Chapter 9

Matrix Formulation of the Linear Variation Method

9-1 Introduction

In Chapter 7 we developed a method for performing linear variational calculations. The method requires solving a determinantal equation for its roots, and then solving a set of simultaneous homogeneous equations for coefficients. This procedure is not the most efficient for programmed solution by computer. In this chapter we describe the *matrix* formulation for the linear variation procedure. Not only is this the basis for many quantum-chemical computer programs, but it also provides a convenient framework for formulating the various quantum-chemical methods we shall encounter in future chapters.

Vectors and matrices may be defined in a formal, algebraic way, but they also may be given geometric interpretations. The formal definitions and rules suffice for quantum-chemical purposes. However, the terminology of matrix algebra is closely connected with the geometric ideas that influenced early development. Furthermore, most chemists are more comfortable if they have a physical or geometric model to carry along with mathematical discussion. Therefore, we append some discussion of geometrical interpretation to the algebraic treatment.[1]

9-2 Matrices and Vectors

9-2.A Definitions

A matrix is an ordered array of elements satisfying certain algebraic rules. We write our matrices with parentheses on the left and right of the array.[2] Unless otherwise stated, we restrict the elements to be numbers (which need not be real). In general, however, as long as the rules of matrix algebra can be observed, there is no restriction on what the elements may be. In expression (9-1), we have written a matrix in three ways:

$$\begin{pmatrix} 1.2 & 3.8 & -4.0 \\ 5.0 & 1.0 & 0.0 \\ 9.1 & 0.0 & -3+4i \\ 6.0 & -1.0 & -1.0 \end{pmatrix} \equiv \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \\ c_{41} & c_{42} & c_{43} \end{pmatrix} \equiv \mathsf{C} \tag{9-1}$$

[1] More thorough discussions of matrix algebra at a level suitable for the nonspecialist are given by Aitken [1] and Birkhoff and MacLane [2].

[2] Some authors use brackets.

On the left, the numerical elements are written explicitly. In the center they are symbolized by a subscripted letter. On the right the entire matrix is indicated by a single symbol. We will use sans serif, upper-case symbols to represent matrices.

Individual matrix elements will be symbolized either by a subscripted lower-case symbol (e.g., c_{12}) or by a subscripted symbol for the matrix in parentheses (e.g., $(\mathsf{C})_{12}$).

It is useful to recognize rows and columns in a matrix. The sample matrix given above has four rows and three columns, so it is said to have *dimensions* 4×3. When subscripts are used to denote position in a matrix, the convention is that the first subscript indicates the row, the second indicates the column. (The order "row-column" is important to remember. The mnemonic "RC," or "Roman Catholic" is helpful.) Rows are numbered from top to bottom, columns from left to right. There is no limit on the dimensions for matrices, but we usually will be concerned in a practical way with finite-dimensional matrices in this book.

If a matrix has only one column or row, it is called a *column vector* or *row vector*, respectively. We will use sans serif, lower-case symbols to denote *column* vectors. Additional symbols, described shortly, will be used to denote row vectors. These two kinds of vector behave differently under the rules of matrix algebra, so it is important to avoid confusing them.

If a matrix has only one row and one column, its behavior under the rules of matrix algebra becomes identical to the familiar behavior of ordinary *scalars* (i.e., numbers), and so a 1×1 matrix is simply a number.

The similarity in appearance between a matrix and a determinant may be deceptive. A determinant is denoted by bounding with vertical straight lines, and is equal to a *number* that can be found by reducing the determinant according to a prescribed procedure (see Appendix 2). For this to be possible, the determinant must be square (i.e., have the same number of rows as columns). A matrix is not equal to a number and need not be square. (However, one can take the determinant of a square matrix A. This *number* is symbolized $|\mathsf{A}|$ and is *not* the same as A without the vertical bars.)

Two matrices are *equal* if all elements in corresponding positions are equal. Thus, $\mathsf{A} = \mathsf{B}$ means $a_{ij} = b_{ij}$ for all i and j.

9-2.B Complex Conjugate, Transpose, and Hermitian Adjoint of a Matrix

We define the *complex conjugate* of a matrix A to be the matrix A^*, formed by replacing every element of A by its complex conjugate. If $\mathsf{A} = \mathsf{A}^*$, A is a *real* matrix. (Every element is real.)

We define the *transpose* of a matrix A to be the matrix $\tilde{\mathsf{A}}$, formed by interchanging row 1 and column 1, row 2 and column 2, etc. The transpose of the 4×3 matrix in expression (9-1) is the 3×4 matrix given in

$$\tilde{\mathsf{C}} = \begin{pmatrix} 1.2 & 5.0 & 9.1 & 6.0 \\ 3.8 & 1.0 & 0.0 & -1.0 \\ -4.0 & 0.0 & -3+4i & -1.0 \end{pmatrix} \tag{9-2}$$

If we denote some column vector as p, we can symbolize the corresponding row vector as $\tilde{\mathsf{p}}$. Thus, the tilde symbol is one device we can use to indicate a row vector.

Transposing a square matrix corresponds to "reflecting" it through its *principal diagonal* (which runs from upper left to lower right) as indicated in

$$\mathsf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad \tilde{\mathsf{A}} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \tag{9-3}$$

If $\mathsf{A} = \tilde{\mathsf{A}}$, A is a *symmetric* matrix.

We define the *hermitian adjoint* of A to be the matrix $\mathsf{A}^\dagger$, formed by taking the transpose of the complex conjugate of A (or the complex conjugate of the transpose. The order of these operations is immaterial.) Hence, $\mathsf{A}^\dagger = (\tilde{\mathsf{A}})^* = (\tilde{\mathsf{A}}^*)$. If $\mathsf{A} = \mathsf{A}^\dagger$, A is a *hermitian* matrix.

