Appendix A

Matrix Algebra

A.1 Notation

A **scalar** a is a single number.

A **vector** $\boldsymbol{a}$ is a $k \times 1$ list of numbers, typically arranged in a column. We write this as

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$$

Equivalently, a vector $\boldsymbol{a}$ is an element of Euclidean k space, written as $\boldsymbol{a} \in \mathbb{R}^k$. If $k = 1$ then $\boldsymbol{a}$ is a scalar.

A **matrix** $\boldsymbol{A}$ is a $k \times r$ rectangular array of numbers, written as

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kr} \end{bmatrix}$$

By convention a_{ij} refers to the element in the i^{th} row and j^{th} column of $\boldsymbol{A}$. If $r = 1$ then $\boldsymbol{A}$ is a column vector. If $k = 1$ then $\boldsymbol{A}$ is a row vector. If $r = k = 1$, then $\boldsymbol{A}$ is a scalar.

A standard convention (which we will follow in this text whenever possible) is to denote scalars by lower-case italics (a), vectors by lower-case bold italics ($\boldsymbol{a}$), and matrices by upper-case bold italics ($\boldsymbol{A}$). Sometimes a matrix $\boldsymbol{A}$ is denoted by the symbol (a_{ij}).

A matrix can be written as a set of column vectors or as a set of row vectors. That is,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_r \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_k \end{bmatrix}$$

where

$$\boldsymbol{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ki} \end{bmatrix}$$

are column vectors and

$$\boldsymbol{\alpha}_j = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jr} \end{bmatrix}$$

are row vectors.

The **transpose** of a matrix $\boldsymbol{A}$, denoted $\boldsymbol{A}'$, $\boldsymbol{A}^\top$, or $\boldsymbol{A}^t$, is obtained by flipping the matrix on its diagonal. (In most of the econometrics literature, and this textbook, we use $\boldsymbol{A}'$, but in the mathematics literature $\boldsymbol{A}^\top$ is the convention.) Thus

$$\boldsymbol{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & & \vdots \\ a_{1r} & a_{2r} & \cdots & a_{kr} \end{bmatrix}$$

Alternatively, letting $\boldsymbol{B} = \boldsymbol{A}'$, then $b_{ij} = a_{ji}$. Note that if $\boldsymbol{A}$ is $k \times r$, then $\boldsymbol{A}'$ is $r \times k$. If $\boldsymbol{a}$ is a $k \times 1$ vector, then $\boldsymbol{a}'$ is a $1 \times k$ row vector.

A matrix is **square** if $k = r$. A square matrix is **symmetric** if $\boldsymbol{A} = \boldsymbol{A}'$, which requires $a_{ij} = a_{ji}$. A square matrix is **diagonal** if the off-diagonal elements are all zero, so that $a_{ij} = 0$ if $i \neq j$. A square matrix is **upper (lower) diagonal** if all elements below (above) the diagonal equal zero.

An important diagonal matrix is the **identity matrix**, which has ones on the diagonal. The $k \times k$ identity matrix is denoted as

$$\boldsymbol{I}_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

A **partitioned matrix** takes the form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \cdots & \boldsymbol{A}_{1r} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \cdots & \boldsymbol{A}_{2r} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{A}_{k1} & \boldsymbol{A}_{k2} & \cdots & \boldsymbol{A}_{kr} \end{bmatrix}$$

where the A_{ij} denote matrices, vectors and/or scalars.

A.2 Complex Matrices*

Scalars, vectors and matrices may contain real or complex numbers as entries. (However, most econometric applications exclusively use real matrices.) If all elements of a vector $\boldsymbol{x}$ are real we say that $\boldsymbol{x}$ is a real vector, and similarly for matrices.

Recall that a complex number can be written as $x = a + b\mathrm{i}$ where where $\mathrm{i} = \sqrt{-1}$ and a and b are real numbers. Similarly a vector with complex elements can be written as $\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{b}\mathrm{i}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are real vectors, and a matrix with complex elements can be written as $\boldsymbol{X} = \boldsymbol{A} + \boldsymbol{B}\mathrm{i}$ where $\boldsymbol{A}$ and $\boldsymbol{B}$ are real matrices.

Recall that the complex conjugate of $x = a + b\mathrm{i}$ is $x^* = a - b\mathrm{i}$. For matrices, the analogous concept is the conjugate transpose. The conjugate transpose of $\boldsymbol{X} = \boldsymbol{A} + \boldsymbol{B}\mathrm{i}$ is $\boldsymbol{X}^* = \boldsymbol{A}' - \boldsymbol{B}'\mathrm{i}$. It is obtained by taking the transpose and taking the complex conjugate of each element.

A.3 Matrix Addition

If the matrices $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ are of the same order, we define the sum

$$\boldsymbol{A} + \boldsymbol{B} = (a_{ij} + b_{ij}).$$

Matrix addition follows the commutative and associative laws:

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$$
$$\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}.$$

A.4 Matrix Multiplication

If $\boldsymbol{A}$ is $k \times r$ and c is real, we define their product as

$$\boldsymbol{A}c = c\boldsymbol{A} = (a_{ij}c).$$

If $\boldsymbol{a}$ and $\boldsymbol{b}$ are both $k \times 1$, then their inner product is

$$\boldsymbol{a}'\boldsymbol{b} = a_1 b_1 + a_2 b_2 + \cdots + a_k b_k = \sum_{j=1}^{k} a_j b_j.$$

Note that $\boldsymbol{a}'\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{a}$. We say that two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are **orthogonal** if $\boldsymbol{a}'\boldsymbol{b} = 0$.

If $\boldsymbol{A}$ is $k \times r$ and $\boldsymbol{B}$ is $r \times s$, so that the number of columns of $\boldsymbol{A}$ equals the number of rows of $\boldsymbol{B}$, we say that $\boldsymbol{A}$ and $\boldsymbol{B}$ are **conformable**. In this event the matrix product $\boldsymbol{AB}$ is defined. Writing $\boldsymbol{A}$ as a set of row vectors and $\boldsymbol{B}$ as a set of column vectors (each of length r), then the matrix product is defined as

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_k' \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_s \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1'\boldsymbol{b}_s \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2'\boldsymbol{b}_s \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_k'\boldsymbol{b}_1 & \boldsymbol{a}_k'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_k'\boldsymbol{b}_s \end{bmatrix}.$$

Matrix multiplication is not commutative: in general $\boldsymbol{AB} \neq \boldsymbol{BA}$. However, it is associative and distributive:

$$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$$
$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}.$$

An alternative way to write the matrix product is to use matrix partitions. For example,

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} & \boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} & \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} \end{bmatrix}.$$

As another example,

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \cdots & \boldsymbol{A}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \\ \vdots \\ \boldsymbol{B}_r \end{bmatrix}$$
$$= \boldsymbol{A}_1\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{B}_2 + \cdots + \boldsymbol{A}_r\boldsymbol{B}_r$$
$$= \sum_{j=1}^{r} \boldsymbol{A}_j\boldsymbol{B}_j.$$

An important property of the identity matrix is that if $\boldsymbol{A}$ is $k \times r$, then $\boldsymbol{A}\boldsymbol{I}_r = \boldsymbol{A}$ and $\boldsymbol{I}_k\boldsymbol{A} = \boldsymbol{A}$.

We say two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are **orthogonal** if $\boldsymbol{A}'\boldsymbol{B} = \boldsymbol{0}$. This means that all columns of $\boldsymbol{A}$ are orthogonal with all columns of $\boldsymbol{B}$.

The $k \times r$ matrix $\boldsymbol{H}$, $r \leq k$, is called **orthonormal** if $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_r$. This means that the columns of $\boldsymbol{H}$ are mutually orthogonal, and each column is normalized to have unit length.

A.5 Trace

The **trace** of a $k \times k$ square matrix $\boldsymbol{A}$ is the sum of its diagonal elements

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}.$$

Some straightforward properties for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and real c are

$$\begin{aligned}
\operatorname{tr}(c\boldsymbol{A}) &= c\operatorname{tr}(\boldsymbol{A}) \\
\operatorname{tr}\left(\boldsymbol{A}'\right) &= \operatorname{tr}(\boldsymbol{A}) \\
\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) &= \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B}) \\
\operatorname{tr}(\boldsymbol{I}_k) &= k.
\end{aligned}$$

Also, for $k \times r$ $\boldsymbol{A}$ and $r \times k$ $\boldsymbol{B}$ we have

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}). \tag{A.1}$$

Indeed,

$$\begin{aligned}
\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) &= \operatorname{tr}\begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1'\boldsymbol{b}_k \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2'\boldsymbol{b}_k \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_k'\boldsymbol{b}_1 & \boldsymbol{a}_k'\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_k'\boldsymbol{b}_k \end{bmatrix} \\
&= \sum_{i=1}^{k} \boldsymbol{a}_i'\boldsymbol{b}_i \\
&= \sum_{i=1}^{k} \boldsymbol{b}_i'\boldsymbol{a}_i \\
&= \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}).
\end{aligned}$$

A.6 Rank and Inverse

The rank of the $k \times r$ matrix $(r \leq k)$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_r \end{bmatrix}$$

is the number of linearly independent columns $\boldsymbol{a}_j$, and is written as $\operatorname{rank}(\boldsymbol{A})$. We say that $\boldsymbol{A}$ has full rank if $\operatorname{rank}(\boldsymbol{A}) = r$.

