CAMBRIDGE

Cambridge University Press
978-1-107-02625-4 — Geometry of Quantum States
2nd Edition
Excerpt
More Information

1

Convexity, colours and statistics

1.1 Convex sets

What picture does one see, looking at a physical theory from a distance, so that the details disappear? Since quantum mechanics is a statistical theory, the most universal picture which remains after the details are forgotten is that of a convex set.

—*Bogdan Mielnik*[1]

Our object is to understand the geometry of the set of all possible states of a quantum system that can occur in nature. This is a very general question; especially since we are not trying to define 'state' or 'system' very precisely. Indeed we will not even discuss whether the state is a property of a thing, or of the preparation of a thing, or of a belief about a thing. Nevertheless we can ask what kind of restrictions are needed on a set if it is going to serve as a space of states in the first place. There is a restriction that arises naturally both in quantum mechanics and in classical statistics: the set must be a *convex set*. The idea is that a convex set is a set such that one can form 'mixtures' of any pair of points in the set. This is, as we will see, how probability enters (although we are not trying to define 'probability' either).

From a geometrical point of view a *mixture* of two states can be defined as a point on the segment of the straight line between the two points that represent the states that we want to mix. We insist that given two points belonging to the set of states, the straight line segment between them must belong to the set too. This is certainly not true of any set. But before we can see how this idea restricts the set of states we must have a definition of 'straight lines' available. One way to proceed is to regard a convex set as a special kind of subset of a flat Euclidean space $\mathbf{E}^n$. Actually we can get by with somewhat less. It is enough to regard a convex set as a subset of an affine space. An *affine space* is just like a vector space, except that no

[1] Reproduced from [659].

special choice of origin is assumed. The *straight line* through the two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined as the set of points

$$\mathbf{x} = \mu_1 \mathbf{x}_1 + \mu_2 \mathbf{x}_2, \qquad \mu_1 + \mu_2 = 1. \tag{1.1}$$

If we choose a particular point $\mathbf{x}_0$ to serve as the origin, we see that this is a one-parameter family of vectors $\mathbf{x} - \mathbf{x}_0$ in the plane spanned by the vectors $\mathbf{x}_1 - \mathbf{x}_0$ and $\mathbf{x}_2 - \mathbf{x}_0$. Taking three different points instead of two in Eq. (1.1) we define a *plane*, provided the three points do not belong to a single line. A k-dimensional *k–plane* is obtained by taking $k + 1$ generic points, where $k < n$. For $k = n$ we describe the entire space $\mathbf{E}^n$. In this way we may introduce *barycentric coordinates* into an n-dimensional affine space. We select $n + 1$ points $\mathbf{x}_i$, so that an arbitrary point $\mathbf{x}$ can be written as

$$\mathbf{x} = \mu_0 \mathbf{x}_0 + \mu_1 \mathbf{x}_1 + \ldots + \mu_n \mathbf{x}_n, \qquad \mu_0 + \mu_1 + \ldots + \mu_n = 1. \tag{1.2}$$

The requirement that the barycentric coordinates μ_i add up to one ensures that they are uniquely defined by the point $\mathbf{x}$. (It also means that the barycentric coordinates are not coordinates in the ordinary sense of the word, but if we solve for μ_0 in terms of the others then the remaining independent set is a set of n ordinary coordinates for the n-dimensional space.) An *affine map* is a transformation that takes lines to lines and preserves the relative length of line segments lying on parallel lines. In equations an affine map is a combination of a linear transformation described by a matrix $\mathbf{A}$ with a translation along a constant vector $\mathbf{b}$, so $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{A}$ is an invertible matrix.

By definition a subset S of an affine space is a *convex set* if for any pair of points $\mathbf{x}_1$ and $\mathbf{x}_2$ belonging to the set it is true that the *mixture* $\mathbf{x}$ also belongs to the set, where

$$\mathbf{x} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2, \qquad \lambda_1 + \lambda_2 = 1, \quad \lambda_1, \lambda_2 \geq 0. \tag{1.3}$$

Here λ_1 and λ_2 are barycentric coordinates on the line through the given pair of points; the extra requirement that they be positive restricts $\mathbf{x}$ to belong to the segment of the line lying between the pair of points.

