This handout is based on Bergantino (1998, 1). Another good reference for econometrics is Greene (2003, Appdx. A). Friedberg et al. (1997) is a good textbook for those wanting to learn this material at a deeper level.

1 Basic Definitions and Properties

1.1 Definition of a Matrix

An $M \times N$ matrix A

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix}$$

$a_{ij} = (A)_{ij}$, the element in the ith row and jth column of A, is a real number.

$A_{M \times N}$ has M rows and N columns.

$A_{M \times 1}$ (M rows and 1 column) is a point in $\mathbb{R}^M$, i.e. an M-dimensional column vector.

$A_{1 \times N}$ (1 row and M columns) is an N-dimensional row vector.

$A_{1 \times 1}$ is just a real number (a scalar).

We write the jth column of A as A_j, i.e.

$$A_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{Mj} \end{pmatrix}$$

This is an $M \times 1$ vector

$\mathcal{M}_{M \times N}$ is the set of all $M \times N$ matrices

If $A \in \mathcal{M}_{N \times N}$ (i.e. same number of rows as columns), then A is a *square* matrix

If A is square, then the elements $a_{11}, \ldots, a_{NN}$ are called the *diagonal* elements (since they are found along the diagonal of the array).

An *upper triangular* matrix is square and has only zeros below the diagonal, i.e. $a_{ij} = 0$ for all $i > j$.

A *lower triangular* matrix is square and has only zeros above the diagonal, i.e. $a_{ij} = 0$ for all $i < j$.

A *triangular* matrix is upper or lower triangular.

A *diagonal matrix* is square and has nonzero elements only on the diagonal, i.e. $a_{ij} = 0$ for all $i \neq j$. A diagonal matrix is both upper and lower triangular.

The *identity matrix* $I_N \in \mathcal{M}_{N\times N}$ is defined as

$$\begin{aligned} (I)_{ii} &= 1 \quad i = 1, \ldots, N \\ (I)_{ij} &= 0 \qquad\quad i \neq j \end{aligned}$$

That is, I_N is an $N \times N$ matrix with 1s on the diagonal and 0s elsewhere:

$$I_N = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

So $I_1 = 1$, $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, etc. I is a diagonal matrix.

1.2 Matrix Addition

If $A, B \in \mathcal{M}_{M\times N}$, $A + B$ is defined by $(A + B)_{ij} = a_{ij} + b_{ij}$. That is,

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1N} + b_{1N} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2N} + b_{2N} \\ \vdots & \vdots & & \vdots \\ a_{M1} + b_{M1} & a_{M2} + b_{M2} & \cdots & a_{MN} + a_{MN} \end{bmatrix}$$

Note that for the sum of A and B to be defined, A and B must have the same dimensions (same number of rows and columns).

1.3 Scalar Multiplication

If $c \in \mathbb{R}$, then cA is defined by $(cA)_{ij} = ca_{ij}$. That is,

$$cA = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1N} \\ ca_{21} & ca_{22} & \cdots & ca_{2N} \\ \vdots & \vdots & & \vdots \\ ca_{M1} & ca_{M2} & \cdots & ca_{MN} \end{bmatrix}$$

1.4 Matrix Multiplication

If $A \in \mathcal{M}_{M \times N}$ and $B \in \mathcal{M}_{N \times P}$, then their product AB is the $M \times P$ matrix defined by

$$(AB)_{ij} = \sum_{n=1}^{N} a_{in} b_{nj}$$

That is, the ijth entry in AB is found by taking the inner product of the ith row of A and the jth row of B.

Note that A and B must be "conformable" for their product to be well-defined. That is, the number of columns of A must equal the number of rows of B.

In general, $AB \neq BA$. Note in particular that if A and B are not both square, then AB and BA cannot both be defined. Furthermore, even if A and B are both square (so that both products are defined), in general $AB \neq BA$.

1.5 Transpose

If $A \in \mathcal{M}_{M \times N}$ then $A' \in \mathcal{M}_{N \times M}$ is defined by $(A')_{ij} = (A)_{ji} = a_{ji}$

This is sometimes written A^T or A^t.

The transpose operator is often used as a shorthand for taking a dot product (also called inner or scalar product). For example, suppose $x, y \in \mathbb{R}^N$. We know that $x \cdot y = \sum_{i=1}^{N} x_i y_i$, but this is just $x'y$ by the definition of matrix multiplication.

1.6 Symmetry

A matrix A is *symmetric* $\iff A = A' \iff a_{ij} = a_{ji}$ for all i, j. Note that only square matrices can be symmetric.

1.7 Trace

If $A \in \mathcal{M}_{N \times N}$ then the trace of A, written $tr\,(A)$, is defined as

$$tr\,(A) = \sum_{i=1}^{N} a_{ii}$$

That is, the trace of A is the sum of the diagonal elements of A. Note that A must be square for the trace to be defined.

1.8 Basic Properties

These are stated without proof unless the proof is interesting or enlightening. Proofs can be found in any linear algebra textbook. If a matrix product is written, it is assumed that the matrices are conformable for multiplication.

1. $(A + B) + C = A + (B + C) = A + B + C$
2. $(AB)\,C = A\,(BC) = ABC$
3. $A\,(B + C) = AB + AC\,;\ (A + B)\,C = AC + BC$
4. If A and B are diagonal matrices, then AB is also diagonal with $(AB)_{ii} = a_{ii} b_{ii}$
5. If A and B are upper triangular matrices, then AB is also upper triangular
6. If A and B are lower triangular matrices, then AB is also lower triangular

7. $(A')' = A$

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

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

10. For any $M \times N$ matrix A, $A'A$ and AA' are both symmetric

    Proof:

$$\begin{aligned} (A'A)' &= A'(A')' \\ &= A'A \\ (AA')' &= (A')'A' \\ &= AA' \end{aligned}$$

11. Trace is a linear operator

    (a) $tr(cA) = ctr(A)$

    (b) $tr(A+B) = tr(A) + tr(B)$

12. $tr(A') = tr(A)$

13. "Cyclic permutations" – this property will turn out to be crucial in the derivation of the Full Information Maximum Likelihood (FIML) estimator in 14.383

    (a) $tr(AB) = tr(BA)$ provided both AB and BA are defined

    (b) $tr(ABC) = tr(BCA)$ provided ABC and BCA are both defined; $tr(ABC) = tr(CAB)$, provided ABC and CAB are both defined.

    (c) If ABC is square, all the products in (b) will be defined.

    Proof: $(ABC)_{N\times N} \Rightarrow A_{N\times M}, B_{M\times P}, C_{P\times N} \Rightarrow (BCA)_{M\times M}, (CAB)_{P\times P}$

14. If A and B are both diagonal $N \times N$ matrices, AB is a diagonal $N \times N$ matrix with $(AB)_{ii} = a_{ii}b_{ii}$

2 Matrix Inverses

2.1 Rank of a Matrix[1]

2.1.1 Linear Independence

Vectors $v_{1,} \ldots, v_K$ are *linearly independent* $\iff$ $c_1 v_1 + \cdots + c_K v_K = 0 \Rightarrow c_1 = \cdots = c_K = 0$

That is, the only representation of the *null vector* 0 as linear combinations of the vectors v_i is *trivial.*

An equivalent definition is that for any vector $v_j \in \{v_{1,} \ldots, v_K\}$, we cannot "reproduce" v_j as a linear combination of the other vectors. Formally,

$$\forall v_j \in \{v_{1,} \ldots, v_K\}, \nexists \{c_i\}_{i \neq j} \text{ such that } v_j = \sum_{i \neq j} c_i v_i.$$

Another way to think about this is that in a set of linearly independent vectors, no member vector is "redundant."

2.1.2 Rank

The *rank* of a matrix A, written $\mathit{rank}(A)$, is the maximum number of linearly independent columns of A.

Result (without proof): the maximum number of linearly independent *columns* of $A =$ the maximum number of linearly independent *rows* of A

Implication: if A is $M \times N$, $\mathit{rank}(A) \leq \max\{M, N\}$

If A is $M \times N$ and $\mathit{rank}(A) = M$, we say A has *full row rank*

If A is $M \times N$ with $\mathit{rank}(A) = M$ and $x \in \mathbb{R}^N$, $Ax = 0$ only if $x = 0$

If A is $M \times N$ with $\mathit{rank}(A) < M$ then there exist $x \in \mathbb{R}^N$, $x \neq 0$ such that $Ax = 0$

If A is $M \times N$ and $\mathit{rank}(A) = N$, we say A has *full column rank*

If A is $M \times N$ with $\mathit{rank}(A) = N$ and $x \in \mathbb{R}^M$, $A'x = 0$ only if $x = 0$

If A is $M \times N$ with $\mathit{rank}(A) < N$ then there exist $x \in \mathbb{R}^M$, $x \neq 0$ such that $A'x = 0$

---

[1] For more details, see Section 2.8.

2.2 Inverse

The *inverse* of a matrix A is the (unique) matrix A^{-1} such that $A^{-1}A = I = AA^{-1}$.

