Notes on Elementary Linear Algebra

C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology
Revised August 2005

1 Vectors

In quantum mechanics, we are generally interested in complex numbers. Let $\mathbf{C}^n$ denote the set of all n-tuples of complex numbers (a *complex n-space*). The elements of $\mathbf{C}^n$ may be called "points" or "vectors" in complex n-space. The elements of the complex numbers $\mathbf{C}$ are *scalars*.

A vector might be denoted by listing all its elements as $(c_1 c_2 c_3 \cdots c_n)$. This format is called a *row vector*. Alternatively, the elements could be arranged as

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{pmatrix},$$

which is a *column vector*. Our default choice will be to write vectors as column vectors, and row vectors will result from taking (complex conjugate) transposes of vectors.

When a vector is multiplied by a scalar (a number), this is the same as multiplying each component of the vector by that scalar, e.g.,

$$2\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 6 \end{pmatrix}. \tag{1}$$

For real numbers, the dot (or inner) product between two vectors $\mathbf{a}$ and $\mathbf{b}$ is given by

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= (a_1 a_2 \cdots a_n) \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \\ &= a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \end{aligned} \tag{2}$$

If $\mathbf{a}$ and $\mathbf{b}$ are complex, then one small complication arises. If $\mathbf{a}$ and $\mathbf{b}$ are given by

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \tag{3}$$

then to take the dot product of $\mathbf{a}$ and $\mathbf{b}$, we must first take the *complex* transpose of $\mathbf{a}$,

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= (a_1^* a_2^* \cdots a_n^*) \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \\ &= a_1^* b_1 + a_2^* b_2 + \cdots + a_n^* b_n. \end{aligned} \tag{4}$$

Recall that the complex conjugate of a complex number like $c + id$ is $c - id$ (the imaginary part has its sign reversed). Any time a row vector is converted into a column vector (or vice versa), if it is complex, then one needs to take complex conjugates of each element.

$$(a_1^* a_2^* \cdots a_n^*) \leftrightarrow \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{5}$$

A vector $\mathbf{a}$ is said to be *normalized* if its dot product with itself, $\mathbf{a} \cdot \mathbf{a}$, is 1. If this is not the case, it is always possible to normalize the vector to force the dot product to come out to 1; one merely needs to divide each component of the vector by the square root of the dot product that results before normalization. For example, the vector

$$\mathbf{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{6}$$

is not normalized, because its dot product with itself is $1 \cdot 1 + 1 \cdot 1 = 2$. However, if we divide the vector by $\sqrt{2}$, then the resulting vector

$$\mathbf{a}' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{7}$$

is normalized:

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} (1 \cdot 1 + 1 \cdot 1) = \frac{1}{2} \cdot 2 = 1. \tag{8}$$

Two vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be *orthogonal* if their dot product $\mathbf{a} \cdot \mathbf{b} = 0$. Two vectors are *orthonormal* if they are orthogonal and each one is normalized.

2 Matrices

A matrix is a rectangular collection of numbers (again, possibly complex). It might represent a collection of row/column vectors, or a transformation that changes a vector into a different vector. For example, the rotation of a real 2D vector by an angle θ about the z axis could be represented as a 2x2 matrix,

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

The effect of this rotation on a 2D vector in the x, y plane can be determined by multiplying the vector by this transformation matrix,

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

where $x' = xcos\theta - ysin\theta$ and $y' = xsin\theta + ycos\theta$.

More generally, the product of a matrix and a column vector may be written as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \tag{10}$$

where

$$\begin{aligned} c_1 &= a_{11}b_1 + a_{12}b_2 + \cdots + a_{1n}b_n \qquad (11) \\ c_2 &= a_{21}b_1 + a_{22}b_2 + \cdots + a_{2n}b_n \\ &\vdots \\ c_n &= a_{n1}b_1 + a_{n2}b_2 + \cdots + a_{nn}b_n. \end{aligned}$$

Notice that each element of the product, c_i, is just the dot product of the ith row of the matrix with the vector. It is also possible to think of the column vector **c** as being a linear combination of the columns of the matrix $\mathbf{A}$, with each column i having a weight b_i.

