# **Mathematical Appendix**

The book assumes a familiarity with basic methods of linear algebra, differential equations, and probability theory, as covered in standard texts. This chapter describes the notation we use and briefly sketches highlights of various techniques. The references provide further information.

## **Linear Algebra**

An operation 0 on a quantity z is called linear if, applied to any two instances  $z_1$  and  $z_2$ ,  $\partial(z_1 + z_2) = \partial(z_1) + \partial(z_2)$ . In this section, we consider linear operator linear operations on vectors and functions. We define a vector **v** as an array of *N* numbers  $(v_1, v_2, \ldots, v_N)$  or equivalantly  $v_a$  for  $a = 1, 2, \ldots, N$ , which are called its components. These are sometimes listed in a single N-row column

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}. \tag{1}$$

When necessary, we write component *a* of **v** as  $[\mathbf{v}]_a = v_a$ . We use **0** to denote zero vector 0 the vector with all its components equal to zero. Spatial vectors, which are related to displacements in space, are a special case, and we donate them by  $\vec{v}$  with components  $v_x$  and  $v_y$  in two-dimensional space or  $v_x$ ,  $v_y$ , and spatial vector  $\vec{v}$  $v_z$  in three-dimensional space.

The length or norm of  $\mathbf{v}$ ,  $|\mathbf{v}|$ , when squared, can be written as a dot product

$$|\mathbf{v}|^{2} = \mathbf{v} \cdot \mathbf{v} = \sum_{a=1}^{N} v_{a}^{2} = v_{1}^{2} + v_{2}^{2} + \ldots + v_{N}^{2}.$$
 (2)

The dot product of two different *N*-component vectors, **v** and **u** is,

$$\mathbf{v} \cdot \mathbf{u} = \sum_{a=1}^{N} v_a u_a \,. \tag{3}$$

Draft: December 17, 2000

**Theoretical Neuroscience** 

vector v

norm

dot product

matrix W

Matrix multiplication is a basic linear operation on vectors. An  $N_r$  by  $N_c$  matrix **W** is an array of  $N_r$  rows and  $N_c$  columns

with elements  $W_{ab}$  for  $a = 1, ..., N_r$  and  $b = 1, ..., N_c$ . In this text, multiplication of a vector by a matrix is written in the somewhat idiosyncratic notation  $\mathbf{W} \cdot \mathbf{v}$ . The dot implies multiplication and summation over a shared index, as it does for the dot product. If  $\mathbf{W}$  is an  $N_r$  by  $N_c$  matrix and  $\mathbf{v}$  is a  $N_c$ -component vector,  $\mathbf{W} \cdot \mathbf{v}$  is an  $N_r$ -component vector with components

$$[\mathbf{W} \cdot \mathbf{v}]_a = \sum_{b=1}^{N_{\rm C}} W_{ab} v_b \,. \tag{5}$$

matrix-vector product

matrix product

In conventional matrix notation, the product of a matrix and a vector is written as Wv, but we prefer to use the dot notation to avoid frequent occurrences of matrix transposes (see below). We similarly denote a matrix product as  $W \cdot M$ . Matrices can only be multiplied in this way if the number of columns of W,  $N_c$ , is equal to the number of rows of M. Then,  $W \cdot M$  is a matrix with the same number of rows as W and the same number of columns as M, and with elements

$$[\mathbf{W} \cdot \mathbf{M}]_{ab} = \sum_{c=1}^{N_{\rm C}} W_{ac} M_{cb} \,. \tag{6}$$

A vector, written as in equation 1, is equivalent to a one-column, *N*-row matrix, and the rules for various matrix operations can thus be applied to vectors as well.

*square matrix* Square matrices are those for which  $N_r = N_c = N$ . An important square matrix is the identity matrix **I** with elements

$$[\mathbf{I}]_{ab} = \delta_{ab} \tag{7}$$

Kronecker delta where the Kronecker delta is defined as

$$\delta_{ab} = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} . \end{cases}$$
(8)

*diagonal matrix* Another important type of square matrix is the diagonal matrix, defined by

$$\mathbf{W} = \operatorname{diag}(h_1, h_2, \dots, h_N) = \begin{pmatrix} h_1 & 0 & \dots & 0 \\ 0 & h_2 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & h_N \end{pmatrix},$$
(9)

Peter Dayan and L.F. Abbott

which has components  $W_{ab} = h_a \delta_{ab}$  for some set of  $h_a$ , a = 1, 2, ..., N.

transpose

The transpose of an  $N_r$  by  $N_c$  matrix **W** is an  $N_c$  by  $N_r$  matrix  $\mathbf{W}^T$  with elements  $[\mathbf{W}^T]_{ab} = W_{ba}$ . The transpose of a column vector is a row vector,  $\mathbf{v}^T = (v_1 v_2 \dots v_N)$ . This is distinguished by the absence of commas from  $(v_1, v_2, \dots, v_N)$  which, for us, is a listing of the components of a column vector. In the following table, we define a number of products involving vectors and matrices. In the definitions, we provide our notation and the corresponding expressions in terms of vector components and matrix elements. We also provide the conventional matrix notation for these quantities as well as the notation used by MATLAB, a computer software package commonly used to perform these operations numerically. For the MAT-LAB notation (which does not use bold or italic symbols), we denote two column vectors by u and v, assuming they are defined within MATLAB by instructions such as  $v = [v(1) \ v(2) \ v(N)]'$ .

Quantity	Definition	Matrix	MATLAB
norm	$ \mathbf{v} ^2 = \mathbf{v} \cdot \mathbf{v} = \sum_a v_a^2$	$\mathbf{v}^{\mathrm{T}}\mathbf{v}$	v'*v
dot product	$\mathbf{v} \cdot \mathbf{u} = \sum_{a} v_{a} u_{a}$	$\mathbf{v}^{\mathrm{T}}\mathbf{u}$	v'*u
outer product	$[\mathbf{v}\mathbf{u}]_{ab}=v_au_b$	<b>vu</b> <sup>T</sup>	v*u'
matrix-vector product	$[\mathbf{W}\cdot\mathbf{v}]_a=\sum_b W_{ab}v_b$	Wv	W*v
vector-matrix product	$[\mathbf{v} \cdot \mathbf{W}]_a = \sum_b v_b W_{ba}$	$\mathbf{v}^{\mathrm{T}}\mathbf{W}$	v'*W
quadratic form	$\mathbf{v} \cdot \mathbf{W} \cdot \mathbf{u} = \sum_{ab} v_a W_{ab} u_b$	<b>v</b> <sup>T</sup> Wu	v'*W*u
matrix-matrix product	$[\mathbf{W}\cdot\mathbf{M}]_{ab}=\sum_{c}W_{ac}M_{cb}$	WM	W*M

Several important definitions for square matrices are:

Operation	Notation	Definition	MATLAB
transpose	$\mathbf{W}^{\mathrm{T}}$	$\mathbf{W}_{ab}^{\mathrm{T}}=W_{ba}$	W'
inverse	$\mathbf{W}^{-1}$	$\mathbf{W} \cdot \mathbf{W}^{-1} = \mathbf{I}$	inv(W)
trace	tr <b>W</b>	$\sum_{a} W_{aa}$	<pre>trace(W)</pre>
determinant	det W	see references	det(W)

A square matrix only has an inverse if its determinant is nonzero. Square matrices with certain properties are given special names:

Property	Definition		
symmetric	$\mathbf{W}^{\mathrm{T}} = \mathbf{W}  \text{or}  W_{ba} = W_{ab}$		
orthogonal	$\mathbf{W}^{\mathrm{T}} = \mathbf{W}^{-1}$ or $\mathbf{W}^{\mathrm{T}} \cdot \mathbf{W} = \mathbf{I}$		
positive-definite	$\mathbf{v} \cdot \mathbf{W} \cdot \mathbf{v} > 0$ for all $\mathbf{v} \neq 0$		
Töplitz	$W_{ab} = f(a-b)$		

Draft: December 17, 2000

**Theoretical Neuroscience** 

	4		
	where $f(a - b)$ is any function of the single variable $a - b$ .		
del operator $ abla$	For any real-valued function $E(\mathbf{v})$ of a vector $\mathbf{v}$ , we can define the derivative (which is sometimes called del) of $E(\mathbf{v})$ as the vector $\nabla$ components	he vector $E(\mathbf{v})$ with	
	$[\nabla E(\mathbf{v})]_a = \frac{\partial E(\mathbf{v})}{\partial v_a} .$	(10)	
directional derivative	The derivative of $E(\mathbf{v})$ in the direction $\mathbf{u}$ is then		
	$\lim_{\epsilon \to 0} \left( \frac{E(\mathbf{v} + \epsilon \mathbf{u}) - E(\mathbf{v})}{\epsilon} \right) = \mathbf{u} \cdot \nabla E(\mathbf{v}) .$	(11)	
	Eigenvectors and Eigenvalues		
eigenvector	An eigenvector of a square matrix ${f W}$ is a non-zero vector ${f e}$ that s	satisfies	
	$\mathbf{W} \cdot \mathbf{e} = \lambda \mathbf{e}$	(12)	
eigenvalue	for some number $\lambda$ called the eigenvalue. Possible values of $\lambda$ mined by solving the polynomial equation	are deter-	
	$\det(\mathbf{W}-\lambda\mathbf{I})=0.$	(13)	
	Typically, but not always, this has <i>N</i> solutions if <b>W</b> is an <i>N</i> by <i>N</i> matrix, and these can be either real or complex. Complex eigenvalues come in complex-conjugate pairs if <b>W</b> has real-valued elements. We use the index $\mu$ to label the different eigenvalues and eigenvectors, $\lambda_{\mu}$ and $\mathbf{e}_{\mu}$ . Note that $\mu$ identifies the eigenvector (and eigenvalue) to which we are referring; it does not signify a component of the eigenvector $\mathbf{e}_{\mu}$ .		

If **e** is an eigenvector,  $\alpha$ **e** is also an eigenvector for any nonzero value of  $\alpha$ . We can use this freedom to normalize eigenvectors so that  $|\mathbf{e}| = 1$ . If two eigenvectors, say  $\mathbf{e}_1$  and  $\mathbf{e}_2$ , have the same eigenvalues  $\lambda_1 = \lambda_2$ , they are termed degenerate, Then,  $\alpha \mathbf{e}_1 + \beta \mathbf{e}_2$  is also an eigenvector with the same eigenvalue, for any  $\alpha$  and  $\beta$  that are not both zero. Apart from such degeneracies, an *N*by *N* matrix can have at most *N* eigenvectors, although some matrices have fewer. If **W** has *N* non-degenerate eigenvalues, the eigenvectors  $\mathbf{e}_1, \ldots, \mathbf{e}_N$  are linearly independent, meaning that  $\sum_{\mu} c_{\mu} \mathbf{e}_{\mu} = \mathbf{0}$  only if the coefficients  $c_{\mu} = \mathbf{0}$  for all  $\mu$ . These eigenvectors can be used to represent any *N* component vector **v** through the relation

$$\mathbf{v} = \sum_{\mu=1}^{N} c_{\mu} \mathbf{e}_{\mu} , \qquad (14)$$

basiswith a unique set of coefficients  $c_{\mu}$ . They are thus said to form a basis.symmetric matrixThe eigenvalues and eigenvectors of symmetric matrices (for which  $\mathbf{W}^{\mathrm{T}} = \mathbf{W}$ ) have special properties, and for the remainder of this section, we con-

Peter Dayan and L.F. Abbott

sider this case. The eigenvalues of a symmetric matrix are real, and the eigenvectors are real and orthogonal (or can be made orthogonal in the case of degeneracy). This means that, if they are normalized to unit length, the eigenvectors satisfy

$$\mathbf{e}_{\mu} \cdot \mathbf{e}_{\nu} = \delta_{\mu\nu} \,. \tag{15}$$

This can be derived by noting that, for a symmetric matrix  $\mathbf{W}$ ,  $\mathbf{e}_{\mu} \cdot \mathbf{W} = \mathbf{W} \cdot \mathbf{e}_{\mu} = \lambda_{\mu} \mathbf{e}_{\mu}$ . Therefore, allowing the matrix to act in both directions we find  $\mathbf{e}_{\nu} \cdot \mathbf{W} \cdot \mathbf{e}_{\mu} = \lambda_{\mu} \mathbf{e}_{\nu} \cdot \mathbf{e}_{\mu} = \lambda_{\nu} \mathbf{e}_{\nu} \cdot \mathbf{e}_{\mu}$ . If  $\lambda_{\mu} \neq \lambda_{\nu}$ , this requires  $\mathbf{e}_{\nu} \cdot \mathbf{e}_{\mu} = \mathbf{0}$ . For orthogonal and normalized (orthonormal) eigenvectors, the coefficients in equation 14 take the values

$$c_{\mu} = \mathbf{v} \cdot \mathbf{e}_{\mu} \,. \tag{16}$$

Let  $\mathbf{E} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_N)$  be an *N* by *N* matrix with columns formed from the orthonormal eigenvectors of a symmetric matrix. From equation 15, this satisfies  $[\mathbf{E}^T \cdot \mathbf{E}]_{\mu\nu} = \mathbf{e}_{\mu} \cdot \mathbf{e}_{\nu} = \delta_{\mu\nu}$ . Thus,  $\mathbf{E}^T = \mathbf{E}^{-1}$ , making  $\mathbf{E}$  an orthogonal matrix.  $\mathbf{E}$  generates a transformation from the original matrix  $\mathbf{W}$  to a diagonal form, which is called matrix diagonalization,

$$\mathbf{E}^{-1} \cdot \mathbf{W} \cdot \mathbf{E} = \mathbf{E}^{\mathrm{T}} \cdot \operatorname{diag}(\lambda_{1} \mathbf{e}_{1}, \dots, \lambda_{N} \mathbf{e}_{N}) = \operatorname{diag}(\lambda_{1}, \dots, \lambda_{N}).$$
(17)