It is natural to use an affine space as the 'container' for the convex sets since convexity properties are preserved by general affine transformations. On the other hand it does no harm to introduce a flat metric on the affine space, turning it into an Euclidean space. There may be no special significance attached to this notion of distance, but it helps in visualizing what is going on. See Figures 1.1 and 1.2. From now on, we will assume that our convex sets sit in Euclidean space, whenever it is convenient to do so.

Intuitively a convex set is a set such that one can always see the entire set from whatever point in the set one happens to be sitting at. Still they can come in a variety of interesting shapes. We will need a few definitions. First, given any subset of the

Cambridge University Press
978-1-107-02625-4 — Geometry of Quantum States
2nd Edition
Excerpt
More Information

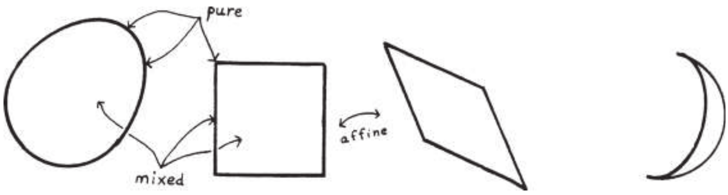

Figure 1.1 Three convex sets, two of which are affine transformations of each other. The new moon is not convex. An observer in Singapore will find the new moon tilted but still not convex, since convexity is preserved by rotations.

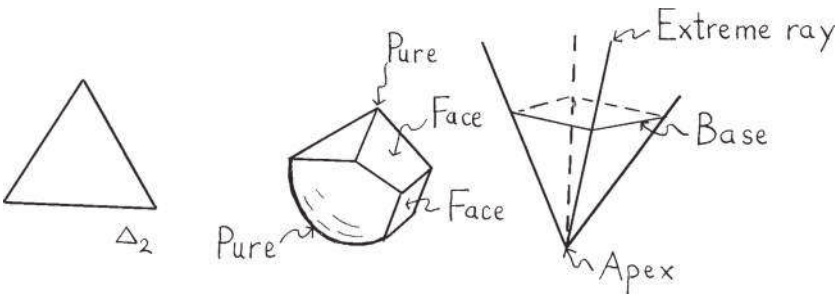

Figure 1.2 The convex sets we will consider are either convex bodies (like the simplex on the left, or the more involved example in the centre) or convex cones with compact bases (an example is shown on the right).

affine space we define the *convex hull* of this subset as the smallest convex set that contains the set. The convex hull of a finite set of points is called a *convex polytope*. If we start with $p+1$ points that are not confined to any $(p-1)$-dimensional subspace then the convex polytope is called a *p-simplex*. The p-simplex consists of all points of the form

$$\mathbf{x} = \lambda_0 \mathbf{x}_0 + \lambda_1 \mathbf{x}_1 + \ldots + \lambda_p \mathbf{x}_p, \qquad \lambda_0 + \lambda_1 + \ldots + \lambda_p = 1, \quad \lambda_i \geq 0. \tag{1.4}$$

(The barycentric coordinates are all non–negative.) The *dimension* of a convex set is the largest number n such that the set contains an n-simplex. When discussing a convex set of dimension n we usually assume that the underlying affine space also has dimension n, to ensure that the convex set possesses interior points (in the sense of point set topology). A closed and bounded convex set that has an interior is known as a *convex body*.

The intersection of a convex set with some lower dimensional subspace of the affine space is again a convex set. Given an n-dimensional convex set S there is also a natural way to increase its dimension with one: choose a point $\mathbf{y}$ not belonging to the n-dimensional affine subspace containing S. Form the union of all the *rays* (in this chapter a ray means a half line), starting from $\mathbf{y}$ and passing through S.

Cambridge University Press
978-1-107-02625-4 — Geometry of Quantum States
2nd Edition
Excerpt
More Information

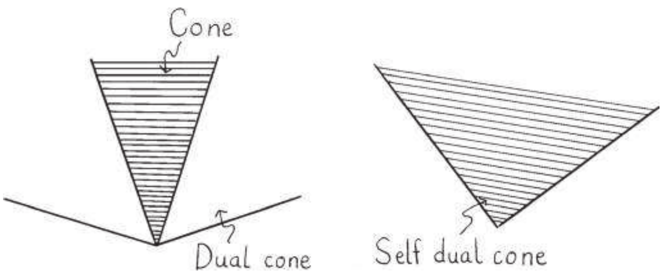

Figure 1.3 Left: a convex cone and its dual, both regarded as belonging to Euclidean 2-space. Right: a self dual cone, for which the dual cone coincides with the original. For an application of this construction see Figure 11.6.

