Steven M Bergantino
14.382
Handout #1

Matrix Algebra

(1) Some Basic Results

(TO)

$(A+B)' = A' + B'$
$(AB)C = A(BC)$
$A(B+C) = AB + AC$
$(AB)' = B'A'$
$(A^{-1})' = (A')^{-1}$
$(AB)^{-1} = B^{-1}A^{-1}$ if B^{-1} and A^{-1} exist
$\left.\begin{array}{l} tr(cA) = c\,tr(A),\ c \in \mathbb{R} \\ tr(A+B) = tr(A) + tr(B) \end{array}\right\} \Rightarrow$ trace is a linear operator
$tr(A') = tr(A)$
$tr(AB) = tr(BA)$
$det(I) = 1.$
$det(cA) = c^n\, det(A)$ for A $n \times n$, $c \in \mathbb{R}$
$det(A') = det(A)$
$\boxed{det(AB) = det(A)\,det(B)}$
$det(A^{-1}) = [det(A)]^{-1}$
A^{-1} exists $\Leftrightarrow det(A) \neq 0$ in which case A is said to be <u>nonsingular</u>

!
$rank(A) = rank(A') = rank(A'A) = rank(AA')$
$rank(AB) \leq \min\{rank(A), rank(B)\}$
$rank(CAB) = rank(A)$ if C, B are nonsingular

A^{-1} exists $\Leftrightarrow rank(A) = n$ for A $n \times n$
where

- $tr(A) = \sum_{i=1}^{n} a_{ii}$ for A $n \times n$ — <u>trace</u> of a (square) matrix

- $det(A) = \begin{cases} a_{11}a_{22} - a_{12}a_{21} & A\ 2\times 2 \\ \sum_{j=1}^{n} a_{ij}(-1)^{i+j} det(A_{ij}) & A\ n \times n \end{cases}$ — <u>determinant</u> of a square matrix

where i is any row of A and A_{ij} is the matrix obtained by deleting the ith row and jth column of A

- $\text{rank}(A)$ = maximum # of linearly independent columns of A ← column rank
  = " rows " ← row rank

- Vectors $x_1, \ldots, x_k$ are linearly independent iff the only solution to the equation $c_1 x_1 + \ldots + c_k x_k = 0$, $c_j \in \mathbb{R}$ is $c_1 = c_2 = \ldots = c_k = 0$.

- $[A^{-1}]_{ij} = \dfrac{(-1)^{i+j} \det(A_{ji})}{\det(A)}$ is the i,j th element of A^{-1}

(2) Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix. Then the $n \times 1$ vector $c \neq 0$ is an eigenvector or characteristic vector of A w/ corresponding eigenvalue (characteristic root) λ if

$$Ac = \lambda c$$

Spectral Decomposition: Let A be an $n \times n$ symmetric matrix. Then A can be written as:

$$A = C \Lambda C' \qquad C'C = CC' = I$$

where $C = [c_1 \quad c_n]$ has n distinct eigenvectors of A as its columns, each of which has unit length ($c_i' c_i = 1$) and is orthogonal to the other columns in C ($c_i' c_j = 0 \quad i \neq j$) and where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix whose ith diagonal element is the eigenvalue λ_i corresponding to the ith column of C.

Note that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are all real ($\lambda_i \in \mathbb{R} \; \forall i$) but need not be distinct, even if A is nonsingular (eg I has $\lambda_i = 1 \; \forall i$).

(T1) Theorem: The rank of a symmetric matrix is the number of nonzero eigenvalues it contains

Proof:

$\text{rank}(A) = \text{rank}(C \Lambda C') = \text{rank}(\Lambda)$ = # of nonzero eigenvalues

← since C, C' nonsingular

(C1) <u>Corollary</u>: rank of any matrix is the number of nonzero eigenvalues in $A'A$ since $\text{rank}(A) = \text{rank}(A'A)$.

(T2) <u>Theorem</u> The nonzero eigenvalues of AA' are the same as those of $A'A$.

<u>Proof:</u>
Let λ be a nonzero eigenvalue of AA' with corresponding eigenvector x. Then $AA'x = \lambda x$ and since $\lambda \neq 0$ and $x \neq 0$, we know that $A'x \neq 0$ also. Premultiply by A' to get $A'AA'x = \lambda A'x$ or $A'Ay = \lambda y$ where $y = A'x \neq 0$ is now an eigenvector of $A'A$ with nonzero eigenvalue λ. We can use the same proof to show that $A'Ay = \lambda y$, $\lambda \neq 0$, $y \neq 0$ implies that $Ay \neq 0$ and hence $AA'(Ay) = \lambda Ay$ or $AA'x = \lambda x$ for $x = Ay \neq 0$. Thus every nonzero eigenvalue of AA' is a nonzero eigenvalue of $A'A$ and vice versa. //

(T3) <u>Theorem:</u> The trace of a symmetric matrix equals the sum of its eigenvalues

