# Mathematical Appendix

This book assumes a familiarity with basic methods of linear algebra, differential equations, and probability theory, as covered in standard texts. Here, we describe the notation we use and briefly sketch highlights of various techniques. The references in the bibliography at the end of this appendix provides further information.

## A.1 Linear Algebra

An operation $\mathcal{O}$ on a quantity $z$ is called linear if, applied to any two instances $z_1$ and $z_2$, $\mathcal{O}(\alpha z_1 + \beta z_2) = \alpha \mathcal{O}(z_1) + \beta \mathcal{O}(z_2)$ for any constants $\alpha$ and $\beta$. In this section, we consider linear operations on vectors and functions. We define a vector $\mathbf{v}$ as an array of $N$ numbers $(v_1, v_2, \ldots, v_N)$. The numbers $v_a$ for $a = 1, 2, \ldots, N$ are called the components of the vector. These are sometimes listed in a single $N$-row column

*linear operator*

*vector* $\mathbf{v}$

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

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

*zero vector* $\mathbf{0}$

*spatial vector* $\vec{v}$

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

*norm*

$$|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v} = \sum_{a=1}^{N} v_a^2 = v_1^2 + v_2^2 + \ldots + v_N^2 . \tag{A.2}$$

The dot product of two different $N$-component vectors, $\mathbf{v}$ and $\mathbf{u}$, is

*dot product*

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

*matrix* **W**

Matrix multiplication is a basic linear operation on vectors. An $N_\mathrm{r}$ by $N_\mathrm{c}$ matrix **W** is an array of $N_\mathrm{r}$ rows and $N_\mathrm{c}$ columns

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1N_\mathrm{c}} \\ W_{21} & W_{22} & \dots & W_{2N_\mathrm{c}} \\ & & \vdots & \\ W_{N_\mathrm{r}1} & W_{N_\mathrm{r}2} & \dots & W_{N_\mathrm{r}N_\mathrm{c}} \end{pmatrix} \tag{A.4}$$

with elements $W_{ab}$ for $a = 1, \dots, N_\mathrm{r}$ and $b = 1, \dots, N_\mathrm{c}$. In this text, the product of a matrix and a vector is written as $\mathbf{W} \cdot \mathbf{v}$. The dot implies multiplication and summation over a shared index, as it does for the dot product. If **W** is an $N_\mathrm{r}$ by $N_\mathrm{c}$ matrix and **v** is a $N_\mathrm{c}$-component vector, $\mathbf{W} \cdot \mathbf{v}$ is an $N_\mathrm{r}$-component vector with components

*matrix-vector product*

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

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 $\mathbf{W} \cdot \mathbf{M}$. Matrices can be multiplied in this way only if the number of columns of **W**, $N_\mathrm{c}$, is equal to the number of rows of **M**. Then, $\mathbf{W} \cdot \mathbf{M}$ is a matrix with the same number of rows as **W** and the same number of columns as **M**, and with elements

*matrix product*

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

A vector, written as in equation A.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 $N_\mathrm{r} = N_\mathrm{c} = N$. An important square matrix is the identity matrix **I** with elements

*identity matrix*

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

*Kronecker delta*

where the Kronecker delta is defined as

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

*diagonal matrix*

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

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

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

The transpose of an $N_r$ by $N_c$ matrix $\mathbf{W}$ 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, $\mathbf{v}^T = (v_1 v_2 \ldots v_N)$. This is distinguished by the absence of commas from $(v_1, v_2, \ldots, v_N)$ which, for us, is a listing of the components of a column vector. In the following table, we define a number of operations 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 MATLAB® 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)]'. *transpose*

