CHAPTER 2. ANALYSIS AND LINEAR ALGEBRA IN A NUTSHELL

2.1. SOME ELEMENTS OF MATHEMATICAL ANALYSIS

2.1.1. 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 $a + \iota b$, where a and b are real numbers and ι is the square root of -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 $r = (a^2+b^2)^{1/2}$ 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, $(a+\iota b)\cdot(c+\iota d) = (ab-cd)+\iota(ad+bc)$.

2.1.2. For sets of objects **A** and **B**, the *union* $\mathbf{A}\cup\mathbf{B}$ is the set of objects in either; the *intersection* $\mathbf{A}\cap\mathbf{B}$ is the set of objects in both; and $\mathbf{A}\backslash\mathbf{B}$ is the set of objects in **A** that are not in **B**. The empty set is denoted $\boldsymbol{\varphi}$. Set inclusion is denoted $\mathbf{A} \subseteq \mathbf{B}$; we say **A** *is contained in* **B**. The complement of a set **A** (which may be relative to a set **B** that contains it) is denoted $\mathbf{A}^c$. A family of sets is *disjoint* if the intersection of each pair is empty. The symbol $a \in \mathbf{A}$ means that a is a member of **A**; and $a \notin \mathbf{A}$ means that a is not a member of **A**. The symbol $\exists$ means "there exists", the symbol $\forall$ means "for all", and the symbol $\ni$ means "such that". A proposition that A implies B is denoted "A $\Longrightarrow$ B", and a proposition that A and B are equivalent is denoted "A $\Longleftrightarrow$ B". The proposition that A implies B, but B does not imply A, is denoted "A $|\!\!\Longrightarrow$ B". The phrase "if and only if" is often abbreviated to "iff".

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

2.1.4. The *supremum* of **A**, denoted sup **A**, is the least upper bound on **A**. A typical application has a function $f:\mathbf{C} \to \mathbb{R}$ and $\mathbf{A} = f(\mathbf{C})$; then $\sup_{c\in\mathbf{C}}$ f© is used to denote sup **A**. If the supremum is achieved by an object $d \in \mathbf{C}$, so $f(d) = \sup_{c\in\mathbf{C}} f(c)$, then we write $f(d) = \max_{c\in\mathbf{C}} f(c)$. When there is a unique maximizing argument, write $d = \text{argmax}_{c\in\mathbf{C}} f(c)$. When there is a non-unique maximizing argument; we will assume that $\mathrm{argmax}_{c \in \mathbf{C}}$ f(c) is a *selection* of any one of the maximizing arguments. Analogous definitions hold for the infimum and minimum, denoted inf, min, and for argmin.

2.1.5. 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 $a_i \to a_o$. 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 $\mathrm{limsup}_{i \to \infty} a_i$ 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.

2.1.6. A real-valued function $\rho(a,b)$ defined for pairs of objects in a set **A** is a *distance function* if it is non-negative, gives a positive distance between all distinct points of **A**, has $\rho(a,b) = \rho(b,a)$, and satisfies the triangle inequality $\rho(a,b) \leq \rho(a,c) + \rho(c,b)$. A set **A** with a distance function ρ is termed a *metric space*. A real-valued function $\|a\|$ defined for objects in a set A is a *norm* if $\|a-b\|$ has the properties of a distance function. 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; then $\mathbb{R}$ is a metric space and a normed space. A (ε-)*neighborhood* of a point *a* in a metric space **A** is a set of the form $\{b \in \mathbf{A} | \rho(a,b) < \varepsilon\}$. A set $\mathbf{C} \subseteq \mathbf{A}$ is *open* if for each point in **C**, some neighborhood of this point is also contained in **C**. A set $\mathbf{C} \subseteq \mathbf{A}$ is *closed* if its complement is open. The *closure* of a set **C** is the intersection of all closed sets that contain **C**. The *interior* of **C** is the union of all open sets contained in **C**; it can be empty. A *covering* of a set **C** is a family of open sets whose union contains **C**. The set **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 **A** is *separable* if there exists a countable subset **B** such that every neighborhood contains a member of **B**. All of the metric spaces encountered in econometrics will be separable. A sequence a_i in a separable metric space **A** is *convergent* (to a point a_o) if the sequence is eventually contained in each neighborhood of *a*; we write $a_i \to a_o$ or $\lim_{i \to \infty} a_i = a_o$ to denote a convergent sequence. A set $\mathbf{C} \subseteq \mathbf{A}$ is compact if and only if every sequence in **C** has a convergent subsequence (which converges to a *cluster point* of the original sequence).

2.1.7. Consider separable metric spaces **A** and **B**, and a function $f{:}\mathbf{A} \to \mathbf{B}$. The function f is *continuous* on **A** if the inverse image of every open set is open. Another characterization of continuity is that for any sequence satisfying $a_i \to a_o$, one has $f(a_i) \to f(a_o)$; the function is said to be continuous on $\mathbf{C} \subseteq \mathbf{A}$ if this property holds for each $a_o \in \mathbf{C}$. Stated another way, f is continuous on **C** if for each $\varepsilon > 0$ and $a \in \mathbf{C}$, 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*). For real valued functions on separable metric spaces, the concepts of supremium and limsup defined earlier for sequences have a natural extension: $\sup_{a \in \mathbf{A}} f(a)$ denotes the least upper bound on the set $\{f(a) | a \in \mathbf{A}\}$, and $\limsup_{a \to b} f(a)$ denotes the limit as $\varepsilon \to 0$ of the suprema of $f(a)$ on ε-neighborhoods of b. Analogous definitions hold for inf and liminf. A real-valued function f is continuous at b if $\limsup_{a \to b} f(a) = \liminf_{a \to b} f(a)$. 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 maximization $\max\{f(a),g(a)\}$ and minimization $\min\{f(a),g(a)\}$. The function f is *uniformly continuous* on **C** if for each $\varepsilon > 0$, there exists $\delta > 0$ such that for all $a \in \mathbf{C}$ and $b \in \mathbf{A}$ 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> $a \in \mathbf{C}$. A function that is continuous on a compact set is uniformly continuous. The function f is *Lipschitz* on **C** if there exist $L > 0$ and $\delta > 0$ such that $|f(b) - f(a)| \le L \cdot \rho(a,b)$ for all $a \in \mathbf{C}$ and $b \in \mathbf{A}$ with b in a δ-neighborhood of a.

2.1.8. Consider a real-valued function f on 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 $\lim_{c \to 0} \varepsilon©$ = 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 for b in this neighborhood 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, and λ is a scalar between zero and one.

If $\lim_{i \to \infty} a_i = a_o$ and f is a continuous function at a_o, then $\lim_{i \to \infty} f(a_i) = f(a_o)$. One useful result for limits is *L'Hopital's rule*, which states that if f(1/n) and g(1/n) are functions that are continuously differentiable at zero with $f(0) = g(0) = 0$, so that $f(n)/g(n)$ approaches the indeterminate expression 0/0, one has $\lim_{n \to \infty} f(n)/g(n) = f'(0)/g'(0)$, provided the last ratio exists.

2.1.9. If a_i for i = 0,1,2,... is a sequence of real numbers, the partial sums $s_n = \sum_{i=0}^{n} a_i$ define a *series*. We say the sequence is *summable*, or that the series is *convergent*, if $\lim_{n \to \infty} s_n$ exists and is finite. An example is the *geometric series* $a_i = r^i$, which has $s_n = (1-r^{n+1})/(1-r)$ if $r \neq 1$. When $|r| < 1$, this series is convergent, with the limit $1/(1-r)$. When $r < -1$ or $r \ge 1$, the series *diverges*. In the borderline case $r = -1$, the series alternates between 0 and 1, so the limit does not exist. Applying the Cauchy criterion, a summable sequence has $\lim_{n \to \infty} a_n = 0$ and $\lim_{n \to \infty} \sum_{i=n}^{\infty} a_i = 0$. A sequence satisfies a more general form of summability, called *Cesaro summability*, if $\lim_{n \to \infty} n^{-1} \sum_{i=0}^{n} a_i$ exists. Summability implies Cesaro summability, but not vice versa. A useful result known as

Kronecker's lemma states that if a_i and b_i are positive series, b_i is monotonically increasing to $+\infty$, and $\sum_{i=0}^{n} a_i/b_i$ is bounded for all n, then $\lim_{n\to\infty} b_n^{-1} \sum_{i=0}^{n} a_i = 0$.

2.1.10. 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 $e^{a+\iota b} = e^a(\cos(b) + \iota\cdot\sin(b))$. 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 $|b| < 1/3$, $|\text{Log}(1+a+b) - a| < 4|b| + 3|a|^2$. Another useful result, obtained by applying L'Hopital's rule to the expression $\log(1+a_n/n)/(1/n)$, is that $\lim_{n\to\infty} (1+a_n/n)^n = \exp(a_0)$ when $\lim_{n\to\infty} a_n = a_0$ exists.

A few specific series appear occasionally in probability theory. The series $a_i = i^\alpha$ for i = 1,2,... is summable for $\alpha < -1$, and divergent otherwise, with $s_n = n(n+1)/2$ for $\alpha = 1$, $s_n = n(n+1)(2n+1)/6$ for $\alpha = 2$, and $s_n = n^2(n+1)^2/4$ for $\alpha = 3$. Differentiating the formula $s_n = (1-r^{n+1})/(1-r)$ for a convergent geometric series leads to the expressions $\sum_{i=1}^{\infty} i\cdot r^i = r/(1-r)^2$ and $\sum_{i=1}^{\infty} i^2\cdot r^i = r(1+r)/(1-r)^3$.