<u>proof:</u> $\text{tr}(A) = \text{tr}(C\Lambda C') = \text{tr}(C'C\Lambda) = \text{tr}(\Lambda) = \sum_{i=1}^{n} \lambda_i$

(T4) <u>Theorem:</u> The determinant of a symmetric matrix equals the product of its eigenvalues

$\det(AB) = \det(A)\det(B)$

<u>proof:</u> $\det(A) = \det(C\Lambda C') = \det(C)\det(\Lambda)\det(C') = \det(C')\det(C)\det(\Lambda)$
$= \det(C'C)\det(\Lambda) = \det(I)\det(\Lambda) = \det(\Lambda) = \prod_i^n \lambda_i$

(T5) <u>Theorem</u> For any symmetric matrix A, the eigenvalues of A^2 are the square of the eigenvalues of A and the eigenvectors are the same

<u>Proof:</u> $A^2 = C\Lambda \underbrace{C'C}_{I} \Lambda C' = C\Lambda^2 C'$ and $[\Lambda^2]_{ii} = \lambda_i^2$

(C5) <u>Corollary</u>: $A^p = C\Lambda^p C'$ $\quad p = 0, 1, 2, \ldots$

(T6) <u>Theorem:</u> <u>For any nonsingular symmetric matrix A, the eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A</u> and the eigenvectors are the same

<u>proof:</u> $A^{-1} = (C\Lambda C')^{-1} = C'^{-1}\Lambda^{-1}C^{-1} = C\Lambda^{-1}C'$ and $[\Lambda^{-1}]_{ii} = 1/\lambda_i$

$CC' = I$

(C6) Corollary: $\boxed{A^p = C\Lambda^p C'}$ $p \in \mathbb{Z} = \{\text{integers}\}$ for A symmetric, nonsingular

For any symmetric matrix A with strictly positive eigenvalues we define $A^r = C\Lambda^r C'$ $\forall r \in \mathbb{R}$. If A has nonnegative eigenvalues then A^r is defined similarly but only for $r \in \mathbb{R}^+$.

Let A be an $n \times n$ matrix. Then A is idempotent if $A^2 = A$. Note that this implies that $A^p = A$ for any $p = 1, 2, \ldots$

(T7) Theorem: The eigenvalues of any symmetric, idempotent matrix A are either 0 or 1.

proof:
Let λ be an eigenvalue of A with eigenvector x. Then by an earlier theorem λ^2 is an eigenvalue of A^2, also having eigenvector x. Thus, we have $\lambda x = Ax = A^2 x = \lambda^2 x \Rightarrow \lambda = \lambda^2$ since $x \neq 0$
$\Rightarrow \lambda \in \{0, 1\}$.

(T8) Theorem: The only full rank, symmetric, idempotent matrix is the identity matrix.

proof.
Let A be an $n \times n$ symmetric, idempotent matrix. Then
$\operatorname{rank}(A) = n \Leftrightarrow \det(A) \neq 0$. But from an earlier theorem
$\det(A) = \prod_{i=1}^{n} \lambda_i \neq 0 \Leftrightarrow \lambda_i = 1 \; \forall i \Leftrightarrow \Lambda = I$. Hence
$\operatorname{rank}(A) = n \Leftrightarrow A = CIC' = CC' = I$.

(C8) Corollary: All symmetric idempotent matrices except the identity matrix are singular

(T9) Theorem: The rank of a symmetric idempotent matrix equals its trace.

rank A = tr A $\quad$ A-sym, idempotent

proof:
$\underbrace{tr(A)}_{= tr(\Lambda)} = \sum_{i=1}^{n} \lambda_i$ = # of nonzero eigenvalues of A (since $\lambda_i \in \{0,1\}$) = $\operatorname{rank}(A)$

$= \operatorname{rank}(C\Lambda C') = \operatorname{rank} \Lambda$

$\Lambda = \begin{bmatrix} \lambda_i & 0 \\ 0 & \ddots \end{bmatrix}$

(T10) 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) $L' = U$.

$$A = LL' = LU = U'U \quad \Rightarrow \quad A^{-1} = (L')^{-1}L^{-1} = U^{-1}L^{-1} = U^{-1}(U')^{-1}$$

where L^{-1} is lower triangular and U^{-1} is upper triangular (check for 2x2 case).

(3) Quadratic Forms and Definite Matrices

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

A is	Negative Definite (ND)	if $x'Ax$	< 0	$\forall x \neq 0$
"	Negative Semi-Definite (NSD)	"	≤ 0	"
"	Positive Semi-Definite (PSD)	"	≥ 0	"
"	Positive Definite (PD)	"	> 0	"

(T11) Theorem: Let A be a symmetric matrix. Then
(1) A is PD (ND) $\Leftrightarrow$ $\lambda_i > 0$ (< 0) $\forall$ eigenvalues λ_i of A
(2) A is PSD (NSD) $\Leftrightarrow$ $\lambda_i \geq 0$ (≤ 0) $\forall$ eigenvalues λ_i of A

Proof:

where $A = C\Lambda C'$

$\Rightarrow$

For each $i = 1, \dots, n$ let $x^i = Ce^i$ where $e^i_i = 1$ and $e^i_j = 0$ $j \neq i$. Then $\forall i$, $x^{i\prime}Ax^i = e^{i\prime}C'C\Lambda C'Ce^i = e^{i\prime}\Lambda e^i = \lambda_i$ and so we have A PD(PSD, NSD, ND) $\Rightarrow$ $0 < (\leq, \geq, >)$ $x^{i\prime}Ax^i = \lambda_i$ $\forall i$.

