A

Matrix Properties

THIS appendix provides a reasonably comprehensive account of matrix properties, which are used in the linear algebra of estimation and control theory. Several theorems are shown, but are not proven here; those proofs given are *constructive* (i.e., suggest an algorithm). The account here is thus not satisfactorily self-contained, but references are provided where rigorous proofs may be found.

A.1 Basic Definitions of Matrices

The system of m linear equations

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{aligned} \tag{A.1}$$

can be written in matrix form as

$$\mathbf{y} = A\,\mathbf{x} \tag{A.2}$$

where $\mathbf{y}$ is an $m \times 1$ vector, $\mathbf{x}$ is an $n \times 1$ vector (see §A.2 for a definition of a vector) and A is an $m \times n$ matrix, with

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{A.3}$$

If $m = n$, then the matrix A is *square*.

Matrix Addition, Subtraction, and Multiplication

Matrices can be added, subtracted, or multiplied. For addition and subtraction, all matrices must of the same dimension. Suppose we wish to add/substract two matrices A and B:

$$C = A \pm B \tag{A.4}$$

Then each element of C is given by $c_{ij} = a_{ij} \pm b_{ij}$. Matrix addition and subtraction are both commutative, $A \pm B = B \pm A$, and associative, $(A \pm B) \pm C = A \pm (B \pm C)$. Matrix multiplication is much more complicated though. Suppose we wish to multiply two matrices A and B:

$$C = A\,B \tag{A.5}$$

This operation is valid only when the number of columns of A is equal to the number of rows of B (i.e., A and B must be *conformable*). The resulting matrix C will have rows equal to the number of rows of A and columns equal to the number of columns of B. Thus, if A has dimension $m \times n$ and B has dimension $n \times p$, then C will have dimension $m \times p$. The c_{ij} element of C can be determined by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{A.6}$$

for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, p$. Matrix multiplication is associative, $A\,(B\,C) = (A\,B)\,C$, and distributive, $A\,(B + C) = A\,B + A\,C$, but not commutative in general, $A\,B \neq B\,A$. In some cases though if $A\,B = B\,A$, then A and B are said to *commute*.

The transpose of a matrix, denoted A^T, has rows that are the columns of A and columns that are the rows of A. The transpose operator has the following properties:

$$(\alpha A)^T = \alpha A^T, \text{ where } \alpha \text{ is a scalar} \tag{A.7a}$$

$$(A + B)^T = A^T + B^T \tag{A.7b}$$

$$(A\,B)^T = B^T A^T \tag{A.7c}$$

If $A = A^T$, then A is said to be a *symmetric* matrix. Also, if $A = -A^T$, then A is said to be a *skew symmetric* matrix.

Matrix Inverse

We now discuss the properties of the matrix inverse. Suppose we are given both $\mathbf{y}$ and A in eqn. (A.2), and we want to determine $\mathbf{x}$. The following terminology should be noted carefully: if $m > n$, the system in eqn. (A.2) is said to be *overdetermined* (there are more equations than unknowns). Under typical circumstances we will find that the exact solution for $\mathbf{x}$ does not exist; therefore algorithms for *approximate* solutions for $\mathbf{x}$ are usually characterized by some measure of *how well* the linear equations are satisfied. If $m < n$, the system in eqn. (A.2) is said to be *underdetermined* (there are fewer equations than unknowns). Under typical circumstances, an infinity of exact solutions for $\mathbf{x}$ exist; therefore solution algorithms have implicit some criterion for selecting a particular solution from the infinity of possible or feasible $\mathbf{x}$ solutions. If $m = n$ the system is said to be *determined*, under typical (but certainly not universal) circumstances, a unique exact solution for $\mathbf{x}$ exists. To determine $\mathbf{x}$ for this case, the matrix inverse of A, denoted by A^{-1}, is used. Let A be an $n \times n$ matrix. The following statements are equivalent:

- A has linearly independent columns.
- A has linearly independent rows.
- The inverse satisfies $A^{-1}A = A\,A^{-1} = I$

where I is an $n \times n$ identity matrix:

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{A.8}$$

A *nonsingular* matrix is a matrix whose inverse exists (likewise A^T is nonsingular):

$$(A^{-1})^{-1} = A \tag{A.9a}$$

$$(A^T)^{-1} = (A^{-1})^T \equiv A^{-T} \tag{A.9b}$$

Furthermore, let A and B be $n \times n$ matrices. The matrix product $A\,B$ is nonsingular if and only if A and B are nonsingular. If this condition is met, then

$$(A\,B)^{-1} = B^{-1}A^{-1} \tag{A.10}$$

Formal proof of this relationship and other relationships are given in Ref. [1]. The inverse of a square matrix A can be computed by

$$A^{-1} = \frac{\text{adj}(A)}{\det(A)} \tag{A.11}$$

where $\text{adj}(A)$ is the *adjoint* of A and $\det(A)$ is the *determinant* of A. The adjoint and determinant of a matrix with large dimension can ultimately be broken down to a series of 2×2 matrix cases, where the adjoint and determinant are given by

$$\text{adj}(A_{2\times 2}) = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{A.12a}$$

$$\det(A_{2\times 2}) = a_{11}a_{22} - a_{12}a_{21} \tag{A.12b}$$

Other determinant identities are given by

$$\det(I) = 1 \tag{A.13a}$$

$$\det(A\,B) = \det(A)\,\det(B) \tag{A.13b}$$

$$\det(A\,B) = \det(B\,A) \tag{A.13c}$$

$$\det(A\,B + I) = \det(B\,A + I) \tag{A.13d}$$

$$\det(A + \mathbf{x}\mathbf{y}^T) = \det(A)\,(1 + \mathbf{y}^T A^{-1}\mathbf{x}) \tag{A.13e}$$

$$\det(A)\,\det(D + C\,A^{-1}B) = \det(D)\,\det(A + B\,D^{-1}C) \tag{A.13f}$$

$$\det(A^\alpha) = [\det(A)]^\alpha,\ \alpha \text{ must be positive if } \det(A) = 0 \tag{A.13g}$$