Not all matrices have inverses. (examples: 0, $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$)

Only square matrices can be invertible, but not all square matrices are invertible.

If a matrix A has an inverse, we say that A is *nonsingular*.

A^{-1} **exists** $\iff$ A **has full rank (**i.e. $A_{N\times N}$ with $rank\,(A) = N$)

If A is a nonsingular matrix and $x \in \mathbb{R}^N$, then $Ax = 0$ only if $x = 0$.

If A is a singular matrix, then there exist $x \in \mathbb{R}^N, x \neq 0$ such that $Ax = 0$

2.3 Orthogonal Matrices

Vectors $x, y \in \mathbb{R}^n$ are orthogonal $\iff$ $x'y = 0$

An $N \times N$ matrix A is *orthogonal* $\iff$ $A^{-1} = A'$. That is, $A'A = I = AA'$

Notice that if A is orthogonal and A_i, A_j are columns of A and therefore $N \times 1$ column vectors, then $A_i'A_j = 0$ for all $i \neq j$. That is, if A is an orthogonal matrix then its columns are orthogonal to each other.

Some texts may call a matrix orthogonal if $A'A$ is diagonal, and use the term *orthonormal* for matrices such that $A'A = I$

2.4 Determinant

We define the *determinant* of an $N \times N$ matrix A iteratively as follows:

$$\begin{aligned} \det(A) &= a_{11} && \text{for } A_{1\times1} \\ &= a_{11}a_{22} - a_{21}a_{12} && \text{for } A_{2\times2} \\ &= \textstyle\sum_{j=1}^{N} (-1)^{i+j}\, a_{ij} \det\left(\tilde{A}_{ij}\right) && \text{for } A_{N\times N} \end{aligned}$$

where i is any row of A and $\tilde{A}_{ij}$ is the $N-1 \times N-1$ matrix obtained from A by deleting the ith row and jth column of A.

Note that the determinant is a *scalar*. Determinants are only defined for square matrices.

2.5 Basic Properties

1. If A is triangular (upper or lower), $\det(A) = \prod_{i=1}^{N} a_{ii}$
2. If A is diagonal, $\det(A) = \prod_{i=1}^{N} a_{ii}$
3. $\det(I) = 1$
4. For any scalar c, $\det(cA) = c^N \det(A)$
5. $\det(A') = \det(A)$
6. $\det(AB) = \det(A)\det(B)$
7. $\det\left(A^{-1}\right) = (\det(A))^{-1}$

   Proof: $\det\left(AA^{-1}\right) = \det(I) = 1$ and $\det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right)$. So $\det(A)\det\left(A^{-1}\right) = 1$ and therefore $\det\left(A^{-1}\right) = (\det(A))^{-1}$
8. A^{-1} exists $\iff \det(A) \neq 0$
9. If A is triangular, A^{-1} is nonsingular $\iff a_{ii} \neq 0$ for all $i = 1, \ldots, N$

   Proof: by #8, A is nonsingular $\iff \det(A) = 0$. By #3, $\det(A) \neq 0 \iff a_{ii} \neq 0$ for all $i = 1, \ldots, N$
10. $rank(A) = rank(A') = rank(AA') = rank(A'A)$ (this will be crucial in the derivation of $\hat{\beta}_{OLS}$)
11. $rank(AB) \leq \min\{rank(A), rank(B)\}$
12. "Multiplication by a nonsingular matrix preserves rank": if C, B nonsingular, then $rank(CAB) = rank(A)$

    Note that C and B must be square (since they are nonsingular), but A need not be.
13. If A is triangular, $rank(A) \leq \#\{a_{11} \neq 0\}$

14. If A is diagonal, $rank\,(A) = \#\{a_{ii} \neq 0\}$

15. If A and B are nonsingular, then AB is nonsingular with $(AB)^{-1} = B^{-1}A^{-1}$

    Proof: $ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I\,;\; B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I$

    Note: you cannot just blindly apply this rule whenever you want to calculate $(AB)^{-1}$: even when AB is nonsingular, A and B might be singular. For example, let X be an $N \times K$ matrix with $N > K$. Since $N \neq K$, we know X is singular, so X^{-1} does not exist. However, if X has full column rank, i.e. $rank\,(X) = K$, then $X'X$ will be a $K \times K$ matrix with rank K and therefore is invertible, i.e. $(X'X)^{-1}$exists, but you cannot say $(X'X)^{-1} = X^{-1}(X')^{-1}$ since these latter two do not exist.

16. $\left(A^{-1}\right)' = (A')^{-1}$

    Proof:

$$\begin{aligned} AA^{-1} &= I \\ \left(AA^{-1}\right)' &= I' \\ \left(A^{-1}\right)' A' &= I \\ (A')^{-1} &= \left(A^{-1}\right)' \end{aligned}$$

17. If A is symmetric and nonsingular, then A^{-1} is symmetric

    Proof: $\left(A^{-1}\right)' = (A')^{-1} = (A)^{-1} = A^{-1}$

18. If A is diagonal and nonsingular, then A^{-1} is diagonal with $\left(A^{-1}\right)_{ii} = \dfrac{1}{a_{ii}}$

2.6 Calculating Matrix Inverses

2.6.1 Simple Cases

If A^{-1} exists, then

$$\left(A^{-1}\right)_{ij} = (-1)^{i+j} \frac{\det\left(\tilde{A}_{ji}\right)}{\det(A)}$$

Note the i and j have switched places in $\tilde{A}_{ji}$

Application to the inverse of a 2x2 matrix: if

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

then

$$\begin{aligned} A^{-1} &= \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} \tilde{A}_{11} & -\tilde{A}_{21} \\ -\tilde{A}_{12} & \tilde{A}_{22} \end{pmatrix} \\ &= \frac{1}{\Delta} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \end{aligned}$$

where

$$\Delta = \det(A) = a_{11}a_{22} - a_{12}a_{21}$$

Knowing this has proved useful on past 14.383 exams.

2.6.2 Partitioned Matrices

Partitioning a matrix A: $A_{M\times N} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, where A_{11} is $M_1 \times N_1$, A_{12} is $M_1 \times N_2$, A_{12} is $M_1 \times N_2$, A_{22} is $M_2 \times N_2$ with $M_1 + M_2 = M$ and $N_1 + N_2 = N$.
$A' = \begin{bmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22} \end{bmatrix}$
Let A and B be conformably partitioned. Then

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix}$$

and

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$

An example: Let $X_{N\times K} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$. Then $X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}$

2.6.3 Inverses of Partitioned Matrices

There are two equivalent ways to compute the inverses of partitioned matrices; choose for convenience.

1. If $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ is nonsingular with A_{11} and A_{22} nonsingular, then

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}D^{-1}A_{11} & -A_{11}^{-1}A_{12}D^{-1} \\ -D^{-1}A_{21}A_{11} & D^{-1} \end{bmatrix}$$

   where

$$D = A_{22} - A_{21}A_{11}^{-1}A_{12}$$

   Note that if $A_{12} = 0_{M_1\times N_2}$ and $A_{21} = 0_{M_2\times N_1}$, then

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0_{M_1\times N_2} \\ 0_{M_2\times N_1} & A_{22}^{-1} \end{bmatrix}$$

2. If $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ is nonsingular with A_{11} and A_{22} nonsingular, then

$$A^{-1} = \begin{bmatrix} E^{-1} & -E^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}E^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}E^{-1}A_{12}A_{22}^{-1} \end{bmatrix}$$

   where

$$E = A_{11} - A_{12}A_{22}^{-1}A_{21}$$

Note that if $A_{12} = 0_{M_1 \times N_2}$ and $A_{21} = 0_{M_2 \times N_1}$, then

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0_{M_1 \times N_2} \\ 0_{M_2 \times N_1} & A_{22}^{-1} \end{bmatrix}$$

I doubt you will ever be asked to prove these or reproduce them from memory. It is important to remember that the formula simplifies greatly if the "off-diagonal blocks" (A_{12} and A_{21}) are zero and you *should* know the formula in this special case. An important application of these formulae is to computing the bias of OLS when explanatory variables are ommitted from your model. This also shows up in proving the conditions under which estimators are asymptotically efficient (i.e. attain the Cramér-Rao lower bound).

2.7 An Application of Partitioned Matrices: An Interesting Way to Multiply Matrices[2]

Let $X_{T\times K}$ be a matrix of N observations of K variables. Let x

$$x_t = \begin{pmatrix} x_{t1} \\ \vdots \\ x_{tK} \end{pmatrix} \in \mathbb{R}^K$$

be the tth observation in X. Then we can write X as a partitioned matrix:

$$X = \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix}$$

---

[2] This section was inspired by a handout for 14.385 written by Peter Hinrichs.