$\Leftarrow$

Suppose that $\lambda_i > 0$ and $\lambda_j < 0$ can't be p.d. for some i, j and wlog assume $i = 1$, $j = 2$. Since C' is nonsingular we can find an $x \neq 0$ such that $C'x = (y_1, y_2, 0, \dots, 0)'$. For any choice of y_1 and y_2 such that $y \neq 0$ (Specifically, $x = Cy \neq 0$ since C nonsingular and $y \neq 0$), in which case $x'Ax = \lambda_1 y_1^2 + \lambda_2 y_2^2$, which can be made positive or negative by choosing (y_1, y_2) appropriately. Thus, if $\{\lambda\}$ violate any of the conditions of the theorem it is impossible for $x'Ax$ to be of one sign $\forall$ $x \neq 0$. //

$\forall x \neq 0$ $y = C'x \neq 0$
$x'Ax = x'C\Lambda C'x = y'\Lambda y = \sum_i \lambda_i y_i^2$

if $\lambda_i > 0 \Rightarrow x'Ax > 0$
$\lambda_i \geq 0 \Rightarrow \geq 0$
$\lambda_i < 0 \Rightarrow < 0$

Some More Results

$\prod \lambda_i$

(T12)
- A symmetric PD (PSD, NSD, ND) $\Rightarrow$ $\det(A) > 0$ (≥ 0, ≤ 0, < 0)
- A symmetric PD (ND) $\Leftrightarrow$ A^{-1} symmetric PD (ND) $\left(\lambda_i^{-1} = \frac{1}{\lambda_i}\right)$
- The identity matrix is PD
- ! Every symmetric, idempotent matrix is PSD $\lambda_i \geq 0$

(T13) Theorem: If A is $n \times k$ w/ $n > k$ and $\text{rank}(A) = k$ then $\underset{k \times k}{A'A}$ is PD and $\underset{n \times n}{AA'}$ is PSD.

Proof:
$\text{rank}(A) = k \Rightarrow$ Columns of A are linearly independent $\Rightarrow Ax \neq 0$ if $x \neq 0$ (Ax is a linear combination of the columns of A). Thus, we have $x'A'Ax = (Ax)'(Ax) > 0$ since $Ax \neq 0$. Also, we see that $y'AA'y = (A'y)'(A'y) \geq 0$ since $\text{rank}(A') = \text{rank}(A) = k < n$ = # of columns in A' implies that $Ay = 0$ for some $y \neq 0$.

(T14) Theorem: If A is PD and B is nonsingular then $B'AB$ is PD

Proof:
$x'B'ABx = (Bx)'A(Bx) > 0$ since $x \neq 0$ and B nonsingular $\Rightarrow Bx \neq 0$ and since A is PD.

We say that a square matrix A is bigger than another matrix B of the same dimension if $A - B$ is PSD. Notationally we write

$$A > (\geq B) \Leftrightarrow A - B \text{ is PD (PSD)} \Leftrightarrow x'Ax > (\geq)\, x'Bx \quad \forall x \neq 0.$$

(T15) Theorem: Let A and B be symmetric, PD matrices. Then $A - B$ is PD (PSD) $\Leftrightarrow$ $B^{-1} - A^{-1}$ is PD (PSD) (Sorry, no proof)

Davidson – MacKinnon Appendix

(4) Kronecker Products: (will show up when we do panel data models)

$A: m \times n$

$B: p \times q$

$$A \otimes B \equiv \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix} : mp \times nq.$$

So $m\times n \otimes p \times q = mp \times nq$. Some properties of Kronecker products are:

(T1.6)
$$\left\{ \begin{array}{ll} (A\otimes B)(C\otimes D) = (AC)\otimes(BD) & \text{assuming conformable dimensions} \\ (A\otimes B)^{-1} = A^{-1}\otimes B^{-1} & \text{for } A, B \text{ invertible} \\ (A\otimes B)' = A'\otimes B' & \\ (A\otimes B) \text{ is symmetric when } A \text{ and } B \text{ are symmetric} & \\ \det(\underset{m\times m}{A}\otimes \underset{p\times p}{B}) = [\det(A)]^m[\det(B)]^p & \text{for } A\ m\times m \text{ and } B\ p\times p \\ \operatorname{tr}(A\otimes B) = \operatorname{tr}(A)\operatorname{tr}(B) & \text{(not true for "normal" product)} \end{array} \right.$$

