APPENDIX A

REVIEW OF MATRIX ALGEBRA

We present here a brief review of some concepts that are assumed as background for the text. Good references include Gantmacher (1977), Brogan (1974), and Strang (1980).

A.1 BASIC DEFINITIONS AND FACTS

The *determinant* of an $n \times n$ matrix is symbolized as $|A|$. If A and B are both square, then

$$|A| = |A^{\mathrm{T}}|, \tag{A.1-1}$$

$$|AB| = |A| \cdot |B|, \tag{A.1-2}$$

where the superscript T represents transpose. If $A \in C^{m \times n}$ and $B \in C^{n \times m}$ (where n can equal m), then

$$\mathrm{trace}(AB) = \mathrm{trace}(BA) \tag{A.1-3}$$

$$|I_m + AB| = |I_n + BA|. \tag{A.1-4}$$

(C represents the complex numbers.)
For any matrices A and B,

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}} A^{\mathrm{T}} \tag{A.1-5}$$

and if A and B are nonsingular, then

$$(AB)^{-1} = B^{-1} A^{-1}. \tag{A.1-6}$$

Lewis bapp01.tex V1 - 10/19/2011 5:31pm Page 519

The *Kronecker product* of two matrices $A = [a_{ij}] \in C^{m\times n}$ and $B = [b_{ij}] \in C^{p\times q}$ is

$$A \otimes B = [a_{ij}B] \in C^{mp\times nq}. \tag{A.1-7}$$

(It is sometimes defined as $A \otimes B = [Ab_{ij}]$.) If $A = [a_1 a_2 \cdots a_n]$, where a_i are the columns of A, the *stacking operator* is defined by

$$s(A) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}. \tag{A.1-8}$$

It converts $A \in C^{m\times n}$ into a vector $s(A) \in C^{mn}$. An identity that is often useful is

$$s(ABD) = (D^{\mathrm{T}} \otimes A)s(B). \tag{A.1-9}$$

If $A \in C^{m\times m}$ and $B \in C^{p\times p}$, then

$$|A \otimes B| = |A|^p \cdot |B|^m. \tag{A.1-10}$$

See Brewer (1978) for other results.

If λ_i is an eigenvalue of A with eigenvector v_i, then $1/\lambda_i$ is an eigenvalue of A^{-1} with the same eigenvector, for

$$Av_i = \lambda_i v_i \tag{A.1-11}$$

implies that

$$\lambda_i^{-1} v_i = A^{-1} v_i. \tag{A.1-12}$$

If λ_i is an eigenvalue of A with eigenvector ω_i, and μ_j is an eigenvalue of B with eigenvector w_j, then $\lambda_i \mu_j$ is an eigenvalue of $A \otimes B$ with eigenvector $v_i \otimes w_j$ (Brewer 1978).

A.2 PARTITIONED MATRICES

If

$$D = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{bmatrix}, \tag{A.2-1}$$

where A_{ij} are matrices, then we write $D = \mathrm{diag}(A_{11}, A_{22}, A_{33})$ and say that D *is block diagonal*. If the A_{ii} are square, then

$$|D| = |A_{11}| \cdot |A_{22}| \cdot |A_{33}|, \tag{A.2-2}$$

and if $|D| \neq 0$, then

$$D^{-1} = \mathrm{diag}(A_{11}^{-1}, A_{22}^{-1}, A_{33}^{-1}). \tag{A.2-3}$$

If

$$D = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix}, \tag{A.2-4}$$

where A_{ij} are matrices, then D is *upper block triangular* and (A.2-2) still holds. *Lower block triangular* matrices have the form of the transpose of (A.2-4).

