Chapter 2. ANALYSIS AND LINEAR ALGEBRA IN A NUTSHELL

1. Some Elements of Mathematical Analysis

Real numbers are denoted by lower case Greek or Roman numbers; the space of real numbers is the real line, denoted by $\mathbb{R}$. The *absolute value* of a real number a is denoted by $|a|$. Complex numbers are rarely required in econometrics before the study of time series and dynamic systems. For future reference, a complex number is written $a + b\iota$, where a and b are real numbers and $\iota = \sqrt{-1}$, with a termed the *real* part and ιb termed the *imaginary* part. The complex number can also be written as $r(\cos\theta + \iota\sin\theta)$, where $r = \sqrt{a^2+b^2}$ is the *modulus* of the number and $\theta = \cos^{-1}(a/r)$. The properties of complex numbers we will need in basic econometrics are the rules for sums, $(a+\iota b) + (c+\iota d) = (a+c)+\iota(b+d)$, and products, $(a+\iota b)\cdot(c+\iota d) = (ab-cd)+\iota(ad+bc)$.

For sets of objects $\mathbb{A}$ and $\mathbb{B}$, the *union* $\mathbb{A}\cup\mathbb{B}$ is the set of objects in either; the *intersection* $\mathbb{A}\cap\mathbb{B}$ is the set of objects in both; and $\mathbb{A}\backslash\mathbb{B}$ is the set of objects in $\mathbb{A}$ that are not in $\mathbb{B}$. The empty set is denoted $\emptyset$. Set inclusion is denoted $\mathbb{A} \subseteq \mathbb{B}$; we say $\mathbb{A}$ *is contained in* $\mathbb{B}$. The complement of a set $\mathbb{A}$ (relative to a set $\mathbb{B}$ that contains it) is denoted $\mathbb{A}^c$. A family of sets is *disjoint* if the intersection of each pair is empty. The symbol $a \in \mathbb{A}$ means that a is a member of $\mathbb{A}$; and $a \notin \mathbb{A}$ means that a is not a member of $\mathbb{A}$. The symbol $\exists$ means "there exists", the symbol $\forall$ means "for all", and the symbol $\ni$ means "such that".

The terms *function*, *mapping*, and *transformation* are used synonymously, and the notation $f:\mathbb{A} \longrightarrow \mathbb{B}$ will mean that each object a in the *domain* $\mathbb{A}$ is mapped into an object $b = f(a)$ in the *range* $\mathbb{B}$. The symbol $f(\mathbb{C})$, termed the *image* of $\mathbb{C}$, is used for the set of all objects $f(a)$ for $a \in \mathbb{C}$. For $\mathbb{D} \subseteq \mathbb{B}$, the symbol $f^{-1}(\mathbb{D})$ denotes the *inverse image* of $\mathbb{D}$: the set of all $a \in \mathbb{A}$ such that $f(a) \in \mathbb{D}$. The function f is *onto* if $\mathbb{B} = f(\mathbb{A})$; it is *one-to-one* if it is onto and if $a,c \in \mathbb{A}$ and $a \neq c$ implies $f(a) \neq f(c)$. When f is one-to-one, the mapping f^{-1} is a function from $\mathbb{B}$ onto $\mathbb{A}$. If $\mathbb{C} \subseteq \mathbb{A}$, define the function $\mathbf{1}_{\mathbb{C}}:\mathbb{A} \longrightarrow \mathbb{R}$ by $\mathbf{1}_{\mathbb{C}}(a) = 1$ for $a \in \mathbb{C}$, and $\mathbf{1}_{\mathbb{C}}(a) = 0$ otherwise; this is called the *indicator function* for the set $\mathbb{C}$. A function is termed *real-valued* if its range is $\mathbb{R}$.

If $\mathbb{A}$ is a set of real numbers, then the *infimum* of $\mathbb{A}$, denoted inf $\mathbb{A}$, is the greatest real number that is less than or equal to every number in $\mathbb{A}$. The *supremum* of $\mathbb{A}$, denoted sup $\mathbb{A}$, is the least real number that is greater than or equal to every number in $\mathbb{A}$. A typical application has a function $f:\mathbb{C} \longrightarrow \mathbb{R}$ and $\mathbb{A} = f(\mathbb{C})$; then $\sup_{c\in\mathbb{C}} f(c)$ is used to denote sup $\mathbb{A}$. If the supremum is achieved by an object $d \in \mathbb{C}$, so $f(d) = \sup_{c\in\mathbb{C}} f(c)$, then we write $f(d) = \max_{c\in\mathbb{C}} f(c)$ and $d = \operatorname{argmax}_{c\in\mathbb{C}} f(c)$. This notation is ambigious when there is a non-unique maximizing argument; we will assume that $\operatorname{argmax}_{c\in\mathbb{C}} f(c)$ is a *selection* of any one of the maximizing arguments. Analogous definitons hold for inf and min.

If a_i is a sequence of real numbers indexed by i = 1,2,..., then the sequence is said to have a *limit* (equal to a_o) if for each $\varepsilon > 0$, there exists n such that $|a_i - a_o| < \varepsilon$ for all $i \geq n$; the notation for a limit is $\lim_{i\to\infty} a_i = a_o$ or $a_i \longrightarrow a_o$. The *Cauchy criterion* says that a sequence a_i has a limit if and only if, for each $\varepsilon > 0$, there exists n such that $|a_i - a_j| < \varepsilon$ for $i,j \geq n$. The notation $\limsup_{i\longrightarrow\infty} a_i$ means the limit of the supremum of the sets $\{a_i,a_{i+1},...\}$; because it is nonincreasing, it always exists (but may equal $+\infty$ or $-\infty$). An analogous definition holds for liminf.

A real-valued function $\rho(a,b)$ defined for pairs of objects in a set $\mathbb{A}$ is a *distance function* if it is non-negative, gives a positive distance between all distinct points of $\mathbb{A}$, has $\rho(a,b) = \rho(b,a)$, and satisfies the triangle inequality $\rho(a,b) \le \rho(a,c) + \rho(c,b)$. A set $\mathbb{A}$ with a distance function ρ is termed a *metric space*. A typical example is the real line $\mathbb{R}$, with the absolute value of the difference of two numbers taken as the distance between them. A (ε-)*neighborhood* of a point a in a metric space $\mathbb{A}$ (for $\varepsilon > 0$) is a set of the form $\{b\in\mathbb{A} \mid \rho(a,b) < \varepsilon\}$. A set $\mathbb{C} \subseteq \mathbb{A}$ is *open* if for each point in $\mathbb{C}$, some neighborhood of this point is also contained in $\mathbb{C}$. A set $\mathbb{C} \subseteq \mathbb{A}$ is *closed* if its complement is open. The *closure* of a set $\mathbb{C}$ is the intersection of all closed sets that contain $\mathbb{C}$. The *interior* of $\mathbb{C}$ is the union of all open sets contained in $\mathbb{C}$. A *covering* of a set $\mathbb{C}$ is a family of open sets whose union contains $\mathbb{C}$. The set $\mathbb{C}$ is said to be *compact* if every covering contains a finite sub-family which is also a covering. A family of sets is said to have the *finite-intersection property* if every finite sub-family has a non-empty intersection. Another characterization of a compact set is that every family of closed subsets with the finite intersection property has a non-empty intersection. A metric space $\mathbb{A}$ is *separable* if there exists a countable subset $\mathbb{B}$ such that every neighborhood contains a member of $\mathbb{B}$. All of the metric spaces encountered in econometrics will be separable. A sequence a_i in a separable metric space $\mathbb{A}$ is *convergent* (to a point a_o) if the sequence is eventually contained in each neighborhood of a; we write $a_i \longrightarrow a_o$ or $\lim_{i\to\infty} a_i = a_o$ to denote a convergent sequence. A set $\mathbb{C} \subseteq \mathbb{A}$ is compact if and only if every sequence in $\mathbb{C}$ has a convergent subsequence (which converges to a *cluster point* of the original sequence).