2.1.11. If a_i and b_i are real numbers and c_i are non-negative numbers for i = 1,2,..., then *Holder's Inequality* states that for $p > 0$, $q > 0$, and $1/p + 1/q = 1$, one has

$$\left|\sum_i c_i\cdot a_i\cdot b_i\right| \le \sum_i c_i\cdot|a_i\cdot b_i| \le \left(\sum_i c_i\cdot|a_i|^p\right)^{1/p}\left(\sum_i c_i\cdot|b_i|^q\right)^{1/q} .$$

When $p = q = 1/2$, this is called the *Cauchy-Schwartz inequality*. Obviously, the inequality is useful only if the sums on the right converge. The inequality also holds in the limiting case where sums are replaced by integrals, and $a(i)$, $b(i)$, and $c(i)$ are functions of a continuous index i.

2.2. VECTORS AND LINEAR SPACES

2.2.1. A finite-dimensional linear space is a set such that (a) 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* $\mathbf{x} = (x_1,x_2,...,x_n)$, given a *basis* for the space $\{\mathbf{b}_1,...,\mathbf{b}_n\}$, where $x_1,...,x_n$ are real numbers such that $\mathbf{x} = x_1\mathbf{b}_1 + ... + x_n\mathbf{b}_n$. The point **x** is called a *vector*, and $x_1,...,x_n$ are called its *components*. The notation $(\mathbf{x})_i$ 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 $\|\mathbf{x}\|_2 = (x_1^2+...+x_n^2)^{1/2}$. This norm can be used to define the distance between vectors, or neighborhoods of a vector. Other possible norms include $\|\mathbf{x}\|_1 = |x_1|+...+|x_n|$, $\|\mathbf{x}\|_\infty = \max\{|x_1|,...,|x_n|\}$, or for $1 \le p < +\infty$, $\|\mathbf{x}\|_p = \left[|x_1|^p+...+|x_n|^p\right]^{1/p}$. 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 $\|\mathbf{x}\|_2$ and associated topology is called *Euclidean* n-*space*.

The *vector product* of **x** and **y** in $\mathbb{R}^n$ is defined as $\mathbf{x}\cdot\mathbf{y} = x_1y_1+...+x_ny_n$. Other notations for vector products are $<\mathbf{x},\mathbf{y}>$ or (when **x** and **y** are interpreted as *row* vectors) $\mathbf{x}\mathbf{y}'$ or (when **x** and **y** are interpreted as column vectors) $\mathbf{x}'\mathbf{y}$.

2.2.2. 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 in $\mathbb{R}^3$ are the plane $\{(a,b,c)|b = 0\}$ and the line $\{(a,b,c)|a = b = 2\cdot c\}$. The linear subspace *spanned* by a set of vectors $\{x_1,...,x_J\}$ is the set of all linear combinations of these vectors, $\mathbf{L} = \{x_1\alpha_1+...+x_J\alpha_J|(\alpha_1,...,\alpha_J) \in \mathbb{R}^J\}$. The vectors $\{\mathbf{x}_1,...,\mathbf{x}_J\}$ 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 **L** is a subspace, then $\mathbf{L}^\perp = \{\mathbf{x}\in\mathbb{R}^n|\ \mathbf{x}\cdot\mathbf{y} = 0$ for all $\mathbf{y}\in\mathbf{L}\}$ is termed the *complementary subspace.* Subspaces **L** and **M** with the property that $\mathbf{x}\cdot\mathbf{y} = 0$ for all $\mathbf{y} \in \mathbf{L}$ and $\mathbf{x} \in \mathbf{M}$ are termed *orthogonal,* and denoted $\mathbf{L}\perp\mathbf{M}$. The *angle* θ between subspaces **L** and **M** is defined by $\cos\theta = \mathrm{Min}\ \{\mathbf{x}\cdot\mathbf{y}|\ \mathbf{y} \in L, \|\mathbf{y}\|_2 = 1, \mathbf{x} \in M, \|\mathbf{x}\|_2 = 1\}$. 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.*

2.2.3. The concept of a finite-dimensional space can be generalized. An example, for $1 \le p < +\infty$, is the family $\mathbf{L}_p(\mathbb{R}^n)$ of real-valued functions f on $\mathbb{R}^n$ such that the integral $\|f\|_p = \left[\int_{\mathbb{R}^n} |f(x)|^p dx\right]^{1/p}$ 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 $L_p(\mathbb{R}^n)$, 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.

2.3. LINEAR TRANSFORMATIONS AND MATRICES

2.3.1. A mapping A from one linear space (its *domain)* into another (its *range*) is a *linear transformation* if it satisfies $A(\mathbf{x}+\mathbf{z}) = A(\mathbf{x}) + A(\mathbf{z})$ for any $\mathbf{x}$ and $\mathbf{z}$ 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 $\mathbf{A}$ with elements a_{ij} for $1 \le i \le m$ and $1 \le j \le n$, with $\mathbf{y} = A(\mathbf{x})$ having components $y_i = \sum_{j=1}^{n} a_{ij}x_j$ for $1 \le i \le m$. In matrix notation, this is written $\mathbf{y} = \mathbf{Ax}$. A matrix $\mathbf{A}$ is *real* if all its elements are real numbers, *complex* if some of its elements are complex numbers. Throughout, matrices are assumed to be real unless explicitly assumed otherwise. The set $\mathbb{N} = \{\mathbf{x}\in\mathbb{R}^n|\mathbf{Ax} = \mathbf{0}\}$ is termed the *null space* of the transformation $\mathbf{A}$. The subspace $\mathbb{N}^\perp$ containing all linear combinations of the column vectors of $\mathbf{A}$ is termed the *column space* of $\mathbf{A}$; it is the complementary subspace to $\mathbb{N}$.

If $\mathbf{A}$ denotes a m×n matrix, then $\mathbf{A}'$ 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 $\mathbf{I}_n$, or $\mathbf{I}$ if the dimension is clear from the context. A matrix of zeros is denoted $\mathbf{0}$, and a n×1 vector of ones is denoted $\mathbf{1}_n$. A *permutation matrix* is obtained by permuting the columns of an identity matrix. If $\mathbf{A}$ is a m×n matrix and $\mathbf{B}$ is a n×p matrix, then the *matrix product* $\mathbf{C} = \mathbf{AB}$ is of dimension m×p with elements $c_{ik} \equiv \sum_{j=1}^{n} a_{ij}b_{jk}$ for $1 \le i \le m$ and $1 \le k \le p$. For the matrix product to be defined, the number of columns in $\mathbf{A}$ must equal the number of rows in $\mathbf{B}$ (i.e., the matrices must be *commensurate*). A matrix $\mathbf{A}$ is *square* if it has the same number of rows and columns. A square matrix $\mathbf{A}$ is *symmetric* if $\mathbf{A} = \mathbf{A}'$, *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 $\mathbf{A}^2 = \mathbf{A}$. A matrix **A** is *column orthonormal* if $\mathbf{A}'\mathbf{A} = \mathbf{I}$; simply *orthonormal* if it is both square and column orthonormal.

A set of linearly independent vectors in $\mathbb{R}^n$ can be recursively orthonormalized; i.e., transformed so they are orthogonal and scaled to have unit length: Suppose vectors $\mathbf{x}_1,...,\mathbf{x}_{J-1}$ have previously been orthonormalized, and **z** is the next vector in the set. Then, $\mathbf{z} - \sum_{j=1}^{J-1} (\mathbf{x}_j'\mathbf{z})\mathbf{x}_j$ is orthogonal to $\mathbf{x}_1,...,\mathbf{x}_{J-1}$, and is non-zero since it is linearly independent. Scale it to unit length; this defines $\mathbf{x}_J$. Each column of a n×m matrix **A** is a vector in $\mathbb{R}^n$. The *rank* of **A**, denoted r = ρ(**A**), is the largest number of columns that are *linearly independent*. Then **A** is of rank m if and only if $\mathbf{x} = \mathbf{0}$ is the only solution to $\mathbf{Ax} = \mathbf{0}$. If A is of rank r, then orthonormalization applied to the linearly independent columns of **A** can be interpreted as defining a r×m lower triangluar matrix **U** such that $\mathbf{AU}'$ is column orthonormal. A n×m matrix **A** is of *full rank* if ρ(**A**) = min(n,m). A n×n matrix **A** of full rank is termed *nonsingular*. A nonsingular n×n matrix **A** has an *inverse* matrix $\mathbf{A}^{-1}$ such that both $\mathbf{AA}^{-1}$ and $\mathbf{A}^{-1}\mathbf{A}$ equal the identity matrix $\mathbf{I}_n$. An orthonormal matrix **A** satisfies $\mathbf{A}'\mathbf{A} = \mathbf{I}_n$, implying that $\mathbf{A}' = \mathbf{A}^{-1}$, and hence $\mathbf{A}'\mathbf{A} = \mathbf{AA}' = \mathbf{I}_n$. The trace tr(**A**) of a square matrix **A** is the sum of its diagonal elements.

2.3.2. The tables in this section summarize useful matrix and vector operations. In addition to the operations in these tables, 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 2.1-2.3 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.