If

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{A.2-5}$$

we define the *Schur complement of* A_{22} as

$$D_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12} \tag{A.2-6}$$

and the *Schur complement of* A_{11} as

$$D_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21}. \tag{A.2-7}$$

The inverse of A can be written

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}D_{22}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}D_{22}^{-1} \\ -D_{22}^{-1}A_{21}A_{11}^{-1} & D_{22}^{-1} \end{bmatrix}, \tag{A.2-8}$$

$$A^{-1} = \begin{bmatrix} D_{11}^{-1} & -D_{11}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}D_{11}^{-1} & A_{22}^{-1} + A_{22}^{-1}A_{21}D_{11}^{-1}A_{12}A_{22}^{-1} \end{bmatrix}, \tag{A.2-9}$$

or

$$A^{-1} = \begin{bmatrix} D_{11}^{-1} & -A_{11}^{-1}A_{12}D_{22}^{-1} \\ -A_{22}^{-1}A_{21}D_{11}^{-1} & D_{22}^{-1} \end{bmatrix}, \tag{A.2-10}$$

depending, of course, on whether $|A_{11}| \neq 0$, $|A_{22}| \neq 0$, or both. These can be verified by checking that $AA^{-1} = A^{-1}A = I$. By comparing these various forms, we obtain the well-known *matrix inversion lemma*

$$(A_{11}^{-1} + A_{12}A_{22}A_{21})^{-1} = A_{11} - A_{11}A_{12}(A_{21}A_{11}A_{12} + A_{22}^{-1})^{-1}A_{21}A_{11}. \tag{A.2-11}$$

The Schur complement arises naturally in the solution of linear simultaneous equations, for if

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 0 \\ Z \end{bmatrix}, \tag{A.2-12}$$

then from the first equation

$$X = -A_{11}^{-1}A_{12}Y,$$

Lewis bapp01.tex V1 - 10/19/2011 5:31pm Page 521

and using this in the second equation yields

$$(A_{22} - A_{21}A_{11}^{-1}A_{12})Y = Z. \tag{A.2-13}$$

If A is given by (A.2-5), then

$$|A| = |A_{11}| \cdot |A_{22} - A_{21}A_{11}^{-1}A_{12}| = |A_{22}| \cdot |A_{11} - A_{12}A_{22}^{-1}A_{21}|. \tag{A.2-14}$$

Therefore, the determinant of A is the product of the determinant of A_{11} (or A_{22}) and the determinant of the Schur complement of A_{22} (or A_{11}).

A.3 QUADRATIC FORMS AND DEFINITENESS

If $x \in R^n$ is a vector, then the square of the Euclidean norm is

$$\|x\|^2 = x^{\mathrm{T}}x. \tag{A.3-1}$$

If S is any nonsingular transformation, the vector Sx has a norm squared of $(Sx)^{\mathrm{T}}Sx = x^{\mathrm{T}}S^{\mathrm{T}}Sx$. Letting $P = S^{\mathrm{T}}S$, we write

$$\|x\|_P^2 = x^{\mathrm{T}}Px \tag{A.3-2}$$

as the norm squared of Sx. We call $||x||_P$ the *norm of* x with respect to P. We call

$$x^{\mathrm{T}}Qx \tag{A.3-3}$$

a *quadratic form*. We shall assume Q is real.

Every real square matrix Q can be decomposed into a *symmetric part* Q_{s} (i.e., $Q_{\mathrm{s}}^{\mathrm{T}} = Q_{\mathrm{s}}$) and an *antisymmetric part* Q_{a} (i.e., $Q_{\mathrm{a}}^{\mathrm{T}} = -Q_{\mathrm{a}}$):

$$Q = Q_{\mathrm{s}} + Q_{\mathrm{a}}, \tag{A.3-4}$$

where

$$Q_{\mathrm{s}} = (Q + Q^{\mathrm{T}})/2, \tag{A.3-5}$$

$$Q_{\mathrm{a}} = (Q - Q^{\mathrm{T}})/2. \tag{A.3-6}$$

If the quadratic form $x^{\mathrm{T}}Ax$ has A antisymmetric, then it must be equal to zero since $x^{\mathrm{T}}Ax$ is a scalar, so that $x^{\mathrm{T}}Ax = (x^{\mathrm{T}}Ax)^{\mathrm{T}} = x^{\mathrm{T}}A^{\mathrm{T}}x = -x^{\mathrm{T}}Ax$. For a general real square Q, then

$$x^{\mathrm{T}}Qx = x^{\mathrm{T}}(Q_{\mathrm{s}} + Q_{\mathrm{a}})x = x^{\mathrm{T}}Q_{\mathrm{s}}x. \tag{A.3-7}$$

We can therefore assume without loss of generality that Q in (A.3-3) is symmetric. Let us do so.

