Geometry of Quantum States

Ingemar Bengtsson and Karol Życzkowski

An Introduction to Quantum Entanglement
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

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 $E^n$. 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 special choice of origin is assumed. The straight line through the two points $x_1$ and $x_2$ is defined as the set of points

$$x = \mu_1 x_1 + \mu_2 x_2 , \quad \mu_1 + \mu_2 = 1 .$$

(1.1)

If we choose a particular point $x_0$ to serve as the origin, we see that this is a one parameter family of vectors $x - x_0$ in the plane spanned by the vectors $x_1 - x_0$.
Convexity, colours and statistics

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.

and \( x_2 - x_0 \). 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 hyperplane is obtained by taking \( k + 1 \) generic points, where \( k < n \). For \( k = n \) we describe the entire space \( \mathbb{E}^n \). In this way we may introduce barycentric coordinates into an \( n \)-dimensional affine space. We select \( n + 1 \) points \( x_i \), so that an arbitrary point \( x \) can be written as

\[
x = \mu_0 x_0 + \mu_1 x_1 + \cdots + \mu_n x_n , \quad \mu_0 + \mu_1 + \cdots + \mu_n = 1.
\]  

(1.2)

The requirement that the barycentric coordinates \( \mu_i \) add up to one ensures that they are uniquely defined by the point \( x \). (It also means that the barycentric coordinates are not coordinates in the ordinary sense of the word, but if we solve for \( \mu_0 \) 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 \( A \) with a translation along a constant vector \( b \), so \( x' = Ax + b \), where \( A \) is an invertible matrix.

By definition a subset \( S \) of an affine space is a convex set if for any pair of points \( x_1 \) and \( x_2 \) belonging to the set it is true that the mixture \( x \) also belongs to the set, where

\[
x = \lambda_1 x_1 + \lambda_2 x_2 , \quad \lambda_1 + \lambda_2 = 1 , \quad \lambda_1, \lambda_2 \geq 0 .
\]  

(1.3)

Here \( \lambda_1 \) and \( \lambda_2 \) are barycentric coordinates on the line through the given pair of points; the extra requirement that they be positive restricts \( x \) 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. 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...
1.1 Convex sets

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 center) or convex cones with compact bases (an example is shown on the right).

come in a variety of interesting shapes. We will need a few definitions. First, given any subset of the affine space we define the \textit{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 \textit{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 \( p \)-\textit{simplex}. The \( p \)-simplex consists of all points of the form

\[
x = \lambda_0 x_0 + \lambda_1 x_1 + \ldots + \lambda_p x_p , \quad \lambda_0 + \lambda_1 + \ldots + \lambda_p = 1 , \quad \lambda_i \geq 0 .
\]  

(1.4)

(The barycentric coordinates are all non-negative.) The \textit{dimension} of a convex set is the largest number \( n \) such that the set contains an \( n \)-simplex. In 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 \textit{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 \( y \) not belonging to the \( n \) dimensional affine subspace containing \( S \). Form the union of all the \textit{rays} (in this chapter a ray means a half line), starting from \( y \) and passing through \( S \). The result is called a \textit{convex cone} and \( y \) is called its \textit{apex}, while \( S \) is its \textit{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

\[
x = \lambda_0 x_0 + \lambda_1 x_1 + \ldots + \lambda_p x_p , \quad \lambda_i \geq 0
\]  

(1.5)

is a convex cone. Convex cones have many nice properties, including an inbuilt partial order among its points: \( x \leq y \) if and only if \( y - x \) 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 \( \mathbb{E}^n \), we can identify the dual vector space with \( \mathbb{E}^n \) itself. If the two cones agree the convex cone is said to be \textit{self dual}. One self
Convexity, colours and statistics

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 Fig. 11.6.

dual convex cone that will appear now and again is the positive orthant or hyperoctant of $E^n$, 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 Fig. 1.4). It is clear that we can make a continuous transformation of the convex body into a ball with radius one and its center at $\mathbf{x}_0$ by moving the points of the container space along the rays.

Figure 1.4. A convex body is homeomorphic to a sphere.

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 $\lambda$ 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

$$\mathbf{x} = \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2 , \quad 0 \leq \lambda \leq 1$$

(1.6)
then \( x \) lies in \( F \) if and only if \( x_1 \) and \( x_2 \) lie in \( F \). A face of dimension \( k \) is a \( k \)-face. A 0-face is an extremal 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.** If \( X \) is a subset of \( \mathbb{R}^n \) then any point in the convex hull of \( X \) can be expressed as a convex combination of at most \( n + 1 \) points in \( X \).