9-2.C Addition and Multiplication of Matrices and Vectors

Multiplication of a matrix by a scalar is equivalent to multiplying every element in the matrix by the scalar. Addition of two matrices is accomplished by adding elements in corresponding positions in the matrices. Thus, for example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + 2 \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1+10 & 2+12 \\ 3+14 & 4+16 \end{pmatrix} = \begin{pmatrix} 11 & 14 \\ 17 & 20 \end{pmatrix}$$

and it is evident that the operation of matrix addition is possible only within sets of matrices of identical dimensions.

Matrix multiplication is a bit more involved. We start by considering multiplication of vectors. Two types of vector multiplication are possible. If we multiply a row vector on the left times a column vector on the right, we take the *product* of the leading element of each *plus* the product of the second element of each, plus . . . , etc., thereby obtaining a *scalar* as a result. Hence, this is called *scalar* multiplication of vectors. For example,

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = 1 \times 4 + 2 \times 5 + 3 \times 6 = 32$$

This kind of multiplication requires an equal number of elements in the two vectors. It is important to retain in mind the basic operation described here: summation of products taken by sweeping *across a row on the left* and *down a column on the right*. Since there exists but *one* row on the left and *one* column on the right, we obtain *one* number as the result.

The other possibility is to multiply a column vector on the left times a row vector on the right. Employing the same basic operation as above, we sweep across row 1 on the left and down column 1 on the right, obtaining a product that we will store in position (1, 1) of a matrix to keep track of its origin. The product of row 1 times column 2 gives us element (1, 2) and so on. In this way, we generate a whole matrix of numbers. For example,

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 \times 3 & 1 \times 4 & 1 \times 5 \\ 2 \times 3 & 2 \times 4 & 2 \times 5 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \end{pmatrix}$$

This is an example of *matrix* multiplication of vectors. Just as before, the number of columns on the left (one) equals the number of rows on the right. Now, however, the number of elements in the vectors may differ, and the dimensions of the matrix reflect the dimensions of the original vectors.

The two types of vector multiplication may be symbolized as follows, using our notation for scalars, row vectors, column vectors, and matrices:

$$\tilde{\mathsf{a}}\mathsf{b} = c \tag{9-4}$$

$$\mathsf{a}\tilde{\mathsf{b}} = \mathsf{C} \tag{9-5}$$

Multiplying two matrices together is most simply viewed as scalar multiplying all the rows in the left matrix by all the columns of the right matrix. Thus, in $\mathsf{AB} = \mathsf{C}$, the element c_{ij} is the (scalar) product of row i in A times column j in B. This process is possible only when the number of columns in A equals the number of rows in B. Thus, AB may exist as a matrix C, while BA may not exist due to a disagreement in number of rows and columns. If A and B are both square matrices and have equal dimension, then AB and BA both exist, but they still need not be equal. That is, *matrix multiplication is not commutative*.

In a triple product of matrices, ABC, one can multiply AB first (call the result D) and then multiply DC to get the final result (call it F). Or, if one takes $\mathsf{BC} = \mathsf{E}$, then one always finds $\mathsf{AE} = \mathsf{F}$. Thus, the result is invariant to the choice between $(\mathsf{AB})\mathsf{C}$ or $\mathsf{A}(\mathsf{BC})$, and so matrix multiplication is *associative*.

9-2.D Diagonal Matrices, Unit Matrices, and Inverse Matrices

A *diagonal matrix* is a square matrix having zeros everywhere except on the principal diagonal. Diagonal matrices of equal dimension commute with each other, but a diagonal matrix does not, in general, commute with a nondiagonal matrix.

A *unit* matrix $\mathsf{1}$ is a special diagonal matrix. Every diagonal element has a value of unity. A unit matrix times any matrix (of appropriate dimension) gives that same matrix as product. That is, $\mathsf{1A} = \mathsf{A1} = \mathsf{A}$. It follows immediately that the unit matrix commutes with any *square* matrix of the same dimension.

We define the *left inverse* of a matrix A to be A^{-1}, satisfying the matrix equation $\mathsf{A}^{-1}\mathsf{A} = \mathsf{1}$. The right inverse is defined to satisfy $\mathsf{A}\mathsf{A}^{-1} = \mathsf{1}$.

In most of our quantum-chemical applications of matrix algebra, we will be concerned only with vectors and *square* matrices. For square matrices, the left and right inverses are identical, and so we refer simply to the inverse of the matrix.

9-2.E Complex Conjugate, Inverse, and Transpose of a Product of Matrices

If $\mathsf{AB} = \mathsf{C}$, then $\mathsf{C}^* = (\mathsf{AB})^* = \mathsf{A}^*\mathsf{B}^*$. In words, the complex conjugate of a product of matrices is equal to the product of the complex conjugate matrices. This is demonstrable from the observation that $(\mathsf{C})_{ij} = (\mathsf{A})_{i1}(\mathsf{B})_{1j} + (\mathsf{A})_{i2}(\mathsf{B})_{2j} + \cdots$ and $(\mathsf{C})^*_{ij} = (\mathsf{A})^*_{i1}(\mathsf{B})^*_{1j} + (\mathsf{A})^*_{i2}(\mathsf{B})^*_{2j} + \cdots$ and so the complex conjugate is produced by taking

the complex conjugate of every element in A and B but not changing their order of combination.