Result:[3]

$$X'X = \sum_{t=1}^{T} x_t x_t'$$

Proof:

$$\begin{aligned}
X'X &= \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix}' \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix} \\
&= \begin{pmatrix} x_1 & \cdots & x_T \end{pmatrix} \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix} && \text{note } x_t \text{ is a } K \times 1 \text{ column vector} \\
&= \textstyle\sum_{t=1}^{T} x_t x_t' && \text{multiplying partitioned matrices}
\end{aligned}$$

Generalization: Suppose $X_{T\times M}$ with tth observation $x_t \in \mathbb{R}^M$ and $Z_{T\times K}$ with tth observation $z_t \in \mathbb{R}^K$. Then:

$$X'Z = \sum_{t=1}^{T} x_t z_t' \text{ and } Z'X = \sum_{t=1}^{T} z_t x_t'$$

Proof:

$$\begin{aligned}
X'Z &= \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix}' \begin{pmatrix} z_1' \\ \vdots \\ z_T' \end{pmatrix} \\
&= \begin{pmatrix} x_1 & \cdots & x_T \end{pmatrix} \begin{pmatrix} z_1' \\ \vdots \\ z_T' \end{pmatrix} && \text{note } x_t \text{ is } M \times 1,\ z_t' \text{ is } 1 \times K \\
&= \textstyle\sum_{t=1}^{T} x_t z_t' && \text{multiplying partitioned matrices}
\end{aligned}$$

---

[3] A caution: you may see this result written as $X'X = \sum_{t=1}^{T} x_t' x_t$ and wonder what the hell is going on. Never fear, just think about what it would take for the dimensions of the LHS and the RHS to agree: Since $X'X$ is $K \times K$, we have to have $x_t' x_t$ a $K \times K$ matrix also. This means x_t must be $1 \times K$, and our mystery is solved: some texts write observations x_t as $1 \times K$ row vectors, which means that the tth row of X is x_t, not x_t' as I have it here. (This second way can be more convenient for writing simultaneous equations models in 14.383.)

$$\begin{aligned}
Z'X &= \begin{pmatrix} z_1' \\ \vdots \\ z_T' \end{pmatrix}' \begin{pmatrix} x_1' \\ \vdots \\ x_T' \end{pmatrix} \\
&= \begin{pmatrix} z_1 & \cdots & z_T \end{pmatrix} \begin{pmatrix} z_1' \\ \vdots \\ z_T' \end{pmatrix} && \text{note } z_t \text{ is } K \times 1,\ x_t' \text{ is } 1 \times M \\
&= \textstyle\sum_{t=1}^T z_t x_t' && \text{multiplying partitioned matrices}
\end{aligned}$$

Note that $X'Z$ is $M \times K$ and $Z'X$ is $K \times M$.

Application: Suppose we have just one "left-hand-side" variable, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} \in \mathbb{R}^T$. Then by the previous results

$$X'y = \sum_{t=1}^T x_t y_t' = x_t y_t$$

where the last equality holds because y_t is a scalar. Suppose we are interested in computing $(X'X)^{-1} X'y$. By our results, we have

$$(X'X)^{-1} X'y = \left(\sum_{t=1}^T x_t x_t' \right)^{-1} \left(\sum_{t=1}^T x_t' y_t \right)$$

We can multiply this by $\dfrac{N}{N}$ to obtain

$$(X'X)^{-1} X'y = \left(\frac{1}{N} \sum_{t=1}^T x_t x_t' \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^T x_t' y_t \right)$$

If x_t is a scalar, we have

$$(X'X)^{-1} X'y = \left(\frac{1}{N} \sum_{t=1}^T x_t^2 \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^T x_t y_t \right)$$

If x_t is a scalar and we have “de-meaned” our data, we have

$$(X'X)^{-1} X'y = \left(\frac{1}{N}\sum_{t=1}^{T} x_t^2\right)^{-1} \left(\frac{1}{N}\sum_{t=1}^{T} x_t y_t\right) = \frac{cov\left(x,y\right)}{var\left(x\right)}$$

This last representation is a convenient way to think about the computed regression coefficient in OLS.

2.8 Solution Sets and the Dimension Theorem

Note: it is not necessary to read this section is not necessary until the discussion of identification in 14.383. However, it does add some depth to the material in Section 2.1.

Given a general *nonhomogeneous* system of equations $A_{M\times N}x_{N\times 1} = b_{M\times 1} \neq 0$, how do we obtain the *solution set* $S = \left\{s \in \mathbb{R}^N : As = b\right\}$?

Let $Ax = 0$ be the corresponding *homogeneous* system, and let $H = \left\{h \in \mathbb{R}^N : Ah = 0\right\}$ be the solution set to the homogeneous system. Note that $H \neq \emptyset$, because $0_N \in H$ always, since $A0_N = 0_M$. We call this the “trivial solution.”

Claim: given a particular solution s to the nonhomogeneous system, s' is also a solution $\iff$ $s' = s + h$ for some $h \in H$

Proof: Sufficiency ($\Rightarrow$): given $s \in S$, spz $\exists s' \in S, s' \neq s$ and $\nexists h \in H$ such that $s' = s + h$. Let $d \in \mathbb{R}^N$ such that $s' = s + d$. Since $s' \neq s$, we know $d \neq 0$, and by hypothesis $d \notin H$. $As' = b$, and $As' = A\left(s + d\right) = As + Ad = b + Ad$. But $d \notin H, d \neq 0 \Rightarrow Ad \neq 0$. So $b + Ad \neq b$, and we have a contradiction. Necessity ($\Leftarrow$): given $s \in S$ and $h \in H$, let $s' = s + h$. Then $As' = A\left(s+h\right) = As + Ah = b + 0 = b$, so $s' \in S$.

Application: Suppose you have a nonhomogeneous system of equations $Ax = b$ and a solution vector s and you wish to show that s is the unique solution to the system. Then it is sufficient to show that $H = \{0\}$, i.e. that the corresponding homogeneous system has only the trivial solution. Then given a proposed solution s' such that $As' = b$, we have (by our claim above) that $s' = s + h = s + 0 = s$.

Theorem 1 (The Dimension Theorem) (Linear Algebra version) *Given vector spaces* V, W *and a linear transformation* $T : V \to W$*, let* $N(T) = \{v \in V : T(v) = 0\}$ *(note that* $N(T) \subseteq V$*) and let* $T(V) = \{w \in W : \exists\ v \in V \text{ s.t. } T(v) = w\}$ *(note that* $T(V) \subseteq W$*). We call* $T(V)$ *the "image of* T*" and* $N(T)$ *the "null space of* T*". Let* $rank(T) = \dim(T(V))$ *and* $null(T) = \dim(N(T))$*.* ***Then:*** $rank(T) + null(T) = \dim(V)$*.*

(Matrix Algebra version) *Given* $A_{M\times N}$ *(which is a linear transformation from* $\mathbb{R}^N$ *to* $\mathbb{R}^M$*),* $N = rank(A) + \dim(H)$*, where* $H = \left\{h \in \mathbb{R}^N : Ah = 0\right\}$ *as above.*

Notes on the matrix algebra version: this just says that the number of columns of A equals the rank plus the dimension of the null space. If we recall that the rank of a matrix equals the maximum number of linearly independent columns, we can rewrite the result as: $\dim(H) = N - rank(A) = N - \max\{\#\text{ of lin. indep. cols.}\}$

Application: If you have a solution s to $As = b$, to prove that s is the unique solution to the system, it is sufficient to show that $rank(A) = N$. Note that since $rank(A) \leq \min\{M, N\}$, if $M < N$, then you definately do *not* have a unique solution. If $M \geq N$, you have a chance.

3 Eigenvectors and Eigenvalues

Let A be an $N \times N$ matrix. A $N \times 1$ vector $x \neq 0$ is an *eigenvector* of A with corresponding *eigenvalue* $\lambda \in \mathbb{R}$ if

$$Ax = \lambda x$$

In econometrics, we are usually only concerned with the eigenvalues of symmetric matrices.

Spectral Decomposition: if A is an $N \times N$ symmetric matrix, A can be decomposed as follows:

$$A = C\Lambda C'$$

with

$$CC' = I = C'C$$

where

$$C = [c_1 \cdots c_N]$$

has N distinct eigenvectors of A as its columns, each of which has unit length (i.e. $c_i' c_i = 1$) and is orthogonal to the other columns of C (i.e. $c_i' c_j = 0$ for al $i \neq j$) and where

$$\Lambda = diag\left(\lambda_1, \ldots, \lambda_N\right)$$

is a diagonal matrix made up of eigenvectors of A, with the ith diagonal element corresponding to the ith column of C, i.e.

$$Ac_i = \lambda_i c_i$$

Note: the N eigenvectors are distinct; the corresponding eigenvalues are all real ($\lambda_i \in \mathbb{R}$ for $i = 1, \ldots, N$) but need not be.distinct, even if A is nonsingular.

