned}$$

Also extremely useful is the *Law of Iterated Expectations (LIE),*

$$E[g(X)] = E\{E[g(X)|Y]\}$$

and the *Law of Total Variance,*

$$Var(X) = E[Var(X|Y)] + Var(E(X|Y).$$

Remark:

Proof of $Cov(BX, DY) = \; BCov(X, Y)D'$:

$$\begin{aligned}
Cov(BX, DY) &= E\left[(BX - E(BX)) \;\; (DY - E(DY))^{'}\right] \\
&= E\left[B\,(X - EX) \;\; (D(Y - EY))^{'}\right] \\
&= E\left[B\,(X - EX) \;\; (Y - EY)D^{'}\right] \\
&= BE\left[(X - EX) \;\; (Y - EY)\right] D^{'} \\
&= BCov(X, Y)D'
\end{aligned}$$

Now, if instead you have $Cov(XB, DY)$:

$$\begin{aligned}
Cov(XB, DY) &= E\left[(XB - E(XB)) \;\; (DY - E(DY))^{'}\right] \\
&= E\left[(X - EX)\,B \;\; (D(Y - EY))^{'}\right] \\
&= E\left[(X - EX)\,B\,(Y - EY)D^{'}\right] \\
&= E\left[(X - EX)\,B\,(Y - EY)\right] D^{'} \\
&= Cov(XB, Y)D'
\end{aligned}$$

Hence, in this case we cannot factor out B. Also, notice that we need the dimensions of X, B, D, and Y to be for instance $q \times p$, $p \times 1$, $p \times q$, $q \times 1$ respectively.

5 Matrix Differentiation

There are a few simple rules for matrix differentiation. These allow much econometrics to be done in matrix form, which can be simpler and far less cumbersome than using nested summation signs.

5.1 Differentiating an inner product with respect to a vector

Let a be a given column vector and let θ be a column choice vector (a vector of values to be chosen). Their transposes can be denoted as a' and θ' . Then the derivative of their inner product, which is a scalar, is a column vector:

$$\frac{\partial a'\theta}{\partial \theta} = \frac{\partial \theta' a}{\partial \theta} = a$$

where the dimensions of a and θ are such that the inner product is well defined.

5.2 Differentiating a quadratic form with respect to a vector

Let A be a matrix, either symmetric or non-symmetric, and consider the quadratic form $\theta' A\theta$, which is itself a scalar. The derivative of this quadratic form with respect to the vector θ is the column vector:

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

But in econometrics, almost always the matrix in the quadratic form will be symmetric. If A is indeed symmetric, the formula can be simplified to

$$\frac{\partial \theta' A\theta}{\partial \theta} = 2A\theta.$$

Finally, let M be a $k \times k$ matrix , and a, b $k \times 1$ vectors, then:

$$\begin{aligned} \frac{\partial (a'Mb)}{\partial M} &= ab' \\ \frac{\partial (a'M'b)}{\partial M} &= ba' \end{aligned}$$

5.3 Application to Ordinary Least Squares

Perhaps the most basic concept in econometrics is ordinary least squares, in which we choose the regression coefficients so as to minimize the sum of squared residuals of the regression. Suppose the regression model is

$$y = X\beta + u,$$

where y is an $n \times 1$ vector of observed values of the dependent variable, X is an $n \times k$ matrix in which each column is an $n \times 1$ vector of observed values of one of the k independent variables ($k < n$), β is the $k \times 1$ vector of parameters to be estimated, and u is the $n \times 1$ error term. We assume that $E(u) = 0$ and $E(u|x) = 0$.

We want to minimize the sum of square residuals:

$$\begin{aligned} u'u &= (y - X\beta)'(y - X\beta) \\ &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta \\ &= y'y - 2\beta'(X'y) + \beta\prime(X'X)\beta \end{aligned}$$

where the last equality holds because $y'X\beta$ is a scalar and thus it is equal to its transpose.

Note that $(X'y)$ is a $k \times 1$ vector and $(X'X)$ is a $k \times k$ matrix. Using the above rules for differentiation, we have:

$$\frac{\partial u'u}{\partial \beta} = -2X'y + 2(X'X)\beta,$$

where we have used the fact that $X'X$ is symmetric.

Remark: Alternatively, you can take derivative directly from the quadratic form $u'u = (y - X\beta)'(y - X\beta)$:

$$\frac{\partial u'u}{\partial \beta} = -2(y - X\beta)X'.$$

5.4 Some useful dimension rules

Whenever you take derivatives w.r.t. a vector it is easy to mess up with the dimensions, thus you should always check that they are as expected.

Let θ be a $k \times 1$ column vector, and $\Psi(\theta)$ be an $H \times 1$ column vector. Two simple rules to remember:

- If you take derivative of the column vector $\Psi(\theta)$ w.r.t. the row vector θ' then your resulting matrix will have k *columns*. That is:

$$\frac{\partial \Psi(\theta)}{\partial \theta'} = \begin{pmatrix} \frac{\partial \Psi(\theta)}{\partial \theta_1} & \cdots & \frac{\partial \Psi(\theta)}{\partial \theta_k} \end{pmatrix} = \begin{pmatrix} \frac{\partial \Psi_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial \Psi_1(\theta)}{\partial \theta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Psi_H(\theta)}{\partial \theta_1} & \cdots & \frac{\partial \Psi_H(\theta)}{\partial \theta_k} \end{pmatrix} \text{ is } H \times k$$

- If you take derivative of the row vector $\Psi'(\theta)$ w.r.t. the column vector θ then your resulting matrix will have k *rows*. That is:

$$\frac{\partial \Psi'(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial \Psi(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \Psi(\theta)}{\partial \theta_k} \end{pmatrix} = \begin{pmatrix} \frac{\partial \Psi_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial \Psi_H(\theta)}{\partial \theta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Psi_1(\theta)}{\partial \theta_k} & \cdots & \frac{\partial \Psi_H(\theta)}{\partial \theta_k} \end{pmatrix} = \left[\frac{\partial \Psi(\theta)}{\partial \theta'}\right]' \text{ is } k \times H$$

5.5 The Jacobian and Hessian

Let θ be a $k \times 1$ column vector, and $\Psi(\theta)$ be an $H \times 1$ column vector.

- The Jacobian of $\Psi(\theta)$ w.r.t. θ is given by:

$$J[\Psi(\theta)] = \frac{\partial \Psi(\theta)}{\partial \theta'}$$

- The Hessian of $\Psi(\theta)$ w.r.t. θ is given by:

$$H[\Psi(\theta)] = \frac{\partial^2 \Psi(\theta)}{\partial \theta \theta'}$$

6 Some other results

- **Thm.1.** Let $X \sim N(0, \Sigma)$, where the $n \times n$ matrix Σ is nonsingular. Then

$$X'\Sigma^{-1}X \sim \chi^2_{(n)}.$$

- **Thm.2.** Let $X \sim N(0, I_n)$ and let A be an $n \times n$ symmetric, idempotent matrix with rank p. Then,

$$X'AX \sim \chi^2_{(p)}$$

- **Thm.3.** Let $X \sim N(0, I_n)$ and let A and B be $n \times n$ symmetric, idempotent matrices with rank p and q, respectively, such that $AB = 0$. Then

$$X'AX \quad \text{and} \quad X'BX \quad \text{are independent}$$

  and hence

$$\frac{X'AX/p}{X'BX/q} \sim F(p, q).$$