We say Q is:

Positive definite ($Q > 0$) if $x_T Qx > 0$ for all nonzero x.

Lewis bapp01.tex V1 - 10/19/2011 5:31pm Page 522

Positive semi-definite ($Q \geq 0$) if $x^{\mathrm{T}}Qx \geq 0$ for all nonzero x.

Negative semi-definite ($Q \leq 0$) if $x^{\mathrm{T}}Qx \leq 0$ for all nonzero x.

Negative definite($Q < 0$) if $x^{\mathrm{T}}Qx < 0$ for all nonzero x.

Indefinite if $x^{\mathrm{T}}Qx > 0$ for some x, $x^{\mathrm{T}}Qx < 0$ for other x.

We can test for definiteness independently of the vectors x. If λ_i are the eigenvalues of Q, then

$$\begin{aligned} Q > 0 \quad &\text{if all} \quad \lambda_i > 0, \\ Q \geq 0 \quad &\text{if all} \quad \lambda_i \geq 0, \\ Q \leq 0 \quad &\text{if all} \quad \lambda_i \leq 0, \\ Q < 0 \quad &\text{if all} \quad \lambda_i < 0. \end{aligned} \tag{A.3-8}$$

Another test is provided as follows. Let $Q = [q_{ij}] \in R^{n\times n}$. The *leading minors* or Q are

$$\begin{aligned} m_1 &= q_{11}, \\ m_2 &= \begin{vmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{vmatrix}, \\ m_3 &= \begin{vmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{vmatrix}, \ldots, \\ m_n &= |Q|. \end{aligned} \tag{A.3-9}$$

In terms of the minors, we have

$$\begin{aligned} Q > 0 \quad &\text{if } m_i > 0, \text{ all } i, \\ Q \geq 0 \quad &\text{if } \textit{all} \text{ principal minor not only leading minors)} \\ &\quad \text{are nonnegative.} \\ Q \leq 0 \quad &\text{if} - Q \geq 0, \\ Q < 0 \quad &\text{if} \begin{cases} m_i < 0, & \text{all odd } i \\ m_i > 0, & \text{all even } i \end{cases} \end{aligned} \tag{A.3-10}$$

Any positive semidefinite matrix Q can be factored into *square roots* either as

$$Q = \sqrt{Q}\sqrt{Q}^{\mathrm{T}} \tag{A.3-11}$$

or as

$$Q = \sqrt{Q}^{\mathrm{T}}\sqrt{Q}. \tag{A.3-12}$$

The ("left" and "right") square roots in (A.3-11) and (A.3-12) are not in general the same. Indeed, Q may have several roots since each of these factorizations is not even unique. If $Q > 0$, then all square roots are nonsingular.

Lewis bapp01.tex V1 - 10/19/2011 5:31pm Page 523

If $P > 0$, then (A.3-2) is a norm. If $P \geq 0$, it is called a *seminorm* since $x^{\mathrm{T}}Px$ may be zero even if x is not.

A.4 MATRIX CALCULUS

Let $x \in C^n = [x_1\, x_2 \cdots x_n]^{\mathrm{T}}$ be a vector, $s \in C$ be a scalar, and $f(x) \in C^m$ be an m-vector function of x. The differential in x is

$$dx = \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}, \tag{A.4-1}$$

and the derivative of x with respect to s (which could be time) is

$$\frac{dx}{ds} = \begin{bmatrix} dx_1/ds \\ dx_2/ds \\ \vdots \\ dx_n/ds \end{bmatrix}. \tag{A.4-2}$$