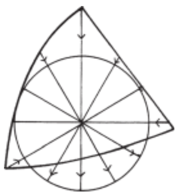

Figure 1.4 A convex body is homeomorphic to a sphere.

The result is called a *convex cone* and $\mathbf{y}$ is called its *apex*, while S is its *base*. A ray is in fact a one dimensional convex cone. A more interesting example is obtained by first choosing a p-simplex and then interpreting the points of the simplex as vectors starting from an origin O not lying in the simplex. Then the $p+1$ dimensional set of points

$$\mathbf{x} = \lambda_0 \mathbf{x}_0 + \lambda_1 \mathbf{x}_1 + \ldots + \lambda_p \mathbf{x}_p, \qquad \lambda_i \geq 0 \tag{1.5}$$

is a convex cone. Convex cones have many nice properties, including an inbuilt partial order among its points: $\mathbf{x} \leq \mathbf{y}$ if and only if $\mathbf{y} - \mathbf{x}$ belongs to the cone. Linear maps to $\mathbb{R}$ that take positive values on vectors belonging to a convex cone form a dual convex cone in the dual vector space. Since we are in the Euclidean vector space $\mathbf{E}^n$, we can identify the dual vector space with $\mathbf{E}^n$ itself. If the two cones agree the convex cone is said to be *self dual*. See Figure 1.3. One self dual convex cone that will appear now and again is the *positive orthant* or *hyperoctant* of $\mathbf{E}^n$, defined as the set of all points whose Cartesian coordinates are non-negative. We use the notation $\mathbf{x} \geq 0$ to denote the fact that $\mathbf{x}$ belongs to the positive orthant.

From a purely topological point of view all convex bodies are equivalent to an n–dimensional ball. To see this choose any point $\mathbf{x}_0$ in the interior and then for every point in the boundary draw a ray starting from $\mathbf{x}_0$ and passing through the boundary point (as in Figure 1.4). It is clear that we can make a continuous transformation of the convex body into a ball with radius one and its centre at $\mathbf{x}_0$ by moving the points of the container space along the rays.

Convex bodies and convex cones with compact bases are the only convex sets that we will consider. Convex bodies always contain some special points that cannot be obtained as mixtures of other points – whereas a half space does not! These points are called *extreme points* by mathematicians and *pure* points by physicists (actually, originally by Weyl), while non-pure points are called *mixed*. In a convex cone the rays from the apex through the pure points of the base are called *extreme rays*; a point $\mathbf{x}$ lies on an extreme ray if and only if $\mathbf{y} \leq \mathbf{x} \Rightarrow \mathbf{y} = \lambda\mathbf{x}$ with λ between zero and one. A subset F of a convex set that is stable under mixing and purification is called a *face* of the convex set. What the phrase means is that if

$$\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \qquad 0 < \lambda < 1 \tag{1.6}$$

then $\mathbf{x}$ lies in F if and only if $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in F. The 'only if' part of the definition forces $\mathbf{x}$ to lie on the boundary of the set. A face of dimension k is a k-face. A 0-face is an extreme point, and an $(n-1)$-face is also known as a *facet*. It is interesting to observe that the set of all faces on a convex body form a partially ordered set; we say that $F_1 \leq F_2$ if the face F_1 is contained in the face F_2. It is a partially ordered set of the special kind known as a *lattice*, which means that a given pair of faces always have a greatest lower bound (perhaps the empty set) and a lowest greater bound (perhaps the convex body itself).

To stem the tide of definitions let us quote two theorems that have an 'obvious' ring to them when they are stated abstractly but which are surprisingly useful in practice:

Minkowski's theorem. *Any convex body is the convex hull of its pure points.*

Carathéodory's theorem. *Any point in an n-dimensional convex set X can be expressed as a convex combination of at most $n+1$ pure points in X.*

Thus any point $\mathbf{x}$ of a convex body S may be expressed as a *convex combination* of pure points:

$$\mathbf{x} = \sum_{i=1}^{p} \lambda_i \mathbf{x}_i, \qquad \lambda_i \geq 0, \qquad p \leq n+1, \qquad \sum_i \lambda_i = 1. \tag{1.7}$$