(5) Some Special Symmetric Idempotent Matrices

Let X be an $n\times k$ matrix with $n>k$ and $\operatorname{rank}(X)=k$. Since the $k\times k$ matrix $\underset{k\times k}{X'X}$ has $\operatorname{rank}(X'X) = \operatorname{rank}(X) = k$, we know that it is invertible. In fact we know from an earlier theorem that it is PD. We can therefore define the following matrices.

$$P \equiv X(X'X)^{-1}X' \quad \text{and} \quad Q \equiv I - X(X'X)^{-1}X' = I - P$$

Claim: P and Q are symmetric, idempotent $\Rightarrow$ psd and $PX = X$, $QX = 0$, $PQ = QP = 0$

proof:

Symmetry of P is obvious

$P^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = P$

$Q' = (I-P)' = I - P' = I - P = Q$

$Q^2 = (I-P)(I-P) = I - P - P + P^2 = I - P = Q$

$PX = X(X'X)^{-1}X'X = X$

$QX = (I-P)X = X - PX = X - X = 0$

$PQ = P(I-P) = P - P^2 = 0 = P - P^2 = (I-P)P = QP$ //

Now let y be an $n \times 1$ vector and observe that we can write

$$y = Py + (I-P)y = Py + Qy = \hat{y} + \hat{\varepsilon} \qquad \hat{y} = Py, \ \hat{\varepsilon} = Qy$$

How should we interpret $\hat{y}$ and $\hat{\varepsilon}$? To help with this observe the following:

- $\hat{\varepsilon}'X = (Qy)'X = y'\cancel{QX}^{0} = 0 \Rightarrow \hat{\varepsilon}$ is orthogonal to each column of X (since $\hat{\varepsilon}'X = [\hat{\varepsilon}'x_1 \ \cdot \ \hat{\varepsilon}'x_k]$ where x_j is the jth column of X) and so is contained in the subspace of $\mathbb{R}^n$ which is orthogonal to the space spanned by the columns of X. This offers us the following interpretation of Q

Q 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.

- $\hat{y} = Py = X(X'X)^{-1}X'y = Xb = x_1b_1 + \cdots x_kb_k$ where $b = (X'X)^{-1}X'y$ $(k \times 1)$ and again x_j is the jth column of X. Thus, $\hat{y}$ is just a linear combination of the columns of X and so is contained in the linear subspace of $\mathbb{R}^n$ spanned by those columns. Hence

P projects $n \times 1$ vectors onto the k-dimensional linear subspace of $\mathbb{R}^n$ spanned by the columns of X.

- $\hat{y}'\hat{\varepsilon} = (Py)'Qy = y'\cancel{PQ}^{0}y = 0 \Rightarrow \hat{y}$ and $\hat{\varepsilon}$ are orthogonal.

Thus, P and Q 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 $\mathbb{R}^n$ spanned by the columns of X and the second contained in the subspace orthogonal to that. In essence what we have done is divide y into a component which can be represented as a linear function of the columns of X ($\hat{y}$) and a component which cannot ($\hat{\varepsilon}$).

(E1) Ex:

Let i be an $n\times 1$ vector of 1's and define $P_0 = i(i'i)^{-1}i'$ $Q_0 = I - P_0$

$P_0 y = i(i'i)^{-1}i'y = i(1/n)\sum_i^n y_i = \bar{y}\, i$

$Q_0 y = y - P_0 y = y - \bar{y}.i \leftarrow$ y in deviation from means form

Thus, the projection of y onto the 1-dimensional space spanned by i is its mean, $\bar{y}$, and the component orthogonal to that projection is the deviations from this mean, $y - \bar{y}\, i$. Further,

$y'Q_0 y = y'Q_0'Q_0 y = (Q_0 y)'(Q_0 y) = (y - \bar{y} i)'(y - \bar{y} i) = \sum_{i=1}^n (y_i - \bar{y})^2$ is the sum of squared deviations from the mean.

(6) Partitioned Matrices

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$ are partitioned matrices w/ partitions A_{ij} A_{ji}

Note: A_{ij} and B_{ij} are unrelated to the minors A_{ij} used in defining the determinant and inverse of a square matrix

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

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$ for conformably partitioned A, B

$$A' = \begin{bmatrix} A_{11}' & A_{21}' \\ A_{12}' & A_{22}' \end{bmatrix}$$

(E2) Ex:

Let X be an $n\times k$ matrix and let X_i' be the $1\times k$ vector representing the ith row of X and X_i its $k\times 1$ transpose. Then,

$$X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} \Rightarrow X'X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}' \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} = [x_1 \cdots x_n] \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} = \sum_{i=1}^n X_i X_i'$$

(E3) Ex:

Let X be $n\times k$ again but this time let X_1 be the $n\times k_1$ matrix containing the first k_1 columns of X and let X_2 be the $n \times k_2$ matrix containing the last k_2 columns of X where $k_1 < K$ and $k_1 + k_2 = K$. Then,

$$X = [X_1 \ X_2] \Rightarrow X'X = [X_1 \ X_2]'[X_1 \ X_2] = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} [X_1 \ X_2] = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}$$

