

## CORRELATION FUNCTIONS, VECTORS AND HIGH-DIMENSIONAL VECTOR SPACES, ORTHONORMAL BASIS VECTORS AND FOURIER SERIES

The fluctuation with time of a variable  $A$  about its mean value is described by the function  $A(t)$ . The time-correlation function for the variable is the classical ensemble average of the product of values of  $A(t)$  at times that differ by the interval  $t$

$$\begin{aligned} C_A(t) &= \langle A(t+\tau) A(\tau) \rangle \\ &= \int \cdots \int dpdq \rho(p,q) A(p,q;t+\tau) A(p,q;\tau) \end{aligned} \quad (1)$$

From a molecular dynamics simulation we obtain a time series of values of a variable, such as a main-chain dihedral angle. The averaging of the product  $A(t+\tau)A(\tau)$  to obtain  $C_A(t)$  can be over the time series. We can view the time series as a vector. The vector is defined by the sequence of values of the variable  $A$  at the sample points of the simulation. The dimension of the vector is equal to the number of sample points. With this picture, the correlation function  $C_A(t)$  is the appropriately normalized inner product (or dot product) of the two vectors,  $A(t+\tau)$  and  $A(\tau)$ ,

$$C_A(t) = \frac{1}{C_{norm}} (A(t+\tau), A(\tau)) \quad (2)$$

The vectors  $A(t+\tau)$  and  $A(\tau)$  are equal-size samples of the signal  $A(t)$ , with the start of the sequence defining the vector  $A(t+\tau)$  displaced by  $t$  from the start of  $A(\tau)$ .

The inner product  $(A(t+\tau), A(\tau))$  is proportional to the projection of the vector  $A(t+\tau)$  on  $A(\tau)$ . For small  $t$  the vectors are near parallel and the projection is large, approximately equal to the squared magnitude of  $A(\tau)$ . For large  $t$ , if the fluctuation in signal is high-frequency random noise, the vectors are orthogonal and the projection and  $C_A(t)$  will be zero. If the signal contains periodic components in addition to random noise, the vectors will oscillate about the perpendicular and the projection and  $C_A(t)$  will be

periodic.

The following develops the above geometrical picture of the inner product on a vector space by examining the two-dimensional case and extending it to higher dimensions and to vector spaces of functions. The notion of orthonormal basis vectors is introduced to flesh out the geometrical picture and to show, within the geometrical picture, the relationship between the inner product, the Fourier series expansion of a function, and the Fourier transform of a function.

I.  $\mathbf{R}^2$ : a two-dimensional vector space.

A vector in two-dimensions is an ordered sequence of two numbers,

$$\mathbf{r} = [x \ y] \tag{3}$$

The inner product of two vectors is the number (scalar) obtained by componentwise multiplication of the two sequences,

$$(\mathbf{r}_1, \mathbf{r}_2) = x_1x_2 + y_1y_2 \tag{4}$$

or by projection of one vector upon the other,

$$(\mathbf{r}_1, \mathbf{r}_2) = |\mathbf{r}_1| |\mathbf{r}_2| \cos(\theta) \tag{4a}$$

where  $\theta$  is the angle between the vectors. Equations (4) can be shown equivalent by expressing  $\cos(\theta)$  in terms of the magnitudes of the three sides of the triangle defined by the two vectors. The inner product or dot product is also written as  $\mathbf{r}_1 \cdot \mathbf{r}_2$ .

We introduce mutually orthogonal unit vectors (orthonormal basis vectors),  $\mathbf{i}$  and  $\mathbf{j}$ ,

$$\begin{aligned}\mathbf{i} &= [1 \ 0] \\ \mathbf{j} &= [0 \ 1]\end{aligned}\tag{5}$$

We can write the vector  $\mathbf{r} = [x \ y]$  as a sum of vectors,

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j}\tag{6}$$

It is seen that  $x = (\mathbf{r}, \mathbf{i})$  and  $y = (\mathbf{r}, \mathbf{j})$ , i.e., the components of  $\mathbf{r}$  are the projections of  $\mathbf{r}$  upon the unit basis vectors.

