Patrick E. Farrell
University of Oxford

2025–2026
March 9, 2026

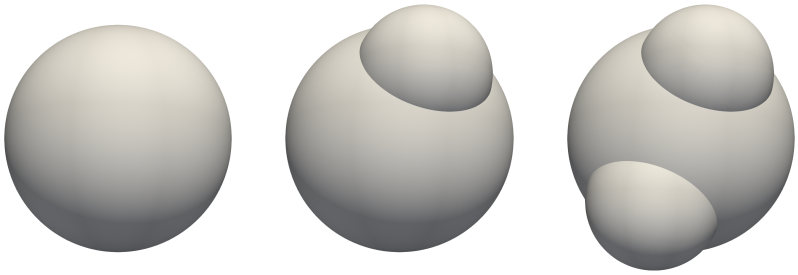
Computational Mathematics Projects

Contents

A 2026A: topological classification

Topology is the study of properties of objects that remain unchanged under continuous deformation. Whereas geometry considers quantities of a shape like lengths, angles, areas, and volumes—all induced by an object called a *metric*—topology considers those properties of the shape[1] that are metric-independent.

[1] More precisely, by 'shape' we mean a manifold, as defined next year in *A13: Geometry*. We will not need the technical definition.



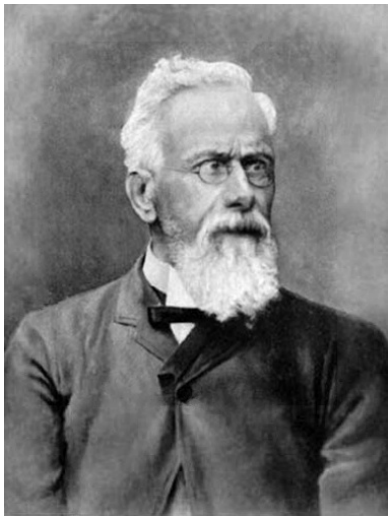
(a) (b) (c)

Figure A.1: Three homeomorphic manifolds: there exists a continuous bijection with continuous inverse that maps any one to any other.

Consider an example. From the geometric point of view, all of the manifolds depicted in Figure A.1 are distinct. But from the topological point of view, they are the same, or more precisely they are *homeomorphic*. This means that there exists a continuous bijection with continuous inverse (a homeomorphism) that maps each manifold to another.

However, the manifolds in Figure A.2 *are* all mutually distinct from the topological point of view. There does not exist a homeomorphism that maps any of the manifolds to any other: there is no invertibly continuous way to deform a manifold having multiple components (A.2b), tunnels (A.2c), or voids (A.2d), to one without (A.2a).

How can we sharpen this notion, that the manifolds in Figure A.2 are all mutually distinct? Given some manifold, can we classify it, so that we know that two manifolds with different classifications cannot be homeomorphic[2]? The key idea was introduced by Enrico Betti in 1870[3], defining a sequence of natural numbers now referred to as the Betti numbers. The k^{th} Betti number b_k is often described as counting the 'number of k-dimensional holes'[4]. In three dimensions, the Betti

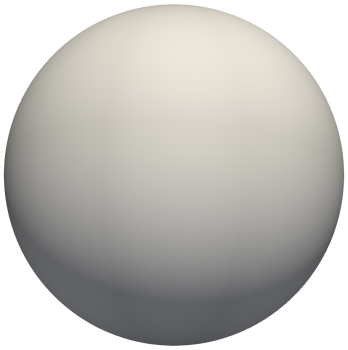


Enrico Betti, 1823–1892

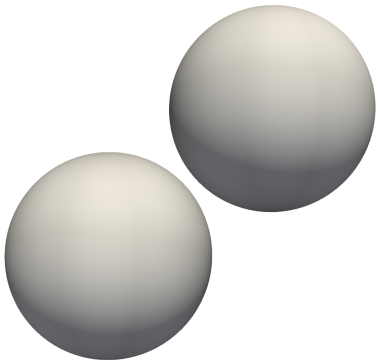
[2] The converse will not be true: equal Betti numbers does not imply that two manifolds are homeomorphic.

[3] E. Betti. Sopra gli spazi di un numero qualunque di dimensioni. *Annali di Matematica Pura ed Applicata*, 4(1):140–158, 1870

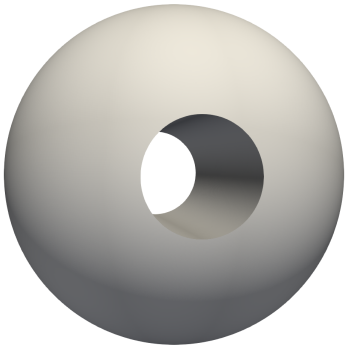
[4] We will give a definition of what exactly this means below.



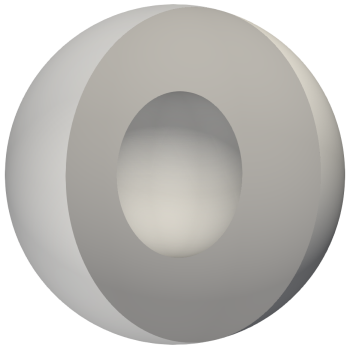
(a) a solid sphere,
$(b_0, b_1, b_2) = (1, 0, 0)$



(b) two solid spheres,
$(b_0, b_1, b_2) = (2, 0, 0)$



(c) a solid sphere with a tunnel,
$(b_0, b_1, b_2) = (1, 1, 0)$



(d) a solid sphere with a void,
$(b_0, b_1, b_2) = (1, 0, 1)$

Figure A.2: Four manifolds, none of which are homeomorphic to any other. Their Betti numbers (b_0, b_1, b_2) are given in the captions. The figure in (d) has been rendered with its front half transparent, so that the void inside can be seen.

numbers count

b_0 the number of connected components (separate pieces),

b_1 the number of tunnels (where light could pass from one side to the other),

b_2 the number of voids (where you could store water as the body rotates).

How can we *compute* these Betti numbers for a given manifold? Do we have to look at the manifold and count, or is there a better way? In 1895 Poincaré showed that there was[5]: by *triangulating* our manifold, we can elegantly reduce the computation of its Betti numbers to linear algebra. In this project we will implement Poincaré's construction.

[5] H. Poincaré. *Papers on Topology: Analysis Situs and Its Five Supplements*, volume 37 of *History of Mathematics*. American Mathematical Society and London Mathematical Society, Providence, RI and London, 2010. Translated and edited by John Stillwell

A.1 *Triangulations of manifolds*

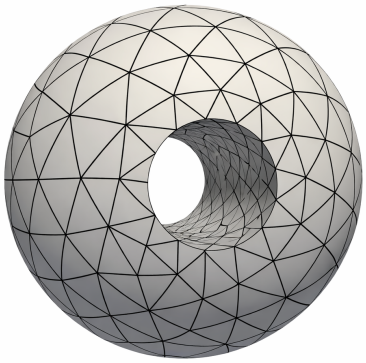


Figure A.3: An example of a simplicial complex for the manifold shown in Figure A.2c.

A *simplicial complex* is a way of building complicated manifolds out of simple building blocks, oriented simplices. An n-simplex is the convex hull of $n+1$ affinely independent points[6]. Concretely, a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a solid triangle, and a 3-simplex is a solid tetrahedron. We can describe a manifold by gluing these simplices together along shared faces. An example of a simplicial complex for Figure A.2c is given in Figure A.3.

Given points $v_0, v_1, \ldots, v_n \in \mathbb{R}^m$, the simplex they generate is denoted $[v_0, v_1, \ldots, v_n]$. These simplices also carry an orientation, so the order matters: for example, the edges $[v_0, v_1]$ and $[v_1, v_0]$ have opposite orientations. We assume we have an ordering of the vertices; throughout this project we fix the convention that simplices written in sorted ascending vertex order are positively oriented.

A simplicial complex (as we use it here) is a collection of simplices in $\mathbb{R}^m$ such that

1. every subsimplex of a simplex in the collection is also in the collection;

[6] This means that e.g. the three points of a triangle do not all lie on a line, that the four points of a tetrahedron do not lie on a plane, etc.

Henri Poincaré, 1854–1912

2. the intersection of any two simplices is either empty or a common subsimplex of both.

So, for example, if a tetrahedron is in a simplicial complex, then so are all of its faces (triangles), edges (line segments), and vertices (points).

It is useful to introduce some notation to discuss subsimplices. We denote $[v_0, \ldots, \hat{v}_i, \ldots, v_n]$ the subsimplex that *excludes* v_i—the hat acts to omit it from the list. This notation is depicted in Figure A.4.

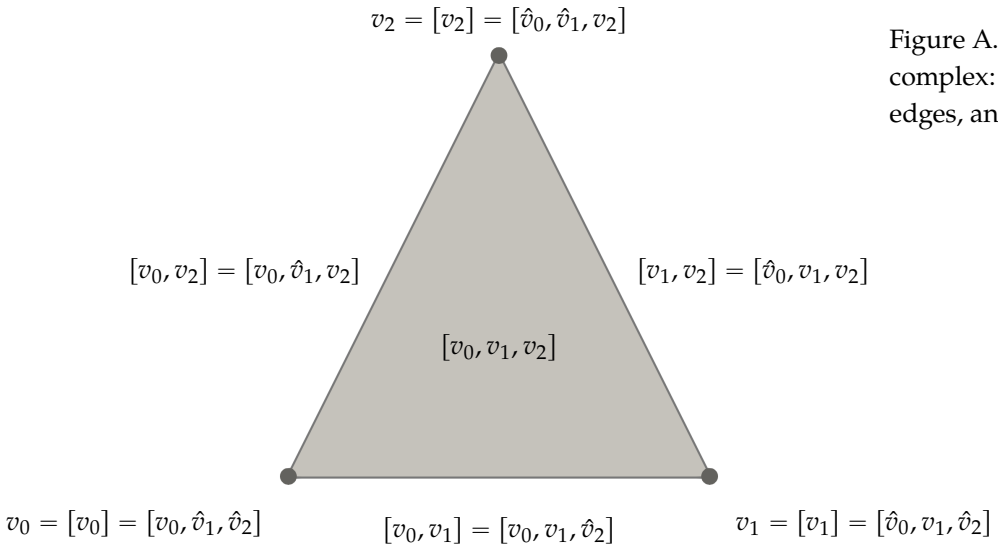


Figure A.4: A simple simplicial complex: one triangle, three edges, and three vertices.

We say that a simplicial complex K triangulates a manifold Ω if the two are homeomorphic. All manifolds in one, two, and three dimensions are triangulisable (such a simplicial complex exists), but there are counterexamples in four dimensions[7].

