APPENDIX A

Vectors and Matrices

Paul D. Groves

This appendix provides a refresher in vector and matrix algebra to support the main body of the book. Introductions to vectors and matrices are followed by descriptions of special matrix types; matrix inversion; vector and matrix calculus; and eigenvalues, eigenvectors and matrix factorization [1–4].

A.1 Introduction to Vectors

A *vector* is a single-dimensional array of single-valued parameters, known as *scalars*. Here scalars are represented as italic and vectors as bold lower case. The scalar components of a vector are denoted by the corresponding italic symbol with a single numerical index and are normally represented together as a bracketed column. Thus,

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \qquad \text{(A.1)}$$

where, in this case, the vector has n components or elements. Vectors may also be represented with an underline, $\underline{a}$, or an arrow, $\vec{a}$, while many authors do not limit them to lower case. Sometimes, it is convenient to represent a vector column on one line. Here, the notation $\mathbf{a} = (a_1, a_2, \cdots, a_n)$ is used. A vector is often used to represent a quantity which has both magnitude and direction; these vectors usually have three components. However, the components of a vector may also be unrelated, with different units. Both types of vector are used here.

Vectors are added and subtracted by adding and subtracting the components:

$$\mathbf{a} = \mathbf{b} + \mathbf{c} \;\Rightarrow\; \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 + c_1 \\ b_2 + c_2 \\ \vdots \\ b_n + c_n \end{pmatrix}. \qquad \text{(A.2)}$$

The corresponding components must have the same units (e.g., meters).

A vector is multiplied by a scalar simply by multiplying each component by that scalar:

$$\mathbf{a} = b\mathbf{c} \;\Rightarrow\; \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} bc_1 \\ bc_2 \\ \vdots \\ bc_n \end{pmatrix}. \tag{A.3}$$

Where two vectors have the same length, a scalar may be obtained by summing the products of the corresponding components. This is known as the *scalar product* or *dot product* and is written as

$$a = \mathbf{b} \cdot \mathbf{c} = \sum_{i=1}^{n} b_i c_i \,. \tag{A.4}$$

Each component product, $b_i c_i$, must have the same units. Scalar products have the properties

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= \mathbf{b} \cdot \mathbf{a} \\ \mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) &= \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c} \end{aligned} . \tag{A.5}$$

Where the scalar product of two vectors is zero, they are said to be orthogonal.

Three-component vectors may also be combined to produce a three-component *vector product* or *cross product*:

$$\mathbf{a} = \mathbf{b} \wedge \mathbf{c} \;\Rightarrow\; \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} b_2c_3 - b_3c_2 \\ b_3c_1 - b_1c_3 \\ b_1c_2 - b_2c_1 \end{pmatrix}, \tag{A.6}$$

where all components within each vector must have the same units. The operator, $\wedge$, is often written as $\times$ or $\otimes$. Vector products have properties

$$\begin{aligned} \mathbf{a} \wedge \mathbf{b} &= -\mathbf{b} \wedge \mathbf{a} \\ \mathbf{a} \wedge (\mathbf{b} \wedge \mathbf{c}) &= (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c} \\ \mathbf{a} \wedge (\mathbf{b} \wedge \mathbf{c}) &\neq (\mathbf{a} \wedge \mathbf{b}) \wedge \mathbf{c} \\ \mathbf{a} \cdot (\mathbf{b} \wedge \mathbf{c}) &= \mathbf{b} \cdot (\mathbf{c} \wedge \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \wedge \mathbf{b}) \\ \mathbf{a} \cdot (\mathbf{a} \wedge \mathbf{b}) &= \mathbf{b} \cdot (\mathbf{a} \wedge \mathbf{b}) = 0 \end{aligned} . \tag{A.7}$$