A square $k \times k$ matrix $\boldsymbol{A}$ is said to be **nonsingular** if it is has full rank, e.g. $\operatorname{rank}(\boldsymbol{A}) = k$. This means that there is no $k \times 1$ $\boldsymbol{c} \neq \boldsymbol{0}$ such that $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$.

If a square $k \times k$ matrix $\boldsymbol{A}$ is nonsingular then there exists a unique matrix $k \times k$ matrix $\boldsymbol{A}^{-1}$ called the **inverse** of $\boldsymbol{A}$ which satisfies

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_k.$$

For non-singular $\boldsymbol{A}$ and $\boldsymbol{C}$, some important properties include

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^{-1} &= \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_k \\
\left(\boldsymbol{A}^{-1}\right)' &= \left(\boldsymbol{A}'\right)^{-1} \\
(\boldsymbol{A}\boldsymbol{C})^{-1} &= \boldsymbol{C}^{-1}\boldsymbol{A}^{-1} \\
(\boldsymbol{A}+\boldsymbol{C})^{-1} &= \boldsymbol{A}^{-1}\left(\boldsymbol{A}^{-1}+\boldsymbol{C}^{-1}\right)^{-1}\boldsymbol{C}^{-1} \\
\boldsymbol{A}^{-1}-(\boldsymbol{A}+\boldsymbol{C})^{-1} &= \boldsymbol{A}^{-1}\left(\boldsymbol{A}^{-1}+\boldsymbol{C}^{-1}\right)^{-1}\boldsymbol{A}^{-1}.
\end{aligned}$$

If a $k\times k$ matrix $\boldsymbol{H}$ is orthonormal (so that $\boldsymbol{H}'\boldsymbol{H}=\boldsymbol{I}_k$), then $\boldsymbol{H}$ is nonsingular and $\boldsymbol{H}^{-1}=\boldsymbol{H}'$. Furthermore, $\boldsymbol{H}\boldsymbol{H}'=\boldsymbol{I}_k$ and $\boldsymbol{H}'^{-1}=\boldsymbol{H}$.

Another useful result for non-singular $\boldsymbol{A}$ is known as the **Woodbury matrix identity**

$$(\boldsymbol{A}+\boldsymbol{B}\boldsymbol{C}\boldsymbol{D})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}\left(\boldsymbol{C}+\boldsymbol{C}\boldsymbol{D}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{C}\right)^{-1}\boldsymbol{C}\boldsymbol{D}\boldsymbol{A}^{-1}. \tag{A.2}$$

In particular, for $\boldsymbol{C}=-1$, $\boldsymbol{B}=\boldsymbol{b}$ and $\boldsymbol{D}=\boldsymbol{b}'$ for vector $\boldsymbol{b}$ we find what is known as the **Sherman–Morrison formula**

$$\left(\boldsymbol{A}-\boldsymbol{b}\boldsymbol{b}'\right)^{-1}=\boldsymbol{A}^{-1}+\left(1-\boldsymbol{b}'\boldsymbol{A}^{-1}\boldsymbol{b}\right)^{-1}\boldsymbol{A}^{-1}\boldsymbol{b}\boldsymbol{b}'\boldsymbol{A}^{-1}. \tag{A.3}$$

The following fact about inverting partitioned matrices is quite useful.

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}^{-1} \overset{def}{=} \begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11\cdot 2}^{-1} & -\boldsymbol{A}_{11\cdot 2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -\boldsymbol{A}_{22\cdot 1}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{22\cdot 1}^{-1} \end{bmatrix} \tag{A.4}$$

where $\boldsymbol{A}_{11\cdot 2}=\boldsymbol{A}_{11}-\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$ and $\boldsymbol{A}_{22\cdot 1}=\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$. There are alternative algebraic representations for the components. For example, using the Woodbury matrix identity you can show the following alternative expressions

$$\begin{aligned}
\boldsymbol{A}^{11} &= \boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22\cdot 1}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} \\
\boldsymbol{A}^{22} &= \boldsymbol{A}_{22}^{-1}+\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11\cdot 2}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\
\boldsymbol{A}^{12} &= -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22\cdot 1}^{-1} \\
\boldsymbol{A}^{21} &= -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11\cdot 2}^{-1}
\end{aligned}$$

Even if a matrix $\boldsymbol{A}$ does not possess an inverse, we can still define the **Moore-Penrose generalized inverse $\boldsymbol{A}^{-}$** as the matrix which satisfies

$$\begin{gathered}
\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A}=\boldsymbol{A} \\
\boldsymbol{A}^{-}\boldsymbol{A}\boldsymbol{A}^{-}=\boldsymbol{A}^{-} \\
\boldsymbol{A}\boldsymbol{A}^{-} \text{ is symmetric} \\
\boldsymbol{A}^{-}\boldsymbol{A} \text{ is symmetric}
\end{gathered}$$

For any matrix $\boldsymbol{A}$, the Moore-Penrose generalized inverse $\boldsymbol{A}^{-}$ exists and is unique.

For example, if

$$\boldsymbol{A}=\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

and $\boldsymbol{A}_{11}^{-1}$ exists then

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

A.7 Determinant

The **determinant** is a measure of the volume of a square matrix. It is written as $\det \boldsymbol{A}$ or $|\boldsymbol{A}|$.

While the determinant is widely used, its precise definition is rarely needed. However, we present the definition here for completeness. Let $\boldsymbol{A} = (a_{ij})$ be a $k \times k$ matrix . Let $\pi = (j_1, ..., j_k)$ denote a permutation of $(1, ..., k)$. There are $k!$ such permutations. There is a unique count of the number of inversions of the indices of such permutations (relative to the natural order $(1, ..., k)$, and let $\varepsilon_\pi = +1$ if this count is even and $\varepsilon_\pi = -1$ if the count is odd. Then the determinant of $\boldsymbol{A}$ is defined as

$$\det \boldsymbol{A} = \sum_\pi \varepsilon_\pi a_{1j_1} a_{2j_2} \cdots a_{kj_k}.$$

For example, if $\boldsymbol{A}$ is 2×2, then the two permutations of $(1, 2)$ are $(1, 2)$ and $(2, 1)$, for which $\varepsilon_{(1,2)} = 1$ and $\varepsilon_{(2,1)} = -1$. Thus

$$\begin{aligned} \det \boldsymbol{A} &= \varepsilon_{(1,2)} a_{11} a_{22} + \varepsilon_{(2,1)} a_{21} a_{12} \\ &= a_{11} a_{22} - a_{12} a_{21}. \end{aligned}$$

For a square matrix $\boldsymbol{A}$, the **minor** M_{ij} of the ij^{th} element a_{ij} is the determinant of the matrix obtained by removing the i^{th} row and j^{th} column of $\boldsymbol{A}$. The **cofactor** of the ij^{th} element is $C_{ij} = (-1)^{i+j} M_{ij}$. An important representation known as Laplace's expansion relates the determinant of $\boldsymbol{A}$ to its cofactors:

$$\det \boldsymbol{A} = \sum_{j=1}^k a_{ij} C_{ij}.$$

This holds for all $i = 1, , .., k$. This is often presented as a method for computation of a determinant.

Theorem A.7.1 *Properties of the determinant*

1. $\det(\boldsymbol{A}) = \det(\boldsymbol{A}')$
2. $\det(c\boldsymbol{A}) = c^k \det \boldsymbol{A}$
3. $\det(\boldsymbol{AB}) = \det(\boldsymbol{BA}) = (\det \boldsymbol{A})(\det \boldsymbol{B})$
4. $\det\left(\boldsymbol{A}^{-1}\right) = (\det \boldsymbol{A})^{-1}$
5. $\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = (\det \boldsymbol{D}) \det\left(\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C}\right)$ if $\det \boldsymbol{D} \neq 0$
6. $\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{A})(\det \boldsymbol{D})$ and $\det \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} = \det(\boldsymbol{A})(\det \boldsymbol{D})$
7. If $\boldsymbol{A}$ is $p \times q$ and $\boldsymbol{B}$ is $q \times p$ then $\det(\boldsymbol{I}_p + \boldsymbol{AB}) = \det(\boldsymbol{I}_q + \boldsymbol{BA})$
8. If $\boldsymbol{A}$ and $\boldsymbol{D}$ are invertible then $\det\left(\boldsymbol{A} - \boldsymbol{BD}^{-1}\boldsymbol{C}\right) = \dfrac{\det(\boldsymbol{A})}{\det(\boldsymbol{D})} \det\left(\boldsymbol{D} - \boldsymbol{CA}^{-1}\boldsymbol{B}\right)$
9. $\det \boldsymbol{A} \neq 0$ if and only if $\boldsymbol{A}$ is nonsingular
10. If $\boldsymbol{A}$ is triangular (upper or lower), then $\det \boldsymbol{A} = \prod_{i=1}^k a_{ii}$
11. If $\boldsymbol{A}$ is orthonormal, then $\det \boldsymbol{A} = \pm 1$
12. $\boldsymbol{A}^{-1} = (\det \boldsymbol{A})^{-1} \boldsymbol{C}$ where $\boldsymbol{C} = (C_{ij})$ is the matrix of cofactors

A.8 Eigenvalues

The **characteristic equation** of a $k \times k$ square matrix $\boldsymbol{A}$ is

$$\det(\lambda \boldsymbol{I}_k - \boldsymbol{A}) = 0.$$

The left side is a polynomial of degree k in λ so it has exactly k roots, which are not necessarily distinct and may be real or complex. They are called the **latent roots**, **characteristic roots**, or **eigenvalues** of $\boldsymbol{A}$. If λ is an eigenvalue of $\boldsymbol{A}$, then $\lambda \boldsymbol{I}_k - \boldsymbol{A}$ is singular so there exists a non-zero vector $\boldsymbol{h}$ such that $(\lambda \boldsymbol{I}_k - \boldsymbol{A})\, \boldsymbol{h} = \boldsymbol{0}$ or

$$\boldsymbol{A}\boldsymbol{h} = \boldsymbol{h}\lambda.$$

The vector $\boldsymbol{h}$ is called a **latent vector**, **characteristic vector**, or **eigenvector** of $\boldsymbol{A}$ corresponding to λ. They are typically normalized so that $\boldsymbol{h}'\boldsymbol{h} = 1$ and thus $\lambda = \boldsymbol{h}'\boldsymbol{A}\boldsymbol{h}$.