For a given simplicial complex K, denote its k-simplices by K_k. The next step is to introduce the vector spaces

$$C_k(K) = \left\{ \sum_{\sigma \in K_k} a_\sigma \sigma \mid a_\sigma \in \mathbb{R} \right\}. \tag{A.1.1}$$

which gather *formal linear combinations* of k-dimensional simplices. These linear combinations are taken over a field[8], which in our case we have chosen to be $\mathbb{R}$. Elements of $C_k(K)$ are called *k-chains*. For example, taking K to be the complex depicted in Figure A.4, a sample element of $C_1(K)$ (a 1-chain) might be

$$5[v_0, v_1] - [v_1, v_2] + 3[v_0, v_2]. \tag{A.1.2}$$

These linear combinations are purely formal: they do not resolve to anything else, except possibly zero. (There is no simpler expres-

[7] A corollary of Michael Freedman's work shows that the so-called E_8 manifold in four dimensions is not triangulisable. Freedman won a Fields medal in 1986, primarily for his work on the topology of 4-manifolds.

M. H. Freedman. The topology of four-dimensional manifolds. *Journal of Differential Geometry*, 17(3), 1982

[8] More generally, the coefficients can be taken over a ring, a more general algebraic structure than a field. In this situation C_k is not a vector space, but is rather an Abelian group, and the homology spaces H_k defined below are also merely groups.

sion for the sum above.) The dimension of $C_k(K)$ is the number of k-simplices in K. For notational convenience we fix $C_{-1}(K) = \varnothing$.

With $C_k(K)$ introduced, we can be more precise about orientations. Permuting the vertices of a simplex with an even number of swaps does not change the orientation, e.g.

$$[v_0, v_1, v_2] = [v_2, v_0, v_1] \tag{A.1.3}$$

while permuting with an odd number of swaps negates the orientation, e.g.

$$[v_0, v_1] = -[v_1, v_0], \text{ and } [v_0, v_1, v_2] = -[v_1, v_0, v_2]. \tag{A.1.4}$$

There are thus two possible orientations for each simplex[9].

[9] Imagine we are building a simplicial complex of a 2-manifold in 3-space, like a hollow sphere or a hollow torus. By consistently orienting each triangle, we can ensure that the normal vector always points outwards.

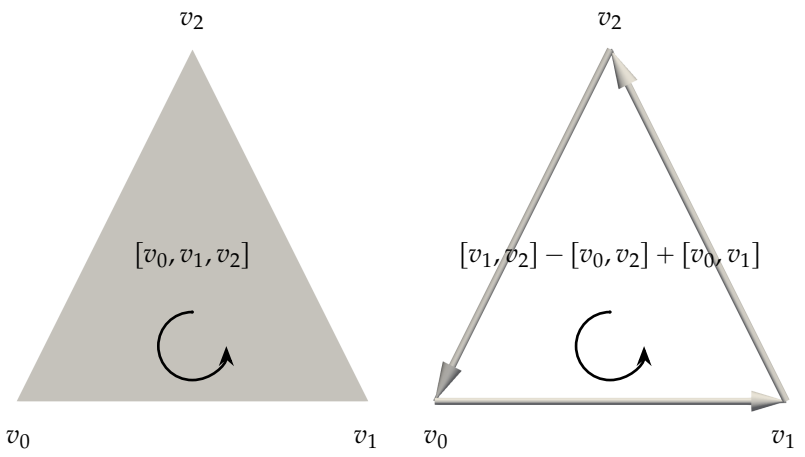


Figure A.5: The boundary of a triangle $\delta_2[v_0, v_1, v_2]$ is the path defined by its oriented edges.

The reason for introducing these formal linear combinations is the following. Define the *boundary operator*

$$\delta_k : C_k(K) \to C_{k-1}(K) \tag{A.1.5}$$

which calculates the boundary of a k-chain. We first specify it on a single simplex:

$$\delta_k[v_0, \ldots, v_k] = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k]. \tag{A.1.6}$$

To make this concrete, let us consider some examples. The boundary of an edge is

$$\delta_1[v_0, v_1] = [v_1] - [v_0], \tag{A.1.7}$$

the difference of the endpoints. The boundary of a triangle is

$$\delta_2[v_0, v_1, v_2] = [v_1, v_2] - [v_0, v_2] + [v_0, v_1] \tag{A.1.8}$$

as shown in Figure A.5.

We then extend the boundary operator δ_k to general k-chains by linearity:

$$\delta_k(a\sigma + b\tau) = a\delta_k\sigma + b\delta_k\tau. \tag{A.1.9}$$

For example, if we calculate the boundary of the boundary of a triangle,

$$\begin{aligned}
\delta_1\delta_2[v_0, v_1, v_2] &= \delta_1\left([v_1, v_2] - [v_0, v_2] + [v_0, v_1]\right) && \text{(A.1.10a)}\\
&= \delta_1[v_1, v_2] - \delta_1[v_0, v_2] + \delta_1[v_0, v_1] && \text{(A.1.10b)}\\
&= [v_2] - [v_1] - [v_2] + [v_0] + [v_1] - [v_0] && \text{(A.1.10c)}\\
&= 0. && \text{(A.1.10d)}
\end{aligned}$$

This is a reflection of a deeper principle: *the boundary of a boundary is always zero*. The boundary of a triangle is a loop of edges; the boundary of a loop is zero. More generally,

$$\delta_k \circ \delta_{k+1} = 0. \tag{A.1.11}$$

This is true not just of simplicial complexes: if a manifold has a boundary, then the boundary of its boundary is always empty.

A.2 Cycles and boundaries

In the language of linear algebra, the fact $\delta_k \circ \delta_{k+1} = 0$ can be expressed as

$$\mathrm{im}(\delta_{k+1}) \subseteq \ker(\delta_k). \tag{A.2.1}$$

This motivates us to define the *k-cycles*

$$Z_k = \ker(\delta_k) \tag{A.2.2}$$

and the *k-boundaries*

$$B_k = \mathrm{im}(\delta_{k+1}), \tag{A.2.3}$$

so that (A.2.1) becomes

$$B_k \subseteq Z_k. \tag{A.2.4}$$

The reason for these names is the following. The k-cycles are closed things (like loops of edges) that have no boundary. The k-boundaries are things that form the boundary of higher-dimensional entities.

So all boundaries are cycles ($B_k \subseteq Z_k$). But are all cycles boundaries ($Z_k \subseteq B_k$)? Let us now consider two topologically different simplicial complexes, shown in Figure A.6. For the complex shown in Figure A.6a you can check by eye that any 1-cycle (a closed loop of edges) you could construct is the boundary of a 2-chain (some set of triangles). However, the complex shown in Figure A.6b has a hole. On this domain *we can construct cycles that are not boundaries* of a set of triangles. Two of them are shown, in blue and green, and there are many others.

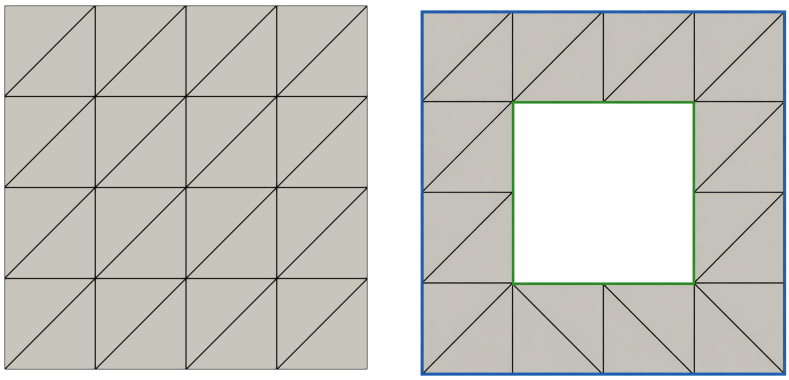


Figure A.6: Two simplicial complexes, one with a hole (right) and one without (left). The complex on the right has cycles (chains with no boundary) that are not themselves boundaries; two of these are shown in blue and green. The complex on the left has no cycles that are not themselves boundaries.

So we see that sometimes $B_k = Z_k$ (Figure A.6a), and sometimes $B_k \subsetneq Z_k$ (Figure A.6b). *The difference between these sets encodes the topology of the manifold.*

There is only one hole in the domain shown in Figure A.6b, and we want some algebraic way of counting this. In other words, we have to treat the blue and green cycles as equivalent. We can do this by agreeing that for the purpose of counting holes, *two cycles are equivalent if the difference between them is the boundary of something else*. For example, the difference between the blue and green cycles is the boundary of the sum of the triangles in the complex. Algebraically, we want to take the quotient space:

$$H_k := Z_k / B_k = \{z_k + B_k \mid z_k \in Z_k\}. \tag{A.2.5}$$

These are the *homology spaces*. Two cycles in H_k are deemed to be the same if their difference is the boundary of something else; effectively, they represent the 'same hole'. The k^{th} Betti number, the number of k-holes in the domain, is precisely the dimension of this quotient space:

$$b_k := \dim H_k. \tag{A.2.6}$$

For example, the complex in Figure A.6a has $b_1 = 0$, while the complex in Figure A.6b has $b_1 = 1$. If it had two holes, there would be two cycles that are not boundaries and whose difference is not the boundary of something else.

Let us do another example to make this concrete. Imagine a domain with three connected components (islands), shown in Figure A.7. With (A.2.5) we say that two vertices are equivalent if their difference is the boundary of a 1-chain (a path of edges). On a fixed island, any two vertices on the island can be connected by a path of edges, so they are equivalent; but two vertices on two different islands cannot be equivalent, because there is no path of edges to connect them. If there are three islands, then there are exactly three distinct cosets in the homology space H_0, i.e. three distinct classes of vertices, and so $b_0 = 3$.

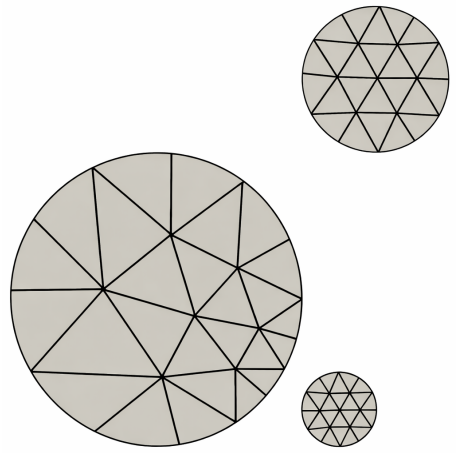


Figure A.7: A simplicial complex with three connected components, $b_0 = 3$.