(T 17) Theorem: Let A be an invertible matrix partitioned into 4 submatrices w/ A_{ii} invertible $i = 1, 2$. Then:

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} F_1 & -F_1 A_{12} A_{22}^{-1} \\ -F_2 A_{21} A_{11}^{-1} & F_2 \end{bmatrix} = \begin{bmatrix} F_1 & -A_{11}^{-1} A_{12} F_2 \\ -A_{22}^{-1} A_{21} F_1 & F_2 \end{bmatrix}$$

where $F_1 = [A_{11} - A_{12} A_{22}^{-1} A_{21}]^{-1}$ and $F_2 = [A_{22} - A_{21} A_{11}^{-1} A_{12}]^{-1}$

proof:
Let $B = A^{-1}$ be conformably partitioned. Then $BA = AB = I$ and we have

$F_2 A_{12} A_{22}^{-1} = A_{11}^{-1} A_{12} F_2$
$A_{11}^{-1} A_{22}^{-1} (A_{22} - A_{21} A_{11}^{-1}$ [illegible] $) = (A_{11} - A_{12} A_{22}^{-1} A_{21}) A_{22}^{-1} A_{12}$
$A_{12} - A_{12} A_{22}^{-1} A_{21} A_{11}^{-1} A_{12} =$ (wet)

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

(1) $B_{11} A_{11} + B_{12} A_{21} = I$
(2) $B_{11} A_{12} + B_{12} A_{22} = 0 \Rightarrow B_{12} = -B_{11} A_{12} A_{22}^{-1}$
(3) $B_{21} A_{11} + B_{22} A_{21} = 0 \Rightarrow B_{21} = -B_{22} A_{21} A_{11}^{-1}$
(4) $B_{21} A_{12} + B_{22} A_{22} = I$

$$(1),(2) \Rightarrow B_{11} A_{11} - B_{11} A_{12} A_{22}^{-1} A_{21} = I \Rightarrow B_{11}[A_{11} - A_{12} A_{22}^{-1} A_{21}] = I$$
$$\Rightarrow \boxed{\begin{aligned} &B_{11} = F_1 \equiv [A_{11} - A_{12} A_{22}^{-1} A_{21}]^{-1} \\ \Rightarrow\ &B_{12} = -F_1 A_{12} A_{22}^{-1} \end{aligned}}$$

$$(3),(4) \Rightarrow -B_{22} A_{21} A_{11}^{-1} A_{12} + B_{22} A_{22} = I \Rightarrow B_{22}[A_{22} - A_{21} A_{11}^{-1} A_{12}] = I$$
$$\Rightarrow \boxed{\begin{aligned} &B_{22} = F_2 \equiv [A_{22} - A_{21} A_{11}^{-1} A_{12}]^{-1} \\ &B_{21} = -F_2 A_{21} A_{11}^{-1} \end{aligned}}$$

This gives the first formula. The second is obtained by post multiplying by B using the same partition as above (this is conformable) and equality of the two matrices follows from the uniqueness of A^{-1}.

(E4) <u>Ex:</u> $X = [X_1 \;\; X_2]$

$$(X'X) = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \quad \text{which implies that} \Rightarrow$$

$$F_1 = [X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1]^{-1} = [X_1'(I - X_2(X_2'X_2)^{-1}X_2')X_1]^{-1} = (X_1'Q_2X_1)^{-1}$$

$$F_2 = [X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]^{-1} = [X_2'(I - \underline{X_1(X_1'X_1)^{-1}X_1'})X_2]^{-1} = (X_2'Q_1X_2)^{-1}$$

$$(X'X)^{-1} = \begin{bmatrix} (X_1'Q_2X_1)^{-1} & -(X_1'Q_2X_1)^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -(X_2'Q_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & (X_2'Q_1X_2)^{-1} \end{bmatrix} \quad \text{First Formula}$$

$$= \begin{bmatrix} (X_1'Q_2X_1)^{-1} & -(X_1'X_1)^{-1}X_1'X_2(X_2'Q_1X_2)^{-1} \\ -(X_2'X_2)^{-1}X_2'X_1(X_1'Q_2X_1)^{-1} & (X_2'Q_1X_2)^{-1} \end{bmatrix} \quad \text{Second Formula}$$