Set $\boldsymbol{H} = [\boldsymbol{h}_1 \; \cdots \; \boldsymbol{h}_k]$ and $\boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1, ..., \lambda_k\}$. A matrix expression is

$$\boldsymbol{A}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{\Lambda}$$

We now state some useful properties.

Theorem A.8.1 *Properties of eigenvalues. Let λ_i and $\boldsymbol{h}_i$, $i = 1, ..., k$, denote the k eigenvalues and eigenvectors of a square matrix $\boldsymbol{A}$.*

1. $\det(\boldsymbol{A}) = \prod_{i=1}^{k} \lambda_i$
2. $\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} \lambda_i$
3. $\boldsymbol{A}$ is non-singular if and only if all its eigenvalues are non-zero.
4. If $\boldsymbol{A}$ has distinct eigenvalues, there exists a nonsingular matrix $\boldsymbol{P}$ such that $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{\Lambda}\boldsymbol{P}$ and $\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1} = \boldsymbol{\Lambda}$.
5. The non-zero eigenvalues of $\boldsymbol{A}\boldsymbol{B}$ and $\boldsymbol{B}\boldsymbol{A}$ are identical.
6. If $\boldsymbol{B}$ is non-singular then $\boldsymbol{A}$ and $\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}$ have the same eigenvalues.
7. If $\boldsymbol{A}\boldsymbol{h} = \boldsymbol{h}\lambda$ then $(\boldsymbol{I} - \boldsymbol{A}) = \boldsymbol{h}(1 - \lambda)$. So $\boldsymbol{I} - \boldsymbol{A}$ has the eigenvalue $1 - \lambda$ and associated eigenvector $\boldsymbol{h}$.

Most eigenvalue applications in econometrics concern the case where the matrix $\boldsymbol{A}$ is real and symmetric. In this case all eigenvalues of $\boldsymbol{A}$ are real and its eigenvectors are mutually orthogonal. Thus $\boldsymbol{H}$ is orthonormal so $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$ and $\boldsymbol{H}\boldsymbol{H}' = \boldsymbol{I}_k$. When the eigenvalues are all real it is conventional to write them in decending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$.

The following is a very important property of real symmetric matrices, which follows directly from the equations $\boldsymbol{A}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{\Lambda}$ and $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$.

Spectral Decomposition. If $\boldsymbol{A}$ is a $k \times k$ real symmetric matrix, then $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ where $\boldsymbol{H}$ contains the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues on the diagaonal. The eigenvalues are all real and the eigenvector matrix satisfies $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$. The decomposition can be alternatively written as $\boldsymbol{H}'\boldsymbol{A}\boldsymbol{H} = \boldsymbol{\Lambda}$.

If $\boldsymbol{A}$ is real, symmetric, and invertible, then by the spectral decomposition and the properties of orthonormal matrices, $\boldsymbol{A}^{-1} = \boldsymbol{H}'^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}^{-1} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}'$. Thus the columns of $\boldsymbol{H}$ are also the eigenvectors of $\boldsymbol{A}^{-1}$, and its eigenvalues are λ_1^{-1}, λ_2^{-1}, ..., λ_k^{-1}.

A.9 Positive Definite Matrices

We say that a $k \times k$ real symmetric square matrix $\boldsymbol{A}$ is **positive semi-definite** if for all $\boldsymbol{c} \neq \mathbf{0}$, $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \geq 0$. This is written as $\boldsymbol{A} \geq 0$. We say that $\boldsymbol{A}$ is **positive definite** if for all $\boldsymbol{c} \neq \mathbf{0}$, $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} > 0$. This is written as $\boldsymbol{A} > 0$.

Some properties include:

Theorem A.9.1 *Properties of positive semi-definite matrices*

1. If $\boldsymbol{A} = \boldsymbol{G}'\boldsymbol{B}\boldsymbol{G}$ with $\boldsymbol{B} \geq 0$ and some matrix $\boldsymbol{G}$, then $\boldsymbol{A}$ is positive semi-definite. (For any $\boldsymbol{c} \neq \mathbf{0}$, $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} = \boldsymbol{\alpha}'\boldsymbol{B}\boldsymbol{\alpha} \geq 0$ where $\boldsymbol{\alpha} = \boldsymbol{G}\boldsymbol{c}$.) If $\boldsymbol{G}$ has full column rank and $\boldsymbol{B} > 0$, then $\boldsymbol{A}$ is positive definite.

2. If $\boldsymbol{A}$ is positive definite, then $\boldsymbol{A}$ is non-singular and $\boldsymbol{A}^{-1}$ exists. Furthermore, $\boldsymbol{A}^{-1} > 0$.

3. $\boldsymbol{A} > 0$ if and only if it is symmetric and all its eigenvalues are positive.

4. By the spectral decomposition, $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ where $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$ and $\boldsymbol{\Lambda}$ is diagonal with non-negative diagonal elements. All diagonal elements of $\boldsymbol{\Lambda}$ are strictly positive if (and only if) $\boldsymbol{A} > 0$.

5. The rank of $\boldsymbol{A}$ equals the number of strictly positive eigenvalues.

6. If $\boldsymbol{A} > 0$ then $\boldsymbol{A}^{-1} = \boldsymbol{H}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}'$.

7. If $\boldsymbol{A} \geq 0$ and $\operatorname{rank}(\boldsymbol{A}) = r \leq k$ then the Moore-Penrose generalized inverse of $\boldsymbol{A}$ is $\boldsymbol{A}^{-} = \boldsymbol{H}\boldsymbol{\Lambda}^{-}\boldsymbol{H}'$ where $\boldsymbol{\Lambda}^{-} = \operatorname{diag}\left(\lambda_1^{-1}, \lambda_2^{-1}, ..., \lambda_r^{-1}, 0, ..., 0\right)$.

8. If $\boldsymbol{A} \geq 0$ we can find a matrix $\boldsymbol{B}$ such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}'$. We call $\boldsymbol{B}$ a **matrix square root** of $\boldsymbol{A}$ and is typically written as $\boldsymbol{B} = \boldsymbol{A}^{1/2}$. The matrix $\boldsymbol{B}$ need not be unique. One matrix square root is obtained using the spectral decomposition $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$. Then $\boldsymbol{B} = \boldsymbol{H}\boldsymbol{\Lambda}^{1/2}\boldsymbol{H}'$ is itself symmetric and positive definite and satisfies $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}$. Another matrix square root is the Cholesky decomposition, described in Section A.14.

A.10 Generalized Eigenvalues

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $k \times k$ matrices. The generalized characteristic equation is

$$\det\left(\mu\boldsymbol{B} - \boldsymbol{A}\right) = 0.$$

The solutions μ are known as **generalized eigenvalues** of $\boldsymbol{A}$ with respect to $\boldsymbol{B}$. Associated with each generalized eigenvalue μ is a **generalized eigenvector** $\boldsymbol{v}$ which satisfies

$$\boldsymbol{A}\boldsymbol{v} = \boldsymbol{B}\boldsymbol{v}\mu.$$

They are typically normalized so that $\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v} = 1$ and thus $\mu = \boldsymbol{v}'\boldsymbol{A}\boldsymbol{v}$.

A matrix expression is

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{M}$$

where $\boldsymbol{M} = \operatorname{diag}\{\mu_1, ..., \mu_k\}$.

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are real and symmetric then the generalized eigenvalues are real.

Suppose in addition that $\boldsymbol{B}$ is invertible. Then the generalized eigenvalues of $\boldsymbol{A}$ with respect to $\boldsymbol{B}$ are equal to the eigenvalues of $\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2\prime}$. The generalized eigenvectors $\boldsymbol{V}$ of $\boldsymbol{A}$ with respect to $\boldsymbol{B}$ are related to the eigenvectors $\boldsymbol{H}$ of $\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2\prime}$ by the relationship $\boldsymbol{V} = \boldsymbol{B}^{-1/2\prime}\boldsymbol{H}$. This implies $\boldsymbol{V}'\boldsymbol{B}\boldsymbol{V} = \boldsymbol{I}_k$. Thus the generalized eigenvectors are orthogonalized with respect to the matrix $\boldsymbol{B}$.

