Massachusetts Institute of Technology Guido Kuersteiner
Department of Economics
14.381 Statistics

Lecture Notes on Matrix Algebra

These lecture notes summarize some basic results on matrix algebra used frequently in statistics and econometrics. Where necessary they also review some more general results on linear algebra. Starred (*) paragraphs are optional. The most important thing is to remember the calculus rules highlighted as **Results** and numbered throughout the notes for easy reference.

1 Basic Concepts

An $m \times n$ matrix A is a rectangular array of real numbers

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

The real numbers a_{ij} are called the elements of A. An $m \times 1$ matrix x is a point in $\mathbb{R}^m$. It is called a column vector. A $1 \times m$ matrix is called a row vector.

The sum of two matrices A, B of dimension $m \times n$ is obtained by forming the sum of all individual elements of A and B such that the i, j-th element of $(A+B)_{ij} = a_{ij} + b_{ij}$

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

The product of a matrix A with a scalar λ is

$$\lambda A = A\lambda$$

with i, j-th element $(\lambda A)_{ij} = \lambda a_{ij}$ for all i, j. The following properties follow easily

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

The null matrix is the matrix that has all elements equal to zero and is denoted by 0.

$$A + (-1)A = 0$$

The matrix product of A and B is defined when A is $m \times n$ and B is $n \times p$. The i, k^{th} element of AB is then given by

$$(AB)_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

and AB is an $m \times p$ matrix. The matrix product satisfies

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

Note that if AB exists, BA need not be defined and even if it exists if is usually not true that $AB = BA$.

Matrices are often used to represent linear maps. An $m \times n$ matrix maps $\mathbb{R}^n \to \mathbb{R}^m$. If $x \in \mathbb{R}^n$ then $AX = y \in \mathbb{R}^m$. The matrix product then represents a sequence of maps applied to x, i.e. $x \longmapsto Ax = y$ and $y \longmapsto By = BAx = z$.

The transpose of an $m \times n$ matrix A is denoted by A'. If A has elements a_{ij} then A' has elements a_{ji}. We have

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

2 The rank of a matrix

Definition 1 (Linear Independence) *Let $x_1 \dots x_n \in \mathbb{R}^n$ be vectors in $\mathbb{R}^n$. Then $x_1 \dots x_n$ are said to be linearly independent if there exists no set of scalars $\alpha_1 \dots \alpha_j$, not all zero, such that*

$$\sum_i \alpha_i x_i = 0.$$

If x_i are not linearly independent then they are said to be linearly dependent.

Definition 2 (Rank) *The column rank of an $m \times n$ matrix A is the largest number of linearly independent column vectors A contains. The row rank is the largest number of linearly independent row vectors A contains.*

It can be shown that the row rank and column rank are equal. The row and column rank of a matrix therefore is often referred to simply as the rank of the matrix. We denote the rank of A by $r(A)$.

Definition 3 (Column Space) *The column space of A is the set of vectors generated by the columns of A, denoted by*

$$\mathcal{M}(A) = \{y : y = Ax,\ x \in \mathbb{R}^n\}.$$

The dimension of $\mathcal{M}(A)$ is $r(A)$.

Definition 4 (Non-Singular) *A square matrix A of dimension $n \times n$ is said to be non singular if $r(A) = n$. A is said to be singular if $r(A) < n$.*

Definition 5 (Null Space) *The null space of A is the set of vectors x such that $Ax = 0$, denoted by $N(A)$.*

If A is non singular then $N(A) = 0$. The dimension of $N(A)$ is $n - r(A)$.

3 Inverses, Partitioned Matrices and Orthogonal Matrices

If A is non-singular then there exists a non-singular matrix A^{-1} such that

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

where

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

is the identity matrix.

