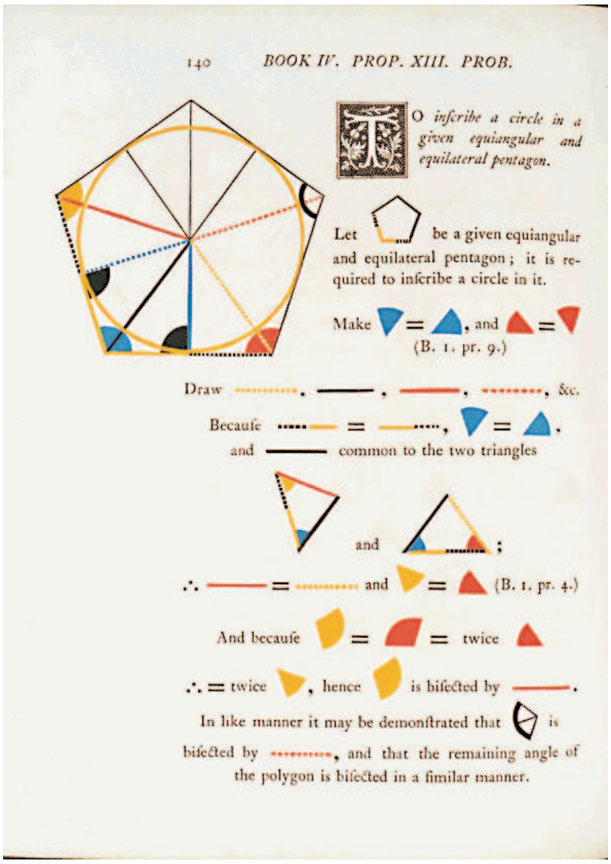
Dimensions – conventional and not so conventional

ACC Coolen
Department of Mathematics, King's College London



140 BOOK IV. PROP. XIII. PROB.

TO inſcribe a circle in a given equiangular and equilateral pentagon.

Let ⬠ be a given equiangular and equilateral pentagon; it is required to inſcribe a circle in it.

Make ▼ = ▲, and ▲ = ▼
(B. 1. pr. 9.)

Draw ⋯⋯, ——, ——, ⋯⋯, &c.

Becauſe ⋯—— = ——⋯, ▼ = ▲,
and —— common to the two triangles

▽ and △ ;

∴ —— = ⋯⋯ and ▶ = ▲ (B. 1. pr. 4.)

And becauſe ◗ = ◗ = twice ▲

∴ = twice ▶, hence ◗ is biſected by ——.

In like manner it may be demonſtrated that ◁ is biſected by ⋯⋯, and that the remaining angle of the polygon is biſected in a ſimilar manner.

Dimensions D=0,1,2,3 – the world of Euclid

From antiquity until roughly 1850, geometry was limited to the study of objects which could be visualized and built: points, lines and solid shapes. The books of Euclid's 'Elements' (the figure above shows a page from a 19th century edition) were the 'gold standard', which more or less continued to be taught with only modest additions and changes in notation. There was little need for definitions of dimension more sophisticated than the following intuitive (so-called topological) one:

> *A set has dimension D if we need D independent variables (coordinates) to describe the neighbourhood of any point of the set.*

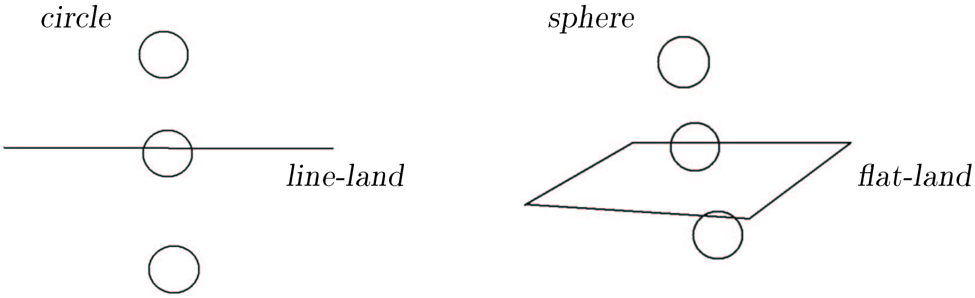


Figure 1. Left picture: a 2D object (a circle) visits line-land. The 'line-landers', who cannot visualize the second dimension, would experience this phenomenon as the appearance out of nothing of a point (the moment the circle touches the line), which subsequently splits in two, followed by the two points merging again and disappearing. Right picture: a 3D object (a sphere) visits flat-land. The 'flat-landers', although more sophisticated, cannot visualize the third dimension and experience this phenomenon as the miraculous appearance of a point (the moment the sphere touches the plane), which grows into a circle, followed by this circle shrinking again and vanishing.