Conversely,

$$\mathbf{W} = \mathbf{E} \cdot \operatorname{diag}(\lambda_1, \dots, \lambda_N) \cdot \mathbf{E}^{-1}.$$
(18)

The transformation to and back from a diagonal form is extremely useful because computations with diagonal matrices are easy. Defining  $\mathbf{L} = \text{diag}(\lambda_1, \ldots, \lambda_N)$  we find, for example, that

$$\mathbf{W}^{n} = (\mathbf{E} \cdot \mathbf{L} \cdot \mathbf{E}^{-1}) \cdot (\mathbf{E} \cdot \mathbf{L} \cdot \mathbf{E}^{-1}) \cdots (\mathbf{E} \cdot \mathbf{L} \cdot \mathbf{E}^{-1})$$
  
=  $\mathbf{E} \cdot \mathbf{L}^{n} \cdot \mathbf{E}^{-1} = \mathbf{E} \cdot \operatorname{diag}(\lambda_{1}^{n}, \dots, \lambda_{N}^{n}) \cdot \mathbf{E}^{-1}.$  (19)

Indeed, for any function *f* that can be written as a power or expanded in a power series (including, for example, exponentials and logarithms),

$$f(\mathbf{W}) = \mathbf{E} \cdot \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_N)) \cdot \mathbf{E}^{-1}.$$
 (20)

#### **Functional Analogs**

A function v(t) can be treated as if it were a vector with a continuous label. *functions as vectors* In other words, the function value v(t) parameterized by the continuously varying argument t takes the place of the component  $v_a$  labeled by the integer-valued index a. In applying this analogy, sums over a for vectors are replaced by integrals over t for functions,  $\sum_a \rightarrow \int dt$ . For example, the functional analog of the squared norm and dot product are

$$\int dt v^2(t)$$
 and  $\int dt v(t) u(t)$ . (21)

Draft: December 17, 2000

*matrix diagonalization* 

orthonormal

eigenvectors

The analog of matrix multiplication for a function is the linear integral operator

$$\int dt' W(t, t')v(t') \tag{22}$$

with the function values W(t, t') playing the role of the matrix elements  $W_{ab}$ . The analog of the identity matrix is the Dirac  $\delta$  function  $\delta(t - t')$  discussed at the end of this section. The analog of a diagonal matrix is a function of two variables that is proportional to a  $\delta$  function,  $W(t, t') = h(t)\delta(t - t')$ , for any function h.

*functional inverse* All of the vector and matrix operations and properties defined above have functional analogs. Of particular importance are the functional inverse (which is not equivalent to an inverse function) that satisfies

$$\int dt'' W^{-1}(t, t'') W(t'', t') = \delta(t - t'), \qquad (23)$$

and the analog of the Töplitz matrix, which is a linear integral operator that is translationally invariant and thus can be written as

$$W(t, t') = K(t - t').$$
 (24)

*linear filter* The linear integral operator then takes the form of a linear filter,

$$\int dt' K(t-t')v(t') = \int d\tau K(\tau)v(t-\tau)$$
(25)

where we have made the replacement  $t' \rightarrow t - \tau$ .

#### The $\delta$ Function

Despite its name, the Dirac  $\delta$  function is not a properly defined function, but rather the limit of a sequence of functions. In this limit, the  $\delta$  function approaches zero everywhere except where its argument is zero, and there it grows without bound. The infinite height and infinitesimal width of this function are matched so that its integral is one. Thus,

$$\int dt \,\delta(t) = 1 \tag{26}$$

provided only that the limits of integration surround the point t=0 (otherwise the integral is zero). The integral of the product of a  $\delta$  function with any continuous function f is

$$\int dt' \,\delta(t-t')\,f(t') = f(t) \tag{27}$$

for any value of t contained within the integration interval (if t is not within this interval, the integral is zero). These two identities normally

linear integral operator

translation invariance provide enough information to use the  $\delta$  function in calculations despite its unwieldy definition.

The sequence of functions used to construct the  $\delta$  function as a limit is not unique. In essence, any function that integrates to one and has a single peak that gets continually narrower and taller as the limit is taken can be used. For example, the  $\delta$  function can be expressed as the limit of a square pulse

$$\delta(t) = \lim_{\Delta t \to 0} \begin{cases} 1/\Delta t & \text{if } -\Delta t/2 < t < \Delta t/2 \\ 0 & \text{otherwise} \end{cases}$$
(28)

or a Gaussian function

$$\delta(t) = \lim_{\Delta t \to 0} \frac{1}{\sqrt{2\pi}\Delta t} \exp\left[-\frac{1}{2}\left(\frac{t}{\Delta t}\right)^2\right].$$
 (29)

It is most often expressed as

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, \exp(i\omega t) \,. \tag{30}$$

This underlies the inverse Fourier transform, as discussed below.

#### **Eigenfunctions**

The functional analog of the eigenvector (equation 12) is the eigenfunction e(t) that satisfies

$$\int dt' W(t, t') e(t') = \lambda e(t) .$$
(31)

For translationally invariant integral operators, W(t, t') = K(t - t'), the eigenfunctions are complex exponentials,

$$\int dt' K(t-t') \exp(i\omega t') = \left(\int d\tau K(\tau) \exp(-i\omega \tau)\right) \exp(i\omega t), \qquad (32)$$

as can be seen by making the change of variables  $\tau = t - t'$ . Here  $i = \sqrt{-1}$ , and the complex exponential is defined by the identity

 $\exp(i\omega t) = \cos(\omega t) + i\sin(\omega t).$ (33)

Comparing equations 31 and 32, we see that the eigenvalue for this eigenfunction is

$$\lambda(\omega) = \int d\tau \, K(\tau) \exp(-i\omega\tau) \,. \tag{34}$$

In this case, the continuous label  $\omega$  takes the place of the discrete label  $\mu$  used to identify the different eigenvalues of a matrix.

Draft: December 17, 2000

**Theoretical Neuroscience** 

complex exponential

δ function definition

A functional analog of expanding a vector using eigenvectors as a basis (equation 14) is the inverse Fourier transform, which expresses a function in an expansion using complex exponential eigenfunctions as a basis. The analog of equation 16 for determining the coefficient functions of this expansion is the Fourier transform.

#### **Fourier Transforms**

As outlined in the previous section, Fourier transforms provide a useful representation for functions when they are acted upon by translation invariant linear operators.