$$\det(\alpha\,A) = \alpha^n \det(A) \tag{A.13h}$$

$$\det(A_{3\times 3}) \equiv \det\left(\begin{bmatrix}\mathbf{a} & \mathbf{b} & \mathbf{c}\end{bmatrix}\right) = \mathbf{a}^T[\mathbf{b}\times]\mathbf{c} = \mathbf{b}^T[\mathbf{c}\times]\mathbf{a} = \mathbf{c}^T[\mathbf{a}\times]\mathbf{b} \tag{A.13i}$$

where the matrices $[\mathbf{a}\times]$, $[\mathbf{b}\times]$, and $[\mathbf{a}\times]$ are defined in eqn. (A.38). The adjoint is given by the transpose of the *cofactor* matrix:

$$\text{adj}(A) = [\text{cof}(A)]^T \tag{A.14}$$

The cofactor is given by

$$C_{ij} = (-1)^{i+j} M_{ij} \tag{A.15}$$

where M_{ij} is the *minor*, which is the determinant of the resulting matrix given by crossing out the row and column of the element a_{ij}. The determinant can be computed using an expansion about row i or column j:

$$\det(A) = \sum_{k=1}^{n} a_{ik} C_{ik} = \sum_{k=1}^{n} a_{kj} C_{kj} \tag{A.16}$$

From eqn. (A.11) A^{-1} exists if and only if the determinant of A is nonzero. Matrix inverses are usually complicated to compute numerically; however, a special case is when the inverse is given by the transpose of the matrix itself. This matrix is then said to be *orthogonal* with the property

$$A^T A = A\, A^T = I \tag{A.17}$$

Also, the determinant of an orthogonal matrix can be shown to be ± 1. An orthogonal matrix preserves the length (norm) of a vector (see eqn. (A.27) for a definition of the norm of a vector). Hence, if A is an orthogonal matrix, then $||A\,\mathbf{x}|| = ||\mathbf{x}||$.

Block Structures and Other Identities

Matrices can also be analyzed using block structures. Assume that A is an $n \times n$ matrix and that C is an $m \times m$ matrix. Then, we have

$$\det\begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \det\begin{bmatrix} A & 0 \\ B & C \end{bmatrix} = \det(A)\,\det(C) \tag{A.18a}$$

$$\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\,\det(P) = \det(D)\,\det(Q) \tag{A.18b}$$

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} Q^{-1} & -Q^{-1} B\, D^{-1} \\ -D^{-1} C\, Q^{-1} & D^{-1}(I + C\, Q^{-1} B\, D^{-1}) \end{bmatrix} \\ &= \begin{bmatrix} A^{-1}(I + B\, P^{-1} C\, A^{-1}) & -A^{-1} B\, P^{-1} \\ -P^{-1} C\, A^{-1} & P^{-1} \end{bmatrix} \end{aligned} \tag{A.18c}$$

where P and Q are *Schur complements* of A and D:

$$P \equiv D - C\, A^{-1} B \tag{A.19a}$$

$$Q \equiv A - B\, D^{-1} C \tag{A.19b}$$

Other useful matrix identities involve the *Sherman-Morrison lemma*, given by

$$(I + A\,B)^{-1} = I - A\,(I + B\,A)^{-1} B \tag{A.20}$$

and the *matrix inversion lemma*, given by

$$(A + B\,C\,D)^{-1} = A^{-1} - A^{-1} B \left(D\,A^{-1} B + C^{-1} \right)^{-1} D\,A^{-1} \tag{A.21}$$

where A is an arbitrary $n \times n$ matrix and C is an arbitrary $m \times m$ matrix. A proof of the matrix inversion lemma is given in §1.3.

Matrix Trace

Another useful quantity often used in estimation theory is the *trace* of a matrix, which is defined only for square matrices:

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} \tag{A.22}$$

Some useful identities involving the matrix trace are given by

$$\mathrm{Tr}(\alpha A) = \alpha \mathrm{Tr}(A) \tag{A.23a}$$

$$\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B) \tag{A.23b}$$

$$\mathrm{Tr}(A\,B) = \mathrm{Tr}(B\,A) \tag{A.23c}$$

$$\mathrm{Tr}(\mathbf{x}\mathbf{y}^T) = \mathbf{x}^T \mathbf{y} \tag{A.23d}$$

$$\mathrm{Tr}(A\,\mathbf{y}\mathbf{x}^T) = \mathbf{x}^T A\,\mathbf{y} \tag{A.23e}$$

$$\mathrm{Tr}(A\,B\,C\,D) = \mathrm{Tr}(B\,C\,D\,A) = \mathrm{Tr}(C\,D\,A\,B) = \mathrm{Tr}(D\,A\,B\,C) \tag{A.23f}$$

Equation (A.23f) shows the cyclic invariance of the trace. The operation $\mathbf{y}\mathbf{x}^T$ is known as the *outer product* (also $\mathbf{y}\mathbf{x}^T \neq \mathbf{x}\mathbf{y}^T$ in general).

Solution of Triangular Systems

An *upper triangular system* of linear equations has the form

$$\begin{aligned} t_{11}x_1 + t_{12}x_2 + t_{13}x_3 + \cdots + t_{1n}x_n &= y_1 \\ t_{22}x_2 + t_{23}x_3 + \cdots + t_{2n}x_n &= y_2 \\ t_{33}x_3 + \cdots + t_{3n}x_n &= y_3 \\ &\vdots \\ t_{nn}x_n &= y_n \end{aligned} \tag{A.24}$$

or

$$T\mathbf{x} = \mathbf{y} \tag{A.25}$$

where

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \cdots & t_{1n} \\ 0 & t_{22} & t_{23} & \cdots & t_{2n} \\ 0 & 0 & t_{33} & \cdots & t_{3n} \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & \cdots & t_{nn} \end{bmatrix} \tag{A.26}$$

The matrix T can be shown to be nonsingular if and only if its diagonal elements are nonzero.[1] Clearly, x_n can be easily determined using the upper triangular form. The x_i coefficients can be determined by a *back substitution algorithm*:

for $i = n, n-1, \ldots, 1$

$$x_i = t_{ii}^{-1}\left(y_i - \sum_{j=i+1}^{n} t_{ij}x_j\right)$$

next i

This algorithm will fail only if $t_{ii} \to 0$. But, this can occur only if T is singular (or nearly singular). Experience indicates that the algorithm is well-behaved for most applications though.

