APPENDIX A

SUMMARY OF VECTOR/MATRIX OPERATIONS

This Appendix summarizes properties of vector and matrices, and vector/matrix operations that are often used in estimation. Further information may be found in most books on estimation or linear algebra; for example, Golub and Van Loan (1996), DeRusso et al. (1965), and Stewart (1988).

A.1 DEFINITION

A.1.1 Vectors

A *vector* is a linear collection of elements. We use a lower case bold letter to denote vectors, which by default are assumed to be column vectors. For example,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

is a three-element column vector. A row vector is (obviously) defined with elements in a row; for example,

$$\mathbf{a} = [a_1 \quad a_2 \quad a_3].$$

A vector is called *unit* or *normalized* when the sum of elements squared is equal to 1: $\sum_{i=1}^{n} a_i^2 = 1$ for an n-element vector $\mathbf{a}$.

A.1.2 Matrices

A *matrix* is a two-dimensional collection of elements. We use bold upper case letters to denote matrices. For example,

Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook, by Bruce P. Gibbs

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

is a matrix with two rows and three columns, or a 2×3 matrix. Individual elements are labeled with the first subscript indicating the row, and the second indicating the column. An n-element column vector may be considered an $n \times 1$ matrix, and an n-element row vector may be considered a $1 \times n$ matrix.

A.1.2.1 Symmetric Matrix A square matrix is symmetric if elements with interchanged subscripts are equal: $A_{ji} = A_{ij}$. For example,

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 9 \\ 5 & 1 & 3 \\ 9 & 3 & 4 \end{bmatrix}$$

is symmetric.

A.1.2.2 Hermitian Matrix A square complex matrix is *Hermitian* if elements with interchanged subscripts are equal to the complex conjugate of each other: $A_{ji} = A_{ij}^*$.

A.1.2.3 Toeplitz Matrix A square matrix is *Toeplitz* if all elements along the upper left to lower right diagonals are equal: $A_{i,j} = A_{i-1,j-1}$. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 & 4 \\ 3 & 1 & 2 & -1 \\ 5 & 3 & 1 & 2 \\ -2 & 5 & 3 & 1 \end{bmatrix}$$

is Toeplitz.

A.1.2.4 Identity Matrix A square matrix that is all zero except for ones along the main diagonal is the identity matrix, denoted as **I**. For example,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is a 4×4 identity matrix. Often a subscript is added to indicate the dimension, as $\mathbf{I}_n$ is an $n \times n$ identity matrix.

A.1.2.5 Triangular Matrix All elements of a lower triangular matrix above the main diagonal are zero. All elements of an upper triangular matrix below the main diagonal are zero. For example,

$$\begin{bmatrix} 1 & 7 & 4 & -4 \\ 0 & 2 & 6 & 1 \\ 0 & 0 & 3 & 9 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

is upper triangular.

A.2 ELEMENTARY VECTOR/MATRIX OPERATIONS

A.2.1 Transpose

The *transpose* of a matrix, denoted with superscript T, is formed by interchanging row and column elements: $\mathbf{B} = \mathbf{A}^T$ is the transpose of $\mathbf{A}$ where $B_{ji} = A_{ij}$. For example,

$$\mathbf{B} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}^T = \begin{bmatrix} A_{11} & A_{21} \\ A_{12} & A_{22} \\ A_{13} & A_{23} \end{bmatrix}.$$

If the matrix is complex, the complex conjugate transpose is denoted as $\mathbf{A}^H = (\mathbf{A}^*)^T$. The matrix is Hermitian if $\mathbf{A} = \mathbf{A}^H$.

A.2.2 Addition

Two or more vectors or matrices of the same dimensions may be added or subtracted by adding/subtracting individual elements. For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & 7 & 5 \\ 2 & 1 & -2 \end{bmatrix}$$

then

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 4 & 9 & 8 \\ 6 & 6 & 4 \end{bmatrix}.$$

Matrix addition is commutative; that is, $\mathbf{C} = \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

A.2.3 Inner (Dot) Product of Vectors

The *dot product* or *inner product* of two vectors of equal size is the sum of the products of corresponding elements. If vectors $\mathbf{a}$ and $\mathbf{b}$ both contain m elements, the dot product is a scalar:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a} \bullet \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^{m} a_i b_i.$$

If $\langle \mathbf{a}, \mathbf{b} \rangle = 0$, the vectors are *orthogonal*.