Now recall that the coefficient vector from our projection problem is $b = (X'X)^{-1}X'Y$ or, <u>using the first formula</u> (Note: this formula was derived by premultiplying $(X'X)$ by $(X'X)^{-1}$ and so will make terms simplify more easily when being substituted into a formula in which $(X'X)^{-1}$ premultiplies X' or $X'X$), we have:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix} = \begin{bmatrix} (X_1'Q_2X_1)^{-1} & -(X_1'Q_2X_1)^{-1}X_1'X_2(X_2'X_2)^{-1} \\ -(X_2'Q_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1} & (X_2'Q_1X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'Q_2X_1)^{-1}X_1'Y - (X_1'Q_2X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'Y \\ (X_2'Q_1X_2)^{-1}X_2'Y - (X_2'Q_1X_2)^{-1}X_2'X_1(X_1'X_1)^{-1}X_1'Y \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'Q_2X_1)^{-1}X_1'[I - X_2(X_2'X_2)^{-1}X_2']Y \\ (X_2'Q_1X_2)^{-1}X_2'[I - X_1(X_1'X_1)^{-1}X_1']Y \end{bmatrix} = \begin{bmatrix} (X_1'Q_2X_1)^{-1}X_1'Q_2Y \\ (X_2'Q_1X_2)^{-1}X_2'Q_1Y \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'Q_2'Q_2X_1)^{-1}X_1'Q_2'Q_2Y \\ (X_2'Q_1'Q_1X_2)^{-1}X_2'Q_1'Q_1Y \end{bmatrix} = \begin{bmatrix} [(Q_2X_1)'(Q_2X_1)]^{-1}(Q_2X_1)'(Q_2Y) \\ [(Q_1X_2)'(Q_1X_2)]^{-1}(Q_1X_2)'(Q_1'Y) \end{bmatrix}$$

Recalling our interpretation of Q from before, we see that b_1 is the same coefficient vector as would be obtained by taking the residual vector from the projection of Y on X_2 and projecting it onto the residual matrix from the projections of X_1 on X_2. b_2 can be interpreted analogously.

(7) Matrix Calculus

Let $F: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^p \times \mathbb{R}^q$ be a matrix-valued function. That is, given a real $m \times n$ matrix X, $F(X)$ is the $p \times q$ matrix:

$$\underset{(p\times q)}{F(X)} = \begin{bmatrix} f_{11}(X) & \cdots & f_{1q}(X) \\ \vdots & & \vdots \\ f_{p1}(X) & \cdots & f_{pq}(X) \end{bmatrix}_{(p\times q)} \qquad \underset{(m\times n)}{X}$$

where each $f_{ij}(X)$ is a scalar-valued function of the matrix X. Hence $f_{ij}: \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R} \;\; \forall ij$. We define the derivative of $F()$ wrt X as follows:

$$\frac{\partial F(X)}{\partial X} = \begin{bmatrix} \frac{\partial F(X)}{\partial x_{11}} & \cdots & \frac{\partial F(X)}{\partial x_{1n}} \\ \vdots & & \\ \frac{\partial F(X)}{\partial x_{m1}} & \cdots & \frac{\partial F(X)}{\partial x_{mn}} \end{bmatrix}_{(mp \times nq)}$$

Note: Each partition of this matrix is itself a $p \times q$ matrix (defined below)

where

$$\frac{\partial F(X)}{\partial X_{ij}} = \begin{bmatrix} \frac{\partial f_{11}(X)}{\partial X_{ij}} & \cdots & \frac{\partial f_{1q}(X)}{\partial X_{ij}} \\ \vdots & & \\ \frac{\partial f_{p1}(X)}{\partial X_{ij}} & \cdots & \frac{\partial f_{pq}(X)}{\partial X_{ij}} \end{bmatrix}_{(p\times q)} \qquad \forall \; i = 1, \ldots, m \quad j = 1, \ldots, n$$

(E5) Ex: Matrix-Valued Function of a Scalar
$m = n = 1 \Rightarrow X = x \in \mathbb{R}$

$$\frac{\partial F(x)}{\partial x} = \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x} & \cdots & \frac{\partial f_{1q}(x)}{\partial x} \\ \vdots & & \\ \frac{\partial f_{p1}(x)}{\partial x} & \cdots & \frac{\partial f_{pq}(x)}{\partial x} \end{bmatrix}$$

For example if $F(x) = xI$, then $\partial F/\partial x = I$ (replace $F(x)$ w/ $V(\varepsilon)$ and x w/ σ^2 and we just took the derivative of the variance-covariance matrix of a random vector ε wrt its common variance term σ^2).

(E6) Ex: Scalar-valued function of a matrix:
$p = q = 1 \Rightarrow F(X) = f(X) \in \mathbb{R}$

$$\frac{\partial F(X)}{\partial X} = \begin{bmatrix} \frac{\partial f(X)}{\partial x_{11}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \vdots & & \\ \frac{\partial f(X)}{\partial x_{m1}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{bmatrix} \quad (m \times n)$$

(E7) Ex: Scalar-Valued Function of a Scalar
$m = n = 1 \Rightarrow X = x \in \mathbb{R}$
$p = q = 1 \Rightarrow F(x) = f(x) \in \mathbb{R}$

$$\frac{\partial F(x)}{\partial x} = \frac{\partial f(x)}{\partial x} = f'(x)$$ ← Standard derivative from univariate calculus.

