# Chapter 1

# Prelude: Love, Hate, and Exponentials

## 1.1. Two sets of travelers

> A topologist is a mathematician who can't tell the difference between a donut and a coffee mug.

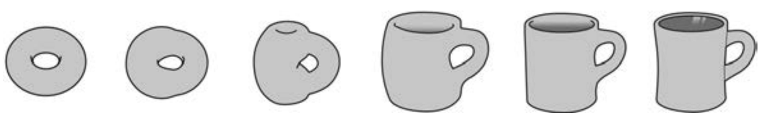

**Figure 1.1.** Transforming a donut into a coffee mug.

This well-known saying expresses the idea that *topology* studies those properties of "spaces" (we will have to say what we mean by that) which are unaffected by "continuous changes" (we will have to say what we mean by that also). Why might mathematicians be interested in such a thing? Doesn't it seem rather, well, *imprecise*?

Here's a story (freely adapted from Chapter 1 of [**3**], where it is attributed to N. N. Konstantinov) that hints at an answer.

> In a certain country there are two cities — call them Aberystwyth and Betws-y-Coed — and two

roads that join them: the "low road" and the "high road".

In A dwell two lovers, Maelon and Dwynwen, who must travel to B: M by the high road, and D by the low. So great is the force of their love that if at any instant they are separated by ten miles or more, they will surely die.

As well as a pair of lovers, our story contains a pair of sworn enemies, Llewelyn and John. As our story begins, L is in A, J is in B, and they must exchange places, L traveling from A to B via the high road while J travels from B to A via the low road. So great is the force of their hatred that if at any instant they are separated by ten miles or less, they will surely die.

Prove that tragedy is inevitable. At least two people will end up dead.

**Remark 1.1.1.** The point about this story is that we are given *no* specific information about the travels of D, M, L, and J: how fast they go, whether they halt on the journey, whether they speed up or slow down or even backtrack. Any mathematical tool effective enough to solve the problem must not care about these kind of "geometrical" specifics: must not care, in fact, about the difference between the donut and the coffee mug. It is wrong to suppose that topology, because it does not care about such distinctions, is somehow imprecise. On the contrary! Only a truly powerful theory can draw precise, specific conclusions from such unspecific initial data.

There are two components to solving the problem. The first is to set up a suitable graphical representation, which turns this picturesque story into a problem in topology. The second is to solve the resulting topological problem.

In the first step, we parameterize the problem by the unit square $S = [0,1] \times [0,1]$. A point $(x,y) \in S$ is thought of as describing the location of a *pair* of characters (either M and D, or L and J) along the high and low roads, respectively. So for instance the point $(0,0)$ represents "both characters are at A", $(1,1)$ represents "both

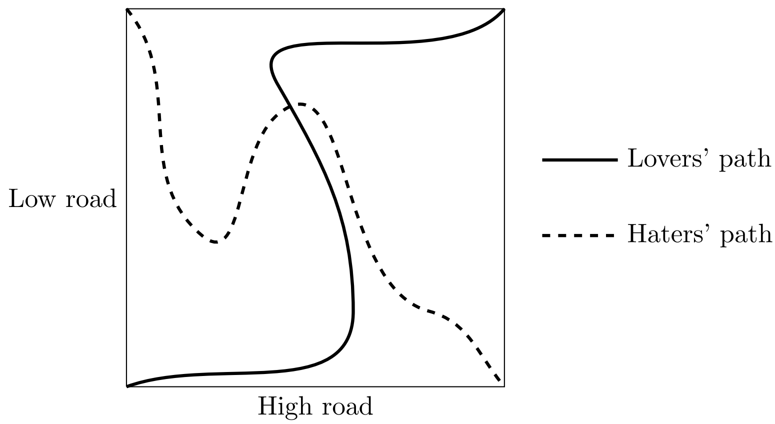


**Figure 1.2.** Parameterizing the lovers and haters problem.

characters are at B", $(0.4, 0.7)$ represents "the first character is 40 percent of the way along the high road from A to B and the second character is 70 percent of the way along the low road", and so on. The travels of a *pair* of characters along the high and low roads are now encoded in the movement of the *single* point $(x, y)$ through $S$.