A.2.4 Outer Product of Vectors

The *outer product* of two vectors (of possibly unequal sizes) is a matrix of products of corresponding vector elements. If vectors **a** and **b** contain m- and n-elements, respectively, then the outer product is an $m \times n$ matrix:

$$\mathbf{C} = \mathbf{a} \otimes \mathbf{b} = \mathbf{a}\mathbf{b}^T$$

or $C_{ij} = a_i b_j$.

A.2.5 Multiplication

Two matrices, where the column dimension of the first (m) is equal to the row dimension of the second, may be multiplied by forming the dot product of the rows of the first matrix and the columns of the second; that is, $C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}$. For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & 2 \\ 0 & 1 \\ 5 & -2 \end{bmatrix}$$

then

$$\mathbf{C} = \mathbf{A}\mathbf{B} = \begin{bmatrix} 18 & -2 \\ 42 & 1 \end{bmatrix}.$$

Matrix multiplication is not commutative; that is, $\mathbf{C} = \mathbf{AB} \neq \mathbf{BA}$. A matrix multiplying or multiplied by the identity $\mathbf{I}$ is unchanged; that is, $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$.

The transpose of the product of two matrices is the reversed product of the transposes: $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

Vector-matrix multiplication is defined as for matrix-matrix multiplication. If matrix $\mathbf{A}$ is $m \times n$ and vector $\mathbf{x}$ has m-elements, $\mathbf{y} = \mathbf{x}^T\mathbf{A}$ or

$$y_j = \sum_{i=1}^{m} x_i A_{ij} \quad \text{for } j = 1, 2, \ldots, n$$

is an n-element row vector. If vector $\mathbf{x}$ has n elements, $\mathbf{y} = \mathbf{Ax}$ is an m-element column vector.

A.3 MATRIX FUNCTIONS

A.3.1 Matrix Inverse

A square matrix that multiplies another square matrix to produce the identity matrix is called the inverse, and is denoted by a superscript −1; that is, if $\mathbf{B} = \mathbf{A}^{-1}$, then $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. Just as scalar division by zero is not defined, a matrix is called *indeterminate* if the inverse does not exist. The matrix inverse may be computed by various methods. Two popular methods for general square matrices are Gauss-Jordan elimination with pivoting, and LU decomposition followed by inversion of the LU factors (see Press et al. 2007, chapter 2). Inversion based on cofactors and determinants (explained below) is also used for small matrices. For symmetric matrices, inversion based on Cholesky factorization is recommended.

The inverse of the product of two matrices is the reversed product of the inverses: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Nonsquare matrices generally do not have an inverse, but left or right inverses can be defined; for example, for $m \times n$ matrix $\mathbf{A}$, $((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{A} = \mathbf{I}_n$, so $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is a left inverse provided that $(\mathbf{A}^T\mathbf{A})^{-1}$ exists, and $\mathbf{A}(\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}) = \mathbf{I}_m$ so $\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$ is a right inverse provided that $(\mathbf{A}\mathbf{A}^T)^{-1}$ exists.

A square matrix is called *orthogonal* when $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$. Thus the transpose is also the inverse: $\mathbf{A}^{-1} = \mathbf{A}^T$. If rectangular matrix $\mathbf{A}$ is $m \times n$, it is called *column orthogonal* when $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ since the columns are orthonormal. This is only possible when $m \geq n$. If $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ for $m \leq n$, matrix $\mathbf{A}$ is called *row orthogonal* because the rows are orthonormal.

A square symmetric matrix must be positive definite for it to be invertible. A *symmetric positive definite* matrix is a square symmetric matrix for which $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all nonzero vectors $\mathbf{x}$. A symmetric *positive semi-definite* or *non-negative definite* matrix is one for which $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$.

A.3.2 Partitioned Matrix Inversion

It is often helpful to compute the inverse of a matrix in partitions. For example, consider the inverse of

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

where the four bold letters indicate smaller matrices. We express the inverse as

$$\begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}$$

and write

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

or

$$\mathbf{AE} + \mathbf{BG} = \mathbf{I} \tag{A3-1}$$

$$\mathbf{AF} + \mathbf{BH} = \mathbf{0} \tag{A3-2}$$

$$\mathbf{CE} + \mathbf{DG} = \mathbf{0} \tag{A3-3}$$

$$\mathbf{CF} + \mathbf{DH} = \mathbf{I}. \tag{A3-4}$$

Using equations (A3-2), (A3-4), (A3-1), and (A3-3) in that order, we obtain:

$$\begin{aligned} \mathbf{F} &= -\mathbf{A}^{-1}\mathbf{BH} \\ \mathbf{H} &= (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ \mathbf{E} &= \mathbf{A}^{-1}(\mathbf{I} - \mathbf{BG}) \quad \text{(intermediate)} \\ \mathbf{G} &= -(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \\ &= -\mathbf{HCA}^{-1} \\ \mathbf{E} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{BHCA}^{-1} \end{aligned} \tag{A3-5}$$