If $\mathsf{AB} = \mathsf{C}$, then $\mathsf{C}^{-1} = (\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}$. In words, the inverse of a product of matrices is equal to the product of inverses, but *with the order reversed*. We can easily show that this satisfies the rules of matrix algebra. $\mathsf{C}^{-1}\mathsf{C} = (\mathsf{AB})^{-1}\mathsf{AB} = \mathsf{B}^{-1}\mathsf{A}^{-1}\mathsf{AB} = \mathsf{B}^{-1}\mathsf{1B} = \mathsf{B}^{-1}\mathsf{B} = \mathsf{1}$. If we failed to reverse the order, we would instead have $\mathsf{A}^{-1}\mathsf{B}^{-1}\mathsf{AB}$ and, because the matrices do not commute, we would be prevented from carrying through the reduction to $\mathsf{1}$.

If $\mathsf{AB} = \mathsf{C}$, then $\tilde{\mathsf{C}} = (\widetilde{\mathsf{AB}}) = \tilde{\mathsf{B}}\tilde{\mathsf{A}}$. The transpose of a product is the product of transposes, again in reverse order. Since $\tilde{\mathsf{C}}$ has c_{ij} and c_{ji} interchanged, it follows that, where we had row i of A times column j of B, we must now have row j of A times column i of B. But this is the same as column j of $\tilde{\mathsf{A}}$ times row i of $\tilde{\mathsf{B}}$. To obtain row on left and column on right for proper multiplication, we must have $\tilde{\mathsf{B}}\tilde{\mathsf{A}}$.

9-2.F A Geometric Model

Consider a vector in two-dimensional space emanating from the origin of a Cartesian system as indicated in Fig. 9-1a. We can summarize the information contained in this vector (magnitude and direction) by writing down the x and y components of the vector terminus, (3 2) in this case. It must be understood that the first number corresponds to the x component and not the y, and so the vector (3 2) carries its information through number *position* as well as number *value*.

If we multiply both numbers in the vector by 2, the result, (6 4), corresponds to a vector collinear to the original but twice as long. Therefore, multiplying a vector by a *number* results in a change of *scale* but no change in direction. Hence, the term "scalar" is often used in place of "number" in vector terminology.

Suppose that we rotated the Cartesian axes counterclockwise through an angle θ, maintaining them orthogonal to each other and not varying the distance scales. We imagine our original vector to remain unrotated during this *coordinate transformation*. (Equivalently, we can imagine rotating the vector *clockwise* by θ, keeping the axes fixed.) We wish to know how to express our vector in the new coordinate system. The situation is depicted in Fig. 9-1b. Inspection reveals that the new coordinates $(x'\,y')$ are related to the old $(x\ y)$ as follows:

$$x' = x\cos\theta + y\sin\theta, \quad y' = -x\sin\theta + y\cos\theta \tag{9-6}$$

If we make use of matrix algebra, we can express Eqs. (9-6) as a matrix equation:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{9-7}$$

Multiplying the two-dimensional vector (call it v) by the 2×2 matrix (call it R) generates a new vector v', that gives the coordinates of our vector in the new coordinate system:

$$\mathsf{v}' = \mathsf{R}\mathsf{v} \tag{9-8}$$

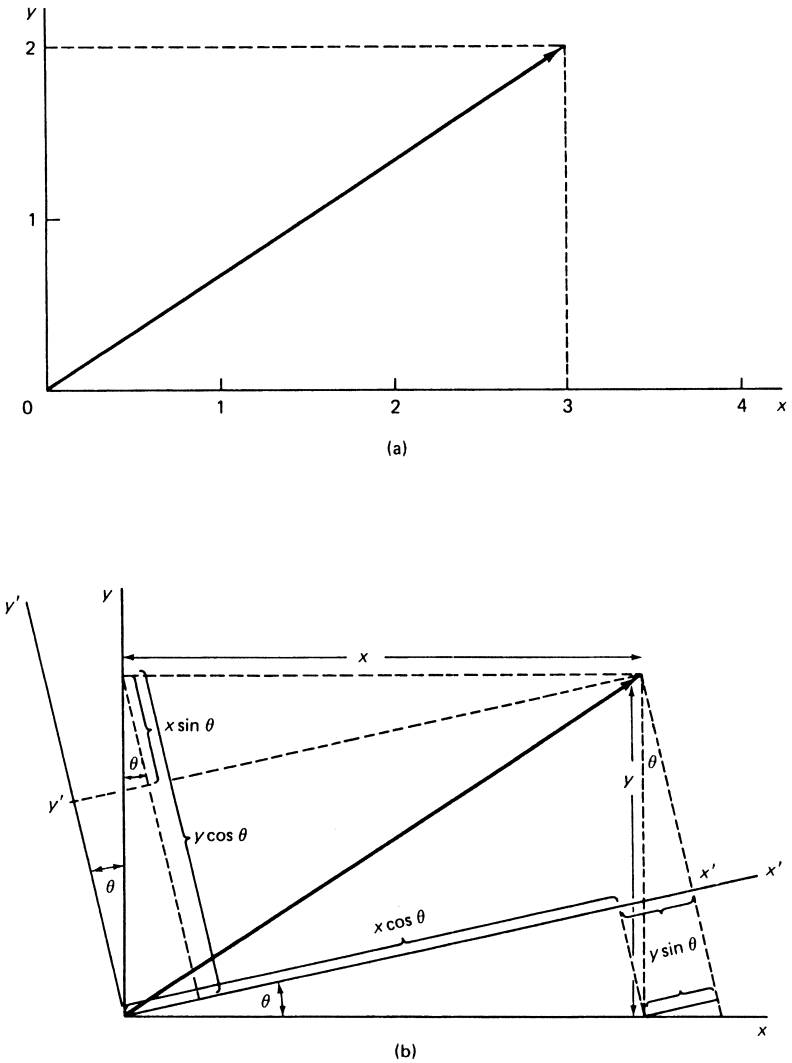

Figure 9-1 ► (a) The vector (3 2). (b) The same vector and its relationship to two Cartesian axis systems.