The following facts about inverses are useful

Result 1 *i)* $(AB)^{-1} = B^{-1}A^{-1}$
ii)$(A')^{-1} = (A^{-1})'$

Proof. For i) note that $ABB^{-1}A^{-1} = AA^{-1} = I$ and $B^{-1}A^{-1}AB = I$. For ii) note that $(A'(A^{-1})')' = A^{-1}A = I$ and $((A^{-1})'A')' = AA^{-1} = I$ ■

Definition 6 *A matrix A is called orthogonal if $A^{-1} = A'$. Two vectors x and y are orthogonal if $x'y = 0$.*

It is sometimes useful to partition a matrix into submatrices. If A is $m \times n$ then

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

where A_{11} is $m_1 \times n_1$, A_{12} is $m_1 \times n_2$, A_{21} is $m_2 \times n_1$, and A_{22} is $m_2 \times n_2$ and $m_1 + m_2 = m$ and $n_1 + n_2 = n$. Then the sum of two matrices that are partitioned conformingly can be obtained as

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

where A_{ij} and B_{ij} for $i, j = 1, 2$ have to have the same dimension $m_i \times n_j$. The product of two conformable matrices A and B can be expressed in partitioned form if B_{11} is $n_1 \times p_1$, B_{12} is $n_1 \times p_2$, B_{21} is $n_2 \times p_1$ and B_{22} is $n_2 \times p_2$ such that

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

and the product is of dimension $m \times (p_1 + p_2)$.

Result 2 *If A is $n \times n$ and both A_{11} and A_{22} are square non singular and $A_{12} = A_{21} = 0$ then*

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

Result 3 *If A is non singular and $D = A_{22} - A_{21}A_{11}^{-1}A_{12}$ is non singular then*

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

Also if A is non singular and E is non singular where $E = A_{11} - A_{12}A_{22}^{-1}A_{21}$, then

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

(*) Proof. We only give a proof for the second representation. The idea behind the proof is to solve the system of matrix equations

$$AX = I$$

for the unknown matrix X. By the properties of the inverse X then has to be the inverse. It is useful to partition X as

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$$

and to write the system of matrix equations more explicitly as the following set of equations

$$\begin{aligned} A_{11}X_{11} + A_{12}X_{21} &= I \\ A_{11}X_{12} + A_{12}X_{22} &= 0 \\ A_{21}X_{11} + A_{22}X_{21} &= 0 \\ A_{21}X_{12} + A_{22}X_{22} &= I. \end{aligned}$$

Since A_{22} is non-sigular we can solve the third matrix equation for X_{21} such that

$$X_{21} = -A_{22}^{-1}A_{21}X_{11}. \tag{1}$$

Substituting this result in the first equation leads to

$$\begin{aligned} A_{11}X_{11} - A_{12}A_{22}^{-1}A_{21}X_{11} &= I \\ X_{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} \qquad (2) \end{aligned}$$

which leads to expressions for the partitioned inverse elements X_{21} and X_{11}. In the same way we now solve the second equation for X_{12} where

$$\begin{aligned} X_{12} &= -A_{11}^{-1}A_{12}X_{22} \qquad (3) \\ -A_{21}A_{11}^{-1}A_{12}X_{22} + A_{22}X_{22} &= I \\ X_{22} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \qquad (4) \end{aligned}$$

such that we have determined all the elements in of the inverse. In order to fully establish our result we need to show that $X_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} = A_{22}^{-1} + A_{11}^{-1}A_{21}E^{-1}A_{12}A_{22}^{-1}$. By Lemma (1) it follows that $(I_m + CB)^{-1} = I_m - C(I_n + BC)^{-1}B$ for matrices B and C such that C is $m \times n$ and B is $n \times m$ and $I_m + CB$ as well as $I_n + BC$ are non singular. Now note that

$$(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} = (I - A_{22}^{-1}A_{21}A_{11}^{-1}A_{12})^{-1}A_{22}^{-1}.$$

Let $C = -A_{22}^{-1}A_{21}$ and $B = A_{11}^{-1}A_{12}$ then

$$\begin{aligned} (I - A_{22}^{-1}A_{21}A_{11}^{-1}A_{12})^{-1} &= (I + A_{22}^{-1}A_{21}(I - A_{11}^{-1}A_{12}A_{22}^{-1}A_{21})^{-1}A_{11}^{-1}A_{12} \\ &= I + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12} \end{aligned}$$

and the result follows. ■

We now state the lemma that we used in the proof of the partitioned inverse formula.