The magnitude of a vector is simply the square root of the scalar product of the vector with itself. Thus

$$a = |\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} \,. \tag{A.8}$$

A vector with magnitude 1 is known as a unit vector and is commonly denoted **u**, **e**, or **1**. A unit vector may be obtained by dividing a vector by its magnitude. Thus

$$\mathbf{u} = \mathbf{a}/a \;\Rightarrow\; \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} a_1/a \\ a_2/a \\ \vdots \\ a_n/a \end{pmatrix}. \tag{A.9}$$

A three-component unit vector can be used to represent a direction in three-dimensional space. The dot product of two unit vectors gives the angle between their directions:

$$\cos\theta_{ab} = \mathbf{u}_a \cdot \mathbf{u}_b \,. \tag{A.10}$$

The direction of a vector product is perpendicular to both the input vectors, while its magnitude is

$$|\mathbf{a} \wedge \mathbf{b}| = ab\sin\theta_{ab}. \tag{A.11}$$

Thus when **a** and **b** are parallel, their vector product is zero.

Finally, a vector may comprise an array of smaller vectors, known as sub-vectors, or a mixture of sub-vectors and scalars. For example,

$$\mathbf{a} = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}, \qquad \mathbf{a} = \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{pmatrix}, \qquad \mathbf{a} = \begin{pmatrix} \mathbf{b} \\ c \\ d \end{pmatrix}. \tag{A.12}$$

A.2 Introduction to Matrices

A *matrix* is a two-dimensional array of scalars. It is represented as an upper case symbol, which here is also bold. The components are denoted by the corresponding italic letter with two indices, the first representing the row and the second the column. Thus,

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}, \tag{A.13}$$

where m is the number of rows, n is the number of columns, and mn is the number of elements. Where the components of a matrix do not have the same units, each row and column must have associated units with the units of each component being the product of the row and column units.

Matrices are added and subtracting by adding and subtracting the components:

$$\mathbf{A} = \mathbf{B} + \mathbf{C} \;\Rightarrow\; \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} = \begin{pmatrix} B_{11}+C_{11} & B_{12}+C_{12} & \cdots & B_{1n}+C_{1n} \\ B_{21}+C_{21} & B_{22}+C_{22} & \cdots & B_{2n}+C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1}+C_{m1} & B_{m2}+C_{m2} & \cdots & B_{mn}+C_{mn} \end{pmatrix}. \tag{A.14}$$

The corresponding components must have the same units.

Multiplication of two matrices produces the matrix of the scalar products of each row of the left matrix with each column of the right. Thus

$$\mathbf{A} = \mathbf{BC} \;\Rightarrow\; \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{l1} & A_{l2} & \cdots & A_{\mathrm{ln}} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} B_{1j}C_{j1} & \sum_{j=1}^{m} B_{1j}C_{j2} & \cdots & \sum_{j=1}^{m} B_{1j}C_{jn} \\ \sum_{j=1}^{m} B_{2j}C_{j1} & \sum_{j=1}^{m} B_{2j}C_{j2} & \cdots & \sum_{j=1}^{m} B_{2j}C_{jn} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{m} B_{lj}C_{j1} & \sum_{j=1}^{m} B_{lj}C_{j2} & \cdots & \sum_{j=1}^{m} B_{lj}C_{jn} \end{pmatrix}, \tag{A.15}$$

where **B** is an l-row by m-column matrix, **C** is an $m \times n$ matrix, and **A** is an $l \times n$ matrix. This operation may be described as pre-multiplication of **C** by **B** or post-multiplication of **B** by **C**. Matrices may only be multiplied where the number of columns of the left matrix matches the number of rows of the right matrix. Furthermore, each component product $B_{ij}C_{jk}$ must have the same units for a given i and k. A key feature of matrix multiplication is that reversing the order of the matrices produces a different result. In formal terms, matrices do not commute. Thus,

$$\mathbf{AB} \neq \mathbf{BA} . \tag{A.16}$$

Consequently,

$$\exp(\mathbf{A}+\mathbf{B}) \neq \exp(\mathbf{A})\exp(\mathbf{B}) . \tag{A.17}$$