Consider separable metric spaces $\mathbb{A}$ and $\mathbb{B}$, and a function $f:\mathbb{A} \longrightarrow \mathbb{B}$. The function f is *continuous* on $\mathbb{A}$ if the inverse image of every open set is open. Another characterization of continuity is that for any sequence satisfying $a_i \longrightarrow a_o$, one has $f(a_i) \longrightarrow f(a_o)$; the function is said to be continuous on $\mathbb{C}$ if this property holds for each $a_o \in \mathbb{C}$. Stated another way, f is continuous on $\mathbb{C}$ if for each $\varepsilon > 0$ and $a \in \mathbb{C}$, there exists $\delta > 0$ such that for each b in a δ-neighborhood of a, f(b) is in a ε-neighborhood of f(a). Continuity of real-valued functions f and g is preserved by the operations of absolute value $|f(a)|$, multiplication $f(a)\cdot g(a)$, addition $f(a)+g(a)$, and maxima or minima $\max\{f(a),g(a)\}$ and $\min\{f(a),g(a)\}$. The function f is *uniformly continuous* on $\mathbb{C}$ if for each $\varepsilon > 0$, there exists $\delta > 0$ such that for all $a \in \mathbb{C}$ and $b \in \mathbb{A}$ with b in a δ-neighborhood of a, one has f(b) in a ε-neighborhood of f(a). The distinction between continuity and uniform continuity is that for the latter a single $\delta > 0$ works for <u>all</u> $a \in \mathbb{C}$. A function that is continuous on a compact set is uniformly continuous.

Consider a real-valued function f on $\mathbb{R}$. The *derivative* of f at a_o, denoted $f'(a_o)$, $\nabla f(a_o)$, or $df(a_o)/da$, has the property if it exists that $|f(b) - f(a_o) - f'(a_o)(b-a_o)| \le \varepsilon(b-a_o)\cdot(b-a_o)$, where $\lim_{c\to 0} \varepsilon(c) = 0$. The function is *continuously differentiable* at a_o if f′ is a continuous function at a_o. If a function is k-times continuously differentiable in a neighborhood of a point a_o, then it has a *Taylor's expansion*

$$f(b) = \sum_{i=0}^{k} f^{(i)}(a_o)\cdot\frac{(b-a_o)^i}{i!} + \left\{f^{(k)}(\lambda b+(1-\lambda)a_o) - f^{(k)}(a_o)\right\}\cdot\frac{(b-a_o)^k}{k!},$$

where $f^{(i)}$ denotes the i-th derivative, λ is a scalar between zero and one, and b is in the neighborhood.

The exponential function e^a, also written exp(a), and natural logarithm log(a) appear frequently in econometrics. The exponential function is defined for both real and complex arguments, and has the properties that $e^{a+b} = e^a e^b$, $e^0 = 1$, and the Taylor's expansion $e^a = \sum_{i=0}^{\infty} \frac{a^i}{i!}$ that is valid for all a. The trigonometric functions cos(a) and sin(a) are also defined for both real and complex arguments, and have Taylor's expansions $\cos(a) = \sum_{i=0}^{\infty} \frac{(-1)^i a^{2i}}{(2i)!}$, and $\sin(a) = \sum_{i=0}^{\infty} \frac{(-1)^i a^{2i+1}}{(2i+1)!}$. These expansions combine to show that $e^{a+\iota b} = e^a(\cos(b) + \iota\sin(b))$. The logarithm is defined for positive arguments, and has the properties that log(1) = 0, $\log(a\cdot b) = \log(a) + \log(b)$, and $\log(e^a) = a$. It has a Taylor's expansion $\log(1+a) = \sum_{i=1}^{\infty} a^i$, valid for $|a| < 1$. A useful bound on logarithms is that for $|a| < 1/3$ and $|b| < 1/3$, $|\text{Log}(1+a+b) - a| < 4|b| + 3|a|^2$.

2. Vectors and Linear Spaces

2.1. A finite-dimensional linear space is a set with the properties (a) that linear combinations of points in the set are defined and are again in the set, and (b) there is a finite number of points in the set (a *basis*) such that every point in the set is a linear combination of this finite number of points. The dimension of the space is the minimum number of points needed to form a basis. A point x in a linear space of dimension n has a *ordinate representation* $x = (x_1,x_2,...,x_n)$, given a *basis* for the space $\{b_1,...,b_n\}$, where $x_1,...,x_n$ are real numbers such that $x = x_1 b_1 + ... + x_n b_n$. The point x is called a *vector*, and $x_1,...,x_n$ are called its *components*. The notation $(x)_i$ will sometimes also be used for component i of a vector x. In econometrics, we work mostly with *finite-dimensional real space*. When this space is of dimension n, it is denoted $\mathbb{R}^n$. Points in this space are vectors of real numbers $(x_1,...,x_n)$; this corresponds to the previous terminology with the *basis* for $\mathbb{R}^n$ being the *unit vectors* (1,0,..,0), (0,1,0,..,0),..., (0,..,0,1). Usually, we assume this representation without being explicit about the basis for the space.

However, it is worth noting that the coordinate representation of a vector depends on the particular basis chosen for a space. Sometimes this fact can be used to choose bases in which vectors and transformations have particularly simple coordinate representations.

The *Euclidean norm* of a vector x is $\|x\|_2 = \sqrt{x_1^2+...+x_n^2}$. This norm can be used to define the distance between vectors, or neighborhoods of a vector. Other possible norms are $\|x\|_1 = |x_1|+...+|x_n|$, $\|x\|_\infty = \max\ \{|x_1|,...,|x_n|\}$, or for $1 \le p < +\infty$, $\|x\|_p = \left[|x_1|^p+...+|x_n|^p\right]^{\frac{1}{p}}$ Each norm defines a *topology* on the linear space, based on neighborhoods of a vector that are less than each positive distance away. The space $\mathbb{R}^n$ with the norm $\|x\|_2$ and associated topology is called *Euclidean* n-*space*.

The *vector product* of x and y in $\mathbb{R}^n$ is defined as

$$x{\cdot}y = x_1y_1+...+x_ny_n.$$

Other notations for vector products are <x,y> or (when x and y are interpreted as *row* vectors) xy′ or (when x and y are interpreted as column vectors) x′y.

2.2. A *linear subspace* of a linear space such as $\mathbb{R}^n$ is a subset that has the property that all linear combinations of its members remain in the subset. Examples of linear subspaces are the plane $\{(a,b,c)|\,b = 0\}$ and the line $\{(a,b,c)|\,a = b = 2{\cdot}c\}$ in $\mathbb{R}^3$. The linear subspace *spanned* by a set of vectors $\{x_1,...,x_J\}$ is the set of all linear combinations of these vectors, $\mathbb{L} = \{x_1\alpha_1+...+x_J\alpha_J|(\alpha_1,...,\alpha_J) \in \mathbb{R}^J\}$. The vectors $\{x_1,...,x_J\}$ are *linearly independent* if and only if one cannot be written as a linear combination of the remainder. The linear subspace that is spanned by a set of J linearly independent vectors is said to be *of dimension* J. Conversely, each linear space of dimension J can be represented as the set of linear combinations of J linearly independent vectors, which are in fact a basis for the subspace. A linear subspace of dimension one is a *line* (through the origin), and a linear subspace of dimension (n-1) is a *hyperplane* (through the origin). If $\mathbb{L}$ is a subspace, then $\mathbb{L}^{\perp} = \{x\in\mathbb{R}^n \mid x\cdot y = 0 \text{ for all } y\in\mathbb{L}\}$ is termed the *complementary subspace.* Subspaces $\mathbb{L}$ and $\mathbb{M}$ with the property that $x\cdot y = 0$ for all $y \in \mathbb{L}$ and $x \in \mathbb{M}$ are termed *orthogonal*, and denoted $\mathbb{L}\perp\mathbb{M}$. The *angle* θ between subspaces $\mathbb{L}$ and $\mathbb{M}$ is defined by

$$\cos\theta = \text{Min } \{x\cdot y \mid y \in \mathbb{L},\ \|y\|_2 = 1,\ x \in \mathbb{M},\ \|x\|_2 = 1\}.$$

Then, the angle between orthogonal subspaces is $\pi/2$, and the angle between subspaces that have a nonzero point in common is zero. A subspace that is translated by adding a nonzero vector c to all points in the subspace is termed an *affine subspace*.