It is also possible to "left-multiply" a matrix by a row vector, like this:

$$(b_1 b_2 \cdots b_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = (c_1 c_2 \cdots c_n), \tag{12}$$

where

$$\begin{aligned} c_1 &= b_1 a_{11} + b_2 a_{21} + \cdots + b_n a_{n1} \\ c_2 &= b_1 a_{12} + b_2 a_{22} + \cdots + b_n a_{n2} \\ &\vdots \\ c_n &= b_1 a_{1n} + b_2 a_{2n} + \cdots + b_n a_{nn}. \end{aligned} \tag{13}$$

This time, each element of the product, c_i, is the dot product of the vector on the left with the ith column of the matrix. The row vector $\mathbf{c}$ may be thought of as a linear combination of the rows of the matrix $\mathbf{A}$, with each row i having a weight of b_i.

From the information given, it should be obvious that a row vector times a matrix times a column vector yields a number (scalar).

Now consider the multiplication of two matrices $\mathbf{A}$ and $\mathbf{B}$, which yields a new matrix, $\mathbf{C}$:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}, \tag{14}$$

where

$$c_{ij} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}. \tag{15}$$

That is, each element c_{ij} of the product matrix $\mathbf{C}$ is the result of a dot product between row i of the matrix $\mathbf{A}$ and column j of the matrix $\mathbf{B}$. It is possible to think of matrix multiplication as a generalization of the product of a matrix times a vector, where now instead of one column vector on the right, we have a series of them. It is also possible to think of matrix multiplication as a generalization of the product of a row vector times a matrix, where now instead of one row vector on the left, we have a series of them.

The *trace* of a matrix, denoted Tr($\mathbf{A}$), is just the sum of the diagonal elements of a matrix, i.e.,

$$\mathrm{Tr}(\mathbf{A}) = \sum_i^n a_{ii} \tag{16}$$

for an n by n matrix.

The *transpose* of a matrix $\mathbf{A}$, typically denoted $\mathbf{A}^T$, is obtained simply by swapping elements across the diagonal, $a_{ij} \rightarrow a_{ji}$. For example,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}. \tag{17}$$

For complex numbers, usually the *complex conjugate transpose*, or *adjoint*, is more useful. The adjoint of a matrix $\mathbf{A}$ is often denoted $\mathbf{A}^\dagger$. For example,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^\dagger = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix}. \tag{18}$$

If a matrix $\mathbf{A}$ is equal to its adjoint $\mathbf{A}^\dagger$, it is said to be a *Hermitian* matrix. Clearly, this can only happen if the diagonal elements are real (otherwise, a will never equal a^*, for example). One can also take the adjoint of a vector, which is effectively what one does in order to take a dot product of two complex vectors (see above):

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}^\dagger = \left(a_1^* a_2^* \cdots a_n^*\right), \qquad (a_1 a_2 \cdots a_n)^\dagger = \begin{pmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{pmatrix}. \tag{19}$$

3 Determinants

The determinant of a matrix $\mathbf{A}$ is denoted $|\mathbf{A}|$ and is a scalar quantity (i.e., a number). This number is involved in computation of inverse matrices (below). For the trivial case of a 1x1 matrix, the determinant is just the number in the matrix. For a 2x2 matrix, the determinant is easily computed as

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}. \tag{20}$$

For a 3x3 matrix, the determinant is again easily computed, being

$$\begin{vmatrix} a_{11}a_{12}a_{13} \\ a_{21}a_{22}a_{23} \\ a_{31}a_{32}a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12}. \tag{21}$$

Note that each term consists of the product of three factors. The positive terms can be obtained by starting on the top row and multiplying by elements going down and to the right diagonally, wrapping around when hitting the edge of the determinant. Similarly, negative terms can be obtained by starting on the bottom row and multiplying by elements going up and to the right diagonally, wrapping around when hitting an edge.

Another simple case for computing determinants is that of diagonal (or triangular) matrices, where the determinant is just the product of the entries on the main diagonal.