Other matrix multiplication properties are

$$\begin{gathered}(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \\ (\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}\end{gathered} . \tag{A.18}$$

A vector is simply a matrix with only one column. Therefore, the same rules apply for multiplying a vector by a matrix:

$$\mathbf{a} = \mathbf{Bc} \;\Rightarrow\; \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} B_{1j}c_j \\ \sum_{j=1}^{n} B_{2j}c_j \\ \vdots \\ \sum_{j=1}^{n} B_{mj}c_j \end{pmatrix}, \tag{A.19}$$

where **a** is an m-element vector, **B** is an $m \times n$ matrix, and **c** an n-element vector.

The transpose of a matrix, denoted by the superscript T, reverses the rows and columns. Thus,

$$\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{pmatrix}. \tag{A.20}$$

If **A** is an $m \times n$ matrix then $\mathbf{A}^{\mathrm{T}}$ is an $n \times m$ matrix. The transpose of a matrix product has the property

$$(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} . \tag{A.21}$$

The transpose of a vector is a single-row matrix:

$$\mathbf{a}^{\mathrm{T}} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}. \tag{A.22}$$

If a vector is pre-multiplied by the transpose of another vector, the result is the scalar product:

$$\mathbf{a}^{\mathrm{T}}\mathbf{b} = \sum_{i=1}^{n} a_i b_i = \mathbf{a} \cdot \mathbf{b} . \tag{A.23}$$

This is known as the inner product. The outer product of two vectors, which may be of different sizes, is

$$\mathbf{ab}^{\mathrm{T}} = \begin{pmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_mb_1 & a_mb_2 & \cdots & a_mb_n \end{pmatrix}, \tag{A.24}$$

where **a** has m components and **b** n.

Those matrix elements which have the same row and column index are known as diagonal elements. The remaining elements are off-diagonals. The sum of the diagonal elements is known as the trace, Tr. Thus

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{\min(m,n)} A_{ii} . \tag{A.25}$$

Matrices may also be made of smaller matrices, known as sub-matrices. For example,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} . \tag{A.26}$$

All sub-matrices in a given column must contain the same number of columns, while all sub-matrices in a given row must contain the same number of rows.

A.3 Special Matrix Types

Special types of matrix include the following:

- The square matrix, which has the same number of rows and columns,
- The zero matrix, $\mathbf{0}$, where all elements are zero;
- The identity matrix, $\mathbf{I}$, a square matrix which has unit diagonal and zero off-diagonal elements, such that $\mathbf{IA} = \mathbf{A}$;
- The diagonal matrix, which has zero off-diagonal elements;
- The lower triangular matrix, which has zero-off diagonal elements above the diagonal;
- The upper triangular matrix, which has zero-off diagonal elements below the diagonal;
- The symmetric matrix, which is a square matrix reflected about the diagonal such that $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$;
- The skew-symmetric, or antisymmetric, matrix, which has the property $\mathbf{A}^{\mathrm{T}} = -\mathbf{A}$.

A 3×3 skew-symmetric matrix may be used to perform the vector product operation

$$\mathbf{Ab} = [\mathbf{a} \wedge]\, \mathbf{b} = \mathbf{a} \wedge \mathbf{b} , \tag{A.27}$$

where

$$\mathbf{A} = [\mathbf{a} \wedge] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} . \tag{A.28}$$

Powers of the skew-symmetric matrix obey the rule

$$\begin{aligned} [\mathbf{a} \wedge]^{2k+1} &= (-1)^k |\mathbf{a}|^{2k} [\mathbf{a} \wedge] \\ [\mathbf{a} \wedge]^{2k+2} &= (-1)^k |\mathbf{a}|^{2k} [\mathbf{a} \wedge]^2 \end{aligned} . \tag{A.29}$$