*Fourier transform* of a function f(t) is a complex function of a real argument  $\omega$  given by

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} dt \, f(t) \exp(i\omega t) \,. \tag{35}$$

The Fourier transform  $\tilde{f}(\omega)$  provides a complete description of the original function f(t) because it can be inverted through,

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, \tilde{f}(\omega) \exp(-i\omega t) \,. \tag{36}$$

This provides an inverse because

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp(-i\omega t) \int_{-\infty}^{\infty} dt' f(t') \exp(i\omega t')$$

$$= \int_{-\infty}^{\infty} dt' f(t') \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp(i\omega(t'-t)) = \int_{-\infty}^{\infty} dt' f(t') \delta(t'-t) = f(t)$$
(37)

by the definition of the  $\delta$  function in equation 30. The function f(t) has to satisfy regularity conditions called the Dirichlet conditions for the inversion of the Fourier transform to be exact.

convolution

The convolution of two functions *f* and *g* is the integral

$$h(t) = \int_{-\infty}^{\infty} d\tau \ f(\tau) g(t-\tau) \,. \tag{38}$$

This is sometimes denoted by h = f \* g. Note that the operation of multiplying a function by a linear filter and integrating, as in equation 25, is a convolution. Fourier transforms are useful for dealing with convolutions because the Fourier transform of a convolution is the product of the Fourier transforms of the two functions being convolved,

$$\hat{h}(\omega) = \hat{f}(\omega)\tilde{g}(\omega).$$
(39)

Peter Dayan and L.F. Abbott

To show this, we note that

$$\tilde{h}(\omega) = \int_{-\infty}^{\infty} dt \exp(i\omega t) \int_{-\infty}^{\infty} d\tau f(\tau) g(t-\tau)$$

$$= \int_{-\infty}^{\infty} d\tau f(\tau) \exp(i\omega \tau) \int_{-\infty}^{\infty} dt g(t-\tau) \exp(i\omega(t-\tau))$$

$$= \int_{-\infty}^{\infty} d\tau f(\tau) \exp(i\omega \tau) \int_{-\infty}^{\infty} dt' g(t') \exp(i\omega t') \quad \text{where } t' = t-\tau ,$$
(40)

which is equivalent to equation 39. A related result is Parseval's theorem,

$$\int_{-\infty}^{\infty} dt f(t)^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega |\tilde{f}(\omega)|^2.$$
(41)

If f(t) is periodic, with period T (which means that f(t + T) = f(t) for all *period*, it can be represented by a Fourier series rather than a Fourier integral. That is

$$f(t) = \sum_{k=-\infty}^{\infty} \tilde{f}_k \exp(-i2\pi kt/T)$$
(42)

where  $\tilde{f}_k$  is given by:

$$\tilde{f}_{k} = \frac{1}{T} \int_{0}^{T} dt \, f(t) \exp(i2\pi kt/T) \,. \tag{43}$$

As in the case of Fourier transforms, regularity conditions have to hold for the series to converge and to be exactly invertible. The Fourier series has properties similar to Fourier transforms, including a convolution theorem and a version of Parseval's theorem. The real and imaginary parts of a Fourier series are often separated giving the alternative form

$$f(t) = \tilde{f}_0 + \sum_{k=1}^{\infty} \left( \tilde{f}_k^c \cos(2\pi kt/T) + \tilde{f}_k^s \sin(2\pi kt/T) \right)$$
(44)

with

$$\tilde{f}_{0} = \frac{1}{T} \int_{0}^{T} dt f(t), \quad \tilde{f}_{k}^{c} = \frac{2}{T} \int_{0}^{T} dt f(t) \cos(2\pi kt/T),$$
  

$$\tilde{f}_{k}^{s} = \frac{2}{T} \int_{0}^{T} dt f(t) \sin(2\pi kt/T).$$
(45)

When computed numerically, a Fourier transform is typically based on a certain number,  $N_t$ , of samples of the function,  $f_n = f(n\delta)$  for  $n = 0, 1, ..., N_t - 1$ . The discrete Fourier transform of these samples is then used as an approximation of the continuous Fourier transform. The discrete Fourier transform is defined as

$$\tilde{f}_m = \sum_{n=0}^{N_t - 1} f_n \exp\left(i2\pi nm/N_t\right) \,. \tag{46}$$

Draft: December 17, 2000

**Theoretical Neuroscience** 

discrete Fourier transform

Parseval's theorem

periodic function Fourier series Note that  $\tilde{f}_{N_t+m} = \tilde{f}_m$ . An approximation of the continuous Fourier transform is provided by the relation  $\tilde{f}(2\pi m/(N_t\delta)) \approx \delta \tilde{f}_m$ . The inverse discrete Fourier transform is

$$f_n = \frac{1}{N_t} \sum_{m=0}^{N_t - 1} \tilde{f}_m \exp\left(-i2\pi mn/N_t\right) \,. \tag{47}$$

This equation implies a periodic continuation of  $f_n$  outside the range  $0 \le n < N_t$ , so that  $f_{n+N_t} = f_n$  for all *n*. Consult the references for an analysis of the properties of the discrete Fourier transform and the quality of its approximation to the continuous Fourier transform. Note in particular that there is a difference between the discrete-time Fourier transform, which is the Fourier transform of a signal that is inherently discrete i.e. is only defined at discrete points) and the discrete Fourier transform, given above, which is based on a finite number of samples of an underlying continuous function. If f(t) is band-limited, meaning that  $\tilde{f}(\omega) = 0$  for  $|\omega| > \pi/\delta$ , the sampling theorem states that f(t) is completely determined by regular samples spaced at intervals  $1/\delta$ .

Fourier transforms of functions of more than one variable involve a direct extension of the equations given above to multi-dimensional integrals. For example,

$$\tilde{f}(\omega_x, \omega_y) = \int dx \int dy \, f(x, y) \exp(i(\omega_x x + \omega_y y)).$$
(48)

The properties of multi-dimensional transforms are similar to those of onedimensional transforms.

## **Finding Extrema and Lagrange Multipliers**

An operation frequently encountered in the text is minimizing a quadratic form. In terms of vectors, this typically amounts to finding the matrix **W** that makes the product  $\mathbf{W} \cdot \mathbf{v}$  closest to another vector **u** when averaged over a number of presentations of **v** and **u**. The function to be minimized is the average squared error  $\langle |\mathbf{u} - \mathbf{W} \cdot \mathbf{v}|^2 \rangle$ , where the brackets denote averaging over all the different samples **v** and **u**. Taking the derivative of this expression with respect to **W** gives the equation

minimization of quadratic form

$$\mathbf{W} \cdot \langle \mathbf{v} \mathbf{v} \rangle = \langle \mathbf{u} \mathbf{v} \rangle \quad \text{or} \quad \sum_{c=1}^{N} W_{ac} \langle v_c v_b \rangle = \langle u_a v_b \rangle.$$
(49)

Many variants of this equation, solved by a number of techniques, appear in the text.