The back substitution algorithm can be modified to compute the inverse, $S = T^{-1}$, of an upper triangular matrix T. We now summarize an algorithm for calculating $S = T^{-1}$ and overwriting T by T^{-1}:

for $k = n, n-1, \ldots, 1$

$$t_{kk} \leftarrow S_{kk} = t_{kk}^{-1}$$

$$t_{ik} \leftarrow S_{ik} = -t_{ii}^{-1}\sum_{j=i+1}^{k} t_{ij}s_{jk}, \quad i = k-1, k-2, \ldots, 1$$

next k

where $\leftarrow$ denotes replacement.* This algorithm requires about $n^3/6$ calculations (note: if only the solution of $\mathbf{x}$ is required and not the explicit form for T^{-1}, then the back substitution algorithm should be solely employed since only $n^2/2$ calculations are required for this algorithm).

A.2 Vectors

The quantities $\mathbf{x}$ and $\mathbf{y}$ in eqn. (A.2) are known as *vectors*, which are a special case of a matrix. Vectors can consist of one row, known as a *row vector*, or one column, known as a *column vector*.

Vector Norm and Dot Product

A measure of the length of a vector is given by the norm:

$$||\mathbf{x}|| \equiv \sqrt{\mathbf{x}^T\mathbf{x}} = \left[\sum_{i=1}^{n} x_i^2\right]^{1/2} \tag{A.27}$$

Also, $||\alpha\mathbf{x}|| = |\alpha|\,||\mathbf{x}||$. A vector with norm one is said to be a *unit vector*. Any

*The symbol $x \leftarrow y$ means "overwrite" x by the current y-value. This notation is employed to indicate how storage may be conserved by overwriting quantities no longer needed.

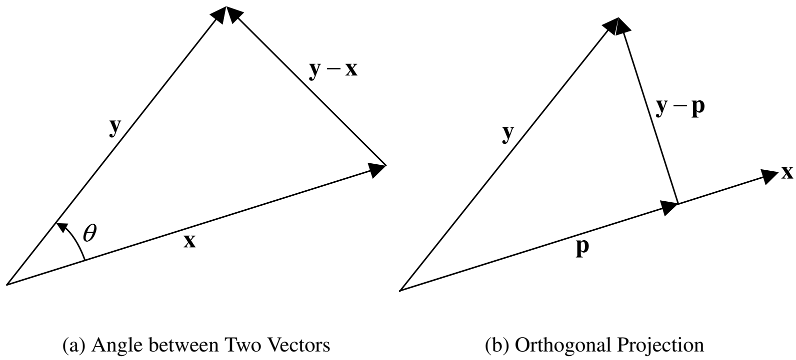

Figure A.1: Depiction of the Angle between Two Vectors and an Orthogonal Projection

nonzero vector can be made into a unit vector by dividing it by its norm:

$$\hat{\mathbf{x}} \equiv \frac{\mathbf{x}}{||\mathbf{x}||} \tag{A.28}$$

Note that the carat is also used to denote estimate in this text. The *dot product* or *inner product* of two vectors of equal dimension, $n \times 1$, is given by

$$\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_{i=1}^{n} x_i y_i \tag{A.29}$$

If the dot product is zero, then the vectors are said to be *orthogonal*. Suppose that a set of vectors $\mathbf{x}_i$ $(i = 1, 2, \ldots, m)$ follows

$$\mathbf{x}_i^T \mathbf{x}_j = \delta_{ij} \tag{A.30}$$

where the Kronecker delta δ_{ij} is defined as

$$\begin{aligned} \delta_{ij} &= 0 \quad \text{if } i \neq j \\ &= 1 \quad \text{if } i = j \end{aligned} \tag{A.31}$$

Then, this set is said to be *orthonormal*. The column and row vectors of an orthogonal matrix, defined by the property shown in eqn. (A.17), form an orthonormal set.

Angle Between Two Vectors and the Orthogonal Projection

Figure A.1(a) shows two vectors, $\mathbf{x}$ and $\mathbf{y}$, and the angle θ which is the angle between them. This angle can be computed from the cosine law:

$$\cos(\theta) = \frac{\mathbf{x}^T \mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||} \tag{A.32}$$

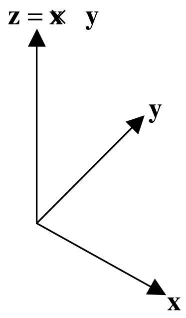

Figure A.2: Cross Product and the Right Hand Rule

Figure A.1(b) shows the *orthogonal projection* of a vector $\mathbf{y}$ to a vector $\mathbf{x}$. The orthogonal projection of $\mathbf{y}$ to $\mathbf{x}$ is given by

$$\mathbf{p} = \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||^2}\mathbf{x} \tag{A.33}$$

This projection yields $(\mathbf{y}-\mathbf{p})^T\mathbf{x} = 0$.

Triangle and Schwartz Inequalities

Some important inequalities are given by the *triangle inequality*:

$$||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \tag{A.34}$$

and the *Schwartz inequality*:

$$|\mathbf{x}^T\mathbf{y}| \leq ||\mathbf{x}||\,||\mathbf{y}|| \tag{A.35}$$

Note that the Schwartz inequality implies the triangle inequality.