Now the terms of the problem say that the path which describes the motion of the pair (M, D) must start at $(0, 0)$ and end at $(1, 1)$. And the path which describes the motion of (L, J) must start at $(0, 1)$ and end at $(1, 0)$. So (and this is the topological bit that we'll have to come back to), "obviously", the two paths have to cross (Figure 1.2).

Okay, what happens at a crossing point $(x_0, y_0)$? This represents a pair of points — one on the high road, one on the low — which are occupied (at different times) both by M and D and by L and J. If that pair of points is 10 miles or more apart, it spells doom for M and D; 10 miles or less, curtains for L and J. Either way, tragedy is inevitable, just as the problem says.

So here is the key topological fact that we have to prove.

**Theorem 1.1.2.** *Two continuous paths in the unit square $S$, one joining $(0, 0)$ to $(1, 1)$ and the other joining $(0, 1)$ to $(1, 0)$, must cross somewhere.*

Surprisingly (perhaps) this is not easy to prove. Let's look at one attempted proof and critique that.

*Attempted Proof #1:* Let the two paths be the graphs of continuous functions $f(x)$ and $g(x)$. Thus $f(0) = 0$, $f(1) = 1$, $g(0) = 1$, $g(1) = 0$. Therefore if we consider the function

$$h(x) = f(x) - g(x),$$

we have $h(0) = -1$ and $h(1) = 1$. By the intermediate value theorem, $h(x_0) = 0$ for some $x_0$. Then $f(x_0) = g(x_0) = y_0$, say, so the two paths cross at $(x_0, y_0)$. (?)

The trouble with this argument is that it *assumes* that our paths can be represented as the graphs of functions — in other words, that there is no "backtracking" in the $x$-direction. But nothing in the statement of the problem requires this, and there are many continuous paths which can*not* be represented as graphs, either in this way where $y$ is a function of $x$ or in the reverse way where $x$ is a function of $y$. In a sense, the "no backtracking assumption" has allowed us to reduce the 2-dimensional problem to a 1-dimensional one, which can then be solved using a 1-dimensional tool, the intermediate value theorem. Without making this assumption we are confronted with a situation which requires essentially 2-dimensional tools.

*Attempted Proof #2:* Consider the *loop* in the plane formed by traveling from $(0,0)$ to $(1,1)$ along the lovers' path and then returning via the circular arc

$$t \mapsto (\cos t, 1 + \sin t), \quad 0 \leqslant t \leqslant 3\pi/2.$$

The point $(1,0)$ is clearly *outside* the loop and $(0,1)$ is *inside* it, so any path — such as the haters' path — from one to the other must cross the loop somewhere.

This argument is correct, but the notions "outside" and "inside" have to be made precise, and this isn't as easy as it may seem — especially if we consider paths that may cross themselves or *self-intersect*. What we will end up doing is defining whether a point $p$ is "outside" or "inside" a loop $\gamma$ by counting how many times $\gamma$ "winds around" $p$. Of course that simply shifts the question to explaining what we mean by "winds around", but this is a question to which it is possible to give a precise answer.

## 1.2. Winding around

Counting revolutions or "windings" is an important and familiar notion, in everyday life as well as in mathematics and science. We measure our days by revolutions of the earth, our months by revolutions of the moon around the earth, and our years by revolutions of the earth around the sun. Computing orbits and their periods is the beginning of the theory of gravitation. The metaphor of life as a "wheel of fortune" resonates through cultures ancient and modern. Jerry Garcia sung in 1972

> The wheel is turning and you can't slow down
> You can't let go and you can't hold on
> You can't go back and you can't stand still
> If the thunder don't get you then the lightning will.

How many times does the wheel turn? If we stipulate that at the end of the story the wheel is in the same position as it was in the beginning, then the answer is an *integer* — a whole number of turns, positive (by convention) for counterclockwise revolutions and negative for clockwise ones. This integer is the *winding number*, the central concept of this book. Notice that to compute it, you have to know the *whole continuous story* of the motion of the wheel: it is not enough to look at snapshots of its beginning and end. In other words, the question "how many times around" is at root a *topological* one, and its answer, the winding number, is a topological notion.