Using this, the exponent of a skew-symmetric matrix is

$$\begin{aligned} \exp[\mathbf{a} \wedge] &= \sum_{r=0}^{\infty} \frac{[\mathbf{a} \wedge]^r}{r!} \\ &= \mathbf{I}_3 + \frac{\sin|\mathbf{a}|}{|\mathbf{a}|}[\mathbf{a} \wedge] + \frac{1-\cos|\mathbf{a}|}{|\mathbf{a}|^2}[\mathbf{a} \wedge]^2 \end{aligned} . \tag{A.30}$$

An *orthonormal* or orthogonal matrix has the property

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I} . \tag{A.31}$$

Furthermore, each row and each column forms a unit vector, and any pair of rows and any pair of columns are orthogonal. The product of two orthonormal matrices is also orthonormal as

$$\mathbf{AB}(\mathbf{AB})^{\mathrm{T}} = \mathbf{ABB}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = \mathbf{AIA}^{\mathrm{T}} = \mathbf{AA}^{\mathrm{T}} = \mathbf{I}\,. \tag{A.32}$$

A.4 Matrix Inversion

The *inverse*, or reciprocal, of a matrix, $\mathbf{A}^{-1}$, fulfils the condition

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}\,. \tag{A.33}$$

Thus, for an orthonormal matrix, the inverse and transpose are the same. Inversion takes the place of division for matrix algebra. Inverse matrices have the properties

$$\begin{aligned} \left(\mathbf{A}^{-1}\right)^{-1} &= \mathbf{A} \\ (\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\ \mathbf{A}^{-1}\mathbf{B} &\neq \mathbf{BA}^{-1} \\ \left(\mathbf{A}^{\mathrm{T}}\right)^{-1} &= \left(\mathbf{A}^{-1}\right)^{\mathrm{T}} \\ (b\mathbf{A})^{-1} &= \frac{1}{b}\mathbf{A}^{-1} \end{aligned}\ . \tag{A.34}$$

Not all matrices have an inverse. The matrix must be square and its rows (or columns) all linearly independent of each other. Where a matrix is not square, a pseudo-inverse, $\mathbf{A}^{\mathrm{T}}\left(\mathbf{AA}^{\mathrm{T}}\right)^{-1}$ or $\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$, may be used instead, provided $\mathbf{AA}^{\mathrm{T}}$ or $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ has an inverse.

The inverse of a matrix is given by

$$\mathbf{A}^{-1} = \frac{\mathrm{adj}\mathbf{A}}{|\mathbf{A}|}, \tag{A.35}$$

where adj**A** and |**A**| are, respectively, the adjoint and determinant of **A**. For an $m \times n$ matrix, these are given by

$$\begin{aligned} \mathrm{adj}\mathbf{A} &= \begin{pmatrix} \alpha_{11} & -\alpha_{21} & \cdots & (-1)^{m+1}\alpha_{n1} \\ -\alpha_{12} & \alpha_{22} & \cdots & (-1)^{m+2}\alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1}\alpha_{1m} & (-1)^{n+2}\alpha_{2m} & \cdots & \alpha_{nm} \end{pmatrix}, \\ |\mathbf{A}| &= \sum_{i=1}^{n} (-1)^{r+i} A_{ri}\alpha_{ri} \end{aligned} \tag{A.36}$$

where r is an arbitrary row and α_{ri} is the minor, the determinant of **A** excluding row r and column i. The solution proceeds iteratively, noting that the determinant of a 2 × 2 matrix is $A_{11}A_{22} - A_{21}A_{12}$. Alternatively, many numerical methods for matrix inversion are available.

Where the rows and columns are not linearly independent, the determinant, |**A**|, is zero, so the matrix does not have an inverse. Such a matrix is called singular. Note that the number of linearly-independent rows of a matrix is known as the rank and is also the number of linearly-independent columns.