Cross Product

The cross product of two vectors yields a vector that is perpendicular to both vectors. The cross product of $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\mathbf{z} = \mathbf{x}\times\mathbf{y} = \begin{bmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{bmatrix} \tag{A.36}$$

The cross product follows the *right hand rule*, which states that the orientation of $\mathbf{z}$ is determined by placing $\mathbf{x}$ and $\mathbf{y}$ tail-to-tail, flattening the right hand, extending it in the direction of $\mathbf{x}$, and then curling the fingers in the direction that the angle $\mathbf{y}$ makes with $\mathbf{x}$. The thumb then points in the direction of $\mathbf{z}$, as shown in Figure A.2. The cross product can also be obtained using matrix multiplication:

$$\mathbf{z} = [\mathbf{x}\times]\mathbf{y} \tag{A.37}$$

where $[\mathbf{x}\times]$ is the *cross product matrix*, defined by

$$[\mathbf{x}\times] \equiv \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \tag{A.38}$$

Note that $[\mathbf{x}\times]$ is a skew symmetric matrix.

The cross product has the following properties:

$$[\mathbf{x}\times]^T = -[\mathbf{x}\times] \tag{A.39a}$$

$$[\mathbf{x}\times]\mathbf{y} = -[\mathbf{y}\times]\mathbf{x} \tag{A.39b}$$

$$[\mathbf{x}\times][\mathbf{y}\times] = -\left(\mathbf{x}^T\mathbf{y}\right)I + \mathbf{y}\mathbf{x}^T \tag{A.39c}$$

$$[\mathbf{x}\times]^3 = -\left(\mathbf{x}^T\mathbf{x}\right)[\mathbf{x}\times] \tag{A.39d}$$

$$[\mathbf{x}\times][\mathbf{y}\times] - [\mathbf{y}\times][\mathbf{x}\times] = \mathbf{y}\mathbf{x}^T - \mathbf{x}\mathbf{y}^T = [(\mathbf{x}\times\mathbf{y})\times] \tag{A.39e}$$

$$\mathbf{x}\mathbf{y}^T[\mathbf{w}\times] + [\mathbf{w}\times]\mathbf{y}\mathbf{x}^T = -[\{\mathbf{x}\times(\mathbf{y}\times\mathbf{w})\}\times] \tag{A.39f}$$

$$(I - [\mathbf{x}\times])(I + [\mathbf{x}\times])^{-1} = \frac{1}{1+\mathbf{x}^T\mathbf{x}}\left\{(1-\mathbf{x}^T\mathbf{x})I + 2\mathbf{x}\mathbf{x}^T - 2[\mathbf{x}\times]\right\} \tag{A.39g}$$

Other useful properties involving an arbitrary 3×3 square matrix M are given by[2]

$$M[\mathbf{x}\times] + [\mathbf{x}\times]M^T + [(M^T\mathbf{x})\times] = \mathrm{Tr}(M)[\mathbf{x}\times] \tag{A.40a}$$

$$M[\mathbf{x}\times]M^T = [\{\mathrm{adj}(M^T)\mathbf{x}\}\times] \tag{A.40b}$$

$$(M\mathbf{x})\times(M\mathbf{y}) = \mathrm{adj}(M^T)(\mathbf{x}\times\mathbf{y}) \tag{A.40c}$$

$$[\{(M\mathbf{x})\times(M\mathbf{y})\}\times] = M[(\mathbf{x}\times\mathbf{y})\times]M^T \tag{A.40d}$$

If we write M in terms of its columns

$$M = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{bmatrix} \tag{A.41}$$

then

$$\det(M) = \mathbf{x}_1^T(\mathbf{x}_2\times\mathbf{x}_3) \tag{A.42}$$

Also, if A is an orthogonal matrix with determinant 1, then from eqn. (A.40b) we have

$$A[\mathbf{x}\times]A^T = [(A\mathbf{x})\times] \tag{A.43}$$

Another important quantity using eqn. (A.39c) is given by

$$[\mathbf{x}\times]^2 = -\left(\mathbf{x}^T\mathbf{x}\right)I + \mathbf{x}\mathbf{x}^T \tag{A.44}$$

This matrix is the projection operator onto the space perpendicular to $\mathbf{x}$. Many other interesting relations involving the cross product are given in Ref. [3].

Table A.1: Matrix and Vector Norms

Norm	**Vector**	**Matrix**
One-norm	$\|\|\mathbf{x}\|\|_1 = \sum_{i=1}^n \|x_i\|$	$\|\|A\|\|_1 = \max_j \sum_{i=1}^n \|a_{ij}\|$
Two-norm	$\|\|\mathbf{x}\|\|_2 = \left[\sum_{i=1}^n x_i^2\right]^{1/2}$	$\|\|A\|\|_2 =$ max singular value of A
Frobenius norm	$\|\|\mathbf{x}\|\|_F = \|\|\mathbf{x}\|\|_2$	$\|\|A\|\|_F = \sqrt{\mathrm{Tr}(A^* A)}$
Infinity-norm	$\|\|\mathbf{x}\|\|_\infty = \max_i \|x_i\|$	$\|\|\|A\|\|_\infty = \max_i \sum_{j=1}^n \|a_{ij}\|$

The angle θ in Figure A.1(a) can be computed from

$$\sin(\theta) = \frac{||\mathbf{x} \times \mathbf{y}||}{||\mathbf{x}||\,||\mathbf{y}||} \tag{A.45}$$

Using $\sin^2(\theta) + \cos^2(\theta) = 1$, eqns. (A.32) and (A.45) also give

$$||\mathbf{x} \times \mathbf{y}|| = \sqrt{(\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y}) - (\mathbf{x}^T\mathbf{y})^2} \tag{A.46}$$

From the Schwartz inequality in eqn. (A.35), the quantity within the square root in eqn. (A.46) is always positive.

A.3 Matrix Norms and Definiteness

Norms for matrices are slightly more difficult to define than for vectors. Also, the definition of a "positive" or "negative" matrix is more complicated for matrices than scalars. Before showing these quantities, we first define the following quantities for any complex matrix A:

- *Conjugate transpose*: defined as the transpose of the conjugate of each element; denoted by A^*.
- *Hermitian*: has the property $A = A^*$ (note: any real symmetric matrix is Hermitian).
- *Normal*: has the property $A^* A = A\, A^*$.
- *Unitary*: inverse is equal to its Hermitian transpose, so that $A^* A = A\, A^* = I$ (note: a real unitary matrix is an orthogonal matrix).