If s is a function of x. Then the *gradient* of s with respect to x is the *column* vector

$$s_x \stackrel{\Delta}{=} \frac{\partial s}{\partial x} = \begin{bmatrix} \partial s/\partial x_1 \\ \partial s/\partial x_2 \\ \vdots \\ \partial s/\partial x_n \end{bmatrix}. \tag{A.4-3}$$

(The gradient is defined as a row vector in some references.) Then the total differential in s is

$$ds = \left(\frac{\partial s}{\partial x}\right)^{\mathrm{T}} dx = \sum_{i=1}^{n} \frac{\partial s}{\partial x_i} dx_i. \tag{A.4-4}$$

If s is a function of two vectors x and y, then

$$ds = \left(\frac{\partial s}{\partial x}\right)^{\mathrm{T}} dx + \left(\frac{\partial s}{\partial y}\right)^{\mathrm{T}} dy. \tag{A.4-5}$$

The *Hessian* of s with respect to x is the second derivative

$$s_{xx} \stackrel{\Delta}{=} \frac{\partial^2 s}{\partial x^2} = \left[\frac{\partial^2 s}{\partial x_i \partial x_j}\right], \tag{A.4-6}$$

which is a symmetric $n \times n$ matrix. In terms of the gradient and the Hessian, the *Taylor series expansion* of *s(x)* about x_0 is

$$s(x) = s(x_0) + \left(\frac{\partial s}{\partial x}\right)^{\mathrm{T}} (x - x_0) + \frac{1}{2}(x - x_0)^{\mathrm{T}} \frac{\partial^2 s}{\partial x^2}(x - x_0) + O(3), \tag{A.4-7}$$

where $O(3)$ represents terms of order 3, and s_x and s_{xx} are evaluated at x_0.

The *Jacobian* of f with respect to x is the $m \times n$ matrix

$$f_x \triangleq \frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n} \right], \tag{A.4-8}$$

so that the total differential of f is

$$df = \frac{\partial f}{\partial x} dx = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i. \tag{A.4-9}$$

We shall use the shorthand notation

$$\frac{\partial f^{\mathrm{T}}}{\partial x} \triangleq \left(\frac{\partial f}{\partial x} \right)^{\mathrm{T}} \in C^{n \times m}. \tag{A.4-10}$$

If y is a vector and A, B, D, Q are matrices, all with dimensions so that the following expressions make sense, then we have the following results:

$$\frac{d}{dt}(A^{-1}) = -A^{-1} \dot{A} A^{-1}. \tag{A.4-11}$$

Some useful gradients are

$$\frac{\partial}{\partial x}(y^{\mathrm{T}} \mathrm{x}) = \frac{\partial}{\partial x}(x^{\mathrm{T}} y) = y, \tag{A.4-12}$$

$$\frac{\partial}{\partial x}(y^{\mathrm{T}} A x) = \frac{\partial}{\partial x}(x^{\mathrm{T}} A^{\mathrm{T}} y) = A^{\mathrm{T}} y, \tag{A.4-13}$$

$$\frac{\partial}{\partial x}(y^{\mathrm{T}} f(x)) = \frac{\partial}{\partial x}(f^{\mathrm{T}}(x) y) = f_x^{\mathrm{T}} y, \tag{A.4-14}$$

$$\frac{\partial}{\partial x}(x^{\mathrm{T}} A x) = A x + A^{\mathrm{T}} x, \tag{A.4-15}$$

and if Q is symmetric, then

$$\frac{\partial}{\partial x}(x^{\mathrm{T}} Q x) = 2 Q x, \tag{A.4-16}$$

$$\frac{\partial}{\partial x}(x - y)^{\mathrm{T}} Q (x - y) = 2 Q (x - y). \tag{A.4-17}$$

The chain rule for two vector functions becomes

$$\frac{\partial}{\partial x}(f^{\mathrm{T}} y) = f_x^{\mathrm{T}} y + y_x^{\mathrm{T}} f. \tag{A.4-18}$$