We have, then, a parallel between the vectors v' and v and matrix R on the one hand, and the two-dimensional "geometrical" vector and rotating coordinate system on the other. R *represents* the rotation and is often referred to as a *rotation* matrix.

If we were to perform the rotation in the opposite direction, the rotation matrix would be the same except for the $\sin\theta$ terms, which would reverse sign:

$$\left.\begin{array}{l}\text{rotation of coordinates}\\ \text{clockwise by } \theta\end{array}\right\} \longrightarrow \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} \tag{9-9}$$

Note that this is just $\tilde{\mathsf{R}}$, the transpose of R.

If we were to rotate counterclockwise by θ and then clockwise by θ, we should end up with our original coordinates for the vector. Thus, we should expect

$$\tilde{\mathsf{R}}\mathsf{R}\mathsf{v} = \mathsf{v} \tag{9-10}$$

or

$$\tilde{\mathsf{R}}\mathsf{R} = \mathsf{1} \tag{9-11}$$

(Note that the *order* of operations is consistent with reading from right to left, just as with differential operators. Thus, $\tilde{\mathsf{R}}\mathsf{R}$ means that first R is performed, then $\tilde{\mathsf{R}}$.) Relation (9-11) is easily verified by explicit multiplication. Thus we see that, if we think of a matrix as representing some coordinate transformation in geometrical space, the inverse of the matrix represents the reverse transformation. In this *particular* example, the *transpose* of the transformation matrix turns out to be the inverse transformation matrix. When this is so, the matrix is said to be *orthogonal*. (Orthogonal transformations do not change the angles between coordinate axes; orthogonal axes remain orthogonal—hence the name "orthogonal.") The analogous transformation for matrices having complex or imaginary coefficients is called a *unitary* transformation. A unitary matrix has its hermitian adjoint as inverse: $\mathsf{A}^\dagger\mathsf{A} = \mathsf{1}$.

While it is easy to visualize a coordinate transformation in two or three dimensions, it is more difficult in higher-dimensional situations. Nevertheless, the mathematics and terminology carry forward to any desired dimension and are very useful. One may, if one wishes, talk of vectors and coordinate transformations in hyperspace, or one can eschew such mental constructs and simply follow the mathematical rules without a mental model.

One must be a bit cautious about inverses of matrices. In the rotation described above, we have a unique way of relating each x, y point in one coordinate system to a point at x', y' in the other. The transformation does not entail any loss of information and can therefore be "undone."

Such transformations (and their matrices) are called *nonsingular*. A nonsingular matrix is recognizable through the fact that its determinant must be nonzero. If we had a transformation which, for example, caused all or some points in one coordinate system to coalesce into a single point in the transformed system, we would lose our ability to back-transform in a unique way. Such a *singular* transformation has no inverse, and the determinant of a singular matrix equals zero.

9-2.G Similarity Transformations

A matrix product of the form $\mathsf{A}^{-1}\mathsf{H}\mathsf{A}$ is called a *similarity transformation* on H. If A is orthogonal, then $\tilde{\mathsf{A}}\mathsf{H}\mathsf{A}$ is a special kind of similarity transformation, called an *orthogonal transformation*. If A is unitary, then $\mathsf{A}^\dagger\mathsf{H}\mathsf{A}$ is a *unitary transformation* on H. There is a physical interpretation for a similarity transformation, which will be discussed in a later chapter. For the present, we are concerned only with the mathematical definition of such a transformation. The important feature is that the eigenvalues, or "latent roots," of H are preserved in such a transformation (see Problem 9-5).

In this section we have quickly presented the salient rules of matrix algebra and hinted at their connection with geometric operations. The results are summarized in Table 9-1 for ease of reference.

TABLE 9-1 ► Some Matrix Rules and Definitions for a Square Matrix A of Dimension n

$\mathsf{A}=\mathsf{B}$	Matrix equality; means $a_{ij}=b_{ij}, \quad i,j=1,n$
$\mathsf{A}+\mathsf{B}=\mathsf{C}$	Matrix addition; $c_{ij}=a_{ij}+b_{ij}, \quad i,j=1,n$
$c\mathsf{A}=\mathsf{B}$	Multiplication of A by scalar; $b_{ij}=c\cdot a_{ij}, \quad i,j=1,n$
$\mathsf{AB}=\mathsf{C}$	Matrix multiplication; $c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj}, \quad i,j=1,n$
$\lvert\mathsf{A}\rvert$	The determinant of the matrix A (see Appendix 2)
A^{-1}	The inverse of A; $\mathsf{A}^{-1}\mathsf{A}=\mathsf{A}\mathsf{A}^{-1}=1$ If A^{-1} exists, A is *nonsingular* and $\lvert\mathsf{A}\rvert\neq 0$.
A^{*}	The complex conjugate of A; $a_{ij}\rightarrow a_{ij}^{*}, \quad i,j=1,n$ If $\mathsf{A}^{*}=\mathsf{A}$, A is *real*.
$\tilde{\mathsf{A}}$	The transpose of A; $(\tilde{\mathsf{A}})_{ij}=a_{ji}$ (rows and columns inter changed) If $\tilde{\mathsf{A}}=\mathsf{A}$, symmetric; if $\tilde{\mathsf{A}}=-\mathsf{A}$, antisymmetric; if $\tilde{\mathsf{A}}=\mathsf{A}^{-1}$, orthogonal.
$\mathsf{A}^{\dagger}$	The hermitian adjoint of A; $(\mathsf{A}^{\dagger})_{ij}=a_{ji}^{*}(\mathsf{A}^{\dagger}=\tilde{\mathsf{A}}^{*})$ If $\mathsf{A}^{\dagger}=\mathsf{A}$, hermitian. If $\mathsf{A}^{\dagger}=\mathsf{A}^{-1}$, unitary.
$(\mathsf{ABC})^{*}=\mathsf{A}^{*}\mathsf{B}^{*}\mathsf{C}^{*}$	Complex conjugate of product
$\widetilde{(\mathsf{ABC})}=\tilde{\mathsf{C}}\tilde{\mathsf{B}}\tilde{\mathsf{A}}$	Transpose of product
$(\mathsf{ABC})^{\dagger}=\mathsf{C}^{\dagger}\mathsf{B}^{\dagger}\mathsf{A}^{\dagger}$	Hermitian adjoint of product
$(\mathsf{ABC})^{-1}=\mathsf{C}^{-1}\mathsf{B}^{-1}\mathsf{A}^{-1}$	Inverse of product
$\lvert\mathsf{ABC}\rvert=\lvert\mathsf{A}\rvert\cdot\lvert\mathsf{B}\rvert\cdot\lvert\mathsf{C}\rvert$	Determinant of product (any order)
$\mathsf{T}^{-1}\mathsf{AT}$	A similarity transformation If $\mathsf{T}^{-1}=\mathsf{T}^{\dagger}$, this is a *unitary* transformation. If $\mathsf{T}^{-1}=\tilde{\mathsf{T}}$, this is an *orthogonal* transformation.