2.3. The concept of a finite-dimensional space can be generalized. For example, consider, for $1 \le p < +\infty$, the family $\mathbb{L}_p(\mathbb{R}^n)$ of real-valued functions f on $\mathbb{R}^n$ such that the integral $\|f\|_p = \int_{\mathbb{R}^n} |f(x)|^p dx$ is well-defined and finite. This is a linear space with norm $\|f\|_p$ since linear combinations of functions that satisfy this property also satisfy (using convexity of the norm function) this property. One can think of the function f as a vector in $\mathbb{L}_p(\mathbb{R}^n)$, and f(x) for a particular value of x as a component of this vector. Many, but not all, of the properties of finite-dimensional space extend to infinite dimensions. In basic econometrics, we will not need the infinite-dimensional generalization. It appears in more advanced econometrics, in stochastic processes in time series, and in nonlinear and nonparametric problems.

3. Linear Transformations and Matrices

3.1. A mapping *A* from one linear space (its *domain)* into another (its *range*) is a *linear transformation* if it satisfies *A*(x+z) = *A*(x) + *A*(z) for any x and z in the domain. When the domain and range are finite-dimensional linear spaces, a linear transformation can be represented as a *matrix*. Specifically, a linear transformation *A* from $\mathbb{R}^n$ into $\mathbb{R}^m$ can be represented by a m×n array A with elements a_{ij} for 1 ≤ i ≤ m and 1 ≤ i ≤ n, with y = *A*(x) having components $y_i = \sum_{j=1}^{n} a_{ij}x_j$ for 1 ≤ *i* ≤ m. In matrix notation, this is written y = Ax. A matrix A is *real* if all its elements are real numbers, *complex* if some of its elements are complex numbers. Throughout these notes, matrices are assumed to be real unless explicitly assumed otherwise. The set $\mathbb{N} = \{x\in\mathbb{R}^n \mid Ax = 0\}$ is termed the *null space* of the transformation A. The set $\mathbb{N}^{\perp}$ of all linear combinations of the column vectors of A is termed the *column space* of A; it is the complementary subspace to $\mathbb{N}$.

If A denotes a m×n matrix, then A′ denotes its n×m transpose (rows become columns and vice versa). The *identity matrix* of dimension n is n×n with one's down the diagonal, zero's elsewhere, and is denoted I_n, or I if the dimension is clear from the context. A *permutation matrix* is obtained by permuting the columns of an identity matrix. If A is a m×n matrix and B is a n×p matrix, then the *matrix product* C = AB is of dimension m×p with elements $c_{ik} \equiv \sum_{j=1}^{n} a_{ij}b_{jk}$ for 1 ≤ i ≤ m and 1 ≤ k ≤ p. For the matrix product to be defined, the number of columns in A must equal the number of rows in B (i.e., the matrices must be *commensurate*). A matrix A is *square* if it has the same number of rows and columns. A square matrix A is *symmetric* if A = A′, *diagonal* if all off-diagonal elements are zero, *upper (lower) triangular* if all its elements below (above) the diagonal are zero, and *idempotent* if it is symmetric and $A^2 = A$. A matrix A is *column orthonormal* if $A'A = I$; simply *orthonormal* if it is both square and column orthonormal.

Each column of a m×n matrix A is a vector in $\mathbb{R}^m$. The *rank* of A, denoted $\rho(A)$, is the largest number of columns that are *linearly independent*. From the definition, A is of rank n if and only if $x = 0$ is the only solution to $Ax = 0$. A n×n matrix A of rank n is termed *nonsingular*; this matrix has an *inverse* matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_m$ if and only if A is *nonsingular*.

3.2. The following tables define useful matrix and vector operations.

Table 1. Basic Operations

	Name	Notation	Definition
1.	Matrix Product	$C = AB$	For m×n A and n×p B, $c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}$
2.	Scalar Multiplication	$C = bA$	For a scalar b, $c_{ij} = ba_{ij}$
3.	Matrix Sum	$C = A+B$	For A and B m×n, $c_{ij} = a_{ij} + b_{ij}$
4.	Matrix Inverse	$C = A^{-1}$	For A n×n nonsingular, $AA^{-1} = I_m$
5.	Trace	$c = tr(A)$	For n×n A, $c = \sum_{i=1}^{n} a_{ii}$

Table 2. Operations on Elements

	Name	Notation	Definition
1.	Element Product	C = A.*B	For A, B m×n, $c_{ij} = a_{ij} \cdot b_{ij}$
2.	Element Division	C = A.÷B	For A, B m×n, $c_{ij} = a_{ij}/b_{ij}$
3.	Logical Condition	C = A.≤B	For A, B m×n, $c_{ij} = \mathbf{1}(a_{ij} \le b_{ij})$ [Note 1]
4.	Row Minimum	C = vmin(A)	For m×n A, $c_{i1} = \min_j a_{ij}$ [Note 2]
5.	Column Minimum	C = min(A)	For m×n A, $c_{ij} = \min_k a_{kj}$ [Note 3]
6.	Cumulative Sum	C = cumsum(A)	For m×n A, $c_{ij} = \sum_{k=1}^{i} a_{kj}$

Table 3. Shaping Operations

	Name	Notation	Definition
1.	Kronecker Product	C = A⊗B	Note 4
2.	Direct Sum	C = A⊕B	Note 5
3.	diag	C = diag(x)	C a diagonal matrix with vector x down the diagonal
4.	vec	c = vec(A)	vector c contains rows of A, by row
5.	vech	c = vech(A)	vector c contains upper triangle of A, row by row, stacked
6.	vecd	c = vecd(A)	vector c contains diagonal of A
7.	horizontal contatination	C = {A,B}	Partitioned matrix C = [A B]
8.	vertical concatination	C = {A;B}	Partitioned matrix C = [A′ B′]′
9.	reshape	C = rsh(A,k)	Note 6

NOTES TO TABLES 2 AND 3:

1. **1**(P) is one of P is true, zero otherwise. The condition is also defined for the logical operations "<", ">", "≥", "=", and "≠".
2. C is a m×1 matrix. The operation is also defined for "max".
3. C is a m×n matrix, with all rows the same. The operation is also defined for "max".
4. Also termed the *direct product*, the Kronecker product is defined for a m×n matrix A and a p×q matrix B as the (mp)×(nq) array

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

5. The *direct sum* is defined for a $m \times n$ matrix A and a $p \times q$ matrix B by the $(m+p) \times (n+q)$ partitioned array $A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$.

6. If A is $m \times n$, then k must be a divisor of $m \cdot n$. The operation takes the elements of A row by row, and rewrites the successive elements as rows of a matrix C that has k rows and $m \cdot n/k$ columns.

In addition to the operations in the tables above, there are statistical operations that can be performed on a matrix when its columns are vectors of observations on various variables. Discussion of these operations is postponed until later. Most of the operations in Tables 1-3 are available as part of the matrix programming languages in econometrics computer packages such as SST, TSP, GAUSS, or MATLAB. The notation in these tables is close to the notation for the corresponding matrix commands in SST and GAUSS.

