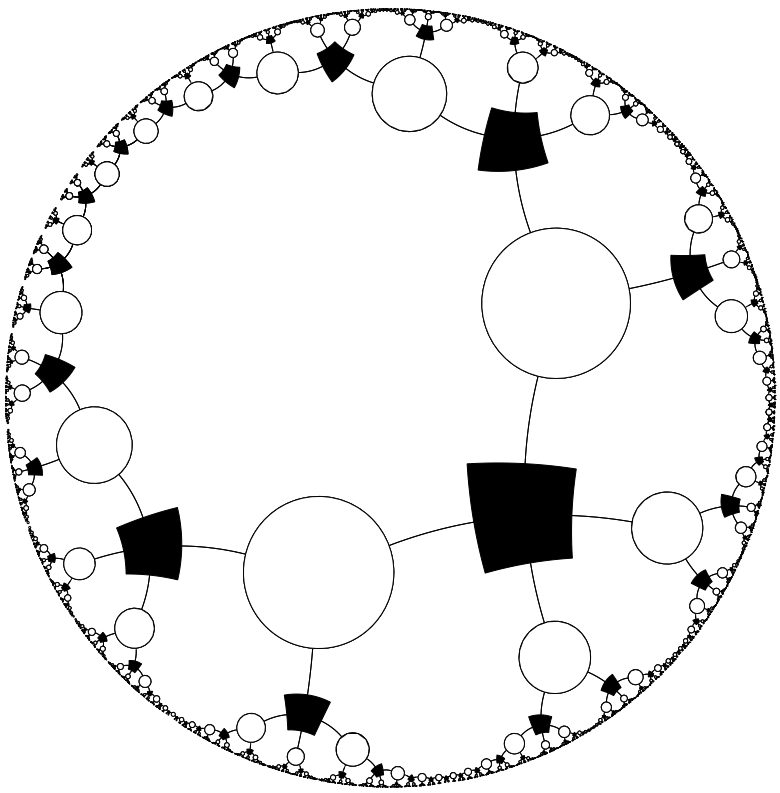
Part VII

Appendices



A

Notation

What does $P(A \mid B, C)$ mean? $P(A \mid B, C)$ is pronounced 'the probability that A is true *given that* B is true *and* C is true'. Or, more briefly, 'the probability of A given B and C'. (See Chapter 2, p.22.)

What do log **and** ln **mean?** In this book, $\log x$ means the base-two logarithm, $\log_2 x$; $\ln x$ means the natural logarithm, $\log_e x$.

What does $\hat{s}$ mean? Usually, a 'hat' over a variable denotes a guess or estimator. So $\hat{s}$ is a guess at the value of s.

Integrals. There is no difference between $\int f(u)\,\mathrm{d}u$ and $\int \mathrm{d}u\, f(u)$. The integrand is $f(u)$ in both cases.

What does $\prod_{n=1}^{N}$ mean? This is like the summation $\sum_{n=1}^{N}$ but it denotes a product. It's pronounced 'product over n from 1 to N'. So, for example,

$$\prod_{n=1}^{N} n = 1 \times 2 \times 3 \times \cdots \times N = N! = \exp\left[\sum_{n=1}^{N} \ln n\right]. \tag{A.1}$$

I like to choose the name of the free variable in a sum or a product – here, n – to be the lower case version of the range of the sum. So n usually runs from 1 to N, and m usually runs from 1 to M. This is a habit I learnt from Yaser Abu-Mostafa, and I think it makes formulae easier to understand.

What does $\binom{N}{n}$ mean? This is pronounced 'N choose n', and it is the number of ways of selecting an unordered set of n objects from a set of size N.

$$\binom{N}{n} = \frac{N!}{(N-n)!\,n!}. \tag{A.2}$$

This function is known as the combination function.

What is $\Gamma(x)$? The *gamma function* is defined by $\Gamma(x) \equiv \int_0^\infty \mathrm{d}u\; u^{x-1} e^{-u}$, for $x > 0$. The gamma function is an extension of the factorial function to real number arguments. In general, $\Gamma(x+1) = x\Gamma(x)$, and for integer arguments, $\Gamma(x+1) = x!$. The digamma function is defined by $\Psi(x) \equiv \frac{\mathrm{d}}{\mathrm{d}x} \ln \Gamma(x)$.

For large x (for practical purposes, $0.1 \leq x \leq \infty$),

$$\ln \Gamma(x) \simeq \left(x - \tfrac{1}{2}\right) \ln(x) - x + \tfrac{1}{2} \ln 2\pi + O(1/x); \tag{A.3}$$

and for small x (for practical purposes, $0 \leq x \leq 0.5$):

$$\ln \Gamma(x) \simeq \ln \frac{1}{x} - \gamma_e x + O(x^2) \tag{A.4}$$

where γ_e is Euler's constant.

What does $H_2^{-1}(1 - R/C)$ mean? Just as $\sin^{-1}(s)$ denotes the inverse function to $s = \sin(x)$, so $H_2^{-1}(h)$ is the inverse function to $h = H_2(x)$.

There is potential confusion when people use $\sin^2 x$ to denote $(\sin x)^2$, since then we might expect $\sin^{-1} s$ to denote $1/\sin(s)$; I therefore like to avoid using the notation $\sin^2 x$.

What does $f'(x)$ mean? The answer depends on the context. Often, a 'prime' is used to denote differentiation:

$$f'(x) \equiv \frac{\mathrm{d}}{\mathrm{d}x} f(x); \tag{A.5}$$

similarly, a dot denotes differentiation with respect to time, t:

$$\dot{x} \equiv \frac{\mathrm{d}}{\mathrm{d}t} x. \tag{A.6}$$

However, the prime is also a useful indicator for 'another variable', for example 'a new value for a variable'. So, for example, x' might denote 'the new value of x'. Also, if there are two integers that both range from 1 to N, I will often name those integers n and n'.

So my rule is: if a prime occurs in an expression that could be a function, such as $f'(x)$ or $h'(y)$, then it denotes differentiation; otherwise it indicates 'another variable'.