9-3 Matrix Formulation of the Linear Variation Method

We have seen that the independent-electron approximation leads to a series of MOs for a molecular system. If the MOs are expressed as a linear combination of n basis functions (which are often approximations to AOs, although this is not necessary), the variation method leads to a set of simultaneous equations:

$$\begin{aligned}(H_{11}-ES_{11})c_1+(H_{12}-ES_{12})c_2+\cdots+(H_{1n}-ES_{1n})c_n &= 0\\ &\vdots\\ (H_{n1}-ES_{n1})c_1+\cdots+(H_{nn}-ES_{nn})c_n &= 0\end{aligned} \qquad (9\text{-}12)$$

All terms have been defined in Chapter 7. Given a value for E that satisfies the associated *determinantal* equations, we can solve this set of simultaneous equations for ratios between the c_i's. Requiring MO normality establishes convenient numerical values for the c_i's.

A matrix equation equivalent to Eq. (9-12) is[3]

$$\begin{pmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} \cdots & H_{1n} - ES_{1n} \\ \vdots & & \\ H_{n1} - ES_{n1} & \cdots & H_{nn} - ES_{nn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{9-13}$$

The matrix in Eq. (9-13) is clearly the difference between two matrices. This enables us to rewrite the equation in the form

$$\begin{pmatrix} H_{11} & H_{12} \cdots & H_{1n} \\ \vdots & & \vdots \\ H_{n1} & \cdots & H_{nn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = E \begin{pmatrix} S_{11} & S_{12} \cdots & S_{1n} \\ \vdots & & \vdots \\ S_{n1} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \tag{9-14}$$

or

$$\mathsf{H}\mathsf{c}_i = E_i \mathsf{S}\mathsf{c}_i, \quad i = 1, 2, \ldots, n \tag{9-15}$$

where we have introduced the subscript i to account for the fact that there are many possible values for E and that each one has its own characteristic set of coefficients. Note that the "*eigenvector*" c_i is a *column* vector and that each element in c_i is (effectively) multiplied by the scalar E_i according to Eq. (9-15).

In general, there are as many MOs as there are basis functions, and so Eq. (9-15) represents n separate matrix equations. We can continue to use matrix notation to reduce these to a single matrix equation. We do this by stacking the n c vectors together, side by side, to produce an $n \times n$ matrix C. The numbers E must also be combined into an appropriate matrix form. We must be careful to do this in such a way that the scalar E_1 still multiplies only c_1 (now column 1 of C) E_2 multiplies only c_2, and so forth. This is accomplished in the following equation

$$\begin{pmatrix} H_{11} \cdots & H_{1n} \\ \vdots & \vdots \\ H_{n1} \cdots & H_{nn} \end{pmatrix} \begin{pmatrix} c_{11} \cdots & c_{1n} \\ \vdots & \vdots \\ c_{n1} \cdots & c_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} S_{11} \cdots & S_{1n} \\ \vdots & \vdots \\ S_{n1} \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} c_{11} \cdots & c_{1n} \\ \vdots & \vdots \\ c_{n1} \cdots & c_{nn} \end{pmatrix} \begin{pmatrix} E_1 & 0 & 0 & \cdots & 0 \\ 0 & E_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & E_n \end{pmatrix} \tag{9-16}$$

or

$$\mathsf{H}\mathsf{C} = \mathsf{S}\mathsf{C}\mathsf{E} \tag{9-17}$$

The matrix E is a diagonal matrix of orbital energies (often referred to as the *matrix* of *eigenvalues*). C is the matrix of coefficients (or *matrix* of *eigenvectors*), and each

[3] Quantum-chemical convention is to use upper case letters for individual elements of the matrices H, S, and E. This differs from the usual convention.

column in C refers to a different MO. The first column refers to the MO having energy E_1. In multiplying E by C from the left, each coefficient in column $\mathsf{1}$ becomes multiplied by E_1. This would not occur if we multiplied E by C_1 from the right. Therefore, $\mathsf{HC} = \mathsf{SCE}$ is correct, whereas $\mathsf{HC} = \mathsf{ESC}$ is incorrect.

9-4 Solving the Matrix Equation

Since we know the basis functions and the effective hamiltonian (in principle, at least), we are in a position to evaluate the elements in H and S. How do we then find C and E?

Let us first treat the simplified situation where our basis set of functions is orthonormal, either by assumption or design. Then all the off-diagonal elements of S (which correspond to overlap between *different* basis functions) are zero, and all the diagonal elements are unity because of normality. In short, $\mathsf{S} = \mathsf{1}$. Therefore, Eq. (9-17) becomes

$$\mathsf{HC} = \mathsf{CE} \tag{9-18}$$

and our problem is, given H, find C and E.

