

Appendix A

Matrix Algebra

A.1 Notation

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

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

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

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

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

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

By convention a_{ij} refers to the element in the i^{th} row and j^{th} column of \mathbf{A} . If $r = 1$ then \mathbf{A} is a column vector. If $k = 1$ then \mathbf{A} is a row vector. If $r = k = 1$, then \mathbf{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 (\mathbf{a}), and matrices by upper-case bold italics (\mathbf{A}). Sometimes a matrix \mathbf{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,

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

where

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

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

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

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

A matrix is **square** if $k = r$. A square matrix is **symmetric** if $\mathbf{A} = \mathbf{A}'$, 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

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

A **partitioned matrix** takes the form

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

where the \mathbf{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 \mathbf{x} are real we say that \mathbf{x} is a real vector, and similarly for matrices.

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

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

A.3 Matrix Addition

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

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

Matrix addition follows the commutative and associative laws:

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

A.4 Matrix Multiplication

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

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

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

$$\mathbf{a}'\mathbf{b} = a_1b_1 + a_2b_2 + \cdots + a_kb_k = \sum_{j=1}^k a_jb_j.$$

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

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

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

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

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

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

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

As another example,

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

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

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

The $k \times r$ matrix \mathbf{H} , $r \leq k$, is called **orthonormal** if $\mathbf{H}'\mathbf{H} = \mathbf{I}_r$. This means that the columns of \mathbf{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 \mathbf{A} is the sum of its diagonal elements

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

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

$$\begin{aligned}\text{tr}(c\mathbf{A}) &= c \text{tr}(\mathbf{A}) \\ \text{tr}(\mathbf{A}') &= \text{tr}(\mathbf{A}) \\ \text{tr}(\mathbf{A} + \mathbf{B}) &= \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}) \\ \text{tr}(\mathbf{I}_k) &= k.\end{aligned}$$

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

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}). \quad (\text{A.1})$$

Indeed,

$$\begin{aligned}\text{tr}(\mathbf{AB}) &= \text{tr} \begin{bmatrix} a'_1 b_1 & a'_1 b_2 & \cdots & a'_1 b_k \\ a'_2 b_1 & a'_2 b_2 & \cdots & a'_2 b_k \\ \vdots & \vdots & & \vdots \\ a'_k b_1 & a'_k b_2 & \cdots & a'_k b_k \end{bmatrix} \\ &= \sum_{i=1}^k a'_i b_i \\ &= \sum_{i=1}^k b'_i a_i \\ &= \text{tr}(\mathbf{BA}).\end{aligned}$$

A.6 Rank and Inverse

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

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_r]$$

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

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

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

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

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

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

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

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

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{BC}(\mathbf{C} + \mathbf{CDA}^{-1}\mathbf{BC})^{-1}\mathbf{CDA}^{-1}. \quad (\text{A.2})$$

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

$$(\mathbf{A} - \mathbf{bb}')^{-1} = \mathbf{A}^{-1} + (1 - \mathbf{b}'\mathbf{A}^{-1}\mathbf{b})^{-1}\mathbf{A}^{-1}\mathbf{bb}'\mathbf{A}^{-1}. \quad (\text{A.3})$$

The following fact about inverting partitioned matrices is quite useful.

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11.2}^{-1} & -\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22.1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22.1}^{-1} \end{bmatrix} \quad (\text{A.4})$$

where $\mathbf{A}_{11.2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ and $\mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. There are alternative algebraic representations for the components. For example, using the Woodbury matrix identity you can show the following alternative expressions

$$\begin{aligned}\mathbf{A}^{11} &= \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22.1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\ \mathbf{A}^{22} &= \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{A}^{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22.1}^{-1} \\ \mathbf{A}^{21} &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{A}_{11.2}^{-1}\end{aligned}$$

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

$$\begin{aligned}\mathbf{A}\mathbf{A}^{-}\mathbf{A} &= \mathbf{A} \\ \mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} &= \mathbf{A}^{-} \\ \mathbf{A}\mathbf{A}^{-} &\text{ is symmetric} \\ \mathbf{A}^{-}\mathbf{A} &\text{ is symmetric}\end{aligned}$$

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

For example, if

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

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

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

A.7 Determinant

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

While the determinant is widely used, its precise definition is rarely needed. However, we present the definition here for completeness. Let $\mathbf{A} = (a_{ij})$ be a $k \times k$ matrix. Let $\pi = (j_1, \dots, j_k)$ denote a permutation of $(1, \dots, k)$. 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 $(1, \dots, k)$), and let $\varepsilon_\pi = +1$ if this count is even and $\varepsilon_\pi = -1$ if the count is odd. Then the determinant of \mathbf{A} is defined as

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

For example, if \mathbf{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

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

For a square matrix \mathbf{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 \mathbf{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 \mathbf{A} to its cofactors:

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

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

Theorem A.7.1 *Properties of the determinant*

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

A.8 Eigenvalues

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

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

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 \mathbf{A} . If λ is an eigenvalue of \mathbf{A} , then $\lambda \mathbf{I}_k - \mathbf{A}$ is singular so there exists a non-zero vector \mathbf{h} such that $(\lambda \mathbf{I}_k - \mathbf{A}) \mathbf{h} = \mathbf{0}$ or

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

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

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

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

We now state some useful properties.