Lemma 1 (*) *Let C $m \times n$ and B $n \times m$ s.t. $I_m + CB$ and $I_n + BC$ are non singular. Then*

$$(I_m + CB)^{-1} = I_m - C(I_n + BC)^{-1}B.$$

Proof. We use the following identity

$$\begin{bmatrix} I_n & -B \\ 0 & I_m \end{bmatrix} \begin{bmatrix} I_n & B \\ -C & I_m \end{bmatrix} = \begin{bmatrix} I_n + BC & 0 \\ -C & I_m \end{bmatrix}.$$

Then

$$\begin{bmatrix} (I_n + BC)^{-1} & 0 \\ C(I_n + BC)^{-1} & I_m \end{bmatrix} = \begin{bmatrix} (I_n + BC)^{-1} & -B(I_m + CB)^{-1} \\ C(I_n + BC)^{-1} & (I_m + CB)^{-1} \end{bmatrix} \begin{bmatrix} I_n & B \\ 0 & I_m \end{bmatrix}$$

by (1), (2), (3) and (4). This implies

$$C(I_n + BC)^{-1}B + (I_m + CB)^{-1} = I_m.$$

■

4 The Determinant

The determinant is defined as a function mapping the vectors of a $n \times n$ matrix into a single number. If A is a $n \times n$ matrix we use $|A|$ to denote the determinant where this causes no confusion with the absolute value of a real number. Alternatively, the notation $\det A$ is used.

The function plays a role in representing inverses of matrices and can be used as a test of whether a matrix is singular or non singular. We first give a formal definition.

Definition 7 *The determinant of a square matrix A with dimension $n \times n$ is a mapping $A \to |A|$ such that*
i) $|\cdot|$ is linear in each row of A
ii) if rank $(A) < n$ then $|A| = 0$ and vice versa
iii) $|I| = 1$

An alternative way of representing the determinant is to view the matrix A as a collection of n row vectors $\{a_1, ..., a_n\}$ where each $a_i \in \mathbb{R}^n$. The determinant is then a function mapping the set of vectors $\{a_1, ..., a_n\}$ into R. We can write $D(a_1, ..., a_n) \to \mathbb{R}$. Property i) of the previous definition the means that

$$D(a_1, ..., \lambda a_i + \mu a_i', a_{i+1}, ..., a_n) = \lambda D(a_1, ..., a_i, ..., a_n) + \mu D(a_1, ..., a_i', ..., a_n)$$

for any scalars $\mu, \lambda \in \mathbb{R}$ and any vectors $a_i, a_i' \in \mathbb{R}$. In particular note that $D(a_1, ..., \lambda a_i, ..., a_n) = \lambda D(a_1, ..., a_i, ..., a_n)$.

A simple geometric interpretation of the determinant can be obtained in the case where $n = 2$. Consider the pair of vectors $a_1, a_2 \in \mathbb{R}^2$. Geometrically, a_1 and a_2 span a parallelogram with edges a_1 and a_2. The absolute value of the determinant $D(a_1, a_2)$ then is the area of the parallelogram. If we stretch one of the vectors by a factor λ then the area changes by a factor

$|\lambda|$. If we stretch both vectors by the factor λ then $D(\lambda a_1, \lambda a_2) = \lambda^2$ such that the area changes by a factor λ^2.

The following theorem establishes (without proof) that there always exists a function that satisfies the above definition.

Theorem 4 *The mapping $|\cdot|$ exists and is unique.*

The discussion so far only established the existence of the determinant and some of its properties. It did not give an explicitly formula for computing the determinant. We note that the determinant can be computed by using the following recursive formulation. Note that

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

where A_{ij} is the matrix obtained from leaving out the ith row and the jth column. This formulation is called the expansion of $|.|$ along the jth column. We rarely need to compute determinants explicitly by hand (computers were invented for that!). On the other hand the following calculus rules are useful.