Matrix Norms

Several possible matrix norms can be defined. Table A.1 lists the most commonly used norms for both vectors and matrices. The one-norm is the largest column sum. The two-norm is the maximum singular value (see §A.4). Also, unless otherwise stated, the norm defined without showing a subscript is the two-norm, as shown by eqn. (A.27). The Frobenius norm is defined as the square root of the sum of the absolute squares of its elements. The infinity-norm is the largest row sum. The matrix norms described in Table A.1 have the following properties:

$$||\alpha A|| = |\alpha|\, ||A|| \tag{A.47a}$$

$$||A + B|| \le ||A|| + ||B|| \tag{A.47b}$$

$$||A\, B|| \le ||A||\, ||B|| \tag{A.47c}$$

Not all norms follow eqn. (A.47c) though (e.g., the maximum absolute matrix element). More matrix norm properties can be found in Refs. [4] and [5].

Definiteness

Sufficiency tests in least squares and the minimization of functions with multiple variables often require that one determine the *definiteness* of the matrix of second partial derivatives. A real and square matrix A is

- *Positive definite* if $\mathbf{x}^T A\, \mathbf{x} > 0$ for all nonzero $\mathbf{x}$.
- *Positive semi-definite* if $\mathbf{x}^T A\, \mathbf{x} \ge 0$ for all nonzero $\mathbf{x}$.
- *Negative definite* if $\mathbf{x}^T A\, \mathbf{x} < 0$ for all nonzero $\mathbf{x}$.
- *Negative semi-definite* if $\mathbf{x}^T A\, \mathbf{x} \le 0$ for all nonzero $\mathbf{x}$.
- *Indefinite* when no definiteness can be asserted.

A simple test for a symmetric real matrix is to check its eigenvalues (see §A.4). This matrix is positive definite if and only if all its eigenvalues are greater than 0. Unfortunately, this condition is only necessary but not sufficient for a non-symmetric real matrix. A real matrix is positive definite if and only if its symmetric part, given by

$$B = \frac{A + A^T}{2} \tag{A.48}$$

is positive definite. Another way to state that a matrix is positive definite is the requirement that all the *leading* principal minors of A are positive.[6] If A is positive definite, then A^{-1} exists and is also positive definite. If A is positive semi-definite, then for any integer $\alpha > 0$ there exists a unique positive semi-definite matrix such that $A = B^\alpha$ (note: A and B commute, so that $A\, B = B\, A$). The following relationship:

$$B > A \tag{A.49}$$

implies $(B - A) > 0$, which states that the matrix $(B - A)$ is positive definite. Also,

$$B \ge A \tag{A.50}$$

implies $(B - A) \geq 0$, which states that the matrix $(B - A)$ is positive semi-definite. The conditions for negative definite and negative semi-definite are obvious from the definitions stated for positive definite and positive semi-definite.

A.4 Matrix Decompositions

Several matrix decompositions are given in the open literature. Many of these decompositions are used in place of a matrix inverse either to simplify the calculations or to provide more numerically robust approaches. In this section we present several useful matrix decompositions that are widely used in estimation and control theory. The methods to compute these decompositions is beyond the scope of the present text. Reference [4] provides all the necessary algorithms and proofs for the interested reader. Before we proceed a short description of the *rank* of a matrix is provided. Several definitions are possible. We will state that the rank of a matrix is given by the dimension of the range of the matrix corresponding to the number of linearly independent rows or columns. An $m \times n$ matrix is *rank deficient* if the rank of A is less than the minimum (m, n). Suppose that the rank of an $n \times n$ matrix A is given by $\text{rank}(A) = r$. Then, a set of $(n - r)$ nonzero unit vectors, $\hat{\mathbf{x}}_i$, can always be found that have the following property for a singular square matrix A:

$$A\,\hat{\mathbf{x}}_i = \mathbf{0}, \quad i = 1, 2, \ldots, n - r \tag{A.51}$$

The value of $(n - r)$ is known as the *nullity*, which is the maximum number of linearly independent null vectors of A. These vectors can form an orthonormal basis (which is how they are commonly shown) for the *null space* of A, and can be computed from the singular value decomposition. If A is nonsingular then no nonzero vector $\hat{\mathbf{x}}_i$ can be found to satisfy eqn. (A.51). For more details on the rank of a matrix see Refs. [4] and [6].

Eigenvalue/Eigenvector Decomposition and the Cayley-Hamilton Theorem

One of the most widely used decompositions for a square $n \times n$ matrix A in the study of dynamical systems is the eigenvalue/eigenvector decomposition. A real or complex number λ is an *eigenvalue* of A if there exists a nonzero (right) *eigenvector* $\mathbf{p}$ such that

$$A\,\mathbf{p} = \lambda \mathbf{p} \tag{A.52}$$

The solution for $\mathbf{p}$ is not unique in general, so usually $\mathbf{p}$ is given as a unit vector. In order for eqn. (A.52) to have a nonzero solution for $\mathbf{p}$, from eqn. (A.51), the matrix $(\lambda I - A)$ must be singular. Therefore, from eqn. (A.11) we have

$$\det(\lambda I - A) = \lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n = 0 \tag{A.53}$$

Equation (A.53) leads to a polynomial of degree n, which is called the *characteristic equation* of A. For example, the characteristic equation for a 3×3 matrix is given by

$$\lambda^3 - \lambda^2 \mathrm{Tr}(A) + \lambda\, \mathrm{Tr}[\mathrm{adj}(A)] - \det(A) = 0 \tag{A.54}$$

If all eigenvalues of A are distinct, then the set of eigenvectors is linearly independent. Therefore, the matrix A can be *diagonalized* as

$$\Lambda = P^{-1} A\, P \tag{A.55}$$

where $\Lambda = \mathrm{diag}\begin{bmatrix}\lambda_1 & \lambda_2 & \cdots & \lambda_n\end{bmatrix}$ and $P = \begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix}$. If A has repeated eigenvalues, then a block diagonal and triangular-form representation must be used, called a *Jordan block*.[6] The eigenvalue/eigenvector decomposition can be used for linear state variable transformations (see §3.1.4). Eigenvalues and eigenvectors can either be real or complex. This decomposition is very useful when A is symmetric, since Λ is always diagonal (even for repeated eigenvalues) and P is orthogonal for this case. A proof is given in Ref. [4]. Also, Ref. [4] provides many algorithms to compute the eigenvalue/eigenvector decomposition.