Using the standard linear algebra fact that for a subspace $V \subset W$

$$\dim(W/V) = \dim(W) - \dim(V), \tag{A.2.7}$$

it follows that

$$b_k = \dim Z_k - \dim B_k = \dim \ker(\delta_k) - \dim \operatorname{im}(\delta_{k+1}). \tag{A.2.8}$$

This is true for all $k \in \mathbb{N}$, so long as we suitably interpret δ_k. In particular, we set $\delta_k = 0$ if there are no k-simplices in the complex. For example, if we have a 2-complex, a complex where the highest dimensional simplices are triangles, then $\delta_3 = 0$, $\dim \operatorname{im}(\delta_3) = 0$, and $b_2 = \dim(\ker \delta_2)$. Similarly, as vertices have no boundary, $\dim Z_0 = \dim \ker(\delta_0) = \dim C_0$, and so $b_0 = \dim C_0 - \dim \operatorname{im}(\delta_1)$.

A.3 *Implementing simplicial complexes*

For our purposes we will represent simplicial complexes with two arrays `cells` and `coordinates`. The first records the connections between the highest-dimensional cells and vertices, with each row corresponding to one cell. For example, for the complex in Figure A.4, the `cells` array is just

(python)
```
cells = [[0, 1, 2]]
```

The second array records the coordinates of each vertex:

(python)
```
coordinates = [[0, 0], [1, 0], [0.5, 0.5]]
```

Note that the decomposition into two arrays mirrors the distinction between topology and geometry: `cells` carries the topology, and `coordinates` carries the geometry. This representation is quite compressed: while it specifies an enumeration of the cells and vertices, it does not specify an enumeration of any other simplices[10].

[10] More precisely, this compressed representation is only able to represent *pure* simplicial complexes, where e.g. in a pure 2-complex every edge present is an edge of a triangle. This data structure could not represent e.g. a triangle with another independent edge adjoined to one of its vertices. We will assume our simplicial complexes are pure.

After installing `pyvista` with

(terminal) `python -m pip install pyvista`

we can plot the complex with the code supplied in `plot_complex.py`[11] via

[11] If this does not work for you, it is not fatal; you do not strictly need to visualise the complexes to answer the following questions.

(python) `from plot_complex import plot_complex`

(python) `plot_complex(cells, coordinates)`

Later on we will load complexes from `.npz` files. These have been generated with `np.save`, so the right way to load them is with the following code block:

Code block A.1. Loading cells and connectivity arrays from a `.npz` file

```
def load_complex_npz(file):
    data = np.load(file)
    cells = data["cells"]
    coordinates = data["coordinates"]
    return (cells, coordinates)
```

The `.npz` files are all contained in a single zipfile `complexes.zip`. You do not need to unzip this manually: you can access all of the `.npz` files within directly from Python:

Code block A.2. Loading `.npz` files from a `.zip` file

```
import zipfile

# Print out files inside the .zip
with zipfile.ZipFile("complexes.zip") as z:
    filenames = z.namelist()
print(f"Available filenames: {filenames}")

# Load a complex from within the .zip
def load_complex_zip(zipname, complexname):
    with zipfile.ZipFile(zipname) as z:
        with z.open(complexname) as f:
            return load_complex_npz(f)
```

All code for this project should be written in a dimension-generic manner, so that it can be applied to simplicial complexes of arbitrary

dimension.

Question A.1. Our first task is to enumerate all the simplices from the compressed representation `cells`.

Write a function `build_complex(cells)` that builds a data structure `complex` that is a list of dictionaries. Each dictionary `complex[dim]` should map a tuple of vertices of length `dim` *with vertices listed in ascending order*[12] to an id number. The id numbers should be natural numbers beginning from zero. The vertices should be plain Python integers (not e.g. `np.int32`). For example, for the complex of Figure A.4 that is stored in `triangle.npz`, the output should be

[12] Recall that we have decided on the convention that a simplex with vertices in sorted ascending order is positively oriented.

Code block A.3. sample output for Question A.1

```
# complex[dim] = {sorted_vertex_tuple: simplex_id}
complex = [
    {(0,): 0, (1,): 1, (2,): 2},          # dimension 0 (vertices)
    {(0, 1): 0, (1, 2): 1, (0, 2): 2},    # dimension 1 (edges)
    {(0, 1, 2): 0}                        # dimension 2 (triangles)
]
```

Only simplices actually present in the complex should be listed in the dictionary: for example, if vertices v_3 and v_8 are not connected by an edge, `(3, 8)` must not be a key in `complex[1]`. The map encoded in `complex[0]` should always be the identity.

Apply your code to the complexes in `triangle.npz` and `triangle_minus_triangle.npz`. Run the following diagnostic code on the two complexes:

Code block A.4. Diagnostics for Question A.1

```
from pprint import pprint

for (dim, c) in enumerate(complex):
    print(f"Printing complex of dimension {dim}")
    pprint(c)
```

[5 marks]

With this, we can now build the matrix representation of the boundary operator $\delta_k : C_k \to C_{k-1}$.

Question A.2. Write a function `boundary_matrix(complex, k)` that builds the matrix representation of the boundary operator $\delta_k : C_k \to C_{k-1}$ (A.1.6) as a numpy array[13].

Print all boundary operators for the complexes stored in `triangle.npz` and `triangle_minus_triangle.npz`.

[5 marks]

[13] Since most entries of the matrix are zero, a production code for simplicial homology would instead use a *sparse matrix format*, one that only explicitly stores the nonzero entries, and employ sophisticated algorithms from sparse linear algebra for the subsequent calculations. These algorithms are not robustly implemented in numpy or scipy, and to avoid further external dependencies, we instead employ *dense* (non-sparse) matrices here.

Question A.3. For each complex in `complexes.zip`, build all boundary operators, and verify that (A.1.11) holds.

[2 marks]

A.4 Computing Betti numbers

We are finally in a position to compute the Betti numbers of a manifold.

Question A.4. Using the rank-nullity theorem, derive a simpler expression for b_k in terms of matrix ranks from (A.2.8).

[3 marks]

Question A.5. Using the expression you derived in Question A.4, implement a function `betti_numbers(cells)` that computes the Betti numbers for a given `cells` array[14].

Compute and print the Betti numbers for each complex stored in `complexes.zip`.

[5 marks]

[*Hint: the rank of a matrix can be computed with* `np.linalg.matrix_rank`.]

[14] Note that this does not require the coordinates at all; the Betti numbers are purely topological.

A.5 Concluding remarks

The boundary of a boundary is zero. We have seen that the relation

$$\delta_k \circ \delta_{k+1} = 0 \tag{A.5.1}$$

encodes the homology of the manifold, in how closely the image of one operator coincides with the kernel of the other.

This is a very deep result; some consider it to be one of the mathematical foundations of reality. In their famous book on general relativ-

ity, Misner, Thorne, & Wheeler[15] write

> ...Thus simply is all of general relativity tied to the principle that the boundary of a boundary is zero. No one has ever discovered a more compelling foundation for the principle of conservation of momentum and energy. No one has ever seen more deeply into that action of matter on space, and space on matter, which one calls gravitation. In summary, *the Einstein theory realizes the conservation of energy-momentum as the identity, ''the boundary of a boundary is zero.''*

[15] Thorne won the Nobel Prize in Physics in 2017 for his contributions to the detection of gravitational waves. The quote is from p. 380.

C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. Princeton University Press, Princeton, USA and Oxford, UK, 2nd edition, 2017

B 2026B: predicting eclipses

Solar eclipses are spectacular and rare. It is no great surprise that different cultures have interpreted them as portending doom or disaster. For example, in the Neo-Assyrian empire (a powerful state in the middle east from c. 911 BC to 609 BC), certain kinds of eclipses indicated that the gods were angry with the king, and that he must therefore be executed.

Figure B.1: Tablet K.2600, on display in the British Museum. This fragment, combined with several others, describes the ritual for appointing a sacrificial king.

Naturally, the king had other ideas. As soon as the eclipse occurred,

the king would appoint a *substitute king*: a condemned criminal, prisoner of war, slave, political adversary, etc. The substitute was then anointed, dressed in the king's robes, and given a companion to serve as queen. The real king went into hiding, until the bad omens of the eclipse were firmly and fatally attached to the substitute. At the end of the process, the unfortunate substitute and queen were executed so that the real king could resume his reign. The process is documented on clay tablets held in collections worldwide, such as tablet K.2600 in the British Museum, shown in Figure B.1[1]; for further details, see Chapter 9 of Bottéro[2].

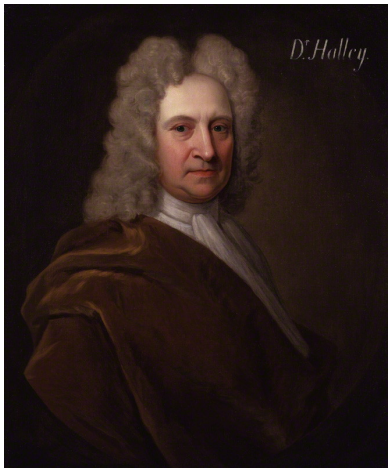
Of course, the Neo-Assyrian court mathematicians did not have Newton's theory of gravity, and could not predict eclipses with any great accuracy. They instead relied on statistical patterns among eclipses, the so-called Saros cycle; in a sense, they were among the first data scientists. Precise predictions were only possible thousands of years later.

[1] W. G. Lambert. A part of the ritual for the substitute king. *Archiv für Orientforschung*, 18:109–112, 1957

[2] J. Bottéro. *Mesopotamia: Writing, Reasoning, and the Gods*. University of Chicago Press, 1995. Translated by Zainab Bahrani and Marc Van De Mieroop.

Perhaps the most famous prediction of an eclipse in the modern era was that of Edmund Halley, Savilian Professor of Geometry at our University, in 1715. Halley had been the primary driver and funder of the publication of Newton's *Principia*, and was foremost in applying its ideas to explain the universe. In 1715 Halley published a prediction of the path of an eclipse to happen two weeks later, in a pamphlet intended for the general public, and reproduced in Figure B.2. Halley's motivations would be familiar to the Neo-Assyrians: as he explains,

> The like Eclipse having not for many Ages been seen in the Southern Parts of Great Britain, I thought it not improper to give the Publick an Account thereof, that the suddain darkness, wherin the Starrs will be visible about the Sun, may give no surprize to the People, who would, if unadvertized, be apt to look upon it as Ominous, and to Interpret it as portending evill to our Sovereign Lord King George and his Government, which God preserve.



Edmund Halley, 1656–1742

Halley's predictions were stunningly accurate, given the limitations of the data[3] and computational power available to him: he predicted the time of totality to within 4 minutes, and the border of the eclipse to within 30 kilometres, as he later recounted in an article for the Royal Society[4].

In this project we shall compute and plot the path of the eclipse, as Halley pioneered.

[3] To achieve better accuracy, modern predictions of eclipses take account of: the Moon is lopsided, and its centre of mass does not coincide with its centre of figure; the Moon's surface is quite rough, and its valleys allow rays from the Sun to pass; the Earth is not spherical, but is better described as an ellipsoid, since its rotation causes a bulge at the equator; the Earth's rotational speed is not constant, but slows over time as the tides impart angular momentum to the Moon; etc.