Often, when a function  $f(\mathbf{v})$  has to be minimized or maximized with respect to a vector  $\mathbf{v}$  there is an additional constraint on  $\mathbf{v}$  that requires another function  $g(\mathbf{v})$  to be held constant. The standard way of dealing with

Peter Dayan and L.F. Abbott

sampling theorem

Lagrange

multiplier

this situation is to find the extrema of the function  $f(\mathbf{v}) + \lambda g(\mathbf{v})$  where  $\lambda$  is a free parameter called a Lagrange multiplier. Once this is done, the value of  $\lambda$  is determined by requiring  $g(\mathbf{v})$  to take the required constant value. This procedure can appear a bit mysterious when first encountered, so we provide a rather extended discussion.

The condition that characterizes an extreme value of the function  $f(\mathbf{v})$  is that small changes  $\Delta \mathbf{v}$  (with components  $\Delta v_a$ ) in the vector  $\mathbf{v}$  should not change the value of the function, to first order in  $\Delta \mathbf{v}$ . This results in the condition

$$\sum_{a=1}^{N} f_a \Delta v_a = 0 \tag{50}$$

where we use the notation

$$f_a = [\nabla f]_a = \frac{\partial f}{\partial v_a} \tag{51}$$

to make the equations more compact. Without a constraint, equation 50 must be satisfied for all  $\Delta v$ , which can only occur if each term in the sum vanishes separately. Thus, we find the usual condition for an extremum

$$f_a = \frac{\partial f}{\partial v_a} = 0 \tag{52}$$

for all *a*. However, with a constraint such as  $g(\mathbf{v}) = \text{constant}$ , equation 50 does not have to hold for all possible  $\Delta \mathbf{v}$ , only for those that satisfy the constraint. The condition on  $\Delta \mathbf{v}$  imposed by the constraint is that it cannot change the value of *g*, to first order in  $\Delta \mathbf{v}$ . Therefore,

$$\sum_{a=1}^{N} g_a \Delta v_a = 0 \tag{53}$$

with the same notation for the derivative used for *g* as for *f*.

The most obvious way to deal with the constraint equation 53 is to solve for one of the components of  $\Delta \mathbf{v}$ , say  $\Delta v_c$ , writing

$$\Delta v_c = -\frac{1}{g_c} \sum_{a \neq c} g_a \Delta v_a \,. \tag{54}$$

Then, we substitute this expression into equation 50 to obtain

$$\sum_{a\neq c} f_a \Delta v_a - \frac{f_c}{g_c} \sum_{a\neq c} g_a \Delta v_a = 0.$$
(55)

Because we have eliminated the constraint, this equation must be satisfied for all values of the remaining components of  $\Delta \mathbf{v}$ , those with  $a \neq c$ , and thus we find

$$f_a - \frac{f_c}{g_c}g_a = 0 \tag{56}$$

Draft: December 17, 2000

**Theoretical Neuroscience** 

for all  $a \neq c$ . The derivatives of *f* and *g* are functions of **v**, so these equations can be solved to determine where the extremum point is located.

In the above derivation, we have singled out component *c* for special treatment. We have no way of knowing until we get to the end of the calculation whether the particular *c* we chose leads to a simple or a complex set of final equations. The clever idea of the Lagrange multiplier is to notice that the whole problem is symmetric with respect to the different components of  $\Delta \mathbf{v}$ . Choosing one *c* value, as we did above, breaks this symmetry and often complicates the algebra. To introduce the Lagrange multiplier we simply define it as

$$\lambda = -\frac{f_c}{g_c}.$$
(57)

With this notation, the final set of equations can be written as

$$f_a + \lambda g_a = 0. \tag{58}$$

Before we had to say that these equations only held for  $a \neq c$  because c was treated differently. Now, however, notice that the above equation when a is set to c is algebraically equivalent to the definition of equation 57. Thus, we can say that equation 58 applies for all a, and this provides a symmetric formulation of the problem of finding an extremum that often results in simpler algebra.

The final realization is that equation 58 for all *a* is precisely what we would have derived if we had set out in the first place to find an extremum of the function  $f(\mathbf{v}) + \lambda g(\mathbf{v})$  and forgot about the constraint entirely. Of course this lunch is not completely free. From equation 58, we derive a set of extremum points parameterized by the undetermined variable  $\lambda$ . To fix  $\lambda$ , we must substitute this family of solutions back into  $g(\mathbf{v})$  and find the value of  $\lambda$  that satisfies the constraint that  $g(\mathbf{v})$  equals the specified constant. This provides the solution to the constrained problem.

## **Differential Equations**

The most general differential equation we consider takes the form

$$\frac{d\mathbf{v}}{dt} = \mathbf{f}(\mathbf{v}) \tag{59}$$

where  $\mathbf{v}(t)$  is an *N*-component vector of time-dependent variables, and **f** is a vector of functions of **v**. Unless it is unstable, allowing the absolute value of one or more of the components of **v** to grow without bound, this type of equation has three classes of solutions. For one class, called fixed points,  $\mathbf{v}(t)$  approaches a time-independent vector  $\mathbf{v}_{\infty}$  ( $\mathbf{v}(t) \rightarrow \mathbf{v}_{\infty}$ ) as  $t \rightarrow \infty$ . In a second class of solutions, called limit cycles,  $\mathbf{v}(t)$  becomes

Peter Dayan and L.F. Abbott

fixed point

limit cycle

chaos

periodic at large times and repeats itself indefinitely. For the third class of solutions, the chaotic ones,  $\mathbf{v}(t)$  never repeats itself but the trajectory of the system lies in a limited subspace of the total space of allowed configurations called a strange attractor. Chaotic solutions are extremely sensitive to initial conditions.

We focus most of our analysis on fixed-point solutions. For  $\mathbf{v}_{\infty}$  to be a time-independent solution of equation 59, which is also called an equilibrium point, we must have  $\mathbf{f}(\mathbf{v}_{\infty}) = \mathbf{0}$ . General solutions of equation 59 when  $\mathbf{f}$  is nonlinear cannot be constructed, but we can use linear techniques to study the behavior of  $\mathbf{v}$  near a fixed point  $\mathbf{v}_{\infty}$ . If  $\mathbf{f}$  is linear, the techniques we use and solutions we obtain as approximations in the nonlinear case are exact. Near the fixed point  $\mathbf{v}_{\infty}$ , we write

$$\mathbf{v}(t) = \mathbf{v}_{\infty} + \boldsymbol{\epsilon}(t) \tag{60}$$

and consider the case when all the components of the vector  $\boldsymbol{\epsilon}$  are small. Then, we can expand **f** in a Taylor series,

$$\mathbf{f}(\mathbf{v}(t)) \approx \mathbf{f}(\mathbf{v}_{\infty}) + \mathbf{J} \cdot \boldsymbol{\epsilon}(t) = \mathbf{J} \cdot \boldsymbol{\epsilon}(t)$$
(61)

where **J** is the called the Jacobian matrix and has elements

$$J_{ab} = \left. \frac{\partial f_a(\mathbf{v})}{\partial v_b} \right|_{\mathbf{v} = \mathbf{v}_{\infty}}.$$
 (62)

In the second equality of equation 61, we have used the fact that  $\mathbf{f}(\mathbf{v}_{\infty}) = \mathbf{0}$ .