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

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

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

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

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

If \mathbf{A} is real, symmetric, and invertible, then by the spectral decomposition and the properties of orthonormal matrices, $\mathbf{A}^{-1} = \mathbf{H}'^{-1}\mathbf{\Lambda}^{-1}\mathbf{H}^{-1} = \mathbf{H}\mathbf{\Lambda}^{-1}\mathbf{H}'$. Thus the columns of \mathbf{H} are also the eigenvectors of \mathbf{A}^{-1} , and its eigenvalues are $\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_k^{-1}$.

A.9 Positive Definite Matrices

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

Some properties include:

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

1. If $\mathbf{A} = \mathbf{G}'\mathbf{B}\mathbf{G}$ with $\mathbf{B} \geq 0$ and some matrix \mathbf{G} , then \mathbf{A} is positive semi-definite. (For any $\mathbf{c} \neq \mathbf{0}$, $\mathbf{c}'\mathbf{A}\mathbf{c} = \boldsymbol{\alpha}'\mathbf{B}\boldsymbol{\alpha} \geq 0$ where $\boldsymbol{\alpha} = \mathbf{G}\mathbf{c}$.) If \mathbf{G} has full column rank and $\mathbf{B} > 0$, then \mathbf{A} is positive definite.
2. If \mathbf{A} is positive definite, then \mathbf{A} is non-singular and \mathbf{A}^{-1} exists. Furthermore, $\mathbf{A}^{-1} > 0$.
3. $\mathbf{A} > 0$ if and only if it is symmetric and all its eigenvalues are positive.
4. By the spectral decomposition, $\mathbf{A} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$ where $\mathbf{H}'\mathbf{H} = \mathbf{I}_k$ and $\boldsymbol{\Lambda}$ is diagonal with non-negative diagonal elements. All diagonal elements of $\boldsymbol{\Lambda}$ are strictly positive if (and only if) $\mathbf{A} > 0$.
5. The rank of \mathbf{A} equals the number of strictly positive eigenvalues.
6. If $\mathbf{A} > 0$ then $\mathbf{A}^{-1} = \mathbf{H}\boldsymbol{\Lambda}^{-1}\mathbf{H}'$.
7. If $\mathbf{A} \geq 0$ and $\text{rank}(\mathbf{A}) = r \leq k$ then the Moore-Penrose generalized inverse of \mathbf{A} is $\mathbf{A}^- = \mathbf{H}\boldsymbol{\Lambda}^-\mathbf{H}'$ where $\boldsymbol{\Lambda}^- = \text{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_r^{-1}, 0, \dots, 0)$.
8. If $\mathbf{A} \geq 0$ we can find a matrix \mathbf{B} such that $\mathbf{A} = \mathbf{B}\mathbf{B}'$. We call \mathbf{B} a **matrix square root** of \mathbf{A} and is typically written as $\mathbf{B} = \mathbf{A}^{1/2}$. The matrix \mathbf{B} need not be unique. One matrix square root is obtained using the spectral decomposition $\mathbf{A} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}'$. Then $\mathbf{B} = \mathbf{H}\boldsymbol{\Lambda}^{1/2}\mathbf{H}'$ is itself symmetric and positive definite and satisfies $\mathbf{A} = \mathbf{B}\mathbf{B}$. Another matrix square root is the Cholesky decomposition, described in Section A.14.

A.10 Generalized Eigenvalues

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

$$\det(\mu\mathbf{B} - \mathbf{A}) = 0.$$

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

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

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

A matrix expression is

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

where $\mathbf{M} = \text{diag}\{\mu_1, \dots, \mu_k\}$.

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

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

If $\mathbf{A}\mathbf{v} = \mathbf{B}\mathbf{v}\mu$ then $(\mathbf{B} - \mathbf{A})\mathbf{v} = \mathbf{B}\mathbf{v}(1 - \mu)$. So a generalized eigenvalue of $\mathbf{B} - \mathbf{A}$ with respect to \mathbf{B} is $1 - \mu$ with associated eigenvector \mathbf{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 \mathbf{B} and \mathbf{C} are invertible $p \times p$ and $r \times r$ matrices, respectively, and \mathbf{A} is $p \times r$. Then the generalized eigenvalue problems*

$$\det(\mu\mathbf{B} - \mathbf{A}\mathbf{C}^{-1}\mathbf{A}') = 0 \quad (\text{A.5})$$

and

$$\det(\mu\mathbf{C} - \mathbf{A}'\mathbf{B}^{-1}\mathbf{A}) = 0 \quad (\text{A.6})$$

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

$$\mathbf{w} = \mu^{-1/2}\mathbf{C}^{-1}\mathbf{A}'\mathbf{v}. \quad (\text{A.7})$$

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

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

Since $\det(\mathbf{B})/\det(\mathbf{C}) \neq 0$ 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

$$\mathbf{B}\mathbf{v}\mu = \mathbf{A}\mathbf{C}^{-1}\mathbf{A}'\mathbf{v} \quad (\text{A.8})$$

and

$$\mathbf{C}\mathbf{w}\mu = \mathbf{A}'\mathbf{B}^{-1}\mathbf{A}\mathbf{w} \quad (\text{A.9})$$

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

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

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

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

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

- $\min_{\mathbf{x}'\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x} = \min_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_k$. The solution is $\mathbf{x} = \mathbf{h}_k$.