[4] E. Halley. Observations of the late Total Eclipse of the Sun on the 22d of April last past, made before the Royal Society at their House in Crane-Court in Fleet-street, London. *Philosophical Transactions of the Royal Society of London*, 29(343):245–262, 1715

B.1 *Besselian elements: a game of cones*

The modern framework for describing eclipses was developed by the Prussian polymath Friedrich Wilhelm Bessel. Bessel left school at the age of 14, but came to the attention of a well-known astronomer by

Figure B.2: The pamphlet published by Halley in 1715 depicting, in advance, the path of the eclipse over London.

refining the calculations of the orbit of Halley's comet (discussed in project 2025B), using data compiled in part by the Oxford mathematician Thomas Harriot. He was appointed director of the Königsberg Observatory in 1810, despite never having attended university.

The *Besselian* elements are a set of numbers that describe an eclipse[5]. They are used to calculate information such as whether a given location will be in total or partial eclipse, at what time will the eclipse occur in a given location, etc. The Besselian elements are themselves computed from high-resolution simulations of the solar system with ordinary differential equations and are published by organisations such as NASA and HM Nautical Almanac Office.

[5] F. W. Bessel. Analyse der Finsternisse. In *Astronomische Untersuchungen*, volume 2. Bornträger, 1841–1842

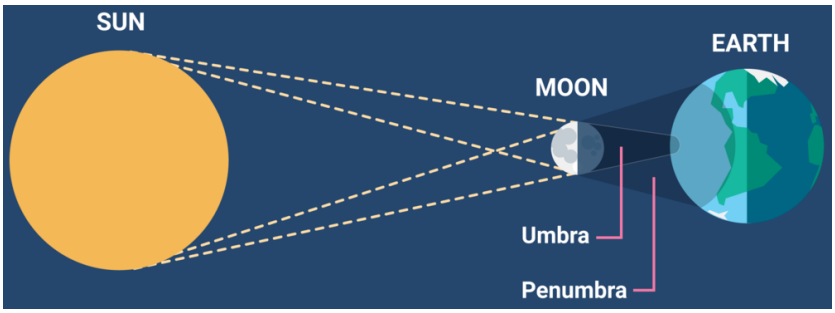


Figure B.3: The umbral and penumbral cone of a solar eclipse. Credit: timeanddate.com.

Assuming the Sun and Moon to be spherical, the shadow cast by the Moon is a cone. There are actually two cones of interest, the umbral cone (total darkness) and the penumbral cone (partial darkness), as shown in Figure B.3. As you know from Prelims Geometry, the intersection of a cone and a plane forms a conic section, so that the boundary of the shadow would be a circle/ellipse/parabola/hyperbola. However, the Earth is not planar, and the shape of the shadow on its surface is thus more complex to describe.

Bessel's key idea was to use a clever coordinate system in which the umbral and penumbral shadows of the Moon are disks. The *fundamental plane* has its origin at the Earth's centre, and its normal direction is defined to be the *shadow axis*, the vector joining the centre of the Moon to the centre of the Sun. Since the fundamental plane is always normal to the shadow, the shadow may be described by a few simple functions of time. The first we meet are the $x(t)$ and $y(t)$ coordinates on the fundamental plane, and the radii $l_1(t)$ and $l_2(t)$ of the penumbral and umbral shadows respectively, see Figure B.4[6]. The unit of length is chosen to be the Earth's equatorial radius. The coordinate system established on the fundamental plane is such that the x-axis is the intersection of the fundamental plane and the plane of the Earth's equator; it is positive to the east (in the direction of the vernal equinox); the z-axis is the shadow axis; the y-axis is constructed to

Friedrich Bessel, 1784–1846

[6] In fact, by convention $l_2(t)$ is the negation of the radius—$l_2(t)$ is positive if the umbral shadow does not reach the Earth, so that the eclipse is annular rather than total. This occurs if the Sun is closer and the Moon is further away than in a total eclipse. We shall not consider this case in this project.

make a right-handed triple, with the positive y-direction to the north.

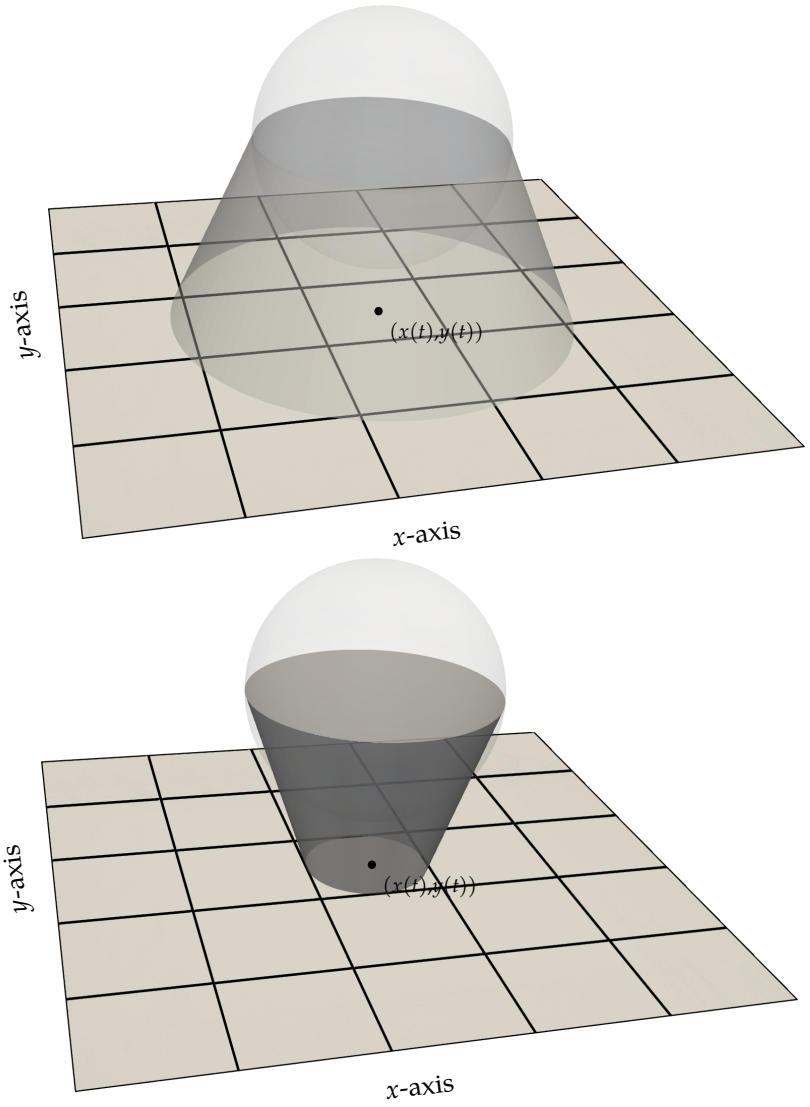


Figure B.4: An illustration of some of the Besselian elements. The Moon (transparent sphere) casts a penumbral shadow (top), where some of the Sun's light is blocked and the eclipse will be partial, and an umbral shadow (bottom), where all of the Sun's light is blocked and the eclipse will be total. By construction, the fundamental plane (the grid) is normal to the cones and their projections onto the fundamental plane are disks. The centre of the disk is $(x(t), y(t))$. The radius of the penumbral disk on the fundamental plane is $l_1(t)$ and the radius of the umbral disk is $|l_2(t)|$. The normal to the plane is described by $d(t)$ and $\mu(t)$.

The functions $x(t)$ and $y(t)$ are described by cubic Taylor expansions around a specified time t_0, typically the nearest hour to the time of greatest eclipse:

$$x(t) \approx x_0 + x_1(t - t_0) + x_2(t - t_0)^2 + x_3(t - t_0)^3, \tag{B.1.1a}$$

$$y(t) \approx y_0 + y_1(t - t_0) + y_2(t - t_0)^2 + y_3(t - t_0)^3, \tag{B.1.1b}$$

with t_0 and $x_0, \ldots, x_3, y_0, \ldots, y_3$ given data. Here $(t - t_0)$ is measured in hours. Since an eclipse can only last a few hours, this cubic expansion is adequate. The other functions are approximated similarly, with $l_1(t)$ and $l_2(t)$ approximated with quadratic Taylor expansions.

The next two functions describe the orientation of the shadow axis with respect to a second coordinate system, geocentric polar

coordinates. This is the polar coordinate system associated to the Cartesian geocentric coordinate system where the origin is fixed at the centre of the Earth, the z-axis points toward the north pole, the x-axis points towards the Greenwich meridian plane[7] at the equator, and the y-axis completes the right-handed triple. In this geocentric polar coordinate system, the direction of the shadow axis is described by two further Besselian elements, $d(t)$ and $\mu(t)$. The quantity $d(t)$ is the *declination* of the shadow axis—the angle it makes north or south of the celestial equator. The quantity $\mu(t)$ is its *hour angle* as measured from the Greenwich meridian plane. More precisely, the hour angle of a direction in the sky is the angle, measured westwards from the Greenwich meridian plane to the meridian plane containing that direction. Together, $d(t)$ and $\mu(t)$ specify the orientation of the shadow axis, just as latitude and longitude specify the orientation of a point on Earth. These are again provided as quadratic polynomials in $t - t_0$ in the published tables.

[7] The meridian plane is a plane passing through a given place on the Earth and the two poles.

The final two Besselian elements, $f_1(t)$ and $f_2(t)$, describe the shape of the penumbral and umbral cones themselves. A cone is not determined only by the radius and centre of its base; we also need to know its height, or equivalently its semi-vertex angle (half the total angle of the vertex). The semi-vertex angles for the penumbral and umbral cones are given as $f_1(t)$ and $f_2(t)$. These change so slowly that they can be treated as constant over the duration of an eclipse. Typical tables of Besselian elements tabulate the tangents of these angles, since that is the quantity actually employed in calculations.

There is one final constant element, ΔT. This relates two different systems for measuring time: Terrestrial Dynamical Time (TDT) and Universal Time (UT). The basic unit of time (the day) derives from the rotation of the Earth. However, the speed of rotation of the Earth is not constant, primarily due to tidal friction between the Earth's oceans and its rocky mantle through the gravitational actions of the Moon and the Sun. UT is based on this rotation and is thus not uniform (the length of the day varies over time). Its uniform counterpart is TDT, where the length of the day is constant; it is this timescale that we will use for calculations. At any given moment of time TDT and UT are out of sync; the difference is known as ΔT. This discrepancy ΔT can only be deduced from observations; its value after the invention of the telescope in 1608 is well-determined, but before 1608 it is highly uncertain.