This equation is quite different from Eq. (1.2) that defined the barycentric coordinates of $\mathbf{x}$ in terms of a fixed set of points $\mathbf{x}_i$, because – with the restriction that all the coefficients be non-negative – it may be impossible to find a finite set of $\mathbf{x}_i$ so that every $\mathbf{x}$ in the set can be written in this new form. An obvious example is a circular disk. Given $\mathbf{x}$ one can always find a finite set of pure points $\mathbf{x}_i$ so that the

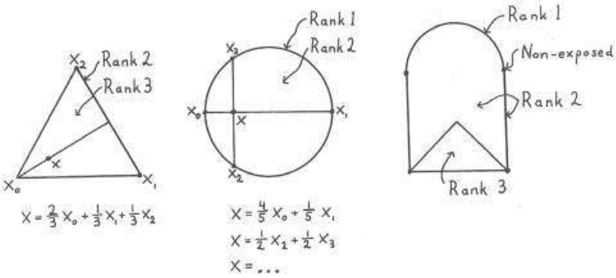

Figure 1.5 In a simplex a point can be written as a mixture in one and only one way. In general the rank of a point is the minimal number of pure points needed in the mixture; the rank may change in the interior of the set as shown in the rightmost example. The set on the right has two non-exposed points which form faces of their own. Note also that its insphere is not unique.

equation holds, but that is a different thing. The exact number of pure points one needs is related to the face structure of the body, as one can see from the proof of Carathéodory's theorem (which we give as Problem 1.1.)

It is evident that the pure points always lie in the boundary of the convex set, but the boundary often contains mixed points as well. The simplex enjoys a very special property, which is that any point in the simplex can be written as a mixture of pure points in one and only one way (as in Figure 1.5). This is because for the simplex the coefficients in Eq. (1.7) are barycentric coordinates and the result follows from the uniqueness of the barycentric coordinates of a point. No other convex set has this property. The *rank* of a point $\mathbf{x}$ is the minimal number p needed in the convex combination (1.7). By definition the pure points have rank one. In a simplex the edges have rank two, the faces have rank three, and so on, while all the points in the interior have maximal rank. From Eq. (1.7) we see that the maximal rank of any point in a convex body in $\mathbb{R}^n$ does not exceed $n+1$. In a ball all interior points have rank two and all points on the boundary are pure, regardless of the dimension of the ball. It is not hard to find examples of convex sets where the rank changes as we move around in the interior of the set (see Figure 1.5).

The simplex has another quite special property, namely that its lattice of faces is *self dual*. We observe that the number of k-faces in an n dimensional simplex is

$$\binom{n+1}{k+1} = \binom{n+1}{n-k}. \tag{1.8}$$

Hence the set of $n-k-1$ dimensional faces can be put in one-to-one correspondence with the set of k-faces. In particular, the pure points ($k = 0$) can be put in one-to-one correspondence with the set of *facets* (by definition, the $n-1$ dimensional faces).

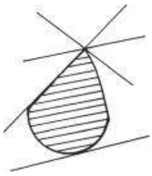

Figure 1.6 Support hyperplanes of a convex set.

For this, and other, reasons its lattice of subspaces will have some exceptional properties, turning it into what is technically known as a *Boolean* lattice.[2]

There is a useful dual description of convex sets in terms of supporting hyperplanes. A *support hyperplane* of S is a hyperplane that intersects the set and which is such that the entire set lies in one of the closed half spaces formed by the hyperplane (see Figure 1.6). Hence a support hyperplane just touches the boundary of S, and one can prove that there is a support hyperplane passing through every point of the boundary of a convex body. By definition a *regular point* is a point on the boundary that lies on only one support hyperplane, a *regular support hyperplane* meets the set in only one point, and the entire convex set is regular if all its boundary points as well as all its support hyperplanes are regular. So a ball is regular, while a convex polytope or a convex cone is not – indeed all the support hyperplanes of a convex cone pass through its apex. A face is said to be *exposed* if it equals the intersection of the convex set and some support hyperplane. Convex polytopes arise as the intersection of a finite number of closed half-spaces in $\mathbb{R}^n$, and any pure point of a convex polytope saturates n of the inequalities that define the half-spaces; again a statement with an 'obvious' ring that is useful in practice.