Alternately, using (A3-3), (A3-1), (A3-4), and (A3-2), we obtain

$$\begin{aligned}
\mathbf{G} &= -\mathbf{D}^{-1}\mathbf{C}\mathbf{E} \\
\mathbf{E} &= (\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \\
\mathbf{H} &= \mathbf{D}^{-1}(\mathbf{I}-\mathbf{C}\mathbf{F}) \\
\mathbf{F} &= -(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\
&= -\mathbf{E}\mathbf{B}\mathbf{D}^{-1} \\
\mathbf{H} &= \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}\mathbf{E}\mathbf{B}\mathbf{D}^{-1}
\end{aligned} . \tag{A3-6}$$

Thus the partitioned inverse can be written in two forms:

$$\boxed{\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\\ -(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\end{bmatrix}} \tag{A3-7}$$

or

$$\boxed{\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}\\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}\end{bmatrix}.} \tag{A3-8}$$

If the matrix to be inverted is symmetric:

$$\boxed{\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^T & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\\ -(\mathbf{D}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{A}^{-1} & (\mathbf{D}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}\end{bmatrix}} \tag{A3-9}$$

or

$$\boxed{\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^T & \mathbf{D}\end{bmatrix}^{-1} = \begin{bmatrix}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{D}^{-1}\\ -\mathbf{D}^{-1}\mathbf{B}^T(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)^{-1} & \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{B}^T(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{D}^{-1}\end{bmatrix}} \tag{A3-10}$$

where **A** and **D** are also symmetric.

A.3.3 Matrix Inversion Identity

The two equivalent expressions for the partitioned inverse suggest a matrix equivalency that is the link between batch least squares and recursive least squares (and Kalman filtering). This formula and variations on it have been attributed to various people (e.g., Woodbury, Ho, Sherman, Morrison), but the relationship has undoubtedly been discovered and rediscovered many times.

From the lower right corner of equations (A3-7) and (A3-8):

$$\boxed{(\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}.} \tag{A3-11}$$

In the symmetric case when $\mathbf{C} = \mathbf{B}^T$:

$$\boxed{(\mathbf{D}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1} = \mathbf{D}^{-1}+\mathbf{D}^{-1}\mathbf{B}^T(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{D}^{-1},} \tag{A3-12}$$

or by changing the sign of **A**:

$$\boxed{(\mathbf{D}+\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}=\mathbf{D}^{-1}-\mathbf{D}^{-1}\mathbf{B}^T(\mathbf{A}+\mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{D}^{-1}.} \tag{A3-13}$$

Equation (A3-13) is the connection between the measurement update of Bayesian least squares and the Kalman filter.

If $\mathbf{D} = \mathbf{I}$ and $\mathbf{C} = \mathbf{B}^T$:

$$\boxed{(\mathbf{I}-\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}=\mathbf{I}+\mathbf{B}^T(\mathbf{A}-\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B}.} \tag{A3-14}$$

or

$$\boxed{(\mathbf{I}+\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B})^{-1}=\mathbf{I}-\mathbf{B}^T(\mathbf{A}+\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{B}.} \tag{A3-15}$$

If **b** is a row vector and a is scalar:

$$(\mathbf{D}-\mathbf{b}^T\mathbf{b}/a)^{-1}=\mathbf{D}^{-1}+(\mathbf{D}^{-1}\mathbf{b}^T)(\mathbf{b}\mathbf{D}^{-1})/(a-\mathbf{b}\mathbf{D}^{-1}\mathbf{b}^T) \tag{A3-16}$$

or

$$(\mathbf{D}+\mathbf{b}^T\mathbf{b}/a)^{-1}=\mathbf{D}^{-1}-(\mathbf{D}^{-1}\mathbf{b}^T)(\mathbf{b}\mathbf{D}^{-1})/(a+\mathbf{b}\mathbf{D}^{-1}\mathbf{b}^T). \tag{A3-17}$$

A.3.4 Determinant

The determinant of a square matrix is a measure of scale change when the matrix is viewed as a linear transformation. When the determinant of a matrix is zero, the matrix is *indeterminate* or *singular*, and cannot be inverted. The *rank* of matrix $|\mathbf{A}|$ is the largest square array in **A** that has nonzero determinant.