Multivariate generalizations can involve either the trace or the determinant.

- $\max_{\mathbf{X}'\mathbf{X}=\mathbf{I}_\ell} \text{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \max_{\mathbf{X}} \text{tr}\left((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{X})\right) = \sum_{i=1}^{\ell} \lambda_i$.

The solution is $\mathbf{X} = [\mathbf{h}_1, \dots, \mathbf{h}_\ell]$.

- $\min_{\mathbf{X}'\mathbf{X}=\mathbf{I}_\ell} \text{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \min_{\mathbf{X}} \text{tr}\left((\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{X})\right) = \sum_{i=1}^{\ell} \lambda_{k-i+1}$.

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

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

Suppose as well that $\mathbf{A} > 0$ with ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ and eigenvectors $[\mathbf{h}_1, \dots, \mathbf{h}_k]$

- $\max_{\mathbf{X}'\mathbf{X}=\mathbf{I}_\ell} \det(\mathbf{X}'\mathbf{A}\mathbf{X}) = \max_{\mathbf{X}} \frac{\det(\mathbf{X}'\mathbf{A}\mathbf{X})}{\det(\mathbf{X}'\mathbf{X})} = \prod_{i=1}^{\ell} \lambda_i$. The solution is $\mathbf{X} = [\mathbf{h}_1, \dots, \mathbf{h}_\ell]$.
- $\min_{\mathbf{X}'\mathbf{X}=\mathbf{I}_\ell} \det(\mathbf{X}'\mathbf{A}\mathbf{X}) = \min_{\mathbf{X}} \frac{\det(\mathbf{X}'\mathbf{A}\mathbf{X})}{\det(\mathbf{X}'\mathbf{X})} = \prod_{i=1}^{\ell} \lambda_{k-i+1}$. The solution is $\mathbf{X} = [\mathbf{h}_{k-\ell+1}, \dots, \mathbf{h}_k]$.
- $\max_{\mathbf{X}'\mathbf{X}=\mathbf{I}_\ell} \det(\mathbf{X}'(\mathbf{I} - \mathbf{A})\mathbf{X}) = \max_{\mathbf{X}} \frac{\det(\mathbf{X}'(\mathbf{I} - \mathbf{A})\mathbf{X})}{\det(\mathbf{X}'\mathbf{X})} = \prod_{i=1}^{\ell} (1 - \lambda_{k-i+1})$. The solution is $\mathbf{X} = [\mathbf{h}_{k-\ell+1}, \dots, \mathbf{h}_k]$.
- $\min_{\mathbf{X}'\mathbf{X}=\mathbf{I}_\ell} \det(\mathbf{X}'(\mathbf{I} - \mathbf{A})\mathbf{X}) = \min_{\mathbf{X}} \frac{\det(\mathbf{X}'(\mathbf{I} - \mathbf{A})\mathbf{X})}{\det(\mathbf{X}'\mathbf{X})} = \prod_{i=1}^{\ell} (1 - \lambda_i)$. The solution is $\mathbf{X} = [\mathbf{h}_1, \dots, \mathbf{h}_\ell]$.

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

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

Let \mathbf{A} and \mathbf{B} be $k \times k$ real symmetric matrices with $\mathbf{B} > 0$. Let $\mu_1 \geq \dots \geq \mu_k$ be the ordered generalized eigenvalues of \mathbf{A} with respect to \mathbf{B} and $\mathbf{v}_1, \dots, \mathbf{v}_k$ the associated ordered eigenvectors.

- $\max_{\mathbf{x}'\mathbf{B}\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x} = \max_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mu_1$. The solution is $\mathbf{x} = \mathbf{v}_1$.
- $\min_{\mathbf{x}'\mathbf{B}\mathbf{x}=1} \mathbf{x}'\mathbf{A}\mathbf{x} = \min_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mu_k$. The solution is $\mathbf{x} = \mathbf{v}_k$.
- $\max_{\mathbf{X}'\mathbf{B}\mathbf{X}=\mathbf{I}_\ell} \text{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \max_{\mathbf{X}} \text{tr}\left((\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{X})\right) = \sum_{i=1}^{\ell} \mu_i$.
The solution is $\mathbf{X} = [\mathbf{v}_1, \dots, \mathbf{v}_\ell]$.
- $\min_{\mathbf{X}'\mathbf{B}\mathbf{X}=\mathbf{I}_\ell} \text{tr}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \min_{\mathbf{X}} \text{tr}\left((\mathbf{X}'\mathbf{B}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{X})\right) = \sum_{i=1}^{\ell} \mu_{k-i+1}$.
The solution is $\mathbf{X} = [\mathbf{v}_{k-\ell+1}, \dots, \mathbf{v}_k]$.

Suppose as well that $\mathbf{A} > 0$.