The Besselian elements for the eclipse predicted by Halley are given in Table B.1.

Question B.1. Use the `numpy.polynomial.Polynomial` class

Variable	expression
t_0	1715-05-03 10:00:00 Terrestrial Dynamical Time (TDT)
ΔT	9.6 s
$x(t)$	$0.0718130 + 0.5682352(t-t_0) + 0.0000137(t-t_0)^2 - 0.0000096(t-t_0)^3$
$y(t)$	$0.7433290 + 0.1231166(t-t_0) - 0.0001459(t-t_0)^2 - 0.0000020(t-t_0)^3$
$l_1(t)$	$0.5330720 + 0.0000319(t-t_0) - 0.0000128(t-t_0)^2$
$l_2(t)$	$-0.0130010 + 0.0000318(t-t_0) - 0.0000127(t-t_0)^2$
$d(t)$	$\pi/180 \times \left(15.5370197 + 0.0120410(t-t_0) - 0.0000030(t-t_0)^2\right)$ rad
$\mu(t)$	$\pi/180 \times \left(330.845947 + 15.002640(t-t_0)\right)$ rad
$\tan f_1$	0.0046323
$\tan f_2$	0.0046092

Table B.1: Besselian elements for the eclipse of 3 May 1715, the one predicted by Halley in Figure B.2. Source: Eclipse Predictions by Fred Espenak, NASA's GSFC.

to represent the different Besselian elements. Plot the coordinates $(x(t), y(t))$ on the Besselian plane for the elements given in Table B.1, for $(t - t_0) \in [-2.5, 2.5]$ hours.

[1 mark]

B.2 *Coordinate systems for the Earth*

In this question there are four coordinate systems of interest:

1. the Besselian coordinate system on the fundamental plane (Cartesian);
2. the geocentric coordinate system (Cartesian);
3. its polar analogue, geocentric latitude and longitude (polar);
4. geographical latitude and longitude (polar).

These last two differ because the Earth is not a perfect sphere. Instead, it is more accurately modelled as an ellipsoid.

Imagine a line connecting your location with the center of the Earth. The geocentric latitude is the angle that line makes with the plane of the equator. Now, imagine a line normal to the Earth at your location and take the angle *that* line makes with the plane of the equator. This angle is the geographical latitude.

On a sphere, the inward normal direction always points towards the centre, so the two notions coincide. But on an ellipsoid, the normal does *not* point toward the centre of the Earth, and the two latitudes differ slightly. This is illustrated in Figure B.5. Geocentric and geographical longitudes always do coincide.

Mathematicians and astronomers have studied the shape of the Earth for centuries[8]. Newton argued in Book III, Proposition XIX of the *Philosophiæ Naturalis Principia Mathematica*[9] that the Earth must

[8] G. Heine. Euler and the flattening of the earth. *Math Horizons*, 21(1):25–29, 2013

[9] I. Newton. *Philosophiæ Naturalis Principia Mathematica*. Royal Society, 1726. Translated by I. B. Cohen, A. Whitman and J. Budenz, University of California Press, 1999

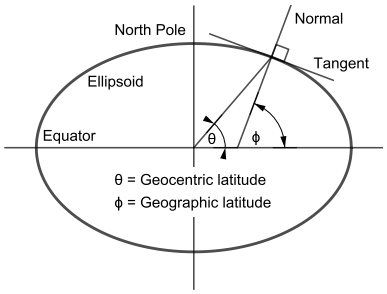


Figure B.5: Two types of latitude, geocentric and geographical latitude. On a sphere they coincide, but on an ellipsoid they do not. Credit: Wikipedia.

bulge at the equator and flatten at the poles. His theoretical predictions were investigated by the French Academy of Sciences, which in the 1730s commissioned two expeditions to measure the length of a degree of latitude, one led by Maupertuis and Celsius to Lapland, and the other led by de la Condamine, Bougeur, and Godin to Ecuador. Their confirmation of the flattening of the Earth was a major scientific event, as it vindicated Newton. Voltaire later mocked Maupertuis as ''the flattener of the Earth''[10].

The modern model for the shape of the Earth is known as the WGS84 standard[11], used e.g. by GPS and all navigation software. The shape is a special case of an ellipsoid, called an oblate spheroid (both equatorial radii assumed equal). The WGS84 model sets the equatorial radius to be $a = 6378137.0$ m, and the shorter polar radius to be $b = 6356752.314245$ m. Important numbers arising in the calculations below include the flattening

$$f = 1 - \frac{b}{a} \tag{B.2.1}$$

and the squared eccentricity

$$e^2 = 2f - f^2 = 1 - \frac{b^2}{a^2}. \tag{B.2.2}$$

With these radii, the geocentric latitude θ and geographical latitude ϕ for a given location are related by

$$\phi = \tan^{-1}\left((a/b)^2 \tan\theta\right), \tag{B.2.3a}$$

$$\theta = \tan^{-1}\left((b/a)^2 \tan\phi\right). \tag{B.2.3b}$$

Question B.2. Plot the difference $\theta - \phi$ as a function of geocentric latitude θ. On the Earth, what is the maximal value of $|\theta - \phi|$, and where does it occur? Report your results in both radians and degrees. What is the maximal value of the difference on Jupiter, where $a \approx 71492$ km and $b \approx 66854$ km[12]? Comment on your results.

[2 marks]

[10] ''Vous devez, mon cher aplatisseur de ce globe, avoir reçu une invitation ...''.

Voltaire. Lettre 1476 à M. de Maupertuis. In T. Besterman, editor, *Correspondance*, volume 36, pages 102–103. Garnier, 1968

[11] National Imagery and Mapping Agency. Department of defense world geodetic system 1984: Its definition and relationships with local geodetic systems. Technical Report TR8350.2, NIMA, St. Louis, MO, 2000

[12] A. Prša, P. Harmanec, G. Torres, E. Mamajek, M. Asplund, N. Capitaine, J. Christensen-Dalsgaard, É. Depagne, M. Haberreiter, S. Hekker, J. Hilton, G. Kopp, V. Kostov, D. W. Kurtz, J. Laskar, B. D. Mason, E. F. Milone, M. Montgomery, M. Richards, W. Schmutz, J. Schou, and S. G. Stewart. Nominal values for selected solar and planetary quantities: IAU 2015 Resolution B3. *The Astronomical Journal*, 152(2):41, 2016

B.3 *Computing the centre line*

Our first task in plotting the eclipse is to compute the centre line of the eclipse. At a given time t, this is the point on the Earth's surface where the centre of the Moon's shadow lies. The following formulae are drawn from section §11.3 of the *Explanatory Supplement to the Astronomical Almanac*[13].

[13] S. Urban and P. K. Seidelmann, editors. *Explanatory Supplement to the Astronomical Almanac*. University Science Books, 2012

Imagine an observer on the surface of the Earth. If the Earth were a sphere, the distance from the observer to the centre of the Earth would be a constant. However, on an ellipsoid, the radius itself is a function of latitude, which is among the variables that we are trying to compute. A naïve approach to calculating the latitude and longitude of the centre line would therefore take a guess for the radius, compute the latitude, update the guess for the radius, re-compute the latitude, and so iterate until convergence. However, one of the major contributions of Bessel's analysis was an approach that can directly map coordinates (σ, η) in the fundamental plane to the geocentric latitude and longitude of the projection of (σ, η) along the shadow axis onto the surface of an ellipsoid without iteration.

The geometric derivation is somewhat complicated. We briefly summarise the key formulae of Bessel's approach here. In the first step, calculate the auxiliary values

$$\tilde{\rho} = \sqrt{1 - e^2 \cos^2 d}, \quad \sin \tilde{d} = \frac{\sin d}{\tilde{\rho}}, \quad \cos \tilde{d} = \frac{(1-f)\cos d}{\tilde{\rho}}, \tag{B.3.1}$$

which account for the flattening of the Earth. Here $\tilde{\rho}$ is a latitudinal scaling factor that normalises the y–coordinate so that it can be treated as if it were on a unit sphere, and $\tilde{d}$ is an auxiliary declination that captures the tilt of the shadow axis after the Earth's ellipsoidal shape has been transformed to a sphere.

Having defined the auxiliary elements to account for the Earth's flattening, we now determine the coordinates of the point where the shadow axis intersects this ''stretched'' spherical model of the Earth. For computing the centre line, the points on the fundamental plane that we wish to map are the coordinates of the centre of the shadow (x, y), so we set $(\sigma, \eta) = (x, y)$. We then compute a normalised y-coordinate $\tilde{\eta}$, normalised by the scaling factor:

$$\tilde{\eta} = \frac{\eta}{\tilde{\rho}}. \tag{B.3.2}$$

To complete the position on the unit sphere, we calculate the auxiliary depth coordinate $\tilde{\zeta}$, which represents the distance of the point orthogonal to the fundamental plane, toward the Moon:

$$\tilde{\zeta} = \sqrt{1 - \sigma^2 - \tilde{\eta}^2}, \text{ for } 1 - \sigma^2 - \tilde{\eta}^2 \geqslant 0. \tag{B.3.3}$$

If $1 - \sigma^2 - \tilde{\eta}^2 < 0$ for a given time t, there is no projection of (σ, η) along the shadow axis onto the ellipsoid. In the context of the centre line, this means that the shadow axis misses the earth completely at that time. The times t for which

$$\sigma(x)^2 + \tilde{\eta}(y)^2 = 1 \tag{B.3.4}$$

are the start and end times of the centre line.

With $(\sigma, \tilde{\eta}, \tilde{\zeta})$ computed, we can now transform these Cartesian coordinates into the observer's normalised latitude $\tilde{\phi}$ and local hour angle θ. This is achieved by reversing the rotation of the coordinate system by the auxiliary declination $\tilde{d}$. The relationship is given by:

$$\begin{bmatrix} \cos\tilde{\phi}\sin\theta \\ \sin\tilde{\phi} \\ \cos\tilde{\phi}\cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\tilde{d} & \sin\tilde{d} \\ 0 & -\sin\tilde{d} & \cos\tilde{d} \end{bmatrix} \begin{bmatrix} \sigma \\ \tilde{\eta} \\ \tilde{\zeta} \end{bmatrix}, \tag{B.3.5}$$

with these equations solved for $\tilde{\phi}$ and θ.

The final step is to convert these intermediate values into standard geographical coordinates. The longitude λ is found by relating the local hour angle to the Greenwich hour angle μ of the shadow axis:

$$\lambda = \theta - \mu - \Omega\Delta T, \tag{B.3.6}$$

where $\Omega \approx 7.29211585 \times 10^{-5}$ is a constant encoding the difference between TDT and UT[14]. The geographical latitude ϕ is then recovered from the parametric latitude $\tilde{\phi}$ by removing the scaling effect of the Earth's eccentricity:

$$\tan\phi = \frac{\tan\tilde{\phi}}{1-f}. \tag{B.3.7}$$

By evaluating these expressions for a sufficiently fine grid of times t between the start and end times of the centre line, its path can be plotted across the Earth's surface.