Result 5 *Let A, B be $n \times n$ matrices. Then*

$$\begin{aligned} |AB| &= |A| |B| \\ \left|A^{-1}\right| &= 1/|A| \end{aligned}$$

(*) Proof:. The idea of the proof is to establish that $|A| |B|$ possesses all the properties of the determinant of AB. Because the determinant is the unique function with these properties this is enough to establish that $|AB| = |A| |B|$. Now first assume that $|B| \neq 0$. But then we can check that $|A| := \frac{|AB|}{|B|}$ has the required properties of a determinant which then establishes $|AB| = |A| |B|$. For this purpose let $\tilde{A}$ be the matrix where the j-th row was replaced by $\alpha a_{jk} \forall k = 1, ..., n$. Also use the notation $[A]_{ij}$ for the element a_{ij} of the matrix A. Then

$$\begin{aligned} \left[\tilde{A}B\right]_{j\ell} &= \alpha \sum_k a_{jk} b_{k\ell} \qquad \forall \ell \\ \left[\tilde{A}B\right]_{i\ell} &= \sum a_{ik} b_{k\ell} \qquad \forall \ell, \text{ and } i \neq j \end{aligned}$$

so that again $\tilde{A}B$ is a matrix where the j^{th} row has been scaled by α. Then $\left|\tilde{A}B\right| = \alpha |AB|$ by the properties of $|.|$. Using our claimed representation of $\left|\tilde{A}\right|$ we note $\left|\tilde{A}\right| = \frac{|\tilde{A}B|}{|B|} = \alpha \frac{|AB|}{|B|} = \alpha |A|$ as required. Also, if rank $A < n$ then rank$(AB) < n$ such that $|A| = \frac{|AB|}{|B|} = 0$ and $|I| = \frac{|B|}{|B|} = 1$ so in other words, $|AB| = |A| |B|$ satisfies all the required properties. If $|B| = 0$ then $|AB| = 0$ and $|A| |B|$ satisfies the requirements trivially.

For the second result we note that $|I| = 1 = \left|A^{-1}A\right| = \left|A^{-1}\right| |A|$. ■

Result 6 *Let A be a $n \times n$ matrix. Then*

$$\left|A'\right| = |A|.$$

Proof. We only need to show that $|\cdot|$ is also linear in the columns. Using the expansion along column j, we find

$$|A| = \sum_i (-1)^{i+j} \alpha a_{ij} \left|\tilde{A}_{ij}\right|$$

where $\left|\tilde{A}_{ij}\right|$ does not depend on the j^{th} column. This shows that the determinant is linear in the columns of A. ■

Result 7 *Let A be a $n \times n$ matrix. Then*

$$|\alpha A| = \alpha^n |A|$$

Proof. Note that $|\alpha A| = D(\alpha a_1, \alpha a_2, ..., \alpha a_n) = \alpha D(a_1, \alpha a_2, ..., \alpha a_n) = \alpha^2 D(a_1, a_2, ..., \alpha a_n)$ by repeated application of the linearity property of $|.|$. ■

Result 8 *Let A be a $n \times n$ matrix, B a $m \times m$ matrix and C a $n \times m$ matrix. Then*

$$\det \begin{bmatrix} A & C \\ 0 & B \end{bmatrix} = |A| |B|.$$

Proof. We need to check that the function $|A| |B|$ satisfies the three condition for a determinant. Let $\tilde{A}$ and $\tilde{C}$ be transformations of A and C by having one row replaced by $\lambda r_i + \mu r_i'$ where λ, μ are scalars and $r_i = (a_i, c_i) \in \mathbb{R}^{n+m}$ and $r_i = (a_i, c_i) \in \mathbb{R}^{n+m}$ are vectors. Let A_i be the matrix A with ith row a_i and A_i' the matrix A with ith row a_i' with corresponding definitions for C. Then

$$\det \begin{vmatrix} \tilde{A} & \tilde{C} \\ 0 & B \end{vmatrix} = \left|\tilde{A}\right| |B| = \left(\lambda |A_i| + \mu \left|A_i'\right|\right) |B| = \lambda \det \begin{bmatrix} A_i & C_i \\ 0 & B \end{bmatrix} + \mu \det \begin{bmatrix} A_i' & C_i' \\ 0 & B \end{bmatrix}.$$

The same argument holds for linear combinations of rows in B. This establishes linearity in the rows. If A is reduced rank then matrix

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$$

is of reduced rank and thus the determinant is zero. The same argument again holds if B is singular. Now assume A, B are full rank then

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

implies $Bx_2 = 0$ and $Ax_1 + CX_2 = 0$. Since $x_2 = 0$ it follows that $Ax_1 = 0$. But because the null space of A is zero, it follows that $x_1 = 0$. So

$$\begin{bmatrix} A & C \\ 0 & B \end{bmatrix}$$

is also full rank and $|A|\,|B| \neq 0$ as required. Finally, if $A = I_n$ and $B = I_m$ and $C = 0$ then $|I_n|\,|I_m| = 1$, establishing the last property. ■

We can establish a more general result for the determinant of a partitioned matrix similar to the result on partitioned inverses.