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

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

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

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

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

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

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

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

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

$$\max_{\mathbf{X}} \frac{\det(\mathbf{X}'\mathbf{C}\mathbf{A}\mathbf{C}\mathbf{X})}{\det(\mathbf{X}'\mathbf{C}\mathbf{B}\mathbf{C}\mathbf{X})} = \max_{\mathbf{X}} \frac{\det(\mathbf{X}'\mathbf{A}\mathbf{X})}{\det(\mathbf{X}'\mathbf{B}\mathbf{X})}$$

and

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

A.12 Idempotent Matrices

A $k \times k$ square matrix \mathbf{A} is **idempotent** if $\mathbf{A}\mathbf{A} = \mathbf{A}$. 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

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

If \mathbf{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 $\mathbf{A} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}'$ where \mathbf{H} is orthonormal and $\mathbf{\Lambda}$ contains the eigenvalues. Then

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

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

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

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

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

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

If \mathbf{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 $\mathbf{A}^- = \mathbf{A}$. This can be verified by checking the conditions for the Moore-Penrose generalized inverse, for example $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}\mathbf{A}\mathbf{A} = \mathbf{A}$.

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

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

$$\mathbf{c}'\mathbf{A}\mathbf{c} \leq \mathbf{c}'\mathbf{c} \quad (\text{A.11})$$

$$\mathbf{c}'(\mathbf{I} - \mathbf{A})\mathbf{c} \leq \mathbf{c}'\mathbf{c} \quad (\text{A.12})$$

To see this, note that

$$\mathbf{c}'\mathbf{c} = \mathbf{c}'\mathbf{A}\mathbf{c} + \mathbf{c}'(\mathbf{I} - \mathbf{A})\mathbf{c}.$$

Since \mathbf{A} and $\mathbf{I} - \mathbf{A}$ are idempotent, they are both positive semi-definite, so both $\mathbf{c}'\mathbf{A}\mathbf{c}$ and $\mathbf{c}'(\mathbf{I} - \mathbf{A})\mathbf{c}$ 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 \mathbf{A} are the positive square roots of the eigenvalues of $\mathbf{A}'\mathbf{A}$. Thus for $j = 1, \dots, r$

$$s_j = \sqrt{\lambda_j(\mathbf{A}'\mathbf{A})}$$

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

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

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

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

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

A.14 Cholesky Decomposition

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

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

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

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

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

The Cholesky decomposition is unique (for positive definite \mathbf{A}). One intuition is that the matrices \mathbf{A} and \mathbf{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 \mathbf{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

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

There are six equations, six knowns (the elements of \mathbf{A}) and six unknowns (the elements of \mathbf{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 $\mathbf{A} > 0$. 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 $\mathbf{x} = (x_1, \dots, x_k)'$ be $k \times 1$ and $g(\mathbf{x}) = g(x_1, \dots, x_k) : \mathbb{R}^k \rightarrow \mathbb{R}$. The vector derivative is

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

and

$$\frac{\partial}{\partial \mathbf{x}'} g(\mathbf{x}) = \left(\frac{\partial}{\partial x_1} g(\mathbf{x}) \quad \cdots \quad \frac{\partial}{\partial x_k} g(\mathbf{x}) \right).$$

Some properties are now summarized.

Theorem A.15.1 *Properties of matrix derivatives*

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

2. $\frac{\partial}{\partial \mathbf{x}'} (\mathbf{A}\mathbf{x}) = \mathbf{A}$
3. $\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}'\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}')\mathbf{x}$
4. $\frac{\partial^2}{\partial \mathbf{x}\partial \mathbf{x}'} (\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathbf{A} + \mathbf{A}'$
5. $\frac{\partial}{\partial \mathbf{A}} \text{tr}(\mathbf{B}\mathbf{A}) = \mathbf{B}'$
6. $\frac{\partial}{\partial \mathbf{A}} \log \det(\mathbf{A}) = (\mathbf{A}^{-1})'$

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

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

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

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

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

For part 6, recall Laplace's expansion

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

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

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

Together this implies

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

where the second equality is Theorem A.7.1.12.

A.16 Kronecker Products and the Vec Operator

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

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

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

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

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

A.17 Vector Norms

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

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

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**

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

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

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

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

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

and the sup-norm

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

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 $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ 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

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

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

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

If $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ 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

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

Loève's c_r Inequality. For $r > 0$,

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

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 \mathbf{a} and \mathbf{b} ,

$$(\mathbf{a} + \mathbf{b})'(\mathbf{a} + \mathbf{b}) \leq 2\mathbf{a}'\mathbf{a} + 2\mathbf{b}'\mathbf{b} \quad (\text{A.17})$$

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

$$\sum_{j=1}^m |a_j b_j| \leq \|\mathbf{a}\|_p \|\mathbf{b}\|_q \quad (\text{A.18})$$

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

$$\|\mathbf{a} + \mathbf{b}\|_p \leq \|\mathbf{a}\|_p + \|\mathbf{b}\|_p \quad (\text{A.19})$$

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

$$|\mathbf{a}'\mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|. \quad (\text{A.20})$$