3.3. The *determinant* of a $n \times n$ matrix A is denoted $|A|$ or det(A), and has the geometric interpretation as the volume of the parallelpiped formed by the column vectors of A. The matrix A is *nonsingular* if and only if $\det(A) \neq 0$. A *minor* of a matrix A (of order r) is the determinant of a submatrix formed by striking out n-r rows and columns. A *principal minor* is formed by striking out symmetric rows and columns of A. A *leading principal minor* (of order r) is formed by striking out the last n-r rows and columns. The *minor* of an element a_{ij} of A is the determinant of the submatrix A^{ij} formed by striking out row i and column j of A. Determinants satisfy the recursion relation

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

with the first equality holding for any j and the second holding for any i. This formula can be used as a recursive definition of determinants, starting from the result that the determinant of a scalar is the scalar. A useful related formula is

$$\sum_{i=1}^{n} (-1)^{i+j} a_{ik} \det(A^{ij})/\det(A) = \delta_{kj},$$

where δ_{kj} is one if k = j and zero otherwise.

3.4. We list without proof a number of useful elementary properties of matrices:

(1) $(A')' = A$.

(2) If A^{-1} exists, then $(A^{-1})^{-1} = A$.

(3) If A^{-1} exists, then $(A')^{-1} = (A^{-1})'$.

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

(5) If A,B are square, nonsingular, and commensurate, then $(AB)^{-1} = B^{-1}A^{-1}$.

(6) If A is m×n, then $\text{Min}\,\{m,n\} \geq \rho(A) = \rho(A') = \rho(A'A) = \rho(AA')$.

(7) If A and B are commensurate, then $\rho(AB) \leq \min(\rho(A),\rho(B))$.

(8) $\rho(A+B) \leq \rho(A) + \rho(B)$.

(9) If A is n×n, then $\det(A) \neq 0$ if and only if $\rho(A) = n$.

(10) If B and C are nonsingular and commensurate with A, then $\rho(BAC) = \rho(A)$.

(11) If A, B are n×n, then $\rho(AB) \geq \rho(A) + \rho(B) - n$.

(12) $\det(AB) = \det(A)\cdot\det(B)$.

(13) If c is a scalar and A is n×n, then $\det(cA) = c^n\det(A)$

(14) The determinant of a matrix is unchanged if a scalar times one column (row) is added to another column (row).

(15) The determinant of a diagonal or upper triangular matrix is the product of the diagonal elements.

(16) $\det(A^{-1}) = 1/\det(A)$.

(17) If A is $n \times n$ and $B = A^{-1}$, then $b_{ij} = (-1)^{i+j}\det(A^{ij})/\det(A)$.

(18) The determinant of an orthonormal matrix is +1 or -1.

(19) If A is $m \times n$ and B is $n \times m$, then $tr(AB) = tr(BA)$.

(20) $tr(I_n) = n$.

(21) $tr(A+B) = tr(A) + tr(B)$.

(22) A permutation matrix P is orthonormal; hence, $P' = P^{-1}$.

(23) The inverse of a (upper) triangular matrix is (upper) triangular, and the inverse of a diagonal matrix D is diagonal, with the reciprocals of the diagonal elements of D down the diagonal of D^{-1}.

(24) The product of orthonormal matrices is orthonormal, and the product of permutation matrices is a permutation matrix.

4. Eigenvalues and Eigenvectors

An eigenvalue of a $n \times n$ matrix A is a scalar λ such that $Ax = \lambda x$ for some vector $x \neq 0$. The vector x is called a (right) eigenvector. The condition $(A-\lambda I)x = 0$ associated with an eigenvalue implies $A-\lambda I$ simgular, and hence $\det(A-\lambda I) = 0$. The determinental equation defines a polynomial in λ of order n; the n roots of this polynomial are the eigenvalues. For each eigenvalue λ, the condition that $A-\lambda I$ is less than rank n implies the existence of one or more linearly independent eigenvectors; the number equals the multiplicity of the root λ. The following basic properties of eigenvalues and eigenvectors are stated without proof:

(1) If A is real and nonsymmetric, then its eigenvalues and eigenvectors in general are complex. If A is real and symmetric, then its eigenvalues and eigenvectors are real.

(2) The number of nonzero eigenvalues of A equals its rank $\rho(A)$.

(3) If λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k, and $1/\lambda$ is an eigenvalue of A^{-1} (if the inverse exists).

(4) If A is real and symmetric, and Λ is a diagonal matrix with the roots of the polynomial $\det(A-\lambda I)$ along the diagonal, then there exists an orthonormal matrix C (whose columns are eigenvectors of A) such that $C'C = I$ and $AC = C\Lambda$, and hence $C'AC = \Lambda$ and $C\Lambda C' = A$. The transformation C is said to *diagonalize* A.

(5) If A is real and nonsymmetric, there exists a nonsingular complex matrix Q and a upper triangular complex matrix T with the eigenvalues of A on its diagonal such that $Q^{-1}AQ = T$.

(6) A real and symmetric implies tr(A) equals the sum of the eigenvalues of A. [Since $A = C\Lambda C'$, $\mathrm{tr}(A) = \mathrm{tr}(C\Lambda C') = \mathrm{tr}(C'C\Lambda) = \mathrm{tr}(\Lambda)$ by 2.3.19.]

(7) If $A_1,\ldots,A_p$ are real and symmetric, then there exists C orthonormal such that $C'A_1C, C'A_2C,\ldots,C'A_pC$ are all diagonal if and only if $A_iA_j = A_jA_i$ for all i and j.

Results (4) and (5) combined with the result 2.3.12 that the determinant of a matrix product is the product of the determinants of the matrices, implies that the determinant of a matrix is the product of its eigenvalues. The transformations in (4) and (5) are called *similarity transformations*, and can be interpreted as representations of the transformation A when the basis of the domain is transformed by C (or Q) and the basis of the range is transformed by C^{-1} (or Q^{-1}). These transformations are used extensively in econometric theory.

5. Partitioned Matrices

It is sometimes useful to *partition* a matrix into submatrices,

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

where A is $m \times n$, A_{11} is $m_1 \times n_1$, A_{12} is $m_1 \times n_2$, A_{21} is $m_2 \times n_1$, A_{22} is $m_2 \times n_2$, and $m_1+m_2 = m$ and $n_1+n_2 = n$. Matrix products can be written for partitioned matrices, applying the usual algorithm to the partition blocks, provided the blocks are commensurate. For example, if B is $n \times p$ and is partitioned $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ where B_1 is $n_1 \times p$ and B_2 is $n_2 \times p$, one has

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_{11}B_1+A_{12}B_2 \\ A_{21}B_1+A_{22}B_2 \end{bmatrix}.$$

Partitioned matrices have the following elementary properties:

(1) A square and A_{11} square and nonsingular implies

$\det(A) = \det(A_{11}) \cdot \det(A_{22}-A_{21}A_{11}^{-1}A_{12})$.

(2) A and A_{11} square and nonsingular implies

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

with $C = A_{22}-A_{21}A_{11}^{-1}A_{12}$.

6. Quadratic Forms

The scalar function $Q(x,A) = x'Ax$, where A is a $n\times n$ matrix and x is a $n\times 1$ vector, is termed a *quadratic form*; we call x the *wings* and A the *center* of the quadratic form. The value of a quadratic form is unchanged if A is replaced by its *symmetrized* version $(A+A')/2$. Therefore, A will be assumed symmetric for the discussion of quadratic forms.

A quadratic form Q(x,A) may fall into one of the classes in the table below:

Class	Defining Condition
Positive Definite	$x\neq 0 \Rightarrow Q(x,A) > 0$
Positive Semidefinite	$x\neq 0 \Rightarrow Q(x,A) \geq 0$
Negative Definite	$x\neq 0 \Rightarrow Q(x,A) < 0$
Negative Semidefinite	$x\neq 0 \Rightarrow Q(x,A) \leq 0$

The basic properties of quadratic forms are listed below:

(1) If B is $m\times n$ and is of rank $\rho(B) = r$, then $B'B$ and BB' are both positive semidefinite; and if $r = m \leq n$, then $B'B$ is positive definite.