Points were sets of dimension zero, curves sets of dimension one, surfaces had dimension two, and solid shapes (volumes) had dimension three. This exhausted the possibilities.

Dimensions D=4,5,... – hyperplanes, hyperspheres, hypercones, etc

In the nineteenth century mathematicians became more adventurous. Complex numbers, for example, had shown that it pays to define new mathematical concepts even if they are hard to visualize, provided the definitions are logically consistent. One began to consider the notion of self-consistent geometries other than that of Euclid. One obvious generalization of traditional geometry was the introduction of objects with integer dimension larger than three. The Victorians were greatly charmed by the notion of a fourth spatial dimension; this was, after all, the age of inventions. One popular 1884 book, 'Flatland' by A. Square (see figure 1) tried to explain the concept of four dimensions by analogy. Try to explain a second dimension to someone who knows only one (an inhabitant of 'line-land', who can image only one position coordinate x), or a third dimension to someone who can visualize only two (the more sophisticated inhabitant of 'flat-land', who thinks in terms of two position coordinates x and y); see figure 1. One can now imagine how we would experience in a world of three coordinates (x, y, z) an encounter with a four dimensional object, such as the 4D sphere $x^2+y^2+z^2+u^2=1$ (assuming the existence a fourth spatial coordinate u). This sphere could travel from one 3D world to another, at any stage intersecting an infinite number of these worlds. When intersecting 'our' 3D world, we would see a 3D sphere appearing out of thin air, growing to full size, and vanishing again in front of our eyes.

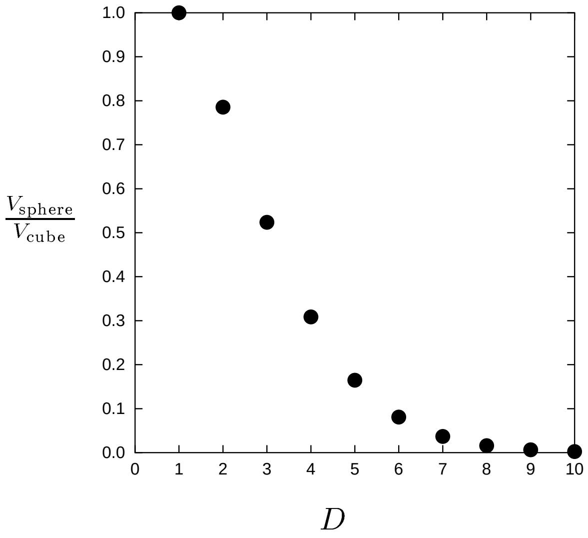


Figure 2. The ratio of the volume V_{sphere} of a (hyper-)sphere over the volume V_{cube} of a (hyper-)cube, both of the same radius, in D dimensions. As D becomes larger, virtually all the volume of the (hyper-)cube is found to be concentrated in its corners.

Visualizations can be helpful at times, but in mathematics we do not actually need them. We can simply proceed to calculate mathematical properties of objects with arbitrary dimension $D > 3$, and gain intuition that way. For example, it is straightforward to calculate the volume of (hyper-)cubes and (hyper-)spheres in any number D of dimensions. For hyper-cubes of diameter $2L$ one simply finds $V_{\text{cube}} = (2L)^D$. For hyper-spheres of radius L the calculation is only slightly more complicated, resulting in the following interesting formula for the ratio $V_{\text{sphere}}/V_{\text{cube}}$:

$$D \text{ even}: \qquad \frac{V_{\text{sphere}}}{V_{\text{cube}}} = \frac{(\pi/4)^{\frac{1}{2}D}}{\frac{1}{2}D(\frac{1}{2}D-1)(\frac{1}{2}D-2)\ldots 1}$$

$$D \text{ odd}: \qquad \frac{V_{\text{sphere}}}{V_{\text{cube}}} = \frac{(\pi/4)^{\frac{1}{2}D}}{\sqrt{\pi}.\frac{1}{2}D(\frac{1}{2}D-1)(\frac{1}{2}D-2)\ldots \frac{1}{2}}$$

These ratios are shown in figure 2. For small D we recover the familiar facts $V_{\text{sphere}} = \pi L^2$ in $D = 2$ and $V_{\text{sphere}} = \frac{4}{3}\pi L^3$ for $D = 3$. We see, without needing images, that for larger D the volume of a (hyper-)cube will be increasingly concentrated in its 2^D corners.