<u>Quick Update</u>

X: $m \times n$ $F(X)$: $p \times q$

x_{ij}: 1×1 $i = 1, .., m$, $j = 1, .., n$ $f_{kl}(X)$: 1×1 $k = 1, .., p$ $l = 1, .., q$

$\frac{\partial F(X)}{\partial X}$: $mp \times nq$ $\frac{\partial F(X)}{\partial x_{ij}}$: $p \times q$ $\frac{\partial f_{kl}(X)}{\partial x_{ij}}$: 1×1

$X_{m\times n}$

(T18) <u>Claim:</u> $\left(\dfrac{\partial F(x)}{\partial x}\right)' = \dfrac{\partial F(x)'}{\partial x'}$

<u>Proof:</u>

$$\left(\frac{\partial F(x)}{\partial x_{ij}}\right)' = \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x_{ij}} & \cdots & \frac{\partial f_{1q}(x)}{\partial x_{ij}} \\ \vdots & & \vdots \\ \frac{\partial f_{p1}(x)}{\partial x_{ij}} & \cdots & \frac{\partial f_{pq}(x)}{\partial x_{ij}} \end{bmatrix}' = \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x_{ij}} & \cdots & \frac{\partial f_{p1}(x)}{\partial x_{ij}} \\ \vdots & & \vdots \\ \frac{\partial f_{1q}(x)}{\partial x_{ij}} & \cdots & \frac{\partial f_{pq}(x)}{\partial x_{ij}} \end{bmatrix} = \frac{\partial F(x)'}{\partial x_{ij}}$$

where $F(x)'$ is the $q\times p$ transpose of $F(x)$. Thus, we have

$$\left(\frac{\partial F(x)}{\partial x}\right)' = \begin{bmatrix} \frac{\partial F(x)}{\partial x_{11}} & \cdots & \frac{\partial F(x)}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial F(x)}{\partial x_{m1}} & \cdots & \frac{\partial F(x)}{\partial x_{mn}} \end{bmatrix}' = \begin{bmatrix} \left(\frac{\partial F(x)}{\partial x_{11}}\right)' & \cdots & \left(\frac{\partial F(x)}{\partial x_{m1}}\right)' \\ \vdots & & \vdots \\ \left(\frac{\partial F(x)}{\partial x_{1n}}\right)' & \cdots & \left(\frac{\partial F(x)}{\partial x_{mn}}\right)' \end{bmatrix}$$

using the formula for the transpose of a partitioned matrix.

$$= \begin{bmatrix} \frac{\partial F(x)'}{\partial x_{11}} & \cdots & \frac{\partial F(x)'}{\partial x_{m1}} \\ \vdots & & \vdots \\ \frac{\partial F(x)'}{\partial x_{1n}} & \cdots & \frac{\partial F(x)'}{\partial x_{mn}} \end{bmatrix} = \frac{\partial F(x)'}{\partial x'} \quad //$$

(T19) <u>Theorem:</u> Given $x: n\times 1$, $b: n\times 1$, $B: m\times n$, $A: n\times n$ we have

(1) $\dfrac{\partial b'x}{\partial x} = b \Rightarrow \dfrac{\partial b'x}{\partial x'} = b'$ $\quad \boxed{\dfrac{\partial (x'B)}{\partial x} = B}$

$x'B = (x'B_1, \ldots, x'B_n)$

(2) $\dfrac{\partial Bx}{\partial x'} = B \Rightarrow \dfrac{\partial (Bx)'}{\partial x} = B'$ $\quad \boxed{\dfrac{\partial x'B}{\partial x} = B}$

(3) $\boxed{\dfrac{\partial x'Ax}{\partial x} = \underbrace{(A+A')}_{\substack{\shortparallel \\ 2A}}x} \Rightarrow \dfrac{\partial x'A'x}{\partial x'} = x'\underbrace{(A+A')}_{\substack{\shortparallel \\ 2A}}$

$2A \leftarrow$ if A is symmetric $\rightarrow 2A$

<u>Note:</u> (3) holds replacing A w/ A' and vice versa in $x'Ax$, $x'A'x$

$$\left(\frac{\partial (x'Ax)}{\partial x_i}\right) = \left(\frac{\partial \sum x_k A_{km} x_m}{\partial x_i}\right) = \left(\sum_m A_{im}x_m + \sum_k A_{ki}x_k\right) = Ax + A'x$$

$x \; (n \times 1)$

! (4) $$\boxed{\frac{\partial x'Ax}{\partial A} = xx'}$$

$$\frac{\partial \sum_{km} x_k A_{km} x_m}{\partial A_{ij}} \Rightarrow x_i x_j$$

$$= \begin{pmatrix} x_1 x_1 & \cdots & x_1 x_n \\ & & \\ x_n x_1 & \cdots & x_n x_n \end{pmatrix} = \cancel{xx}' = \begin{pmatrix} x_1 x' \\ \\ x_n x' \end{pmatrix} = xx'$$

(5) $$\boxed{\frac{\partial \det(A)}{\partial A} = \det(A)\, A^{-1\prime}}$$

$\det A = \sum_j a_{ij} (-1)^{i+j} \det(A_{ij})$ по i-ой строке → алгебраическое дополнение

$$\frac{\partial \det A}{\partial a_{ik}} = (-1)^{i+k} \det(A_{ik}) \Rightarrow \frac{(-1)^{i+k} \det(A')_{ki}}{\det A'} \cdot \det A \to (A^{-1})' \det A$$