The determinant of matrix **A** is denoted as det(**A**) or $|\mathbf{A}|$. Laplace's method for computing determinants uses *cofactors*, where a cofactor of a given matrix element ij is $C_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|$ and $\mathbf{M}_{ij}$, called the *minor* of ij, is the matrix formed by deleting the i row and j column of matrix **A**. For 2×2 matrix

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

the cofactors are

$$C_{11} = A_{22}, \quad C_{12} = -A_{21}, \quad C_{21} = -A_{12}, \quad C_{22} = A_{11}.$$

The determinant is the sum of the products of matrix elements and cofactors for any row or column. Thus

$$\begin{aligned}|\mathbf{A}| &= A_{11}C_{11}+A_{12}C_{12} = A_{11}C_{11}+A_{21}C_{21} = A_{21}C_{21}+A_{22}C_{22} = A_{12}C_{12}+A_{22}C_{22} \\ &= A_{11}A_{22}-A_{12}A_{21}\end{aligned}.$$

For a 3×3 matrix **A**,

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix},$$

the cofactors are

$$C_{11} = A_{22}A_{33} - A_{23}A_{32}, \quad C_{21} = -(A_{12}A_{33} - A_{13}A_{32}), \quad C_{31} = A_{12}A_{23} - A_{13}A_{22}, \quad \ldots$$

so using the first column,

$$\begin{aligned} |\mathbf{A}| &= A_{11}C_{11} + A_{21}C_{21} + A_{31}C_{31} \\ &= A_{11}(A_{22}A_{33} - A_{23}A_{32}) - A_{21}(A_{12}A_{33} - A_{13}A_{32}) + A_{31}(A_{12}A_{23} - A_{13}A_{22}) \end{aligned}.$$

A matrix inverse can be computed from the determinant and cofactors as

$$\boxed{\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}[\mathbf{C}]^T = \frac{1}{|\mathbf{A}|}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix},} \tag{A3-18}$$

where $[\mathbf{C}]$ is the matrix of cofactors for $\mathbf{A}$.

Computation of determinants using cofactors is cumbersome, and is seldom used for dimensions greater than three. The determinant is more easily computed by factoring $\mathbf{A} = \mathbf{LU}$ using Crout reduction, where $\mathbf{L}$ is unit lower triangular and $\mathbf{U}$ is upper triangular. Since the determinant of the product of matrices is equal to the product of determinants,

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|, \tag{A3-19}$$

we have $|\mathbf{A}| = |\mathbf{L}||\mathbf{U}|$, which is simply equal to the product of the diagonals of $\mathbf{U}$ because $|\mathbf{L}| = 1$. A number of other methods may also be used to compute determinants. Often the determinant is a bi-product of matrix inversion algorithms.

A.3.5 Matrix Trace

The *trace* of a square matrix is the sum of the diagonal elements. This property is often useful in least-squares or minimum variance estimation, as the sum of squared elements in an n-vector can be written as

$$\boxed{\sum_{i=1}^{n} a_i^2 = \mathbf{a}^T\mathbf{a} = tr\left[\mathbf{a}\mathbf{a}^T\right].} \tag{A3-20}$$

This rearrangement of vector order often allows solutions for minimum variance problems: it has been used repeatedly in previous chapters.

Since the trace only involves diagonal elements, $tr(\mathbf{A}^T) = tr(\mathbf{A})$. Also,

$$\boxed{tr(\mathbf{A}+\mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})} \tag{A3-21}$$

and for scalar c,

$$\boxed{tr(c\mathbf{A}) = c \cdot tr(\mathbf{A}).} \tag{A3-22}$$

Unlike the determinant, the trace of the matrix products is not the product of traces. If $\mathbf{A}$ is an $n \times m$ matrix and $\mathbf{B}$ is an $m \times n$ matrix,

$$\boxed{\begin{aligned} tr(\mathbf{AB}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}B_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{n} B_{ji}A_{ij} \\ &= tr(\mathbf{BA}) \end{aligned}.} \tag{A3-23}$$

However, this commutative property only works for pairs of matrices, or interchange of "halves" of the matrix product:

$$\boxed{\begin{aligned} tr(\mathbf{ABC}) &= tr(\mathbf{C}(\mathbf{AB})) = tr(\mathbf{BCA}) \\ &\neq tr(\mathbf{ACB}) \\ &\neq tr(\mathbf{CBA}) \end{aligned}.} \tag{A3-24}$$

When the three individual matrices are square and symmetric, any permutation works:

$$\boxed{tr(\mathbf{ABC}) = tr(\mathbf{C}^T\mathbf{B}^T\mathbf{A}^T) = tr(\mathbf{CBA}).} \tag{A3-25}$$

This permutation does not work with four or more symmetric matrices.