| Quantity | Definition | Matrix | MATLAB® |
|---|---|---|---|
| norm | $\lvert\mathbf{v}\rvert^2 = \mathbf{v}\cdot\mathbf{v} = \sum_a v_a^2$ | $\mathbf{v}^T\mathbf{v}$ | v'*v |
| dot product | $\mathbf{v}\cdot\mathbf{u} = \sum_a v_a u_a$ | $\mathbf{v}^T\mathbf{u}$ | v'*u |
| outer product | $[\mathbf{vu}]_{ab} = v_a u_b$ | $\mathbf{v}\mathbf{u}^T$ | v*u' |
| matrix-vector product | $[\mathbf{W}\cdot\mathbf{v}]_a = \sum_b W_{ab} v_b$ | $\mathbf{Wv}$ | W*v |
| vector-matrix product | $[\mathbf{v}\cdot\mathbf{W}]_a = \sum_b v_b W_{ba}$ | $\mathbf{v}^T\mathbf{W}$ | v'*W |
| quadratic form | $\mathbf{v}\cdot\mathbf{W}\cdot\mathbf{u} = \sum_{ab} v_a W_{ab} u_b$ | $\mathbf{v}^T\mathbf{Wu}$ | v'*W*u |
| matrix-matrix product | $[\mathbf{W}\cdot\mathbf{M}]_{ab} = \sum_c W_{ac} M_{cb}$ | $\mathbf{WM}$ | W*M |
| transpose | $[\mathbf{W}^T]_{ab} = W_{ba}$ | $\mathbf{W}^T$ | W' |

Several important definitions for square matrices are given below.

| Operation | Notation | Definition | MATLAB® |
|---|---|---|---|
| inverse | $\mathbf{W}^{-1}$ | $\mathbf{W}\cdot\mathbf{W}^{-1} = \mathbf{I}$ | `inv`(W) |
| trace | tr$\mathbf{W}$ | $\sum_a W_{aa}$ | `trace`(W) |
| determinant | det$\mathbf{W}$ | see references | `det`(W) |

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

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

where $f(a-b)$ is any function of the single variable $a-b$.

*del operator* $\nabla$

For any real-valued function $E(\mathbf{v})$ of a vector $\mathbf{v}$, we can define the vector derivative (which is sometimes called del) of $E(\mathbf{v})$ as the vector $\nabla E(\mathbf{v})$ with components

$$[\nabla E(\mathbf{v})]_a = \frac{\partial E(\mathbf{v})}{\partial v_a}\,. \tag{A.10}$$

*directional derivative*

The derivative of $E(\mathbf{v})$ in the direction $\mathbf{u}$ is then

$$\lim_{\epsilon\to 0}\left(\frac{E(\mathbf{v}+\epsilon\mathbf{u})-E(\mathbf{v})}{\epsilon}\right) = \mathbf{u}\cdot\nabla E(\mathbf{v})\,. \tag{A.11}$$

## Eigenvectors and Eigenvalues

*eigenvector*

An eigenvector of a square matrix $\mathbf{W}$ is a nonzero vector $\mathbf{e}$ that satisfies

$$\mathbf{W}\cdot\mathbf{e} = \lambda\mathbf{e} \tag{A.12}$$

*eigenvalue*

for some number $\lambda$ called the eigenvalue. Possible values of $\lambda$ are determined by solving the polynomial equation

$$\det(\mathbf{W}-\lambda\mathbf{I}) = 0\,. \tag{A.13}$$

Typically, but not always, this has $N$ solutions if $\mathbf{W}$ is an $N$ by $N$ matrix, and these can be either real or complex. Complex eigenvalues come in complex-conjugate pairs if $\mathbf{W}$ 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 $\mathbf{e}_\mu$.

*degeneracy*

*linear independence*

If $\mathbf{e}$ is an eigenvector, $\alpha\mathbf{e}$ 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 0. Apart from such degeneracies, an $N$ by $N$ matrix can have at most $N$ eigenvectors, although some matrices have fewer. If $\mathbf{W}$ has $N$ nondegenerate 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 $c_\mu = 0$ for all $\mu$. These eigenvectors can be used to represent any $N$ component vector $\mathbf{v}$ through the relation

$$\mathbf{v} = \sum_{\mu=1}^{N} c_\mu\mathbf{e}_\mu\,, \tag{A.14}$$

*basis*

with a unique set of coefficients $c_\mu$. They are thus said to form a basis for the set of vectors $\mathbf{v}$.