For more general matrices, there is a procedure for computing determinants which involves cofactor matrices

$$|\mathbf{A}| = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}, \tag{22}$$

where a_{ij} are the elements of matrix $\mathbf{A}$ and A_{ij} are the *cofactors*, which are $(-1)^{i+j}$ times the determinant of a smaller $(n-1)\times(n-1)$ matrix found by eliminating row i and column j [$A_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|$]. Obviously there is a problem if we are defining a determinant in terms of other determinants! However, we can apply these rules iteratively until we get to 3x3 or 2x2 matrices, for which we can take determinants using the simple rules given above. You may wish to confirm that the cofactor method, when applied to the case of a 3x3 matrix, yields equation (21) when the cofactors of the smaller 2x2 matrices are evaluated using the aid of equation (20).

4 Inverse Matrices

The *inverse* of a matrix is another matrix which, when multiplied by the first matrix, yields the unit matrix $\mathbf{I}$ (a matrix with all zeroes except 1's down the diagonal).

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}. \tag{23}$$

In the general case, the inverse may be written

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathbf{A}^T_{cof}, \tag{24}$$

where $\mathbf{A}^T_{cof}$ is the transpose of the matrix of cofactors $A_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|$. For example:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{25}$$

Clearly, there are major problems in finding the inverse of a matrix if it has a determinant equal to zero, since the formula for $\mathbf{A}^{-1}$ involves dividing by the determinant. In such cases, we say that $\mathbf{A}$ is *singular* and has no inverse.

5 Eigenvectors and Eigenvalues

The eigenvectors of a matrix $\mathbf{A}$ are those special vectors $\mathbf{v}$ for which $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, where λ is an associated constant (possibly complex) called the eigenvalue. Let us rearrange the eigenvalue equation to the form $(\mathbf{A}-\lambda)\mathbf{v} = \mathbf{0}$, where $\mathbf{0}$ represents a vector of all zeroes (the zero vector). We may rewrite this expression using the identity matrix $\mathbf{I}$ to yield $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$, which will be more convienient for the next step. Now to solve for $\mathbf{v}$, multiply the left and right sides of the equation by $(\mathbf{A} - \lambda\mathbf{I})$, if it exists. This yields $\mathbf{v} = (\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{0}$.

This last equation presents a challenge. Anything multiplied by the zero vector yields the zero vector, but clearly that is a trivial solution for $\mathbf{v}$ that we aren't interested in. Thus, our assumption that $(\mathbf{A} - \lambda\mathbf{I})^{-1}$ exists must be wrong. The eigenvalue equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ therefore has non-trivial solutions only when $(\mathbf{A} - \lambda\mathbf{I})^{-1}$ does not exist. In our previous discussion of determinants, we noted that a matrix $\mathbf{A}$ does not have an inverse if its determinant $|\mathbf{A}|$ is zero. Thus, we can satisfy the eigenvalue equation for those special values of λ such that $|\mathbf{A} - \lambda\mathbf{I}| = 0$. This is called the *secular determinant*, and expanding the $n \times n$ determinant gives an n-th degree polynomial in λ called the *secular equation* or the *characteristic equation*. Once the roots of this equation are determined to give n eigenvalues λ, these eigenvalues may be inserted into the eigenvalue equation, one at a time, to yield n eigenvectors.

As an example, let us find the eigenvalues and eigenvectors for the 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \tag{26}$$

We begin with the secular determinant $|\mathbf{A} - \lambda\mathbf{I}| = 0$, which in this case becomes

$$\begin{vmatrix} 1-\lambda & 0 & 1 \\ 0 & 1-\lambda & 0 \\ 1 & 0 & 1-\lambda \end{vmatrix} = 0. \tag{27}$$

Expanding out this determinant using the rules given above for the determinants of 3×3 matrices, we obtain the following characteristic equation:

$$\begin{aligned} (1-\lambda)^3 - (1-\lambda) &= 0 \\ (1-\lambda)[(1-\lambda)^2 - 1] &= 0 \\ (1-\lambda)(\lambda)(\lambda - 2) &= 0, \end{aligned} \tag{28}$$

which has solutions $\lambda = 1, 0, 2$. These are the three eigenvalues of $\mathbf{A}$. What are the corresponsing eigenvectors? We substitute each of these eigenvalues, one at a time, into the eigenvalue equation, and solve for the 3×3 system of equations that result.