Theorem 2 *If A is symmetric, $rank\left(A\right) = \#\left\{\lambda_{ii}^{A} \neq 0\right\}$*

In words: the rank of a symmetric matrix is the number of nonzero eigenvalues it has

Proof:

$$\begin{aligned} rank\left(A\right) &= rank\left(C\Lambda C'\right) && A \text{ is symmetric} \\ &= rank\left(\Lambda\right) && C, C' \text{ nonsingular} \\ &= \#\left\{\lambda_{ii}^{A} \neq 0\right\} && \Lambda \text{ is diagonal} \end{aligned}$$

Corollary 3 *A symmetric matrix is nonsingular $\Longleftrightarrow$ all its eigenvectors are nonzero*

Corollary 4 *The rank of any (not necessarily symmetric) matrix A equals the number of nonzero eigenvectors of $A'A$.*

Proof:

$$\begin{aligned} rank\left(A\right) &= rank\left(A'A\right) && 10 \\ &= \#\left\{\lambda_{ii}^{A'A} \neq 0\right\} && A'A \text{ is symmetric} \end{aligned}$$

Theorem 5 *The nonzero eigenvalues of AA' are the same as those of $A'A$*

Proof: To show that the sets of eigenvalues are the same, we must first we show that an arbitrary eigenvalue of AA' is also an eigenvalue of $A'A$, then we must show that an arbitrary eigenvalue of $A'A$ is also an eigenvalue of AA'. First, let λ be a nonzero eigenvalue of $AA'$with corresponding eigenvector x. We wish to show that λ is also an eigenvalue of $A'A$, i.e. that there is a vector y such that $A'Ay = \lambda y$.

$$\begin{aligned} AA'x &= \lambda x && \lambda \text{ an eigenvector of } AA' \\ A'AA'x &= A'\lambda x = \lambda A'x && \text{premultiply first line by } A' \\ A'Ay &= \lambda y && \text{let } y = A'x \end{aligned}$$

We need to verify that y is nonzero:

$$\begin{aligned} AA'x &= \lambda x && \\ \lambda x &\neq 0 && \lambda \neq 0 \text{ and } x \neq 0 \\ AA'x &\neq 0 && \\ A'x &\neq 0 && \text{if } A'x = 0, \text{ then } AA'x = 0 \\ y &\neq 0 && A'x = y \end{aligned}$$

Note that $y \neq x$ and in fact need not even be the same dimension (A need not be square, although $A'A$ and AA' will be).

Next, let λ be a nonzero eigenvalue of $A'A$ with corresponding eigenvector x. We wish to show that λ is also an eigenvalue of AA', i.e. that there is a vector y such that $AA'y = \lambda y$.

$$\begin{aligned} A'Ax &= \lambda x && \lambda \text{ an eigenvector of } A'A \\ AA'Ax &= A\lambda x = \lambda Ax && \text{premultiply first line by } A' \\ AA'y &= \lambda y && \text{let } y = Ax \end{aligned}$$

We need to verify that y is nonzero:

$$\begin{aligned} A'Ax &= \lambda x \\ \lambda x &\neq 0 && \lambda \neq 0 \text{ and } x \neq 0 \\ A'Ax &\neq 0 \\ Ax &\neq 0 && \text{if } Ax = 0 \text{, then } AA'x = 0 \\ y &\neq 0 && Ax = y \end{aligned}$$

Theorem 6 *The trace of a symmetric matrix equals the sum of its eigenvalues, i.e.*

$$\textit{if } A \textit{ is a symmetric } N \times N \textit{ matrix, } tr\,(A) = \sum_{i=1}^{N} \lambda_i$$

Proof:

$$\begin{aligned} tr\,(A) &= tr\,(C\Lambda C') && \text{since } A \text{ is symmetric, } A = C\Lambda C' \\ &= tr\,(C'C\Lambda) \\ &= tr\,(I\Lambda) && C'C = I \\ &= tr\,(\Lambda) \\ &= \textstyle\sum_{i=1}^{N} \lambda_i && \text{def. of } \Lambda \end{aligned}$$

Theorem 7 *The determinant of a symmetric matrix equals the product of its eigenvalues, i.e.*

$$\textit{if } A \textit{ is a symmetric } N \times N \textit{ matrix, } \det(A) = \prod_{i=1}^{N} \lambda_i$$

Proof:

$$\begin{aligned}
\det(A) &= \det(C\Lambda C') && \text{since } A \text{ is symmetric, } A = C\Lambda C' \\
&= \det(C)\det(\Lambda)\det(C') && \det(MN) = \det(M)\det(N) \\
&= \det(C)\det(C')\det(\Lambda) && \text{determinant is a scalar} \\
&= \det(CC')\det(\Lambda) && \det(MN) = \det(M)\det(N) \\
&= \det(I)\det(\Lambda) && CC' = I \\
&= \det(\Lambda) && \det(I) = 1 \\
&= \textstyle\prod_{i=1}^{N} \lambda_i && \Lambda \text{ is diagonal}
\end{aligned}$$

Theorem 8 *For any symmetric matrix A, the eigenvalues of $A^2 = AA$ are the squares of the eigenvalues of A and the eigenvectors are the same, i.e.*

$$\textit{if } A \textit{ is symmetric and } x \neq 0 \textit{ with } Ax = \lambda x, \textit{ then } A^2 x = \lambda^2 x$$

Proof:

$$\begin{aligned}
Ax &= \lambda x \\
A^2 x &= AAx \\
&= A\lambda x \\
&= \lambda A x && \lambda \text{ a scalar} \\
&= \lambda^2 x && Ax = \lambda x \\
A^2 x &= \lambda^2 x
\end{aligned}$$

Alternative proof:

$$\begin{aligned}
A^2 &= (C\Lambda C')(C\Lambda C') && A = C\Lambda C' \\
&= C\Lambda C' C\Lambda C' \\
&= C\Lambda\Lambda C' && C'C = I \\
&= C\left(\Lambda^2\right)C' \\
\left(\Lambda^2\right)_{ii} &= (\Lambda_{ii})^2 && \Lambda \text{ is diagonal} \\
\lambda_i^{A^2} &= \left(\lambda_i^{A}\right)^2
\end{aligned}$$

Corollary 9 *Natural powers of symmetric matrices: if A is symmetric, $A^p = C\Lambda^p C'$ for $p \in \mathbb{N}$*

Proof:

$$\begin{aligned}
A^p &= AA\cdots A \\
&= (C\Lambda C')\,(C\Lambda C')\cdots(C\Lambda C') \\
&= C\Lambda C'C\Lambda C'\cdots C\Lambda C' \\
&= C\Lambda\Lambda\cdots\Lambda C' & C'C = I \\
&= C\Lambda^p C^{'}
\end{aligned}$$

Theorem 10 *For any nonsingular symmetric matrix A, the eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A and the eigenvectors are the same, i.e.*

$$\textit{if } A \textit{ is nonsingular and symmetric and } Ax = \lambda x \textit{ for some } x \neq 0, \textit{ then } A^{-1}x = \frac{1}{\lambda}x$$

Proof:

$$\begin{aligned}
A^{-1} &= (C\Lambda C')^{-1} \\
&= (C')^{-1}\Lambda^{-1}C^{-1} \\
&= \left(C^{-1}\right)^{-1}\Lambda^{-1}C^{-1} & C' = C^{-1} \\
&= C\Lambda^{-1}C'
\end{aligned}$$

Since A^{-1} is symmetric, Λ^{-1} is the matrix of eigenvalues of A^{-1} and C is the matrix of eigenvectors. Since Λ is diagonal, $\left(\Lambda^{-1}\right)_{ii} = \dfrac{1}{\lambda_{ii}}$.

Corollary 11 *Integer powers of symmetric nonsingular matrices: If A is symmetric and nonsingular, $A^p = C\Lambda^p C'$ for all $p \in \mathbb{Z}$*

Proof: already proved for $p = 0, 1, 2, \ldots$. Proof for $p = -1, -2, \ldots$:

$$\begin{aligned}
A^p &= A^{-1}A^{-1}\cdots A^{-1} \\
&= \left(C\Lambda^{-1}C'\right)\left(C\Lambda^{-1}C'\right)\cdots\left(C\Lambda^{-1}C'\right) \\
&= C\Lambda^{-1}C'C\Lambda^{-1}C'\cdots C\Lambda^{-1}C' \\
&= C\Lambda^{-1}\Lambda^{-1}\cdots\Lambda^{-1}C' & C'C = I \\
&= C\Lambda^p C^{'}
\end{aligned}$$

Real-valued powers of symmetric, nonsingular matrices with strictly postitive eigenvalues.