Result 9 *Let*

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

where A_{11} *and* A_{22} *are full rank* $n \times n, m \times m$ *matrices. Then*

$$\begin{aligned} |A| &= |A_{11}|\left|A_{22} - A_{21}A_{11}^{-1}A_{12}\right| \\ &= |A_{22}|\left|A_{11} - A_{12}A_{22}^{-1}A_{21}\right| \end{aligned}$$

Proof. Define the matrix

$$B = \begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{bmatrix}$$

then

$$AB = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$$

such that

$$\begin{aligned} |AB| &= |A|\,|B| \\ &= |A_{11}|\left|A_{22} - A_{21}A_{11}^{-1}A_{12}\right| \end{aligned}$$

where the first equality follows from Result (5) and the second equality follows from Result (8) and the fact that $|B| = 1$. ■

5 Quadratic Forms

Let A be an $n \times n$ matrix and $x \in \mathbb{R}^n$. Then

$$x'Ax = \sum_{i,j} a_{ij}x_i x_j$$

is a quadratic form. Note that $x'Ax$ is a scalar. Therefore, it follows that $x'Ax = x'A'x$ such that we can write $x'Ax = 1/2x'(A+A')x$ where $(A+A')/2$ is symmetric. Thus assume without loss of generality that A is symmetric, i.e. $a_{ij} = a_{ji}$ or $A = A'$. Then A is

positive	definite	if $x'Ax > 0$	$\forall\, x \neq 0$
positive	semidefinite	if $x'Ax \geq 0$	$\forall\, x$
negative	definite	if $x'Ax < 0$	$\forall\, x \neq 0$
negative	semidefinite	if $x'Ax \leq 0$	$\forall\, x$

Result 10 *For any matrix B, $A = B'B$ is positive semi-definite.*

Proof. Let $y = Bx$. Then $x'Ax = x'B'Bx = y'y = \sum_i y_i^2 \geq 0$. Note that $y'y$ could be zero if x is chosen to lie in the null space of B. ■

Result 11 *A positive definite matrix is non singular.*

Proof. Assume that A is positive definite and A is singular. Then there exists an $x \neq 0$ such that $Ax = 0$. But then $x'Ax = 0$ which contradicts that A is positive definite. ■

Result 12 *If B is $m \times n$ with $m \geq n$ and $r(B) = n$ then $B'B$ is positive definite.*

Proof. Let $y = Bx$ where Bx is only zero if $x = 0$ because B is of full column rank n. Then $y \neq 0$ such that $x'B'Bx = y'y > 0$. ■

Definition 8 *A $n \times n$ matrix M is said to be idempotent if*

$$MM = M.$$

Result 13 *If M is symmetric then M is also positive semi-definite.*

Proof. For any $x \in \mathbb{R}^n$, let $y = Mx$. Then $x'Mx = x'MMx$ by idempotency. By symmetry it then follows that $x'MMx = x'M'Mx = y'y \geq 0$ as shown before where $y'y = 0$ only if $y = 0$ which occurs if $Mx = 0$. (This is possible because usually M is of reduced rank). ■

6 Eigenvalues and Eigenvectors

In this section we study eigenvalues and eigenvectors of a given matrix A. These eigenvectors and eigenvalues can be used to transform the matrix A into a simpler form which in turn is useful to solve systems of linear equations and to analyze the properties of the mapping described by A. We say that λ is an eigenvalue of an $n \times n$ matrix A with corresponding eigenvector x if

$$Ax = \lambda x \tag{5}$$

for some $x \neq 0$. An $x \neq 0$ satisfying this equation is an eigenvector corresponding to the eigenvalue λ.