Some useful Hessians are

$$\frac{\partial^2 x^{\mathrm{T}} A x}{\partial x^2} = A + A^{\mathrm{T}}, \tag{A.4-19}$$

and if Q is symmetric

$$\frac{\partial^2 x^{\mathrm{T}} Q x}{\partial x^2} = 2Q, \tag{A.4-20}$$

$$\frac{\partial^2}{\partial x^2}(x-y)^{\mathrm{T}} Q(x-y) = 2Q. \tag{A.4-21}$$

Some useful Jacobians are

$$\frac{\partial}{\partial x}(Ax) = A \tag{A.4-22}$$

(contrast this with (A.4-12)), and the chain rule

$$\frac{\partial}{\partial x}(sf) = \frac{\partial}{\partial x}(fs) = sf_x + fs_x^{\mathrm{T}} \tag{A.4-23}$$

(contrast this with (A.4-18)).

Some useful derivatives involving the trace and determinant are

$$\frac{\partial}{\partial A}\mathrm{trace}(A) = I, \tag{A.4-24}$$

$$\frac{\partial}{\partial A}\mathrm{trace}(BAD) = B^{\mathrm{T}} D^{\mathrm{T}}, \tag{A.4-25}$$

$$\frac{\partial}{\partial A}\mathrm{trace}(ABA^{\mathrm{T}}) = 2AB, \text{ if } B = B^{\mathrm{T}} \tag{A.4-26}$$

$$\frac{\partial}{\partial A}|BAD| = |BAD|A^{-\mathrm{T}}, \tag{A.4-27}$$

where $A^{-\mathrm{T}} \triangleq (A^{-1})^{\mathrm{T}}$.

A.5 THE GENERALIZED EIGENVALUE PROBLEM

Consider the generalized eigenvalue problem

$$Gz = \mu F z, \tag{A.5-1}$$

where

$$\det(\mu F - G) \equiv 0. \tag{A.5-2}$$

Then the finite generalized eigenvalues are the roots of $\det(\mu F - G)$. Let μ_i be the roots of $\det(\mu F - G)$ and define

$$\eta_i = \mathrm{dimker}(\mu_i F - G). \tag{A.5-3}$$

Then the rank 1 finite generalized eigenvectors are defined by

$$(\mu_i F - G) z_{ij}^1 = 0,\ j \in \hat{\eta}_i \tag{A.5-4}$$

(where $\hat{\eta}_i = \{1, 2, \ldots, \eta_i\}$) and the rank k finite eigenvectors for $k > 1$ and each i and j by

$$(\mu_i F - G) z_{ij}^{k+1} = -F z_{ij}^{k}, k \geq 1. \tag{A.5-5}$$

If F is nonsingular, the above equation can be used to solve recursively for the z_{ij}^{k} beginning with the highest rank eigenvector in each chain. In that case this construction provides the eigenstructure of $F^{-1}G$. In the case where F in singular, the above equation cannot generally be used to recursively generate the z_{ij}^{k}. Furthermore, there exist eigenvalues at infinity and corresponding eigenvectors that can be constructed as follows. Define $= \dim \ker(F)$. Then the rank 1 infinite eigenvectors are defined by

$$F z_{\infty j}^{1} = 0, j = \hat{\eta} \tag{A.5-6}$$

and the rank k infinite eigenvectors for $k > 1$ and each j by

$$F z_{\infty j}^{k+1} = G z_{\infty j}^{k}, k \geq 1. \tag{A.5-7}$$

By arranging the eigenvectors as the columns of two nonsingular matrices according to

$$Z = [z_{ij}^{k} | z_{\infty j}^{k}], W = [F z_{ij}^{k} | G z_{\infty j}^{k}] \tag{A.5-8}$$

with *i, j, k* incrementing in odometer order, then

$$W^{-1} F V = \begin{bmatrix} I & 0 \\ 0 & N \end{bmatrix}, W^{-1} G V = \begin{bmatrix} M & 0 \\ 0 & I \end{bmatrix}, \tag{A.5-9}$$

where M is a Jordan form matrix containing the finite generalized eigenvalues of (G, F) and N is a nilpotent Jordan matrix representing the infinite generalized eigenvalues. The above canonical form is also known as the *Weierstrass form*.