II.  $\mathbf{R}^n$ : an  $n$ -dimensional vector space.

The extension from two to three dimensions is obvious and familiar. The extension to more than three dimensions, where the vector space is a hyperspace, may be less familiar but also should be obvious. We change notation from that for two or three dimensions as used above to a notation suitable to vectors with a larger number of components. In  $\mathbf{R}^n$  as in  $\mathbf{R}^2$  a vector is a sequence of  $n$  numbers,

$$\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]\tag{7}$$

The inner product of two vectors is defined as for two dimensions, as the componentwise product of the two sequences,

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i\tag{8}$$

A sequence of orthonormal basis vectors for  $\mathbf{R}^n$  is

$$\begin{aligned}\mathbf{e}_1 &= [1 \ 0 \ \cdots \ 0] \\ \mathbf{e}_2 &= [0 \ 1 \ \cdots \ 0] \\ &\vdots \\ &\vdots \\ \mathbf{e}_n &= [0 \ 0 \ \cdots \ 1]\end{aligned}\tag{9}$$

As for two dimensions, a vector  $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]$  can be expressed as a sum of scaled basis vectors,

$$\mathbf{x} = \sum_{i=1}^n (\mathbf{x}, \mathbf{e}_i) \mathbf{e}_i \quad (10)$$

where the inner product coefficients are the projections of the vector upon the unit basis vectors.

The geometrical picture, e.g., of basis vectors, of the inner product, and of projection, is for higher dimensions the same as for two or three dimensions. The geometrical parallels serve as the basis for an intuitive understanding of the properties of a hyperspace.

A signal sampled discretely, as in a molecular dynamics simulation, can be represented as a vector in  $\mathbf{R}^n$ , with the dimension  $n$  being the size of the sample. The picture given for  $C_A(t)$  in the first two paragraphs immediately follows from the development in this subsection. If the first two paragraphs were not clear on first reading, perhaps they should be reviewed now.

For a molecular dynamics simulation of a protein with  $N$  atoms, a snapshot of the trajectory taken at a sample point, with values for  $3N$  coordinates and  $3N$  velocities, can be viewed as a vector in a  $3N$ -dimensional hyperspace describing the conformation or as a vector in the  $6N$ -dimensional phase space of statistical mechanics. As the system evolves with time, the vector changes direction and magnitude. The trajectory determined by the simulation is represented by a sequence of such vectors. The vectors define a set of points in configuration or phase space. The points will be clustered with varying density, but will be most dense about one or several regions. Note that we are thinking here as though in two or three dimensions. Although a  $3N$  or  $6N$ -dimensional vector space cannot be visualized, we note again, that geometrical concepts extended from two and three dimensions provide an intuitive and comfortable understanding of the more complex systems.

III.  $\mathbf{L}^2(a,b)$ : the infinite-dimensional vector space of functions square integrable over the range  $a$  to  $b$ .

The geometrical picture extends to vector spaces of well-behaved functions and infinite sequences of functions.

The inner product is defined on  $\mathbf{L}^2(a,b)$  as

$$(\mathbf{f}, \mathbf{g}) = \int_a^b f(t) \overline{g(t)} dt \quad (11)$$

This expression is similar to the inner product in  $\mathbf{R}^n$ , which is obtained by componentwise multiplication of the vectors. To see this similarity, consider the definition of an integral, where the integrand is divided into small segments that are summed. In  $\mathbf{L}^2(a,b)$  as in  $\mathbf{R}^n$ , the inner product is related to the projection of the function  $f(t)$  upon the function  $g(t)$ . The expression and picture developed for  $C_A(t)$  in the first two paragraphs holds equally well for  $A(t + \tau)$  and  $A(\tau)$  being functions rather than discrete sequences.

In  $\mathbf{L}^2$  as in  $\mathbf{R}^n$ , it should be possible to find a complete orthonormal system of basis vectors,  $\mathbf{e}_n$ . Each basis vector in  $\mathbf{L}^2$  will be itself a function, which is different from but parallels the situation in  $\mathbf{R}^n$ , a vector space of ordered sequences, where each basis vector is itself an ordered sequence  $[0 \cdots 1 \cdots 0]$ . As in  $\mathbf{R}^n$ , a function can be expressed as a sum of terms, the projection of the function upon each basis vector times the basis vector,

$$\mathbf{f} = \sum_{n=-\infty}^{\infty} (\mathbf{f}, \mathbf{e}_n) \mathbf{e}_n \quad (12)$$