*symmetric matrix*

The 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-

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

*orthonormal eigenvectors*

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

To derive this result we note that, if $\mathbf{W}$ is a symmetric matrix, we can write $\mathbf{e}_\mu \cdot \mathbf{W} = \mathbf{W} \cdot \mathbf{e}_\mu = \lambda_\mu \mathbf{e}_\mu$. 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 $\mathbf{e}_\nu \cdot \mathbf{e}_\mu = 0$. For orthogonal and normalized (orthonormal) eigenvectors, the coefficients in equation A.14 take the values

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

Let $\mathbf{E} = (\mathbf{e}_1\ \mathbf{e}_2\ \ldots\ \mathbf{e}_N)$ be an $N$ by $N$ matrix with columns formed from the orthonormal eigenvectors of a symmetric matrix. From equation A.15, this satisfies $[\mathbf{E}^{\mathrm{T}} \cdot \mathbf{E}]_{\mu\nu} = \mathbf{e}_\mu \cdot \mathbf{e}_\nu = \delta_{\mu\nu}$. Thus, $\mathbf{E}^{\mathrm{T}} = \mathbf{E}^{-1}$, 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,

*matrix diagonalization*

$$\mathbf{E}^{-1} \cdot \mathbf{W} \cdot \mathbf{E} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N) . \tag{A.17}$$

Conversely,

$$\mathbf{W} = \mathbf{E} \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_N) \cdot \mathbf{E}^{-1} . \tag{A.18}$$

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

$$\begin{aligned} \mathbf{W}^n &= (\mathbf{E} \cdot \mathbf{L} \cdot \mathbf{E}^{-1}) \cdot (\mathbf{E} \cdot \mathbf{L} \cdot \mathbf{E}^{-1}) \cdots (\mathbf{E} \cdot \mathbf{L} \cdot \mathbf{E}^{-1}) \\ &= \mathbf{E} \cdot \mathbf{L}^n \cdot \mathbf{E}^{-1} = \mathbf{E} \cdot \mathrm{diag}(\lambda_1^n, \ldots, \lambda_N^n) \cdot \mathbf{E}^{-1} . \end{aligned} \tag{A.19}$$

This result serves as a basis for defining functions of matrices. For any function $f$ that can be written as a power or expanded in a power series (including, for example, exponentials and logarithms),

$$f(\mathbf{W}) = \mathbf{E} \cdot \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_N)) \cdot \mathbf{E}^{-1} . \tag{A.20}$$

## Functional Analogs

A function $v(t)$ can be treated as if it were a vector with a continuous label. 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, $\sum_a \to \int dt$. For example, the functional analog of the squared norm and dot product are

*functions as vectors*

$$\int_{-\infty}^{\infty} dt\, v^2(t) \quad \text{and} \quad \int_{-\infty}^{\infty} dt\, v(t) u(t) . \tag{A.21}$$

*linear integral operator*

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

$$\int_{-\infty}^{\infty} dt'\, W(t,t')v(t') \tag{A.22}$$

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 some 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

$$\int_{-\infty}^{\infty} dt''\, W^{-1}(t,t'')W(t'',t') = \delta(t-t')\,, \tag{A.23}$$

*translation invariance*

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

$$W(t,t') = K(t-t')\,. \tag{A.24}$$

*linear filter*

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

$$\int_{-\infty}^{\infty} dt'\, K(t-t')v(t') = \int_{-\infty}^{\infty} d\tau\, K(\tau)v(t-\tau)\,, \tag{A.25}$$

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

### 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 0 everywhere except where its argument is 0, and there it grows without bound. The infinite height and infinitesimal width of this function are matched so that its integral is 1. Thus,

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

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

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

for any value of $t$ contained within the integration interval (if $t$ is not within this interval, the integral is 0). These two identities normally 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 1 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

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

or a Gaussian function

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

*$\delta$ function definition*

It is most often expressed as

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

This underlies the inverse Fourier transform, as discussed below.