We've already seen in the previous section an example of how any kind of *continuous motion* can be conceptualized as a *path* in a suitable abstract space, that is, a mapping from the unit interval into that space. Similarly, a continuous motion that *returns to its starting point* can be conceptualized as a *loop*, that is, a mapping of the unit circle into a space. The winding number provides a way to classify and distinguish such loops. As we hinted above, it is the key to such intuitively natural notions as the distinction between the "inside" and the "outside" of a closed curve in the plane.

Many students will first meet the winding number in a course on complex analysis, rather than topology. This is because of the beautiful way the winding number enters into Cauchy's residue theorem, which allows one to compute certain integrals of a function $f(z)$ in terms of the behavior of $f$ at certain special points, its so-called *poles* or singularities, and the winding numbers of loops around these singularities. That powerful subject is not emphasized here, however (in particular, one does not need any prior acquaintance with complex analysis in order to read this book). Why? Because important as complex analysis is (with its applications throughout mathematics, physics, and engineering), the notion of winding number turns out to have ramifications far beyond even that field. In fact, it's not really too much of a stretch to see the winding number as the golden cord which guides the student through the labyrinth of classical mathematics: connecting algebra and analysis, potential theory and cohomology, complex numbers and just about everything.

In this book, we will look at some of the many ways that winding numbers show up in mathematics. The settings are quite diverse: topology, geometry, functional analysis, complex analysis, algebraic systems, and even Lie groups. However, underneath it all is a simple idea: *winding around.*

Let's get started.

## 1.3. The most important function in mathematics

We'll begin by renewing our acquaintance with a familiar object — the function $e^x$ — from the viewpoint of the complex plane. As Euler discovered, the exponential and trigonometric functions are closely related in the complex domain, and in particular the exponential function can be used to describe the unit circle in $\mathbb{C}$. It should therefore be no surprise that exponentials are going to be closely involved in our discussion of the winding number, which is all about continuous travel on the unit circle.

The *exponential function* $\exp(z)$, or $e^z$, is defined by

$$\exp(z) = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots .$$

Rudin [**34**] begins his classic book *Real and Complex Analysis* with the statement "This is the most important function in mathematics." Before we can start looking at its properties, though, we need to remind ourselves what kind of thing $z$ is here.

Remember that a *complex number* is a formal expression of the sort

$$z = x + yi$$

where $x$ and $y$ are real numbers and $i^2 = -1$. (We call $x$ the *real part* of $z$ and $y$ the *imaginary part*, and we use the notation $x = \operatorname{Re} z$, $y = \operatorname{Im} z$.) We'll think of $x + yi$ as represented by the point $(x, y)$ of the plane (sometimes called the *complex plane* or the *Argand diagram* in this context).

There is no problem in adding, subtracting, or multiplying complex numbers by the usual rules. However, the following is a nontrivial fact.

**Theorem 1.3.1.** *The complex numbers form a field; i.e., every nonzero complex number has a multiplicative inverse.*

**Proof.** We write an explicit formula for the inverse. If $z = x+yi$ is a complex number, then its *absolute value* or *modulus* $|z|$ is the positive real number defined by

$$|z|^2 = x^2 + y^2.$$

The *complex conjugate* of $z$ is

$$\bar{z} = x - yi.$$

One computes

$$z\bar{z} = x^2 + y^2 = |z|^2.$$

Thus, if $z \neq 0$, one has $|z| > 0$ and

$$\frac{\bar{z}}{|z|^2} = \frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2} i$$

is the multiplicative inverse of $z$. □

**Remark 1.3.2.** The field $\mathbb{C}$ may be considered as a *vector space* over $\mathbb{R}$ (see Appendix A to review the theory of vector spaces and linear algebra). When so considered, $\mathbb{C}$ is 2-dimensional: any $\mathbb{R}$-basis has two elements (the canonical example is of course the basis consisting of 1 and $i$). The possible $\mathbb{R}$-bases $(z, w)$ of $\mathbb{C}$ fall into two classes, *right-handed* like $(1, i)$ and *left-handed* like $(1, -i)$. Formally we may say that $(z, w)$ is a *right-handed basis* if $\operatorname{Im}(\bar{z}w) > 0$ and a *left-handed basis* if $\operatorname{Im}(\bar{z}w) < 0$. (If $\operatorname{Im}(\bar{z}w) = 0$, then $z$ and $w$ do not form a basis.)