TABLE 2.1. BASIC OPERATIONS

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

TABLE 2.2. OPERATIONS ON ELEMENTS

	Name	Notation	Definition
1.	Element Product	**C** = **A**.***B**	For **A**, **B** m×n: $c_{ij} = a_{ij}\, b_{ij}$
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**: $c_i = \min_{1\le k\le m} a_{ik}$ (Note 2)
5.	Row Min Replace	**C** = rmin(**A**)	For m×n **A**: $c_{ij} = \min_{1\le k\le m} a_{ik}$ (Note 3)
6.	Column Min Replace	**C** = cmin(**A**)	For m×n **A**: $c_{ij} = \min_{1\le k\le n} a_{ik}$ (Note 4)
7.	Cumulative Sum	**C** = cumsum(**A**)	For m×n **A**: $c_{ij} = \sum_{k=1}^{i} a_{kj}$

NOTES:
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 vector. The operation is also defined for "max".
3. **C** is a m×n matrix, with all columns the same. The operation is also defined for "max"
4. **C** is a m×n matrix, with all rows the same. The operation is also defined for "max".

TABLE 2.3. SHAPING OPERATIONS

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

NOTES:
1. Also termed the *direct product*, the Kronecker product creates an array made up of blocks, with each block the product of an element of **A** and the matrix **B**; see Section 2.11.

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

3. If **A** is m×n, then k must be a divisor of m·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·n/k columns.

2.3.3. The *determinant* of a n×n matrix **A** is denoted $|\mathbf{A}|$ or det(**A**), and has a geometric interpretation as the volume of the parallelepiped formed by the column vectors of **A**. The matrix **A** is *nonsingular* if and only if det(**A**) ≠ 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 $\mathbf{A}^{ij}$ formed by striking out row i and column j of **A**. Determinants satisfy the recursion relation

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

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

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

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

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

(1) $(\mathbf{A}')' = \mathbf{A}$.
(2) If $\mathbf{A}^{-1}$ exists, then $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
(3) If $\mathbf{A}^{-1}$ exists, then $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
(4) $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.
(5) If **A**,**B** are square, nonsingular, and commensurate, then $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
(6) If **A** is m×n, then Min {m,n} $\geq \rho(\mathbf{A}) = \rho(\mathbf{A}') = \rho(\mathbf{A}'\mathbf{A}) = \rho(\mathbf{AA}')$.
(7) If **A** is m×n and **B** is m×r, then $\rho(\mathbf{AB}) \leq \min(\rho(\mathbf{A}),\rho(\mathbf{B}))$.
(8) If **A** is m×n with $\rho(\mathbf{A}) = m$, and **B** is m×r, then $\rho(\mathbf{AB}) = \rho(\mathbf{B})$.
(9) $\rho(\mathbf{A}+\mathbf{B}) \leq \rho(\mathbf{A}) + \rho(\mathbf{B})$.
(10) If **A** is n×n, then det(**A**) ≠ 0 if and only if $\rho(\mathbf{A}) = n$.
(11) If **B** and **C** are nonsingular and commensurate with **A**, then $\rho(\mathbf{BAC}) = \rho(\mathbf{A})$.
(12) If **A**, **B** are n×n, then $\rho(\mathbf{AB}) \geq \rho(\mathbf{A}) + \rho(\mathbf{B}) - n$.
(13) $\det(\mathbf{AB}) = \det(\mathbf{A})\cdot\det(\mathbf{B})$.
(14) If *c* is a scalar and **A** is n×n, then $\det(c\mathbf{A}) = c^n\det(\mathbf{A})$
(15) The determinant of a matrix is unchanged if a scalar times one column (row) is added to another column (row).
(16) If **A** is n×n and diagonal or triangular, then det(**A**) is the product of the diagonal elements.
(17) $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$.

(18) If **A** is n×n and **B** = $\mathbf{A}^{-1}$, then $b_{ij} = (-1)^{i+j}\det(\mathbf{A}^{ij})/\det(\mathbf{A})$.
(19) The determinant of an orthonormal matrix is +1 or -1.
(20) If **A** is m×n and **B** is n×m, then tr(**AB**) = tr(**BA**).
(21) $\mathrm{tr}(\mathbf{I}_n) = n$.
(22) tr(**A**+**B**) = tr(**A**) + tr(**B**).
(23) A permutation matrix **P** is orthonormal; hence, $\mathbf{P}' = \mathbf{P}^{-1}$.
(24) The inverse of a (upper) triangular matrix is (upper) triangular, and the inverse of a diagonal matrix **D** is diagonal, with $(\mathbf{D}^{-1})_{ii} = 1/\mathbf{D}_{ii}$.
(25) The product of orthonormal matrices is orthonormal, and the product of permutation matrices is a permutation matrix.

2.4. EIGENVALUES AND EIGENVECTORS

An *eigenvalue* of a n×n matrix **A** is a scalar λ such that $\mathbf{Ax} = \lambda\mathbf{x}$ for some vector $\mathbf{x} \neq \mathbf{0}$. The vector **x** is called a (right) *eigenvector*. The condition $(\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ associated with an eigenvalue implies $\mathbf{A}-\lambda\mathbf{I}$ simgular, and hence $\det(\mathbf{A}-\lambda\mathbf{I}) = 0$. This determanental equation defines a polynomial in λ of order n, and the n roots of this polynomial are the eigenvalues. For each eigenvalue λ, the condition that $\mathbf{A}-\lambda\mathbf{I}$ 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 of a n×n matrix **A** are stated without proof:

(1) If **A** is real and symmetric, then its eigenvalues and eigenvectors are real. However, if **A** is nonsymmetric, then its eigenvalues and eigenvectors in general are complex.
(2) The number of nonzero eigenvalues of **A** equals its rank ρ(**A**).
(3) If λ is an eigenvalue of **A**, then λ^k is an eigenvalue of $\mathbf{A}^k$, and 1/λ is an eigenvalue of $\mathbf{A}^{-1}$ (if the inverse exists).
(4) If **A** is real and symmetric, then the eigenvalues corresponding to distinct roots are orthogonal. [$\mathbf{Ax}_i = \lambda_i\mathbf{x}_i$ implies $\mathbf{x}_i'\mathbf{Ax}_j = \lambda_i\mathbf{x}_i'\mathbf{x}_j = \lambda_j\mathbf{x}_i'\mathbf{x}_j$, which can be true for $i \neq j$ only if $\mathbf{x}_i'\mathbf{x}_j = 0$.]
(5) If **A** is real and symmetric, and **Λ** is a diagonal matrix with the roots of the polynomial $\det(\mathbf{A}-\lambda\mathbf{I})$ along the diagonal, then there exists an orthonormal matrix **C** such that $\mathbf{C}'\mathbf{C} = \mathbf{I}$ and $\mathbf{AC} = \mathbf{C}\boldsymbol{\Lambda}$, and hence $\mathbf{C}'\mathbf{AC} = \boldsymbol{\Lambda}$ and $\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' = \mathbf{A}$. The transformation **C** is said to *diagonalize* **A**. [Take **C** to be an array whose columns are eigenvectors of **A**, scaled to unit length. In the case of a multiple root, orthonormalize the eigenvectors corresponding to this root.].
(6) 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 $\mathbf{Q}^{-1}\mathbf{AQ} = \mathbf{T}$.
(7) **A** real and symmetric implies tr(**A**) equals the sum of the eigenvalues of **A**. [Since $\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$, $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}') = \mathrm{tr}(\mathbf{C}'\mathbf{C}\boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{\Lambda})$ by 2.3.20.]

(8) If $\mathbf{A}_i$ are real and symmetric for i = 1,...,p, then there exists $\mathbf{C}$ orthonormal such that $\mathbf{C}'\mathbf{A}_i\mathbf{C}$, are all diagonal if and only if $\mathbf{A}_i\mathbf{A}_j = \mathbf{A}_j\mathbf{A}_i$ for i,j = 1,...,p.

Results (5) and (6) combined with the result 2.3.13 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 (5) and (6) are called *similarity transformations*, and can be interpreted as representations of the transformation $\mathbf{A}$ when the basis of the domain is transformed by $\mathbf{C}$ (or $\mathbf{Q}$) and the basis of the range is transformed by $\mathbf{C}^{-1}$ (or $\mathbf{Q}^{-1}$). These transformations are used extensively in econometric theory.

2.5. PARTITIONED MATRICES

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

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

where $\mathbf{A}$ is m×n, $\mathbf{A}_{11}$ is $m_1 \times n_1$, $\mathbf{A}_{12}$ is $m_1 \times n_2$, $\mathbf{A}_{21}$ is $m_2 \times n_1$, $\mathbf{A}_{22}$ 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 $\mathbf{B}$ is n×p and is partitioned $\mathbf{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}$ where $\mathbf{B}_1$ is $n_1 \times p$ and $\mathbf{B}_2$ is $n_2 \times p$, one has $\mathbf{AB} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_1 + \boldsymbol{A}_{12}\boldsymbol{B}_2 \\ \boldsymbol{A}_{21}\boldsymbol{B}_1 + \boldsymbol{A}_{22}\boldsymbol{B}_2 \end{bmatrix}$.