**Eigenfunctions**

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

$$\int dt' \, W(t, t') e(t') = \lambda e(t) . \tag{A.31}$$

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

$$\int dt' \, K(t - t') \exp(i\omega t') = \left( \int d\tau \, K(\tau) \exp(-i\omega\tau) \right) \exp(i\omega t) , \tag{A.32}$$

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

*complex exponential*

$$\exp(i\omega t) = \cos(\omega t) + i \sin(\omega t) . \tag{A.33}$$

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

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

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

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

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

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

*inverse Fourier transform*

The Fourier transform $\tilde{f}(\omega)$ provides an alternative representation of the original function $f(t)$ because it can be inverted through

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

This provides an inverse because

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \exp(-i\omega t) \int_{-\infty}^{\infty} dt'\, f(t') \exp(i\omega t') \tag{A.37}$$

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

by the definition of the $\delta$ function in equation A.30. The function $f(t)$ has to satisfy a set of criteria 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

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

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 A.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,

$$\tilde{h}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega)\,. \tag{A.39}$$

To show this, we note that

$$\tilde{h}(\omega) = \int_{-\infty}^{\infty} dt \exp(i\omega t) \int_{-\infty}^{\infty} d\tau\, f(\tau) g(t-\tau) \tag{A.40}$$

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

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

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

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

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

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

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

$$\tilde{f}_k = \frac{1}{T} \int_0^T dt\, f(t) \exp(i2\pi kt/T) \,. \tag{A.43}$$

As in the case of Fourier transforms, certain 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

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

with

$$\begin{aligned} \tilde{f}_0 &= \frac{1}{T} \int_0^T dt\, f(t) \,, \quad \tilde{f}_k^{\mathrm{c}} = \frac{2}{T} \int_0^T dt\, f(t) \cos(2\pi kt/T) \,, \\ \tilde{f}_k^{\mathrm{s}} &= \frac{2}{T} \int_0^T dt\, f(t) \sin(2\pi kt/T) \,. \end{aligned} \tag{A.45}$$

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 $n = 0, 1, \ldots N_t - 1$. 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 *discrete Fourier transform*

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

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

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

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 in the bibliography 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 defined only 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 $1/\delta$.

*sampling theorem*

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

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

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

## A.2 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 $\mathbf{W}$ that makes the product $\mathbf{W} \cdot \mathbf{v}$ closest to another vector $\mathbf{u}$ when averaged over a number of presentations of $\mathbf{v}$ and $\mathbf{u}$. The function to be minimized is the average squared error $\langle |\mathbf{W} \cdot \mathbf{v} - \mathbf{u}|^2 \rangle$, where the brackets denote averaging over all the different samples $\mathbf{v}$ and $\mathbf{u}$. Setting the derivative of this expression with respect to $\mathbf{W}$ (or equivalently its elements $W_{ab}$) to 0 gives the equation

*minimization of quadratic form*

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

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 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 specified value. This procedure can appear a bit mysterious when first encountered, so we provide a rather extended discussion.

*Lagrange multiplier*

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

$$\sum_{a=1}^{N}[\nabla f]_a \Delta v_a = 0\,, \tag{A.50}$$

where we use the notation

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

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

$$[\nabla f]_a = 0 \tag{A.52}$$

for all $a$. However, with a constraint such as $g(\mathbf{v})$=constant, equation A.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 $g$ cannot change to first order in $\Delta\mathbf{v}$. Therefore,

$$\sum_{a=1}^{N}[\nabla g]_a \Delta v_a = 0 \tag{A.53}$$

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

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

$$\Delta v_c = -\frac{1}{[\nabla g]_c}\sum_{a\neq c}[\nabla g]_a \Delta v_a\,. \tag{A.54}$$