What is the error function? Definitions of this function vary. I define it to be the cumulative probability of a standard (variance = 1) normal distribution,

$$\Phi(z) \equiv \int_{-\infty}^{z} \exp(-z^2/2)/\sqrt{2\pi}\, \mathrm{d}z. \tag{A.7}$$

What does $\mathcal{E}(r)$ mean? $\mathcal{E}[r]$ is pronounced 'the expected value of r' or 'the expectation of r', and it is the mean value of r. Another symbol for 'expected value' is the pair of angle-brackets, $\langle r \rangle$.

What does $|x|$ mean? The vertical bars '$|\cdot|$' have two meanings. If $\mathcal{A}$ is a set, then $|\mathcal{A}|$ denotes the number of elements in the set; if x is a number, then $|x|$ is the absolute value of x.

What does $[\mathbf{A}|\mathbf{P}]$ mean? Here, $\mathbf{A}$ and $\mathbf{P}$ are matrices with the same number of rows. $[\mathbf{A}|\mathbf{P}]$ denotes the double-width matrix obtained by putting $\mathbf{A}$ alongside $\mathbf{P}$. The vertical bar is used to avoid confusion with the product $\mathbf{AP}$.

What does $\mathbf{x}^{\mathsf{T}}$ mean? The superscript T is pronounced 'transpose'. Transposing a row-vector turns it into a column vector:

$$(1, 2, 3)^{\mathsf{T}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \tag{A.8}$$

and *vice versa*. [Normally my vectors, indicated by bold face type ($\mathbf{x}$), are column vectors.]

Similarly, matrices can be transposed. If M_{ij} is the entry in row i and column j of matrix $\mathbf{M}$, and $\mathbf{N} = \mathbf{M}^{\mathsf{T}}$, then $N_{ji} = M_{ij}$.

What are Trace $\mathbf{M}$ **and** det $\mathbf{M}$**?** The trace of a matrix is the sum of its diagonal elements,

$$\text{Trace}\,\mathbf{M} = \sum_i M_{ii}. \tag{A.9}$$

The determinant of $\mathbf{M}$ is denoted $\det \mathbf{M}$.

What does δ_{mn} mean? The δ matrix is the identity matrix.

$$\delta_{mn} = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n. \end{cases}$$

Another name for the identity matrix is $\mathbf{I}$ or $\mathbf{1}$. Sometimes I include a subscript on this symbol – $\mathbf{1}_K$ – which indicates the size of the matrix $(K \times K)$.

What does $\delta(x)$ mean? The delta function has the property

$$\int \mathrm{d}x\, f(x)\delta(x) = f(0). \tag{A.10}$$

Another possible meaning for $\delta(S)$ is the truth function, which is 1 if the proposition S is true but I have adopted another notation for that. After all, the symbol δ is quite busy already, with the two roles mentioned above in addition to its role as a small real number δ and an increment operator (as in δx)!

What does $\mathbb{1}[S]$ mean? $\mathbb{1}[S]$ is the truth function, which is 1 if the proposition S is true and 0 otherwise. For example, the number of positive numbers in the set $T = \{-2, 1, 3\}$ can be written

$$\sum_{x \in T} \mathbb{1}[x > 0]. \tag{A.11}$$

What is the difference between ':=' and '='? In an algorithm, $x := y$ means that the variable x is updated by assigning it the value of y.

In contrast, $x = y$ is a proposition, a statement that x is equal to y.

See Chapters 23 and 29 for further definitions and notation relating to probability distributions.

B

Some Physics

B.1 About phase transitions

A system with states $\mathbf{x}$ in contact with a heat bath at temperature $T = 1/\beta$ has probability distribution

$$P(\mathbf{x} \,|\, \beta) = \frac{1}{Z(\beta)} \exp(-\beta E(\mathbf{x})). \tag{B.1}$$

The partition function is

$$Z(\beta) = \sum_{\mathbf{x}} \exp(-\beta E(\mathbf{x})). \tag{B.2}$$

The inverse temperature β can be interpreted as defining an exchange rate between entropy and energy. $(1/\beta)$ is the amount of energy that must be given to a heat bath to increase its entropy by one nat.

Often, the system will be affected by some other parameters such as the volume of the box it is in, V, in which case Z is a function of V too, $Z(\beta, V)$.

For any system with a finite number of states, the function $Z(\beta)$ is evidently a continuous function of β, since it is simply a sum of exponentials. Moreover, all the derivatives of $Z(\beta)$ with respect to β are continuous too.

What phase transitions are all about, however, is this: phase transitions correspond to values of β and V (called critical points) at which the derivatives of Z have discontinuities or divergences.

Immediately we can deduce:

> Only systems with an infinite number of states can show phase transitions.

Often, we include a parameter N describing the size of the system. Phase transitions may appear in the limit $N \to \infty$. Real systems may have a value of N like 10^{23}.

If we make the system large by simply grouping together N independent systems whose partition function is $Z_{(1)}(\beta)$, then nothing interesting happens. The partition function for N independent identical systems is simply

$$Z_{(N)}(\beta) = [Z_{(1)}(\beta)]^N. \tag{B.3}$$

Now, while this function $Z_{(N)}(\beta)$ may be a very rapidly varying function of β, that doesn't mean it is showing phase transitions. The natural way to look at the partition function is in the logarithm

$$\ln Z_{(N)}(\beta) = N \ln Z_{(1)}(\beta). \tag{B.4}$$

Duplicating the original system N times simply scales up all properties like the energy and heat capacity of the system by a factor of N. So if the original system showed no phase transitions then the scaled up system won't have any either.

Only systems with long-range correlations show phase transitions.

Long-range correlations do not require long-range energetic couplings; for example, a magnet has only short-range couplings (between adjacent spins) but these are sufficient to create long-range order.

Why are points at which derivatives diverge interesting?

The derivatives of $\ln Z$ describe properties like the heat capacity of the system (that's the second derivative) or its fluctuations in energy. If the second derivative of $\ln Z$ diverges at a temperature $1/\beta$, then the heat capacity of the system diverges there, which means it can absorb or release energy without changing temperature (think of ice melting in ice water); when the system is at equilibrium at that temperature, its energy fluctuates a lot, in contrast to the normal law-of-large-numbers behaviour, where the energy only varies by one part in $\sqrt{N}$.