(2) If A is symmetric and positive semidefinite (positive definite), then the eigenvalues of A are nonnegative (positive). Similarly, if A is symmetric and negative semidefinite (negative definite), then the eigenvalues of A are nonpositive (negative).

(3) Every symmetric positive semidefinite matrix A has a symmetric positive semidefinite square root $A^{1/2}$ [By 2.4.4, $C'AC = D$ for some C orthonormal and D a diagonal matrix with the nonnegative eigenvalues down the diagonal. Then, $A = CDC'$ and $A^{1/2} = CD^{1/2}C'$ with $D^{1/2}$ a diagonal matrix of the positive square roots of the diagonal of D.]

(4) If A is positive definite, then A^{-1} is positive definite.

(5) B positive definite and A - B positive semidefinite imply B^{-1} - A^{-1} positive semidefinite.

(6) The following conditions are equivalent:

(i) A is positive definite
(ii) The principal minors of A are positive
(iii) The leading principal minors of A are positive.

7. The LDU and Cholesky Factorizations of a Matrix

A n×n matrix A has a LDU factorization if it can be written A = LDU, where D is a diagonal matrix, L is a lower triangular matrix, and U is a upper triangular matrix. This factorization is useful for computation of inverses, as triangular matrices are easily inverted by recursion.

Theorem. For each n×n matrix A, A = PLDUQ′, where P and Q are permutation matrices, L is a lower triangular matrix and U is a upper triangular matrix, each with ones on the diagonal, and D is a diagonal matrix. If the leading principal minors of A are all non-zero, then P and Q can be taken to be identity matrices.

Proof: First assume that the leading principal minors of A are all nonzero. We give a recursive construction of the required L and U. Suppose the result has been established for matrices up to order n-1. Then, write the required decomposition A = LDU for a n×n matrix in partitioned form

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} \cdot \begin{bmatrix} U_{11} & U_{12} \\ 0 & 1 \end{bmatrix},$$

where A_{11}, L_{11}, D_{11}, and U_{11} are (n-1)×(n-1), L_{21} is 1×(n-1), U_{12} is (n-1)×1, and A_{22} and D_{22} are 1×1. Assume that L_{11}, D_{11}, and U_{11} have been defined so that A_{11} =

$L_{11}D_{11}U_{11}$, and that L_{11}^{-1} and U_{11}^{-1} also exist and have been computed. Let $S = L^{-1}$ and $T = U^{-1}$, and partition S and T commensurately with L and U. Then, the remaining elements satisfy the equations

$$A_{21} = L_{21}D_{11}U_{11} \quad \Rightarrow \quad L_{21} = A_{21}U_{11}^{-1}D_{11}^{-1} \equiv A_{21}T_{11}D_{11}^{-1}$$
$$A_{12} = L_{11}D_{11}U_{12} \quad \Rightarrow \quad U_{12} = D_{11}^{-1}L_{11}^{-1}A_{12} \equiv D_{11}^{-1}S_{11}A_{12}$$
$$A_{22} = L_{21}D_{11}U_{12} + D_{22} \ \Rightarrow \ D_{22} = A_{22} - A_{21}T_{11}D_{11}^{-1}S_{11}A_{12} = A_{22} - A_{21}A_{11}^{-1}A_{12}$$
$$S_{21} = -L_{21}S_{11}$$
$$T_{12} = -\ T_{11}U_{12}$$

where $\det(A) = \det(A_{11})\cdot\det(A_{22} - A_{21}A_{11}^{-1}A_{12}) \neq 0$ implies $D_{22} \neq 0$. Since the decomposition is trivial for n = 1, this recursion establishes the result, and furthermore gives the triangular matrices S and T from the same recursion that can be multiplied to give $A^{-1} = TD^{-1}S$.

Now assume that A is of rank r < n, and that the first r columns of A are linearly independent, with non-zero leading principal minors up to order r. Partition

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & I \end{bmatrix} \cdot \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & I \end{bmatrix},$$

where A_{11} is r×r and the remaining blocks are commensurate. Then, $U_{12} = D_{11}^{-1}S_{11}A_{12}$ and $L_{21} = A_{21}T_{11}D_{11}^{-1}$, and one must satisfy $A_{22} = L_{21}D_{11}U_{12} = A_{21}T_{11}D_{11}^{-1}S_{11}A_{12} = A_{21}A_{11}^{-1}A_{12}$. But the rank condition requires that the last n-r columns can be written as linear combinations of the first r columns, or $\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}$C for some r×(n-r) matrix C. But $A_{12} = A_{11}C$ implies $C = A_{11}^{-1}A_{12}$ and hence $A_{22} = A_{21}C = A_{21}A_{11}^{-1}A_{12}$ as required.

Finally, consider any real matrix A of rank r. By column permutations, the first r columns can be made linearly independent. Then, by row permutations, the first r rows of these r columns can be made linearly independent. Repeat this process recursively on the remaining northwest principal submatrices to obtain products of permutation matrices that give nonzero leading principal minors up to order r. This defines P and Q, and completes the proof of the theorem. ▫

Corollary 1. If A is symmetric, then L = U′.

Corollary 2. (LU Factorization) If A has nonzero leading principal minors, then A can be written $A = L\tilde{U}$, where $\tilde{U} = DU$ is upper triangular with a diagonal coinciding with that of D.

Corollary 3. (Cholesky Factorization) If A is symmetric and positive definite, then A can be written $A = \tilde{U}'\tilde{U}$, where $\tilde{U}$ is upper triangular with a positive diagonal.

Corollary 4. A symmetric positive semidefinite implies $A = P\tilde{U}'\tilde{U}P'$, with $\tilde{U}$ upper triangular with a nonnegative diagonal, P a permutation matrix.

Corollary 5. If A is m×n with m ≥ n, then there exists a factorization A = PLDUQ′, with D n×n diagonal, P a m×m permutation matrix, Q a n×n permutation matrix, U a n×n upper triangular matrix with ones on the diagonal, and L a m×n lower triangular matrix with ones on the diagonal (i.e., L has the form $L = \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix}$ with L_{11}

$n \times n$ and lower triangular with ones on the diagonal, and L_{21} $(m-n) \times n$. Further, if $\rho(A) = n$, then $(A'A)^{-1}A' = QU^{-1}D^{-1}(L'L)^{-1}L'P'$.

To show Corollary 3, note that a positive definite matrix has positive leading principal minors, and note from the proof of the theorem that this implies that the diagonal of D is positive. Take $\tilde{U} = D^{1/2}U$, where $D^{1/2}$ is the positive square root. The same construction applied to the LDU factorization of A after permutation gives Corollary 4.

To show Corollary 5, note first that the rows of A can be permuted so that the first n rows are of maximum rank $\rho(A)$. Suppose $A = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}$ is of this form, and apply the theorem to obtain $A_{11} = P_{11}L_{11}DUQ'$. The rank condition implies that $A_{21} = FA_{11}$ for some $(m-n) \times n$ array F. Then, $A_{21} = L_{21}DUQ'$, with $L_{21} = FP_{11}L_{11}$, so that

$$A = \begin{bmatrix} P_{11} & 0 \\ 0 & I_{m-n} \end{bmatrix} \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix} DUQ'.$$

To complete the proof, apply a left permutation if necessary to undo the initial row permutation of A. An implication of the last result in Corollary 5 is that if the system of equations $Ax = y$ with A $m \times n$ of rank n has a solution, then then solution is given by $x = (A'A)^{-1}A'y = QU^{-1}D^{-1}(L'L)^{-1}L'P'y$.

The recursion in the proof of the theorem is called Crout's algorithm, and is the method for matrix inversion used in many computer programs. It is unnecessary to do the permutations in advance of the factorizations; they can also be carried out recursively, bringing in rows (in what is termed a *pivot*) to make the successive elements of D as large in magnitude as possible. This pivot step is important for numerical accuracy.