! (6) $$\boxed{\frac{\partial \ln \det(A)}{\partial A} = A^{-1\prime}}$$

$$= \frac{1}{\det A} \frac{\partial \det A}{\partial A} = A^{-1\prime}$$

Proof:

(1) $$b'x = \sum_{i=1}^n b_i x_i \Rightarrow \frac{\partial b'x}{\partial x_j} = \frac{\partial \sum_{i=1}^n b_i x_i}{\partial x_j} = b_j \quad j = 1, \ldots, n$$

$$\Rightarrow \frac{\partial b'x}{\partial x} = \begin{bmatrix} \frac{\partial b'x}{\partial x_1} \\ \vdots \\ \frac{\partial b'x}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b \Rightarrow \frac{\partial b'x}{\partial x'} = \left(\frac{\partial b'x}{\partial x}\right)' = b'$$

since $(b'x)' = x'b = b'x$

(2) Notice that, using our results from partitioned matrices, we can write Bx as:

$$Bx = [B_1 \cdots B_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n B_i x_i \Rightarrow \frac{\partial Bx}{\partial x_j} = \frac{\partial \sum_{i=1}^n B_i x_i}{\partial x_j} = B_j \quad j = 1, \ldots, n$$

where B_i is the ith column of B. Thus, we have

$$\frac{\partial Bx}{\partial x'} = \left[\frac{\partial Bx}{\partial x_1} \cdots \frac{\partial Bx}{\partial x_n} \right] = [B_1 \cdots B_n] = B \Rightarrow \frac{\partial (Bx)'}{\partial x} = \left(\frac{\partial Bx}{\partial x'}\right)' = B'$$

(3) $$x'Ax = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i=1}^n \sum_{j \neq i} a_{ij} x_i x_j$$

$$\frac{\partial x'Ax}{\partial x_k} = 2 a_{kk} x_k + \sum_{j \neq k} a_{kj} x_j + \sum_{i \neq k} a_{ik} x_i$$

$$= \sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n a_{ik} x_i$$

$$= \sum_{j=1}^{n} [A]_{kj}\, x_j + \sum_{i=1}^{n} [A']_{ki}\, x_i = (Ax)_k + (A'x)_k$$

where $[A]_{kj} = a_{kj}$ is the element from the kth row and jth column of A while $[A']_{ki} = a_{ik}$ is the element from the kth row and ith column of A', and where $(Ax)_k$ and $(A'x)_k$ are the kth elements from the $n\times 1$ vectors Ax and A'x. Thus, we have

$$\frac{\partial x'Ax}{\partial x} = \begin{bmatrix} \frac{\partial x'Ax}{\partial x_1} \\ \vdots \\ \frac{\partial x'Ax}{\partial x_n} \end{bmatrix} = \begin{bmatrix} (Ax)_1 + (A'x)_1 \\ \vdots \\ (Ax)_n + (A'x)_n \end{bmatrix} = \underbrace{\begin{bmatrix} (Ax)_1 \\ \vdots \\ (Ax)_n \end{bmatrix}}_{Ax} + \underbrace{\begin{bmatrix} (A'x)_1 \\ \vdots \\ (A'x)_n \end{bmatrix}}_{A'x} = (A+A')x$$

$$\Rightarrow \frac{\partial x'A'x}{\partial x'} = \left(\frac{\partial x'Ax}{\partial x}\right)' = x'(A+A')$$

Also, $\frac{\partial x'A'x}{\partial x} = (A+A')x$ and $\frac{\partial x'Ax}{\partial x'} = x'(A+A')$

(4) $x'Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j \Rightarrow \frac{\partial x'Ax}{\partial a_{ij}} = x_ix_j \quad \forall\, i,j = 1,..,n$

$$\Rightarrow \frac{\partial x'Ax}{\partial A} = \begin{bmatrix} \frac{\partial x'Ax}{\partial a_{11}} & \cdots & \frac{\partial x'Ax}{\partial a_{1n}} \\ \vdots & & \vdots \\ \frac{\partial x'Ax}{\partial a_{n1}} & \cdots & \frac{\partial x'Ax}{\partial a_{nn}} \end{bmatrix} = \begin{bmatrix} x_1x_1 & \cdots & x_1x_n \\ & & \\ x_nx_1 & \cdots & x_nx_n \end{bmatrix} = \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}}_{= xx'}$$

(5) Recall from the beginning of this handout (many pages ago) that

$$\det(A) = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(A_{ij}) \Rightarrow \frac{\partial \det(A)}{\partial a_{ij}} = (-1)^{i+j}\det(A_{ij}),$$

for any $i = 1,..,n$

where A_{ij} is the $(n-1)\times(n-1)$ matrix formed by deleting the ith row and jth colum from A. Recall also that