Thus any point \( x \) of a convex body \( S \) may be expressed as a convex combination of pure points:

\[
x = \sum_{i=1}^{p} \lambda_i x_i, \quad \lambda_i \geq 0, \quad p \leq n + 1, \quad \sum_{i} \lambda_i = 1.
\]  

(1.7)

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

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 Fig. 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 \( x \) 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 Fig. 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
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 of \( n \) \( k \)-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). 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.\(^1\)

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 Fig. 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. 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 \( \mathbf{x} \rightarrow a \cdot \mathbf{x} \), where \( a \) is some constant vector. Geometrically, this defines a family of parallel hyperplanes. We have the important

\[ \binom{n + 1}{k + 1} = \binom{n + 1}{n - k}. \]  

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). 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.\(^1\)

\[ \text{Figure 1.5. In a simplex a point can be written as a mixture in one and only one way. In general the rank of a point is the minimal number of pure points needed in the mixture; the rank may change in the interior of the set as shown in the rightmost example.} \]

\( ^1 \)Because it is related to what George Boole thought were the laws of thought; see Varadarajan’s book [7] on quantum logic for these things.
1.1 Convex sets

Hahn–Banach separation theorem. Given a convex body and a point $x_0$ that does not belong to it. Then one can find a linear function $f$ that takes positive values for all points belonging to the convex body, while $f(x_0) < 0$.

This is again almost obvious if one thinks in terms of hyperplanes.\(^2\)

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

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y). \quad (1.9)$$

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

$$f \left( \sum_{j=1}^{k} \lambda_j x_j \right) \leq \sum_{j=1}^{k} \lambda_j f(x_j), \quad (1.10)$$

where $x_j \in X$ 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

$$f(y) - f(x) \geq (y - x) f'(x). \quad (1.11)$$

Figure 1.6. Support hyperplanes of a convex set

Figure 1.7. (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.

\(^2\) 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 (1958) [?].
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 \textit{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 \textit{inspheres} and \textit{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 \textit{sphere} \( S^n \). In equations a sphere of radius \( r \) is given by the set

\[
X_0^2 + X_1^2 + \ldots + X_n^2 = r^2
\]

in an \( n + 1 \) dimensional flat space \( E^{n+1} \). A sphere of radius one is denoted \( S^n \).

The sphere can be parametrized by the angles \( \phi, \theta_1, \ldots, \theta_{n-1} \) according to

\[
\begin{align*}
X_0 &= r \cos \phi \sin \theta_1 \sin \theta_2 \ldots \sin \theta_{n-1} \\
X_1 &= r \sin \phi \sin \theta_1 \sin \theta_2 \ldots \sin \theta_{n-1} \\
X_2 &= r \cos \theta_1 \sin \theta_2 \ldots \sin \theta_{n-1} \quad 0 < \theta_i < \pi \\
& \quad \ldots \\
X_n &= r \cos \theta_{n-1} \\
\end{align*}
\]

The volume element \( dA \) on the unit sphere then becomes

\[
dA = \, d\phi \, d\theta_1 \ldots d\theta_{n-1} \sin \theta_1 \sin^2 \theta_2 \ldots \sin^{n-1} \theta_{n-1} .
\]

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

\[
I = \int e^{-X_0^2} e^{-X_1^2} \ldots e^{-X_n^2} \, dX_0 \, dX_1 \ldots \, dX_n = (\sqrt{\pi})^{n+1} .
\]
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, 
\[ \Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \]
and the other is the one we want to do:
\[ I = \int_0^\infty dr \int_{S^n} dA e^{-r^2 r^n} = \frac{1}{2} \Gamma \left( \frac{n+1}{2} \right) \text{vol}(S^n) . \] 

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)
\[ \text{vol}(S^n) = 2 \frac{\pi^{\frac{n+1}{2}}}{\Gamma \left( \frac{n+1}{2} \right)} = \begin{cases} \frac{2(2\pi)^p}{(2p-1)!!} & \text{if } n = 2p \\ \frac{(2\pi)^{p+1}}{(2p)!!} & \text{if } n = 2p + 1 \end{cases} \] 

where double factorial is the product of every other number, 5!! = 5 · 3 · 1 and 6!! = 6 · 4 · 2. An alarming thing happens as the dimension grows. For large \( x \) we can approximate the Gamma function using Stirling’s formula
\[ \Gamma(x) = \sqrt{2\pi} e^{-\frac{1}{2}} \left( 1 + \frac{1}{12x} + o \left( \frac{1}{x^2} \right) \right) . \] 