8. The Singular Value Decomposition of a Matrix

A factorization that is useful as a tool for finding the eigenvalues and eigenvectors of a symmetric matrix, and for calculation of inverses of moment matrices of data with high multicollinearity, is the *singular value decomposition* (SVD): Every real $m \times n$ matrix A of rank r can be decomposed into a product $A = UDV'$, where D is a $r \times r$ diagonal matrix with positive nonincreasing elements down the diagonal, U is $m \times r$, V is $n \times r$, and U and V are column-orthonormal; i.e., $U'U = I_r = V'V$.

To prove that the SVD is possible, note first that the $m \times m$ matrix AA' is symmetric and positive semidefinite. Then, there exists a $m \times m$ orthonormal matrix W, partitioned $W = [W_1\ W_2]$ with W_1 of dimension $m \times r$, such that $W_1'(AA')W_1 = \Lambda$ is diagonal with positive, non-increasing diagonal elements, and $W_2'(AA')W_2 = 0$, implying $A'W_2 = 0$. Define D from Λ by replacing the diagonal elements of Λ by their positive square roots. Note that $W'W = I = WW' \equiv W_1W_1' + W_2W_2'$. Define $U = W_1$ and $V' = D^{-1}U'A$. Then, $U'U = I_r$ and $V'V = D^{-1}U'AA'UD^{-1} = D^{-1}\Lambda D^{-1} = I_r$. Further, $A = (I_m - W_2W_2')A = UU'A = UDV'$. This establishes the decomposition.

If A is symmetric, then U is the array of eigenvectors of A corresponding to the non-zero roots, so that $A'U = UD_1$, with D_1 the $r \times r$ diagonal matrix with the non-zero eigenvalues in decending magnitude down the diagonal. In this case, $V = A'UD^{-1} = UD_1D^{-1}$. Since the elements of D_1 and D are identical except possibly for sign, the columns of U and V are either equal (for positive roots) or reversed in sign (for negative roots). Then, the SVD of a square symmetric nonsingular matrix provides the pieces necessary to write down its eigenvalues and eigenvectors. For a positive definite matrix, the connection is direct.

When the m×n matrix A is of rank n, so that A′A is symmetric and positive definite, the SVD provides a method of calculating $(A'A)^{-1}$ that is particularly numerically accurate: Substituting the form A = UDV′, one obtains $(A'A)^{-1} = VD^{-2}V'$. One also obtains convenient forms for a square root of A′A and its inverse, $(A'A)^{1/2} = VDV'$ and $(A'A)^{-1/2} = VD^{-1}V'$.

The numerical accuracy of the SVD is most advantageous when m is large and some of the rows of A are nearly linearly dependent. Then, roundoff errors in the matrix product A′A can lead to quite inaccurate results when a matrix inverse of A′A is computed directly. The SVD extracts the required information from A before the roundoff errors in A′A are introduced.

Computer programs for the Singular Value Decomposition can be found in Press *et al*, Numerical Recipes, Cambridge University Press, 1986.

9. Projections and Idempotent Matrices

Let x be a vector in $\mathbb{R}^n$, expressed as a n×1 array, and let $\mathbb{L}$ be a subspace of $\mathbb{R}^n$. The *projection* of x on the subspace $\mathbb{L}$ is the point $\tilde{y} = \underset{z\in\mathbb{L}}{\mathrm{argmin}}\ \|z-x\|_2$ that minimizes the Euclidean distance between x and $\mathbb{L}$. Suppose $\mathbb{L}$ is of dimension k and $\{b_1,...,b_k\}$ is a basis for $\mathbb{L}$, and let B be the n×k matrix whose columns are these basis vectors. Then $\mathbb{L} = \{Bw \mid w \in \mathbb{R}^k\}$. The projection of x on $\mathbb{L}$ can then be calculated by solving $\min_w (x-Bw)'(x-Bw)$. One can show by writing this out and using calculus that the minimum occurs at $\tilde{w} = (B'B)^{-1}B'x$ and that the point in $\mathbb{L}$ closest to x is $\tilde{y} = B\tilde{w} \equiv B(B'B)^{-1}B'x$. To verify that $\tilde{y}$ achives the minimum, note that $(x-\tilde{y})'B = x'[I-B(B'B)^{-1}B']B = 0$; i.e., $x-\tilde{y}$ is orthogonal to $\mathbb{L}$. Let $c = w - \tilde{w}$, and rewrite the criterion

$$(x-Bw)'(x-Bw) = (x-B\tilde{w}-Bc)'(x-B\tilde{w}-Bc)$$

$$= (x-B\tilde{w})'(x-B\tilde{w}) + c'B'Bc - 2(x-B\tilde{w})'Bc$$

$$= (x-B\tilde{w})'(x-B\tilde{w}) + c'B'Bc \geq (x-B\tilde{w})'(x-B\tilde{w}).$$

The linear transformation $A = B(B'B)^{-1}B'$ is termed a *projection matrix*, and is sometimes written P_B or $P_{\mathbb{L}}$ to emphasize that it is a transformation that projects vectors in $\mathbb{R}^n$ onto the subspace $\mathbb{L}$ or $\mathbb{L}_B$ (i.e., the subspace spanned by B). It has the properties that it is symmetric with $A^2 = A$, so that it is *idempotent*. Geometrically, this says that once a vector is projected onto a subspace, than repeated applications of the same projection leave it unchanged. Conversely, every idempotent matrix A can be interpreted as a projection matrix onto some subspace. An important and useful property of projections is that they depend only on the subspace, and not on the particular choice of basis for this subspace. Thus, if B and C are matrices that span the common k-dimensional subspace $\mathbb{L}$, then $P_B = P_C = P_{\mathbb{L}}$. Further, it is irrelevant whether the columns of B or C are all linearly independent, as one can simply discard linearly dependent columns before forming the projection matrix $A = B(B'B)^{-1}B'$.

Some of the properties of an $n \times n$ *idempotent* matrix A are listed below:

(1) The eigenvalues of A are either zero or one.

(2) The rank of A is the dimension of the linear space into which the projection is made, and $\rho(A) = tr(A)$.

(3) The matrices I, 0, and I-A are idempotent.

(4) If B is an orthonormal matrix, then B′AB is idempotent.

(5) If $\rho(A) = r$, then there exists a $n \times r$ matrix B of rank r such that $A = B(B'B)^{-1}B'$, and thus $A = P_B$.

(6) A, B idempotent implies AB = 0 if and only if A+B idempotent.

(7) A, B idempotent and AB = BA implies AB idempotent.

(8) A, B idempotent implies A-B idempotent if and only if BA = B.

Suppose X is n×r of rank r and Z is n×s of rank s, and consider the subspaces $\mathbb{L}_X$ and $\mathbb{L}_Z$ and the projections $A = P_X = X(X'X)^{-1}X'$ and $B = P_Z = Z(Z'Z)^{-1}Z'$. If X and Z are column-orthogonal (i.e., X′Z = 0), then AB = 0 and (6) implies that A+B is the projection onto $\mathbb{L}_{[X\ Z]}$. If X contains all the columns of Z, so that $\mathbb{L}_Z \subseteq \mathbb{L}_X$, then a projection in one stage onto $\mathbb{L}_Z$ is equivalent to a projection onto $\mathbb{L}_X$ followed in a second stage by a further projection onto $\mathbb{L}_Z$, so that B = AB, and a projection onto $\mathbb{L}_Z$ is left invariant by a further projection onto $\mathbb{L}_X$, so that B = BA. From (8), this implies that A - B is idempotent; this is the projection onto the subspace of $\mathbb{L}_X$ that is orthogonal to $\mathbb{L}_Z$; i.e., the subspace $\mathbb{L}_X \cap \mathbb{L}_Z^{\perp}$. Every vector $w \in \mathbb{R}^n$ has a unique decomposition w = x + z + y, where $x = Bw \in \mathbb{L}_X$, $z = (A-B)w = A(I-B)w \in \mathbb{L}_X \cap \mathbb{L}_Z^{\perp}$, and $y \in (I-A)w \in \mathbb{L}_X^{\perp}$. The projections B and I-A are orthogonal, implying that I-A+B is also a projection, onto $\mathbb{L}_X^{\perp} \cup \mathbb{L}_Z$. A direct matrix manipulation gives an alternative demonstration of (8). Suppose X = [Z W] is a partitioned matrix where Z is n×s, W is n×t, and X is of rank r = n+t < n. Define $A = X(X'X)^{-1}X'$ and $B = Z(Z'Z)^{-1}Z'$. Then

$$AB = [Z\ W]\begin{bmatrix} Z'Z & Z'W \\ W'Z & W'W \end{bmatrix}^{-1}\begin{bmatrix} Z' \\ W' \end{bmatrix} Z(Z'Z)^{-1}Z' = [Z\ W]\begin{bmatrix} I \\ 0 \end{bmatrix}(Z'Z)^{-1}Z' = Z(Z'Z)^{-1}Z' = B\ ,$$

so that (8) implies A-B is idempotent, thus also I-A+B.