Definition 12 *For any symmetric matrix A with strictly postive eigenvalues, we define $A^r = C\Lambda^r C'$ for all $r \in \mathbb{R}$. If A is symmetric with non-negative but not necessarily strictly positive, A^r is defined as above for all $r \in \mathbb{R}_+$*

Definition 13 *An $N \times N$ matrix A is **idempotent** if $A^2 = A$.*

Note that if $A^2 = A$, then $A^p = A$ for all $p = 3, 4, \ldots$

Theorem 14 *If A is a symmetric, idempotent matrix, then all the eigenvalues of A are 0 or 1*

Proof: Let λ be an eigenvalue of A with corresponding eigenvector $x \neq 0$. Then

$$\begin{aligned} Ax &= \lambda x \\ A^2x &= A\lambda x \\ Ax &= \lambda A x \\ Ax &= \lambda\lambda x \\ Ax &= \lambda^2 x \\ \lambda x &= \lambda^2 x \\ \lambda &= \lambda^2 \qquad \text{since } x \neq 0 \\ &\lambda = \lambda^2 \text{ only if } \lambda \in \{0,1\} \end{aligned}$$

Theorem 15 *The only full rank, symmetric, idempotent matrix is the identity matrix*

Proof: Let A be a full rank, symmetric, idempotent matrix. Since A is full rank and symmetric, A is invertible and the eigenvalues of A^{-1} are the inverses of the eigenvalues of A. Since A is symmetric and idempotent, the eigenvalues of A are either 0 or 1. But $\frac{1}{0}$ does not exist, so all eigenvalues of A must be 1. Since A is symmetric, $A = C\Lambda C'$ and since all eigenvalues are 1, $\Lambda = I$, so $A = CIC' = CC' = I$.

Alternate proof: Let A be a full rank, symmetric, idempotent matrix. A is full rank $\iff$ $\det(A) = 0$. Since A is symmetric, we have $\det(A) = \prod_{i=1}^{N} \lambda_i$. But $\det(A) = \prod_{i=1}^{N} \lambda_i \iff \lambda_i \neq 0$ for all i.

Since A is symmetric and idempotent, $\lambda_i \in \{0,1\}$ for all i, so we know $\lambda_i = 1$ for all i. Since A is symmetric, $A = C\Lambda C'$ and since all eigenvalues are 1, $\Lambda = I$, so $A = CIC' = CC' = I$.

Theorem 16 *The rank of a symmetric, idempotent matrix equals its trace, i.e.*

$$\textit{if } A \textit{ is symmetric and idempotent, } rank\,(A) = tr\,(A)$$

Proof:

$$\begin{aligned} rank\,(A) &= \#\{\lambda_{ii} \neq 0\} && \text{theorem 2 above} \\ tr\,(A) &= tr\,(C\Lambda C') \\ &= tr\,(C'C\Lambda) \\ &= tr\,(\Lambda) && C'C = I \\ &= \#\{\lambda_i \neq 0\} && \text{since } \lambda \in \{0,1\} \\ rank\,(A) &= tr\,(A) \end{aligned}$$

Theorem 17 *Cholesky Factorization: Any symmetric matrix A with positive eigenvalues can be written as the product of a lower triangular matrix L and its transpose (which is upper triangular) $L' = U$. That is,*

$$A = LL' = LU = U'U$$

It follows that

$$A^{-1} = (L')^{-1} L^{-1} = U^{-1}L^{-1} = U^{-1}(U')^{-1}$$

where $L^{-1}$is lower triangular and U^{-1} is upper triangular.

4 Quadratic Forms and Definite Matrices

A *quadratic form* in the $N \times N$ matrix A and the $N \times 1$ vector x is the scalar $x'Ax$

A is negative definite (ND) if $x'Ax < 0$ for all $x \neq 0$

A is positive definite (PD) if $x'Ax > 0$ for all $x \neq 0$

A is definite if A is PD or ND

A is negative semi-definite (NSD) if $x'Ax \leq 0$ for all x

A is positive semi-definite (PSD) if $x'Ax \geq 0$ for all x

Theorem 18 *If A is definite, A is nonsingular*

Proof (by contrapositive): suppose A is singular. Then $\exists x \neq 0$ such that $Ax = 0$. But then $x'Ax = 0$, so A is neither PD nor ND.

Theorem 19 *Let A be a symmetric matrix. Then*

$$\begin{aligned} A \text{ is PD (ND)} &\iff \lambda_i > 0 \ (< 0) \text{ for all eigenvalues } \lambda_i \text{ of } A \\ A \text{ is PSD (NSD)} &\iff \lambda_i \geq 0 \ (\leq 0) \text{ for all eigenvalues } \lambda_i \text{ of } A \end{aligned}$$

Theorem 20 *If A is symmetric,*

$$\begin{aligned} A \text{ is PD} &\iff \det(A) > 0 \\ A \text{ is ND} &\iff \det(A) < 0 \\ A \text{ is PSD} &\iff \det(A) \geq 0 \\ A \text{ is NSD} &\iff \det(A) \leq 0 \\ A \text{ is PD} &\iff A^{-1} \text{ is PD} \\ A \text{ is ND} &\iff A^{-1} \text{ is ND} \end{aligned}$$

If A is symmetric, A is PD or ND $\iff$ A is nonsingular

The identity matrix I is PD

Every symmetric, idempotent matrix is PSD

Theorem 21 *For any matrix A, $A'A$ is PSD. If A is nonsingular, $A'A$ is PD.*

Proof: $x'A'Ax = (Ax)'(Ax) = y'y \geq 0$. If A is PD and $x \neq 0$, then $y = Ax \neq 0$, so $y'y > 0$ and therefore A

Theorem 22 *If A is an $N \times K$ matrix with $N > K$ and $rank\,(A) = K$, then $A'A$ is PD and AA' is PSD*

Proof: $rank\,(A) = K \Rightarrow$columns of A are linearly independent$\Rightarrow$ $Ax \neq 0$ for all nonzero $K \times 1$ vectors x. To prove $A'A$ is PD, we need to show that $x'A'Ax > 0$ for all $K \times 1$ nonzero vectors . $x'A'Ax = (Ax)' Ax > 0$ since $Ax \neq 0$. Similarly, $y'AA'y = (A'y)' A'y \geq 0$ for all $N \times 1$ nonzero vectors y. However, we do not have strict positivity, because $rank\,(A) = K < N$, so there exist $y \neq 0$ such that $A'y = 0$.

Theorem 23 *If A is PD and B is nonsingular, then $B'AB$ is PD*

Proof: to prove: $x'B'ABx > 0$ for all $x \neq 0$

$$\begin{aligned}
x'B'ABx &= (Bx)' A (Bx) \\
&= y'Ay \qquad y = Bx \neq 0 \text{ since } B \text{ is nonsingular and } x \neq 0 \\
&> 0 \qquad \text{since } A \text{ is PD}
\end{aligned}$$

Definition 24 *An $N \times N$ matrix A is "bigger than" ("no smaller than") another $N \times N$ matrix B if and only if $A - B$ PD (PSD)*

Notation 25 $A > (\geq) B \iff A - B$ *is PD (PSD)* $\iff x'(A - B)x > (\geq) 0$ *for all* $x \neq 0$

Notation 26 *A is PD (PSD) is often written shorthand as $A > 0$ ($A \geq 0$)*

Theorem 27 *Let A and B be symmetric, PD matrices. Then $A - B$ is PD (PSD) $\iff$ $B^{-1} - A^{-1}$ is PD (PSD)*

5 Special Symmetric Idempotent Matrices

Suppose X is an $N \times K$ matrix with $N > K$ and $rank\,(X) = K$. Since the $K \times K$ matrix $X'X$ has $rank\,(X'X) = rank\,(X) = K$, $X'X$ is full rank and therefore invertible and is PD. Therefore, the following $N \times N$ matrices are defined

$$P_X \equiv X (X'X)^{-1} X' \text{ and } Q_X \equiv I - P_X = I - X (X'X)^{-1} X'$$

Claim: P_X and Q_X are symmetric and idempotent (and therefore PSD) and $P_X X = X$, $Q_X X = 0$, $P_X Q_X = Q_X P_X = 0$

Proof:

Symmetry of P_X:

$$\begin{aligned} P_X' &= \left(X\left(X'X\right)^{-1}X'\right)' \\ &= \left(X'\right)'\left(\left(X'X\right)^{-1}\right)'\left(X\right)' \\ &= X\left(X'X\right)^{-1}X' \\ &= P_X \end{aligned}$$

Idempotence of P_X:

$$\begin{aligned} P_X^2 &= \left(X\left(X'X\right)^{-1}X'\right)\left(X\left(X'X\right)^{-1}X'\right) \\ &= X\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1}X \\ &= X\left(X'X\right)^{-1}X' && X'X\left(X'X\right)^{-1} = I \\ &= P_X \end{aligned}$$

Symmetry of Q_X:

$$\begin{aligned} Q_X' &= \left(I - X\left(X'X\right)^{-1}X'\right)' \\ &= I' - \left(X'\right)'\left(\left(X'X\right)^{-1}\right)'\left(X\right)' \\ &= I - X\left(X'X\right)^{-1}X' \\ &= Q_X \end{aligned}$$

Idempotence of Q_X:

$$\begin{aligned} Q_X^2 &= \left(I - P_X\right)\left(I - P_X\right) \\ &= I^2 - 2P_X^2 + P_X^2 \\ &= I - P_X \\ &= Q_X \end{aligned}$$