A toy system that shows a phase transition

Imagine a collection of N coupled spins that have the following energy as a function of their state $\mathbf{x} \in \{0,1\}^N$.

$$E(\mathbf{x}) = \begin{cases} -N\epsilon & \mathbf{x} = (0,0,0,\ldots,0) \\ 0 & \text{otherwise.} \end{cases} \tag{B.5}$$

This energy function describes a ground state in which all the spins are aligned in the zero direction; the energy per spin in this state is $-\epsilon$. if any spin changes state then the energy is zero. This model is like an extreme version of a magnetic interaction, which encourages pairs of spins to be aligned.

We can contrast it with an ordinary system of N *independent* spins whose energy is:

$$E^0(\mathbf{x}) = \epsilon \sum_n (2x_n - 1). \tag{B.6}$$

Like the first system, the system of independent spins has a single ground state $(0,0,0,\ldots,0)$ with energy $-N\epsilon$, and it has roughly 2^N states with energy very close to 0, so the low-temperature and high-temperature properties of the independent-spin system and the coupled-spin system are virtually identical.

The partition function of the coupled-spin system is

$$Z(\beta) = e^{\beta N\epsilon} + 2^N - 1. \tag{B.7}$$

The function

$$\ln Z(\beta) = \ln \left(e^{\beta N\epsilon} + 2^N - 1 \right) \tag{B.8}$$

is sketched in figure B.1a along with its low temperature behaviour,

$$\ln Z(\beta) \simeq N\beta\epsilon, \quad \beta \to \infty, \tag{B.9}$$

and its high temperature behaviour,

$$\ln Z(\beta) \simeq N \ln 2, \quad \beta \to 0. \tag{B.10}$$

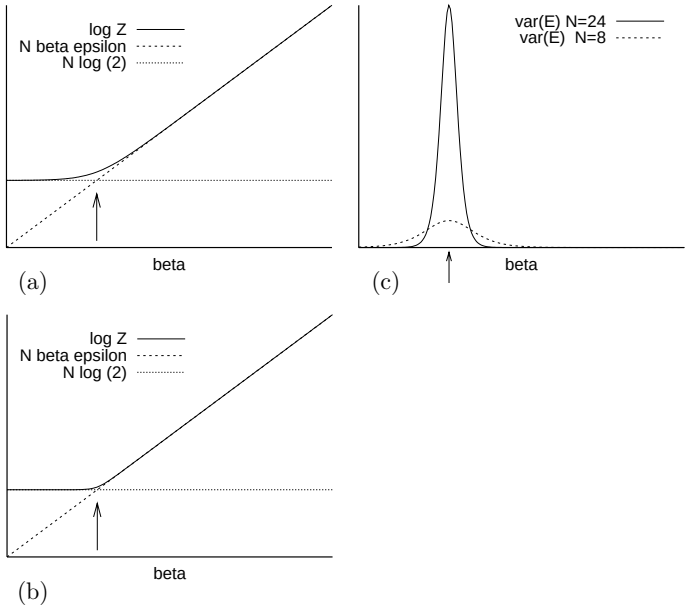


Figure B.1. (a) Partition function of toy system which shows a phase transition for large N. The arrow marks the point $\beta_c = \log 2/\epsilon$. (b) The same, for larger N. (c) The variance of the energy of the system as a function of β for two system sizes. As N increases the variance has an increasingly sharp peak at the critical point β_c. Contrast with figure B.2.

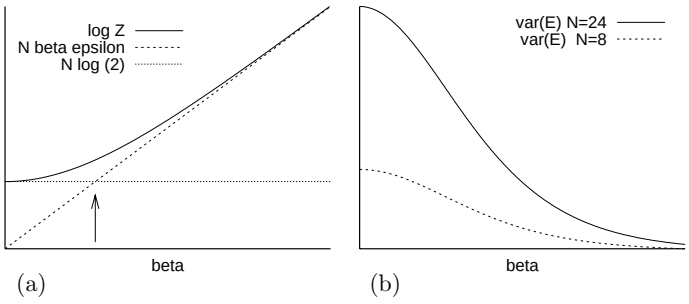


Figure B.2. The partition function (a) and energy-variance (b) of a system consisting of N independent spins. The partition function changes gradually from one asymptote to the other, regardless of how large N is; the variance of the energy does not have a peak. The fluctuations are largest at high temperature (small β) and scale linearly with system size N.

The arrow marks the point

$$\beta = \frac{\ln 2}{\epsilon} \tag{B.11}$$

at which these two asymptotes intersect. In the limit $N \to \infty$, the graph of $\ln Z(\beta)$ becomes more and more sharply bent at this point (figure B.1b).

The second derivative of $\ln Z$, which describes the variance of the energy of the system, has a peak value, at $\beta = \ln 2/\epsilon$, roughly equal to

$$\frac{N^2\epsilon^2}{4}, \tag{B.12}$$

which corresponds to the system spending half of its time in the ground state and half its time in the other states.

At this critical point, the heat capacity of this system is thus proportional to N^2; the heat capacity per spin is proportional to N, which, for infinite N, is infinite, in contrast to the behaviour of systems away from phase transitions, whose capacity per atom is a finite number.

For comparison, figure B.2 shows the partition function and energy-variance of the ordinary independent-spin system.

More generally

Phase transitions can be categorized into 'first-order' and 'continuous' transitions. In a first-order phase transition, there is a discontinuous change of one

or more order-parameters; in a continuous transition, all order-parameters change continuously. [What's an order-parameter? – a scalar function of the state of the system; or, to be precise, the expectation of such a function.]

In the vicinity of a critical point, the concept of 'typicality' defined in Chapter 4 does not hold. For example, our toy system, at its critical point, has a 50% chance of being in a state with energy $-N\epsilon$, and roughly a $1/2^{N+1}$ chance of being in each of the other states that have energy zero. It is thus not the case that $\ln 1/P(\mathbf{x})$ is very likely to be close to the entropy of the system at this point, unlike a system with N i.i.d. components.

Remember that information content ($\ln 1/P(\mathbf{x})$) and energy are very closely related. If typicality holds, then the system's energy has negligible fluctuations, and *vice versa*.

C

Some Mathematics

C.1 Finite field theory

Most linear codes are expressed in the language of Galois theory

Why are Galois fields an appropriate language for linear codes? First, a definition and some examples.