Hence for large \( n \) we obtain
\[ \text{vol}(S^n) \sim \sqrt{2} \left( \frac{2\pi e}{n} \right)^{\frac{n}{2}} . \] 

This is small if \( n \) is large! In fact the “biggest” unit sphere—in the sense that it has the largest hyperarea—is \( S^6 \), which has
\[ \text{vol}(S^6) = \frac{16}{15} \pi^3 \approx 33.1 . \] 

Incidentally Stirling’s formula gives 31.6, which is already rather good. We hasten to add that \( \text{vol}(S^2) \) is measured in square meters and \( \text{vol}(S^6) \) in (meter)\(^6\), so that the direct comparison makes no sense.

There is another funny thing to be noticed. If we compute the volume of the \( n \)-sphere without any clever tricks, simply by integrating the volume element \( dA \) using angular coordinates, then we find that
\[ \text{vol}(S^n) = 2\pi \int_0^\pi d\theta \sin \theta \int_0^\pi d\theta \sin^2 \theta \cdots \int_0^\pi d\theta \sin^n \theta = \] 

\[ \text{vol}(S^{n-1}) \int_0^\pi d\theta \sin^{n-1} \theta . \] 

As \( n \) grows the integrand of the final integral has an increasingly sharp peak close to the equator \( \theta = \pi/2 \). Hence we conclude that when \( n \) is high most of the hyperarea of the sphere is confined to a “band” close to the equator.

What about the volume of an \( n \)-dimensional unit ball \( B^n \)? By definition it
has unit radius and its boundary is $S^{n-1}$. Its volume, using the radial integral
\[ \int_0^1 r^{n-1} dr = 1/n \] and the fact that $\Gamma(x + 1) = x\Gamma(x)$, is
\[ \text{vol}(B^n) = \frac{\text{vol}(S^{n-1})}{n} = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} \sim \frac{1}{\sqrt{2\pi}} \left( \frac{2\pi e}{n} \right)^{\frac{n}{2}}. \quad (1.22) \]
Again, as the dimension grows the denominator grows faster than the numerator and therefore the volume of a unit ball is small when the dimension is high. We can turn this around if we like: a ball of unit volume has a large radius if the dimension is high. Indeed since the volume is proportional to $r^n$, where $r$ is the radius, it follows that the radius of a ball of unit volume grows like $\sqrt{n}$ when Stirling’s formula applies.

The fraction of the volume of a unit ball that lies inside a radius $r$ is $r^n$. We assume $r < 1$, so this is a quickly shrinking fraction as $n$ grows. The curious conclusion of this is that when the dimension is high almost all of the volume of a ball lies very close to its surface. In fact this is a crucial observation in statistical mechanics. It is also the key property of $n$-dimensional geometry: when $n$ is large the “amount of space” available grows very fast with distance from the origin.

In some ways it is easier to see what is going on if we consider hypercubes $\Box_n$ rather than balls. Take a cube of unit volume. In $n$ dimensions it has $2^n$ corners, and the longest straight line that we can draw inside the hypercube connects two opposite corners. It has length $L = \sqrt{1^2 + \ldots + 1^2} = \sqrt{n}$. Or expressed in another way, a straight line of any length fits into a hypercube of unit volume if the dimension is large enough. The reason why the longest line segment fitting into the cube is large is clearly that we normalized the volume to one. If we normalize $L = 1$ instead we find that the volume goes to zero like $(1/\sqrt{n})^n$. Concerning the insphere—the largest inscribed sphere, with inradius $r_n$—and the outsphere—the smallest circumscribed sphere, with outradius $R_n$—we observe that
\[ R_n = \frac{\sqrt{n}}{2} = \sqrt{n}r_n. \quad (1.23) \]
The ratio between the two grows with the dimension, $\zeta_n \equiv R_n/r_n = \sqrt{n}$. Incidentally, the somewhat odd statement that the volume of a sphere goes to zero when the dimension $n$ goes to infinity can now be interpreted: since $\text{vol}(\Box_n) = 1$ the real statement is that $\text{vol}(S^n)/\text{vol}(\Box_n)$ goes to zero when $n$ goes to infinity.