Partitioned matrices have the following elementary properties:

(1) $\mathbf{A}$ square and $\mathbf{A}_{11}$ square and nonsingular implies $\det(\mathbf{A}) = \det(\mathbf{A}_{11})\cdot\det(\mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})$.
(2) $\mathbf{A}$ and $\mathbf{A}_{11}$ square and nonsingular implies

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

with $\mathbf{C} = \mathbf{A}_{22}-\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. When $\mathbf{A}_{22}$ is nonsingular, the northwest matrix in this partition can also be written as $(\mathbf{A}_{11}-\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}$.

2.6. QUADRATIC FORMS

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

A quadratic form $Q(\mathbf{x},\mathbf{A})$ may fall into one of the classes in the table below:

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

A quadratic form that is not in one of these four classes is termed *indefinite*. The basic properties of quadratic forms are listed below:

(1) If $\mathbf{B}$ is m×n and is of rank $\rho(\mathbf{B}) = r$, then $\mathbf{B}'\mathbf{B}$ and $\mathbf{B}\mathbf{B}'$ are both positive semidefinite; and if $r = m \leq n$, then $\mathbf{B}'\mathbf{B}$ is positive definite.
(2) If $\mathbf{A}$ is symmetric and positive semidefinite (positive definite), then the eigenvalues of $\mathbf{A}$ are nonnegative (positive). Similarly, if $\mathbf{A}$ is symmetric and negative semidefinite (negative definite), then the eigenvalues of $\mathbf{A}$ are nonpositive (negative).
(3) Every symmetric positive semidefinite matrix $\mathbf{A}$ has a symmetric positive semidefinite square root $\mathbf{A}^{1/2}$ [By 2.4.4, $\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}$ for some $\mathbf{C}$ orthonormal and $\mathbf{D}$ a diagonal matrix with the nonnegative eigenvalues down the diagonal. Then, $\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}'$ and $\mathbf{A}^{1/2} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{C}'$ with $\mathbf{D}^{1/2}$ a diagonal matrix of the positive square roots of the diagonal of $\mathbf{D}$.]
(4) If $\mathbf{A}$ is positive definite, then $\mathbf{A}^{-1}$ is positive definite.
(5) If $\mathbf{A}$ and $\mathbf{B}$ are real, symmetric n×n matrices and $\mathbf{B}$ is positive definite, then there exists a n×n matrix $\mathbf{Q}$ that simultaneously diagonalizes $\mathbf{A}$ and $\mathbf{B}$: $\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{\Lambda}$ diagonal and $\mathbf{Q}'\mathbf{B}\mathbf{Q} = \mathbf{I}$. [From 2.4(5), there exists a n×n orthonormal matrix $\mathbf{U}$ such that $\mathbf{U}'\mathbf{B}\mathbf{U} = \mathbf{D}$ is diagonal. Let $\mathbf{G}$ be an orthonormal matrix that diagonalizes $\mathbf{D}^{-1/2}\mathbf{U}'\mathbf{A}\mathbf{U}\mathbf{D}^{-1/2}$, and define $\mathbf{Q} = \mathbf{U}\mathbf{D}^{-1/2}\mathbf{G}$.]
(6) $\mathbf{B}$ positive definite and $\mathbf{A} - \mathbf{B}$ positive semidefinite imply $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ positive semidefinite. [For a vector $\mathbf{z}$, let $\mathbf{x} = \mathbf{Q}^{-1}\mathbf{z}$, where $\mathbf{Q}$ is the diagonalizing matrix from (5). Then $\mathbf{z}'(\mathbf{B} - \mathbf{A})\mathbf{z} = \mathbf{x}'\mathbf{Q}'(\mathbf{B} - \mathbf{A})\mathbf{Q}\mathbf{x} = \mathbf{x}'(\mathbf{\Lambda} - \mathbf{I})\mathbf{x} \geq 0$, so no diagonal element of $\mathbf{\Lambda}$ is less than one. Alternately, let $\mathbf{x} = \mathbf{Q}'\mathbf{z}$. Then $\mathbf{z}'(\mathbf{B}^{-1} - \mathbf{A}^{-1})\mathbf{z} = \mathbf{x}'\mathbf{Q}^{-1}(\mathbf{B}^{-1} - \mathbf{A}^{-1})(\mathbf{Q}')^{-1}\mathbf{x} = \mathbf{x}'(\mathbf{I} - \mathbf{\Lambda}^{-1})\mathbf{x}$ must be non-negative.]
(7) The following conditions are equivalent:
  (i) $\mathbf{A}$ is positive definite
  (ii) The principal minors of $\mathbf{A}$ are positive
  (iii) The leading principal minors of $\mathbf{A}$ are positive.

2.7. 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 and **L** and **U** are lower triangular matrices. This factorization is useful for computation of inverses, as triangular matrices are easily inverted by recursion.

Theorem 2.1. Each n×n matrix **A** can be written as **A** = **PLDU**′**Q**′, where **P** and **Q** are permutation matrices, **L** and **U** are lower triangular matrices, 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

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{D}_{11} & \mathbf{0} \\ \mathbf{0} & D_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{U}_{11}' & \mathbf{U}_{21}' \\ 0 & 1 \end{bmatrix}' ,$$

where $\mathbf{A}_{11}$, $\mathbf{L}_{11}$, $\mathbf{D}_{11}$, and $\mathbf{U}_{11}'$ are (n-1)×(n-1), $\mathbf{L}_{21}$ is 1×(n-1), $\mathbf{U}_{21}$ is 1×(n-1), and A_{22} and D_{22} are 1×1. Assume that $\mathbf{L}_{11}$, $\mathbf{D}_{11}$, and $\mathbf{U}_{11}$ have been defined so that $\mathbf{A}_{11} = \mathbf{L}_{11}\mathbf{D}_{11}\mathbf{U}_{11}'$, and that $\mathbf{L}_{11}^{-1}$ and $\mathbf{U}_{11}^{-1}$ also exist and have been computed. Let $\mathbf{S} = \mathbf{L}^{-1}$ and $\mathbf{T} = \mathbf{U}^{-1}$, and partition **S** and **T** commensurately with **L** and **U**. Then, $\mathbf{A}_{11}^{-1} = \mathbf{U}_{11}'^{-1}\mathbf{D}_{11}^{-1}\mathbf{L}_{11}^{-1}$ and the remaining elements must satisfy the equations

$$\begin{aligned}
&\mathbf{A}_{21} = \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{U}_{11}' \;\Rightarrow\; \mathbf{L}_{21} = \mathbf{A}_{21}\mathbf{U}_{11}'^{-1}\mathbf{D}_{11}^{-1} \equiv \mathbf{A}_{21}\mathbf{T}_{11}'\mathbf{D}_{11}^{-1} \\
&\mathbf{A}_{12} = \mathbf{L}_{11}\mathbf{D}_{11}\mathbf{U}_{21}' \;\Rightarrow\; \mathbf{U}_{21}' = \mathbf{D}_{11}^{-1}\mathbf{L}_{11}^{-1}\mathbf{A}_{12} \equiv \mathbf{D}_{11}^{-1}\mathbf{S}_{11}\mathbf{A}_{12} \\
&A_{22} = \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{U}_{21}' + D_{22} \;\Rightarrow\; D_{22} = A_{22} - \mathbf{A}_{21}\mathbf{U}_{11}'^{-1}\mathbf{D}_{11}^{-1}\mathbf{L}_{11}^{-1}\mathbf{A}_{12} = A_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\
&\mathbf{S}_{21} = -\mathbf{L}_{21}\mathbf{S}_{11} \qquad \mathbf{S}_{22} = 1 \\
&\mathbf{T}_{21}' = -\mathbf{T}_{11}'\mathbf{U}_{21}' \qquad \mathbf{T}_{22} = 1
\end{aligned}$$

where $\det(\mathbf{A}) = \det(\mathbf{A}_{11})\cdot\det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) \neq 0$ 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 $\mathbf{A}^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{S}$.

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

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{D}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{U}_{11}' & \mathbf{U}_{21}' \\ \mathbf{0} & \mathbf{I} \end{bmatrix} ,$$

where $\mathbf{A}_{11}$ is r×r and the remaining blocks are commensurate. Then, $\mathbf{U}_{21}' = \mathbf{D}_{11}^{-1}\mathbf{S}_{11}\mathbf{A}_{12}$ and $\mathbf{L}_{21} = \mathbf{A}_{21}\mathbf{T}_{11}'\mathbf{D}_{11}^{-1}$, and one must satisfy $\mathbf{A}_{22} = \mathbf{L}_{21}\mathbf{D}_{11}\mathbf{U}_{12}' = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. But the rank condition implies the last n-r columns of **A** can be written as a linear combination $\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} C$ of the first r columns, where **C** is some r×(n-r) matrix. But $\mathbf{A}_{12} = \mathbf{A}_{11}\mathbf{C}$ implies $\mathbf{C} = \mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ and hence $\mathbf{A}_{22} = \mathbf{A}_{21}\mathbf{C} = \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ 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 2.1.1. If **A** is symmetric, then **L** = **U**.
Corollary 2.1.2. (LU Factorization) If **A** has nonzero leading principal minors, then **A** can be written $\mathbf{A} = \mathbf{L}\mathbf{V}'$, where $\mathbf{V}' = \mathbf{D}\mathbf{U}'$ is upper triangular with a diagonal coinciding with that of **D**.