Note that Equation (5) can be written as $(A - \lambda I)x = 0$. This homogeneous system of equations has a nontrivial ($x \neq 0$) solution if and only if the matrix $A - \lambda I$ is singular. The matrix $A - \lambda I$ is singular if and only if the determinant $|A - \lambda I| = 0$. This leads to a characterization of the eigenvalues as solutions to the equation $|A - \lambda I| = 0$.Note that $|A - \lambda I|$ is a polynomial of degree n in λ. To see this consider the representation of the determinant as

$$\det |A - \lambda I| = \sum (-1)^{\phi(\partial_1,\dots,\partial_n)} \prod_{i=1}^{n} (a_{ij_i} - \lambda 1 \{i = j_i\})$$

where the summation is taken over all permutations $(\partial_1, ..., \partial_n)$ of the set of integers $(1, ..., n)$, and $\phi(j_1, ..., j_n)$ is the number of transpositions required to change $(1, ..., n)$ into $(j_1, ..., j_n)$.

The polynomial $|A - \lambda I|$ has at most n, possibly complex, distinct roots that can appear with multiplicities, and at least one (complex) root. Note, that the number of distinct roots can be less than n. In that case one or more roots occur with multiplicities.

Also note that the eigenvectors associated with a given eigenvalue need not be unique. A simple example is the case where $A = I$. Then, the only eigenvalue of I is 1, occurring with multiplicity n. In this case all nonzero vectors $x \in \mathbb{R}^n$ are eigenvectors of I.

We now normalize the eigenvector x such that $x'x = 1$. If x is an eigenvector for λ and y an eigenvector for μ then

$$y'Ax = \lambda y'x$$

and

$$y'Ax = \mu y'x$$

thus

$$0 = (\lambda - \mu)y'x$$

implying that $y'x = 0$. This shows that eigenvectors of distinct eigenvalues are orthogonal. If $\lambda = \mu$ and $x \neq \alpha y$ then any vector in the span of x and y is also a solution to $(A - \lambda I)x = 0 = (A - \mu I)y$. Thus we can always choose x and y to be orthogonal.

We have so far established that any $n \times n$ matrix A has eigenvalues $\lambda_1, ..., \lambda_n$ with orthogonal eigenvectors v_i. We now use these results to transform a given matrix A into a simpler form. The next theorem states that for any $n \times n$ matrix A there exists possibly complex valued matrices T and M such that M is upper triangular (all elements below the diagonal are zero) and T is unitary. Note that T is unitary if $T^*T = I$ where $T = X + iy$ for two matrices x and y of dimension $n \times n$, $i = \sqrt{-1}$ and $T^* = x' - iy'$. The matrix T^* is called the complex conjugate of T.

Result 14 (Schur Decomposition Theorem) *Let A be an $n \times n$ matrix. Then there exists a unitary $n \times n$ matrix T and an upper triangular matrix M whose diagonal elements are the* *eigenvalues of A such that*

$$P^*AP = M.$$

(*) Proof. Let us proceed by induction, beginning with the 2×2 case. Let λ_1 be a characteristic root of A and c^1 an associated characteristic vector normalized by the condition $c^{*1}c^1 = 1$ where c^{*1} is the complex conjugate transpose of c^1. Let T be a matrix whose first column is c^1 and whose second column is chosen so that T is unitary. Then evaluating the expression $T^{-1}AT$ as the product $T^{-1}(AT)$, we see that

$$T^{-1}AT = \begin{bmatrix} \lambda_1 & b_{12} \\ 0 & b_{22} \end{bmatrix}$$

The quantity b_{22} must equal λ_2 since $T^{-1}AT$ has the same characteristic roots as A.

Let us now show that we can use the reduction for the N^{th}-order matrices to demonstrate the result for $(N+1)$-dimensional matrices. As before, let c^1 be a normalized characteristic vector associated with λ_1, and let N other vectors $a^1, a^2, ..., a^N$, be chosen so that the matrix T_1, whose columns are $c^1, a^1, a^2, ..., a^N$, is unitary. Then, as for $N = 2$, we have

$$T_1^{-1}AT_1 = \begin{bmatrix} \lambda_1 & b_{12}^1 & \cdots & b_{1N+1}^1 \\ 0 & & & \\ \vdots & & B_N & \\ 0 & & & \end{bmatrix}$$

where B_N is an $N \times N$ matrix.