Then we substitute this expression into equation A.50 to obtain

$$\sum_{a\neq c}[\nabla f]_a \Delta v_a - \frac{[\nabla f]_c}{[\nabla g]_c}\sum_{a\neq c}[\nabla g]_a \Delta v_a = 0\,. \tag{A.55}$$

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

$$[\nabla f]_a - \frac{[\nabla f]_c}{[\nabla g]_c}[\nabla g]_a = 0 \tag{A.56}$$

for all $a \neq c$. The derivatives of $f$ and $g$ are functions of $\mathbf{v}$, 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

$$\lambda = -\frac{[\nabla f]_c}{[\nabla g]_c}\,. \tag{A.57}$$

With this notation, the final set of equations (A.56) can be written as

$$[\nabla f]_a + \lambda[\nabla g]_a = 0\,. \tag{A.58}$$

Before, we had to say that these equations held only 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 in equation A.57. Thus, we can say that equation A.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 A.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 forgotten about the constraint entirely. Of course this lunch is not completely free. From equation A.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 value. This provides the solution to the constrained problem.

## A.3 Differential Equations

The most general differential equation we consider takes the form

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

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

*fixed point*

*limit cycle*

*chaos*

*strange attractor*

We focus most of our analysis on fixed point solutions, which are also called equilibrium points. For $\mathbf{v}_\infty$ to be a time-independent solution

*equilibrium point*

of equation A.59, we must have $\mathbf{f}(\mathbf{v}_\infty) = 0$. General solutions of equation A.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 the solutions we obtain as approximations in the nonlinear case are exact. Near the fixed point $\mathbf{v}_\infty$, we write

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

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

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

where $\mathbf{J}$ is the called the Jacobian matrix and has elements *Jacobian matrix*

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

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

Using the approximation of equation A.61, equation A.59 becomes

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

The temporal evolution of $\mathbf{v}(t)$ is best understood by expanding $\boldsymbol{\epsilon}$ in the basis provided by the eigenvectors of $\mathbf{J}$. Assuming that $\mathbf{J}$ 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

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

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

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

This produces the solution

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

where $\boldsymbol{\epsilon}(0) = \sum_\mu c_\mu(0)\mathbf{e}_\mu$. The individual terms in the sum on the right side of equation A.66 are called modes. This solution is exact for equation A.63, but is only a valid approximation when applied to equation A.59 if $\boldsymbol{\epsilon}$ 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 modes come *modes*

in conjugate pairs that combine to form a real function. Expression A.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 A.66 determines how the evolution of $\mathbf{v}(t)$ in the neighborhood of $\mathbf{v}_\infty$ depends on the eigenvalues of $\mathbf{J}$. If we write $\lambda_\mu = \alpha_\mu + i\omega_\mu$,

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

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 $\mathbf{J}$ is a symmetric matrix. Modes with negative real eigenvalues ($\alpha_\mu < 0$ and $\omega_\mu = 0$) decay exponentially to 0, 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 0 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 $\omega_\mu \neq 0$).

*attractor*

*unstable fixed point*

*marginal stability*

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 starts from any point in the neighborhood of $\mathbf{v}_\infty$. If any real part is positive ($\alpha_\mu > 0$ 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 $\mathbf{f}$ is linear, the exponential growth of $|\mathbf{v}(t) - \mathbf{v}_\infty|$ never stops in this case. For a nonlinear $f$, equation A.66 determines what happens only 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 0, the 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

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

This can be solved by setting $z = v - v_\infty$, rewriting the equation as $dz/z = -dt/\tau$, and integrating both sides:

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

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

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

In some cases, we consider discrete rather than continuous dynamics defined over discrete steps $n = 1, 2, \ldots$ 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*

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

The strategy for solving this equation is similar to that for solving differential equations. Assuming $\mathbf{W}$ 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

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

where $\mathbf{v}(0) = \sum_\mu c_\mu \mathbf{e}_\mu$. This has characteristics similar to equation A.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

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

If this is greater than 1 for any value of $\mu$, $|\mathbf{v}(n)| \to \infty$ as $n \to \infty$. If it is less than 1 for all $\mu$, $\mathbf{v}(n) \to \mathbf{0}$ in this limit.