All this still leaves the question of whether results such as the above are mere ‘recreational mathematics’. The answer is no. The notion of dimensions $D > 3$ was highly fruitful, especially in the theory of functions. To see why this is the case, let us define the following family of simple real-valued functions :

$$x \in \mathbb{R}: \qquad f(x) = a + bx + cx^2 + dx^3 + ex^4$$

with coefficients $a, b, c, d, e \in \mathbb{R}$. For $d = e = 0$ we return to the family of parabolas. The above family is a set of objects (functions) in which we need exactly five real numbers to characterize each member. Thus, we assign to each such function five coordinates

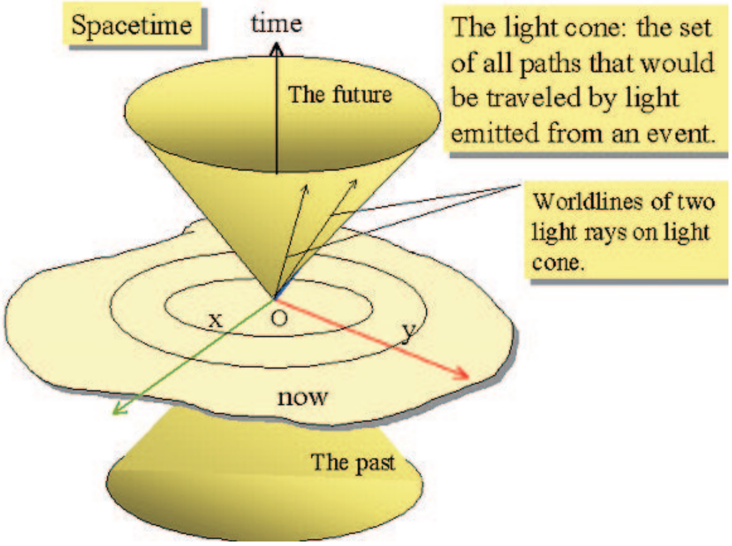


Figure 3. Since Einstein, physical theories are formulated in 4-dimensional space-time (of which the above is a 3-dimensional caricature). The 'life' of a particle is described by a curve through this space, with physically possible and physically impossible curves separated by the 'light-cone' (this reflects the impossibility of material objects to exceed or even reach the speed of light).

(a, b, c, d, e), and may interpret (a, b, c, d, e) as a point in a five-dimensional space. It will be clear that there is no limit on the number of coordinates which might be required to build more complicated functions (for instance, just add higher powers of x), and thus in function theory one usually does calculations in spaces of dimension $D = \infty$.

Soon after 1900 it became clear that our naive notion that the physical world we live in has three immutable spatial dimensions also needed revision. Einstein's special relativity theory (1905, dealing with physics at high velocities) and general relativity (1916, dealing with physics at high accelerations or under strong gravitational forces) showed that the division into three spatial and one time coordinate is not absolute (clocks may and will go at different speeds for different observers), and physics became formulated in terms of geometry in so-called 4-dimensional space-time (see figure 3).

Physics underwent a further revolution with the birth of quantum theory in 1925 (physics at the level of elementary particles). Since then, theoretical physics has been mostly concerned with 'unification', reconciling relativity theory with quantum theory. This proved possible for special relativity (high velocities or energies), leading to so-called quantum field theory. Reconciling general relativity (gravitation) with quantum theory, however, proved more difficult, and is still at the centre of research. The presently accepted best candidate for a unified physical theory ('M-Theory') is based on a representation of elementary particle as tiny vibrating 'strings' (see figure 4). Such theories, however, force us to accept that we must live in a world with $D = 11$ space-time coordinates (although most of these are invisible to us).