Permutation can be used to express the weighted quadratic form $\mathbf{a}^T\mathbf{Wa}$ as

$$\boxed{\begin{aligned} \mathbf{a}^T\mathbf{Wa} &= \mathbf{a}^T\mathbf{W}^{T/2}\mathbf{W}^{1/2}\mathbf{a} = tr(\mathbf{W}^{1/2}\mathbf{aa}^T\mathbf{W}^{T/2}) = tr(\mathbf{W}^{T/2}\mathbf{W}^{1/2}\mathbf{aa}^T) \\ &= tr\left(\mathbf{W}(\mathbf{aa}^T)\right) \end{aligned},} \tag{A3-26}$$

where $\mathbf{a}$ is a vector and matrix $\mathbf{W} = \mathbf{W}^{T/2}\mathbf{W}^{1/2}$ is symmetric.

For a transformation of the form $\mathbf{T}^{-1}\mathbf{AT}$ (called a similarity transformation), the trace is unchanged:

$$\boxed{tr(\mathbf{T}^{-1}\mathbf{AT}) = tr(\mathbf{TT}^{-1}\mathbf{A}) = tr(\mathbf{A}).} \tag{A3-27}$$

A.3.6 Derivatives of Matrix Functions

The derivative of matrix $\mathbf{A}$ with respect to scalar variable t is $\partial\mathbf{A}/\partial t$, which is a matrix with the same dimensions as $\mathbf{A}$. The derivative of matrix product $\mathbf{BA}$ with respect to t can be written as

$$\boxed{\frac{\partial(\mathbf{BA})}{\partial t} = \frac{\partial(\mathbf{BA})}{\partial\mathbf{A}}\frac{\partial\mathbf{A}}{\partial t},} \tag{A3-28}$$

where $\partial(\mathbf{BA})/\partial\mathbf{A}$ is a four-dimensional variable:

$$\frac{\partial(\mathbf{BA})}{\partial A_{11}}, \frac{\partial(\mathbf{BA})}{\partial A_{12}}, \frac{\partial(\mathbf{BA})}{\partial A_{13}}, \ldots.$$

Thus matrix derivatives for a function of $\mathbf{A}$, $\mathbf{C}(\mathbf{A})$, can be written using $\partial\mathbf{C}/\partial\mathbf{A}$ if the four dimensions are properly handled.

The derivative of the inverse of a matrix is obtained by differentiating $\mathbf{AB} = \mathbf{I}$, which gives $(d\mathbf{A})\mathbf{B} + \mathbf{A}(d\mathbf{B}) = \mathbf{0}$. Thus

$$d(\mathbf{A}^{-1}) = d\mathbf{B} = -\mathbf{A}^{-1}(d\mathbf{A})\mathbf{A}^{-1}$$

is two-dimensional, but has perturbations with respect to all elements of $d\mathbf{A}$; that is, each i, j element $(d\mathbf{B})_{ij}$ is a matrix of derivatives for all perturbations $d\mathbf{A}$. Thus

$$\boxed{\frac{\partial(\mathbf{A}^{-1})}{\partial t} = -\mathbf{A}^{-1}\left(\frac{\partial \mathbf{A}}{\partial t}\right)\mathbf{A}^{-1}.} \tag{A3-29}$$

The derivative of the trace of matrix products is computed by examining the derivative with respect to an individual element:

$$\frac{\partial tr(\mathbf{AB})}{\partial B_{ij}} = \frac{\partial}{\partial B_{ij}}\left(\sum_{k=1}^{m}\sum_{l=1}^{m} A_{kl}B_{lk}\right) = A_{ji}.$$

Thus

$$\boxed{\frac{\partial tr(\mathbf{AB})}{\partial \mathbf{B}} = \mathbf{A}^T.} \tag{A3-30}$$

For products of three matrices,

$$\boxed{\frac{\partial tr(\mathbf{ABC})}{\partial \mathbf{B}} = \frac{\partial tr(\mathbf{CAB})}{\partial \mathbf{B}} = (\mathbf{CA})^T.} \tag{A3-31}$$