A.5 Calculus

The derivative of a vector or matrix with respect to a scalar simply comprises the derivatives of the elements. Thus,

$$\frac{d\mathbf{a}}{db} = \begin{pmatrix} da_1/db \\ da_2/db \\ \vdots \\ da_n/db \end{pmatrix}, \qquad \frac{d\mathbf{A}}{db} = \begin{pmatrix} dA_{11}/db & dA_{12}/db & \cdots & dA_{1n}/db \\ dA_{21}/db & dA_{22}/db & \cdots & dA_{2n}/db \\ \vdots & \vdots & \ddots & \vdots \\ dA_{m1}/db & dA_{m2}/db & \cdots & dA_{mn}/db \end{pmatrix}. \qquad \text{(A.37)}$$

The derivative of a scalar with respect to a vector is written as the transposed vector of the partial derivatives with respect to each vector component. Thus

$$\frac{da}{d\mathbf{b}} = \begin{pmatrix} \dfrac{\partial a}{\partial b_1} & \dfrac{\partial a}{\partial b_2} & \cdots & \dfrac{\partial a}{\partial b_n} \end{pmatrix}. \qquad \text{(A.38)}$$

Post-multiplying this by the vector **b** then produces a scalar with the same units as a.

The derivative of one vector with respect to another is then a matrix of the form

$$\frac{d\mathbf{a}}{d\mathbf{b}} = \begin{pmatrix} \partial a_1/\partial b_1 & \partial a_1/\partial b_2 & \cdots & \partial a_1/\partial b_n \\ \partial a_2/\partial b_1 & \partial a_2/\partial b_2 & \cdots & \partial a_2/\partial b_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial a_m/\partial b_1 & \partial a_m/\partial b_2 & \cdots & \partial a_m/\partial b_n \end{pmatrix}. \qquad \text{(A.39)}$$

The derivative of a scalar with respect to a matrix follows a different convention. The components of the derivative simply match the components of the matrix. Thus,

$$\frac{da}{d\mathbf{B}} = \begin{pmatrix} da/dB_{11} & da/dB_{12} & \cdots & da/dB_{1n} \\ da/dB_{21} & da/dB_{22} & \cdots & da/dB_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ da/dB_{m1} & da/dB_{m2} & \cdots & da/dB_{mn} \end{pmatrix} \qquad \text{(A.40)}$$

From this, it may be shown that

$$\frac{\partial}{\partial \mathbf{A}} \mathrm{Tr}(\mathbf{AB}) = \mathbf{B}^{\mathrm{T}}, \qquad \text{(A.41)}$$

where **AB** is symmetric and

$$\frac{\partial}{\partial \mathbf{A}} \mathrm{Tr}(\mathbf{ABA}^{\mathrm{T}}) = 2\mathbf{AB}, \qquad \text{(A.42)}$$

where **B** is symmetric.

From (A.23) and (A.38),

$$\frac{d(\mathbf{a}^{\mathrm{T}}\mathbf{b})}{d\mathbf{b}} = \frac{d(\mathbf{b}^{\mathrm{T}}\mathbf{a})}{d\mathbf{b}} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} = \mathbf{a}^{\mathrm{T}}. \qquad \text{(A.43)}$$

Similarly, as $\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{b} = \sum_{ij} A_{ij} b_i b_j$,

$$\frac{d(\mathbf{b}^{\mathrm{T}}\mathbf{A}\mathbf{b})}{d\mathbf{b}} = \begin{pmatrix} \sum_i (A_{i1} + A_{1i}) b_i & \sum_i (A_{i2} + A_{2i}) b_i & \cdots & \sum_i (A_{in} + A_{ni}) b_i \end{pmatrix} = \mathbf{b}^{\mathrm{T}}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}). \qquad \text{(A.44)}$$

The integral of a function across a vector or matrix is the multiple integral across each component of that vector or matrix. For example,

$$\int f(\mathbf{a})d\mathbf{a} = \int \cdots \int \int f(\mathbf{a})da_1 da_2 \cdots da_n \ . \tag{A.45}$$