One of the most useful properties used in linear algebra is the *Cayley-Hamilton theorem*, which states that a matrix satisfies its own characteristic equation, so that

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I = 0 \tag{A.56}$$

This theorem is useful for computing powers of A that are larger than n, since A^{n+1} can be written as a linear combination of $(A,\ A^2,\ \ldots,\ A^n)$.[6]

QR Decomposition

The QR decomposition is especially useful in least squares (see §1.6.1) and the Square Root Information Filter (SRIF) (see §5.7.1). The QR decomposition of an $m \times n$ matrix A, with $m \geq n$, is given by

$$A = \mathcal{Q}\mathcal{R} \tag{A.57}$$

where $\mathcal{Q}$ is an $m \times m$ orthogonal matrix, and $\mathcal{R}$ is an upper triangular $m \times n$ matrix with all elements $\mathcal{R}_{ij} = 0$ for $i > j$. If A has full column rank, then the first n columns of $\mathcal{Q}$ form an orthonormal basis for the range of A.[4] Therefore, the "thin" QR decomposition is often used:

$$A = Q\, R \tag{A.58}$$

where Q is an $m \times n$ matrix with orthonormal columns and R is an upper triangular $n \times n$ matrix. Since the QR decomposition is widely used throughout the present text, we present a numerical algorithm to compute this decomposition by the *modified Gram-Schmidt* method.[4] Let A and Q be partitioned by columns $\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}$ and $\begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n\end{bmatrix}$, respectively. To begin the algorithm we set $Q = A$, and then

for $k = 1, 2, \ldots, n$

  $r_{kk} = ||\mathbf{q}_k||_2$

  $\mathbf{q}_k \leftarrow \mathbf{q}_k / r_{kk}$

    $r_{kj} = \mathbf{q}_k^T \mathbf{q}_j, \quad j = k+1, \ldots, n$

$$\mathbf{q}_j \leftarrow \mathbf{q}_j - r_{kj}\mathbf{q}_k, \quad j = k+1, \ldots, n$$

next k

where $\leftarrow$ denotes replacement (note: r_{kk} and r_{kj} are elements of the matrix R). This algorithm works even when A is complex. The QR decomposition is useful to invert an $n \times n$ matrix A, which is given by $A^{-1} = R^{-1}Q^T$ (note: the inverse of an upper triangular matrix is also a triangular matrix). Other methods, based on the Householder transformation and Givens rotations, can be used for the QR decomposition.[4]

Singular Value Decomposition

Another decomposition of an $m \times n$ matrix A is the *singular-value decomposition*,[4, 7] which decomposes a matrix into a diagonal matrix and two orthogonal matrices:

$$A = \mathcal{U}\,\mathcal{S}\,\mathcal{V}^* \tag{A.59}$$

where $\mathcal{U}$ is an $m \times m$ unitary matrix, $\mathcal{S}$ is an $m \times n$ diagonal matrix such that $\mathcal{S}_{ij} = 0$ for $i \neq j$, and $\mathcal{V}$ is an $n \times n$ unitary matrix. Many efficient algorithms can be used to determine the singular value decomposition.[4] Note that the zeros below the diagonal in $\mathcal{S}$ (with $m > n$) imply that the elements of columns $(n+1)$, $(n+2)$, $\ldots$, m of $\mathcal{U}$ are arbitrary. So, we can define the following reduced singular value decomposition:

$$A = U\,S\,V^* \tag{A.60}$$

where U is the $m \times n$ subset matrix of $\mathcal{U}$ (with the $(n+1)$, $(n+2)$, $\ldots$, m columns eliminated), S is the upper $n \times n$ matrix of $\mathcal{S}$, and $V = \mathcal{V}$. Note that $U^*U = I$, but it is no longer possible to make the same statement for $U\,U^*$. The elements of $S = \text{diag}\left[s_1 \ \cdots \ s_n\right]$ are known as the *singular values* of A, which are ordered from the smallest singular value to the largest singular value. These values are extremely important since they can give an indication of "how well" we can invert a matrix.[8] A common measure of the invertability of a matrix is the *condition number*, which is usually defined as the ratio of its largest singular value to its smallest singular value:

$$\text{Condition Number} = \frac{s_n}{s_1} \tag{A.61}$$

Large condition numbers may indicate a near singular matrix, and the minimum value of the condition number is unity (which occurs when the matrix is orthogonal). The rank of A is given by the number of nonzero singular values. Also, the singular value decomposition is useful to determine various norms (e.g., $||A||_F^2 = s_1^2 + \cdots + s_p^2$, where $p = \min(m, n)$, and the two-norm as shown in Table A.1).

Gaussian Elimination

Gaussian elimination is a classical reduction procedure by which a matrix A can be reduced to upper triangular form. This procedure involves pre-multiplications of a square matrix A by a sequence of *elementary lower triangular* matrices, each chosen to introduce a column with zeros below the diagonal (this process is often called "annihilation"). Several possible variations of Gaussian elimination can be derived.