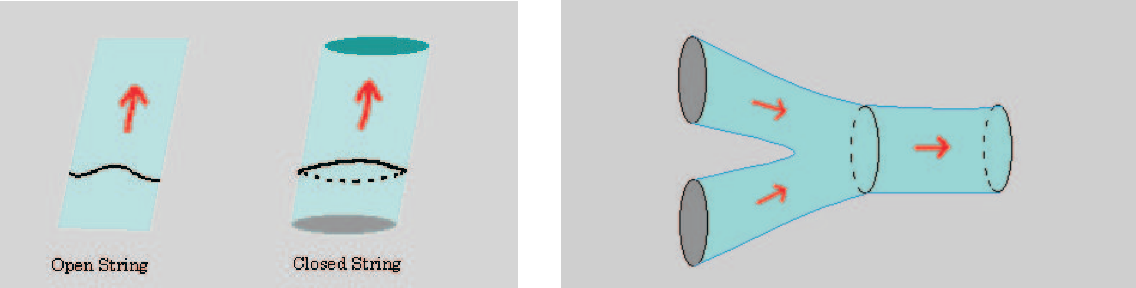


Figure 4. Strings in elementary particle physics. In contrast to point particles, the life of a string is not represented by a curve in space-time: strings sweep out surfaces as they move. Open strings have different ways of vibrating than closed ones, leading to differences in physical properties (left), and strings can interact and merge (right).

Proper mathematical definitions of dimension

Precision in mathematical definitions came to be regarded as vital by the mathematicians of the 19th century. For instance, although limits play a central role in the calculus developed by Newton and Leibniz in the 17th century, their proper definition had to wait for more than hundred years. Indeed, sharpening of the definition of dimension was prompted by the emergence of examples of nasty geometrical objects which are generated by a limiting process.

Many definitions of dimension have been proposed over the years, all meeting the constraint that they reduce to the traditional one when we return to points, curves, planes and volumes. One has, for instance, Lebesgue Covering Dimension, Brouwer Dimension, Hausdorff Dimension (or Hausdorff-Besicovitch Dimension), Minkowski-Bouligand Dimension, Fractal Dimension (or Capacity Dimension), Box-counting Dimension, Information Dimension, and more. Most of these definitions are based on the following idea: if we cover all elements of a set by tiny marbles, how many marbles do we need ? And, crucially: how does this number change if we make our marbles smaller and smaller ? Let us illustrate this for the so-called Box-Counting Dimension, see figure 5. Here we cover our set with tiny boxes of diameter ϵ (i.e. every point in the set has to be inside at least one box), and we define N_ϵ as the *minimum* number of boxes needed to achieve covering. We expect N_ϵ to increase as ϵ decreases. Working out N_ϵ for the simple cases of points ($D = 0$), lines of length L ($D = 1$), rectangular planes of area S ($D = 2$), and cubes of volume V ($D = 2$), reveals a pattern:

$$\begin{array}{lllll}
\text{point,} & D = 0: & N_\epsilon = \epsilon^0 & \Rightarrow & \log(N_\epsilon) = 0 \\
\text{line,} & D = 1: & N_\epsilon = L/\epsilon^1 & \Rightarrow & \log(N_\epsilon) = 1.[\log(L) - \log(\epsilon)] \\
\text{plane,} & D = 2: & N_\epsilon = S/\epsilon^2 & \Rightarrow & \log(N_\epsilon) = 2.[\log(S^{1/2}) - \log(\epsilon)] \\
\text{cube,} & D = 3: & N_\epsilon = V/\epsilon^3 & \Rightarrow & \log(N_\epsilon) = 3.[\log(V^{1/3}) - \log(\epsilon)]
\end{array}$$

In order to deal also with curved sets, we must require $\epsilon \to 0$. This, however, will imply $\log(\epsilon) \to -\infty$ in the above expressions for $\log(N_\epsilon)$, so we must first divide the

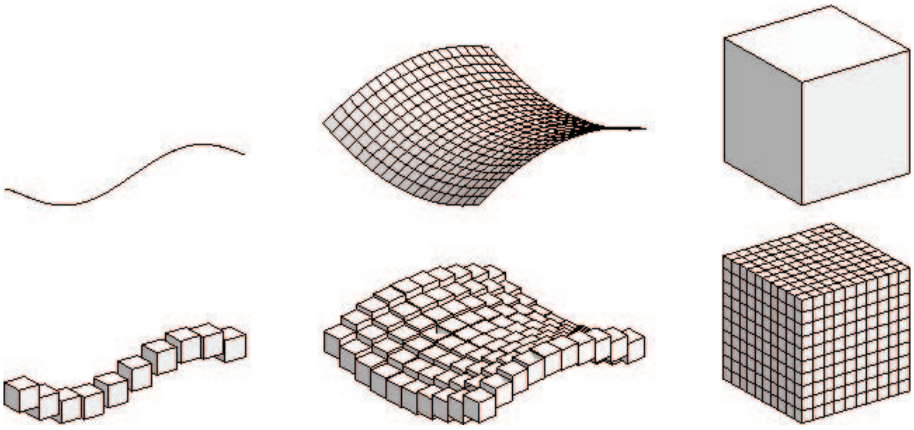


Figure 5. The idea behind box-counting dimension. If a set is covered by tiny boxes, then the minimum number of boxes required scales with the diameter of the set (equivalently: with the diameter of our boxes) in a way which depends crucially on the dimension of the object covered. This dependence can be used to define the dimension of the covered set. Most definitions of dimension are based on similar ideas.