A.6 Eigenvalues, Eigenvectors and Matrix Factorization

For an $n \times n$ square matrix, $\mathbf{A}$, there exist n-component unit vectors, $\mathbf{e}$, known as normalized eigenvectors, which are solutions to

$$\mathbf{A}\mathbf{e} = \lambda \mathbf{e}, \qquad \mathbf{e}^{\mathrm{T}}\mathbf{e} = 1, \tag{A.46}$$

where λ is a scalar, known as the eigenvalue. For each non-trivial solution of the normalized eigenvector, there exists a corresponding eigenvalue. Where $\mathbf{A}$ is non-singular, the number of ($\mathbf{e}$, λ) solutions is n. Otherwise, the number of solutions corresponds to the rank of the matrix. The eigenvalues alone may be determined by solving

$$\left|\mathbf{A} - \lambda \mathbf{I}_n\right| = 0, \tag{A.47}$$

simplifying the determination of the eigenvectors using (A.46).

A matrix is positive definite where all eigenvalues are real and positive and the corresponding eigenvectors satisfy $\mathbf{e}^{\mathrm{T}}\mathbf{A}\mathbf{e} > 0$. A matrix is positive semidefinite or non-negative definite when all eigenvalues are real and non-negative and the corresponding eigenvectors satisfy $\mathbf{e}^{\mathrm{T}}\mathbf{A}\mathbf{e} \geq 0$.

The matrices $\mathbf{E}$ and $\mathbf{D}_\lambda$ may be defined as

$$\mathbf{E} = \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix}, \qquad \mathbf{D}_\lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \tag{A.48}$$

where normalized eigenvector λ_i corresponds to eigenvector $\mathbf{e}_i$. Substituting these into (A.46) gives

$$\mathbf{A}\mathbf{E} = \mathbf{E}\mathbf{D}_\lambda \ . \tag{A.49}$$

The transformations from $\mathbf{A}$ to $\mathbf{D}_\lambda$ and $\mathbf{D}_\lambda$ to $\mathbf{A}$ are thus

$$\begin{aligned} \mathbf{D}_\lambda &= \mathbf{E}^{-1}\mathbf{A}\mathbf{E} \\ \mathbf{A} &= \mathbf{E}\mathbf{D}_\lambda\mathbf{E}^{-1} \end{aligned}, \tag{A.50}$$

where the former transformation is known as diagonalization. Where $\mathbf{A}$ is symmetric, the normalized eigenvector matrix, $\mathbf{E}$, is orthonormal and the transformation is known as symmetric QR decomposition.

Another way of diagonalizing a positive semidefinite symmetric matrix is Cholesky factorization:

$$\begin{aligned} \mathbf{D}_U &= \mathbf{U}^{-1}\mathbf{A}\left(\mathbf{U}^{-1}\right)^{\mathrm{T}}, & \mathbf{D}_L &= \mathbf{L}^{-1}\mathbf{A}\left(\mathbf{L}^{-1}\right)^{\mathrm{T}} \\ \mathbf{A} &= \mathbf{U}\mathbf{D}_U\mathbf{U}^{\mathrm{T}}, & \mathbf{A} &= \mathbf{L}\mathbf{D}_L\mathbf{L}^{\mathrm{T}} \end{aligned}, \tag{A.51}$$

where $\mathbf{U}$ is an upper triangular matrix and $\mathbf{L}$ is a lower triangular matrix, both with unit diagonal elements. The process is also known as UDU or LDL factorization. Note that the diagonal matrices, $\mathbf{D}_U$, $\mathbf{D}_L$, and $\mathbf{D}_\lambda$, are all different. The matrices are calculated using