Using the approximation of equation 61, equation 59 becomes

$$\frac{d\boldsymbol{\epsilon}}{dt} = \mathbf{J} \cdot \boldsymbol{\epsilon} \,. \tag{63}$$

The temporal evolution of  $\mathbf{v}(t)$  is best understood by expanding  $\epsilon$  in the basis provided by the eigenvectors of **J**. Assuming that **J** is real and has *N* linearly independent eigenvectors  $\mathbf{e}_1, \ldots, \mathbf{e}_N$  with different eigenvalues  $\lambda_1, \ldots, \lambda_N$ , we write

$$\boldsymbol{\epsilon}(t) = \sum_{\mu=1}^{N} c_{\mu}(t) \mathbf{e}_{\mu} \,. \tag{64}$$

Substituting this into equation 63, we find that the coefficients must satisfy

$$\frac{dc_{\mu}}{dt} = \lambda_{\mu} c_{\mu} . \tag{65}$$

This produces the solution

$$\boldsymbol{\epsilon}(t) = \sum_{\mu=1}^{N} c_{\mu}(0) \exp(\lambda_{\mu} t) \mathbf{e}_{\mu}$$
(66)

Draft: December 17, 2000

Taylor series

Jacobian matrix

strange attractor

equilibrium point

where  $\epsilon(0) = \sum_{\mu} c_{\mu}(0) \mathbf{e}_{\mu}$ . The individual terms in the sum on the right side of equation 66 are called modes. This solution is exact for equation 63, but is only a valid approximation when applied to equation 59 if  $\epsilon$  is small. Note that the different coefficients  $c_{\mu}$  evolve over time independently of each other. This does not require the eigenvectors to be orthogonal. If the eigenvalues and eigenvectors are complex,  $\mathbf{v}(t)$  will nonetheless remain real if  $\mathbf{v}(0)$  is real, because the complex parts of the conjugate pairs cancel appropriately. Expression 66 is not the correct solution if some of the eigenvalues are equal. The reader should consult the references for the solution in this case.

Equation 66 determines how the evolution of  $\mathbf{v}(t)$  in the neighborhood of  $\mathbf{v}_{\infty}$  depends on the eigenvalues of **J**. If we write  $\lambda_{\mu} = \alpha_{\mu} + i\omega_{\mu}$ ,

$$\exp(\lambda_{\mu} t) = \exp(\alpha_{\mu} t) \left( \cos(\omega_{\mu} t) + i \sin(\omega_{\mu} t) \right).$$
(67)

This implies that modes with real eigenvalues ( $\omega_{\mu} = 0$ ) evolve exponentially over time, and modes with complex eigenvalues ( $\omega_{\mu} \neq 0$ ) oscillate with a frequency  $\omega_{\mu}$ . Recall that the eigenvalues are always real if **J** is a symmetric matrix. Modes with negative real eigenvalues ( $\alpha_{\mu} < 0$  and  $\omega_{\mu} = 0$ ) decay exponentially to zero, while those with positive real eigenvalues ( $\alpha_{\mu} > 0$  and  $\omega_{\mu} = 0$ ) grow exponentially. Similarly, the oscillations for modes with complex eigenvalues are damped exponentially to zero if the real part of the eigenvalue is negative ( $\alpha_{\mu} < 0$  and  $\omega_{\mu} \neq 0$ ) and grow exponentially if the real part is positive ( $\alpha_{\mu} > 0$  and  $\omega_{\mu} \neq 0$ ).

Stability of the fixed point  $\mathbf{v}_{\infty}$  requires the real parts of all the eigenvalues to be negative ( $\alpha_{\mu} < 0$  for all  $\mu$ ). In this case, the point  $\mathbf{v}_{\infty}$  is a stable fixed-point attractor of the system, meaning that  $\mathbf{v}(t)$  will approach  $\mathbf{v}_{\infty}$  if it attractor starts from any point in the neighborhood of  $\mathbf{v}_{\infty}$ . If any real part is positive unstable fixed point  $(\alpha_{\mu} > 0 \text{ for any } \mu)$ , the fixed point is unstable. Almost any  $\mathbf{v}(t)$  initially in the neighborhood of  $\mathbf{v}_{\infty}$  will move away from that neighborhood. If **f** is linear, the exponential growth of  $|\mathbf{v}(t) - \mathbf{v}_{\infty}|$  never stops in this case. For a nonlinear f, equation 66 only determines what happens in the neighborhood of  $\mathbf{v}_{\infty}$ , and the system may ultimately find a stable attractor away from  $\mathbf{v}_{\infty}$ , either a fixed point, a limit cycle, or a chaotic attractor. In all these cases, the mode for which the real part of  $\lambda_{\mu}$  takes the largest value dominates the dynamics as  $t \to \infty$ . If this real part is equal to zero, the marginal stability fixed point is called marginally stable.

> As mentioned previously, the analysis presented above as an approximation for nonlinear differential equations near a fixed point is exact if the original equation is linear. In the text, we frequently encounter linear equations of the form

$$\tau \frac{dv}{dt} = v_{\infty} - v \,. \tag{68}$$

This can be solved by setting  $z = v - v_{\infty}$ , rewriting the equation as dz/z =

Peter Dayan and L.F. Abbott

 $-dt/\tau$  and integrating both sides

$$\tau \int_{z(0)}^{z(t)} dz' \frac{1}{z'} = \ln\left(\frac{z(t)}{z(0)}\right) = -\frac{t}{\tau}.$$
(69)

This gives  $z(t) = z(0) \exp(-t/\tau)$  or

$$v(t) = v_{\infty} + (v(0) - v_{\infty}) \exp(-t/\tau).$$
(70)

In some cases, we consider discrete rather than continuous dynamics defined over discrete steps n = 1, 2, ... through a difference rather than a differential equation. Linearization about equilibrium points can be used to analyze nonlinear difference equations as well as differential equations, and this reveals similar classes of behavior. We illustrate difference equations by analyzing a linear case,

difference equation

Ohm's law

$$\mathbf{v}(n+1) = \mathbf{v}(n) + \mathbf{W} \cdot \mathbf{v}(n).$$
(71)