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

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

This implies

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

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

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

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

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

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

Applying the exponential yields (A.15). \blacksquare

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

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

\blacksquare

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

$$\begin{aligned} (\mathbf{a} + \mathbf{b})'(\mathbf{a} + \mathbf{b}) &= \sum_{j=1}^m (a_j + b_j)^2 \\ &\leq 2 \sum_{j=1}^m a_j^2 + 2 \sum_{j=1}^m b_j^2 \\ &= 2\mathbf{a}'\mathbf{a} + 2\mathbf{b}'\mathbf{b} \end{aligned}$$

■

Proof of Hölder's Inequality. Set $u_j = |a_j|^p / \|\mathbf{a}\|_p^p$ and $v_j = |b_j|^q / \|\mathbf{b}\|_q^q$ 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 $1/p + 1/q = 1$

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

which is (A.18). ■

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

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

Solving, we find (A.19). ■

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

$$|\mathbf{a}'\mathbf{b}| \leq \sum_{j=1}^m |a_j b_j| \leq \|\mathbf{a}\| \|\mathbf{b}\|$$

■

A.18 Matrix Norms

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

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

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

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

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

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

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

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

$$\|\mathbf{a}\mathbf{b}'\|_F = \text{tr}(\mathbf{b}\mathbf{a}'\mathbf{a}\mathbf{b}')^{1/2} = (\mathbf{b}'\mathbf{b}\mathbf{a}'\mathbf{a})^{1/2} = \|\mathbf{a}\| \|\mathbf{b}\| \quad (\text{A.23})$$

and in particular

$$\|\mathbf{a}\mathbf{a}'\|_F = \|\mathbf{a}\|^2. \quad (\text{A.24})$$

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

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

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

$$\lambda_{\max}(\mathbf{A}'\mathbf{A}) = \|\mathbf{A}'\mathbf{A}\|_2$$

so

$$\|\mathbf{A}\|_2 = \|\mathbf{A}'\mathbf{A}\|_2^{1/2}.$$

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

$$\|\mathbf{A}\|_2 = \max_{j \leq m} |\lambda_j|.$$

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

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

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

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

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

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

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

$$\|\mathbf{A} + \mathbf{B}\| = \sup_{\mathbf{x}'\mathbf{x}=1} \|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x}\| \leq \sup_{\mathbf{x}'\mathbf{x}=1} \|\mathbf{A}\mathbf{x}\| + \sup_{\mathbf{x}'\mathbf{x}=1} \|\mathbf{B}\mathbf{x}\| = \|\mathbf{A}\| + \|\mathbf{B}\|.$$

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

$$\|\mathbf{Ax}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|$$

a property which is called consistent norms.

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

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

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,

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

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 \mathbf{A} and \mathbf{B} , and either the Frobenius or spectral norm,

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|. \quad (\text{A.25})$$

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

$$\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|. \quad (\text{A.26})$$

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

$$\text{tr}(\mathbf{AB}) \leq \|\mathbf{A}\|_2 \text{tr}(\mathbf{B}). \quad (\text{A.27})$$

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

$$\mathbf{b}'\mathbf{Ab} \leq \|\mathbf{A}\|_2 \mathbf{b}'\mathbf{b} \quad (\text{A.28})$$

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

$$\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_F. \quad (\text{A.29})$$

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

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{r} \|\mathbf{A}\|_2. \quad (\text{A.30})$$

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

$$\lambda_{\min}(\mathbf{A}) \lambda_{\min}(\mathbf{B}) \leq \lambda_\ell(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A}) \lambda_{\max}(\mathbf{B}) \quad (\text{A.31})$$

(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 $\mathbf{A}' = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ and $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]$.

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

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

■

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

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

■

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

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{\Lambda C}) = \sum_{j=1}^m \lambda_j C_{jj} \leq \max_j |\lambda_j| \sum_{j=1}^m C_{jj} = \|\mathbf{A}\|_2 \text{tr}(\mathbf{C})$$

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

$$\text{tr}(\mathbf{C}) = \text{tr}(\mathbf{H}'\mathbf{B}\mathbf{H}) = \text{tr}(\mathbf{H}\mathbf{H}'\mathbf{B}) = \text{tr}(\mathbf{B})$$

since \mathbf{H} is orthonormal. Thus $\text{tr}(\mathbf{AB}) \leq \|\mathbf{A}\|_2 \text{tr}(\mathbf{B})$ as stated. ■

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

Proof of Strong Schwarz Matrix Inequality. By the definition of the Frobenius norm, the property of the trace, the Trace Inequality (noting that both $\mathbf{A}'\mathbf{A}$ and $\mathbf{B}\mathbf{B}'$ are symmetric and

positive semi-definite), and the Schwarz matrix inequality

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

■

Appendix B

Probability Inequalities

The following bounds are used frequently in econometric theory, predominantly in asymptotic analysis.

Monotone Probability Inequality. For any events A and B such that $A \subset B$,

$$\Pr(A) \leq \Pr(B). \quad (\text{B.1})$$

Union Equality. For any events A and B ,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \quad (\text{B.2})$$

Boole's Inequality (Union Bound). For any events A and B ,

$$\Pr(A \cup B) \leq \Pr(A) + \Pr(B). \quad (\text{B.3})$$

Bonferroni's Inequality. For any events A and B ,

$$\Pr(A \cap B) \geq \Pr(A) + \Pr(B) - 1. \quad (\text{B.4})$$

Jensen's Inequality. If $g(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$ is convex, then for any random vector \mathbf{x} for which

$$\mathbb{E} \|\mathbf{x}\| < \infty \text{ and } \mathbb{E} |g(\mathbf{x})| < \infty,$$

$$g(\mathbb{E}(\mathbf{x})) \leq \mathbb{E}(g(\mathbf{x})). \quad (\text{B.5})$$

If $g(\cdot)$ concave, then the inequality is reversed.