The derivative of the determinant is computed by rearranging equation (A3-18) as

$$|\mathbf{A}|\mathbf{I} = \mathbf{A}[\mathbf{C}]^T. \tag{A3-32}$$

Thus for any diagonal element $i = 1, 2, \ldots, n$,

$$|\mathbf{A}| = \sum_{k=1}^{n} A_{ik}C_{ik}.$$

The partial derivative of $|\mathbf{A}|$ with respect to any $\mathbf{A}$ element in row i is

$$\frac{\partial|\mathbf{A}|}{\partial A_{ij}} = \frac{\partial}{\partial A_{ij}}\left(\sum_{k=1}^{n} A_{ik}C_{ik}\right) = C_{ij},$$

where $\partial C_{ik}/\partial A_{ij} = 0$ from cofactor definitions. Hence

$$\boxed{\frac{\partial|\mathbf{A}|}{\partial \mathbf{A}} = [\mathbf{C}] = |\mathbf{A}|\mathbf{A}^{-T}.} \tag{A3-33}$$

By a similar development

$$\boxed{\frac{\partial|\mathbf{A}|}{\partial t} = |\mathbf{A}|tr\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial t}\right).} \tag{A3-34}$$

A.3.7 Norms

Norm of vectors or matrices is often useful when analyzing growth of numerical errors. The Hölder p-norms for vectors are defined as

$$\boxed{\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.} \tag{A3-35}$$

The most frequently used p-norms are

$$\boxed{\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|,} \quad \boxed{\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2},} \quad \boxed{\|\mathbf{x}\|_\infty = \max_i |x_i|.} \tag{A3-36}$$

The l_2-norm $\|\mathbf{x}\|_2$ is also called the Euclidian or root-sum-squared norm.

An l_2-like norm can be defined for matrices by treating the matrix elements as a vector. This leads to the *Frobenius norm*

$$\boxed{\|\mathbf{A}\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2\right)^{1/2}.} \tag{A3-37}$$

Induced matrix norms measure the ability of matrix $\mathbf{A}$ to modify the magnitude of a vector; that is,

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1}\left(\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}\right).$$

The l_1-norm is $\|\mathbf{A}\|_1 = \max_j \|\mathbf{a}_{:j}\|_1$ where $\mathbf{a}_{:j}$ is the j-th column of $\mathbf{A}$, and the l_∞-norm is $\|\mathbf{A}\|_\infty = \max_i \|\mathbf{a}_{i:}\|_1$ where $\mathbf{a}_{i:}$ is the i-th row of $\mathbf{A}$. It is more difficult to compute an l_2-norm based on this definition than a Frobenius norm. $\|\mathbf{A}\|_2$ is equal to the square root of the maximum eigenvalue of $\mathbf{A}^T\mathbf{A}$—or equivalently the largest singular value of $\mathbf{A}$. These terms are defined in Section A.4.

Norms of matrix products obey inequality conditions:

$$\boxed{\|\mathbf{A}\mathbf{x}\| \le \|\mathbf{A}\|\|\mathbf{x}\|} \quad \text{or} \quad \boxed{\|\mathbf{A}\mathbf{B}\| \le \|\mathbf{A}\|\|\mathbf{B}\|.} \tag{A3-38}$$

The l_2 and Frobenius matrix norms are unchanged by orthogonal transformations, that is,

$$\boxed{\left.\begin{aligned}\|\mathbf{A}\mathbf{B}\|_2 &= \|\mathbf{A}\|_2\\ \|\mathbf{B}\mathbf{A}\|_2 &= \|\mathbf{A}\|_2\end{aligned}\right\} \quad \text{if} \quad \mathbf{B}^T\mathbf{B} = \mathbf{I}.} \tag{A3-39}$$

A.4 MATRIX TRANSFORMATIONS AND FACTORIZATION

A.4.1 LU Decomposition

LU decomposition has been mentioned previously. Crout reduction is used to factor square matrix $\mathbf{A} = \mathbf{LU}$ where $\mathbf{L}$ is unit lower triangular and $\mathbf{U}$ is upper triangular. This is often used for matrix inversion or when repeatedly solving equations of the form $\mathbf{Ax} = \mathbf{y}$ for $\mathbf{x}$. The equation $\mathbf{Lz} = \mathbf{y}$ is first solved for $\mathbf{z}$ using forward substitution, and then $\mathbf{Ux} = \mathbf{z}$ is solved for $\mathbf{x}$ using backward substitution.