## A.4 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 A.1A and A.1B show the standard symbols for resistors and capacitors, and define the relevant voltages and currents. A resistor (figure A.1A) satisfies Ohm's law, which states that the voltage $V_\mathrm{R} = V_1 - V_2$ across a resistance $R$ carrying a current $I_\mathrm{R}$ is *Ohm's law*

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

Resistance is measured in ohms ($\Omega$); 1 ohm is the resistance through which 1 ampere of current causes a voltage drop of 1 volt (1 V = 1 A $\times$ 1 $\Omega$).

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

$$CV_\mathrm{C} = Q_\mathrm{C} \,, \tag{A.75}$$

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 A.75 and use the fact that current is equal to the rate of

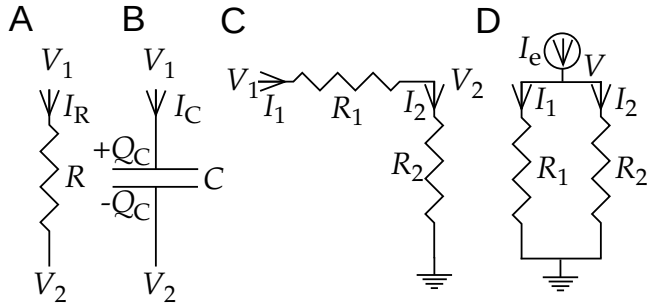


Figure A.1 Electrical circuit elements and resistor circuits. (A) Current $I_R$ flows through a resistance $R$, producing a voltage drop $V_1 - V_2 = V_R$. (B) Charge $\pm Q_C$ is stored across a capacitance $C$, leading to a voltage $V_C = V_1 - V_2$ and a current $I_C$. (C) Series resistor circuit called a voltage divider. (D) Parallel resistor circuit. $I_e$ represents an external current source. The lined triangle symbol at the bottom of the circuits in C and D represents an electrical ground, which is defined to be at 0 voltage.

change of charge, $I_C = dQ_C/dt$, to obtain the basic voltage-current relationship for a capacitor,

$$C\frac{dV_C}{dt} = I_C\,. \tag{A.76}$$

*V-I relation for capacitor*

Capacitance is measured in units of farads (F), defined as the capacitance for which 1 ampere of current causes a voltage change of 1 volt per second (1 F × 1 V/s = 1 A).

*Kirchhoff's laws*

The voltages at different points in a circuit and the currents flowing through various circuit elements can be computed using equations A.74 and A.76 and rules called Kirchhoff's laws. These state that voltage differences around any closed loop in a circuit must sum to 0, and that the sum of all the currents entering any point in a circuit must be 0. Applying the second of these rules to the circuit in figure A.1C, we find that $I_1 = I_2$. Ohm's law tells us that $V_1 - V_2 = I_1R_1$ and $V_2 = I_2R_2$. Solving these gives $V_1 = I_1(R_1 + R_2)$, which tells us that resistors arranged in series add, and $V_2 = V_1R_2/(R_1 + R_2)$, which is why this circuit is called a voltage divider.

In the circuit of figure A.1D, we have added an external source passing the current $I_e$. For this circuit, Kirchhoff'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, $V = I_eR_1R_2/(R_1 + R_2)$.