latter by $\log(\epsilon)$. We then observe that *en passant* we lose the unwanted example specific properties L, S, and V, and we end up with the appealing general expression

$$D = -\lim_{\epsilon\to 0} \frac{\log(N_\epsilon)}{\log(\epsilon)}$$

This definition of dimension reduces to the traditional one for conventional geometric objects, but can be used more generally, and need not be limited to integer values.

Non-integer dimensions D

We can approach the problem of assigning meaning to non-integer dimensions in two ways. The first is to simply carry out mathematical calculations and allow D to take non-integer values wherever possible. The second is to work with mathematical definitions such as given above, and try to construct sets for which D comes out non-integer.

The most well-known example of the former is a procedure known as 'dimensional regularization' in quantum field theory of elementary particles, dating from the middle of the 20th century. This theory was plagued by infinities: many integrals occurring in the various equations were ill-defined, requiring messy repair mechanisms to make the various unwanted infinities cancel. The simplest such so-called Feynman integrals, for a theory describing interacting particles in D space-time dimensions, is

$$I = \int_0^\infty dx\, \frac{x^{D-1}}{(1+x^2)^2}$$

This integral can become infinite either because the argument explodes near $x = 0$, or because the argument decays too slowly as $x \to \infty$. Let us briefly inspect these two possibilities. Close to $x = 0$ we have $x^{D-1}/(1+x^2)^2 \approx x^{D-1}$; integration of the latter

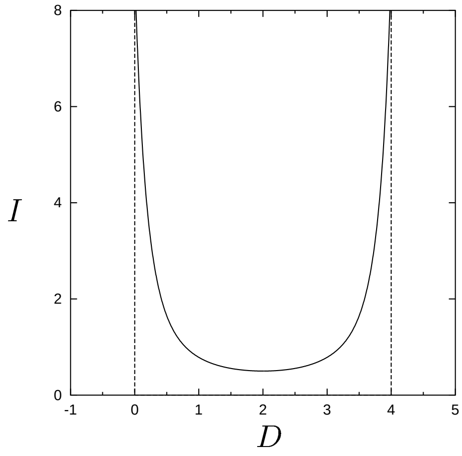


Figure 6. Example of a Feynman integral (see main text) in quantum field theory, for D space-time dimensions. It diverges at $D = 0$ and $D = 4$, prompting the introduction of calculations in $D = 4 - \epsilon$ dimensions, followed by $\epsilon \to 0$ in the end.

from $x = 0$ upwards gives infinity if $D \leq 0$. For large x we have $x^{D-1}/(1+x^2)^2 \approx x^{D-5}$; integration of the latter from any finite positive x to $x = \infty$ gives infinity if $D \geq 4$. Thus the integral I appears to be finite only for $0 < D < 4$. Evaluation of I on a computer gives figure 6, showing the divergencies at $D \downarrow 0$ and $D \uparrow 4$. Since the theory was intended to be used for $D = 4$ space-time dimensions, there was a serious problem.

One way out of trouble was to carry out *all* calculations in $D = 4 - \epsilon$ dimensions, with $0 < \epsilon < 1$. Now the offending integrals were finite, and the theory would make sense. Moreover, those terms which would explode as $\epsilon \to 0$ could be carefully made to cancel; at the end one could take the limit $\epsilon \to 0$ and still retain meaningful experimentally testable predictions. Here it was found both possible and fruitful to do mathematical calculations in a non-integer number of dimensions, again without visualizations.

Let us now turn to the construction of sets with non-integer dimension. If we allow objects to be the result of a limiting process, matters can become tricky. Take a simple function like $f(x) = \sin(kx)$ for $x \in [0, 2]$. For increasing values of k the function oscillates faster and faster, and its graph will start resembling a black square. For finite k the set of points on the paper is clearly one-dimensional, but what about $k \to \infty$? Around 1900 such questions were very fashionable, and the first so-called self-similar sets were designed, specifically to test concepts and definitions. I will discuss two of these: the Sierpinski triangle (1916) and the Von Koch curve (1904).