$$
\begin{aligned}
D_{L,11} &= A_{11} \\
L_{11} &= 1 \\
L_{i1} &= \frac{A_{i1}}{D_{L,11}} && \forall i \\
D_{L,jj} &= A_{jj} - \sum_{k=1}^{j-1} L_{jk}^2 D_{L,kk} && j > 1 \\
L_{jj} &= 1 && \forall j \\
L_{ij} &= \frac{1}{D_{L,jj}}\left(A_{ij} - \sum_{k=1}^{i-1} L_{ik} L_{jk} D_{L,kk} \right) && i > j, i > 1 \\
L_{ij} &= 0 && i < j
\end{aligned}
\tag{A.52}
$$

for LDL factorization where $A_{11} \neq 0$, and

$$
\begin{aligned}
D_{U,nn} &= A_{nn} \\
U_{nn} &= 1 \\
U_{in} &= \frac{A_{in}}{D_{U,nn}} && \forall i \\
D_{U,jj} &= A_{jj} - \sum_{k=j+1}^{n} U_{jk}^2 D_{U,kk} && j < n \\
U_{jj} &= 1 && \forall j \\
U_{ij} &= \frac{1}{D_{U,jj}}\left(A_{ij} - \sum_{k=i+1}^{n} U_{ik} U_{jk} D_{U,kk} \right) && i < j, i < n \\
U_{ij} &= 0 && i < j
\end{aligned}
\tag{A.53}
$$

for UDU factorization where $A_{nn} \neq 0$. Each column of $\mathbf{L}$ or $\mathbf{U}$ is calculated after the corresponding diagonal element of $\mathbf{D}_L$ or $\mathbf{D}_U$. For LDL factorization, the calculation starts at column 1 and ends at column n, whereas for UDU factorization, the calculation proceeds in reverse from column n to column 1. One application of these transformations is the decorrelation of a vector of Gaussian-distributed random variables (see Section B.3.2 of Appendix B on CD).

Cholesky factorization can also be used to determine the square root of a positive semidefinite matrix:

$$
\mathbf{A} = \mathbf{U}\mathbf{U}^{\mathrm{T}}, \qquad \mathbf{A} = \mathbf{L}\mathbf{L}^{\mathrm{T}}, \tag{A.54}
$$

where the matrices are calculated using

$$
\begin{aligned}
L_{11} &= \sqrt{A_{11}} \\
L_{i1} &= \frac{A_{i1}}{L_{11}} \\
L_{jj} &= \sqrt{A_{jj} - \sum_{k=1}^{j-1} L_{jk}^2} && j > 1 \\
L_{ij} &= \frac{1}{L_{jj}}\left(A_{ij} - \sum_{k=1}^{i-1} L_{ik} L_{jk} \right) && i > j, i > 1 \\
L_{ij} &= 0 && i < j
\end{aligned}
\tag{A.55}
$$

for lower diagonal factorization and

$$
\begin{aligned}
U_{nn} &= \sqrt{A_{nn}} \\
U_{in} &= \frac{A_{in}}{U_{nn}} \\
U_{jj} &= \sqrt{A_{jj} - \sum_{k=j+1}^{n} U_{jk}^2} && j < n \\
U_{ij} &= \frac{1}{U_{jj}}\left(A_{ij} - \sum_{k=i+1}^{n} U_{ik} U_{jk} \right) && i < j, i < n \\
U_{ij} &= 0 && i < j
\end{aligned}
\tag{A.56}
$$

for upper diagonal factorization. The elements are calculated in the same order as for LDL and UDU factorization.

References

[1] Farrell, J. A., *Aided Navigation: GPS with High Rate Sensors*, New York: McGraw Hill, 2008.

[2] Golub, G. H. and C. F. Van Loan, *Matrix Computations*, Baltimore, MD: Johns Hopkins University Press, 1983.

[3] Grewal, M. S., L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation, and Integration*, 2nd ed., New York: Wiley, 2007.

[4] Stephenson, G., *Mathematical Methods for Science Students*, 2nd ed., London, UK: Longman, 1973.