We present a very robust algorithm called *Gaussian elimination with complete pivoting*. This approach requires data movements such as the interchange of two matrix rows. These interchanges can be tracked by using "permutation matrices," which are just identity matrices with rows or columns reordered. For example, consider the following matrix:

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{A.62}$$

So $P\,A$ is a row permutated version of A and $A\,P$ is a column permutated version of A. Permutation matrices are orthogonal. This algorithm computes the complete pivoting factorization

$$A = P\,L\,U\,Q^T \tag{A.63}$$

where P and Q are permutation matrices, L is a *unit* (with ones along the diagonal) lower triangular matrix, and U is an upper triangular matrix. The algorithm begins by setting $P = Q = I$, which are partitioned into column vectors as

$$P = \begin{bmatrix} \mathbf{p}_1 \; \mathbf{p}_2 \; \cdots \; \mathbf{p}_n \end{bmatrix}, \quad Q = \begin{bmatrix} \mathbf{q}_1 \; \mathbf{q}_2 \; \cdots \; \mathbf{q}_n \end{bmatrix} \tag{A.64}$$

The algorithm for Gaussian elimination with complete pivoting is given by overwriting the A matrix:

for $k = 1, 2, \ldots, n-1$
  Determine μ with $k \le \mu \le n$ and λ with $k \le \lambda \le n$ so
    $|a_{\mu\lambda}| = \max\{|a_{ij}|,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, n\}$
  if $\mu \ne k$
    $\mathbf{p}_k \leftrightarrow \mathbf{p}_\mu$
    $a_{kj} \leftrightarrow a_{\mu j}, \quad j = 1, 2, \ldots, n$
  end if
  if $\lambda \ne k$
    $\mathbf{q}_k \leftrightarrow \mathbf{q}_\lambda$
    $a_{jk} \leftrightarrow a_{j\lambda}, \quad j = 1, 2, \ldots, n$
  end if
  if $a_{kk} \ne 0$
    $a_{jk} \leftarrow a_{kj}/a_{kk}, \quad j = k+1, \ldots, n$
    $a_{jj} \leftarrow a_{jj} - a_{jk}a_{kj}, \quad j = k+1, \ldots, n$
  end if
next k

where $\leftarrow$ denotes replacement and $\leftrightarrow$ denotes "interchange the value assigned to." The matrix U is given by the upper triangular part (including the diagonal elements) of the overwritten A matrix, and the matrix L is given by the lower triangular part (replacing the diagonal elements with ones) of the overwritten A matrix. More details on Gaussian elimination can be found in Ref. [4].

LU and Cholesky Decompositions

The LU decomposition factors an $n \times n$ matrix A into a product of a lower triangular matrix L and an upper triangular matrix U, so that

$$A = L\,U \tag{A.65}$$

Gaussian elimination is a foremost example of LU decompositions. In general, the LU decomposition is not unique. This can be seen by observing that for an arbitrary nonsingular diagonal matrix D, setting $L' = L\,D$ and $U' = D^{-1}U$ yield new upper and lower triangular matrices that satisfy $L'U' = L\,D\,D^{-1}U = L\,U = A$. The fact that the decomposition is not unique suggests the possible wisdom of forming the *normalized* decomposition

$$A = L\,D\,U \tag{A.66}$$

in which L and U are unit lower and upper triangular matrices and D is a diagonal matrix. The question of existence and uniqueness is addressed by Stewart[1] who proves that the $A = L\,D\,U$ decomposition is unique, provided the leading diagonal sub-matrices of A are nonsingular.

There are three important variants of the $L\,D\,U$ decomposition; the first associates D with the lower triangular part to give the factorization

$$A = \mathcal{L}\,U \tag{A.67}$$

where $\mathcal{L} \equiv L\,D$. This is known as the *Crout reduction*. The second variant associates D with the upper triangular factor as

$$A = L\,\mathcal{U} \tag{A.68}$$

where $\mathcal{U} \equiv D\,U$. This reduction is exactly that obtained by Gaussian elimination.

The third variation is possible only for symmetric positive definite matrices, in which case

$$A = L\,D\,L^T \tag{A.69}$$

Thus A can be written as

$$A = \mathcal{L}\,\mathcal{L}^T \tag{A.70}$$

where now $\mathcal{L} \equiv L\,D^{1/2}$ is known as the *matrix square root*, and the factorization in eqn. (A.69) is known as the *Cholesky decomposition*. Efficient algorithms to compute the LU and Cholesky decompositions can be found in Ref. [4].

A.5 Matrix Calculus

In this section several relations are given for taking partial or time derivatives of matrices.[†] Before providing a list of matrix calculus identities, we first will define

[†]Most of these relations can be found in a website given by Mike Brooks, Imperial College, London, UK. As of this writing this website is given by http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html.

the *Jacobian* and *Hessian* of a scalar function $f(\mathbf{x})$, where $\mathbf{x}$ is an $n \times 1$ vector. The Jacobian of $f(\mathbf{x})$ is an $n \times 1$ vector given by

$$\nabla_{\mathbf{x}} f \equiv \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix} \tag{A.71}$$

The Hessian of $f(\mathbf{x})$ is an $n \times n$ matrix given by

$$\nabla^2_{\mathbf{x}} f \equiv \frac{\partial^2 f}{\partial \mathbf{x}\, \partial \mathbf{x}^T} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1 \partial x_1} & \dfrac{\partial f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial f}{\partial x_1 \partial x_n} \\ \dfrac{\partial f}{\partial x_2 \partial x_1} & \dfrac{\partial f}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f}{\partial x_n \partial x_1} & \dfrac{\partial f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n \partial x_n} \end{bmatrix} \tag{A.72}$$

Note that the Hessian of a scalar is a symmetric matrix. If $\mathbf{f}(\mathbf{x})$ is an $m \times 1$ vector and $\mathbf{x}$ is an $n \times 1$ vector, then the Jacobian matrix is given by

$$\nabla_{\mathbf{x}} \mathbf{f} \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{A.73}$$

Note that the Jacobian matrix is an $m \times n$ matrix. Also, there is a slight inconsistency between eqn. (A.71) and eqn. (A.73) when $m = 1$, since eqn. (A.71) gives an $n \times 1$ vector, while eqn. (A.73) gives a $1 \times n$ vector. This should pose no problems for the reader though since the context of this notation is clear for the particular system shown in this text.