10. Generalized Inverses

A k×m matrix A^- is a *Moore-Penrose generalized inverse* of a m×k matrix A if it has three properties:

(i) $AA^-A = A$, (ii) $A^-AA^- = A^-$, (iii) AA^- and A^-A are symmetric.

There are other generalized inverse definitions that have some, but not all, of these properties. Results on generalized inverses, particularly for partitioned or bordered matrices, can be found in R.M. Pringle and A. Rayner <u>Generalized Inverse Matrices</u>, Griffin, 1971.

Let r denote the rank of A. Recall that A has a singular value decomposition A = UDV′, where D is a r×r diagonal matrix with a positive nonincreasing numbers down the diagonal, U is k×r and column-orthonormal, and V is m×r and column-orthonormal. The Moore-Penrose generalized inverse of A is then the matrix $A^- = VD^{-1}U'$. It is easy to check that this definition satisfies properties (i)-(iii) above, as well as the following properties:

(1) $A^- = A^{-1}$ if A is square and non-singular.

(2) AA^- and A^-A are idempotent, and are projection matrices, respectively, onto the subspace of $\mathbb{R}^m$ spanned by the columns of A and onto the subspace of $\mathbb{R}^k$ spanned by the rows of A.

(3) The system of equations Ax = y, with A m×k, has a solution if and only if $y = AA^-y$ (i.e., y is in the linear subspace spanned by the columns of A). If it has a solution, then the affine linear subspace of all solutions is

$$\{x\in\mathbb{R}^k \mid x = A^-y + [I - A^-A]z \text{ for all } z \in \mathbb{R}^k\}.$$

A solution is unique if and only if $A^-A = I$ (i.e., A is of rank k.)

(4) If A is square, symmetric, and positive semidefinite, then there exists Q positive definite, R idempotent, such that A = QRQ and $A^- = Q^{-1}RQ^{-1}$.

(5) If A is idempotent, then $A = A^-$.

(6) If A = BCB′ with B nonsingular, then $A^- = (B')^{-1} C^-B^{-1}$.

(7) If A is k×k and positive semidefinite with rank r, then there exists a k×r column-orthonormal matrix U such that $U'U = I_r$, $UAU' = D$ is a diagonal matrix with a positive diagonal, $A = UDU'$, and $A^- = UD^{-1}U'$.

(8) If A is square, symmetric, and positive semidefinite, then it has a symmetric square root $B = A^{1/2}$, and $A^- = B^-B^-$.

(9) $(A^-)^- = A = AA'(A^-)' = (A^-)'A'A$.

(10) $(A')^- = (A^-)'$

(11) $(A'A)^- = A^-(A^-)'$

(12) If $A = \sum A_i$ with $A_i'A_j = 0$ and $A_iA_j' = 0$ for $i \neq j$, then $A^- = \sum A_i^-$.

Most of these results are easy consequences of the singular value decomposition, or can be checked by verifying that conditions (i)-(iii) hold. To prove (4), let W = [U V] be an orthonormal matrix diagonalizing A. Then, U′AU = D, a diagonal matrix of positive eigenvalues, and AV = 0. Define $Q = W\begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix}W'$ and $R = W\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}W'$.

11. Kronecker Products

If A is a m×n matrix and B is a p×q matrix, then the *Kronecker (direct) product* of A and B is the (mp)×(nq) partitioned array

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

In general, A⊗B ≠ B⊗A. The Kronecker product has the following properties:

(1) For a scalar c, $(cA)\otimes B = A\otimes(cB) = c(A\otimes B)$.

(2) $(A\otimes B)\otimes C = A\otimes(B\otimes C)$.

(3) $(A \otimes B)' = (A') \otimes (B')$.

(4) $tr(A \otimes B) = (tr(A)) \cdot (tr(B))$.

(5) If the matrix products AC and BF are defined, then

$(A \otimes B)(C \otimes F) = (AC) \otimes (BF)$.

(6) If A and B are square and nonsingular, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

(7) If A and B are orthonormal, then $A \otimes B$ is orthonormal.

(8) If A and B are positive semidefinite, then $A \otimes B$ is positive semidefinite.

(9) If A is $k \times k$ and B is $n \times n$, then $\det(A \otimes B) = \det(A)^n \cdot \det(B)^k$.

(10) $\rho(A \otimes B) = \rho(A) \cdot \rho(B)$.

(11) $(A+B) \otimes C = A \otimes C + B \otimes C$.

12. Shaping Operations

The most common operations used to reshape vectors and matrices are (1) C = diag(x) which creates a diagonal matrix with the elements of the vector x down the diagonal; (2) c = vec(A) which creates a vector by stacking the columns of A; (3) c = vech(A) which creates a vector by stacking the proportions of the columns of A that are in the upper triangle of the matrix; and (4) c = vecd(A) which creates a vector containing the diagonal of A. There are a few rules that can be used to manipulate these operations:

(1) If x and y are commensurate vectors, diag(x+y) = diag(x) + diag(y).

(2) vec(A+B) = vec(A) + vec(B).

(3) If A is $m \times k$ and B is $k \times n$, then $vec(AB) = (I_n \otimes A)vec(B) = (B' \otimes I_m)vec(A)$.

(4) If A is $m \times k$, B is $k \times n$, C is $n \times p$, then $vec(ABC) = (I_p \otimes (AB))vec(C) = (C' \otimes A)vec(B) = ((C'B') \otimes I_m)vec(A)$.

(5) If A is $n \times n$, then vech(A) is of length $n(n+1)/2$.

(6) vecd(diag(x)) = x and diag(vecd(A)) = A.

13. Vector and Matrix Derivatives

The derivatives of functions with respect to the elements of vectors or matrices can sometimes be expressed in a convenient matrix form. First, a scalar function of a $n \times 1$ vector of variables, f(x), has partial derivatives that are usually written as the arrays

$$\partial f/\partial x = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ | \\ \partial f/\partial x_n \end{bmatrix}, \quad \partial f^2/\partial x \partial x' = \begin{bmatrix} \partial f^2/\partial x_1{}^2 & \partial f^2/\partial x_1 \partial x_2 & \ldots & \partial f^2/\partial x_1 \partial x_n \\ \partial f^2/\partial x_2 \partial x_1 & \partial f^2/\partial x_2{}^2 & \ldots & \partial f^2/\partial x_2 \partial x_n \\ | & | & & | \\ \partial f^2/\partial x_n \partial x_1 & \partial f^2/\partial x_n \partial x_2 & \ldots & \partial f^2/\partial x_n{}^2 \end{bmatrix}.$$

Other common notation is $f_x(x)$ or $\nabla_x f(x)$ for the vector of first derivatives, and $f_{xx}(x)$ or $\nabla_x^2 f(x)$ for the matrix of second derivatives. Sometimes, the vector of first derivatives will be interpreted as a row vector rather than a column vector. Some examples of scalar functions of a vector are the linear function $f(x) = a'x$, which has $\nabla_x f = a$, and the quadratic function $f(x) = x'Ax$, which has $\nabla_x f = 2Ax$.