The Sierpinski triangle results from the iteration process described in the caption of figure 7. If we iterate an *infinite number* of times, every sub-triangle in the set will be identical in shape to the whole triangle ('self-similarity'). Furthermore, we see that the total remaining black surface $S(n)$ after n iterations will be given by the expression $S(n) = (\frac{3}{4})^n S(0)$. Since $lim_{n\to\infty} S(n) = 0$, we end up with an object with finite extension in two orthogonal directions, but with zero surface.

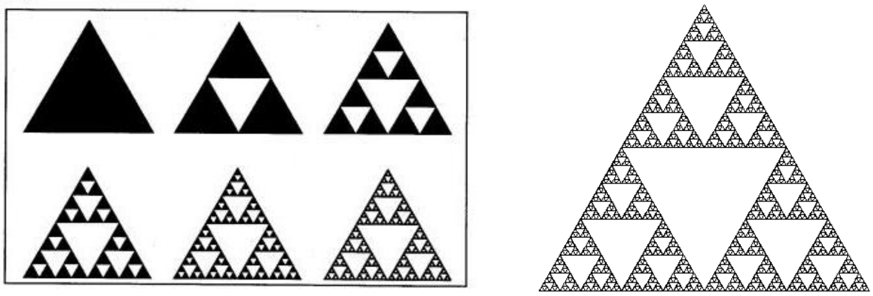


Figure 7. Left: construction of the Sierpinski triangle. One starts with a simple black triangle, with equal sides. Then, at every stage of the iteration, one cuts *each* black triangle into four similar triangles of equal size, and eliminates the middle one. This iteration is repeated an infinite number of times. Right: the final result.

step 0 *step 1* *step 2* *step 3*

Figure 8. Construction of the Von Koch curve. One starts with a simple line segment. Then, at every stage of the iteration, one cuts *each* line segment into three, and replaces the middle piece by two new connected segments (sticking outward), each of the same size as the two remaining original pieces. This is repeated an infinite number of times.

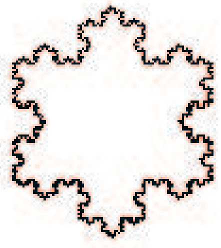


Figure 9. The Von Koch snowflake, obtained by connecting three Von Koch curves (see previous figure) in a triangle.

A second self-similar object is the Von Koch curve, again constructed by an iterative rule; see the caption of figure 8. Again, if we iterate an *infinite number* of times, every part of the set will be identical in shape to the whole set ('self-similarity'). Furthermore, we see that the length $L(n)$ of the curve after n iterations will be given by the expression $L(n) = (\frac{4}{3})^n L(0)$. Since $lim_{n\to\infty} L(n) = \infty$, we end up with a curve with finite extension in the plane, yet of infinite length. When used as a building block in larger structures, one can obtain further interesting objects, such as the Von Koch snowflake in figure 9.

I will now sketch the calculation of the box counting dimension introduced earlier, viz. $D = -\lim_{\epsilon\to 0}\log(N_\epsilon)/\log(\epsilon)$, for the Von Koch curve. We start with an initial line segment of length 1. After n iterations, the total curve length $L(n)$ and the resolution $d(n)$ (i.e. the length of the smallest segment in the curve) will be

$$\text{step}\ \ n: \qquad L(n) = (\frac{4}{3})^n, \qquad d(n) = (\frac{1}{3})^n$$

Boxes of size ϵ will thus be able to follow all details of the curve up until those corresponding to the step n with $d(n) \geq \epsilon$, i.e. to (modulo rounding off)

$$n = \log(\epsilon)/\log(1/3)$$

Curve parts with finer details will automatically be covered, so, modulo rounding off, the number of boxes needed will be the length of the curve at stage n divided by the size of the covering boxes:

$$\log(N_\epsilon) = \log(\frac{L(n)}{\epsilon}) = n\log(\frac{4}{3}) - \log(\epsilon) = \frac{\log(\epsilon).\log(4/3)}{\log(1/3)} - \log(\epsilon)$$

Hence

$$D = -\lim_{\epsilon\to 0}\frac{\log(N_\epsilon)}{\log(\epsilon)} = -\frac{\log(4/3)}{\log(1/3)} + 1 = \frac{\log(4)}{\log(3)}$$

Thus, the Von Koch curve is an object with so-called ‘fractal dimension’ (whose definition for the present class of set coincides with our box counting dimension) equal to $D = \log(4)/\log(3) \approx 1.26186$. It is somewhere between a line and a plane. Similarly one can calculate the fractal dimension of the Sierpinski triangle, which comes out as $D = \log(3)/\log(2) \approx 1.58496$ (again between a line and a plane, but closer to the plane than the Von Koch curve). Clearly, non-integer dimensions do occur in reality.