[14] More precisely, it is the uniform angular velocity of the Earth ($2\pi/86400$ rad s^{-1}) multiplied by the ratio of a sidereal day to a solar day (86400/86164.0905 s/s).

Question B.3. Write a function `compute_latitude_longitude` that implements the projection of points (σ, η) on the fundamental plane along the shadow axis given by (B.3.2)–(B.3.7). For later use, the function should take σ and η as matrices of shape $n_{\text{pts}} \times n_{\text{times}}$, where n_{pts} is the number of points at a given time $t - t_0$ and n_{times} is the number of different times to consider.

What is the geographical latitude and longitude of the centre line at the time of maximal eclipse,

$$t_{\max} = \text{1715-05-03 09:39:30 TDT?} \tag{B.3.8}$$

[4 marks]

[*Hint: use* `np.arctan2` *or similar to calculate* θ.]

Question B.4. Using numerical rootfinding, determine the start and end times for the 1715 eclipse. Express your answer in hours from t_0.

[3 marks]

[*Hint: an eclipse on Earth cannot last more than 8 hours. Use this to generate good initial guesses, then apply a standard rootfinding algorithm.*]

Question B.5. Plot the centre line of the 1715 eclipse on a map of the Earth. The map should be drawn with the `cartopy` package and should render coastline boundaries. Consult the `cartopy` documentation and examples online to learn how to use it. Choose a suitable (i) map projection (ii) colour scheme for land/ocean (iii) resolution the coastline (iv) map region so that the entire centre line is visible.

[5 marks]

B.4 *Computing the region of total eclipse*

With the centre line now computed, we next consider the region of total eclipse. This is the intersection of the umbral cone with the curved surface of the Earth. An analogous computation holds for the penumbral cone.

Recall that the radius of the umbral cone on the fundamental plane is given by the Besselian element $l_2(t)$. However, because the Earth's surface is at a distance ζ from the fundamental plane, the local radius L_2 of the shadow must be adjusted. The radius at a height ζ is:

$$L_2 = l_2 - \zeta \tan f_2 \tag{B.4.1}$$

where recall that $\tan f_2$ is the semi-vertex angle of the umbral cone. By convention, $l_2 < 0$ for a total eclipse. A point (σ, η, ζ) on the surface of the Earth (described in Besselian coordinates) lies within the total eclipse if:

$$(\sigma - x(t))^2 + (\eta - y(t))^2 < L_2(t, \zeta)^2. \tag{B.4.2}$$

The boundary of the total eclipse can be computed by finding the intersection of the umbral cone and the Earth's ellipsoid. At a given time t, the boundary on the fundamental plane is a circle centered at $(x(t), y(t))$ with radius L_2. We can parameterize this boundary using

an angle $\alpha \in [0, 2\pi]$:

$$\sigma(\alpha, t) = x(t) + L_2(t)\cos\alpha \tag{B.4.3}$$
$$\eta(\alpha, t) = y(t) + L_2(t)\sin\alpha \tag{B.4.4}$$

For each α, we find the corresponding ζ by requiring the point to lie on the Earth's ellipsoid:

$$\sigma^2 + \eta^2 + \frac{\zeta^2}{1-e^2} = 1 \implies \zeta = \sqrt{(1-e^2)(1-\sigma^2-\eta^2)} \tag{B.4.5}$$

Note that L_2 itself depends on ζ. In practice, since $\tan f_2$ is very small, one can solve this by fixed point iteration: start with $\zeta = \sqrt{(1-e^2)(1-x^2-y^2)}$, compute L_2, update σ and η, and recompute ζ until convergence.

Question B.6. Write a function that compute N points uniformly distributed in angle α on the boundary of the umbral cone in Besselian coordinates at a given time $t - t_0$ using the fixed-point iteration (B.4.1)–(B.4.5). Terminate each fixed point iteration when the update to ζ is sufficiently small. What is the Besselian coordinates of the point on the umbral boundary with $\alpha = \pi/2$ at $t - t_0 = -0.87$?

[2 marks]

Question B.7. Convert the Besselian coordinates computed in Question B.6 for $N = 100$, $t - t_0 = -0.87$ to geographical latitude and longitude using your code from Question B.3. Plot the resulting closed loop on a map of southern England with `cartopy`, along with the centre line. Does the shadow encompass London (geographical latitude 51.5074°, longitude -0.1278°)?

[3 marks]

B.5 *Concluding remarks*

Halley offered a lovely description of the eclipse[15]:

> From this time the Eclipse advanced, and by Nine of the Clock was about Ten Digits, when the Face and Colour of the Sky began to change from perfect serene azure blew, to a more dusky livid Colour having an eye of Purple intermixt, and grew darker and darker till the total Immersion of the Sun, which happened at $9^h 9' \cdot 17''$ by the Clock ...This Moment was determinable with great nicety, the Sun's light being extinguish'd at once; and yet more so was that of the Emersion, for the Sun came out in an Instant with so much Lustre that it surprized the Beholders, and in a Moment restored the Day.

[15] E. Halley. Observations of the late Total Eclipse of the Sun on the 22d of April last past, made before the Royal Society at their House in Crane-Court in Fleet-street, London. *Philosophical Transactions of the Royal Society of London*, 29(343):245–262, 1715

Halley also took care to report observations of the eclipse in other locations. Both Oxford and Cambridge hold lessons for us:

> Our Professors of Astronomy in both Universities were not so fortunate: My worthy Colleague Dr. John Keill by Reason of Clouds saw nothing distinctly at Oxford but the End …And the Reverend Mr. Roger Cotes at Cambridge had the misfortune to be opprest by too much Company, so that, though the Heavens were very favourable, yet he miss'd both the time of the Beginning of the Eclipse and that of total Darkness.

If you ever get the chance to observe an eclipse, go somewhere with good weather, and make sure to be ready on time!

C 2026C: *irreversibility from reversible dynamics*

Do you believe the atomic theory of matter?

Of course you know the theory to be true; you have been taught it in school. But do you *believe* it? It is not obvious that it is true; it is not at all straightforward to reconcile the atomic theory of matter with our everyday experience of how the world works. Aristotle argued that the materials we interact with are continua when he rejected the atomism of Leucippus and Democritus in Book I, Chapter 2 of *On Generation and Corruption*[1]. His incorrect argument was to mislead for over 2000 years.

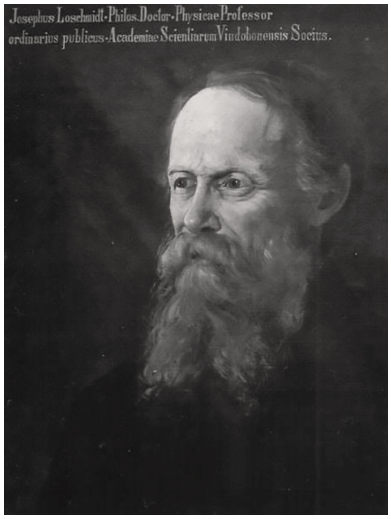
A more sophisticated difficulty with the atomic theory of matter arose with the development of thermodynamics. As described in Gottwald & Oliver[2],

> All theories of microscopic physics are governed by laws that are invariant under reversal of time—the evolution of the system can be traced back into the past by the same evolution equation that governs the prediction into the future. Processes on macroscopic scales, however, are manifestly irreversible. Fluids mix, but cannot be unmixed by the same process ... Steam engines use differences in temperature to do useful mechanical work; in doing so, heat flows from hot to cold, eventually equilibrates, and reduces the potential for doing further mechanical work.

How can a process that is completely reversible at the microscale induce irreversible phenomena at the macroscale? Imagine the air in your bedroom: the microscale description would involve knowing all of the positions and velocities of the roughly 10^{27} particles; the macroscale description would describe the gas in a *coarse-grained* manner, in terms of its density, velocity, and temperature at every point in the continuum of the room. Both viewpoints are correct and accurate descriptions of the same system: how can it possibly be that one of these is reversible, and the other irreversible? This is known as Loschmidt's paradox, after Josef Loschmidt, who raised it in the context of the fundamental work of Maxwell and Boltzmann on the irreversible generation of entropy in the statistical mechanics of gases[3].

[1] Αριστοτελης. *On Generation and Corruption*. Oxford University Press, Oxford, 335–323 B.C. Translated by H. H. Joachim, published in 1922

[2] G. A. Gottwald and M. Oliver. Boltzmann's dilemma: an introduction to statistical mechanics via the Kac ring. *SIAM Review*, 51(3):613–635, 2009



Josef Loschmidt, 1821–1895

[3] J. J. Loschmidt. Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Classe*, 73:128–142, 1876

The short answer to this paradox is that the macroscopic laws of physics are a fundamentally statistical description: they represent the most probable behaviour of the system. As the number of particles in the system increases, the probability of deviating from this most probable behaviour becomes vanishingly small. In this sense the equations governing the macroscopic behaviour of continua should be understood as *exceptionally strong suggestions* for how the system ought to behave. While this resolution is now widely accepted, rigorously proving and quantifying this for real physical systems is extremely difficult, and many open mathematical questions remain in this area.

In this project we will build intuition for this resolution of Loschmidt's paradox. We will do so by studing the Kac ring, a microscopically-reversible model proposed by Mark Kac in 1956 that exhibits the central thermodynamical difficulty of irreversible macroscopic phenomena while being simple enough to permit explicit computation[4].

Mark Kac, 1914–1984

[4] M. Kac. Some remarks on the use of probability in classical statistical mechanics. *Bulletin de la Classe des Sciences, Académie Royale de Belgique*, 42:356–361, 1956

[5] The following discussion draws heavily from the wonderful exposition of Gottwald & Oliver.

G. A. Gottwald and M. Oliver. Boltzmann's dilemma: an introduction to statistical mechanics via the Kac ring. *SIAM Review*, 51(3):613–635, 2009

C.1 *The Kac ring*

A Kac ring consists of N equidistant sites, arranged around a circle[5]. Neighbouring sites are joined by an *edge*. Each site is occupied by either a white ball or a black ball. A number $n < N$ of the edges are *marked*.

The system evolves on a discrete set of clock ticks $t \in \mathbb{Z}$ from state t to state $t + 1$ as follows. Each ball moves to the clockwise neighbouring site. When a ball passes a marked edge, it changes colour. The initial state with all white balls, and its evolution after one step, are shown in Figure C.1.