Now we turn to simplices, whose properties will be of some importance later on. We concentrate on regular simplices $\Delta_n$, for which the distance between any pair of corners is one. For $n = 1$ this is the unit interval, for $n = 2$ a regular triangle, for $n = 3$ a regular tetrahedron, and so on. Again we are interested in the volume, the radius $r_n$ of the insphere, and the radius $R_n$ of the circumsphere. We will also compute $\chi_n$, the angle between the lines from the “center of mass” to a pair of corners. For a triangle it is $\arccos(-1/2) = 2\pi/3 = 120$ degrees, but it drops to $\arccos(-1/3) \approx 110^\circ$ for the tetrahedron. A practical way to go about all this is to think of $\Delta_n$ as a (part of) a cone
1.2 High dimensional geometry

Figure 1.8. Regular simplices in two, three and four dimensions. For $\Delta_2$ we also show the insphere, the circumsphere, and the angle discussed in the text.

having $\Delta_{n-1}$ as its base. It is then not difficult to show that

$$R_n = nr_n = \sqrt{\frac{n}{2(n+1)}}$$

and

$$r_n = \sqrt{\frac{1}{2(n+1)n}} ,$$

so their ratio grows linearly, $\zeta = R_n/r_n = n$. The volume of a cone is $V = Bh/n$, where $B$ is the area of the base, $h$ is the height of the cone and $n$ is the dimension. For the simplex we obtain

$$\text{vol}(\Delta_n) = \frac{1}{n!} \sqrt{\frac{n+1}{2^n}}$$

We can check that the ratio of the volume of the largest inscribed sphere to the volume of the simplex goes to zero. Hence most of the volume of the simplex sits in its corners, as expected. The angle $\chi_n$ subtended by an edge as viewed from the center is given by

$$\sin \frac{\chi_n}{2} = \frac{1}{2R_n} = \sqrt{\frac{n+1}{2n}} \quad \Leftrightarrow \quad \cos \chi_n = \frac{1}{n} ;$$

When $n$ is large we see that $\chi_n$ tends to a right angle. This is as it should be. The corners sit on the outsphere, and for large $n$ almost all the volume of the circumsphere lies close to the equator—hence, if we pick one corner and let it play the role of the North Pole, all the other corners are likely to lie close to the equator. Finally it is interesting to observe that it is known for convex bodies in general that the radius of the circumsphere is bounded by

$$R_n \leq L \sqrt{\frac{n}{2(n+1)}} ,$$

where $L$ is the length of the longest line segment $L$ contained in the body. The regular simplex saturates this bound.

The effects of increasing dimension are clearly seen if we look at the ratio between surface (hyper) area and volume for bodies of various shapes. Rather than fixing the scale, let us study the dimensionless quantities $\zeta_n = R_n/r_n$ and $\eta(X) \equiv R \frac{\text{vol}(\partial X)}{\text{vol}(X)}$, where $X$ is the body, $\partial X$ its boundary, and $R$ its
outradius. For \(n\)-balls we receive
\[
\eta_n(B^n) = R \frac{\text{vol}(\partial B^n)}{\text{vol}(B^n)} = R \frac{\text{vol}(S^{n-1})}{\text{vol}(B^n)} = \frac{Rn}{R} = n.
\] (1.28)

Next consider a hypercube of edge length \(L\). Its boundary consists of \(2n\) facets, that are themselves hypercubes of dimension \(n - 1\). This gives
\[
\eta_n(\Box_n) = R \frac{\text{vol}(\partial \Box_n)}{\text{vol}(\Box_n)} = \frac{\sqrt{n}L 2n \text{vol}(\Box_{n-1})}{2 \text{vol}(\Box_n)} = \frac{n^{3/2}L}{L} = n^{3/2}.
\] (1.29)

A regular simplex of edge length \(L\) has a boundary consisting of \(n + 1\) regular simplices of dimension \(n - 1\). We obtain the ratio
\[
\eta_n(\Delta_n) = R \frac{\text{vol}(\partial \Delta_n)}{\text{vol}(\Delta_n)} = L \sqrt{n \frac{(n+1)\text{vol}(\Delta_{n-1})}{\text{vol}(\Delta_n)}} = n^2.
\] (1.30)

In this case the ratio \(\eta_n\) grows quadratically with \(n\), reflecting the fact that simplices have sharper corners than those of the cube.

The reader may know about the five regular Platonic solids in three dimensions. When \(n > 4\) there are only three kinds of regular solids, namely the simplex, the hypercube, and the cross-polytope. The latter is the generalization to arbitrary dimension of the octahedron. It is dual to the cube; while the cube has \(2^n\) corners and \(2^n\) facets, the cross-polytope has \(2^n\) corners and \(2^n\) facets. The two polytopes have the same values of \(\zeta_n\) and \(\eta_n\).