Corollary 2.1.3. (Cholesky Factorization) If **A** is symmetric and positive definite, then **A** can be written $\mathbf{A} = \mathbf{V}\mathbf{V}'$, where $\mathbf{V} = \mathbf{L}\mathbf{D}^{1/2}$ is lower triangular with a positive diagonal.
Corollary 2.1.4. **A** symmetric positive semidefinite implies $\mathbf{A} = \mathbf{P}\mathbf{V}\mathbf{V}'\mathbf{P}'$, with **V** lower triangular with a nonnegative diagonal, **P** a permutation matrix.
Corollary 2.1.5. If **A** is m×n with m ≥ n, then there exists a factorization $\mathbf{A} = \mathbf{PLDU}'\mathbf{Q}'$, with **D** n×n diagonal, **P** a m×m permutation matrix, **Q** a n×n permutation matrix, **U** a n×n lower 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 $\mathbf{L}' = [\mathbf{L}_{11}' \;\; \mathbf{L}_{21}']$ with $\mathbf{L}_{11}$ n×n and lower triangular with ones on the diagonal, and $\mathbf{L}_{21}$ (m-n)×n. Further, if ρ(A) = n, then $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{Q}\mathbf{U}'^{-1}\mathbf{D}^{-1}(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\mathbf{P}'$.
Corollary 2.1.6. If the system of equations $\mathbf{Ax} = \mathbf{y}$ with **A** m×n of rank n has a solution, then the solution is given by $\mathbf{x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} = \mathbf{Q}\mathbf{U}'^{-1}\mathbf{D}^{-1}(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'\mathbf{P}'\mathbf{y}$.

Proof outline: 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 $\mathbf{V}' = \mathbf{D}^{1/2}\mathbf{U}'$, where $\mathbf{D}^{1/2}$ 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 ρ(**A**). Suppose **A** =

$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ is of this form, and apply the theorem to obtain $\mathbf{A}_{11} = \mathbf{P}_{11}\mathbf{L}_{11}\mathbf{D}\mathbf{U}'\mathbf{Q}'$. The rank condition implies that $\mathbf{A}_{21} = \mathbf{F}\mathbf{A}_{11}$ for some (m-n)×n array $\mathbf{F}$. Then, $\mathbf{A}_{21} = \mathbf{L}_{21}\mathbf{D}\mathbf{U}'\mathbf{Q}'$, with $\mathbf{L}_{21} = \mathbf{F}\mathbf{P}_{11}\mathbf{L}_{11}$, so that

$$\mathbf{A} = \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix} \mathbf{D}\mathbf{U}'\mathbf{Q}'.$$

To complete the proof, apply a left permutation if necessary to undo the initial row permutation of **A**. Corollary 6 is an implication of the last result. □

The recursion in the proof of the theorem is called *Crout's algorithm*, and is the method for matrix inversion of positive definite matrices 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.

The Cholesky factorization of a n×n positive definite matrix **A** that was obtained above as a corollary of the **LDU** decomposition states that **A** can be written as $\mathbf{A} = \mathbf{L}\mathbf{L}'$, where **L** is lower triangular with a positive diagonal. This factorization is readily computed and widely used in econometrics. We give a direct recursive construction of **L** that forms the basis for its computation. Write the factorization in partitioned form

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ U_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix}' .$$

Also, let $\mathbf{V} = \mathbf{L}^{-1}$, and partition it commensurately, so that

$$\begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} V_{11} & 0 & 0 \\ V_{21} & V_{22} & 0 \\ V_{31} & V_{32} & V_{33} \end{bmatrix} .$$

Then $\mathbf{A}_{11} = \mathbf{L}_{11}\mathbf{L}_{11}'$, $\mathbf{A}_{12} = \mathbf{L}_{11}\mathbf{L}_{21}'$, $\mathbf{A}_{22} = \mathbf{L}_{21}\mathbf{L}_{21}' + \mathbf{L}_{22}\mathbf{L}_{22}'$, $\mathbf{V}_{11} = \mathbf{L}_{11}^{-1}$, $\mathbf{V}_{22} = \mathbf{L}_{22}^{-1}$, and $\mathbf{0} = \mathbf{L}_{21}\mathbf{V}_{11} + \mathbf{L}_{22}'\mathbf{V}_{22}$. Note first that if $\mathbf{A}_{11}$ is 1×1, then $\mathbf{L}_{11} = \mathbf{A}_{11}^{1/2}$ and $\mathbf{V}_{11} = 1/\mathbf{L}_{11}$. Now suppose that one has proceeded recursively from the northwest corner of these matrices, and that $\mathbf{L}_{11}$ and $\mathbf{V}_{11}$ have already been computed up through dimension n_1. Suppose that $\mathbf{A}_{22}$ is 1×1. Then, compute in sequence $\mathbf{L}_{21}' = \mathbf{V}_{11}'\mathbf{A}_{12}$, $\mathbf{L}_{22} = (\mathbf{A}_{22} - \mathbf{L}_{12}\mathbf{L}_{12}')^{1/2}$, $\mathbf{V}_{22} = 1/\mathbf{L}_{22}$, and $\mathbf{V}_{12} = -\mathbf{V}_{11}\mathbf{L}_{21}'\mathbf{V}_{22}$. This gives the required factors up through dimension n_1+1. Repeat this for each dimension in turn to construct the full **L** and **V** matrices.

An extension of the Cholesky decomposition holds for an n×n positive semidefinite matrix **A** of rank r, which can be written as $\mathbf{A} = \mathbf{PLL}'\mathbf{P}'$ with **P** a permutation matrix and **L** a lower triangular matrix whose first r diagonal elements are positive. The construction proceeds recursively as before, but at each stage one may have to search among remaining columns to find one for which $\mathbf{L}_{22} > 0$, determining the **P** matrix. Once dimension r is reached, all remaining columns will have $\mathbf{L}_{22} = 0$. Now reinterpret $\mathbf{L}_{21}$ and $\mathbf{L}_{22}$ as a partition corresponding to all the remaining columns and compute $\mathbf{L}_{12}' = \mathbf{V}_{11}'\mathbf{A}_{12}$ and $\mathbf{L}_{22} = \mathbf{0}$ to complete the Cholesky factor.

2.8. 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):

Theorem 2.2. Every real m×n matrix **A** of rank r can be decomposed into a product $\mathbf{A} = \mathbf{UDV}'$, where **D** is a r×r diagonal matrix with positive nonincreasing elements down the diagonal, **U** is m×r, **V** is n×r, and **U** and **V** are column-orthonormal; i.e., $\mathbf{U}'\mathbf{U} = \mathbf{I}_r = \mathbf{V}'\mathbf{V}$.

Proof: Note that the SVD is an extension of the LDU decomposition to non-square matrices. To prove that the SVD is possible, note first that the m×m matrix $\mathbf{AA}'$ is symmetric and positive semidefinite. Then, there exists a m×m orthonormal matrix **W** whose columns are eigenvectors of $\mathbf{AA}'$ arranged in non-increasing order for the eigenvalues, partitioned $\mathbf{W} = [\mathbf{W}_1\ \mathbf{W}_2]$ with $\mathbf{W}_1$ of dimension m×r, such that $\mathbf{W}_1'(\mathbf{AA}')\mathbf{W}_1 = \mathbf{\Lambda}$ is diagonal with positive, non-increasing diagonal elements, and $\mathbf{W}_2'(\mathbf{AA}')\mathbf{W}_2 = 0$, implying $\mathbf{A}'\mathbf{W}_2 = \mathbf{0}$. Define **D** from **Λ** by replacing the diagonal elements of **Λ** by their positive square roots. Note that $\mathbf{W}'\mathbf{W} = \mathbf{I} = \mathbf{WW}' \equiv \mathbf{W}_1\ \mathbf{W}_1' + \mathbf{W}_2\mathbf{W}_2'$. Define $\mathbf{U} = \mathbf{W}_1$ and $\mathbf{V}' = \mathbf{D}^{-1}\mathbf{U}'\mathbf{A}$. Then, $\mathbf{U}'\mathbf{U} = \mathbf{I}_r$ and $\mathbf{V}'\mathbf{V} = \mathbf{D}^{-1}\mathbf{U}'\mathbf{AA}'\mathbf{UD}^{-1} = \mathbf{D}^{-1}\mathbf{\Lambda}\mathbf{D}^{-1} = \mathbf{I}_r$. Further, $\mathbf{A} = (\mathbf{I}_m\text{-}\mathbf{W}_2\ \mathbf{W}_2')\mathbf{A} = \mathbf{UU}'\mathbf{A} = \mathbf{UDV}'$. 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 $\mathbf{A}'\mathbf{U} = \mathbf{UD}_1$, with $\mathbf{D}_1$ the r×r diagonal matrix with the non-zero eigenvalues in descending magnitude down the diagonal. In this case, $\mathbf{V} = \mathbf{A}'\mathbf{UD}^{-1} = \mathbf{UD}_1\mathbf{D}^{-1}$. Since the elements of $\mathbf{D}_1$ 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 $\mathbf{A}'\mathbf{A}$ is symmetric and positive definite, the SVD provides a method of calculating $(\mathbf{A}'\mathbf{A})^{-1}$ that is particularly numerically accurate: Substituting the form $\mathbf{A} = \mathbf{UDV}'$, one obtains $(\mathbf{A}'\mathbf{A})^{-1} = \mathbf{VD}^{-2}\mathbf{V}'$. One also obtains convenient forms for a square root of $\mathbf{A}'\mathbf{A}$ and its inverse, $(\mathbf{A}'\mathbf{A})^{1/2} = \mathbf{VDV}'$ and $(\mathbf{A}'\mathbf{A})^{-1/2} = \mathbf{VD}^{-1}\mathbf{V}'$.