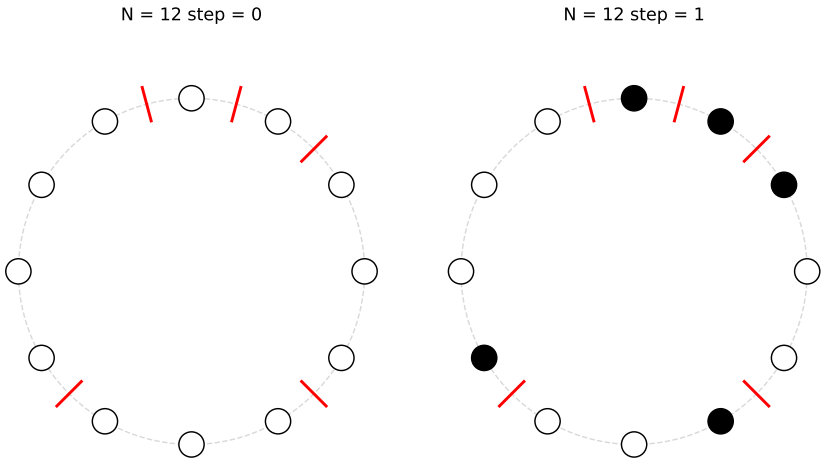


Figure C.1: Initial configuration of a Kac ring (left) and the state after one step (right). Marked edges are shown in red.

The dynamics of the Kac ring are time-reversible. When the direction of movement is changed from clockwise to counterclockwise, the balls retrace their past colour sequence without change to the rules of

the game. Moreover, after N clock ticks, each ball has reached its initial site and changed colour n times. Thus, if n is even the initial state recurs; if n is odd, it takes at most $2N$ clock ticks for the initial state to recur[6].

Let $W(t)$ denote the number of white balls and $B(t) := N - W(t)$ the number of black balls. The core question is what will happen to $W(t)$ and the difference

$$\Delta(t) := B(t) - W(t) \tag{C.1.1}$$

for a large number of moves (N large, $0 \ll t \ll 2N$). Think of

$$\delta(t) := \Delta(t)/N \tag{C.1.2}$$

as the colour of the system: if $\delta(t) = -1$ the system is all-white, if $\delta(t) = 1$ the system is all-black, and if $\delta(t) = 0$ the system is gray.

Let $w(t)$ denote the number of white balls in front of a marker, and $b(t)$ denote the number of black balls in front of a marker. Then

$$W(t+1) = W(t) + b(t) - w(t), \quad B(t+1) = B(t) + w(t) - b(t), \tag{C.1.3}$$

and hence

$$\Delta(t+1) = B(t+1) - W(t+1) = \Delta(t) + 2w(t) - 2b(t). \tag{C.1.4}$$

Here we think of W, B, Δ and δ as *macroscopic* quantities, describing a global feature of the system state[7]. We can observe and measure these quantities (we assume we know N for any given observed ring). On the other hand, w and b are *microscopic* quantities: they cannot be computed without full knowledge of the location of each marker and the colour of the ball at every site. We assume these microscopic quantities are not observable or controllable, just as we cannot in practice measure or set the position and momentum of each atom in a gas. As (C.1.3)–(C.1.4) show, the evolution of the macroscopic quantities *is not computable* only from macroscopic state information: it is not possible to eliminate w and b from the equations.

Question C.1. Implement the dynamics of one step of the Kac ring as a Python function.

[2 marks]

[*Hint: there is no need to optimise the efficiency of this function, as the heavy lifting will be done in a different function below.*]

Question C.2. Write a Python function to render the current configuration of a Kac ring as a figure like Figure C.1.

[6] This recurrence property is a consequence of the Poincaré recurrence theorem, a deep theorem of dynamical systems theory. It was mentioned in passing by Poincaré in the context of a special case of the three-body problem and later proven in generality by Birkhoff.

H. Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890

[7] They are all equivalent—if you know one, you know them all, if you know N.

[1 mark]

Question C.3. Initialise a Kac ring with $N = 12$, $W(0) = N$, and $B(0) = 0$ (all balls white). Mark each edge independently with probability $\mu = 1/3$. Render the initial configuration and three subsequent steps of the dynamics.

[1 mark]

If we only possess macroscopic information, how are we to close the system? *We need to make an additional assumption that the dynamics are 'typical'.* Denote by μ the probability that any particular edge is marked:

$$\mu := \frac{n}{N}. \tag{C.1.5}$$

We can assume that the fraction of black balls before a marker, and the fraction of white balls, are both μ:

$$\frac{b}{B} = \frac{w}{W} = \mu. \tag{C.1.6}$$

This (C.1.6) is the key assumption that allows us to close the system macroscopically. It effectively disregards the history of the system evolution: there is no memory of where the balls originated and which markers they have passed up to time t. For any particular realisation and configuration of a Kac ring it will not hold, and for small N it will be wildly wrong. But for large N we hope that (C.1.6) represents, in some sense, the typical behaviour of such rings[8].

Under the closure assumption (C.1.6), the evolution law (C.1.4) becomes

$$\Delta(t+1) = \Delta(t) + 2\mu W(t) - 2\mu B(t) = (1-2\mu)\Delta(t). \tag{C.1.7}$$

By induction, this recurrence yields

$$\delta(t) = (1-2\mu)^t \delta(0). \tag{C.1.8}$$

Clearly, this equation cannot describe the dynamics of any one particular ring exactly. Since $\mu < 1$, (C.1.8) dictates that $\delta(t) \to 0$ as $t \to \infty$ (the ring becomes gray). Furthermore, δ is monotonically-decreasing in magnitude and therefore not time-reversible; this is in contrast to what we know about the microscopic 'laws of physics'[9]. This is the manifestation of Loschmidt's paradox in the context of the Kac ring.

[8] This assumption is the analogue in the Kac ring of a famous assumption of Maxwell and Boltzmann in statistical mechanics, the 'molecular chaos' assumption. The molecular chaos assumption is that before two particles in a gas collide, their velocities and positions are statistically uncorrelated. This is reasonable in dilute gases at low densities, but fails for denser fluids or those with long-range interactions. This assumption is what breaks time-reversibility and introduces the 'arrow of time' in statistical mechanics: before collision the particles are uncorrelated, but after collision they are.

[9] In statistical mechanics, the analogue of this result is the famous H-theorem of Boltzmann, which states that the entropy of a gas can never decrease.

C.2 *Ensemble averages*

Our task is to give a meaning to the macroscopic evolution equation (C.1.8) on the basis of the microscopic dynamics. Boltzmann suggested that the macroscopic law he derived could only be valid in a

statistical sense, describing the most probable behaviour of a large ensemble of systems rather than the exact behaviour of any particular member of the ensemble. As we will see, for the Kac ring this suggestion can be made precise.

Define an ensemble of Kac rings to be a collection of rings with the same number of sites N. Each member of the ensemble is initialised with the same configuration of black and white balls. The markers, however, are placed at random on the edges such that μ is the probability that any one edge is marked. If X is some function of the configuration (e.g. $X = \delta(t)$), then X is thought of as a random variable, with expectation $\mathbb{E}[X]$.

Some elementary analysis (see (10)–(19) of Gottwald & Oliver) shows that for $0 \leqslant t < N$,

$$\mathbb{E}[\delta(t)] = (1 - 2\mu)^t \delta(0), \tag{C.2.1}$$

while for $N \leqslant t \leqslant 2N$,

$$\mathbb{E}[\delta(t)] = (1 - 2\mu)^{(2N-t)} \delta(0). \tag{C.2.2}$$

So at least for the early timesteps we are interested in (before the Poincaré recurrence theorem foils our plans) our macroscopic evolution law (C.1.8) derived with our closure assumption reflects the average behaviour of the system.

Of course, this result does not in any way imply that (C.1.8) represents 'typical' members of the ensemble, or that it is close to any individual system trajectory. For example, at the half-recurrence time $t = N$, each ball is back at its initial position with a possible global change of colour whenever the total number of markers is odd, so that $\delta(N) = \pm\delta(0)$ while, by (C.2.2), $\mathbb{E}[\delta(N)]$ is close to zero.

To quantify how much we expect the individual members of the ensemble to differ from the mean, we compute the variance

$$\mathbb{V}[\delta(t)] \coloneqq \mathbb{E}[\delta^2(t)] - \mathbb{E}[\delta(t)]^2 \tag{C.2.3a}$$

$$\tag{C.2.3b}$$

which after some calculation ((22)–(30) of Gottwald & Oliver) we find to be

$$\mathbb{V}[\delta(t)] \leqslant \frac{1}{N}\left(\frac{1}{2\mu(1-\mu)} - 1\right), \tag{C.2.4}$$

for early times $0 \leqslant t < N/2$. The bound holds with equality if $|\Delta(0)| = N$. Put loosely, we expect deviations from the mean of magnitude roughly $1/\sqrt{N}$.

Question C.4. Write a function `simulate_ensemble(M, N, mu, T)`, where M is the number of rings, N is the size of each ring, μ the

probability of an edge being marked, and T the final time. All balls should be initialised to be white. You are free to choose the representation of the configuration of the ensemble members and how the update rule is coded to maximise elegance, clarity, and computational efficiency. Your code must be efficient enough to execute the subsequent simulation tasks in a reasonable amount of time.

Take an ensemble of $M = 100$ rings with $N = 500$ and marked edge probability $\mu = 0.009$. For $0 \leqslant t \leqslant 1000$, plot $\delta(t)$ against t for

1. the output of the macroscopic evolution law (C.1.8);
2. the sample mean;
3. the trajectories themselves (with some transparency so that the plot remains comprehensible).

[6 marks]

[*Hint: be sure to label the plot, add a legend, use suitable colours, etc. Marks are awarded for clarity.*]

Question C.5. Explain the behaviour of the trajectories at $t = 500$ in the plot you produced in Question C.4. [*No coding required for this question.*

[1 mark]

Question C.6. For $N = 100, 800, 16000, 32000, 64000$, take an ensemble of $M = 1000$ rings. Initialise all balls as white ($W(0) = N$). For each N, set

$$\mu = \frac{1}{N^{1/2}}. \tag{C.2.5}$$

For each N, plot $\delta(t)$ against t for $0 \leqslant t \leqslant N/2$. Plot the following quantities:

1. the output of the macroscopic evolution law (C.1.8), $\pm$ the bound for standard deviation $\sqrt{\mathbb{V}[\delta(t)]}$;
2. the sample mean;
3. the first 100 trajectories.

Use the same scale on the y-axis for all plots.

Comment on your results, and their relevance to Loschmidt's paradox.