The *exponential series*

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

converges for all values of $z$ and defines a differentiable function on the whole complex plane (such a function is called an *entire function*). We also use the notation $e^z$ for this function.

By term-by-term multiplication and differentiation one verifies (treat these as exercises)

(a) addition law: $e^{z+w} = e^z e^w$;

(b) differentiation law: the function $z \mapsto e^z$ is its own derivative.

The sine and cosine functions are defined in terms of the exponential by

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!},$$

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}.$$

The exponential, sine, and cosine functions are real-valued for real arguments, and we have

$$e^{iz} = \cos z + i \sin z$$

for all $z$. Moreover, since the power series for the exponential function has real coefficients, $e^{\bar{z}} = \overline{e^z}$. It follows that

$$|e^z|^2 = e^z \overline{e^z} = e^{z+\bar{z}} = e^{2 \operatorname{Re} z}$$

so $|e^z| = e^{\operatorname{Re} z}$, for all complex numbers $z$. In particular, $|e^{iy}| = 1$ for all real $y$.

The addition law for the exponential function yields the corresponding laws for sine and cosine,

$$\sin(z+w) = \sin z \cos w + \cos z \sin w,$$
$$\cos(z+w) = \cos z \cos w - \sin z \sin w.$$

In particular $\sin^2 z + \cos^2 z = 1$ — the special case $w = -z$ of the second identity. One sees by computation that cos has a positive real zero; define $\pi$ by letting $\pi/2$ be the smallest positive real zero of cos. We have $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$. The identities now give

$$\sin(z + \pi/2) = \cos(z), \quad \cos(z + \pi/2) = -\sin(z).$$

Iterating these we find that cos and sin are $2\pi$-periodic, so the exponential function is $2\pi i$-periodic. In particular we get the famous formulae

$$e^{2\pi i} = 1, \quad e^{\pi i} = -1.$$

**Remark 1.3.3.** It's often useful to represent a complex number $z = x + iy$ in polar coordinate form as

$$z = re^{i\theta} = r\cos\theta + ir\sin\theta.$$

Here $r = |z|$, and $\theta$ is a "residue class modulo $2\pi$" (an equivalence class of real numbers, two numbers being considered as equivalent if they differ by a multiple of $2\pi$). One calls $r$ the *modulus* of $z$ and $\theta$ the *argument*. In polar coordinates the law for multiplication of complex numbers takes the simple form

$$(r_1 e^{i\theta_1})(r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1+\theta_2)};$$

you multiply the moduli and add the arguments.

The identity $e^z \cdot e^{-z} = 1$ shows that the exponential function never takes the value 0. However, it *can take any other value*. Indeed if $w = r(\cos\theta + i\sin\theta)$ is a nonzero complex number written in polar form, then $z = \log(r) + i\theta$ has $e^z = w$. There are of course infinitely many such $z$, differing by integer multiples of $2\pi i$. The really interesting story begins when we ask how these infinitely many possibilities for the preimage of $w$ fit together as *$w$ varies continuously* in $\mathbb{C} \setminus \{0\}$. Let's start to get a grip on this by considering only a limited range of values of $w$, those that lie on the unit circle ($|w| = 1$).

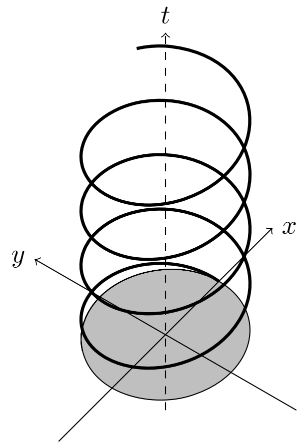


**Figure 1.3.** The exponential map illustrated — from the picture, you can guess that "winding around" will be involved somehow. This is the graph of $x + iy = e^{it}$.