Since P_X and Q_X are symmetric and idempotent, they are both PSD. We will often use this (especially in 14.385) to prove that a matrix A is no smaller than another matrix B by transforming $A - B$ into a form similar either P_X or Q_X. This often occurs when we want to prove that one estimator $\hat{\theta}_0$ is more efficient than another estimator $\hat{\theta}_1$, which we do by showing that $V\left[\hat{\theta}_1\right]-V\left[\hat{\theta}_1\right]$ is PSD, or using the convenient notation, $V\left[\hat{\theta}_1\right]-V\left[\hat{\theta}_1\right]\geq 0$

$P_X X = X\left(X'X\right)^{-1}X'X = XI = X$

$Q_X X = \left(I - P_X\right)X = X - X = 0$

$P_X Q_X = P_X\left(I - P_X\right) = P_X - P_X = 0$

An application of these special idempotent matrices:

Let y be a $N \times 1$ vector. Note

$$y = Iy + P_X y - P_X y = P_X y + \left(I - P_X\right)y = P_X y + Q_X y = \hat{y} + \hat{\varepsilon}$$

where

$$\hat{y} = P_X y \in \mathbb{R}^N \text{ and } \hat{\varepsilon} = Q_X y \in \mathbb{R}^N$$

Interpretation of $\hat{y}$ and $\hat{\varepsilon}$:

$\hat{\varepsilon}'X = \left(Q_X y\right)'X = y'Q_X'X = y'Q_X X = y'0 = 0$

So $\hat{\varepsilon}$ is orthogonal to each column of X, since $\hat{\varepsilon}'X = \left[\hat{\varepsilon}'X_1 \cdots \hat{\varepsilon}'X_K\right] = [0 \cdots 0] = 0$. So $\hat{\varepsilon}$ is contained in the subspace of $\mathbb{R}^N$ which is *orthogonal* to the space spanned by the columns of X.

We can interpret Q_X as follows:

Q_X projects $N \times 1$ vectors onto the subspace of $\mathbb{R}^N$ which is orthogonal to the K-dimensional linear subspace spanned by the columns of X.

$\hat{y} = P_X y = X\left(X'X\right)^{-1}X'y = X\hat{\beta} = X_1\hat{\beta}_1 + \cdots + X_K\hat{\beta}_K$, where $\hat{\beta} = \left(X'X\right)^{-1}X'y$ is a $K \times 1$ vector and $X_1, \ldots, X_K$ are the columns of X. So $\hat{y}$ is a linear combination of the columns of $X$and therefore is contained in the linear subspace of $\mathbb{R}^N$spanned by the columns of X. We can interpret P_X as follows:

P_X projects $N \times 1$ vectors onto the K-dimensional linear supspace of $\mathbb{R}^N$ spanned by the columns of X.

$\hat{y}'\hat{\varepsilon} = \left(P_X X\right)'\left(Q_X X\right) = X'P_X Q_X X = 0$, so $\hat{y}$ and $\hat{\varepsilon}$ are orthogonal vectors in $\mathbb{R}^N$.

P_X and Q_X allow us to write y as the sum of two orthogonal components $\hat{y}$ and $\hat{\varepsilon}$, the first contained in the linear subspace of $\not{R}^N$ spanned by the columns of X (the "column space of X") and the second contained in the subspace orthogonal to the column space of X. We have split y into two components, one which can be written as a linear function of the columns of X and a second which cannot be.

An example:

Let i be an $N \times 1$ vector of ones. Define $P_i = i\left(i'i\right)^{-1} i' = \frac{1}{N} ii'$. Note that P_i is an $N \times N$ matrix with all entries equal to $\frac{1}{N}$, since ii' is an $N \times N$ matrix of ones. Define $Q_i = I - P_i = I - \frac{1}{N} ii'$. Note that Q_i is an $N \times N$ matrix with diagonal entries $1 - \frac{1}{N}$ and off-diagonal entries $-\frac{1}{N}$.

$P_i y = \frac{1}{N} ii'y = i \frac{\sum_{i=1}^{N} y_i}{N} = \overline{y} i$, which is a $N \times 1$ vector with all components equal to the mean of y

$Q_i y = \left(I - P_i\right) y = y - \overline{y} i$, which is an $N \times 1$ vector of deviations of y from its own mean.

Thus, the projection of y onto the 1-dimensional spaced spanned by i is its mean, $\overline{y}$, which the component orthogonal to this projection is the vector of deviations from this mean, $y - \overline{y} i$

Furthermore, $y'Q_i y = y'Q_i Q_i y = y'Q_i' Q_i y = \left(Q_i y\right)' Q_i y = \left(y - \overline{y} i\right)' \left(y - \overline{y} i\right) = \sum_{i=1}^{N} \left(y_i - \overline{y}\right)^2$, which is the sum of squared deviations from the mean.

6 Matrix Calculus

We will consider the most general case of the matrix derivative of a matrix-valued function of a matrix, then simplify to the cases we will use in applications.

Let $F : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}^P \times \mathbb{R}^Q$ be a matrix-valued function of a matrix. That is, given a real-valued $M \times N$ matrix X, $F\left(X\right)$ is the $P \times Q$ matrix

$$F\left(X\right) = \begin{bmatrix} f_{11}\left(X\right) & \cdots & f_{1Q}\left(X\right) \\ \vdots & & \vdots \\ f_{P1}\left(X\right) & \cdots & f_{PQ}\left(X\right) \end{bmatrix}$$

where each $f_{ij}\left(X\right)$ is a scalar-valued function of the matrix X, i.e. $f_{ij} : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}^1$ for all i, j.

Definition 28 *We define the derivative of the matrix-valued function $F\left(\right)$ with respect to its argu-*

ment, the matrix X*, as*

$$\frac{\partial F(X)}{\partial X} \equiv \begin{bmatrix} \frac{\partial F(X)}{\partial x_{11}} & \cdots & \frac{\partial F(X)}{\partial x_{1N}} \\ \vdots & & \vdots \\ \frac{\partial F(X)}{\partial x_{M1}} & \cdots & \frac{\partial F(X)}{\partial x_{MN}} \end{bmatrix}_{MP \times NQ}$$

where

$$\frac{\partial F(X)}{\partial x_{ij}} = \begin{bmatrix} \frac{f_{11}(X)}{\partial x_{ij}} & \cdots & \frac{f_{1Q}(X)}{\partial x_{ij}} \\ \vdots & & \vdots \\ \frac{f_{P1}(X)}{\partial x_{ij}} & \cdots & \frac{f_{PQ}(X)}{\partial x_{ij}} \end{bmatrix}_{P \times Q}$$

for all i, j*. Note that* $\frac{\partial F(X)}{\partial X}$ *has dimension* $MP \times NQ$*, because it is an* $M \times N$ *array of* $P \times Q$ *matrices (each* $\frac{\partial F(X)}{\partial x_{ij}}$ *is a* $P \times Q$ *matrix, and* $\frac{\partial F(X)}{\partial X}$ *is an* $M \times N$ *array of* $\frac{\partial F(X)}{\partial x_{ij}}$*).*

Example 29 *Matrix-valued function of a scalar:* $M = N = 1$

We have $F : \mathbb{R} \to \mathbb{R}^P \times \mathbb{R}^Q$, that is $F(x)$ is a $P \times Q$ matrix, while $x \in \mathbb{R}$ is a scalar. Applying our definition from above, we have

$$\frac{\partial F(x)}{\partial x} = \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x} & \cdots & \frac{\partial f_{1Q}(x)}{\partial x} \\ \vdots & & \vdots \\ \frac{\partial f_{P1}(x)}{\partial x} & \cdots & \frac{\partial f_{PQ}(x)}{\partial x} \end{bmatrix}_{P \times Q}$$

which is an $P \times Q$ matrix.

An application: if $F(x) = xI$, then

$$\frac{\partial F(x)}{\partial x} = \begin{bmatrix} \frac{\partial x}{\partial x} & \frac{\partial 0}{\partial x} & \cdots & \cdots & \frac{\partial 0}{\partial x} \\ \frac{\partial 0}{\partial x} & \frac{\partial x}{\partial x} & \frac{\partial 0}{\partial x} & \cdots & \frac{\partial 0}{\partial x} \\ \vdots & \frac{\partial 0}{\partial x} & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \frac{\partial 0}{\partial x} \\ \frac{\partial 0}{\partial x} & \cdots & \cdots & \frac{\partial 0}{\partial x} & \frac{\partial x}{\partial x} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} = I$$

So $\frac{\partial F(x)}{\partial x} = I$. As an application, suppose

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

is a random vector with $\varepsilon_i \sim ? \left(0, \sigma^2\right)$ for all i and $E[\varepsilon_i \varepsilon_j] = 0$ for all $i \neq j$ (or, even stronger, ε_i and ε_j are independent for all $i \neq j$. Then $V[\varepsilon] = E[\varepsilon\varepsilon'] = \sigma^2 I$, so $\frac{\partial V[\varepsilon]}{\partial \sigma^2} = I$.