A field F is a set $F = \{0, F'\}$ such that

1. F forms an Abelian group under an addition operation '+', with 0 being the identity; [Abelian means all elements commute, i.e., satisfy $a + b = b + a$.]
2. F' forms an Abelian group under a multiplication operation '$\cdot$'; multiplication of any element by 0 yields 0;
3. these operations satisfy the distributive rule $(a + b) \cdot c = a \cdot c + b \cdot c$.

For example, the real numbers form a field, with '+' and '$\cdot$' denoting ordinary addition and multiplication.

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Table C.1. Addition and multiplication tables for $GF(2)$.

A Galois field $GF(q)$ is a field with a finite number of elements q.

A unique Galois field exists for any $q = p^m$, where p is a prime number and m is a positive integer; there are no other finite fields.

$GF(2)$. The addition and multiplication tables for $GF(2)$ are shown in table C.1. These are the rules of addition and multiplication modulo 2.

$GF(p)$. For any prime number p, the addition and multiplication rules are those for ordinary addition and multiplication, modulo p.

$GF(4)$. The rules for $GF(p^m)$, with $m > 1$, are *not* those of ordinary addition and multiplication. For example the tables for $GF(4)$ (table C.2) are *not* the rules of addition and multiplication modulo 4. Notice that $1 + 1 = 0$, for example. So how can $GF(4)$ be described? It turns out that the elements can be related to *polynomials*. Consider polynomial functions of x of degree 1 and with coefficients that are elements of $GF(2)$. The polynomials shown in table C.3 obey the addition and multiplication rules of $GF(4)$ *if* addition and multiplication are modulo the polynomial $x^2 + x + 1$, and the coefficients of the polynomials are from $GF(2)$. For example, $B \cdot B = x^2 + (1 + 1)x + 1 = x = A$. Each element may also be represented as a bit pattern as shown in table C.3, with addition being bitwise modulo 2, and multiplication defined with an appropriate carry operation.

$+$	0	1	A	B
0	0	1	A	B
1	1	0	B	A
A	A	B	0	1
B	B	A	1	0

$\cdot$	0	1	A	B
0	0	0	0	0
1	0	1	A	B
A	0	A	B	1
B	0	B	1	A

Table C.2. Addition and multiplication tables for $GF(4)$.

Element	Polynomial	Bit pattern
0	0	00
1	1	01
A	x	10
B	$x + 1$	11

Table C.3. Representations of the elements of $GF(4)$.

$GF(8)$. We can denote the elements of $GF(8)$ by $\{0, 1, A, B, C, D, E, F\}$. Each element can be mapped onto a polynomial over $GF(2)$. The multiplication and addition operations are given by multiplication and addition of the polynomials, modulo $x^3 + x + 1$. The multiplication table is given below.

element	polynomial	binary representation
0	0	000
1	1	001
A	x	010
B	$x+1$	011
C	x^2	100
D	x^2+1	101
E	x^2+x	110
F	x^2+x+1	111

$\cdot$	0	1	A	B	C	D	E	F
0	0	0	0	0	0	0	0	0
1	0	1	A	B	C	D	E	F
A	0	A	C	E	B	1	F	D
B	0	B	E	D	F	C	1	A
C	0	C	B	F	E	A	D	1
D	0	D	1	C	A	F	B	E
E	0	E	F	1	D	B	A	C
F	0	F	D	A	1	E	C	B

Why are Galois fields relevant to linear codes? Imagine generalizing a binary generator matrix $\mathbf{G}$ and binary vector $\mathbf{s}$ to a matrix and vector with elements from a larger set, and generalizing the addition and multiplication operations that define the product $\mathbf{Gs}$. In order to produce an appropriate input for a symmetric channel, it would be convenient if, for random $\mathbf{s}$, the product $\mathbf{Gs}$ produced all elements in the enlarged set with equal probability. This uniform distribution is easiest to guarantee if these elements form a group under both addition and multiplication, because then these operations do not break the symmetry among the elements. When two random elements of a multiplicative group are multiplied together, all elements are produced with equal probability. This is not true of other sets such as the integers, for which the multiplication operation is more likely to give rise to some elements (the composite numbers) than others. Galois fields, by their definition, avoid such symmetry-breaking effects.

▶ C.2 Eigenvectors and eigenvalues

A *right-eigenvector* of a square matrix $\mathbf{A}$ is a non-zero vector $\mathbf{e}_{\mathsf{R}}$ that satisfies

$$\mathbf{A}\mathbf{e}_{\mathsf{R}} = \lambda \mathbf{e}_{\mathsf{R}}, \tag{C.1}$$

where λ is the eigenvalue associated with that eigenvector. The eigenvalue may be a real number or complex number and it may be zero. Eigenvectors may be real or complex.

A *left-eigenvector* of a matrix $\mathbf{A}$ is a vector $\mathbf{e}_{\mathsf{L}}$ that satisfies

$$\mathbf{e}_{\mathsf{L}}^{\mathsf{T}}\mathbf{A} = \lambda \mathbf{e}_{\mathsf{L}}^{\mathsf{T}}. \tag{C.2}$$

The following statements for right-eigenvectors also apply to left-eigenvectors.

- If a matrix has two or more linearly independent right-eigenvectors with the same eigenvalue then that eigenvalue is called a degenerate eigenvalue of the matrix, or a repeated eigenvalue. Any linear combination of those eigenvectors is another right-eigenvector with the same eigenvalue.
- The principal right-eigenvector of a matrix is, by definition, the right-eigenvector with the largest associated eigenvalue.
- If a real matrix has a right-eigenvector with complex eigenvalue $\lambda = x + yi$ then it also has a right-eigenvector with the conjugate eigenvalue $\lambda^* = x - yi$.

Symmetric matrices

If $\mathbf{A}$ is a real symmetric $N \times N$ matrix then

1. all the eigenvalues and eigenvectors of $\mathbf{A}$ are real;
2. every left-eigenvector of $\mathbf{A}$ is also a right-eigenvector of $\mathbf{A}$ with the same eigenvalue, and *vice versa*;
3. a set of N eigenvectors and eigenvalues $\{\mathbf{e}^{(a)}, \lambda_a\}_{a=1}^{N}$ can be found that are orthonormal, that is,

$$\mathbf{e}^{(a)} \cdot \mathbf{e}^{(b)} = \delta_{ab}; \tag{C.3}$$

the matrix can be expressed as a weighted sum of outer products of the eigenvectors:

$$\mathbf{A} = \sum_{a=1}^{N} \lambda_a [\mathbf{e}^{(a)}][\mathbf{e}^{(a)}]^{\mathsf{T}}. \tag{C.4}$$

(Whereas I often use i and n as indices for sets of size I and N, I will use the indices a and b to run over eigenvectors, even if there are N of them. This is to avoid confusion with the components of the eigenvectors, which are indexed by n, e.g. $e_n^{(a)}$.)