A.4.2 Cholesky Factorization

Cholesky factorization is essentially LU decomposition for symmetric matrices. Usually it implies $\mathbf{A} = \mathbf{LL}^T$ where $\mathbf{L}$ is lower triangular, but it is sometimes used to indicate $\mathbf{A} = \mathbf{L}^T\mathbf{L}$. Cholesky factorization for symmetric matrices is much more efficient and accurate than LU decomposition.

A.4.3 Similarity Transformation

A similarity transformation on square matrix $\mathbf{A}$ is one of the form $\mathbf{B} = \mathbf{T}^{-1}\mathbf{AT}$. Since $\mathbf{TB} = \mathbf{AT}$, matrices $\mathbf{A}$ and $\mathbf{B}$ are called similar. Similarity transformations can be used to reduce a given matrix to an equivalent canonical form; for example, eigen decomposition and singular value decomposition discussed below.

A.4.4 Eigen Decomposition

The vectors $\mathbf{x}_i$ for which $\mathbf{Ax}_i = \lambda_i \mathbf{x}_i$, where λ_i are scalars, are called the eigenvectors of square matrix $\mathbf{A}$. For an $n \times n$ matrix $\mathbf{A}$, there will be n eigenvectors $\mathbf{x}_i$ with corresponding eigenvalues λ_i. The λ_i may be complex (occurring in complex conjugate pairs) even when $\mathbf{A}$ is real, and some eigenvalues may be repeated. The eigenvectors are the directions that are invariant with pre-multiplication by $\mathbf{A}$.

The eigenvector/eigenvalue relationship can be written in matrix form as

$$\boxed{\mathbf{AM} = \mathbf{M\Lambda}} \tag{A4-1}$$

or

$$\boxed{\mathbf{A} = \mathbf{M\Lambda M}^{-1}} \tag{A4-2}$$

where

$$\mathbf{M} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n], \quad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Matrix $\mathbf{M}$ is called the modal matrix and λ_i are the eigenvalues. The eigenvalues are the roots of the *characteristic polynomial* $p(s) = |s\mathbf{I} - \mathbf{A}|$, and they define the spectral response of the linear system $\dot{\mathbf{x}}(t) = \mathbf{A}\,\mathbf{x}(t)$ when $\mathbf{x}(t)$ is the system state vector (not eigenvectors). Eigen decomposition is a similarity transformation, and thus

$$\boxed{|\mathbf{A}| = \lambda_1 \lambda_2 \ldots \lambda_n.} \tag{A4-3}$$

Also

$$\boxed{tr(\mathbf{A}) = \lambda_1 + \lambda_2 + \ldots + \lambda_n.} \tag{A4-4}$$

When real $\mathbf{A}$ is symmetric and nonsingular, the λ_i are all real, and the eigenvectors are distinct and orthogonal. Thus $\mathbf{M}^{-1} = \mathbf{M}^T$ and $\mathbf{A} = \mathbf{M\Lambda M}^T$.

Eigenvectors and eigenvalues are computed in LAPACK using either generalized QR decomposition or a divide-and-conquer approach.

A.4.5 Singular Value Decomposition (SVD)

While eigen decomposition is used for square matrices, SVD is used for either square or rectangular matrices. The SVD of $m \times n$ matrix $\mathbf{A}$ is

$$\boxed{\mathbf{A} = \mathbf{USV}^T} \tag{A4-5}$$

where $\mathbf{U}$ is an $m \times m$ orthogonal matrix ($\mathbf{U}\mathbf{U}^T = \mathbf{I}_m$), $\mathbf{V}$ is an $n \times n$ orthogonal matrix ($\mathbf{V}\mathbf{V}^T = \mathbf{I}_n$), and $\mathbf{S}$ is an $m \times n$ upper diagonal matrix of singular values. In least-squares problems $m > n$ is typical, so

$$\mathbf{A} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_n & \mathbf{u}_{n+1} & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} S_1 & 0 & \cdots & 0 \\ 0 & S_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \\ \mathbf{v}_n^T \end{bmatrix}$$

where $\mathbf{u}_i$ are the left singular vectors and $\mathbf{v}_i$ are the right singular vectors. The left singular vectors $\mathbf{u}_{n+1}, \mathbf{u}_{n+2}, \dots, \mathbf{u}_m$ are the nullspace of $\mathbf{A}$ as they multiply zeroes in $\mathbf{S}$. The above SVD is called "full," but some SVD utilities have the option to omit computation of $\mathbf{u}_{n+1}, \mathbf{u}_{n+2}, \dots, \mathbf{u}_m$. These are called "thin" SVDs.