The numerical accuracy of the SVD is most advantageous when m is large and some of the columns of A are nearly linearly dependent. Then, roundoff errors in the matrix product $\mathbf{A}'\mathbf{A}$ can lead to quite inaccurate results when a matrix inverse of $\mathbf{A}'\mathbf{A}$ is computed directly. The SVD extracts the required information from **A** before the roundoff errors in $\mathbf{A}'\mathbf{A}$ are introduced. Computer programs for the Singular Value Decomposition can be found in Press *et al*, Numerical Recipes, Cambridge University Press, 1986.

2.9. IDEMPOTENT MATRICES AND GENERALIZED INVERSES

A symmetric n×n matrix **A** is *idempotent* if $\mathbf{A}^2 = \mathbf{A}$. Examples of idempotent matrices are **0**, **I**, and for any n×r matrix **X** of rank r, $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Idempotent matrices are intimately related to projections, discussed in the following section. Some of the properties of an n×n idempotent matrix **A** are listed below:

(1) The eigenvalues of **A** are either zero or one.
(2) The rank of **A** equals tr(**A**).
(3) The matrix **I**-**A** is idempotent.
(4) If **B** is an orthonormal matrix, then $\mathbf{B}'\mathbf{AB}$ is idempotent.
(5) If $\rho(\mathbf{A}) = r$, then there exists a n×r matrix **B** of rank r such that $\mathbf{A} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$. [Let **C** be an orthonormal matrix that diagonalizes **A**, and take **B** to be the columns of **C** corresponding to the non-zero elements in the diagonalization.]
(6) **A**, **B** idempotent implies **AB** = **0** if and only if **A**+**B** is 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**.

Recall that a n×n non-singular matrix **A** has an inverse $\mathbf{A}^{-1}$ that satisfies $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. It is useful to extend the concept of an inverse to matrices that are not necessarily non-singular, or even square. For an m×k matrix **A** (of rank r), define its *Moore-Penrose generalized inverse* $\mathbf{A}^{-}$ to be a k×m matrix with the following three properties:

(i) $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$,
(ii) $\mathbf{A}^{-}\mathbf{AA}^{-} = \mathbf{A}^{-}$
(iii) $\mathbf{AA}^{-}$ and $\mathbf{A}^{-}\mathbf{A}$ are symmetric

The next theorem shows that the Moore-Penrose generalized inverse always exists, and is unique. Conditions (i) and (ii) imply that the matrices $\mathbf{AA}^-$ and $\mathbf{A}^-\mathbf{A}$ are idempotent. There are other generalized inverse definitions that have some, but not all, of the properties (i)-(iii); in particular $\mathbf{A}^+$ will denote any matrix that satisfies (i), or $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$.

Theorem 2.3. *The Moore-Penrose generalized inverse of a* m×k matrix $\mathbf{A}$ of rank r (which has a SVD $\mathbf{A} = \mathbf{UDV}'$, where $\mathbf{U}$ is m×r, $\mathbf{V}$ is k×r, $\mathbf{U}$ and $\mathbf{V}$ are column-orthogonal, and $\mathbf{D}$ is r×r diagonal with positive diagonal elements) is the matrix $\mathbf{A}^- = \mathbf{VD}^{-1}\mathbf{U}'$. Let $\mathbf{A}^+$ denote any matrix, including $\mathbf{A}^-$, that satisfies $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$. These matrices satisfy:

(1) $\mathbf{A}^- = \mathbf{A}^+ = \mathbf{A}^{-1}$ if $\mathbf{A}$ is square and non-singular.
(2) The system of equations $\mathbf{Ax} = \mathbf{y}$ has a solution if and only if $\mathbf{y} = \mathbf{AA}^+\mathbf{y}$, and the linear subspace of all solutions is the set of vectors $\mathbf{x} = \mathbf{A}^+\mathbf{y} + [\mathbf{I} - \mathbf{A}^+\mathbf{A}]\mathbf{z}$ for $\mathbf{z} \in \mathbb{R}^k$.
(3) $\mathbf{AA}^+$ and $\mathbf{A}^+\mathbf{A}$ are idempotent.
(4) If $\mathbf{A}$ is idempotent, then $\mathbf{A} = \mathbf{A}^-$.
(5) If $\mathbf{A} = \mathbf{BCD}$ with $\mathbf{B}$ and $\mathbf{D}$ nonsingular, then $\mathbf{A}^- = \mathbf{D}^{-1}\,\mathbf{C}^-\,\mathbf{B}^{-1}$, and any matrix $\mathbf{A}^+ = \mathbf{D}^{-1}\mathbf{C}^+\mathbf{B}^{-1}$ satisfies $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$.
(6) $(\mathbf{A}')^- = (\mathbf{A}^-)'$
(7) $(\mathbf{A}'\mathbf{A})^- = \mathbf{A}^-(\mathbf{A}^-)'$
(8) $(\mathbf{A}^-)^- = \mathbf{A} = \mathbf{AA}'(\mathbf{A}^-)' = (\mathbf{A}^-)'\mathbf{A}'\mathbf{A}$.

(9) If $\mathbf{A} = \sum_i \mathbf{A}_i$ with $\mathbf{A}_i'\mathbf{A}_j = \mathbf{0}$ and $\mathbf{A}_i\mathbf{A}_j' = \mathbf{0}$ for i ≠ j, then $\mathbf{A}^- = \sum_i \mathbf{A}_i^-$.

Theorem 2.4. If $\mathbf{A}$ is m×m, symmetric, and positive semidefinite of rank r, then
(1) There exist $\mathbf{Q}$ positive definite and $\mathbf{R}$ idempotent of rank r such that $\mathbf{A} = \mathbf{QRQ}$ and $\mathbf{A}^- = \mathbf{Q}^{-1}\mathbf{RQ}^{-1}$.
(2) There exists an m×r column-orthonormal matrix $\mathbf{U}$ such that $\mathbf{U}'\mathbf{AU} = \mathbf{D}$ is positive diagonal, $\mathbf{A} = \mathbf{UDU}'$, $\mathbf{A}^- = \mathbf{UD}^{-1}\mathbf{U}' = \mathbf{U}(\mathbf{U}'\mathbf{AU})^{-1}\mathbf{U}'$, and any matrix $\mathbf{A}^+$ satisfying condition (i) for a generalized inverse, $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$, has $\mathbf{U}'\mathbf{A}^+\mathbf{U} = \mathbf{D}^{-1}$.
(3) $\mathbf{A}$ has a symmetric square root $\mathbf{B} = \mathbf{A}^{1/2}$, and $\mathbf{A}^- = \mathbf{B}^-\mathbf{B}^-$.

Proof: Let $\mathbf{U}$ be an m×r column-orthonormal matrix of eigenvectors of $\mathbf{A}$ corresponding to the positive characteristic roots, and $\mathbf{W}$ be a m×(m-r) column-orthonormal matrix of eigenvectors corresponding to the zero characteristic roots. Then $[\mathbf{U}\ \mathbf{W}]$ is an orthonormal matrix diagonalizing $\mathbf{A}$, with $\begin{bmatrix} U' \\ W' \end{bmatrix} A \,[U\ \ W] = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and $\mathbf{D}$ positive diagonal. Define $\mathbf{Q} = [U\ \ W]\begin{bmatrix} D^{1/2} & \mathbf{0} \\ \mathbf{0} & I_{m-r} \end{bmatrix}\begin{bmatrix} U' \\ W' \end{bmatrix}$,

and $\mathbf{R} = \mathbf{UU}'$. The diagonalizing transformation implies $\mathbf{U}'\mathbf{AU} = \mathbf{D}$ and $\mathbf{AW} = \mathbf{0}$. One has $\mathbf{U}'\mathbf{U} = \mathbf{I}_r$,

$\mathbf{W}'\mathbf{W} = \mathbf{I}_{m-r}$, and $\mathbf{UU}' + \mathbf{WW}' = \mathbf{I}_m$. Since $\mathbf{AW} = \mathbf{0}$, $\mathbf{A} = \mathbf{A}[\mathbf{UU}' + \mathbf{WW}'] = \mathbf{AUU}'$ and $\mathbf{D} = \mathbf{U}'\mathbf{AU} = \mathbf{U}'\mathbf{AA}^+\mathbf{AU} = \mathbf{UAUU}'\mathbf{A}^+\mathbf{UU}'\mathbf{AU} = \mathbf{DU}'\mathbf{A}^+\mathbf{UD}$, implying $\mathbf{U}'\mathbf{A}^+\mathbf{U} = \mathbf{D}^{-1}$. Define $\mathbf{B} = \mathbf{UD}^{1/2}\mathbf{U}'$. □

2.10. PROJECTIONS

Consider a Euclidean space $\mathbb{R}^n$ of dimension n, and suppose $\mathbf{X}$ is a n×p array with columns that are vectors in this space. Let $\boldsymbol{\mathcal{X}}$ denote the linear subspace of $\mathbb{R}^n$ that is *spanned* or *generated* by $\mathbf{X}$; i.e., the space formed by all linear combinations of the vectors in $\mathbf{X}$. Every linear subspace can be identified with an array such as $\mathbf{X}$. The dimension of the subspace is the rank of $\mathbf{X}$. (The array $\mathbf{X}$ need not be of full rank, although if it is not, then a subarray of linearly independent columns also generates $\boldsymbol{\mathcal{X}}$.) A given $\mathbf{X}$ determines a unique subspace, so that $\mathbf{X}$ characterizes the subspace. However, any set of vectors contained in the subspace that form an array with the rank of the subspace, in particular any array $\mathbf{XA}$ with rank equal to the dimension of $\mathbf{X}$, also generates $\boldsymbol{\mathcal{X}}$. Then, $\mathbf{X}$ is not a unique characterization of the subspace it generates.