General square matrices

An $N \times N$ matrix can have up to N distinct eigenvalues. Generically, there are N eigenvalues, all distinct, and each has one left-eigenvector and one right-eigenvector. In cases where two or more eigenvalues coincide, for each distinct eigenvalue that is non-zero there is at least one left-eigenvector and one right-eigenvector.

Left- and right-eigenvectors that have different eigenvalue are orthogonal, that is,

$$\text{if } \lambda_a \neq \lambda_b \text{ then } \mathbf{e}_{\mathsf{L}}^{(a)} \cdot \mathbf{e}_{\mathsf{R}}^{(b)} = 0. \tag{C.5}$$

Non-negative matrices

Definition. If all the elements of a non-zero matrix $\mathbf{C}$ satisfy $C_{mn} \geq 0$ then $\mathbf{C}$ is a non-negative matrix. Similarly, if all the elements of a non-zero vector $\mathbf{c}$ satisfy $c_n \geq 0$ then $\mathbf{c}$ is a non-negative vector.

Properties. A non-negative matrix has a principal eigenvector that is non-negative. It may also have other eigenvectors with the same eigenvalue that are not non-negative. But if the principal eigenvalue of a non-negative matrix is not degenerate, then the matrix has only one principal eigenvector $\mathbf{e}^{(1)}$, and it is non-negative.

Generically, all the other eigenvalues are smaller in absolute magnitude. [There can be several eigenvalues of identical magnitude in special cases.]

Transition probability matrices

An important example of a non-negative matrix is a transition probability matrix $\mathbf{Q}$.

Definition. A transition probability matrix $\mathbf{Q}$ has columns that are probability vectors, that is, it satisfies $\mathbf{Q} \geq 0$ and

$$\sum_i Q_{ij} = 1 \text{ for all } j. \tag{C.6}$$

Matrix	Eigenvalues and eigenvectors $\mathbf{e}_{\mathsf{L}}, \mathbf{e}_{\mathsf{R}}$									
$\begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	2.41		1		−0.41					
	$\begin{bmatrix} .58 \\ .82 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .82 \\ .58 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -.58 \\ .82 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -.82 \\ .58 \\ 0 \end{bmatrix}$				
$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$	1.62		−0.62							
	$\begin{bmatrix} .53 \\ .85 \end{bmatrix}$	$\begin{bmatrix} .53 \\ .85 \end{bmatrix}$	$\begin{bmatrix} .85 \\ -.53 \end{bmatrix}$	$\begin{bmatrix} .85 \\ -.53 \end{bmatrix}$						
$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$	1.62		0.5+0.9i		0.5−0.9i		−0.62			
	$\begin{bmatrix} .60 \\ .37 \\ .37 \\ .60 \end{bmatrix}$	$\begin{bmatrix} .60 \\ .37 \\ .37 \\ .60 \end{bmatrix}$	$\begin{bmatrix} .1-.5i \\ -.3-.4i \\ .3+.4i \\ -.1+.5i \end{bmatrix}$	$\begin{bmatrix} .1-.5i \\ .3+.4i \\ -.3-.4i \\ -.1+.5i \end{bmatrix}$	$\begin{bmatrix} .1+.5i \\ -.3+.4i \\ .3-.4i \\ -.1-.5i \end{bmatrix}$	$\begin{bmatrix} .1+.5i \\ .3-.4i \\ -.3+.4i \\ -.1-.5i \end{bmatrix}$	$\begin{bmatrix} .37 \\ -.60 \\ -.60 \\ .37 \end{bmatrix}$	$\begin{bmatrix} .37 \\ -.60 \\ -.60 \\ .37 \end{bmatrix}$		

Table C.4. Some matrices and their eigenvectors.

Matrix	Eigenvalues and eigenvectors $\mathbf{e}_{\mathsf{L}}, \mathbf{e}_{\mathsf{R}}$									
$\begin{bmatrix} 0 & .38 \\ 1 & .62 \end{bmatrix}$	1		−0.38							
	$\begin{bmatrix} .71 \\ .71 \end{bmatrix}$	$\begin{bmatrix} .36 \\ .93 \end{bmatrix}$	$\begin{bmatrix} -.93 \\ .36 \end{bmatrix}$	$\begin{bmatrix} -.71 \\ .71 \end{bmatrix}$						
$\begin{bmatrix} 0 & .35 & 0 \\ 0 & 0 & .46 \\ 1 & .65 & .54 \end{bmatrix}$	1		−0.2−0.3i		−0.2+0.3i					
	$\begin{bmatrix} .58 \\ .58 \\ .58 \end{bmatrix}$	$\begin{bmatrix} .14 \\ .41 \\ .90 \end{bmatrix}$	$\begin{bmatrix} -.8+.1i \\ -.2-.5i \\ .2+.2i \end{bmatrix}$	$\begin{bmatrix} .2-.5i \\ -.6+.2i \\ .4+.3i \end{bmatrix}$	$\begin{bmatrix} -.8-.1i \\ -.2+.5i \\ .2-.2i \end{bmatrix}$	$\begin{bmatrix} .2+.5i \\ -.6-.2i \\ .4-.3i \end{bmatrix}$				

Table C.5. Transition probability matrices for generating random paths through trellises.