Conditional Jensen's Inequality. If $g(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$ is convex, then for any random vectors (\mathbf{y}, \mathbf{x}) for which $\mathbb{E} \|\mathbf{y}\| < \infty$ and $\mathbb{E} \|g(\mathbf{y})\| < \infty$,

$$g(\mathbb{E}(\mathbf{y} \mid \mathbf{x})) \leq \mathbb{E}(g(\mathbf{y}) \mid \mathbf{x}). \quad (\text{B.6})$$

If $g(\cdot)$ concave, then the inequality is reversed.

Conditional Expectation Inequality. For any $r \geq 1$ such that $\mathbb{E} |y|^r < \infty$, then

$$\mathbb{E} (|\mathbb{E}(y \mid \mathbf{x})|^r) \leq \mathbb{E} (|y|^r) < \infty. \quad (\text{B.7})$$

Expectation Inequality. For any random matrix \mathbf{Y} for which $\mathbb{E} \|\mathbf{Y}\| < \infty$,

$$\|\mathbb{E}(\mathbf{Y})\| \leq \mathbb{E} \|\mathbf{Y}\|. \quad (\text{B.8})$$

Hölder's Inequality. If $p > 1$ and $q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then for any random $m \times n$ matrices \mathbf{X} and \mathbf{Y} ,

$$\mathbb{E} \|\mathbf{X}'\mathbf{Y}\| \leq (\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} (\mathbb{E}(\|\mathbf{Y}\|^q))^{1/q}. \quad (\text{B.9})$$

Cauchy-Schwarz Inequality. For any random $m \times n$ matrices \mathbf{X} and \mathbf{Y} ,

$$\mathbb{E} \|\mathbf{X}'\mathbf{Y}\| \leq \left(\mathbb{E}(\|\mathbf{X}\|^2) \right)^{1/2} \left(\mathbb{E}(\|\mathbf{Y}\|^2) \right)^{1/2}. \quad (\text{B.10})$$

Matrix Cauchy-Schwarz Inequality. Tripathi (1999). For any random $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^\ell$,

$$\mathbb{E}(\mathbf{y}\mathbf{x}') (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1} \mathbb{E}(\mathbf{x}\mathbf{y}') \leq \mathbb{E}(\mathbf{y}\mathbf{y}') \quad (\text{B.11})$$

Minkowski's Inequality. For any random $m \times n$ matrices \mathbf{X} and \mathbf{Y} ,

$$(\mathbb{E}(\|\mathbf{X} + \mathbf{Y}\|^p))^{1/p} \leq (\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} + (\mathbb{E}(\|\mathbf{Y}\|^p))^{1/p} \quad (\text{B.12})$$

Liapunov's Inequality. For any random $m \times n$ matrix \mathbf{X} and $1 \leq r \leq p$,

$$(\mathbb{E}(\|\mathbf{X}\|^r))^{1/r} \leq (\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} \quad (\text{B.13})$$

Markov's Inequality (standard form). For any random vector \mathbf{x} and non-negative function $g(\mathbf{x}) \geq 0$,

$$\Pr(g(\mathbf{x}) > \alpha) \leq \alpha^{-1} \mathbb{E}(g(\mathbf{x})). \quad (\text{B.14})$$

Markov's Inequality (strong form). For any random vector \mathbf{x} and non-negative function $g(\mathbf{x}) \geq 0$,

$$\Pr(g(\mathbf{x}) > \alpha) \leq \alpha^{-1} \mathbb{E}(g(\mathbf{x}) \mathbf{1}(g(\mathbf{x}) > \alpha)). \quad (\text{B.15})$$

Chebyshev's Inequality. For any random variable x ,

$$\Pr(|x - \mathbb{E}x| > \alpha) \leq \frac{\text{var}(x)}{\alpha^2}. \quad (\text{B.16})$$

Proof of Monotone Probability Inequality. Since $A \subset B$ then $B = A \cup \{B \cap A^c\}$ where A^c is the complement of A . The sets A and $\{B \cap A^c\}$ are disjoint. Thus

$$\Pr(B) = \Pr(A \cup \{B \cap A^c\}) = \Pr(A) + \Pr(B \cap A^c) \geq \Pr(A)$$

since probabilities are non-negative. Thus $\Pr(A) \leq \Pr(B)$ as claimed. \blacksquare

Proof of Union Equality. $\{A \cup B\} = A \cup \{B \cap A^c\}$ where A and $\{B \cap A^c\}$ are disjoint. Also $B = \{B \cap A\} \cup \{B \cap A^c\}$ where $\{B \cap A\}$ and $\{B \cap A^c\}$ are disjoint. These two relationships imply

$$\begin{aligned} \Pr(A \cup B) &= \Pr(A) + \Pr(B \cap A^c) \\ \Pr(B) &= \Pr(B \cap A) + \Pr(B \cap A^c). \end{aligned}$$