If $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{B}\boldsymbol{v}\mu$ then $(\boldsymbol{B} - \boldsymbol{A})\,\boldsymbol{v} = \boldsymbol{B}\boldsymbol{v}(1-\mu)$. So a generalized eigenvalue of $\boldsymbol{B} - \boldsymbol{A}$ with respect to $\boldsymbol{B}$ is $1-\mu$ with associated eigenvector $\boldsymbol{v}$.

Generalized eigenvalue equations have an interesting dual property. The following is based on Lemma A.9 of Johansen (1995).

Theorem A.10.1 *Suppose that $\boldsymbol{B}$ and $\boldsymbol{C}$ are invertible $p \times p$ and $r \times r$ matrices, respectively, and $\boldsymbol{A}$ is $p \times r$. Then the generalized eigenvalue problems*

$$\det\left(\mu \boldsymbol{B} - \boldsymbol{A}\boldsymbol{C}^{-1}\boldsymbol{A}'\right) = 0 \tag{A.5}$$

and

$$\det\left(\mu \boldsymbol{C} - \boldsymbol{A}'\boldsymbol{B}^{-1}\boldsymbol{A}\right) = 0 \tag{A.6}$$

have the same non-zero generalized eigenvalues. Furthermore, for any such generalized eigenvalue μ, if $\boldsymbol{v}$ and $\boldsymbol{w}$ are the associated generalized eigenvectors of (A.5) and (A.6), then

$$\boldsymbol{w} = \mu^{-1/2}\boldsymbol{C}^{-1}\boldsymbol{A}'\boldsymbol{v}. \tag{A.7}$$

Proof:. Let $\mu \neq 0$ be an eigenvalue of (A.5). Then using Theorem A.7.1.8

$$\begin{aligned} 0 &= \det\left(\mu \boldsymbol{B} - \boldsymbol{A}\boldsymbol{C}^{-1}\boldsymbol{A}'\right) \\ &= \frac{\det(\mu \boldsymbol{B})}{\det(\boldsymbol{C})}\det\left(\boldsymbol{C} - \boldsymbol{A}'(\mu \boldsymbol{B})^{-1}\boldsymbol{A}\right) \\ &= \frac{\det(\boldsymbol{B})}{\det(\boldsymbol{C})}\det\left(\mu \boldsymbol{C} - \boldsymbol{A}'\boldsymbol{B}^{-1}\boldsymbol{A}\right). \end{aligned}$$

Since $\det(\boldsymbol{B}) / \det(\boldsymbol{C}) \neq 0$ this implies (A.7) holds. Hence μ is an eigenvalue of (A.6), as claimed.

We next show that (A.7) is an eigenvector of (A.6). Note that the solutions to (A.5) and (A.6) satisfy

$$\boldsymbol{B}\boldsymbol{v}\mu = \boldsymbol{A}\boldsymbol{C}^{-1}\boldsymbol{A}'\boldsymbol{v} \tag{A.8}$$

and

$$\boldsymbol{C}\boldsymbol{w}\mu = \boldsymbol{A}'\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{w} \tag{A.9}$$

and are normalized so that $\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v} = 1$ and $\boldsymbol{w}'\boldsymbol{C}\boldsymbol{w} = 1$. We show that (A.7) satisfies (A.9). Using (A.7), we find that the left-side of (A.9) equals

$$\boldsymbol{C}\left(\mu^{-1/2}\boldsymbol{C}^{-1}\boldsymbol{A}'\right)\mu = \boldsymbol{A}'\mu^{1/2} = \boldsymbol{A}'\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{v}\mu^{1/2} = \boldsymbol{A}'\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{C}^{-1}\boldsymbol{A}'\boldsymbol{v}\mu^{-1/2} = \boldsymbol{A}'\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{w}$$

The third equality is (A.8) and the final is (A.7). This shows that (A.9) holds and thus (A.7) is an eigenvector of (A.6) as stated. ■

A.11 Extrema of Quadratic Forms

The extrema of quadratic forms in real symmetric matrices can be conveniently be written in terms of eigenvalues and eigenvectors.

Let $\boldsymbol{A}$ denote a $k \times k$ real symmetric matrix. Let $\lambda_1 \geq \cdots \geq \lambda_k$ be the ordered eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{h}_1, ..., \boldsymbol{h}_k$ the associated ordered eigenvectors.

We start with results for the extrema of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$. Throughout this Section, when we refer to the "solution" of an extremum problem, it is the solution to the normalized expression.

- $\max_{\boldsymbol{x}'\boldsymbol{x}=1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \max_{\boldsymbol{x}} \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \lambda_1$. The solution is $\boldsymbol{x} = \boldsymbol{h}_1$. (That is, the maximizer of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ over $\boldsymbol{x}'\boldsymbol{x} = 1$.)

- $\min_{x'x=1} x'Ax = \min_x \frac{x'Ax}{x'x} = \lambda_k$. The solution is $x = h_k$.

Multivariate generalizations can involve either the trace or the determinant.

- $\max_{X'X=I_\ell} \operatorname{tr}(X'AX) = \max_X \operatorname{tr}\left((X'X)^{-1}(X'AX)\right) = \sum_{i=1}^{\ell} \lambda_i$.
  The solution is $X = [h_1, ..., h_\ell]$.
- $\min_{X'X=I_\ell} \operatorname{tr}(X'AX) = \min_X \left((X'X)^{-1}(X'AX)\right) = \sum_{i=1}^{\ell} \lambda_{k-i+1}$.
  The solution is $X = [h_{k-\ell+1}, ..., h_k]$.

For a proof, see Theorem 11.13 of Magnus and Neudecker (1988).

Suppose as well that $A > 0$ with ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ and eigenvectors $[h_1, ..., h_k]$

- $\max_{X'X=I_\ell} \det(X'AX) = \max_X \frac{\det(X'AX)}{\det(X'X)} = \prod_{i=1}^{\ell} \lambda_i$. The solution is $X = [h_1, ..., h_\ell]$.
- $\min_{X'X=I_\ell} \det(X'AX) = \min_X \frac{\det(X'AX)}{\det(X'X)} = \prod_{i=1}^{\ell} \lambda_{k-i+1}$. The solution is $X = [h_{k-\ell+1}, ..., h_k]$.
- $\max_{X'X=I_\ell} \det(X'(I-A)X) = \max_X \frac{\det(X'(I-A)X)}{\det(X'X)} = \prod_{i=1}^{\ell}(1-\lambda_{k-i+1})$. The solution is $X = [h_{k-\ell+1}, ..., h_k]$.
- $\min_{X'X=I_\ell} \det(X'(I-A)X) = \min_X \frac{\det(X'(I-A)X)}{\det(X'X)} = \prod_{i=1}^{\ell}(1-\lambda_i)$. The solution is $X = [h_1, ..., h_\ell]$.

For a proof, see Theorem 11.15 of Magnus and Neudecker (1988).

We can extend the above results to incorporate generalized eigenvalue equations.

Let A and B be $k \times k$ real symmetric matrices with $B > 0$. Let $\mu_1 \geq \cdots \geq \mu_k$ be the ordered generalized eigenvalues of A with respect to B and $v_1, ..., v_k$ the associated ordered eigenvectors.

- $\max_{x'Bx=1} x'Ax = \max_x \frac{x'Ax}{x'Bx} = \mu_1$. The solution is $x = v_1$.
- $\min_{x'Bx=1} x'Ax = \min_x \frac{x'Ax}{x'Bx} = \mu_k$. The solution is $x = v_k$.
- $\max_{X'BX=I_\ell} \operatorname{tr}(X'AX) = \max_X \operatorname{tr}\left((X'BX)^{-1}(X'AX)\right) = \sum_{i=1}^{\ell} \mu_i$.
  The solution is $X = [v_1, ..., v_\ell]$.
- $\min_{X'BX=I_\ell} \operatorname{tr}(X'AX) = \min_X \operatorname{tr}\left((X'BX)^{-1}(X'AX)\right) = \sum_{i=1}^{\ell} \mu_{k-i+1}$.
  The solution is $X = [v_{k-\ell+1}, ..., v_k]$.

Suppose as well that $A > 0$.

- $\max_{\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}=\boldsymbol{I}_\ell} \det\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right) = \max_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right)} = \prod_{i=1}^{\ell} \mu_i.$

  The solution is $\boldsymbol{X} = [\boldsymbol{v}_1, ..., \boldsymbol{v}_\ell]$.

- $\min_{\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}=\boldsymbol{I}_\ell} \det\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right) = \min_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right)} = \prod_{i=1}^{\ell} \mu_{k-i+1}.$

  The solution is $\boldsymbol{X} = [\boldsymbol{v}_{k-\ell+1}, ..., \boldsymbol{v}_k]$.

- $\max_{\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}=\boldsymbol{I}_\ell} \det\left(\boldsymbol{X}'\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{X}\right) = \max_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right)} = \prod_{i=1}^{\ell} \left(1-\mu_{k-i+1}\right).$

  The solution is $\boldsymbol{X} = [\boldsymbol{v}_{k-\ell+1}, ..., \boldsymbol{v}_k]$.