The strategy for solving this equation is similar to that for solving differential equations. Assuming **W** has a complete set of linearly independent eigenvectors  $\mathbf{e}_1, \ldots, \mathbf{e}_N$  with different eigenvalues  $\lambda_1, \ldots, \lambda_N$ , the modes separate, and the general solution is

$$\mathbf{v}(n) = \sum_{\mu=1}^{N} c_{\mu} (1+\lambda_{\mu})^{n} \mathbf{e}_{\mu}$$
(72)

where  $\mathbf{v}(0) = \sum_{\mu} c_{\mu} \mathbf{e}_{\mu}$ . This has characteristics similar to equation 66. Writing  $\lambda_{\mu} = \alpha_{\mu} + i\omega_{\mu}$ , mode  $\mu$  is oscillatory if  $\omega_{\mu} \neq 0$ . In the discrete case, stability of the system is controlled by the magnitude

$$|1 + \lambda_{\mu}|^{2} = (1 + \alpha_{\mu})^{2} + (\omega_{\mu})^{2}.$$
(73)

If this is greater than one for any value of  $\mu$ ,  $|\mathbf{v}(n)| \to \infty$  as  $n \to \infty$ . If it is less than one for all  $\mu$ ,  $\mathbf{v}(n) \to \mathbf{0}$  in this limit.

### **Electrical Circuits**

Biophysical models of single cells involve equivalent circuits composed of resistors, capacitors, and voltage and current sources. We review here basic results for such circuits. Figures 1A & B show the standard symbols for resistors and capacitors, and define the relevant voltages and currents. A resistor (figure 1A) satisfies Ohm's law, which states that the voltage  $V_{\rm R} = V_1 - V_2$  across a resistance *R* carrying a current  $I_{\rm R}$  is

$$V_{\rm R} = I_{\rm R} R \,. \tag{74}$$



Figure 1: Electrical circuit elements and resistor circuits. A) Current  $I_{\rm R}$  flows through a resistance *R* producing a voltage drop  $V_1 - V_2 = V_{\rm R}$ . B) Charge  $\pm Q_{\rm C}$  is stored across a capacitance *C* leading to a voltage  $V_{\rm C} = V_1 - V_2$  and a current  $I_{\rm C}$ . C) Series resistor circuit called a voltage divider. D) Parallel resistor circuit.  $I_{\rm e}$  represents an external current source. The lined triangle symbol at the bottom of the circuits in C & D represents an electrical ground, which is defined to be at zero voltage.

Resistance is measured in ohms ( $\Omega$ ) defined as the resistance through which one ampere of current causes a voltage drop of one volt (1 V = 1 A × 1  $\Omega$ ).

A capacitor (figure 1B) stores charge across an insulating medium, and the voltage across it  $V_{\rm C} = V_1 - V_2$  is related to the charge it stores  $Q_{\rm C}$  by

$$CV_{\rm C} = Q_{\rm C} \tag{75}$$

where *C* is the capacitance. Electrical current cannot cross the insulating medium, but charges can be redistributed on each side of the capacitor, which leads to the flow of current. We can take a time derivative of both sides of equation 75 and use the fact that current is equal to the rate of change of charge,  $I_{\rm C} = dQ_{\rm C}/dt$ , to obtain the basic voltage-current relationship for a capacitor,

 $C\frac{dV_{\rm C}}{dt} = I_{\rm C} \,. \tag{76}$ 

Capacitance is measured in units of farads (F) defined as the capacitance for which one ampere of current causes a voltage change of one volt per second ( $1 \text{ F} \times 1 \text{ V/s} = 1 \text{ A}$ ).

The voltages at different points in a circuit and the currents flowing through various circuit elements can be computed using equations 74 and 76 and rules called Kirchoff's laws. These state that: 1) voltage differences around any closed loop in a circuit must sum to zero, and 2) the sum of all the currents entering any point in a circuit must be zero. Applying the second of these rules to the circuit in figure 1C, we find that  $I_1 = I_2$ . Ohm's law tells us that  $V_1 - V_2 = I_1 R_1$  and  $V_2 = I_2 R_2$ . Solving these gives

Peter Dayan and L.F. Abbott

Draft: December 17, 2000

V-I relation for capacitor

Kirchoff's laws

 $V_1 = I_1(R_1 + R_2)$ , which tells us that resistors arranged in series add, and  $V_2 = V_1 R_2/(R_1 + R_2)$ , which is why this circuit is called a voltage divider.

In the circuit of figure 1D, we have added an external source passing the current  $I_e$ . For this circuit, Kirchoff's and Ohm's laws tells us that  $I_e = I_1 + I_2 = V/R_1 + V/R_2$ . This indicates how resistors add in parallel,  $V = I_e R_1 R_2/(R_1 + R_2)$ .

Next, we consider the electrical circuit in figure 2A, in which a resistor and capacitor are connected together. Kirkoff's laws require that  $I_{\rm C} + I_{\rm R} = 0$ . Putting this together with equations 74 and 76, we find

$$C\frac{dV}{dt} = I_{\rm C} = -I_{\rm R} = -\frac{V}{R} \,. \tag{77}$$

Solving this, gives

$$V(t) = V(0) \exp(-t/RC)$$
 (78)

showing the exponential decay (with time constant  $\tau = RC$ ) of the initial voltage V(0) as the charge on the capacitor leaks out through the resistor.



Figure 2: RC circuits. A) Current  $I_{\rm C} = -I_{\rm R}$  flows in the resistor-capacitor circuit as the stored charge is released. B) Simple passive membrane model including a potential *E* and current source  $I_{\rm e}$ . As in figure 1, the lined triangles represent a ground or point of zero voltage.

Figure 2B includes two extra components needed to build a simple model neuron, the voltage source *E* and the current source  $I_e$ . Using Kirchoff's laws,  $I_e - I_C - I_R = 0$ , and the equation for the voltage *V* is

$$C\frac{dV}{dt} = \frac{E - V}{R} + I_{\rm e} \,. \tag{79}$$

If  $I_{\rm e}$  is constant, the solution of this equation is

$$V(t) = V_{\infty} + (V(0) - V_{\infty}) \exp(-t/\tau)$$
(80)

where  $V_{\infty} = E + RI_{\rm e}$  and  $\tau = RC$ . This shows exponential relaxation from the initial potential V(0) to the equilibrium potential  $V_{\infty}$  at a rate governed by the time constant  $\tau$  of the circuit.

For the case  $I_e = I\cos(\omega t)$ , once an initial transient has decayed to zero, we find

$$V(t) = E + \frac{RI\cos(\omega t - \phi)}{\sqrt{1 + \omega^2 \tau^2}}$$
(81)

where  $\tan(\phi) = \omega \tau$ . Equation 81 shows that the cell membrane acts as a low pass filter, because the higher the frequency  $\omega$  of the input current, the more the attenuation of the oscillations of the potential due to the factor  $1/\sqrt{1 + \omega^2 \tau^2}$ . The phase shift  $\phi$  is an increasing function of frequency that approaches  $\pi/2$  as  $\omega \to \infty$ .