**Lemma 1.3.4.** *The complex number* $w = \exp(z)$ *lies on the unit circle if and only if* $z$ *is* purely imaginary*, which is to say* $z = it$ *for some* $t \in \mathbb{R}$*.*

**Proof.** This follows from the formula $|\exp(z)| = e^{\mathrm{Re}\, z}$ which we observed before: $|\exp(z)| = 1$ if and only if $\mathrm{Re}\, z = 0$, which is to say that $z$ is purely imaginary. □

For $t \in \mathbb{R}$ we have Euler's formula

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

As $t$ moves along the real axis, the point $w = \exp(it)$ rotates with unit speed around the circle. (If you like, think of $t$ as time, measured in suitable units, and $w$ as the position of the tip of the minute hand of a clock whose center is at the origin.[1]) Many different $t$-values correspond to the same $w$, just as many different time-values all have the minute hand pointing to 6. We can think of the exponential function

[1] Unfortunately for this illustration, mathematical convention is that the positive direction of rotation is *counterclockwise*, so you should think of the clock as running backwards.

as "wrapping up" the imaginary axis into a spiral that projects to the unit circle, as shown in Figure 1.3.

It is fundamentally important that, although each point $w$ on the unit circle corresponds to many different $t$-values, there is no way to choose those $t$-values for the whole unit circle in a "continuous" manner. More precisely,

**Lemma 1.3.5.** *There is no continuous function* $\theta\colon S^1 \to \mathbb{R}$ *such that* $w = \exp(i\theta(w))$ *for all* $w \in S^1$*. In fact, there is no continuous function* $s\colon S^1 \to S^1$ *such that* $s(w)^2 = w$ *for all* $w \in S^1$*.*

**Proof.** The second statement clearly implies the first since if we could find a function $\theta$ having the required properties, we could then define

$$s(w) = \exp(i\theta(w)/2)$$

and we would have $s(w)^2 = w$.

Suppose then that a continuous function $s$ exists having $s(w)^2 = w$. Consider the function

$$u(t) = s(e^{it})s(e^{-it}), \quad t \in \mathbb{R}.$$

This is a continuous function on $\mathbb{R}$. We have

$$u(t)^2 = s(e^{it})^2 s(e^{-it})^2 = e^{it}e^{-it} = 1.$$

Thus $u(t) = \pm 1$ for each $t$. A continuous integer-valued function on $\mathbb{R}$ is constant, so $u$ is constant. But then

$$-1 = s(-1)^2 = u(\pi) = u(0) = s(0)^2 = 1,$$

which is an obvious contradiction[2]. □

**Remark 1.3.6.** It's interesting to contrast the situation for square roots of *complex* numbers, revealed by Lemma 1.3.5, with the corresponding situation for *real* numbers. When we look in the real field, we find two problems: an existence problem (some numbers, the negative ones, don't have any square roots) and a uniqueness problem (other numbers, the positive ones, have more than one, so the symbol $\sqrt{x}$ can be ambiguous). In the real case it's easy to resolve the uniqueness problem by executive order: we just decree, as is done

[2]This argument is adapted from Beardon's book [**8**].

in high school algebra, that for $x > 0$, the symbol $\sqrt{x}$ should stand for the *positive* square root of $x$. Lemma 1.3.5 tells us that the introduction of *complex* numbers, which fixes the existence problem (every complex number has a square root), at the same time makes it impossible to come up with *any* sort of "executive order" which resolves the uniqueness problem (the ambiguity of square roots) in a reasonable (read: continuous) way.

It follows from Lemma 1.3.5 that there is no "complex logarithm" function $\ell$ defined and continuous on all of $\mathbb{C} \setminus \{0\}$ and such that $\exp(\ell(z)) = z$. Functions with this property can, however, be found on some smaller domains. Here is an important example.