- $\min_{\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}=\boldsymbol{I}_\ell} \det\left(\boldsymbol{X}'\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{X}\right) = \min_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\left(\boldsymbol{I}-\boldsymbol{A}\right)\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right)} = \prod_{i=1}^{\ell} \left(1-\mu_i\right).$

  The solution is $\boldsymbol{X} = [\boldsymbol{v}_1, ..., \boldsymbol{v}_\ell]$..

By change-of-variables, we can re-express one eigenvalue problem in terms of another. For example, let $\boldsymbol{A} > 0$, $\boldsymbol{B} > 0$, and $\boldsymbol{C} > 0$. Then

$$\max_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}\boldsymbol{X}\right)} = \max_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right)}$$

and

$$\min_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}\boldsymbol{X}\right)} = \min_{\boldsymbol{X}} \frac{\det\left(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}\right)}{\det\left(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}\right)}.$$

A.12 Idempotent Matrices

A $k \times k$ square matrix $\boldsymbol{A}$ is **idempotent** if $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$. When $k = 1$ the only idempotent numbers are 1 and 0. For $k > 1$ there are many possibilities. For example, the following matrix is idempotent

$$\boldsymbol{A} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

If $\boldsymbol{A}$ is idempotent and symmetric with rank r, then it has r eigenvalues which equal 1 and $k-r$ eigenvalues which equal 0. To see this, by the spectral decomposition we can write $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ where $\boldsymbol{H}$ is orthonormal and $\boldsymbol{\Lambda}$ contains the eigenvalues. Then

$$\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}' = \boldsymbol{H}\boldsymbol{\Lambda}^2\boldsymbol{H}'.$$

We deduce that $\boldsymbol{\Lambda}^2 = \boldsymbol{\Lambda}$ and $\lambda_i^2 = \lambda_i$ for $i = 1, ..., k$. Hence each λ_i must equal either 0 or 1. Since the rank of $\boldsymbol{A}$ is r, and the rank equals the number of positive eigenvalues, it follows that

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_{k-r} \end{bmatrix}.$$

Thus the spectral decomposition of an idempotent matrix $\boldsymbol{A}$ takes the form

$$\boldsymbol{A} = \boldsymbol{H} \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_{k-r} \end{bmatrix} \boldsymbol{H}' \tag{A.10}$$

with $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_k$. Additionally, $\operatorname{tr}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$ and $\boldsymbol{A}$ is positive semi-definite.

If $\boldsymbol{A}$ is idempotent and symmetric with rank $r < k$ then it does not possess an inverse, but its Moore-Penrose generalized inverse takes the simple form $\boldsymbol{A}^{-} = \boldsymbol{A}$. This can be verified by checking the conditions for the Moore-Penrose generalized inverse, for example $\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$.

If $\boldsymbol{A}$ is idempotent then $\boldsymbol{I} - \boldsymbol{A}$ is also idempotent.

One useful fact is that if $\boldsymbol{A}$ is idempotent then for any conformable vector $\boldsymbol{c}$,

$$\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \leq \boldsymbol{c}'\boldsymbol{c} \tag{A.11}$$

$$\boldsymbol{c}'\left(\boldsymbol{I} - \boldsymbol{A}\right)\boldsymbol{c} \leq \boldsymbol{c}'\boldsymbol{c} \tag{A.12}$$

To see this, note that

$$\boldsymbol{c}'\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} + \boldsymbol{c}'\left(\boldsymbol{I} - \boldsymbol{A}\right)\boldsymbol{c}.$$

Since $\boldsymbol{A}$ and $\boldsymbol{I} - \boldsymbol{A}$ are idempotent, they are both positive semi-definite, so both $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c}$ and $\boldsymbol{c}'\left(\boldsymbol{I} - \boldsymbol{A}\right)\boldsymbol{c}$ are non-negative. Thus they must satisfy (A.11)-(A.12).

A.13 Singular Values

The singular values of a $k \times r$ real matrix $\boldsymbol{A}$ are the positive square roots of the eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$. Thus for $j = 1, ..., r$

$$s_j = \sqrt{\lambda_j\left(\boldsymbol{A}'\boldsymbol{A}\right)}$$

Since $\boldsymbol{A}'\boldsymbol{A}$ is positive semi-definite, its eigenvalues are non-negative. Thus singular values are always real and non-negative.

The non-zero singular values of $\boldsymbol{A}$ and $\boldsymbol{A}'$ are the same.

When $\boldsymbol{A}$ is positive semi-definite then the singular values of $\boldsymbol{A}$ correspond to its eigenvalues.

The singular value decomposition of a $k \times r$ real matrix $\boldsymbol{A}$ takes the form $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}'$ where $\boldsymbol{U}$ is $k \times k$, $\boldsymbol{\Lambda}$ is $k \times r$ and $\boldsymbol{V}$ is $r \times r$, with $\boldsymbol{U}$ and $\boldsymbol{V}$ orthonormal ($\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}_k$ and $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_r$) and $\boldsymbol{\Lambda}$ is a diagonal matrix with the singular values of $\boldsymbol{A}$ on the diagonal.

It is convention to write the singular values in decending order $s_1 \geq s_2 \geq \cdots \geq s_r$.

A.14 Cholesky Decomposition

For a $k \times k$ positive definite matrix $\boldsymbol{A}$, its **Cholesky decomposition** takes the form

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}'$$

where $\boldsymbol{L}$ is **lower triangular**, and thus takes the form

$$\boldsymbol{L} = \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ L_{21} & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{k1} & L_{k2} & \cdots & L_{kk} \end{bmatrix}.$$

The diagonal elements of $\boldsymbol{L}$ are all strictly positive.

The Cholesky decomposition is unique (for positive definite $\boldsymbol{A}$). One intuition is that the matrices $\boldsymbol{A}$ and $\boldsymbol{L}$ each have $k(k+1)/2$ free elements.

The decomposition is very useful for a range of computations, especially when a matrix square root is required. Algorithms for computation are available in standard packages (for example, `chol` in either MATLAB or R).

Lower triangular matrices such as $\boldsymbol{L}$ have special properties. One is that its determinant equals the product of the diagonal elements.

Proofs of uniqueness are algorithmic. Here is one such argument for the case $k = 3$. Write out

$$\begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{21} & A_{22} & A_{32} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}' = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{bmatrix}$$
$$= \begin{bmatrix} L_{11}^2 & L_{11}L_{21} & L_{11}L_{31} \\ L_{11}L_{21} & L_{21}^2 + L_{22}^2 & L_{31}L_{21} + L_{32}L_{22} \\ L_{11}L_{31} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{bmatrix}$$

There are six equations, six knowns (the elements of $\boldsymbol{A}$) and six unknowns (the elements of $\boldsymbol{L}$). We can solve for the latter by starting with the first column, moving from top to bottom. The first element has the simple solution

$$L_{11} = \sqrt{A_{11}}.$$

This has a real solution since $A_{11} > 0$. Moving down, since L_{11} is known, for the entries beneath L_{11} we solve and find

$$L_{21} = \frac{A_{21}}{L_{11}} = \frac{A_{21}}{\sqrt{A_{11}}}$$
$$L_{31} = \frac{A_{31}}{L_{11}} = \frac{A_{31}}{\sqrt{A_{11}}}$$

Next we move to the second column. We observe that L_{21} is known. Then we solve for L_{22}

$$L_{22} = \sqrt{A_{22} - L_{21}^2} = \sqrt{A_{22} - \frac{A_{21}^2}{A_{11}}}.$$

This has a real solution since $\boldsymbol{A} > 0$. Then since L_{22} is known we can move down the column to find

$$L_{32} = \frac{A_{32} - L_{31}L_{21}}{L_{22}} = \frac{A_{32} - \frac{A_{31}A_{21}}{A_{11}}}{\sqrt{A_{22} - \frac{A_{21}^2}{A_{11}}}}.$$

Finally we take the third column. All elements except L_{33} are known. So we solve to find

$$L_{33} = \sqrt{A_{33} - L_{31}^2 - L_{32}^2} = \sqrt{A_{33} - \frac{A_{31}^2}{A_{11}} - \frac{\left(A_{32} - \frac{A_{31}A_{21}}{A_{11}}\right)^2}{A_{22} - \frac{A_{21}^2}{A_{11}}}}.$$

A.15 Matrix Calculus

Let $\boldsymbol{x} = (x_1, ..., x_k)'$ be $k \times 1$ and $g(\boldsymbol{x}) = g(x_1, ..., x_k) : \mathbb{R}^k \to \mathbb{R}$. The vector derivative is

$$\frac{\partial}{\partial \boldsymbol{x}} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(\boldsymbol{x}) \\ \vdots \\ \frac{\partial}{\partial x_k} g(\boldsymbol{x}) \end{pmatrix}$$

and

$$\frac{\partial}{\partial \boldsymbol{x}'} g(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} g(\boldsymbol{x}) & \cdots & \frac{\partial}{\partial x_k} g(\boldsymbol{x}) \end{pmatrix}.$$

Some properties are now summarized.