Since the characteristic equation of the right-hand side is

$$(\lambda_1 - \lambda)\,|B_N - \lambda I| = 0$$

it follows that the characteristic roots of B_N are $\lambda_2, \lambda_3, ..., \lambda_{N+1}$, the remaining N characteristic roots of A. The inductive hypothesis asserts that there exists a unitary matrix T_N such that

$$T_N^{-1}B_NT_N = \begin{bmatrix} \lambda_2 & c_{12} & \cdots & c_{1N} \\ & \lambda_3 & & c_{2N} \\ & & \ddots & \vdots \\ 0 & & & \lambda_{N+1} \end{bmatrix}$$

Let T_{N+1} be the unitary $(N+1) \times (N+1)$ matrix formed as follows:

$$T_{N+1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & T_N & \\ 0 & & & \end{bmatrix}$$

Consider the expression

$$
\begin{aligned}
(T_1 T_{N+1})^{-1} A (T_1 T_{N+1}) &= T_{N+1}^{-1} (T_1^{-1} A T_1) T_{N+1} \\
&= \begin{bmatrix} \lambda_1 & b_{12} & \cdots & b_{1,N+1} \\ & \lambda_1 & & \\ & & \ddots & \vdots \\ & & & \lambda_{N+1} \end{bmatrix}
\end{aligned}
$$

The matrix $T_1 T_{N+1}$ is thus the required unitary matrix which reduces A to semidiagonal form. ■

Our next result shows that under certain conditions on the matrix A we can find transformations that reduce A to a diagonal matrix. In order to prove this result we first show that the eigenvalues of a real, symmetric matrix are also real valued.

Result 15 *The eigenvalues of a real symmetric matrix A are real.*

Proof. Let x be an eigenvalue of A and $x = u + iv$ its associated eigenvector. Then

$$A(u + iv) = \lambda(u + iv)$$

and

$$(u - iv)' A(u + iv) = \lambda(u - iu)'(u + iv)$$

because of symmetry of A it follows that $iv'Au = u'Aiv$ and thus

$$u'Au + v'Av = \lambda(u'u + v'v)$$

which implies that λ is real. If λ is real then x can be chosen to be real as well. ■

Result 16 *Let A be a real symmetric matrix. Then there exists an orthogonal matrix S and a diagonal matrix Λ whose diagonal elements are the eigenvalues of A such that*

$$S'AS = \Lambda.$$

Proof. By the previous result the eigenvalues are real and the eigenvectors can be chosen real. By the Schur decomposition theorem there exists an upper diagonal matrix M such that

$$S^* AS = M.$$

By symmetry $S^*AS = M^*$ such that $M' = M$ which proves that M is diagonal. From the previous result we also know that the diagonal elements of M are real. Thus, write $M = \Lambda$. Since $AS = S\Lambda$ it follows that the rows of S contain the eigenvectors corresponding to Λ. But we know that the eigenvectors can be chosen real by the previous result. This establishes the claim. ■

Result 17 *If A is idempotent then the eigenvalues of A are 0 or 1.*

Proof. $Ax = \lambda x \Rightarrow Ax = AAx = \lambda Ax = \lambda^2 x$ so $\lambda^2 = \lambda$ which implies $\lambda = 0$ or $\lambda = 1$. ■

Result 18 *If A is real symmetric and idempotent $n \times n$ matrix of rank $k < n$ then $\exists\, S_1, S_2$ $n \times k$ and $n \times n - k$ s.t. $S_1' S_2 = 0, S_2' S_1 = 0, S_1^1 S_1 = I$ and $S_2' S_2 = I$ such that $A = S_1 S_1'$.*