In a flat Euclidean space a linear function to the real numbers takes the form $\mathbf{x} \to \mathbf{a} \cdot \mathbf{x}$, where $\mathbf{a}$ is some constant vector. Geometrically, this defines a family of parallel hyperplanes. We have the important

Hahn–Banach separation theorem. *Given a convex body and a point $\mathbf{x_0}$ that does not belong to it. Then one can find a linear function f and a constant k such that $f(\mathbf{x}) > k$ for all points belonging to the convex body, while $f(\mathbf{x_0}) < k$.*

This is again almost obvious if one thinks in terms of hyperplanes.

It is useful to know a bit more about dual convex sets. For definiteness let us start out with a three dimensional vector space in which a point is represented by a vector $\mathbf{y}$. Then its dual plane is the set of vectors $\mathbf{x}$ such that

$$\mathbf{x} \cdot \mathbf{y} = -1. \tag{1.9}$$

[2] Because it is related to what George Boole thought were the laws of thought; see Varadarajan's book [916] on quantum logic for these things.

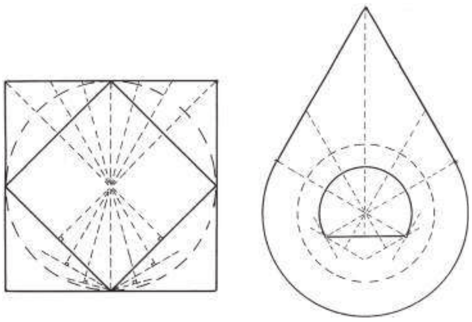

Figure 1.7 A square is dual to another square; on the left we see how the points on the edge at the top define a corner at the bottom of the dual square. To the right we see a more complicated convex set with non-exposed faces (that are points). Its dual has non-polyhedral corners. In both cases the unit circle is shown dashed.

The constant on the right hand side was set to -1 for convenience. The dual of a line is the intersection of a one-parameter family of planes dual to the points on the line. This is in itself a line. The dual of a plane is a point, while the dual of a curved surface is another curved surface – the envelope of the planes that are dual to the points on the original surface. To define the dual of a convex body with a given boundary we change the definition slightly, and include all points on one side of the dual planes in the dual. Thus the *dual* X° of a convex body X is defined to be

$$X^\circ = \{\mathbf{x} \mid c + \mathbf{x} \cdot \mathbf{y} \ \geq \ 0 \ \ \forall \mathbf{y} \in X\}, \tag{1.10}$$

where c is a number that was set equal to 1 above (and also when drawing Figure 1.7). The dual of a convex body including the origin is the intersection of the half-spaces defined by the pure points $\mathbf{y}$ of X. The dual of the dual of a body that includes the origin is equal to the convex hull of the original body. If we enlarge a convex body the conditions on the dual become more stringent, and hence the dual shrinks. The dual of a sphere centred at the origin is again a sphere, so a sphere (of suitable radius) is self dual. The dual of a cube is an octahedron. The dual of a regular tetrahedron is another copy of the original tetrahedron, possibly of a different size. The copy can be made to coincide with the original by means of an affine transformation, hence the tetrahedron is a self dual body.[3]

We will find much use for the concept of *convex functions*. A real function $f(\mathbf{x})$ defined on a closed convex subset X of $\mathbb{R}^n$ is called *convex*, if for any $\mathbf{x}, \mathbf{y} \in X$ and $\lambda \in [0, 1]$ it satisfies

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}). \tag{1.11}$$

[3] To readers who wish to learn more about convex sets – or who wish to see proofs of the various assertions that we left unproved – we recommend the book by Eggleston [283].

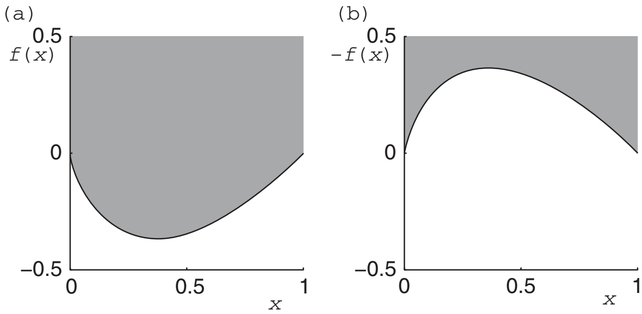

Figure 1.8 (a): the convex function $f(x) = x \ln x$ (b): the concave function $g(x) = -x \ln x$. The names stem from the shaded epigraphs of the functions which are convex and concave, respectively.