This property can be rewritten in terms of the all-ones vector $\mathbf{n} = (1, 1, \ldots, 1)^{\mathsf{T}}$:

$$\mathbf{n}^{\mathsf{T}}\mathbf{Q} = \mathbf{n}^{\mathsf{T}}. \tag{C.7}$$

So $\mathbf{n}$ is the principal left-eigenvector of $\mathbf{Q}$ with eigenvalue $\lambda_1 = 1$.

$$\mathbf{e}_{\mathsf{L}}^{(1)} = \mathbf{n}. \tag{C.8}$$

Because it is a non-negative matrix, $\mathbf{Q}$ has a principal right-eigenvector that is non-negative, $\mathbf{e}_{\mathsf{R}}^{(1)}$. Generically, for Markov processes that are ergodic, this eigenvector is the only right-eigenvector with eigenvalue of magnitude 1 (see table C.6 for illustrative exceptions). This vector, if we normalize it such that $\mathbf{e}_{\mathsf{R}}^{(1)} \cdot \mathbf{n} = 1$, is called the invariant distribution of the transition probability matrix. It is the probability density that is left unchanged under $\mathbf{Q}$. Unlike the principal left-eigenvector, which we explicitly identified above, we can't usually identify the principal right-eigenvector without computation.

The matrix may have up to $N - 1$ other right-eigenvectors all of which are orthogonal to the left-eigenvector $\mathbf{n}$, that is, they are zero-sum vectors.

▶ C.3 Perturbation theory

Perturbation theory is not used in this book, but it is useful in this book's fields. In this section we derive first-order perturbation theory for the eigenvectors and eigenvalues of square, *not necessarily symmetric*, matrices. Most presentations of perturbation theory focus on symmetric matrices, but non-symmetric matrices (such as transition matrices) also deserve to be perturbed!

	Matrix	Eigenvalues and eigenvectors $\mathbf{e}_{\mathsf{L}}, \mathbf{e}_{\mathsf{R}}$							
		1		1		0.70		0.70	
(a)	$\begin{bmatrix} .90 & .20 & 0 & 0 \\ .10 & .80 & 0 & 0 \\ 0 & 0 & .90 & .20 \\ 0 & 0 & .10 & .80 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ .71 \\ .71 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ .89 \\ .45 \end{bmatrix}$	$\begin{bmatrix} .71 \\ .71 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .89 \\ .45 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .45 \\ -.89 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} .71 \\ -.71 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ -.45 \\ .89 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ -.71 \\ .71 \end{bmatrix}$
		1		0.98		0.70		0.69	
(a′)	$\begin{bmatrix} .90 & .20 & 0 & 0 \\ .10 & .79 & .02 & 0 \\ 0 & .01 & .88 & .20 \\ 0 & 0 & .10 & .80 \end{bmatrix}$	$\begin{bmatrix} .50 \\ .50 \\ .50 \\ .50 \end{bmatrix}$	$\begin{bmatrix} .87 \\ .43 \\ .22 \\ .11 \end{bmatrix}$	$\begin{bmatrix} -.18 \\ -.15 \\ .66 \\ .72 \end{bmatrix}$	$\begin{bmatrix} -.66 \\ -.28 \\ .61 \\ .33 \end{bmatrix}$	$\begin{bmatrix} .20 \\ -.40 \\ -.40 \\ .80 \end{bmatrix}$	$\begin{bmatrix} .63 \\ -.63 \\ -.32 \\ .32 \end{bmatrix}$	$\begin{bmatrix} -.19 \\ .41 \\ -.44 \\ .77 \end{bmatrix}$	$\begin{bmatrix} -.61 \\ .65 \\ -.35 \\ .30 \end{bmatrix}$
		1		0.70		−0.70		−1	
(b)	$\begin{bmatrix} 0 & 0 & .90 & .20 \\ 0 & 0 & .10 & .80 \\ .90 & .20 & 0 & 0 \\ .10 & .80 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} .50 \\ .50 \\ .50 \\ .50 \end{bmatrix}$	$\begin{bmatrix} .63 \\ .32 \\ .63 \\ .32 \end{bmatrix}$	$\begin{bmatrix} -.32 \\ .63 \\ -.32 \\ .63 \end{bmatrix}$	$\begin{bmatrix} .50 \\ -.50 \\ .50 \\ -.50 \end{bmatrix}$	$\begin{bmatrix} .32 \\ -.63 \\ -.32 \\ .63 \end{bmatrix}$	$\begin{bmatrix} -.50 \\ .50 \\ .50 \\ -.50 \end{bmatrix}$	$\begin{bmatrix} .50 \\ .50 \\ -.50 \\ -.50 \end{bmatrix}$	$\begin{bmatrix} .63 \\ .32 \\ -.63 \\ -.32 \end{bmatrix}$

Table C.6. Illustrative transition probability matrices and their eigenvectors showing the two ways of being non-ergodic. (a) More than one principal eigenvector with eigenvalue 1 because the state space falls into two unconnected pieces. (a′) A small perturbation breaks the degeneracy of the principal eigenvectors. (b) Under this chain, the density may oscillate between two parts of the state space. In addition to the invariant distribution, there is another right-eigenvector with eigenvalue −1. In general such circulating densities correspond to complex eigenvalues with magnitude 1.

We assume that we have an $N \times N$ matrix $\mathbf{H}$ that is a function $\mathbf{H}(\epsilon)$ of a real parameter ϵ, with $\epsilon = 0$ being our starting point. We assume that a Taylor expansion of $\mathbf{H}(\epsilon)$ is appropriate:

$$\mathbf{H}(\epsilon) = \mathbf{H}(0) + \epsilon \mathbf{V} + \cdots \tag{C.9}$$

where

$$\mathbf{V} \equiv \frac{\partial \mathbf{H}}{\partial \epsilon}. \tag{C.10}$$

We assume that for all ϵ of interest, $\mathbf{H}(\epsilon)$ has a complete set of N right-eigenvectors and left-eigenvectors, and that these eigenvectors and their eigenvalues are continuous functions of ϵ. This last assumption is not necessarily a good one: if $\mathbf{H}(0)$ has degenerate eigenvalues then it is possible for the eigenvectors to be discontinuous in ϵ; in such cases, degenerate perturbation theory is needed. That's a fun topic, but let's stick with the non-degenerate case here.