Proof. By the previous result we know that the eigenvalues of A are either 0 or 1. Since A has rank k there must exist exactly $n-k$ linearly independent vectors x such that $Ax = 0$. This shows that A has $n-k$ zero eigenvalues with the remaining nonzero eigenvalues equal to 1.We know that $S'AS = \Lambda$ where S is an orthogonal matrix of eigenvectors and Λ is the corresponding matrix of eigenvalues. Now arrange the eigenvalues such that

$$\Lambda = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Then arrange the eigenvectors accordingly such that

$$A = [S_1 \; S_2] \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S_1' \\ S_2' \end{bmatrix} = S_1 S_1^1.$$

■

The square root of a real symmetric matrix A is defined as $A^{1/2} A^{1/2} = A$. Note that $A^{1/2}$ is not the same as taking the square root of all elements in A! We want to find an explicit representation of $A^{1/2}$. By the previous result we can diagonalize the matrix A such that

$$S'AS = \Lambda$$

or

$$A = S \Lambda S' = S \Lambda^{1/2} \Lambda^{1/2} S',$$

where now we are allowed to define

$$\Lambda^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{bmatrix}$$

because $\Lambda^{1/2}$, so defined, satisfies

$$\Lambda^{1/2} \Lambda^{1/2} = \Lambda$$

because Λ is a diagonal matrix. Then define

$$A^{1/2} = S \Lambda^{1/2} S'.$$

It can be checked that $A^{1/2} A^{1/2} = S \Lambda^{1/2} S' S \Lambda^{1/2} S' = S \Lambda^{1/2} \Lambda^{1/2} S' = A$. Also note that $A^{-1} = (S \Lambda S')^{-1} = S \Lambda^{-1} S'$ such that $A^{-1/2} = S \Lambda^{-1/2} S^1$.

Result 19 *If Σ is a symmetric real $n \times n$ matrix then there exists a matrix A such that $AA' = \Sigma$.*

Proof. Let $\Sigma = S\Lambda S'$. Then $A = S\Lambda^{1/2}$. ■

Result 20 *If A is a symmetric real $n \times n$ matrix then $|A| = \prod_i \lambda_i$ where λ_i are the eigenvalues of A.*

Proof. We diagonalize the matrix A such that $|A| = \left|S\Lambda S'\right| = |S|\,|\Lambda|\,|S'| = |SS'|\,|\Lambda| = |I|\,|\Lambda| = \prod_i \lambda_i$. The last equality can be shown by using a column expansion of the determinant of Λ. ■

Result 21 *If A is a symmetric real $n \times n$ matrix and if A is positive definite then $\lambda_i > 0$.*

Proof. Let x be an eigenvector of A such that $Ax = \lambda x$ where $x \neq 0$. Then $x'Ax = \lambda > 0$ by positive definiteness of A. ■

The trace of a square matrix A is defined as the sum of all the diagonal elements. Let A be a $n \times n$ matrix. Then define

$$\text{tr } A = \sum_{i=1}^{n} a_{ii}.$$

Result 22 *i)* $\text{tr}\,\alpha A = \alpha\,\text{tr}\,A$ *for α scalar*
ii) $\text{tr } A' = \text{tr } A$
iii) $\text{tr}(A + B) = \text{tr } A + \text{tr } B$ *where A and B are $n \times n$ matrices*
iv) $\text{tr}(AB) = \text{tr}(BA)$ *if AB is square*
v) Let a be a $n \times 1$ vector then

$$a'a = \text{tr}(a'a) = \text{tr } aa'.$$

vi) if A is a $n \times n$ matrix then $\text{tr}\,A = \Sigma\lambda_i$.
vii) if A is real symmetric and idempotent then $\text{tr}\,A = \text{rank}\,A$.

Proof. Most of these results can be immediately verified with some algebra. We show vi): By the Schur decomposition theorem and iv) we can write $\text{tr}\,A = \text{tr } SMS^* = \text{tr } S^*SM = \text{tr } M = \sum_i \lambda_i$. Next we show vii): From previous results we know that

$$\begin{aligned} \text{tr } A &= \text{tr}(S\Lambda S') = \text{tr}\left([S_1 S_2] \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S_1' \\ S_2' \end{bmatrix} \right) \\ &= \text{tr}(S_1 S_1') = \text{tr } I_k = k \end{aligned}$$

where k is the number of non-zero eigenvalues. ■