When f is a column vector of scalar functions, $f(x) = [f^1(x)\ f^2(x)\ \ldots\ f^k(x)]'$, then the array of first partial derivatives is called the *Jacobean matrix* and is written

$$J(x) = \begin{bmatrix} \partial f^1/\partial x_1 & \partial f^1/\partial x_2 & \ldots & \partial f^1/\partial x_n \\ \partial f^2/\partial x_1 & \partial f^2/\partial x_2 & \ldots & \partial f^2/\partial x_n \\ \\ \partial f^k/\partial x_1 & \partial f^k/\partial x_2 & \ldots & \partial f^k/\partial x_n \end{bmatrix}.$$

When calculating multivariate integrals of the form $\int_{\mathbb{A}} g(y)dy$, where $y \in \mathbb{R}^n$, $\mathbb{A} \subseteq \mathbb{R}^n$, and g is a scalar or vector function of y, one may want to make a nonlinear one-to-one transformation of variables y = f(x). In terms of the transformed variables, the integral becomes

$$\int_{\mathbb{A}} g(y)dy = \int_{f^{-1}(\mathbb{A})} g(f(x))\cdot|\det(\mathcal{J}(x))|dx \ ,$$

where $f^{-1}(\mathbb{A})$ is the set of x vectors that map onto $\mathbb{A}$, and the Jacobean matrix is square and nonsingular for well-behaved one-to-one transformations. The intuition for the presence of the determinant of the Jacobean in the transformed integral is that "dy" is the volumn of a small rectangle in y-space, and because determinants give the volumn of the parallelpiped formed by the columns of a linear transformation, $\det(\mathcal{J}(x))dx$ gives the volumn (with a plus or minus sign) of the image in x-space of the "dy" rectangle in y-space.

It is useful to define the derivative of a scalar function with respect to a matrix as an array of commensurate dimensions. Consider the bilinear form f(A) = $x'Ay$, where x is n×1, y is m×1, and A is n×m. By collecting the individual terms $\partial f/\partial A_{ij} = x_i y_j$, one obtains the result $\partial f/\partial A = xy'$. Another example for a n×n matrix A is f(A) = tr(A), which has $\partial f/\partial A = I_n$. There are a few other derivatives that are particularly useful for statistical applications. In these formulas, A is a square nonsingular matrix. We do not require that A be symmetric, and the derivatives do not impose symmetry. One will still get valid calculations involving derivatives when these expressions are evaluated at matrices that hapen to be symmetric. There are alternative, and somewhat more complicated, derivative formulas that hold when symmetry is imposed. For analysis, it is unnecessary to introduce this complication.

(1) If det(A) > 0, then $\partial \log(\det(A))/\partial A = A^{-1}$.

(2) If A is nonsingular, then $\partial(x'A^{-1}y)/\partial A = -A^{-1}xy'A^{-1}$.

(3) If A = TT′, with T square and nonsingular, then

$$\partial(x'A^{-1}y)/\partial T = -2A^{-1}xy'A^{-1}T.$$

We prove the formulas in order. Recall that $\det(A) = \sum_k (-1)^{i+k} a_{ik} \det(A^{ik})$, where A^{ik} is the minor of a_{ik}. Then, $\partial\det(A)/\partial A_{ij} = (-1)^{i+j}\det(A^{ij})$. From 2.3.17, the ij element of A^{-1} is $(-1)^{i+j}\det(A^{ij})/\det(A)$. This establishes (1). To show (2), first apply the chain rule to the identity $AA^{-1} \equiv I$ to get $\Delta_{ij}A^{-1} + A\cdot\partial A^{-1}/\partial A_{ij} \equiv 0$, where Δ_{ij} denotes a matrix with a one in row i and column j, zeros elsewhere. Then, $\partial x'A^{-1}y/\partial A_{ij} = -x'A^{-1}\Delta_{ij}A^{-1}y = (A^{-1}x)_i(A^{-1}y)_j$. This establishes (2). To show (3), first note that

$\partial A_{ij}/\partial T_{rs} = \delta_{ir}T_{js}+\delta_{jr}T_{is}$. Combine this with (2) to get

$$\partial x'A^{-1}y/\partial T_{rs} = \sum_i\sum_j (A^{-1}x)_i(A^{-1}y)_j(\delta_{ir}T_{js}+\delta_{jr}T_{is})$$

$$= \sum_j (A^{-1}x)_r(A^{-1}y)_jT_{js} + \sum_i (A^{-1}x)_i(A^{-1}y)_rT_{is}$$

$$= 2(A^{-1}xy'A^{-1}T)_{rs}.$$

14. Updating and Backdating Matrix Operations

Often in statistical applications, one needs to modify the calculation of a matrix inverse or other matrix operation to accomodate the addition of data, or deletion of data in bootstrap methods. It is convenient to have quick methods for these calculations. Some of the useful formulas are given below:

(1) If A is $n\times n$ and nonsingular, and A^{-1} has been calculated, and if B and C are arrays that are $n\times k$ of rank k, then

$$(A+BC')^{-1} = A^{-1} - A^{-1}B(I_k+C'A^{-1}B)^{-1}C'A^{-1},$$

provided $I_k+C'A^{-1}B$ is nonsingular. If $k = 1$, then no matrix inversion is required in the updating.

(2) If A is $m\times n$ with $m \ge n$ and $\rho(A) = n$, so that it has a LDU factori-zation $A = PLDUQ'$ with D $n\times n$ diagonal, P and Q permutation matrices, L lower triangular, U upper triangular, then the array $\begin{bmatrix} A \\ B \end{bmatrix}$, with B $k\times n$, has the LDU factorization $\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} P & 0 \\ 0 & I_k \end{bmatrix}\begin{bmatrix} L \\ C \end{bmatrix}DUQ'$, where $C = BQU^{-1}D^{-1}$.

(3) Suppose A is $m\times n$ of rank n, and $B = (A'A)^{-1}A'y$. Suppose $\tilde{A} = \begin{bmatrix} A \\ C \end{bmatrix}$ with C $k\times n$, $\tilde{y} = \begin{bmatrix} y \\ w \end{bmatrix}$ with w $k\times 1$, and $\tilde{B} = (\tilde{A}'\tilde{A})^{-1}\tilde{A}'\tilde{y}$. Then,

$$\tilde{B}-B = (A'A)^{-1}C'[I_k+C(A'A)^{-1}C']^{-1}(w-CB)$$
$$= (\tilde{A}'\tilde{A})^{-1}C'[I_k-C(\tilde{A}'\tilde{A})^{-1}C']^{-1}(w-C\tilde{B}).$$

One can verify (1) by multiplication. To show (2), use Corollary 5 of Theorem 5.1. To show (3), apply (1) to $\tilde{A}'\tilde{A} = A'A + C'C$, or to $A'A = \tilde{A}'\tilde{A} - C'C$, and use $\tilde{A}'\tilde{y} = Ay + Cw$.

NOTES AND COMMENTS

The basic results of linear algebra, including the results stated without proof in this summary, can be found in standard linear algebra texts, such as G. Hadley (1961) Linear Algebra, Addison-Wesley or F. Graybill (1983) Matrices with Applications in Statistics, Wadsworth. The organization of this summary is based on the admirable synopsis of matrix theory in the first chapter of F. Graybill (1961) An Introduction to Linear Statistical Models, McGraw-Hill. For computations involving matrices, W. Press *et al* (1986) Numerical Recipes, Cambridge Univ. Press, provides a good discussion of algorithms and accompanying computer code. For numerical issues in statistical computation, see R.Thisted (1988) Elements of Statistical Computing, Chapman and Hall.