The *projection* of a vector y in $\mathbb{R}^n$ into the subspace $\boldsymbol{\mathcal{X}}$ is defined as the point $\mathbf{v}$ in $\boldsymbol{\mathcal{X}}$ that is the minimum Euclidean distance from $\mathbf{y}$. Since each vector $\mathbf{v}$ in $\boldsymbol{\mathcal{X}}$ can be represented as a linear combination $\mathbf{X}\boldsymbol{\alpha}$ of an array $\mathbf{X}$ that generates $\boldsymbol{\mathcal{X}}$, the projection is characterized by the value of $\boldsymbol{\alpha}$ that minimizes the squared Euclidean distance of $\mathbf{y}$ from $\boldsymbol{\mathcal{X}}$, $(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})$. The solution to this problem is the vector $\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ giving $\mathbf{v} = \mathbf{X}\hat{\alpha} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$. In these formulas, we use the Moore-Penrose generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$ rather than $(\mathbf{X}'\mathbf{X})^{-1}$ so that the solution is defined even if $\mathbf{X}$ is not of full rank. The array $\boldsymbol{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is termed the *projection matrix* for the subspace $\boldsymbol{\mathcal{X}}$; it is the linear transformation in $\mathbb{R}^n$ that maps any vector in the space into its projection $\mathbf{v}$ in $\boldsymbol{\mathcal{X}}$. The matrix $\boldsymbol{P}_X$ is *idempotent* (i.e., $\boldsymbol{P}_X\boldsymbol{P}_X = \boldsymbol{P}_X$ and $\boldsymbol{P}_X = \boldsymbol{P}_X'$), and every idempotent matrix can be interpreted as a projection matrix. These observations have two important implications: First, the projection matrix is uniquely determined by $\mathbf{X}$, so that starting from a different array that generates $\boldsymbol{\mathcal{X}}$, say an array $\mathbf{S} = \mathbf{XA}$, implies $\boldsymbol{P}_X = \boldsymbol{P}_S$. (One could use the notation $\boldsymbol{P}_{\boldsymbol{\mathcal{X}}}$ rather than $\boldsymbol{P}_X$ to emphasize that the projection matrix depends only on the subspace, and not on any particular set of vectors that generate $\boldsymbol{\mathcal{X}}$.) Second, if a vector $\mathbf{y}$ is contained in $\boldsymbol{\mathcal{X}}$, then the projection into $\boldsymbol{\mathcal{X}}$ leaves it unchanged, $\boldsymbol{P}_X\mathbf{y} = \mathbf{y}$.

Define $\boldsymbol{Q}_X = \mathbf{I} - \boldsymbol{P}_X = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$; it is the projection to the subspace orthogonal to that spanned by $\mathbf{X}$. Every vector $\mathbf{y}$ in $\mathbb{R}^n$ is uniquely decomposed into the sum of its projection $\boldsymbol{P}_X\mathbf{y}$ onto $\boldsymbol{\mathcal{X}}$ and its projection $\boldsymbol{Q}_X\mathbf{y}$ onto the subspace orthogonal to $\boldsymbol{\mathcal{X}}$. Note that $\boldsymbol{P}_X\boldsymbol{Q}_X = \mathbf{0}$, a property that holds in general for two projections onto orthogonal subspaces.

If $\boldsymbol{\mathcal{X}}$ is a subspace generated by an array $\mathbf{X}$ and $\boldsymbol{\mathcal{W}}$ is a subspace generated by a more inclusive array $\mathbf{W} = [\mathbf{X}\ \mathbf{Z}]$, then $\boldsymbol{\mathcal{X}} \subseteq \boldsymbol{\mathcal{W}}$. This implies that $\boldsymbol{P}_X\boldsymbol{P}_W = \mathbf{P}_{\mathcal{W}}\mathbf{P}_{\mathcal{X}} = \boldsymbol{P}_X$; i.e., a projection onto a subspace is left invariant by a further projection onto a larger subspace, and a two-stage projection onto a large subspace followed by a projection onto a smaller one is the same as projecting directly onto the smaller one. The subspace of $\boldsymbol{\mathcal{W}}$ that is orthogonal to $\boldsymbol{\mathcal{X}}$ is generated by $\boldsymbol{Q}_X\mathbf{W}$; i.e., it is the set of linear combinations of the residuals, orthogonal to $\boldsymbol{X}$, obtained by the difference of $\mathbf{W}$ and its projection onto $\boldsymbol{X}$. Note that any $\mathbf{y}$ in $\mathbb{R}^n$ has a unique decomposition $\boldsymbol{P}_X\mathbf{y} + \boldsymbol{Q}_X\boldsymbol{P}_W\mathbf{y} + \boldsymbol{Q}_W\mathbf{y}$ into the sum of projections onto three mutually orthogonal subspaces, $\boldsymbol{X}$, the subspace of $\boldsymbol{W}$ orthogonal to $\boldsymbol{X}$, and the subspace orthogonal to $\boldsymbol{W}$. The projection $\boldsymbol{Q}_X\boldsymbol{P}_W$ can be rewritten $\boldsymbol{Q}_X\boldsymbol{P}_W = \boldsymbol{P}_W - \boldsymbol{P}_X = \boldsymbol{P}_W\boldsymbol{Q}_X = \boldsymbol{Q}_X\boldsymbol{P}_W\boldsymbol{Q}_X$, or since $\boldsymbol{Q}_X\mathbf{W} = \boldsymbol{Q}_X[\mathbf{X}\ \ \mathbf{Z}] = [\mathbf{0}\ \ \boldsymbol{Q}_X\mathbf{Z}]$, $\boldsymbol{Q}_X\boldsymbol{P}_W = \boldsymbol{P}_{\boldsymbol{Q}_X W} = \boldsymbol{P}_{\boldsymbol{Q}_X Z} = \boldsymbol{Q}_X\mathbf{Z}(\mathbf{Z}'\boldsymbol{Q}_X\mathbf{Z})^{-}\,\mathbf{Z}'\boldsymbol{Q}_X$.

This establishes that $\boldsymbol{P}_W$ and $\boldsymbol{Q}_X$ commute. This condition is necessary and sufficient for the product of two projections to be a projection; equivalently, it implies that $\boldsymbol{Q}_X\mathbf{P}_W$ is idempotent since $(\boldsymbol{Q}_X\boldsymbol{P}_W)(\boldsymbol{Q}_X\boldsymbol{P}_W) = \boldsymbol{Q}_X(\boldsymbol{P}_W\boldsymbol{Q}_X)\boldsymbol{P}_W = \boldsymbol{Q}_X(\boldsymbol{Q}_X\boldsymbol{P}_W)\boldsymbol{P}_W = \boldsymbol{Q}_X\boldsymbol{P}_W$.

2.11. KRONECKER PRODUCTS

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

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1m}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2m}\boldsymbol{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1}\boldsymbol{B} & a_{n2}\boldsymbol{B} & \dots & a_{nm}\boldsymbol{B} \end{bmatrix}.$$

In general, $\mathbf{A}\otimes\mathbf{B} \neq \mathbf{B}\otimes\mathbf{A}$. The Kronecker product has the following properties:

(1) For a scalar c, $(c\mathbf{A})\otimes\mathbf{B} = \mathbf{A}\otimes(c\mathbf{B}) = c(\mathbf{A}\otimes\mathbf{B})$.
(2) $(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})$.
(3) $(\mathbf{A}\otimes\mathbf{B})' = (\mathbf{A}')\otimes(\mathbf{B}')$.
(4) $\mathrm{tr}(\mathbf{A}\otimes\mathbf{B}) = (\mathrm{tr}(\mathbf{A}))\cdot(\mathrm{tr}(\mathbf{B}))$ when $\mathbf{A}$ and $\mathbf{B}$ are square.
(5) If the matrix products $\mathbf{AC}$ and $\mathbf{BF}$ are defined, then $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{F}) = (\mathbf{AC})\otimes(\mathbf{BF})$.
(6) If $\mathbf{A}$ and $\mathbf{B}$ are square and nonsingular, then $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$.
(7) If $\mathbf{A}$ and $\mathbf{B}$ are orthonormal, then $\mathbf{A}\otimes\mathbf{B}$ is orthonormal.
(8) If $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite, then $\mathbf{A}\otimes\mathbf{B}$ is positive semidefinite.
(9) If $\mathbf{A}$ is k×k and $\mathbf{B}$ is n×n, then $\det(\mathbf{A}\otimes\mathbf{B}) = \det(\mathbf{A})^n\cdot\det(\mathbf{B})^k$.
(10) $\rho(\mathbf{A}\otimes\mathbf{B}) = \rho(\mathbf{A})\cdot\rho(\mathbf{B})$.
(11) $(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes\mathbf{C} + \mathbf{B}\otimes\mathbf{C}$.