The name refers to the fact that the *epigraph* of a convex function, that is the region lying above the curve $f(\mathbf{x})$ in the graph, is convex. Applying the inequality $k-1$ times we see that

$$f\left(\sum_{j=1}^{k} \lambda_j \mathbf{x}_j\right) \leq \sum_{j=1}^{k} \lambda_j f(\mathbf{x}_j), \tag{1.12}$$

where $\mathbf{x}_j \in X$ and the nonnegative weights sum to unity, $\sum_{j=1}^{k} \lambda_j = 1$. If a function f from $\mathbb{R}$ to $\mathbb{R}$ is differentiable, it is convex if and only if

$$f(y) - f(x) \geq (y-x) f'(x). \tag{1.13}$$

If f is twice differentiable it is convex if and only if its second derivative is non-negative. For a function of several variables to be convex, the matrix of second derivatives must be positive definite. In practice, this is a very useful criterion. A function f is called *concave* if $-f$ is convex.

One of the main advantages of convex functions is that it is (comparatively) easy to study their minima and maxima. A minimum of a convex function is always a global minimum, and it is attained on some convex subset of the domain of definition X. If X is not only convex but also compact, then the global maximum sits at an extreme point of X.

1.2 High dimensional geometry

In quantum mechanics the spaces we encounter are often of very high dimension; even if the dimension of Hilbert space is small the dimension of the space of density matrices will be high. Our intuition on the other hand is based on two and three dimensional spaces, and frequently leads us astray. We can improve ourselves by asking some simple questions about convex bodies in flat space. We choose to look at balls, cubes and simplices for this purpose. A flat metric is assumed. Our questions will concern the *inspheres* and *outspheres* of these bodies (defined as the largest inscribed sphere and the smallest circumscribed sphere, respectively). For any convex body the outsphere is uniquely defined, while the insphere is not – one can show that the upper bound on the radius of inscribed spheres is always attained by some sphere, but there may be several of those.

Let us begin with the surface of a ball, namely the n-dimensional *sphere*. In equations a sphere of radius r is given by the set

$$X_0^2 + X_1^2 + \cdots + X_n^2 = r^2 \tag{1.14}$$

in an $n+1$ dimensional flat space $\mathbf{E}^{n+1}$. A sphere of radius one is denoted $\mathbf{S}^n$. The sphere can be parametrized by the angles $\phi, \theta_1, \ldots, \theta_{n-1}$ according to

$$\begin{cases} X_0 = r\cos\phi\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-1} \\ X_1 = r\sin\phi\sin\theta_1\sin\theta_2\ldots\sin\theta_{n-1} \\ X_2 = r\cos\theta_1\sin\theta_2\ldots\sin\theta_{n-1} \\ \cdots \qquad \cdots \\ X_n = r\cos\theta_{n-1} \end{cases} \qquad \begin{array}{l} 0 < \theta_i < \pi \\ 0 \leq \phi < 2\pi \end{array} . \tag{1.15}$$

The volume element $\mathrm{d}A$ on the unit sphere then becomes

$$\mathrm{d}A = \mathrm{d}\phi\mathrm{d}\theta_1\ldots\mathrm{d}\theta_{n-1}\sin\theta_1\sin^2\theta_2\ldots\sin^{n-1}\theta_{n-1}. \tag{1.16}$$

We want to compute the volume vol($\mathbf{S}^n$) of the n-sphere, that is to say its ‘hyperarea’ – meaning that vol($\mathbf{S}^2$) is measured in square metres, vol($\mathbf{S}^3$) in cubic metres, and so on. A clever trick simplifies the calculation: Consider the well known Gaussian integral

$$I = \int e^{-X_0^2 - X_1^2 - \ldots - X_n^2}\mathrm{d}X_0\mathrm{d}X_1 \ldots \mathrm{d}X_n = (\sqrt{\pi})^{n+1}. \tag{1.17}$$

Using the spherical polar coordinates introduced above our integral splits into two, one of which is related to the integral representation of the Euler Gamma function, $\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\mathrm{d}t$, and the other is the one we want to do:

$$I = \int_0^\infty \mathrm{d}r \int_{\mathbf{S}^n} \mathrm{d}A e^{-r^2} r^n = \frac{1}{2}\Gamma\left(\frac{n+1}{2}\right)\mathrm{vol}(\mathbf{S}^n). \tag{1.18}$$

We do not have to do the integral over the angles. We simply compare these results and obtain (recalling the properties of the Gamma function)