Subtracting,

$$\Pr(A \cup B) - \Pr(B) = \Pr(A) - \Pr(B \cap A)$$

which is (B.2) upon rearrangement. ■

Proof of Boole's Inequality. From the Union Equality and $\Pr(A \cap B) \geq 0$,

$$\begin{aligned}\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\ &\leq \Pr(A) + \Pr(B)\end{aligned}$$

as claimed. ■

Proof of Bonferroni's Inequality. Rearranging the Union Equality and using $\Pr(A \cup B) \leq 1$

$$\begin{aligned}\Pr(A \cap B) &= \Pr(A) + \Pr(B) - \Pr(A \cup B) \\ &\geq \Pr(A) + \Pr(B) - 1\end{aligned}$$

which is (B.4). ■

Proof of Jensen's Inequality. Since $g(\mathbf{u})$ is convex, at any point \mathbf{u} there is a nonempty set of subderivatives (linear surfaces touching $g(\mathbf{u})$ at \mathbf{u} but lying below $g(\mathbf{u})$ for all \mathbf{u}). Let $a + \mathbf{b}'\mathbf{u}$ be a subderivative of $g(\mathbf{u})$ at $\mathbf{u} = \mathbb{E}(\mathbf{x})$. Then for all \mathbf{u} , $g(\mathbf{u}) \geq a + \mathbf{b}'\mathbf{u}$ yet $g(\mathbb{E}(\mathbf{x})) = a + \mathbf{b}'\mathbb{E}(\mathbf{x})$. Applying expectations, $\mathbb{E}(g(\mathbf{x})) \geq a + \mathbf{b}'\mathbb{E}(\mathbf{x}) = g(\mathbb{E}(\mathbf{x}))$, as stated. ■

Proof of Conditional Jensen's Inequality. The same as the proof of Jensen's Inequality, but using conditional expectations. The conditional expectations exist since $\mathbb{E}\|\mathbf{y}\| < \infty$ and $\mathbb{E}\|g(\mathbf{y})\| < \infty$. ■

Proof of Conditional Expectation Inequality. As the function $|u|^r$ is convex for $r \geq 1$, the Conditional Jensen's inequality implies

$$|\mathbb{E}(y \mid \mathbf{x})|^r \leq \mathbb{E}(|y|^r \mid \mathbf{x}).$$

Taking unconditional expectations and the law of iterated expectations, we obtain

$$\mathbb{E}(|\mathbb{E}(y \mid \mathbf{x})|^r) \leq \mathbb{E}(\mathbb{E}(|y|^r \mid \mathbf{x})) = \mathbb{E}(|y|^r) < \infty$$

as required. ■

Proof of Expectation Inequality. By the Triangle inequality, for $\lambda \in [0, 1]$,

$$\|\lambda \mathbf{U}_1 + (1 - \lambda) \mathbf{U}_2\| \leq \lambda \|\mathbf{U}_1\| + (1 - \lambda) \|\mathbf{U}_2\|$$

which shows that the matrix norm $g(\mathbf{U}) = \|\mathbf{U}\|$ is convex. Applying Jensen's Inequality (B.5) we find (B.8). ■

Proof of Hölder's Inequality. Since $\frac{1}{p} + \frac{1}{q} = 1$ an application of the discrete Jensen's Inequality (A.13) shows that for any real a and b

$$\exp \left[\frac{1}{p}a + \frac{1}{q}b \right] \leq \frac{1}{p} \exp(a) + \frac{1}{q} \exp(b).$$

Setting $u = \exp(a)$ and $v = \exp(b)$ this implies

$$u^{1/p} v^{1/q} \leq \frac{u}{p} + \frac{v}{q}$$

and this inequality holds for any $u > 0$ and $v > 0$.

Set $u = \|\mathbf{X}\|^p / \mathbb{E}(\|\mathbf{X}\|^p)$ and $v = \|\mathbf{Y}\|^q / \mathbb{E}(\|\mathbf{Y}\|^q)$. Note that $\mathbb{E}(u) = \mathbb{E}(v) = 1$. By the matrix Schwarz Inequality (A.25), $\|\mathbf{X}'\mathbf{Y}\| \leq \|\mathbf{X}\| \|\mathbf{Y}\|$. Thus

$$\begin{aligned} \frac{\mathbb{E} \|\mathbf{X}'\mathbf{Y}\|}{(\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} (\mathbb{E}(\|\mathbf{Y}\|^q))^{1/q}} &\leq \frac{\mathbb{E}(\|\mathbf{X}\| \|\mathbf{Y}\|)}{(\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} (\mathbb{E}(\|\mathbf{Y}\|^q))^{1/q}} \\ &= \mathbb{E}\left(u^{1/p} v^{1/q}\right) \\ &\leq \mathbb{E}\left(\frac{u}{p} + \frac{v}{q}\right) \\ &= \frac{1}{p} + \frac{1}{q} \\ &= 1, \end{aligned}$$

which is (B.9). \blacksquare

Proof of Cauchy-Schwarz Inequality. Special case of Hölder's with $p = q = 2$.