A list of derivatives involving linear products is given by

$$\frac{\partial}{\partial \mathbf{x}}(A\mathbf{x}) = A \tag{A.74a}$$

$$\frac{\partial}{\partial A}(\mathbf{a}^T A\,\mathbf{b}) = \mathbf{a}\,\mathbf{b}^T \tag{A.74b}$$

$$\frac{\partial}{\partial A}(\mathbf{a}^T A^T \mathbf{b}) = \mathbf{b}\,\mathbf{a}^T \tag{A.74c}$$

$$\frac{d}{dt}(A\,B) = A\left[\frac{d}{dt}(B)\right] + \left[\frac{d}{dt}(A)\right]B \tag{A.74d}$$

A list of derivatives involving quadratic and cubic products is given by

$$\frac{\partial}{\partial \mathbf{x}}(A\,\mathbf{x}+\mathbf{b})^T C\,(D\,\mathbf{x}+\mathbf{e}) = A^T C\,(D\,\mathbf{x}+\mathbf{e}) + D^T C^T (A\,\mathbf{x}+\mathbf{b}) \tag{A.75a}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T C\,\mathbf{x}) = (C + C^T)\,\mathbf{x} \tag{A.75b}$$

$$\frac{\partial}{\partial A}(\mathbf{a}^T A^T A\,\mathbf{b}) = A\,(\mathbf{a}\,\mathbf{b}^T + \mathbf{b}\,\mathbf{a}^T) \tag{A.75c}$$

$$\frac{\partial}{\partial A}(\mathbf{a}^T A^T C\,A\,\mathbf{b}) = C^T A\,\mathbf{a}\,\mathbf{b}^T + C\,A\,\mathbf{b}\,\mathbf{a}^T \tag{A.75d}$$

$$\frac{\partial}{\partial A}(A\,\mathbf{a}+\mathbf{b})^T C\,(A\,\mathbf{a}+\mathbf{b}) = (C + C^T)\,(A\,\mathbf{a}+\mathbf{b})\,\mathbf{a}^T \tag{A.75e}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T A\,\mathbf{x}\,\mathbf{x}^T) = (A + A^T)\,\mathbf{x}\,\mathbf{x}^T + (\mathbf{x}^T A\,\mathbf{x})\,I \tag{A.75f}$$

A list of derivatives involving the inverse of a matrix is given by

$$\frac{d}{dt}(A^{-1}) = -A^{-1}\left[\frac{d}{dt}(A)\right]A^{-1} \tag{A.76a}$$

$$\frac{\partial}{\partial A}(\mathbf{a}^T A^{-1}\mathbf{b}) = -A^{-T}\,\mathbf{a}\,\mathbf{b}^T A^{-T} \tag{A.76b}$$

A list of derivatives involving the trace of a matrix is given by

$$\frac{\partial}{\partial A}\mathrm{Tr}(A) = \frac{\partial}{\partial A}\mathrm{Tr}(A^T) = I \tag{A.77a}$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(A^\alpha) = \alpha\left(A^{\alpha-1}\right)^T \tag{A.77b}$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(C\,A^{-1}B) = -A^{-T}C\,B\,A^{-T} \tag{A.77c}$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(C^T A\,B^T) = \frac{\partial}{\partial A}\mathrm{Tr}(B\,A^T C) = C\,B \tag{A.77d}$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(C\,A\,B\,A^T D) = C^T D^T A\,B^T + D\,C\,A\,B \tag{A.77e}$$

$$\frac{\partial}{\partial A}\mathrm{Tr}(C\,A\,B\,A) = C^T A^T B^T + B^T A^T C^T \tag{A.77f}$$

A list of derivatives involving the determinant of a matrix is given by

$$\frac{\partial}{\partial A}\det(A) = \frac{\partial}{\partial A}\det(A^T) = [\mathrm{adj}(A)]^T \tag{A.78a}$$

$$\frac{\partial}{\partial A}\det(C\,A\,B) = \det(C\,A\,B)A^{-T} \tag{A.78b}$$

$$\frac{\partial}{\partial A}\ln[\det(C\,A\,B)] = A^{-T} \tag{A.78c}$$

$$\frac{\partial}{\partial A}\det(A^\alpha) = \alpha\det(A^\alpha)\,A^{-T} \tag{A.78d}$$

$$\frac{\partial}{\partial A}\ln[\det(A^\alpha)] = \alpha A^{-T} \tag{A.78e}$$

$$\frac{\partial}{\partial A}\det(A^T C\,A) = \det(A^T C\,A)\,(C+C^T)\,A\,(A^T C\,A)^{-1} \tag{A.78f}$$

$$\frac{\partial}{\partial A}\ln[\det(A^T C\,A)] = (C+C^T)\,A\,(A^T C\,A)^{-1} \tag{A.78g}$$

Relations involving the Hessian matrix are given by

$$\frac{\partial^2}{\partial\mathbf{x}\,\partial\mathbf{x}^T}(A\,\mathbf{x}+\mathbf{b})^T C\,(D\,\mathbf{x}+\mathbf{e}) = A^T C\,D + D^T C^T A \tag{A.79a}$$

$$\frac{\partial^2}{\partial\mathbf{x}\,\partial\mathbf{x}^T}(\mathbf{x}^T C\,\mathbf{x}) = C + C^T \tag{A.79b}$$

References

[1] Stewart, G.W., *Introduction to Matrix Computations*, Academic Press, New York, NY, 1973.

[2] Shuster, M.D., "A Survey of Attitude Representations," *Journal of the Astronautical Sciences*, Vol. 41, No. 4, Oct.-Dec. 1993, pp. 439–517.

[3] Tempelman, W., "The Linear Algebra of Cross Product Operations," *Journal of the Astronautical Sciences*, Vol. 36, No. 4, Oct.-Dec. 1988, pp. 447–461.

[4] Golub, G.H. and Van Loan, C.F., *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 3rd ed., 1996.

[5] Zhang, F., *Linear Algebra: Challenging Problems for Students*, The Johns Hopkins University Press, Baltimore, MD, 1996.

[6] Chen, C.T., *Linear System Theory and Design*, Holt, Rinehart and Winston, New York, NY, 1984.

[7] Horn, R.A. and Johnson, C.R., *Matrix Analysis*, Cambridge University Press, Cambridge, MA, 1985.

[8] Nash, J.C., *Compact Numerical Methods for Computers: linear algebra and function minimization*, Adam Hilger Ltd., Bristol, 1979.