Now, we want a set of coefficients that correspond to normalized MOs. We have seen earlier that, for an orthonormal basis set, this requires each MO to have coefficients satisfying the equation (assuming real coefficients)

$$c_{1i}^2 + c_{2i}^2 + \cdots + c_{ni}^2 = 1 \tag{9-19}$$

We can write this as a vector equation

$$\tilde{\mathsf{c}}_i \mathsf{c}_i \equiv \begin{pmatrix} c_{i1} & c_{i2} & \cdots & c_{in} \end{pmatrix} \begin{pmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{ni} \end{pmatrix} = 1 \tag{9-20}$$

Furthermore, we know that any two different MOs must be orthogonal to each other. That is $\tilde{\mathsf{c}}_i \mathsf{c}_j = 0$, $i \neq j$. All this may be summarized in the matrix equation

$$\tilde{\mathsf{C}}\mathsf{C} = \mathsf{1} \tag{9-21}$$

Hence, the coefficient matrix is *orthogonal*. In the more general case in which coefficients may be complex, C is *unitary*; i.e., $\mathsf{C}^\dagger\mathsf{C} = \mathsf{1}$. Our problem then is, given H, find a unitary matrix C such that $\mathsf{HC} = \mathsf{CE}$ with E diagonal.

We can multiply both sides of a matrix equation by the same matrix and preserve the equality. However, because matrices do not necessarily commute, we must be careful to carry out the multiplication from the left on both sides, or from the right on both sides. Thus, multiplying Eq. (9-18) from the left by $\mathsf{C}^\dagger$, we obtain

$$\mathsf{C}^\dagger\mathsf{HC} = \mathsf{C}^\dagger\mathsf{CE} = \mathsf{1E} = \mathsf{E} \tag{9-22}$$

where we have used the fact that C is unitary. Now our problem may be stated as, given H, find a unitary matrix C such that $\mathsf{C}^\dagger\mathsf{HC}$ is diagonal.[4] Several techniques

[4] Not every matrix (not even every square matrix) can be diagonalized by a unitary transformation, but every *hermitian* matrix can be so diagonalized.

exist for finding such a matrix C. These are generally much more suitable for machine computation than are determinantal manipulations.

We can illustrate that the allyl radical energies and coefficients already found by the HMO method do in fact satisfy the relations $\mathsf{C}^\dagger\mathsf{C} = \mathsf{1}$ and $\mathsf{C}^\dagger\mathsf{H}\mathsf{C} = \mathsf{E}$. The matrix C can be constructed from the HMO coefficients and is

$$\mathsf{C} = \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix} \tag{9-23}$$

Therefore

$$\begin{aligned} \mathsf{C}^\dagger\mathsf{C} &= \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathsf{1} \end{aligned} \tag{9-24}$$

The matrix H for the allyl radical is, in HMO theory,

$$\mathsf{H} = \begin{pmatrix} \alpha & \beta & 0 \\ \beta & \alpha & \beta \\ 0 & \beta & \alpha \end{pmatrix} \tag{9-25}$$

The reader should verify that

$$\mathsf{C}^\dagger\mathsf{H}\mathsf{C} = \begin{pmatrix} \alpha + \sqrt{2}\beta & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha - \sqrt{2}\beta \end{pmatrix} = \mathsf{E} \tag{9-26}$$

The diagonal elements can be seen to correspond to the HMO energies. Note that the energy in the (1 1) position of E corresponds to the MO with coefficients appearing in *column* 1 of C, illustrating the positional correlation of eigenvalues and eigenvectors referred to earlier.

If the basis functions are not orthogonal, $\mathsf{S} \neq 1$ and the procedure is slightly more complicated. Basically, one first transforms to an *orthogonal* basis to obtain an equation of the form $\mathsf{H}'\mathsf{C}' = \mathsf{C}'\mathsf{E}$. One diagonalizes H' as indicated above to find E and C', where C' is the matrix of coefficients in the *orthogonalized* basis. Then one back transforms C' into the original basis set to obtain C. There are many choices available for the orthogonalizing transformation. The Schmidt transformation, based on the Schmidt orthogonalization procedure described in Chapter 6, is popular because it is very rapidly performed by a computer. Here we will simply indicate the matrix algebra involved. Let the matrix that transforms a nonorthonormal basis to an orthonormal one be symbolized A. This matrix satisfies the relation

$$\mathsf{A}^\dagger\mathsf{S}\mathsf{A} = \mathsf{1} \tag{9-27}$$

Furthermore, $|\mathbf{A}| \neq 0$ and so $\mathbf{A}^{-1}$ exists. We can insert the unit matrix (in the form $\mathbf{A}\mathbf{A}^{-1}$) wherever we please in the matrix equation $\mathbf{HC} = \mathbf{SCE}$ without affecting the equality. Thus,

$$\mathbf{HAA}^{-1}\mathbf{C} = \mathbf{SAA}^{-1}\mathbf{CE} \tag{9-28}$$

Multiplying from the left by $\mathbf{A}^{\dagger}$ gives

$$\mathbf{A}^{\dagger}\mathbf{HAA}^{-1}\mathbf{C} = \mathbf{A}^{\dagger}\mathbf{SAA}^{-1}\mathbf{CE} \tag{9-29}$$