We write the eigenvectors and eigenvalues as follows:

$$\mathbf{H}(\epsilon)\mathbf{e}_{\mathsf{R}}^{(a)}(\epsilon) = \lambda^{(a)}(\epsilon)\mathbf{e}_{\mathsf{R}}^{(a)}(\epsilon), \tag{C.11}$$

and we Taylor-expand

$$\lambda^{(a)}(\epsilon) = \lambda^{(a)}(0) + \epsilon\mu^{(a)} + \cdots \tag{C.12}$$

with

$$\mu^{(a)} \equiv \frac{\partial \lambda^{(a)}(\epsilon)}{\partial \epsilon} \tag{C.13}$$

and

$$\mathbf{e}_{\mathsf{R}}^{(a)}(\epsilon) = \mathbf{e}_{\mathsf{R}}^{(a)}(0) + \epsilon \mathbf{f}_{\mathsf{R}}^{(a)} + \cdots \tag{C.14}$$

with

$$\mathbf{f}_{\mathsf{R}}^{(a)} \equiv \frac{\partial \mathbf{e}_{\mathsf{R}}^{(a)}}{\partial \epsilon}, \tag{C.15}$$

and similar definitions for $\mathbf{e}_{\mathsf{L}}^{(a)}$ and $\mathbf{f}_{\mathsf{L}}^{(a)}$. We define these left-vectors to be row vectors, so that the 'transpose' operation is not needed and can be banished.

We are free to constrain the magnitudes of the eigenvectors in whatever way we please. Each left-eigenvector and each right-eigenvector has an arbitrary magnitude. The natural constraints to use are as follows. First, we constrain the inner products with:

$$\mathbf{e}_{\mathsf{L}}^{(a)}(\epsilon)\mathbf{e}_{\mathsf{R}}^{(a)}(\epsilon) = 1, \quad \text{for all } a. \tag{C.16}$$

Expanding the eigenvectors in ϵ, equation (C.19) implies

$$(\mathbf{e}_{\mathsf{L}}^{(a)}(0) + \epsilon \mathbf{f}_{\mathsf{L}}^{(a)} + \cdots)(\mathbf{e}_{\mathsf{R}}^{(a)}(0) + \epsilon \mathbf{f}_{\mathsf{R}}^{(a)} + \cdots) = 1, \tag{C.17}$$

from which we can extract the terms in ϵ, which say:

$$\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mathbf{f}_{\mathsf{L}}^{(a)}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = 0 \tag{C.18}$$

We are now free to choose the two constraints:

$$\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} = 0, \quad \mathbf{f}_{\mathsf{L}}^{(a)}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = 0, \tag{C.19}$$

which in the special case of a symmetric matrix correspond to constraining the eigenvectors to be of constant length, as defined by the Euclidean norm.

OK, now that we have defined our cast of characters, what do the defining equations (C.11) and (C.9) tell us about our Taylor expansions (C.13) and (C.15)? We expand equation (C.11) in ϵ.

$$(\mathbf{H}(0)+\epsilon\mathbf{V}+\cdots)(\mathbf{e}_{\mathsf{R}}^{(a)}(0)+\epsilon\mathbf{f}_{\mathsf{R}}^{(a)}+\cdots) = (\lambda^{(a)}(0)+\epsilon\mu^{(a)}+\cdots)(\mathbf{e}_{\mathsf{R}}^{(a)}(0)+\epsilon\mathbf{f}_{\mathsf{R}}^{(a)}+\cdots). \tag{C.20}$$

Identifying the terms of order ϵ, we have:

$$\mathbf{H}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = \lambda^{(a)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mu^{(a)}\mathbf{e}_{\mathsf{R}}^{(a)}(0). \tag{C.21}$$

We can extract interesting results from this equation by hitting it with $\mathbf{e}_{\mathsf{L}}^{(b)}(0)$:

$$\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{H}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = \mathbf{e}_{\mathsf{L}}^{(b)}(0)\lambda^{(a)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mu^{(a)}\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{e}_{\mathsf{R}}^{(a)}(0).$$

$$\Rightarrow \ \lambda^{(b)}\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = \lambda^{(a)}(0)\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} + \mu^{(a)}\delta_{ab}. \tag{C.22}$$

Setting $b = a$ we obtain

$$\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = \mu^{(a)}. \tag{C.23}$$

Alternatively, choosing $b \neq a$, we obtain:

$$\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0) = \left[\lambda^{(a)}(0) - \lambda^{(b)}(0)\right] \mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} \tag{C.24}$$

$$\Rightarrow \mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{f}_{\mathsf{R}}^{(a)} = \frac{1}{\lambda^{(a)}(0) - \lambda^{(b)}(0)}\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0). \tag{C.25}$$

Now, assuming that the right-eigenvectors $\{\mathbf{e}_{\mathsf{R}}^{(b)}(0)\}_{b=1}^{N}$ form a complete basis, we must be able to write

$$\mathbf{f}_{\mathsf{R}}^{(a)} = \sum_b w_b \mathbf{e}_{\mathsf{R}}^{(b)}(0), \tag{C.26}$$

where

$$w_b = \mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{f}_{\mathsf{R}}^{(a)}, \tag{C.27}$$

so, comparing (C.25) and (C.27), we have:

$$\mathbf{f}_{\mathsf{R}}^{(a)} = \sum_{b \neq a} \frac{\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0)}{\lambda^{(a)}(0) - \lambda^{(b)}(0)} \mathbf{e}_{\mathsf{R}}^{(b)}(0). \tag{C.28}$$

Equations (C.23) and (C.28) are the solution to the first-order perturbation theory problem, giving respectively the first derivative of the eigenvalue and the eigenvectors.

Second-order perturbation theory

If we expand the eigenvector equation (C.11) to second order in ϵ, and assume that the equation

$$\mathbf{H}(\epsilon) = \mathbf{H}(0) + \epsilon\mathbf{V} \tag{C.29}$$

is exact, that is, $\mathbf{H}$ is a purely linear function of ϵ, then we have:

$$\begin{aligned} &(\mathbf{H}(0) + \epsilon\mathbf{V})(\mathbf{e}_{\mathsf{R}}^{(a)}(0) + \epsilon\mathbf{f}_{\mathsf{R}}^{(a)} + \frac{1}{2}\epsilon^2\mathbf{g}_{\mathsf{R}}^{(a)} + \cdots) \\ = \;& (\lambda^{(a)}(0) + \epsilon\mu^{(a)} + \tfrac{1}{2}\epsilon^2\nu^{(a)} + \cdots)(\mathbf{e}_{\mathsf{R}}^{(a)}(0) + \epsilon\mathbf{f}_{\mathsf{R}}^{(a)} + \tfrac{1}{2}\epsilon^2\mathbf{g}_{\mathsf{R}}^{(a)} + \cdots) \end{aligned} \tag{C.30}$$

where $\mathbf{g}_{\mathsf{R}}^{(a)}$ and $\nu^{(a)}$ are the second derivatives of the eigenvector and eigenvalue. Equating the second-order terms in ϵ in equation (C.30),

$$\mathbf{V}\mathbf{f}_{\mathsf{R}}^{(a)} + \frac{1}{2}\mathbf{H}(0)\mathbf{g}_{\mathsf{R}}^{(a)} = \frac{1}{2}\lambda^{(a)}(0)\mathbf{g}_{\mathsf{R}}^{(a)} + \frac{1}{2}\nu^{(a)}\mathbf{e}_{\mathsf{R}}^{(a)}(0) + \mu^{(a)}\mathbf{f}_{\mathsf{R}}^{(a)}. \tag{C.31}$$