Theorem A.15.1 *Properties of matrix derivatives*

1. $\frac{\partial}{\partial \boldsymbol{x}} (\boldsymbol{a}'\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}} (\boldsymbol{x}'\boldsymbol{a}) = \boldsymbol{a}$

2. $\frac{\partial}{\partial \boldsymbol{x}'} (\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}$
3. $\frac{\partial}{\partial \boldsymbol{x}} (\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}')\, \boldsymbol{x}$
4. $\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}'} (\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A} + \boldsymbol{A}'$
5. $\frac{\partial}{\partial \boldsymbol{A}} \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = \boldsymbol{B}'$
6. $\frac{\partial}{\partial \boldsymbol{A}} \log\det(\boldsymbol{A}) = \left(\boldsymbol{A}^{-}\right)'$

The final two results require some justification. Recall from Section A.5 that we can write out explicitly

$$\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = \sum_i \sum_j a_{ij} b_{ji}.$$

Thus if we take the derivative with respect to a_{ij} we find

$$\frac{\partial}{\partial a_{ij}} \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = b_{ji}.$$

which is the ij^{th} element of $\boldsymbol{B}'$, establishing part 5.

For part 6, recall Laplace's expansion

$$\det \boldsymbol{A} = \sum_{j=1}^{k} a_{ij} C_{ij}.$$

where C_{ij} is the ij^{th} cofactor of $\boldsymbol{A}$. Set $\boldsymbol{C} = (C_{ij})$. Observe that C_{ij} for $j = 1, ..., k$ are not functions of a_{ij}. Thus the derivative with respect to a_{ij} is

$$\frac{\partial}{\partial a_{ij}} \log\det(\boldsymbol{A}) = (\det \boldsymbol{A})^{-1} \frac{\partial}{\partial a_{ij}} \det \boldsymbol{A} = (\det \boldsymbol{A})^{-1} C_{ij}$$

Together this implies

$$\frac{\partial}{\partial \boldsymbol{A}} \log\det(\boldsymbol{A}) = (\det \boldsymbol{A})^{-1} \boldsymbol{C} = \boldsymbol{A}^{-1}$$

where the second equality is Theorem A.7.1.12.

A.16 Kronecker Products and the Vec Operator

Let $\boldsymbol{A} = [\boldsymbol{a}_1 \ \boldsymbol{a}_2 \ \cdots \ \boldsymbol{a}_n]$ be $m \times n$. The **vec** of $\boldsymbol{A}$, denoted by $\operatorname{vec}(\boldsymbol{A})$, is the $mn \times 1$ vector

$$\operatorname{vec}(\boldsymbol{A}) = \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{pmatrix}.$$

Let $\boldsymbol{A} = (a_{ij})$ be an $m \times n$ matrix and let $\boldsymbol{B}$ be any matrix. The **Kronecker product** of $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted $\boldsymbol{A} \otimes \boldsymbol{B}$, is the matrix

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

Some important properties are now summarized. These results hold for matrices for which all matrix multiplications are conformable.

Theorem A.16.1 *Properties of the Kronecker product*

1. $(\boldsymbol{A}+\boldsymbol{B})\otimes\boldsymbol{C}=\boldsymbol{A}\otimes\boldsymbol{C}+\boldsymbol{B}\otimes\boldsymbol{C}$
2. $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D})=\boldsymbol{AC}\otimes\boldsymbol{BD}$
3. $\boldsymbol{A}\otimes(\boldsymbol{B}\otimes\boldsymbol{C})=(\boldsymbol{A}\otimes\boldsymbol{B})\otimes C$
4. $(\boldsymbol{A}\otimes\boldsymbol{B})'=\boldsymbol{A}'\otimes\boldsymbol{B}'$
5. $\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})=\operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$
6. If $\boldsymbol{A}$ is $m\times m$ and $\boldsymbol{B}$ is $n\times n$, $\det(\boldsymbol{A}\otimes\boldsymbol{B})=(\det(\boldsymbol{A}))^n(\det(\boldsymbol{B}))^m$
7. $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$
8. If $\boldsymbol{A}>0$ and $\boldsymbol{B}>0$ then $\boldsymbol{A}\otimes\boldsymbol{B}>0$
9. $\operatorname{vec}(\boldsymbol{ABC})=(\boldsymbol{C}'\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{B})$
10. $\operatorname{tr}(\boldsymbol{ABCD})=\operatorname{vec}(\boldsymbol{D}')'(\boldsymbol{C}'\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{B})$

A.17 Vector Norms

Given any vector space V (such as Euclidean space $\mathbb{R}^m$) a **norm** on V is a function $\rho : V\to\mathbb{R}$ with the properties

1. $\rho(c\boldsymbol{a})=|c|\,\rho(\boldsymbol{a})$ for any complex number c and $\boldsymbol{a}\in V$
2. $\rho(\boldsymbol{a}+\boldsymbol{b})\leq\rho(\boldsymbol{a})+\rho(\boldsymbol{b})$
3. If $\rho(\boldsymbol{a})=0$ then $\boldsymbol{a}=\boldsymbol{0}$

A seminorm on V is a function which satisfies the first two properties. The second property is known as the triangle inequality, and it is the one property which typically needs a careful demonstration (as the other two properties typically hold by inspection).

The typical norm used for Euclidean space $\mathbb{R}^m$ is the **Euclidean norm**

$$\begin{aligned}\|\boldsymbol{a}\| &= \left(\boldsymbol{a}'\boldsymbol{a}\right)^{1/2}\\ &= \left(\sum_{i=1}^m a_i^2\right)^{1/2}.\end{aligned}$$

An alternative norm is the $p-$norm (for $p\geq 1$)

$$\|\boldsymbol{a}\|_p=\left(\sum_{i=1}^m |a_i|^p\right)^{1/p}.$$

Special cases include the Euclidean norm ($p=2$), the $1-$norm

$$\|\boldsymbol{a}\|_1=\sum_{i=1}^m |a_i|$$

and the sup-norm

$$\|\boldsymbol{a}\|_\infty=\max(|a_1|,...,|a_m|).$$

For real numbers ($m=1$) these norms coincide.

Some standard inequalities for Euclidean space are now given. The Minkowski inequality given below establishes that any p-norm with $p \geq 1$ (including the Euclidean norm) satisfies the triangle inequality and is thus a valid norm.

Jensen's Inequality. If $g(\cdot) : \mathbb{R} \to \mathbb{R}$ is convex, then for any non-negative weights a_j such that $\sum_{j=1}^m a_j = 1$, and any real numbers x_j

$$g\left(\sum_{j=1}^m a_j x_j\right) \leq \sum_{j=1}^m a_j g(x_j). \tag{A.13}$$

In particular, setting $a_j = 1/m$, then

$$g\left(\frac{1}{m}\sum_{j=1}^m x_j\right) \leq \frac{1}{m}\sum_{j=1}^m g(x_j). \tag{A.14}$$

If $g(\cdot) : \mathbb{R} \to \mathbb{R}$ is concave then the inequalities in (A.13) and (A.14) are reversed.

Weighted Geometric Mean Inequality. For any non-negative real weights a_j such that $\sum_{j=1}^m a_j = 1$, and any non-negative real numbers x_j

$$x_1^{a_1} x_2^{a_2} \cdots x_m^{a_m} \leq \sum_{j=1}^m a_j x_j \tag{A.15}$$

Loève's c_r Inequality. For $r > 0$,

$$\left|\sum_{j=1}^m a_j\right|^r \leq c_r \sum_{j=1}^m |a_j|^r \tag{A.16}$$

where $c_r = 1$ when $r \leq 1$ and $c_r = m^{r-1}$ when $r \geq 1$.

c_2 **Inequality**. For any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,

$$(\boldsymbol{a} + \boldsymbol{b})' (\boldsymbol{a} + \boldsymbol{b}) \leq 2\boldsymbol{a}'\boldsymbol{a} + 2\boldsymbol{b}'\boldsymbol{b} \tag{A.17}$$

Hölder's Inequality. If $p > 1$, $q > 1$, and $1/p + 1/q = 1$, then for any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,

$$\sum_{j=1}^m |a_j b_j| \leq \|\boldsymbol{a}\|_p \|\boldsymbol{b}\|_q \tag{A.18}$$

Minkowski's Inequality. For any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, if $p \geq 1$, then

$$\|\boldsymbol{a} + \boldsymbol{b}\|_p \leq \|\boldsymbol{a}\|_p + \|\boldsymbol{b}\|_p \tag{A.19}$$

Schwarz Inequality. For any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$,

$$\left|\boldsymbol{a}'\boldsymbol{b}\right| \leq \|\boldsymbol{a}\| \|\boldsymbol{b}\|. \tag{A.20}$$

Proof of Jensen's Inequality (A.13). By the definition of convexity, for any $\lambda \in [0, 1]$

$$g(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda g(x_1) + (1 - \lambda) g(x_2). \tag{A.21}$$