By Eq. (9-27), this reduces to

$$(\mathbf{A}^{\dagger}\mathbf{HA})(\mathbf{A}^{-1}\mathbf{C}) = (\mathbf{A}^{-1}\mathbf{C})\mathbf{E} \tag{9-30}$$

where the parentheses serve only to make the following discussion clearer. If we *define* $\mathbf{A}^{\dagger}\mathbf{HA}$ to be $\mathbf{H}'$, and $\mathbf{A}^{-1}\mathbf{C}$ to be $\mathbf{C}'$, Eq. (9-30) becomes

$$\mathbf{H}'\mathbf{C}' = \mathbf{C}'\mathbf{E} \tag{9-31}$$

Since we know the matrix $\mathbf{H}$ and can compute $\mathbf{A}$ from knowledge of $\mathbf{S}$, it is possible to write down an *explicit* $\mathbf{H}'$ matrix for a given problem. Then, knowing $\mathbf{H}'$ (which is just the *hamiltonian matrix* for the problem in the orthonormal basis), we can seek the unitary matrix $\mathbf{C}'$ such that $\mathbf{C}'^{\dagger}\mathbf{H}'\mathbf{C}'$ is diagonal. These diagonal elements are our orbital energies. (Note that $\mathbf{E}$ in Eq. (9-31) is the same as $\mathbf{E}$ in $\mathbf{HC} = \mathbf{SCE}$.) To find the coefficients for the MOs in terms of the *original* basis (i.e., to find $\mathbf{C}$), we use the relation

$$\mathbf{AC}' = \mathbf{A}(\mathbf{A}^{-1}\mathbf{C}) = \mathbf{1C} = \mathbf{C} \tag{9-32}$$

One nice feature of this procedure is that, even though we use the inverse matrix $\mathbf{A}^{-1}$ in our *formal* development, we never need to actually compute it. ($\mathbf{A}$ and $\mathbf{A}^{\dagger}$ are used to find $\mathbf{H}'$, and $\mathbf{A}$ is used to find $\mathbf{C}$.) This is fortunate because calculating inverse matrices is a relatively slow process.

A few more words should be said about the process of diagonalizing a hermitian matrix $\mathbf{H}$ with a unitary transformation. Two methods are currently in wide use. The older, slower method, known as the *Jacobi* method, requires a series of steps on the starting matrix. In the first step, a matrix $\mathbf{O}_1$ is constructed that causes the largest off-diagonal pair of elements of $\mathbf{H}$ to vanish in the transformation $\mathbf{H}_1 = \tilde{\mathbf{O}}_1\mathbf{HO}_1$. Now a second transformation matrix $\mathbf{O}_2$ is constructed to force the largest off-diagonal pair of elements in $\mathbf{H}_1$ to vanish in the transformation $\tilde{\mathbf{O}}_2\mathbf{H}_1\mathbf{O}_2 = \tilde{\mathbf{O}}_2\tilde{\mathbf{O}}_1\mathbf{HO}_1\mathbf{O}_2$. This procedure is continued. However, since each transformation affects more elements in the matrix than just the biggest pair, we eventually "unzero" the pair that was zeroed in forming $\mathbf{H}_1$ or $\mathbf{H}_2$, etc. This means that many more transformations are required than there are off-diagonal pairs. Eventually, however, the off-diagonal elements will have been nibbled away (while the diagonal elements have been building up) until they are all smaller in magnitude than some preselected value, and so we stop the process. The transformation matrix $\mathbf{C}$ corresponds to the accumulated product $\mathbf{O}_1\mathbf{O}_2\mathbf{O}_3\ldots$.

A more recently discovered, faster procedure is the Givens– Householder–Wilkinson method. Here, H is first *tridiagonalized*, which means that all elements are made to vanish except those on the main diagonal *as well as on the codiagonals above and below the main diagonal*. This similarity transformation can be done in a few steps, each step zeroing all the necessary elements in an entire row and column. The eigenvalues for the tridiagonal matrix (and hence for the original matrix) may be found one at a time as desired. If only the third lowest eigenvalue is of interest, that one alone can be computed. This is a useful degree of freedom which results in substantial savings of time. Once an eigenvalue is found, its corresponding eigenvector may be computed.

9-5 Summary

The steps to be performed in a matrix solution for a linear variation calculation are:

1. From the basis set, calculate the overlap matrix S.

2. From the basis set and hamiltonian operator, calculate the hamiltonian matrix H.

3. If $\mathsf{S} \neq \mathsf{1}$, find an orthogonalization procedure. In the Schmidt method, A is such that $\mathsf{A}^\dagger\mathsf{SA} = \mathsf{1}$. The matrix equation may now be written in the form $\mathsf{H}'\mathsf{C}' = \mathsf{C}'\mathsf{E}$.

4. Find C' such that $\mathsf{C}'^\dagger\mathsf{H}'\mathsf{C}'$ is a diagonal matrix. The diagonal elements are the roots E.

5. If necessary, back transform: $\mathsf{AC}' = \mathsf{C}$. The columns of C contain the MO coefficients appropriate for the original basis set.

9-5.A Problems

9-1. Evaluate the following according to the rules of matrix algebra:

a) $\begin{pmatrix} 6 & 7 & 8 \end{pmatrix} \begin{pmatrix} 9 \\ 10 \\ 11 \end{pmatrix}$

b) $\begin{pmatrix} 6 \\ 7 \end{pmatrix} \begin{pmatrix} \mathrm{a} & \mathrm{b} & \mathrm{c} \end{pmatrix}$

c) $\begin{pmatrix} 4 & 6 \\ i & -3 \end{pmatrix} + 7 \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}$

d) $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$

e) $\begin{pmatrix} i & 4 \\ 1 & 7 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 4 & 7 \end{pmatrix}$

f) $\begin{pmatrix} 3 & 2 \\ 4 & 7 \end{pmatrix} \begin{pmatrix} i & 4 \\ 1 & 7 \\ 0 & -3 \end{pmatrix}$

g) $\begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix}$

9-2. If $H_{ij} = \int \chi_i^* \hat{H} \chi_j \, d\tau$ and $\hat{H}$ is hermitian, show that H is a hermitian matrix.