Hitting this equation on the left with $\mathbf{e}_{\mathsf{L}}^{(a)}(0)$, we obtain:

$$\begin{aligned} &\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{V}\mathbf{f}_{\mathsf{R}}^{(a)} + \frac{1}{2}\lambda^{(a)}\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{g}_{\mathsf{R}}^{(a)} \\ = \;& \tfrac{1}{2}\lambda^{(a)}(0)\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{g}_{\mathsf{R}}^{(a)} + \tfrac{1}{2}\nu^{(a)}\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{e}_{\mathsf{R}}^{(a)}(0) + \mu^{(a)}\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{f}_{\mathsf{R}}^{(a)}. \end{aligned} \tag{C.32}$$

The term $\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{f}_{\mathsf{R}}^{(a)}$ is equal to zero because of our constraints (C.19), so

$$\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{V}\mathbf{f}_{\mathsf{R}}^{(a)} = \frac{1}{2}\nu^{(a)}, \tag{C.33}$$

so the second derivative of the eigenvalue with respect to ϵ is given by

$$\frac{1}{2}\nu^{(a)} = \mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{V}\sum_{b \neq a} \frac{\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0)}{\lambda^{(a)}(0) - \lambda^{(b)}(0)} \mathbf{e}_{\mathsf{R}}^{(b)}(0) \tag{C.34}$$

$$= \sum_{b \neq a} \frac{[\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0)][\mathbf{e}_{\mathsf{L}}^{(a)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(b)}(0)]}{\lambda^{(a)}(0) - \lambda^{(b)}(0)}. \tag{C.35}$$

This is as far as we will take the perturbation expansion.

Summary

If we introduce the abbreviation V_{ba} for $\mathbf{e}_{\mathsf{L}}^{(b)}(0)\mathbf{V}\mathbf{e}_{\mathsf{R}}^{(a)}(0)$, we can write the eigenvectors of $\mathbf{H}(\epsilon) = \mathbf{H}(0) + \epsilon\mathbf{V}$ to first order as

$$\mathbf{e}_{\mathsf{R}}^{(a)}(\epsilon) = \mathbf{e}_{\mathsf{R}}^{(a)}(0) + \epsilon\sum_{b \neq a} \frac{V_{ba}}{\lambda^{(a)}(0) - \lambda^{(b)}(0)} \mathbf{e}_{\mathsf{R}}^{(b)}(0) + \cdots \tag{C.36}$$

and the eigenvalues to second order as

$$\lambda^{(a)}(\epsilon) = \lambda^{(a)}(0) + \epsilon V_{aa} + \epsilon^2 \sum_{b \neq a} \frac{V_{ba}V_{ab}}{\lambda^{(a)}(0) - \lambda^{(b)}(0)} + \cdots. \tag{C.37}$$

C.4 Some numbers

	2^{8192}	10^{2466}	Number of distinct 1-kilobyte files
	2^{1024}	10^{308}	Number of states of a 2D Ising model with 32×32 spins
2^{1000}		10^{301}	Number of binary strings of length 1000
2^{500}		3×10^{150}	
	2^{469}	10^{141}	Number of binary strings of length 1000 having 100 1s and 900 0s
	2^{266}	10^{80}	Number of electrons in universe
2^{200}		1.6×10^{60}	
	2^{190}	10^{57}	Number of electrons in solar system
	2^{171}	3×10^{51}	Number of electrons in the earth
2^{100}		10^{30}	
	2^{98}	3×10^{29}	Age of universe/picoseconds
	2^{58}	3×10^{17}	Age of universe/seconds
2^{50}		10^{15}	
2^{40}		10^{12}	
		10^{11}	Number of neurons in human brain
		10^{11}	Number of bits stored on a DVD
		3×10^{10}	Number of bits in the wheat genome
		6×10^{9}	Number of bits in the human genome
	2^{32}	6×10^{9}	Population of earth
2^{30}		10^{9}	
		2.5×10^{8}	Number of fibres in the corpus callosum
		2×10^{8}	Number of bits in *C. Elegans* (a worm) genome
		2×10^{8}	Number of bits in *Arabidopsis thaliana* (a flowering plant related to broccoli) genome
	2^{25}	3×10^{7}	One year/seconds
		2×10^{7}	Number of bits in the compressed PostScript file that is this book
		2×10^{7}	Number of bits in `unix` kernel
		10^{7}	Number of bits in the *E. Coli* genome, or in a floppy disk
		4×10^{6}	Number of years since human/chimpanzee divergence
2^{20}		10^{6}	1 048 576
		2×10^{5}	Number of generations since human/chimpanzee divergence
		3×10^{4}	Number of genes in human genome
		3×10^{4}	Number of genes in *Arabidopsis thaliana* genome
		1.5×10^{3}	Number of base pairs in a gene
2^{10}	e^{7}	10^{3}	$2^{10} = 1024$; $e^{7} = 1096$
2^{0}		10^{0}	1
	2^{-2}	2.5×10^{-1}	Lifetime probability of dying from smoking one pack of cigarettes per day.
		10^{-2}	Lifetime probability of dying in a motor vehicle accident
2^{-10}		10^{-3}	
		10^{-5}	Lifetime probability of developing cancer because of drinking 2 litres per day of water containing 12 p.p.b. benzene
2^{-20}		10^{-6}	
		3×10^{-8}	Probability of error in transmission of coding DNA, per nucleotide, per generation
2^{-30}		10^{-9}	
2^{-60}		10^{-18}	Probability of undetected error in a hard disk drive, after error correction