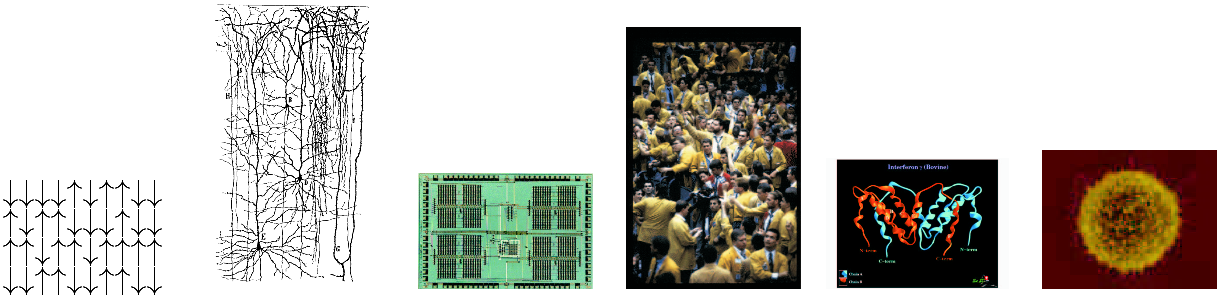
Negative dimensions D

Integer dimensions $D > 3$ and non-integer dimensions are both mathematically allowed and part of physical reality. Let us now take the final step and inspect the possibility of $D < 0$. Calculations involving a negative number of dimensions emerged only relatively recently. To appreciate their origin we will first need a detour.

The field of disordered many-particle theory deals with the analysis of large systems of structurally similar interacting microscopic elements, with noisy dynamics and (pseudo-)randomness in their mutual forces. There are many examples, see figure 10. The formalism with which to analyze many-particle systems is ‘statistical mechanics’. Since ±1870 we known that, under certain conditions, the probability to find such systems in a given microscopic state can be written in terms of the energy of this state:

$$\text{Prob[state]} = Z^{-1}e^{-E(\text{state})} \qquad E: \ \text{energy of a given state}$$

Z is a constant, needed to make the probabilities add up to one. Given the above, it turns out that all *macroscopic* mathematical laws characterizing equilibrium can be obtain from the so-called ‘free energy’ $f = -\log[\sum_{\text{states}} e^{-E}]$. In a nutshell, solving a



10^{24} magnets

$10^8 - 10^{12}$ neurons

$10^4 - 10^6$ processors

$10^2 - 10^5$ traders

$10^3 - 10^4$ amino-acids

$10^8 - 10^{12}$ lymphocytes

Figure 10. Examples of disordered many-particle problems in various disciplines. In physics one finds certain magnetic systems ('spin-glasses') consisting of some 10^{24} atomic magnets with nearly random mutual interactions, such that individual pairs of magnets can either prefer opposite (SN and NS) or identical orientations (SS and NN, in contrast to macroscopic magnets the latter is possible for atomic ones). Examples in biology are neural networks (networks of some $10^8 - 10^{12}$ interacting brain cells, or neurons), proteins (where some $10^3 - 10^4$ amino-acids interact in a complicated way to produce a specific 3D folded shape for the protein), or the immune system (with $10^8 - 10^{12}$ different interacting cell types). In computer science one has machine learning in parallel distributed systems (here the 'particles' are processors). Finally, in economics one finds models of markets where thousands of traders interact in complicated ways, leading to irregular time-series for prices, stocks or exchange rates.

model means calculating f. For homogeneous systems the energy E is a simple function of the state variables. For disordered systems, in contrast, it is complicated and messy, making it impossible to calculate f directly. In the latter case, the rescue comes from the observation that, for sufficiently large systems, many microscopic details become irrelevant, and the macroscopic laws (or the free energy which generates them) will depend only on the statistical properties of the forces in the system. Hence we may simply calculate the *average* of f over all realizations of the random forces.