9-3. Let

$$\mathsf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

Show that, in general, $\mathsf{AB} \neq \mathsf{BA}$.

9-4. Let

$$\mathsf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad \mathsf{B} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 6 \end{pmatrix}, \quad \text{and } \mathsf{C} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Show that $\mathsf{AB} = \mathsf{BA}$, but $\mathsf{AC} \neq \mathsf{CA}$. Compare the matrix AC with CA. Do these matrices show any simple relationship? Can you relate this to properties of A and C mathematically?

9-5. The "latent roots" λ_i of A are solutions to the equation $|\mathsf{A} - \lambda_i \mathsf{1}| = 0$, $i = 1, 2, \ldots, n$, where n is the dimension of A.

a) Show that, under a similarity transformation $\mathsf{B} = \mathsf{T}^{-1}\mathsf{AT}$, the latent roots are preserved.
b) Demonstrate that diagonalization of A *via* a similarity transformation produces the latent roots as the diagonal elements.

9-6. Show that, if a matrix has any latent roots equal to zero, it has no inverse.

9-7. The *trace* (or *spur*) of a matrix is the sum of the elements on the principal diagonal. Thus, tr $\mathsf{A} = \Sigma_{i=1}^{n} a_{ii}$.

a) Show that the trace of a triple product of matrices is invariant under cyclic permutation. That is, $\mathrm{tr}(\mathsf{ABC}) = \mathrm{tr}(\mathsf{CAB}) = \mathrm{tr}(\mathsf{BCA})$ but not $\mathrm{tr}(\mathsf{CBA})$.
b) Show that the trace of a matrix is invariant under a similarity transformation.

9-8. The *norm* of a matrix is the positive square root of the sum of the absolute squares of all the elements.

For a real matrix A,

$$\text{norm } \mathsf{A} = \left[\sum_{i,j=1}^{n} a_{ij}^2 \right]^{1/2} = \left[\sum_i \sum_j \left(\tilde{\mathsf{A}} \right)_{i,j} (\mathsf{A})_{j,i} \right]^{1/2}$$

Prove that the norm of a real matrix is preserved in an orthogonal transformation (or, you may prefer to prove that the norm of any matrix is preserved in a unitary transformation).

9-9. Use the facts that the trace, the determinant, and the norm of a matrix are invariant under an orthogonal transformation to find the eigenvalues of the following matrices:

a) $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

b) $\begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix}$

c) $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

9-10. Consider the matrix

$$\begin{pmatrix} \cos\theta & 0 \\ -\sin\theta & 0 \end{pmatrix}$$

What is the effect of this transformation on $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$? On $\begin{pmatrix} 3 \\ 3 \end{pmatrix}$? Can the transformation be uniquely reversed? (That is, for, say, $\theta = 0$, and given a transformed vector $\begin{pmatrix} 3 \\ 0 \end{pmatrix}$, can one uniquely determine the vector this was transformed from?) Does the matrix have an inverse? Evaluate its determinant.

9-11. What are the eigenvectors for the matrix

$$\mathsf{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

9-12. Show that, if A and B have "simultaneous eigenvectors" (i.e., both diagonalized by the same similarity transformation), then A and B commute.

9-13. If $\mathsf{HC} = \mathsf{CE}$, and $\mathsf{C}^\dagger\mathsf{C} = \mathsf{1}$, then $\mathsf{C}^\dagger\mathsf{HC} = \mathsf{E}$, and we seek a unitary transformation that diagonalizes H. If $\mathsf{HC} = \mathsf{SCE}$, and $\mathsf{C}^\dagger\mathsf{SC} = \mathsf{1}$, then $\mathsf{C}^\dagger\mathsf{HC} = \mathsf{C}^\dagger\mathsf{SCE} = \mathsf{1E}$, and $\mathsf{C}^\dagger\mathsf{HC} = \mathsf{E}$. Since this is the same working equation as the one we found above, why do we not proceed in the same way? Why do we bother orthogonalizing our basis first?

9-14. We have mentioned that a matrix may be used to represent the rotation of coordinates by some angle θ. Such a rotation is a geometric *operation*, so we have,

in effect, represented an *operator* with a matrix. It is possible to represent other operators in a similar way. Indeed, an alternative approach to quantum mechanics exists in which the whole formalism is based on matrices and their properties (matrix mechanics, as opposed to wave mechanics). A particularly interesting example is provided by the matrices constructed by Pauli to represent spin operators and functions. It was mentioned in Chapter 5 that spin functions α and β satisfy rules similar to those for orbital angular momentum. Two of these are

$$\hat{S}_z\alpha = \frac{1}{2}\alpha, \quad \hat{S}_z\beta = -\frac{1}{2}\beta$$

But it was pointed out that α and β could not be expressed in terms of spherical harmonics. Pauli represented this operator and functions by

$$\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \hat{S}_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Using these definitions, show that

$$\int \alpha^\dagger \beta \, d\omega = \int \beta^\dagger \alpha \, d\omega = 0, \quad \int \alpha^\dagger \alpha \, d\omega = \int \beta^\dagger \beta \, d\omega = 1,$$
$$\hat{S}_z\alpha = \frac{1}{2}\alpha, \quad \hat{S}_z\beta = -\frac{1}{2}\beta$$

[*Note*: since α and β are essentially the Dirac delta functions in the spin coordinate ω, the process of integration reduces here to scalar multiplication of vectors.]

References

[1] A. C. Aitken, *Determinants and Matrices*, 9th ed. Wiley (Interscience), New York, 1958.

[2] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 5th ed. Macmillan, New York, 1995.