2.12. 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** = vecc(**A**) which creates a vector by stacking the columns of **A**, and vecr(**A**) = vecc(**A**′); (3) **c** = vech(**A**) which creates a vector by stacking the portions of the rows 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**. (In some computer matrix languages, vec(**A**) stacks by row rather than by column.) 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) vecc(**A**+**B**) = vecc(**A**) + vecc(**B**).
(3) If **A** is m×k and **B** is k×n, then vecr(**AB**) = ($\mathbf{I}_n \otimes \mathbf{A}$)vecr(**B**) = ($\mathbf{B}' \otimes \mathbf{I}_m$)vecr(**A**).
(4) If **A** is m×k, **B** is k×n, **C** is n×p, then vecr(**ABC**) = ($\mathbf{I}_p \otimes (\mathbf{AB})$)vecr© = ($\mathbf{C}' \otimes \mathbf{A}$)vecr(**B**) = (($\mathbf{C}'\mathbf{B}'$)$\otimes \mathbf{I}_m$)vecr(**A**).
(5) If **A** is n×n, then vech(**A**) is of length n(n+1)/2.
(6) vecd(diag(**x**)) = **x**.

2.13. 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×1 vector of variables, f(**x**), has partial derivatives that are usually written as the arrays

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

Other common notation is $f_{\mathbf{x}}(\mathbf{x})$ or $\nabla_{\mathbf{x}} f(\mathbf{x})$ for the vector of first derivatives, and $f_{\mathbf{xx}}(\mathbf{x})$ or $\nabla_{\mathbf{xx}} f(\mathbf{x})$ 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 $\nabla_{\mathbf{x}} f = \mathbf{a}$, and the quadratic function f(**x**) = **x**′**Ax**, which has $\nabla_{\mathbf{x}} f = 2\mathbf{Ax}$.

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

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

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

$$\int_{\boldsymbol{A}} g(\mathbf{y})d\mathbf{y} = \int_{\boldsymbol{f}^{-1}(\boldsymbol{A})} g(\mathbf{f}(\mathbf{x}))\cdot|\det(\boldsymbol{J}(\mathbf{x}))|d\mathbf{x},$$

where $\mathbf{f}^{-1}(\mathbf{A})$ is the set of **x** vectors that map onto **A**, and the Jacobean matrix is square and nonsingular for well-behaved one-to-one transformations. The intuition for the presence of the Jacobean determinant in the transformed integral is that "d**y**" is the volume of a small rectangle in **y**-space, and because determinants give the volume of the parallelepiped formed by the columns of a linear transformation, "det(***J***(**x**))d**x**" gives the volume (with a plus or minus sign) of the image in **x**-space of the "d**y**" 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 $f(\mathbf{A}) = \mathbf{x}'\mathbf{A}\mathbf{y}$, 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 $\partial f/\partial \mathbf{A} = \mathbf{x}\mathbf{y}'$. Another example for a n×n matrix **A** is $f(\mathbf{A}) = tr(\mathbf{A})$, which has $\partial f/\partial \mathbf{A} = \mathbf{I}_n$. 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 happen 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 $\det(\mathbf{A}) > 0$, then $\partial\log(\det(\mathbf{A}))/\partial\mathbf{A} = \mathbf{A}^{-1}$.
(2) If **A** is nonsingular, then $\partial(\mathbf{x}'\mathbf{A}^{-1}\mathbf{y})/\partial\mathbf{A} = -\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}$.
(3) If $\mathbf{A} = \mathbf{T}\mathbf{T}'$, with **T** square and nonsingular, then $\partial(\mathbf{x}'\mathbf{A}^{-1}\mathbf{y})/\partial\mathbf{T} = -2\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}\mathbf{T}$.

We prove the formulas in order. For (1), recall that $\det(\mathbf{A}) = \sum_k (-1)^{i+k} a_{ik} \det(\mathbf{A}^{ik})$, where $\mathbf{A}^{ik}$ is the minor of a_{ik}. Then, $\partial\det(\mathbf{A})/\partial\mathbf{A}_{ij} = (-1)^{i+j}\det(\mathbf{A}^{ij})$. From 2.3.17, the ij element of $\mathbf{A}^{-1}$ is $(-1)^{i+j}\det(\mathbf{A}^{ij})/\det(\mathbf{A})$. For (2), apply the chain rule to the identity $\mathbf{A}\mathbf{A}^{-1} \equiv I$ to get $\boldsymbol{\Delta}_{ij}\mathbf{A}^{-1} + \mathbf{A}\cdot\partial\mathbf{A}^{-1}/\partial\mathbf{A}_{ij}$

$\equiv \mathbf{0}$, where $\boldsymbol{\Delta}_{ij}$ denotes a matrix with a one in row i and column j, zeros elsewhere. Then, $\partial\mathbf{x}'\mathbf{A}^{-1}\mathbf{y}/\partial\mathbf{A}_{ij} = -\,\mathbf{x}'\mathbf{A}^{-1}\boldsymbol{\Delta}_{ij}\mathbf{A}^{-1}\mathbf{y} = (\mathbf{A}^{-1}\mathbf{x})_i(\mathbf{A}^{-1}\mathbf{y})_j$. For (3), first note that $\partial\mathbf{A}_{ij}/\partial\mathbf{T}_{rs} = \delta_{ir}\mathbf{T}_{js}+\delta_{jr}\mathbf{T}_{is}$. Combine this with (2) to get

$$\partial\mathbf{x}'\mathbf{A}^{-1}\mathbf{y}/\partial\mathbf{T}_{rs} = \sum_j (\mathbf{A}^{-1}\mathbf{x})_i(\mathbf{A}^{-1}\mathbf{y})_j(\delta_{ir}\mathbf{T}_{js}+\delta_{jr}\mathbf{T}_{is})$$

$$= \sum_j (\mathbf{A}^{-1}\mathbf{x})_r(\mathbf{A}^{-1}\mathbf{y})_j\mathbf{T}_{js} + \sum_i (\mathbf{A}^{-1}\mathbf{x})_i(\mathbf{A}^{-1}\mathbf{y})_r\mathbf{T}_{is} = 2(\mathbf{A}^{-1}\mathbf{x}\mathbf{y}'\mathbf{A}^{-1}\mathbf{T})_{rs}.$$

2.14. UPDATING AND BACKDATING MATRIX OPERATIONS

Often in statistical applications, one needs to modify the calculation of a matrix inverse or other matrix operation to accommodate 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 $\mathbf{A}$ is n×n and nonsingular, and $\mathbf{A}^{-1}$ has been calculated, and if $\mathbf{B}$ and $\mathbf{C}$ are arrays that are n×k of rank k, then $(\mathbf{A}+\mathbf{B}\mathbf{C}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I}_k+\mathbf{C}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}'\mathbf{A}^{-1}$, provided $\mathbf{I}_k+\mathbf{C}'\mathbf{A}^{-1}\mathbf{B}$ is nonsingular. No matrix inversion is required if k = 1.

(2) If $\mathbf{A}$ is m×n with m $\ge$ n and $\rho(\mathbf{A})$ = n, so that it has a LDU factorization $\mathbf{A} = \mathbf{PLDU}'\mathbf{Q}'$ with $\mathbf{D}$ n×n diagonal, $\mathbf{P}$ and $\mathbf{Q}$ permutation matrices, $\mathbf{L}$ and $\mathbf{U}$ lower triangular, then the array $\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{bmatrix}$, with $\mathbf{B}$ k×n, has the LDU factorization $\begin{bmatrix} \boldsymbol{P} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_k \end{bmatrix}\begin{bmatrix} \boldsymbol{L} \\ \boldsymbol{C} \end{bmatrix}\boldsymbol{DU}'\boldsymbol{Q}'$, where $\mathbf{C} = \mathbf{BQU}'^{-1}\mathbf{D}^{-1}$.

(3) Suppose $\mathbf{A}$ is m×n of rank n, and $\mathbf{b} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$. Suppose $\mathbf{A}^* = \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix}$ and $\boldsymbol{y}^* = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{w} \end{bmatrix}$ with $\mathbf{C}$ k×n and $\mathbf{w}$ k×1, and $\mathbf{b}^* = (\mathbf{A}^{*\prime}\mathbf{A}^*)^{-1}\,\mathbf{A}^{*\prime}\mathbf{y}^*$. Then,

$$\mathbf{b}^* - \mathbf{b} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}'[\mathbf{I}_k+\mathbf{C}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{C}']^{-1}(\mathbf{w}-\mathbf{Cb}) = (\mathbf{A}^{*\prime}\mathbf{A}^*)^{-1}\mathbf{C}'[\mathbf{I}_k-\mathbf{C}(\mathbf{A}^{*\prime}\mathbf{A}^*)^{-1}\mathbf{C}']^{-1}(\mathbf{w}-\mathbf{Cb}^*).$$

One can verify (1) by multiplication. To show (2), use Corollary 5 of Theorem 2.1. To show (3), apply (1) to $\mathbf{A}^{*\prime}\mathbf{A}^* = \mathbf{A}'\mathbf{A} + \mathbf{C}'\mathbf{C}$, or to $\mathbf{A}'\mathbf{A} = \mathbf{A}^{*\prime}\mathbf{A}^* - \mathbf{C}'\mathbf{C}$, and use $\mathbf{A}^{*\prime}\mathbf{y}^* = \mathbf{Ay} + \mathbf{Cw}$.

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.