The archetypal orthogonal expansion of this kind is the Fourier series, for which the basis vectors are

$$\mathbf{e}_n(t) = (2\pi)^{-1/2} e^{int} \quad -\pi < t < \pi \quad (13)$$

$(\mathbf{e}_n)_{-\infty}^{\infty}$  is a complete orthonormal sequence in  $\mathbf{L}^2(-\pi, \pi)$ . Thus we can write for a real  $2\pi$ -periodic function

$$\mathbf{f}(t) = (2\pi)^{-1/2} \sum_{n=-\infty}^{\infty} c_n e^{int} \quad n \in \mathbf{Z} \quad (14a)$$

$$\begin{aligned} c_n &= (\mathbf{f}, \mathbf{e}_n) \\ &= (2\pi)^{-1/2} \int_{-\pi}^{\pi} \mathbf{f}(t) e^{-int} dt \end{aligned} \quad (14b)$$

Each inner product coefficient  $c_n$  is a number which gives the size of a contribution of frequency  $n$  to the function  $\mathbf{f}(t)$ . That  $e^{int}$  describes a contribution of frequency  $n$  is apparent from the Euler formula

$$e^{int} = \cos(nt) + i \sin(nt) \quad (15)$$

The Fourier transform is similar to the Fourier series expansion just described. In the series expansion the indices  $n$  of the system of basis functions are integers, ranging from  $-\infty$  to  $\infty$ . Each basis function makes a contribution of integer frequency  $n$ . For the Fourier transform, we allow the "index" to be continuous, ranging now over the real numbers from  $-\infty$  to  $\infty$ , and we make a conforming change in notation, replacing the index  $n$  by the variable  $\omega$ . We also remove the restriction on the range of  $\mathbf{f}(t)$ . The relevant vector space is  $\mathbf{L}^2(-\infty, \infty)$ .

The Fourier transform of  $\mathbf{f}(t)$  is

$$\hat{\mathbf{f}}(\omega) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \mathbf{f}(t) e^{-i\omega t} dt \quad \omega \in \mathbf{R} \quad (16)$$

The original function  $\mathbf{f}(t)$  is recovered from  $\hat{\mathbf{f}}(\omega)$  by the inverse Fourier transform

$$\mathbf{f}(t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \hat{\mathbf{f}}(\omega) e^{i\omega t} d\omega \quad (17)$$

The transform  $\hat{\mathbf{f}}(\omega)$  is a vector space inner product, projecting the function  $\mathbf{f}(t)$  on the function  $(2\pi)^{-1/2} e^{i\omega t}$ . The transform is similar to the coefficients  $c_n$  of the Fourier series expansion, except the index has become continuous and can take any value. The inverse transform, returning the function, is similar to the summation of the Fourier series.

Taking the Fourier transform of a function  $\mathbf{f}(t)$  is a mapping from the  $t$  to the  $\omega$  domain, i.e., from the time to the frequency domain. The Fourier transform changes the basis from a function of time to a function of frequency.

## SUGGESTED READING

1. N. Young, *An introduction to Hilbert space*, Cambridge University Press, Cambridge, 1988, soft cover, ISBN 0-521-33717-8, QA322.4.Y68 1988.

A readable but mathematical treatment of vector spaces of sequences and functions and of linear operators. There is a list of references on page 225.

2. A. Szabo and N. S. Ostlund, *Modern quantum chemistry: Introduction to advanced electronic structure theory*, 1st edition, revised, McGraw-Hill, New York, 1989, hard cover, ISBN 0-07-062739-8, QD462.S95 1989

Chapter 1 has an excellent concise review of the mathematical background for quantum mechanics, including: vector spaces over  $\mathbf{C}^n$ , linear (matrix) algebra, basis functions, eigenvalues, extensions to vector spaces of functions, and linear operators. Standard quantum mechanics texts (e.g., Schiff) also give this background.

3. E. Kreyszig, *Advanced engineering mathematics*, 7th edition, John Wiley, New York, 1993, hard cover, ISBN 0-471-55380-8, QA401.K7 1993.

A text that is a good general reference. Not mathematical. The treatment of linear algebra is thorough. The treatment of vector spaces is sketchy.

4. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C*, 2nd edition, Cambridge University Press, Cambridge, 1992, hard cover, ISBN 0-521-43108-5, QA297.N866 1992.

A book of algorithms for computations, with code in C that can be obtained on diskette. The book covers a wide range of topics, including function minimization, eigenvalue problems and matrix manipulations, statistical treatment of data, differential equations, Fourier transform, etc. Useful and comprehensible background is given for each topic, with pointers to



references. A good book to find out what numerical methods are available for work with a problem. Purists and perhaps some others nitpick the code.