This implies

$$g\left(\sum_{j=1}^{m} a_j x_j\right) = g\left(a_1 x_1 + (1-a_1)\sum_{j=2}^{m}\frac{a_j}{1-a_1}x_j\right)$$
$$\leq a_1 g(x_1) + (1-a_1)\, g\left(\sum_{j=2}^{m} b_j x_j\right)$$

where $b_j = a_j/(1-a_1)$ and $\sum_{j=2}^{m} b_j = 1$. By another application of (A.21) this is bounded by

$$a_1 g(x_1) + (1-a_1)\left(b_2 g(x_2) + (1-b_2) g\left(\sum_{j=2}^{m} c_j x_j\right)\right)$$
$$= a_1 g(x_1) + a_2 g(x_2) + (1-a_1)(1-b_2) g\left(\sum_{j=2}^{m} c_j x_j\right)$$

where $c_j = b_j/(1-b_2)$. By repeated application of (A.21) we obtain (A.13). ■

Proof of Weighted Geometric Mean Inequality. Since the logarithm is strictly concave, by Jensen's inequality

$$\log\left(x_1^{a_1} x_2^{a_2} \cdots x_m^{a_m}\right) = \sum_{j=1}^{m} a_j \log x_j \leq \log\left(\sum_{j=1}^{m} a_j x_j\right).$$

Applying the exponential yields (A.15). ■

Proof of Loève's c_r Inequality. For $r \geq 1$ this is simply a rewriting of the finite form Jensen's inequality (A.14) with $g(u) = u^r$. For $r < 1$, define $b_j = |a_j| / \left(\sum_{j=1}^{m} |a_j|\right)$. The facts that $0 \leq b_j \leq 1$ and $r < 1$ imply $b_j \leq b_j^r$ and thus

$$1 = \sum_{j=1}^{m} b_j \leq \sum_{j=1}^{m} b_j^r$$

which implies

$$\left(\sum_{j=1}^{m} |a_j|\right)^r \leq \sum_{j=1}^{m} |a_j|^r.$$

The proof is completed by observing that

$$\left(\sum_{j=1}^{m} a_j\right)^r \leq \left(\sum_{j=1}^{m} |a_j|\right)^r.$$

■

Proof of c_2 Inequality. By the c_r inequality, $(a_j + b_j)^2 \leq 2a_j^2 + 2b_j^2$. Thus

$$(\boldsymbol{a}+\boldsymbol{b})'(\boldsymbol{a}+\boldsymbol{b}) = \sum_{j=1}^{m} (a_j + b_j)^2$$
$$\leq 2\sum_{j=1}^{m} a_j^2 + 2\sum_{j=1}^{m} b_j^2$$
$$= 2\boldsymbol{a}'\boldsymbol{a} + 2\boldsymbol{b}'\boldsymbol{b}$$

■

Proof of Hölder's Inequality. Set $u_j = |a_j|^p \,/\, \|\boldsymbol{a}\|_p^p$ and $u_j = |b_j|^q \,/\, \|\boldsymbol{b}\|_q^q$ and observe that $\sum_{j=1}^m u_j = 1$ and $\sum_{j=1}^m v_j = 1$. By the weighted geometric mean inequality,

$$u_j^{1/p} v_j^{1/q} \leq \frac{u_j}{p} + \frac{v_j}{q}.$$

Then since $\sum_{j=1}^m u_j = 1$, $\sum_{j=1}^m v_j = 1$ and $1/p + 1/q = 1$

$$\frac{\sum_{j=1}^m |a_j b_j|}{\|\boldsymbol{a}\|_p \|\boldsymbol{b}\|_q} = \sum_{j=1}^m u_j^{1/p} v_j^{1/q} \leq \sum_{j=1}^m \left(\frac{u_j}{p} + \frac{v_j}{q} \right) = 1$$

which is (A.18). ■

Proof of Minkowski's Inequality. Se $q = p/(p-1)$ so that $1/p + 1/q = 1$. Using the triangle inequality for real numbers and two applications of Hölder's inequality

$$\begin{aligned}
\|\boldsymbol{a} + \boldsymbol{b}\|_p^p &= \sum_{j=1}^m |a_j + b_j|^p \\
&= \sum_{j=1}^m |a_j + b_j| \, |a_j + b_j|^{p-1} \\
&\leq \sum_{j=1}^m |a_j| \, |a_j + b_j|^{p-1} + \sum_{j=1}^m |b_j| \, |a_j + b_j|^{p-1} \\
&\leq \|\boldsymbol{a}\|_p \left(\sum_{j=1}^m |a_j + b_j|^{(p-1)q} \right)^{1/q} + \|\boldsymbol{b}\|_p \left(\sum_{j=1}^m |a_j + b_j|^{(p-1)q} \right)^{1/q} \\
&= \left(\|\boldsymbol{a}\|_p + \|\boldsymbol{b}\|_p \right) \|\boldsymbol{a} + \boldsymbol{b}\|_p^{p-1}
\end{aligned}$$

Solving, we find (A.19). ■

Proof of Schwarz Inequality. Using Hölder's inequality with $p = q = 2$

$$\left| \boldsymbol{a}'\boldsymbol{b} \right| \leq \sum_{j=1}^m |a_j b_j| \leq \|\boldsymbol{a}\| \, \|\boldsymbol{b}\|$$

■

A.18 Matrix Norms

Two common norms used for matrix spaces are the **Frobenius norm** and the **spectral norm**. We can write either as $\|\boldsymbol{A}\|$, but may write $\|\boldsymbol{A}\|_F$ or $\|\boldsymbol{A}\|_2$ when we want to be specific.

The **Frobenius norm** of an $m \times k$ matrix $\boldsymbol{A}$ is the Euclidean norm applied to its elements

$$\begin{aligned}
\|\boldsymbol{A}\|_F &= \|\mathrm{vec}\,(\boldsymbol{A})\| \\
&= \left(\mathrm{tr}\left(\boldsymbol{A}'\boldsymbol{A} \right) \right)^{1/2} \\
&= \left(\sum_{i=1}^m \sum_{j=1}^k a_{ij}^2 \right)^{1/2}.
\end{aligned}$$

When $m \times m$ $\boldsymbol{A}$ is real symmetric then

$$\|\boldsymbol{A}\|_F = \left(\sum_{\ell=1}^{m} \lambda_\ell^2\right)^{1/2}$$

where λ_ℓ, $\ell = 1, ..., m$ are the eigenvalues of $\boldsymbol{A}$. To see this, by the spectral decomposition $\boldsymbol{A} = \boldsymbol{H\Lambda H}'$ with $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}$ and $\boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1, ..., \lambda_m\}$, so

$$\|\boldsymbol{A}\|_F = \left(\operatorname{tr}\left(\boldsymbol{H\Lambda H}'\boldsymbol{H\Lambda H}'\right)\right)^{1/2} = \left(\operatorname{tr}\left(\boldsymbol{\Lambda\Lambda}\right)\right)^{1/2} = \left(\sum_{\ell=1}^{m} \lambda_\ell^2\right)^{1/2}. \tag{A.22}$$

A useful calculation is for any $m \times 1$ vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, using (A.1),

$$\left\|\boldsymbol{ab}'\right\|_F = \operatorname{tr}\left(\boldsymbol{ba}'\boldsymbol{ab}'\right)^{1/2} = \left(\boldsymbol{b}'\boldsymbol{ba}'\boldsymbol{a}\right)^{1/2} = \|\boldsymbol{a}\|\,\|\boldsymbol{b}\| \tag{A.23}$$

and in particular

$$\left\|\boldsymbol{aa}'\right\|_F = \|\boldsymbol{a}\|^2. \tag{A.24}$$

The **spectral norm** of an $m \times k$ real matrix $\boldsymbol{A}$ is its largest singular value

$$\|\boldsymbol{A}\|_2 = s_{\max}(\boldsymbol{A}) = \left(\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2}$$

where $\lambda_{\max}(\boldsymbol{B})$ denotes the largest eigenvalue of the matrix $\boldsymbol{B}$. Notice that

$$\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right) = \left\|\boldsymbol{A}'\boldsymbol{A}\right\|_2$$

so

$$\|\boldsymbol{A}\|_2 = \left\|\boldsymbol{A}'\boldsymbol{A}\right\|_2^{1/2}.$$

If $\boldsymbol{A}$ is $m \times m$ and symmetric with eigenvalues λ_j then

$$\|\boldsymbol{A}\|_2 = \max_{j \le m} |\lambda_j|.$$

The Frobenius and spectral norms are closely related. They are equivalent when applied to a matrix of rank 1, since $\left\|\boldsymbol{ab}'\right\|_2 = \|\boldsymbol{a}\|\,\|\boldsymbol{b}\| = \left\|\boldsymbol{ab}'\right\|_F$. In general, for $m \times k$ matrix $\boldsymbol{A}$ with rank r

$$\|\boldsymbol{A}\|_2 = \left(\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2} \le \left(\sum_{j=1}^{k} \lambda_j\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2} = \|\boldsymbol{A}\|_F.$$

Since $\boldsymbol{A}'\boldsymbol{A}$ also has rank at most r, it has at most r non-zero eigenvalues, and hence

$$\|\boldsymbol{A}\|_F = \left(\sum_{j=1}^{k} \lambda_j\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2} = \left(\sum_{j=1}^{r} \lambda_j\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2} \le \left(r\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right)\right)^{1/2} = \sqrt{r}\,\|\boldsymbol{A}\|_2.$$

Given any vector norm $\|\boldsymbol{a}\|$ the **induced matrix norm** is defined as

$$\|\boldsymbol{A}\| = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{Ax}\| = \sup_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|}.$$

To see that this is a norm we need to check that it satisfies the triangle inequality. Indeed

$$\|\boldsymbol{A}+\boldsymbol{B}\| = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{Ax}+\boldsymbol{Bx}\| \le \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{Ax}\| + \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{Bx}\| = \|\boldsymbol{A}\| + \|\boldsymbol{B}\|.$$

For any vector $\boldsymbol{x}$, by the definition of the induced norm

$$\|\boldsymbol{Ax}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{x}\|$$

a property which is called consistent norms.