Let us begin with the eigenvalue $\lambda = 1$. Substituting this into $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$, we obtain

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{29}$$

This is three equations in three unknowns, which we may rewrite as

$$
\begin{aligned}
v_3 &= 0 & (30)\\
0 &= 0 & (31)\\
v_1 &= 0. & (32)
\end{aligned}
$$

The middle equation is, of course, not particularly useful. However, we know that the components v_1 and v_3 of the eigenvector corresponding to $\lambda = 1$ are both zero, and there is no equation governing the choice of v_2. We are therefore free to chose *any* value for v_2, and a valid eigenvector will result. Note that any eigenvector times a constant will yield another valid eigenvector. Most frequently, we chose *normalized* eigenvectors by convention (such that $\mathbf{v} \cdot \mathbf{v} = 1$), so in this case we will choose $v_2 = 1$. This gives the final eigenvector

$$\mathbf{v} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \tag{33}$$

We can verify that this is indeed an eigenvector corresponding to the eigenvalue $\lambda = 1$ by multiplying this eigenvector by the original matrix $\mathbf{A}$:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \tag{34}$$

that is, multiplication of $\mathbf{A}$ times the eigenvector yields the eigenvector again times a constant (the eigenvalue, $\lambda = 1$).

By a similar procedure, one can obtain the other two eigenvectors, which, when normalized, are

$$\mathbf{v}(\lambda = 0) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad , \quad \mathbf{v}(\lambda = 2) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{35}$$

Although an $n \times n$ matrix $\mathbf{A}$ has n eigenvalues, they are not necessarily distinct. That is, one or more of the roots λ of the characteristic equation may be identical. In this case, we say that those eigenvalues are *degenerate*. Determination of eigenvectors is somewhat more complicated in such a case, because there will be additional flexibility in selecting them. Consider the matrix

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \tag{36}$$

which has the characteristic equation

$$(\lambda - 2)^2 \lambda = 0, \tag{37}$$

with solutions $\lambda = 0, 2, 2$. Although it is no difficulty to find the eigenvector corresponding to $\lambda = 0$, the doubly-degenerate eigenvalue $\lambda = 2$ presents an additional complication. Upon substituting this value into the eigenvalue equation, we obtain

$$\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{38}$$

or

$$\begin{aligned} -v_1 + v_3 &= 0 \\ 0 &= 0 \\ v_1 - v_3 &= 0. \end{aligned} \tag{39}$$

In this case, the first and third equations are equivalent (one equation is just the minus of the other), and so they are not independent. Additionally, the second equation gives no information. We therefore have only one equation to determine the three coefficients of the eigenvector. Recall that only two valid equations resulted above for our previous non-degenerate case, because any multiple of an eigenvector still yields a valid eigenvector. Here, a double degeneracy has lost us one of our equations. Hence, we only know that $v_1 = v_3$ and v_2 is arbitrary. Any eigenvector satisfying these rules will be satisfactory, and clearly there are an infinite number of ways to chose them, even if we require them to be normalized. Let us, somewhat arbitrarily, pick the normalized vector

$$\mathbf{v}(\lambda = 2, \mathrm{first}) = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}. \tag{40}$$

Then, it is traditional to try to pick the second eigenvector for $\lambda = 2$ as orthogonal to the first (there are reasons for doing this, most commonly because we might wish to use these vectors as a new orthonormal basis). In that case, the following (normalized) vector will be suitable:

$$\mathbf{v}(\lambda = 2, \mathrm{second}) = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}. \tag{41}$$

You can verify that both of these vectors are (a) orthonormal, and (b) satisfy the eigenvalue equation for $\lambda = 2$.