Proof of Matrix Cauchy-Schwarz Inequality. Define $e = \mathbf{y} - (\mathbb{E}(\mathbf{y}\mathbf{x}'))(\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-}\mathbf{x}$. Note that $\mathbb{E}(ee') \geq 0$ is positive semi-definite. We can calculate that

$$\mathbb{E}(ee') = \mathbb{E}(\mathbf{y}\mathbf{y}') - (\mathbb{E}(\mathbf{y}\mathbf{x}'))(\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-}\mathbb{E}(\mathbf{x}\mathbf{y}').$$

Since the left-hand-side is positive semi-definite, so is the right-hand-side, which means $\mathbb{E}(\mathbf{y}\mathbf{y}') \geq (\mathbb{E}(\mathbf{y}\mathbf{x}'))(\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-}\mathbb{E}(\mathbf{x}\mathbf{y}')$ as stated. \blacksquare

Proof of Liapunov's Inequality. The function $g(u) = u^{p/r}$ is convex for $u > 0$ since $p \geq r$. Set $u = \|\mathbf{X}\|^r$. By Jensen's inequality, $g(\mathbb{E}(u)) \leq \mathbb{E}(g(u))$ or

$$(\mathbb{E}(\|\mathbf{X}\|^r))^{p/r} \leq \mathbb{E}\left((\|\mathbf{X}\|^r)^{p/r}\right) = \mathbb{E}(\|\mathbf{X}\|^p).$$

Raising both sides to the power $1/p$ yields $(\mathbb{E}(\|\mathbf{X}\|^r))^{1/r} \leq (\mathbb{E}(\|\mathbf{X}\|^p))^{1/p}$ as claimed. \blacksquare

Proof of Minkowski's Inequality. Note that by rewriting, using the triangle inequality (A.26), and then Hölder's Inequality to the two expectations

$$\begin{aligned} \mathbb{E}(\|\mathbf{X} + \mathbf{Y}\|^p) &= \mathbb{E}\left(\|\mathbf{X} + \mathbf{Y}\| \|\mathbf{X} + \mathbf{Y}\|^{p-1}\right) \\ &\leq \mathbb{E}\left(\|\mathbf{X}\| \|\mathbf{X} + \mathbf{Y}\|^{p-1}\right) + \mathbb{E}\left(\|\mathbf{Y}\| \|\mathbf{X} + \mathbf{Y}\|^{p-1}\right) \\ &\leq (\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} \mathbb{E}\left(\left(\|\mathbf{X} + \mathbf{Y}\|^{q(p-1)}\right)^{1/q}\right) \\ &\quad + (\mathbb{E}(\|\mathbf{Y}\|^p))^{1/p} \mathbb{E}\left(\left(\|\mathbf{X} + \mathbf{Y}\|^{q(p-1)}\right)^{1/q}\right) \\ &= \left((\mathbb{E}(\|\mathbf{X}\|^p))^{1/p} + (\mathbb{E}(\|\mathbf{Y}\|^p))^{1/p}\right) \mathbb{E}\left((\|\mathbf{X} + \mathbf{Y}\|^p)^{(p-1)/p}\right) \end{aligned}$$

where the second equality picks q to satisfy $1/p + 1/q = 1$, and the final equality uses this fact to make the substitution $q = p/(p-1)$ and then collects terms. Dividing both sides by $\mathbb{E}\left((\|\mathbf{X} + \mathbf{Y}\|^p)^{(p-1)/p}\right)$, we obtain (B.12). \blacksquare

Proof of Markov's Inequality. Let F denote the distribution function of \mathbf{x} . Then

$$\begin{aligned} \Pr(g(\mathbf{x}) \geq \alpha) &= \int_{\{g(\mathbf{u}) \geq \alpha\}} dF(\mathbf{u}) \\ &\leq \int_{\{g(\mathbf{u}) \geq \alpha\}} \frac{g(\mathbf{u})}{\alpha} dF(\mathbf{u}) \\ &= \alpha^{-1} \int 1(g(\mathbf{u}) > \alpha) g(\mathbf{u}) dF(\mathbf{u}) \\ &= \alpha^{-1} \mathbb{E}(g(\mathbf{x}) 1(g(\mathbf{x}) > \alpha)) \end{aligned}$$

the inequality using the region of integration $\{g(\mathbf{u}) > \alpha\}$. This establishes the strong form (B.15). Since $1(g(\mathbf{x}) > \alpha) \leq 1$, the final expression is less than $\alpha^{-1} \mathbb{E}(g(\mathbf{x}))$, establishing the standard form (B.14). ■

Proof of Chebyshev's Inequality. Define $y = (x - \mathbb{E}x)^2$ and note that $\mathbb{E}(y) = \text{var}(x)$. The events $\{|x - \mathbb{E}x| > \alpha\}$ and $\{y > \alpha^2\}$ are equal, so by an application Markov's inequality we find

$$\Pr(|x - \mathbb{E}x| > \alpha) = \Pr(y > \alpha^2) \leq \alpha^{-2} \mathbb{E}(y) = \alpha^{-2} \text{var}(x)$$

as stated. ■