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be conformable and $\|\boldsymbol{A}\|$ an induced matrix norm. Then using the property of consistent norms

$$\|\boldsymbol{AB}\| = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{ABx}\| \leq \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{A}\| \, \|\boldsymbol{Bx}\| = \|\boldsymbol{A}\| \, \|\boldsymbol{B}\| \, .$$

A matrix norm which satisfies this property is called a **sub-multiplicative norm**, and is a matrix form of the Schwarz inequality.

Of particular interest, the matrix norm induced by the Euclidean vector norm is the spectral norm. Indeed,

$$\sup_{\boldsymbol{x}'\boldsymbol{x}=1} \|\boldsymbol{Ax}\|^2 = \sup_{\boldsymbol{x}'\boldsymbol{x}=1} \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{Ax} = \lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right) = \|\boldsymbol{A}\|_2^2 \, .$$

It follows that the spectral norm is consistent with the Euclidean norm, and is sub-multiplicative.

A.19 Matrix Inequalities

Schwarz Matrix Inequality: For any $m \times k$ and $k \times m$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and either the Frobenius or spectral norm,

$$\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{B}\| \, . \tag{A.25}$$

Triangle Inequality: For any $m \times k$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and either the Frobenius or spectral norm,

$$\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\| \, . \tag{A.26}$$

Trace Inequality. For any $m \times m$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\boldsymbol{A}$ is symmetric and $\boldsymbol{B} \geq 0$

$$\operatorname{tr}(\boldsymbol{AB}) \leq \|\boldsymbol{A}\|_2 \operatorname{tr}(\boldsymbol{B}) \, . \tag{A.27}$$

Quadratic Inequality. For any $m \times 1$ $\boldsymbol{b}$ and $m \times m$ symmetric matrix $\boldsymbol{A}$

$$\boldsymbol{b}'\boldsymbol{Ab} \leq \|\boldsymbol{A}\|_2 \, \boldsymbol{b}'\boldsymbol{b} \tag{A.28}$$

Strong Schwarz Matrix Inequality. For any conformable matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

$$\|\boldsymbol{AB}\|_F \leq \|\boldsymbol{A}\|_2 \, \|\boldsymbol{B}\|_F \, . \tag{A.29}$$

Norm Equivalence. For any $m \times k$ matrix $\boldsymbol{A}$ of rank r

$$\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_F \leq \sqrt{r} \, \|\boldsymbol{A}\|_2 \, . \tag{A.30}$$

Eigenvalue Product Inequality. For any $m \times m$ real symmetric matrices $\boldsymbol{A} \geq 0$ and $\boldsymbol{B} \geq 0$, the eigenvalues $\lambda_\ell(\boldsymbol{AB})$ are real and satisfy

$$\lambda_{\min}(\boldsymbol{A}) \, \lambda_{\min}(\boldsymbol{B}) \leq \lambda_\ell(\boldsymbol{AB}) \leq \lambda_{\max}(\boldsymbol{A}) \, \lambda_{\max}(\boldsymbol{B}) \tag{A.31}$$

(Zhang and Zhang, 2006, Corollary 11)

Proof of Schwarz Matrix Inequality: The inequality holds for the spectral norm since it is an induced norm. Now consider the Frobenius norm. Partition $\boldsymbol{A}' = [\boldsymbol{a}_1, ..., \boldsymbol{a}_n]$ and $\boldsymbol{B} = [\boldsymbol{b}_1, ..., \boldsymbol{b}_n]$.

Then by partitioned matrix multiplication, the definition of the Frobenius norm and the Schwarz inequality for vectors

$$
\begin{aligned}
\|\boldsymbol{A}\boldsymbol{B}\|_F &= \left\| \begin{matrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 & \cdots \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 & \cdots \\ \vdots & \vdots & \ddots \end{matrix} \right\|_F \\
&\le \left\| \begin{matrix} \|\boldsymbol{a}_1\|\,\|\boldsymbol{b}_1\| & \|\boldsymbol{a}_1\|\,\|\boldsymbol{b}_2\| & \cdots \\ \|\boldsymbol{a}_2\|\,\|\boldsymbol{b}_1\| & \|\boldsymbol{a}_2\|\,\|\boldsymbol{b}_2\| & \cdots \\ \vdots & \vdots & \ddots \end{matrix} \right\|_F \\
&= \left(\sum_{i=1}^m \sum_{j=1}^m \|\boldsymbol{a}_i\|^2 \|\boldsymbol{b}_j\|^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^m \|\boldsymbol{a}_i\|^2 \right)^{1/2} \left(\sum_{i=1}^m \|\boldsymbol{b}_i\|^2 \right)^{1/2} \\
&= \left(\sum_{i=1}^k \sum_{j=1}^m \boldsymbol{a}_{ji}^2 \right)^{1/2} \left(\sum_{i=1}^m \sum_{j=1}^k \|\boldsymbol{b}_{ji}\|^2 \right)^{1/2} \\
&= \|\boldsymbol{A}\|_F \, \|\boldsymbol{B}\|_F
\end{aligned}
$$

■

Proof of Triangle Inequality: The inequality holds for the spectral norm since it is an induced norm. Now consider the Frobenius norm. Let $\boldsymbol{a} = \text{vec}(\boldsymbol{A})$ and $\boldsymbol{b} = \text{vec}(\boldsymbol{B})$. Then by the definition of the Frobenius norm and the Schwarz Inequality for vectors

$$
\begin{aligned}
\|\boldsymbol{A}+\boldsymbol{B}\|_F &= \|\text{vec}(\boldsymbol{A}+\boldsymbol{B})\|_F \\
&= \|\boldsymbol{a}+\boldsymbol{b}\| \\
&\le \|\boldsymbol{a}\| + \|\boldsymbol{b}\| \\
&= \|\boldsymbol{A}\|_F + \|\boldsymbol{B}\|_F
\end{aligned}
$$

■

Proof of Trace Inequality. By the spectral decomposition for symmetric matices, $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ where $\boldsymbol{\Lambda}$ has the eigenvalues λ_j of $\boldsymbol{A}$ on the diagonal and $\boldsymbol{H}$ is orthonormal. Define $\boldsymbol{C} = \boldsymbol{H}'\boldsymbol{B}\boldsymbol{H}$ which has non-negative diagonal elements C_{jj} since $\boldsymbol{B}$ is positive semi-definite. Then

$$
\text{tr}(\boldsymbol{A}\boldsymbol{B}) = \text{tr}(\boldsymbol{\Lambda}\boldsymbol{C}) = \sum_{j=1}^m \lambda_j C_{jj} \le \max_j |\lambda_j| \sum_{j=1}^m C_{jj} = \|\boldsymbol{A}\|_2 \,\text{tr}(\boldsymbol{C})
$$

where the inequality uses the fact that $C_{jj} \ge 0$. But note that

$$
\text{tr}(\boldsymbol{C}) = \text{tr}\left(\boldsymbol{H}'\boldsymbol{B}\boldsymbol{H}\right) = \text{tr}\left(\boldsymbol{H}\boldsymbol{H}'\boldsymbol{B}\right) = \text{tr}(\boldsymbol{B})
$$

since $\boldsymbol{H}$ is orthonormal. Thus $\text{tr}(\boldsymbol{A}\boldsymbol{B}) \le \|\boldsymbol{A}\|_2 \,\text{tr}(\boldsymbol{B})$ as stated. ■

Proof of Quadratic Inequality: In the Trace Inequality set $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$ and note $\text{tr}(\boldsymbol{A}\boldsymbol{B}) = \boldsymbol{b}'\boldsymbol{A}\boldsymbol{b}$ and $\text{tr}(\boldsymbol{B}) = \boldsymbol{b}'\boldsymbol{b}$. ■

Proof of Strong Schwarz Matrix Inequality. By the definition of the Frobenius norm, the property of the trace, the Trace Inequality (noting that both $\boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{B}\boldsymbol{B}'$ are symmetric and positive semi-definite), and the Schwarz matrix inequality

$$
\begin{aligned}
\|\boldsymbol{A}\boldsymbol{B}\|_F &= \left(\operatorname{tr}\left(\boldsymbol{B}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{B}\right)\right)^{1/2} \\
&= \left(\operatorname{tr}\left(\boldsymbol{A}'\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}'\right)\right)^{1/2} \\
&\leq \left(\left\|\boldsymbol{A}'\boldsymbol{A}\right\|_2 \operatorname{tr}\left(\boldsymbol{B}\boldsymbol{B}'\right)\right)^{1/2} \\
&= \|\boldsymbol{A}\|_2 \|\boldsymbol{B}\|_F .
\end{aligned}
$$

■