Example 30 *Scalar-valued function of a matrix:* $P = Q = 1$

We have $f : \mathbb{R}^M \times \mathbb{R}^N \to \mathbb{R}$, that is $f(X)$ is a scalar, while $X \in \mathbb{R}^M \times \mathbb{R}^N$ is an $M \times N$ matrix. Applying our definition from above, we have

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \cdots & \frac{\partial f(x)}{\partial x_{1N}} \\ \vdots & & \vdots \\ \frac{\partial f(x)}{\partial x_{M1}} & \cdots & \frac{\partial f(x)}{\partial x_{MN}} \end{bmatrix}_{M \times N}$$

which is an $M \times N$ matrix

Example 31 *Scalar-valued function of a scalar:* $M = N = P = Q = 1$

We have $f : \mathbb{R} \to \mathbb{R}$, i.e. $f(x)$ and x are both scalars. Applying our definition from above, we have

$$\frac{\partial f\left(x\right)}{\partial x} = \left[\frac{\partial f\left(x\right)}{\partial x}\right] = f'\left(x\right)$$

which is just the standard derivative from univariate calculus.

Theorem 32 $\left(\dfrac{\partial F\left(X\right)}{\partial X}\right)' = \dfrac{\partial\left(F\left(X\right)\right)'}{\partial X'}$

Proof:

$$\begin{aligned}
\frac{\partial F\left(X\right)}{\partial X} &\equiv \begin{bmatrix} \frac{\partial F\left(X\right)}{\partial x_{11}} & \cdots & \frac{\partial F\left(X\right)}{\partial x_{1N}} \\ \vdots & & \vdots \\ \frac{\partial F\left(X\right)}{\partial x_{M1}} & \cdots & \frac{\partial F\left(X\right)}{\partial x_{MN}} \end{bmatrix} \\
\left(\frac{\partial F\left(X\right)}{\partial X}\right)' &= \begin{bmatrix} \frac{\partial F\left(X\right)}{\partial x_{11}} & \cdots & \frac{\partial F\left(X\right)}{\partial x_{1N}} \\ \vdots & & \vdots \\ \frac{\partial F\left(X\right)}{\partial x_{M1}} & \cdots & \frac{\partial F\left(X\right)}{\partial x_{MN}} \end{bmatrix}' \\
&= \begin{bmatrix} \left(\frac{\partial F\left(X\right)}{\partial x_{11}}\right)' & \cdots & \left(\frac{\partial F\left(X\right)}{\partial x_{M1}}\right)' \\ \vdots & & \vdots \\ \left(\frac{\partial F\left(X\right)}{\partial x_{1N}}\right)' & \cdots & \left(\frac{\partial F\left(X\right)}{\partial x_{MN}}\right)' \end{bmatrix} \quad \text{transpose of partitioned matrix}
\end{aligned}$$

Consider the matrices $\left(\dfrac{\partial F\left(X\right)}{\partial x_{ij}}\right)'$ individually:

$$\begin{aligned}
\left(\frac{\partial F\left(X\right)}{\partial x_{ij}}\right)' &= \begin{bmatrix} \frac{f_{11}\left(X\right)}{\partial x_{ij}} & \cdots & \frac{f_{1Q}\left(X\right)}{\partial x_{ij}} \\ \vdots & & \vdots \\ \frac{f_{P1}\left(X\right)}{\partial x_{ij}} & \cdots & \frac{f_{PQ}\left(X\right)}{\partial x_{ij}} \end{bmatrix}' \\
&= \begin{bmatrix} \frac{f_{11}\left(X\right)}{\partial x_{ij}} & \cdots & \frac{f_{P1}\left(X\right)}{\partial x_{ij}} \\ \vdots & & \vdots \\ \frac{f_{1Q}\left(X\right)}{\partial x_{ij}} & \cdots & \frac{f_{PQ}\left(X\right)}{\partial x_{ij}} \end{bmatrix} \\
&= \frac{\partial\left(F\left(X\right)\right)'}{\partial x_{ij}}
\end{aligned}$$

So we have

$$\begin{aligned}
\left(\frac{\partial F(X)}{\partial X}\right)' &= \begin{bmatrix} \left(\frac{\partial F(X)}{\partial x_{11}}\right)' & \cdots & \left(\frac{\partial F(X)}{\partial x_{M1}}\right)' \\ \vdots & & \vdots \\ \left(\frac{\partial F(X)}{\partial x_{1N}}\right)' & \cdots & \left(\frac{\partial F(X)}{\partial x_{MN}}\right)' \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial (F(X))'}{\partial x_{11}} & \cdots & \frac{\partial (F(X))'}{\partial x_{M1}} \\ \vdots & & \vdots \\ \frac{\partial (F(X))'}{\partial x_{N1}} & \cdots & \frac{\partial (F(X))'}{\partial x_{MN}} \end{bmatrix} \\
&= \frac{\partial (F(X))'}{\partial X'} \qquad \text{since } X' = \begin{pmatrix} x_{11} & \cdots & x_{M1} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{MN} \end{pmatrix}
\end{aligned}$$

Theorem 33 *If* $x, b \in \mathbb{R}^N$*, then*

$$\frac{\partial b'x}{\partial x} = b\,; \ \frac{\partial x'b}{\partial x} = b\,; \ \frac{\partial b'x}{\partial x'} = b\,; \ \frac{\partial x'b}{\partial x'} = b'$$

Proof: $\frac{\partial b'x}{\partial x} = b$

$$\begin{aligned}
b'x &= \textstyle\sum_{i=1}^N b_i x_i \\
\frac{\partial b'x}{\partial x} &= \frac{\partial \sum_{i=1}^N b_i x_i}{\partial x} \\
&= \begin{pmatrix} \frac{\partial \sum_{i=1}^N b_i x_i}{\partial x_1} \\ \vdots \\ \frac{\partial \sum_{i=1}^N b_i x_i}{\partial x_N} \end{pmatrix} \quad \text{by Def. 28} \\
&= \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} \\
&= b
\end{aligned}$$

$\dfrac{\partial x'b}{\partial x} = b$ follows immediately, since $x'b = b'x$

Proof: $\dfrac{\partial b'x}{\partial x'} = b'$

$$\begin{aligned}
\frac{\partial b'x}{\partial x'} &= \frac{\partial \sum_{i=1}^{N} b_i x_i}{\partial x'} \\
&= \begin{pmatrix} \frac{\partial \sum_{i=1}^{N} b_i x_i}{\partial x_1} & \cdots & \frac{\partial \sum_{i=1}^{N} b_i x_i}{\partial x_1} \end{pmatrix} \quad \text{by Def. 28} \\
&= \begin{pmatrix} b_1 & \cdots & b_N \end{pmatrix} \\
&= b'
\end{aligned}$$

$\dfrac{\partial x'b}{\partial x'} = b'$ follows immediately, since $x'b = b'x$

Theorem 34 *If $x \in \mathbb{R}^N$ and $B_{M\times N}$, then*

$$\frac{\partial Bx}{\partial x'} = B\,; \; \frac{\partial \left(Bx\right)'}{\partial x} = \frac{\partial x'B'}{\partial x} = B'$$

Proof: $\dfrac{\partial Bx}{\partial x'} = B$

$$\begin{aligned}
Bx &= \begin{pmatrix} b_{11}x_1 + \cdots b_{1N}x_N \\ \vdots \\ b_{M1}x_1 + \cdots b_{MN}x_N \end{pmatrix}_{M\times 1} \\
\frac{\partial Bx}{\partial x'} &= \frac{\partial}{\partial x'}\begin{pmatrix} b_{11}x_1 + \cdots b_{1N}x_N \\ \vdots \\ b_{M1}x_1 + \cdots b_{MN}x_N \end{pmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial x_1}(b_{11}x_1 + \cdots b_{1N}x_N) & \cdots & \frac{\partial}{\partial x_N}(b_{11}x_1 + \cdots b_{1N}x_N) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1}(b_{M1}x_1 + \cdots b_{MN}x_N) & \cdots & \frac{\partial}{\partial x_N}(b_{M1}x_1 + \cdots b_{MN}x_N) \end{bmatrix}_{N\times N} \\
&= \begin{bmatrix} b_{11} & \cdots & b_{1N} \\ \vdots & & \vdots \\ b_{M1} & \cdots & b_{MN} \end{bmatrix} \\
&= B
\end{aligned}$$

Proof: $\dfrac{\partial (Bx)'}{\partial x} = B'$: this can be proved as above by writing down $(Bx)'$ (a $1 \times N$ row vector) and taking the derivative with respect to $x \in \mathbb{R}^N$ or by applying Theorem 32. Recall that Theorem 32 states that $\left(\dfrac{\partial F(X)}{\partial X}\right)' = \dfrac{\partial (F(X))'}{\partial X'}$. Here, we take $X = x'$ (so $X' = x$) and $F(X) = Bx$. Since we just proved that $\dfrac{\partial Bx}{\partial x'} = B$, we have the following:

$$\begin{aligned}
\frac{\partial (Bx)'}{\partial x} &= \left(\frac{\partial Bx}{\partial x'}\right)' \quad \text{Thm. 32} \\
&= (B)' \\
&= B'
\end{aligned}$$