**Lemma 1.3.7.** *Let $S = \mathbb{C}\setminus\mathbb{R}^-$ be the complex plane with the negative real axis removed (this is sometimes called a "slit plane"). There exists a continuous function $\ell\colon S \to \mathbb{C}$ such that $\exp(\ell(z)) = z$ for all $z \in S$.*

**Proof.** Each $z \in S$ has a unique polar coordinate representation

$$z = re^{i\theta} = r\cos\theta + ir\sin\theta, \quad -\pi < \theta < \pi,$$

and the polar coordinates $r, \theta$ depend continuously on $z \in S$. Put

$$\ell(z) = \log(r) + i\theta,$$

where log is the usual natural logarithm for positive real numbers. □

**Remark 1.3.8.** A function $\ell$ having the property asserted by the lemma is called a *branch of the logarithm* defined on the slit plane $\mathbb{C} \setminus \mathbb{R}^-$. Notice that such branches are not unique: if $z \mapsto \ell(z)$ has the property of the lemma, then so does $z \mapsto \ell(z) + 2k\pi i$ for any integer $k$. Later, we will see that this integer ambiguity is related to the winding number in a simple way.

## 1.4. Exercises

**Exercise 1.4.1.** Calculate the quotient $(3 + 2i)/(1 - 2i)$. Find two complex roots of the quadratic equation

$$2z^2 - 3z - 5i = 0.$$

**Exercise 1.4.2.** Show that the modulus obeys the *triangle inequality*

$$|z \pm w| \leqslant |z| + |w|.$$

This allows us to make the complex plane into a *metric space* (see later, Definition B.1.1) and thus to introduce topological notions such as open and closed sets, continuity, etc.

**Exercise 1.4.3.** Let $a = 1 + i$ and $b = \sqrt{3} - i$. Express each of the complex numbers

$$a + b, \quad a - b, \quad ab, \quad a/b$$

in the form $x + yi$ and in the form $re^{i\theta}$, simplifying your answers as much as possible.

**Exercise 1.4.4.** Let $z = e^{i\theta}$ where $\theta = 2\pi/5$. Prove that $1 + z + z^2 + z^3 + z^4 = 0$. By considering the real part of this expression prove that

$$\cos\theta = \frac{-1 + \sqrt{5}}{4}.$$

**Exercise 1.4.5.** (a) Show that the mapping $z \mapsto 1/z$ sends the circle $|z - 1| = 1$ (in the complex plane) into a straight line.

(b) Let $A$, $B$, $C$, and $D$ be four points on a circle in the (Euclidean) plane, and let the symbol $d(X, Y)$ denote the Euclidean distance between two points $X$ and $Y$. Let $p = d(A, B)d(C, D)$, let $q = d(A, C)d(B, D)$, and let $r = d(A, D)d(B, C)$. Show that one of $p, q, r$ is equal to the sum of the other two. (This result is due to Ptolemy of Alexandria, nearly 2,000 years ago. To prove it using complex numbers, take the circle to be the one in the first part of the question, and take $A$ to be the origin. Use the transformation $z \mapsto 1/z$ to relate the theorem to the distances between points on a straight line.)

**Exercise 1.4.6.** In the 1840s, William Rowan Hamilton spent much effort trying to find a 3-dimensional field of "hypercomplex" numbers, i.e., of symbols of the form $x + yi + zj$, with $x, y, z \in \mathbb{R}$, which can be added, subtracted, multiplied, and divided in the same way that complex numbers can. Show that his quest was hopeless: no matter how we define $i^2$ and $j^2$, we will not obtain a 3-dimensional system of the desired sort. (HINT: Use linear algebra. Let $V$ denote the proposed system. Fix a specific nonreal element $\alpha \in V$ and let $m_\alpha$ denote the operation of multiplication by $\alpha$, which is an $\mathbb{R}$-linear transformation from $V$ to $V$. This transformation must have a real eigenvalue because $V$ is odd-dimensional. From this, deduce that one can find two nonzero elements of $V$ whose product is zero, which contradicts the desired existence of division in $V$.)