## **Probability Theory**

Probability distributions and densities are discussed extensively in the<br/>text. Here, we present a slightly more formal treatment. At the heart of<br/>probability measureprobability measureProbability theory lie two sets; a sample space,  $\Omega$  and a measure. We be-<br/>gin by considering the simplest case in which the sample space is finite. In<br/>this case, each element  $\omega$  of the full sample space  $\Omega$  can be thought of as<br/>one of the possible outcomes of a random process, for example one results<br/>of rolling five dice. The measure assigns a number  $\gamma_{\omega}$  to each outcome  $\omega$ ,<br/>and these must satisfy  $0 \le \gamma_{\omega} \le 1$  and  $\sum_{\omega} \gamma_{\omega} = 1$ .random variableWe are primarily interested in random variables (which are infamously<br/>neither random nor variable). A random variable is a mapping from a<br/>random outcome  $\omega$  to a space such as the space of integers. An example is<br/>the number of ones that appear when five dice are rolled. Typically, a capi-

the number of ones that appear when five dice are rolled. Typically, a capital letter, such as *S*, is used for the random variable, and the corresponding lower case letter, *s* in this case, is used for a particular value it might take. The probability that *S* takes the value *s* is then written as P[S = s]. In the text, we typically shorten this to P[s], but here we keep the full notation (except in the following table). P[S = s] is determined by the measures of the events for which S = s and takes the value

$$P[S=s] = \sum_{\substack{\omega \text{ with} \\ S(\omega) = s}} \gamma_{\omega} .$$
(82)

The notation  $S(\omega)$  refers to the value of *S* generated by the random event labeled by  $\omega$ , and the sum is over all events for which  $S(\omega) = s$ .

Quantity	Definition	Alias
mean	$\langle s \rangle = \sum_{s} P[s]s$	$\overline{s}, E[S]$
variance	$\operatorname{var}(S) = \langle s^2 \rangle - \langle s \rangle^2 = \sum_s P[s]s^2 - \langle s \rangle^2$	$\sigma_s^2, V[S]$
covariance	$\langle s_1 s_2 \rangle - \langle s_1 \rangle \langle s_2 \rangle = \sum_{s_1 s_2} P[s_1, s_2] s_1 s_2 - \langle s_1 \rangle \langle s_2 \rangle$	$\operatorname{cov}(S_1, S_2)$

Some key statistics for discrete random variables include:

Peter Dayan and L.F. Abbott

independence

continuous

random variable

probability density

where  $S_1$  and  $S_2$  are two random variables defined over the same sample space. This links the two random variables, in that

$$P[S_1 = s_1, S_2 = s_2] = \sum_{\substack{\omega \text{ with} \\ S_1(\omega) = s_1, \\ S_2(\omega) = s_2}} \gamma_{\omega},$$
(83)

and provides a basis for them to be correlated. Means are additive,

$$\langle s_1 + s_2 \rangle = \langle s_1 \rangle + \langle s_2 \rangle, \qquad (84)$$

but other quantities are typically not, for example

$$var(S_1 + S_2) = var(S_1) + var(S_2) + 2cov(S_1, S_2).$$
(85)

Two random variables are independent if  $P[S_1 = s_1, S_2 = s_2] = P[S_1 = s_1]P[S_2 = s_2]$  for all  $s_1$  and  $s_2$ . If  $S_1$  and  $S_2$  are independent,  $cov(S_1, S_2) = 0$ , but the converse is not generally true.

Sample spaces can be infinite or uncountable. In the latter case, there are technical complications that are discussed in the references, but all the sums in the expressions for discrete sample spaces turn into integrals. Under suitable regularity conditions, a continuous random variable *S*, which is a mapping from a sample space to a continuous space such as the real numbers, has a probability density function p[s] defined by

$$p[s] = \lim_{\Delta s \to 0} \left( \frac{P[s \le S \le s + \Delta s]}{\Delta s} \right).$$
(86)

Quantities such as the mean and variance of a continuous random variable are defined as for a discrete random variable but involve integrals over probability densities rather than sums over probabilities.

Name	Range of <i>s</i>	Probability	Mean	Variance
Bernoulli	0 or 1	$p^s(1-p)^{1-s}$	р	p(1-p)
Poisson	positive integer	$\alpha^{s} \exp(-\alpha)/s!$	α	α
Exponential	<i>s</i> > 0	$\alpha \exp(-\alpha s)$	$1/\alpha$	$1/\alpha^2$
Gaussian	$-\infty < s < \infty$	$N[s; g, \Sigma]$	g	Σ
Cauchy	$-\infty < s < \infty$	$\frac{\beta}{\pi((s-\alpha)^2+\beta^2)}$	α	$\infty$

Some commonly used discrete and continuous distributions are:

where

$$N(s; g, \Sigma) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{(s-g)^2}{2\Sigma}\right).$$
(87)

Here, we use  $\Sigma$  to denote the variance of the Gaussian distribution, which is more often written as  $\sigma^2$ . The Cauchy distribution has such heavy tails that the integral defining its variance does not converge.

The Gaussian distribution is particularly important because of the central limit theorem. Consider *m* continuous random variables  $S_1, S_2, S_3, \ldots S_m$  that are independent and have identical distributions with finite mean *g* and variance  $\sigma^2$ . Defining

$$z_m = \frac{1}{m} \sum_{k=1}^m S_k,$$
 (88)

the central limit theorem states that, under rather general conditions,

$$\lim_{m \to \infty} P\left[\frac{\sqrt{m}(z_m - g)}{\sigma} \le s\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} dz \exp(-z^2/2)$$
(89)

for every *s*. This means that, for large *m*,  $z_m$  should be approximately Gaussian distributed with mean *g* and variance  $\sigma^2/m$ .

# **Annotated Bibliography**

Most of the material in this chapter is covered in standard texts on mathematical methods such as **Mathews & Walker (1971)**; **Boas (1996)**. Discussion of relevant computational techniques, and code for implementing them, is available in **Press et al. (1992)**. Linear algebra is covered by **Strang (1976)**; linear and non-linear differential equations by **Jordan & Smith (1977)**; probability theory by **Feller (1968)**; and Fourier transforms and the analysis of linear systems and electrical circuits by **Siebert (1986)**; **Oppenheim & Willsky (1997)**. Mathematical approaches to biological problems are described in **Edelstein-Keshet (1988)**; **Murray (1993)**. Modern techniques of mathematical modeling are described by **Gershenfeld (1999)**.

General references for the other bodies of techniques used in the book include, for statistics, **Lindgren (1993)** and **Cox & Hinckley (1974)**; and, for information theory, **Cover & Thomas (1991)**.

central limit

theorem