It can be useful to write this in the equivalent form

$$\frac{\partial x'B'}{\partial x} = B'$$

Theorem 35 *If* $x \in \mathbb{R}^N$ *and* $A_{N \times N}$*, then*

$$\begin{aligned} \frac{\partial x'Ax}{\partial x} &= (A + A')\,x\,;\ \frac{\partial x'A'x}{\partial x'} = x'\,(A + A') \\ \frac{\partial x'A'x}{\partial x} &= (A + A')\,x\,;\ \frac{\partial x'Ax}{\partial x} = x'\,(A + A') \end{aligned}$$

Proof of $\dfrac{\partial x'Ax}{\partial x} = (A + A')\,x$: First, note that

$$x'Ax \quad = \quad \textstyle\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_i x_j \qquad \text{Note: } x'Ax \in \mathbb{R}$$

The derivative is calculated as follows:

$$
\begin{aligned}
\frac{\partial x'Ax}{\partial x} &= \frac{\partial}{\partial x}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}x_ix_j\right) \\
&= \frac{\partial}{\partial x}\left(\sum_{i=1}^{N} a_{ii}x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} a_{ij}x_ix_j\right) \\
&= \begin{pmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{N} a_{ii}x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} a_{ij}x_ix_j\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\sum_{i=1}^{N} a_{ii}x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} a_{ij}x_ix_j\right) \end{pmatrix} \\
&= \begin{pmatrix} 2a_{11}x_1 + \sum_{j\neq 1} a_{1j}x_1x_j + \sum_{i\neq 1} a_{i1}x_ix_1 \\ \vdots \\ 2a_{NN}x_1 + \sum_{j\neq N} a_{Nj}x_Nx_j + \sum_{i\neq N} a_{iN}x_ix_N \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^{N} a_{1j}x_1 + \sum_{i=1}^{N} a_{i1}x_i \\ \vdots \\ \sum_{j=1}^{N} a_{Nj}x_N + \sum_{i=1}^{N} a_{iN}x_i \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^{N} (A)_{1j}\, x_j + \sum_{i=1}^{N} (A')_{1i}\, x_i \\ \vdots \\ \sum_{j=1}^{N} (A)_{Nj}\, x_j + \sum_{i=1}^{N} (A')_{Ni}\, x_i \end{pmatrix} \\
&= \begin{pmatrix} (Ax)_1 + (A'x)_1 \\ \vdots \\ (Ax)_N + (A'x)_N \end{pmatrix} \\
&= Ax + A'x \\
& \quad\ (A + A')\, x
\end{aligned}
$$

Proof of $\dfrac{\partial x' A' x}{\partial x'} = x' \left(A + A' \right)$

$$\begin{array}{rcl} x' A' x & = & \left(x' A x \right)' \\ \dfrac{\partial x' A' x}{\partial x'} & = & \dfrac{\partial \left(x' A x \right)'}{\partial x'} \\ & = & \left(\dfrac{\partial x' A x}{\partial x} \right)' \\ & = & \left(\left(A + A' \right) x \right)' \\ & = & x' \left(A + A' \right) \end{array}$$

Corollary 36 *If* $x \in \mathbb{R}^N$ *and* $A_{N \times N}$ *is symmetric, then*

$$\begin{array}{rcl} \dfrac{\partial x' A x}{\partial x} & = & 2Ax \, ; \, \dfrac{\partial x' A' x}{\partial x'} = 2x' A \\ \dfrac{\partial x' A' x}{\partial x} & = & 2Ax \, ; \, \dfrac{\partial x' A x}{\partial x'} = 2x' A \end{array}$$

Theorem 37 *If* $x \in \mathbb{R}^N$ *and* $A_{N \times N}$*, then*

$$\frac{\partial x' A x}{\partial A} = x x'$$

Proof:

$$\begin{array}{rcl} x' A x & = & \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_i x_j \\ \dfrac{\partial x' A x}{\partial a_{ij}} & = & x_i x_j \\ \dfrac{\partial x' A x}{\partial A} & = & \begin{pmatrix} \dfrac{\partial x' A x}{\partial a_{11}} & \cdots & \dfrac{\partial x' A x}{\partial a_{1N}} \\ \vdots & & \vdots \\ \dfrac{\partial x' A x}{\partial a_{N1}} & \cdots & \dfrac{\partial x' A x}{\partial a_{NN}} \end{pmatrix} \\ & = & \begin{pmatrix} x_1 x_1 & \cdots & x_1 x_N \\ \vdots & & \vdots \\ x_N x_1 & \cdots & x_N x_N \end{pmatrix} \\ & = & x x' \end{array}$$

Theorem 38 *If* $x \in \mathbb{R}^N$ *and* $A_{N \times N}$*, then*

$$\frac{\partial \det(A)}{\partial A} = \det(A) A^{-1'}$$

Consider case $N = 1$. $A = a\,;\ \det(A) = a\,;\ \dfrac{\partial \det(A)}{\partial A} = \dfrac{\partial a}{\partial a} = 1\,;\ \det(A) A^{-1'} = \dfrac{a}{a} = 1$

Consider case $N = 2$.

$$\begin{aligned}
A &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\
\det(A) &= a_{11}a_{22} - a_{12}a_{21} \\
\frac{\partial \det(A)}{\partial A} &= \frac{\partial(a_{11}a_{22} - a_{12}a_{21})}{\partial A} \\
&= \begin{pmatrix} \frac{\partial(a_{11}a_{22} - a_{12}a_{21})}{\partial a_{11}} & \frac{\partial(a_{11}a_{22} - a_{12}a_{21})}{\partial a_{12}} \\ \frac{\partial(a_{11}a_{22} - a_{12}a_{21})}{\partial a_{21}} & \frac{\partial(a_{11}a_{22} - a_{12}a_{21})}{\partial a_{22}} \end{pmatrix} \\
&= \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix} \\
A^{-1} &= \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \\
\det(A) A^{-1'} &= \frac{\det(A)}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}' \\
&= \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}
\end{aligned}$$

Theorem 39 *If* $x \in \mathbb{R}^N$ *and* $A_{N \times N}$*, then*

$$\frac{\partial \ln \det(A)}{\partial A} = A^{-1'}$$

Consider case $N = 1$. $A = a\,;\ \det(A) = a\,;\ \ln \det(A) = \ln a\,;\ \dfrac{\partial \ln \det(A)}{\partial A} = \dfrac{\partial \ln a}{\partial a} = \dfrac{1}{a}\,;\ A^{-1'} = \dfrac{1}{a}$

Consider case $N = 2$.

$$
\begin{aligned}
A &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\
\det(A) &= a_{11}a_{22} - a_{12}a_{21} \\
\frac{\partial \ln \det(A)}{\partial A} &= \frac{\partial \ln(a_{11}a_{22} - a_{12}a_{21})}{\partial A} \\
&= \begin{pmatrix} \frac{\partial(a_{11}a_{22} - a_{12}a_{21})/\partial a_{11}}{(a_{11}a_{22} - a_{12}a_{21})} & \frac{\partial(a_{11}a_{22} - a_{12}a_{21})/\partial a_{12}}{(a_{11}a_{22} - a_{12}a_{21})} \\ \frac{\partial(a_{11}a_{22} - a_{12}a_{21})/\partial a_{21}}{(a_{11}a_{22} - a_{12}a_{21})} & \frac{\partial(a_{11}a_{22} - a_{12}a_{21})/\partial a_{22}}{(a_{11}a_{22} - a_{12}a_{21})} \end{pmatrix} \\
&= \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix} \\
&= A^{-1'}
\end{aligned}
$$

7 Kronecker Products

Used in panel data models

If A is $M \times N$ and B is $P \times Q$ then

$$
A_{M\times N} \otimes B_{P\times Q} \equiv \begin{bmatrix} a_{11}B & \cdots & a_{N1}B \\ \vdots & & \vdots \\ a_{M1}B & \cdots & a_{MN}B \end{bmatrix}_{MP\times NQ}
$$

Properties of Kronecker products

1. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ assuming dimensions conformable
2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ provided A, B nonsingular
3. $(A \otimes B)' = A' \otimes B'$
4. If A, B are symmetric, then $A \otimes B$

5. $\det\left(A \otimes B\right) = \left(\det\left(A\right)\right)^{M} \left(\det\left(B\right)\right)^{N}$

6. $tr\left(A \otimes B\right) = tr\left(A\right) tr\left(B\right)$

References

Steven M. Bergantino. Handouts (various). Unpublished course notes for 14.382. Handout 1: Matrix Algebra; Handout 2: Distribution Theory; Handout 3: Statistical Inference. Year is approximate., 1998.

Stephen H. Friedberg, Arnold J. Isel, and Lawrence E. Spence. *Linear Algebra*. Prentice-Hall, 3rd edition, 1997.

William H. Greene. *Econometric Analysis*. Prentice-Hall, 5th edition, 2003. URL `http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm`.