Averaging the logarithm of a sum of exponentials is usually very hard, but a trick was devised to deal with this, giving birth to 'replica theory' (±1975). Let us denote the (many-particle) state by the symbol σ, and the realization of all mutual forces by ξ (drawn randomly according to probabilities $p(\xi)$). We wish to calculate

$$\langle f\rangle_{\rm forces} = \int d\xi\; p(\xi)\left\{-\log[\sum_\sigma e^{-E(\sigma,\xi)}]\right\}$$

The replica trick uses the identity $\langle \log Z\rangle = \lim_{n\to 0}\frac{1}{n}\log\langle Z^n\rangle$ to write the above as

$$\begin{aligned}\langle f\rangle_{\rm forces} &= -\lim_{n\to 0}\frac{1}{n}\log\int d\xi\; p(\xi)\left[\sum_\sigma e^{-E(\sigma,\xi)}\right]^n\\ &= -\lim_{n\to 0}\frac{1}{n}\log\int d\xi\; p(\xi)\sum_{\sigma_1}\ldots\sum_{\sigma_n} e^{-E(\sigma_1,\xi)-\ldots-E(\sigma_n,\xi)}\\ &= -\lim_{n\to 0}\frac{1}{n}\log[\sum_{\sigma_1\ldots\sigma_n} e^{-E(\sigma_1,\ldots,\sigma_n)}]\end{aligned}$$

with

$$E(\sigma_1,\ldots,\sigma_n) = -\log\left[\int d\xi\; p(\xi)e^{-E(\sigma_1,\xi)-\ldots-E(\sigma_n,\xi)}\right]$$

The 'trick' consists in having used $Z^n = Z.Z...Z$ (n times) to go from line 1 to line 2, which is true only for integer n. If for now we accept this step, we see that (apart from a factor $\frac{1}{n}$) our final expression for $\langle f\rangle$ can be interpreted as the free energy of a new system, consisting of n interacting copies ('replicas') of the original one, described by a *non-disordered* energy function $E(\sigma_1,\ldots,\sigma_n)$. The price paid for the elimination of the disorder is having to take the strange limit $n \to 0$ at the end.

From this stage onwards it is generally plain mathematical sailing, and one usually arrives in the end at an expression of the following form, where we have to determine the minimum of a function $F[\boldsymbol{q}]$ of an $n\times n$ matrix $\boldsymbol{q}$ (a table with n rows and n columns of real numbers) which is restricted to having zero diagonal elements:

$$\langle f\rangle = \lim_{n\to 0}\ \min F[\boldsymbol{q}] \qquad\qquad \boldsymbol{q} = \begin{pmatrix} 0 & q_{1,2} & \cdots & q_{1,n-1} & q_{1,n} \\ q_{2,1} & 0 & & & q_{2,n} \\ \vdots & & \ddots & & \vdots \\ q_{n-1,1} & & & 0 & q_{n-1,n} \\ q_{n,1} & q_{n,2} & \cdots & q_{n,n-1} & 0 \end{pmatrix}$$

However, the number of nonzero entries of $\boldsymbol{q}$ is $n(n-1)$, which is *negative* for $n < 1$. Since $n \to 0$ we are to find the minimum of a function in a space of *negative* dimension. This was not greeted with joy; the above procedure was only accepted (although not by all) after it turned out to produce results which agreed perfectly with experiments ...

Thus we now find ourselves reluctantly with calculations in spaces of negative dimension. There are no examples (yet) of sets with $D < 0$, but one can easily see that sets and points in spaces with $D < 0$ must have peculiar properties. For instance

- The length of a real-valued vector in $D < 0$ dimensions will be negative:
  $$x_i \in \mathbb{R}:\quad \sum_i x_i^2 \leq 0 \qquad e.g.\quad x_i = x\ \text{ for all } i:\quad \sum_i x_i^2 = Dx^2 \leq 0$$
  This suggests that the relevant quantity to work with might be the average over components, since the latter would come out as $D^{-1}\sum_i x_i^2 \geq 0$.
- Hyper-cubes of size L would decrease in volume with increasing size:
  $$C = \{\boldsymbol{x} \in \mathbb{R}^D |\ x_i \in [0,L] \text{ for all } i\}$$
  Volume: $V(L) = L^D = L^{-|D|}$, so $\lim_{L\to\infty} V(L) = 0$ and $\lim_{L\to 0} V(L) = \infty$.
- A Hyper-sphere of radius L in dimension $D < 0$ would have to be defined as
  $$S = \{\boldsymbol{x} \in \mathbb{R}^D |\ \sum_i x_i^2 = -L^2\}$$

All this seems crazy, and for now there is no pretense that we know what we are doing, yet if we simply persist and work our way through the calculations without the ambition to understand what it all means, the above replica formalism continues to produce accurate and beautiful mathematical results. This suggests strongly that there might well be a consistent definition of negative dimension around the corner !