When $\mathbf{A}$ is a real symmetric matrix formed as $\mathbf{A} = \mathbf{H}^T\mathbf{H}$ (such as the information matrix in least-squares estimation), the eigenvalues of $\mathbf{A}$ are the squares of the singular values of $\mathbf{H}$, and the eigenvectors are the right singular vectors $\mathbf{v}_i$. This is seen from

$$\mathbf{A} = \mathbf{H}^T\mathbf{H} = (\mathbf{V}\mathbf{S}^T\mathbf{U}^T)(\mathbf{U}\mathbf{S}\mathbf{V}^T) = \mathbf{V}(\mathbf{S}^T\mathbf{S})\mathbf{V}^T.$$

The accuracy of the SVD factorization is much greater than that for eigen decomposition of $\mathbf{A} = \mathbf{H}^T\mathbf{H}$ because the "squaring" operation doubles the condition number (defined below). This doubles sensitivity to numerical round-off errors.

A.4.6 Pseudo-Inverse

When $\mathbf{A}$ is rectangular or singular, $\mathbf{A}$ does not have an inverse. However, Penrose (1955) defined a *pseudo-inverse* $\mathbf{A}^\#$ uniquely determined by four properties:

$$\boxed{\begin{aligned} &\mathbf{A}\mathbf{A}^\#\mathbf{A} = \mathbf{A} \\ &\mathbf{A}^\#\mathbf{A}\mathbf{A}^\# = \mathbf{A}^\# \\ &(\mathbf{A}\mathbf{A}^\#)^T = \mathbf{A}\mathbf{A}^\# \\ &(\mathbf{A}^\#\mathbf{A})^T = \mathbf{A}^\#\mathbf{A} \end{aligned}}. \tag{A4-6}$$

This Moore-Penrose pseudo-inverse is sometimes used in least-squares problems when there is insufficient measurement information to obtain a unique solution. A pseudo-inverse for the least-squares problem based on measurement equation $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{r}$ can be written using the SVD of $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Then the normal equation least-squares solution $\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$ is obtained using the pseudo-inverse of $\mathbf{H}$,

$$\begin{aligned} \mathbf{H}^\# &= (\mathbf{V}\mathbf{S}^T\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{S}^T\mathbf{U}^T \\ &= \mathbf{V}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{V}^T\mathbf{V}\mathbf{S}^T\mathbf{U}^T \\ &= \mathbf{V}(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\mathbf{U}^T \\ &= \mathbf{V}\begin{bmatrix}\mathbf{S}_1^\# & \mathbf{0}\end{bmatrix}\mathbf{U}^T \end{aligned},$$

where $\mathbf{S}_1^{\#}$ is the pseudo-inverse of the nonzero square portion of $\mathbf{S}$. For singular values that are exactly zero, they are replaced with zero in the same location when forming $\mathbf{S}_1^{\#}$. Thus

$$\hat{\mathbf{x}} = \mathbf{V}\begin{bmatrix}\mathbf{S}_1^{\#} & \mathbf{0}\end{bmatrix}\mathbf{U}^T\mathbf{y}.$$

This shows that a pseudo-inverse can be computed even when $\mathbf{H}^T\mathbf{H}$ is singular. The pseudo-inverse provides the minimal norm solution for a *rank-deficient* (rank < min(m,n)) $\mathbf{H}$ matrix.

A.4.7 Condition Number

The *condition number* of square matrix $\mathbf{A}$ is a measure of sensitivity of errors in $\mathbf{A}^{-1}$ to perturbations in $\mathbf{A}$. This is used to analyze growth of numerical errors. The condition number is denoted as

$$\boxed{\kappa_p(\mathbf{A}) = \|\mathbf{A}\|_p \|\mathbf{A}^{-1}\|_p} \tag{A4-7}$$

when using the an l_p induced matrix norm for $\mathbf{A}$. Because it is generally not convenient to compute condition numbers by inverting a matrix, they are most often computed from the singular values of matrix $\mathbf{A}$. Decomposing $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal and S_i are the singular values in $\mathbf{S}$, then

$$\boxed{\kappa_2(\mathbf{A}) = \frac{\max_i(S_i)}{\min_i(S_i)}.} \tag{A4-8}$$