[6 marks]

C.3 Entropy

Entropy is a fundamental concept in thermodynamics, statistical mechanics, and information theory. It arises in the mathematical description of thermodynamic processes as a quantity that can only increase in time in closed systems. It accounts for the accumulation of irreversible phenomena. As energy is exchanged between different forms (e.g. from kinetic energy of motion to internal energy of heat) it is conserved, but it tends to become stored in less useful forms. The unstoppable increase in entropy records the fact that the energy of the system is less available to do useful work.

Entropy arises whenever a system is coarse-grained—whenever there is a microscopic configuration (thought of as unmeasurable/uncontrollable), and an associated macroscopic description that is measurable/controllable, but does not capture the full microscopic details. For example, in a gas in a room treated with Newtonian mechanics, the microscopic configuration at a given time would be the position and momentum of every particle in the system, while the macroscopic description would be given in terms of the density, velocity, and temperature at every point in the room.

In popular accounts entropy is often described as a measure of 'disorder'. This is imprecise and potentially confusing. It is better to think of entropy as a quantitative measure of non-injectivity of the coarse-graining map. One advantage of studying the Kac ring is that its entropy can be quantified simply and precisely.

In the Kac ring, our microstate is the full detailed description of the colour state of all balls. Our macrostate is specified equivalently by any of W, B, Δ or δ (from any one of these all others can be recovered). We now introduce the *partition function* Ω, which quantifies the number of microstates possible for any given macrostate. The state with W white balls and $B = N - W$ black balls can be realised in

$$\Omega(W) = \binom{N}{W} = \frac{N!}{W!B!} \tag{C.3.1}$$

different ways. The entropy S is defined to be the logarithm of this quantity:

$$S := \log \Omega. \tag{C.3.2}$$

This is because we want the entropy to be *additive*. In general, if we have two independent systems (think two Kac rings) each with respective partition functions Ω_A and Ω_B, then the number of microstates of the combined system is $\Omega_{AB} = \Omega_A \times \Omega_B$. By taking the logarithm, we ensure that the entropy of the combined system is $S_{AB} = S_A + S_B$. Note that the partition function and the entropy are

functions only of the macrostate: they quantify the non-injectivity of the many-to-one coarse-graining map.

To make further progress, we employ Stirling's approximation, stated by De Moivre in 1721 and sharpened in 1730 by James Stirling[10]. Stirling's approximation is that

$$\log k! = k\log k - k + \frac{1}{2}\log(2\pi k) + O(k^{-1}) \sim k\log k - k \sim k\log k \tag{C.3.3}$$

where the notation $a \sim b$ means that the ratio a/b tends to 1 as $k \to \infty$. Applying this simplification to (C.3.2), we find

$$\begin{aligned} S = \log\Omega &= \log\frac{N!}{W!B!} && \text{(C.3.4a)}\\ &= \log N! - \log W! - \log B! && \text{(C.3.4b)}\\ &\sim (N\log N - N) - (W\log W - W) - (B\log B - B) && \text{(C.3.4c)}\\ &= N\log N - W\log W - B\log B. && \text{(C.3.4d)} \end{aligned}$$

Introducing $p := B/N$ and $q := W/N$, and noting that $p + q = 1$, we have

$$\begin{aligned} S &\sim N\log N - qN\log(qN) - pN\log(pN) && \text{(C.3.5a)}\\ &= N\log N - qN(\log q + \log N) - pN(\log p + \log N) && \text{(C.3.5b)}\\ &= N\log N - (p+q)N\log N - qN\log q - pN\log p && \text{(C.3.5c)}\\ &= -N(p\log p + q\log q). && \text{(C.3.5d)} \end{aligned}$$

Since we want something comparable across rings of different sizes, we consider the entropy per site $s := S/N$. Dividing (C.3.5d) by N and rewriting in terms of the colour δ, we find

$$s \sim \log 2 - \frac{1}{2}\left[(1+\delta)\log(1+\delta) + (1-\delta)\log(1-\delta)\right]. \tag{C.3.6}$$

Question C.7. Plot (C.3.6) as a function of $\delta \in [-1, 1]$, with the convention that $x\log x = 0$ if $x = 0$. At what value is the entropy maximised?

[1 mark]

Question C.8. For large N, it follows from (C.2.1) and (C.2.4) that entropy should generally increase. Show, however, that this is not strictly necessary, by devising a special set of initial conditions which start at maximal entropy and for which the entropy subsequently decreases, at least for early times. (You can choose N and the marked edges.)

[10] James Stirling was a Scottish nobleman. He studied at Oxford from 1710 to 1715, when he was expelled for having corresponded with his cousins, who were noted Jacobites. (Jacobites rejected the Protestant Hanoverian king George I in favour of the exiled Catholic James Stuart.) He then settled in Venice, before again having to flee, as he feared assassination by the guild of glassmakers. He made major contributions to the mining industry in Scotland, and was involved in a project of deepening the Clyde river to Glasgow for establishing a seaport. The approximation that now bears his name is given in Example 2 to Proposition 28 of

J. Stirling. *Methodus Differentialis: sive Tractatus de Summatione et Interpolatione Serierum Infinitarum*. G. Strahan, London, 1730

Tabulate the entropy as a function of time t to show that entropy decreases for at least two timesteps. Plot the configurations using your code from Question C.2.

[2 marks]

C.4 *Concluding remarks*

Let us summarise what we have studied in this project.

For large N, the Kac ring is a system with a very large microscopic state space. At this level, the dynamics are entirely deterministic and time-reversible. However, the microscopic state is assumed to be unobservable and uncontrollable (just as an experimenter cannot actually measure the position and momentum of each molecule in the air of a laboratory). We thus introduce a coarse-graining function, a many-to-one map from the microscopic state space to a much smaller macroscopic state space; the output of this coarse-graining map is an experimentally accessible quantity (like density in a gas, or the colour δ in the Kac ring). Since this is the quantity we can measure, we wish to devise an evolution law from the macroscopic description of the system.

To close the macroscopic system we must make some assumption, e.g. that the dynamics are 'typical'. For small systems, this closure is not appropriate, and the macroscopic evolution law will not accurately predict the actual dynamics. But for many systems we observe that the macroscopic law describes the expected value of the observable, and that the variance from this expected value collapses as the size of the system increases. Thus, for large systems, the macroscopic evolution law does become an accurate description of what we expect to see—in the sense that we will probably have to wait many times the length of the universe to observe any deviation from the macroscopic law.

It is precisely this coarse-graining assumption that the dynamics are typical that breaks the time-reversible symmetry of the system. As Carlo Rovelli[11] puts it,

> Boltzmann has shown that entropy exists because we describe the world in a blurred fashion. He has demonstrated that entropy is precisely the quantity that counts *how many* are the different configurations that our blurred vision does *not* distinguish between. ... This is the disconcerting conclusion that emerges from Boltzmann's work: the difference between the past and the future refers only to *our own* blurred vision of the world. It's a conclusion that leaves us flabbergasted: is it really possible that a perception so vivid, basic, existential—my perception of the passage of time—depends on the fact that I cannot apprehend the world in all of its minute detail?

[11] C. Rovelli. *The Order of Time*. Riverhead Books, New York, NY, 2018. Originally published in Italian as L'Ordine del Tempo, 2017

Bibliography

[1] F. W. Bessel. Analyse der Finsternisse. In *Astronomische Untersuchungen*, volume 2. Bornträger, 1841–1842.

[2] E. Betti. Sopra gli spazi di un numero qualunque di dimensioni. *Annali di Matematica Pura ed Applicata*, 4(1):140–158, 1870.

[3] J. Bottéro. *Mesopotamia: Writing, Reasoning, and the Gods*. University of Chicago Press, 1995. Translated by Zainab Bahrani and Marc Van De Mieroop.

[4] M. H. Freedman. The topology of four-dimensional manifolds. *Journal of Differential Geometry*, 17(3), 1982.

[5] G. A. Gottwald and M. Oliver. Boltzmann's dilemma: an introduction to statistical mechanics via the Kac ring. *SIAM Review*, 51(3):613–635, 2009.

[6] E. Halley. Observations of the late Total Eclipse of the Sun on the 22d of April last past, made before the Royal Society at their House in Crane-Court in Fleet-street, London. *Philosophical Transactions of the Royal Society of London*, 29(343):245–262, 1715.

[7] G. Heine. Euler and the flattening of the earth. *Math Horizons*, 21(1):25–29, 2013.

[8] National Imagery and Mapping Agency. Department of defense world geodetic system 1984: Its definition and relationships with local geodetic systems. Technical Report TR8350.2, NIMA, St. Louis, MO, 2000.

[9] M. Kac. Some remarks on the use of probability in classical statistical mechanics. *Bulletin de la Classe des Sciences, Académie Royale de Belgique*, 42:356–361, 1956.

[10] W. G. Lambert. A part of the ritual for the substitute king. *Archiv für Orientforschung*, 18:109–112, 1957.

[11] J. J. Loschmidt. Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft.

Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Classe, 73:128–142, 1876.

[12] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. Princeton University Press, Princeton, USA and Oxford, UK, 2nd edition, 2017.

[13] I. Newton. *Philosophiæ Naturalis Principia Mathematica*. Royal Society, 1726. Translated by I. B. Cohen, A. Whitman and J. Budenz, University of California Press, 1999.

[14] H. Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–270, 1890.

[15] H. Poincaré. *Papers on Topology: Analysis Situs and Its Five Supplements*, volume 37 of *History of Mathematics*. American Mathematical Society and London Mathematical Society, Providence, RI and London, 2010. Translated and edited by John Stillwell.

[16] A. Prša, P. Harmanec, G. Torres, E. Mamajek, M. Asplund, N. Capitaine, J. Christensen-Dalsgaard, É. Depagne, M. Haberreiter, S. Hekker, J. Hilton, G. Kopp, V. Kostov, D. W. Kurtz, J. Laskar, B. D. Mason, E. F. Milone, M. Montgomery, M. Richards, W. Schmutz, J. Schou, and S. G. Stewart. Nominal values for selected solar and planetary quantities: IAU 2015 Resolution B3. *The Astronomical Journal*, 152(2):41, 2016.

[17] C. Rovelli. *The Order of Time*. Riverhead Books, New York, NY, 2018. Originally published in Italian as L'Ordine del Tempo, 2017.

[18] J. Stirling. *Methodus Differentialis: sive Tractatus de Summatione et Interpolatione Serierum Infinitarum*. G. Strahan, London, 1730.

[19] S. Urban and P. K. Seidelmann, editors. *Explanatory Supplement to the Astronomical Almanac*. University Science Books, 2012.

[20] Voltaire. Lettre 1476 à M. de Maupertuis. In T. Besterman, editor, *Correspondance*, volume 36, pages 102–103. Garnier, 1968.

[21] Αριστοτελης. *On Generation and Corruption*. Oxford University Press, Oxford, 335–323 B.C. Translated by H. H. Joachim, published in 1922.