Next, we consider the electrical circuit in figure A.2A, in which a resistor and capacitor are connected together. Kirchhoff's laws require that $I_C + I_R = 0$. Putting this together with equations A.74 and A.76, we find

$$C\frac{dV}{dt} = I_C = -I_R = -\frac{V}{R}\,. \tag{A.77}$$

Solving this gives

$$V(t) = V(0)\exp(-t/RC)\,, \tag{A.78}$$

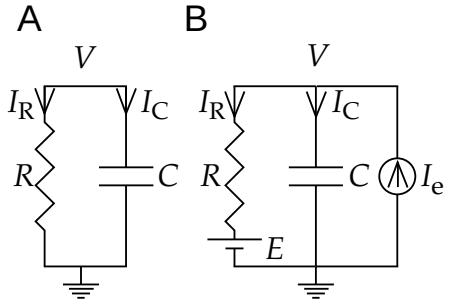


Figure A.2 RC circuits. (A) Current $I_C = -I_R$ flows in the resistor-capacitor circuit as the stored charge is released. (B) Simple passive membrane model including a potential $E$ and current source $I_e$. As in figure A.1, the lined triangles represent a ground or point of 0 voltage.

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

$$C\frac{dV}{dt} = \frac{E - V}{R} + I_e\,. \tag{A.79}$$

If $I_e$ is constant, the solution of this equation is

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

where $V_\infty = E + RI_e$ 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 0, we find

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

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

## A.5 Probability Theory

Probability distributions and densities are discussed extensively in the text. Here, we present a slightly more formal treatment. At the heart of

probability theory lie two objects: a sample space, $\Omega$, and a measure. We begin by considering the simplest case of a finite sample space. Here, each element $\omega$ of the full sample space $\Omega$ can be thought of as one of the possible outcomes of a random process, for example, one of the $6^5$ possible results of rolling five dice. The measure assigns a number $\gamma_\omega$ to each outcome $\omega$, and these must satisfy $0 \leq \gamma_\omega \leq 1$ and $\sum_\omega \gamma_\omega = 1$.

*sample space*

*probability measure*

*random variable*

We are primarily interested in random variables (which are infamously neither random nor variable). A random variable is a mapping from a random outcome $\omega$ to a space such as the space of integers. An example is 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 lowercase 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

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

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 $S(\omega) = s$.

Some key statistics for discrete random variables include the following.

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

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

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

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

$$\langle s_1 + s_2 \rangle = \langle s_1 \rangle + \langle s_2 \rangle \,, \tag{A.84}$$

but other quantities typically are not, for example,

$$\mathrm{var}(S_1 + S_2) = \mathrm{var}(S_1) + \mathrm{var}(S_2) + 2\mathrm{cov}(S_1, S_2) \,. \tag{A.85}$$

*independence*

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, $\mathrm{cov}(S_1, S_2) = 0$, but the converse is not true in general.

In addition to finite, sample spaces can be either countably or uncountably infinite. In the latter case, there are various technical complications that are discussed in the references. Under suitable 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

*continuous random variable*

*probability density*

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

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.

Some commonly used discrete and continuous distributions are listed in the table below.

| Name | Range of $s$ | Probability | Mean | Variance |
|---|---|---|---|---|
| Bernoulli | $s = 0$ or 1 | $p^s(1-p)^{1-s}$ | $p$ | $p(1-p)$ |
| Poisson | $s = 0, 1, 2, \ldots$ | $\alpha^s \exp(-\alpha)/s!$ | $\alpha$ | $\alpha$ |
| Exponential | $s > 0$ | $\alpha \exp(-\alpha s)$ | $1/\alpha$ | $1/\alpha^2$ |
| Gaussian | $-\infty < s < \infty$ | $\mathcal{N}[s; g, \Sigma]$ | $g$ | $\Sigma$ |
| Cauchy | $-\infty < s < \infty$ | $\beta/(\pi((s-\alpha)^2 + \beta^2))$ | $*\,\alpha\,*$ | $*\,1/\beta^2\,*$ |

where

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

Here, we use $\Sigma$ to denote the variance of the Gaussian distribution, which is more often written as $\sigma^2$. The asterisks in the entries for the Cauchy distribution reflect the fact that it has such heavy tails that the integrals defining its mean and variance do not converge. Nevertheless, $\alpha$ and $1/\beta^2$ play similar roles, and are called location and scale parameters respectively.

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

*central limit theorem*

$$Z_m = \frac{1}{m} \sum_{k=1}^{m} S_k , \tag{A.88}$$

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

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

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