These results are collected in Tab. 14.2. We observe that \(\eta_n = n \zeta_n\) for all these bodies. There is a reason for this. When Archimedes computed volumes, he did so by breaking them up into cones and using the formula \(V = Bh/n\), where \(V\) is the volume of the cone \(B\) is the area of its base. Then we get
\[
\eta_n = R \frac{\sum \text{cones } B}{(\sum \text{cones } B) h/n} = \frac{nR}{h}.
\] (1.31)

If the height \(h\) of the cones is equal to the inradius of the body, the result follows.\(^3\)

### 1.3 Colour theory

How do convex sets arise? An instructive example occurs in colour theory, and more particularly in the psychophysical theory of colour. (This means that we will leave out the interesting questions about how our brain actually processes the visual information until it becomes a percept.) In a way tradition suggests that colour theory should be studied before quantum mechanics, because this is what Schrödinger was doing before inventing his wave equation.\(^4\) The object of our attention is colour space, whose points are the colours. Naturally one might worry that the space of colours may differ from person to person but in

---

\(^3\) For more information on the subject of this section, consult Ball (1997) \[?\]. For a discussion of rotations in higher dimensions consult section 8.3.

\(^4\) Schrödinger (1926) wrote a splendid review of the subject \[?\]. Readers who want a more recent discussion may enjoy the book by Williamson and Cummins (1983) \[?\].
1.3 Colour theory

Figure 1.9. Left: the chromaticity diagram, and the part of it that can be obtained by mixing red, green and blue. Right: when the total illumination is taken into account, colour space becomes a convex cone.

fact it does not. The perception of colour is remarkably universal for human beings, colour blind persons not included. What has been done experimentally is to shine mixtures of light of different colours on white screens; say that three reference colours consisting of red, green and blue light are chosen. Then what one finds is that by adjusting the mixture of these colours the observer will be unable to distinguish the resulting mixture from a given colour $C$. To simplify matters, suppose that the overall brightness has been normalized in some way. Then a colour $C$ is a point on a two dimensional chromaticity diagram. Its position is determined by the equation

$$C = \lambda_0 R + \lambda_1 G + \lambda_2 B.$$  \hspace{1cm} (1.32)

The barycentric coordinates $\lambda_i$ will naturally take positive values only in this experiment. This means that we only get colours inside the triangle spanned by the reference colours $R$, $G$ and $B$. Note that the “feeling of redness” does not enter into the experiment at all.

But colour space is not a simplex, as designers of TV screens learn to their chagrin. There will always be colours $C'$ that cannot be reproduced as a mixture of three given reference colours. To get out of this difficulty one shines a certain amount of red (say) on the sample to be measured. If the result is indistinguishable from some mixture of $G$ and $B$ then $C'$ is determined by the equation

$$C' + \lambda_0 R = \lambda_1 G + \lambda_2 B.$$  \hspace{1cm} (1.33)

If not, repeat with $R$ replaced by $G$ or $B$. If necessary, move one more colour to the left hand side. The empirical finding is that all colours can be assigned a position on the chromaticity diagram in this way. If we take the overall intensity into account we find that the full colour space is a three dimensional convex cone with the chromaticity diagram as its base and complete darkness as its apex (of course this is to the extent that we ignore the fact that very intense light will cause the eyes to boil rather than make them see a colour). The pure colours are those that cannot be obtained as a mixture of different
colours; they form the curved part of the boundary. The boundary also has a planar part made of purple.

How can we begin to explain all this? We know that light can be characterized by its spectral distribution, which is some positive function \( I \) of the wave length \( \lambda \). It is therefore immediately apparent that the space of spectral distributions is a convex cone, and in fact an infinite dimensional convex cone since a general spectral distribution \( I(\lambda) \) can be defined as a convex combination

\[
I(\lambda) = \int d\lambda' I(\lambda') \delta(\lambda - \lambda') , \quad I(\lambda') \geq 0 .
\] (1.34)

The delta functions are the pure states. But colour space is only three dimensional. The reason is that the eye will assign the same colour to many different spectral distributions. A given colour corresponds to an equivalence class of spectral distributions, and the dimension of colour space will be given by the dimension of the space of equivalence classes. Let us denote the equivalence classes by \( [I(\lambda)] \), and the space of equivalence classes as colour space. Since we know that colours can be mixed to produce a third quite definite colour, the equivalence classes must be formed in such a way that the equation

\[
[I(\lambda)] = [I_1(\lambda)] + [I_2(\lambda)]
\] (1.35)

is well defined. The point here is that whatever representatives of \([I_1(\lambda)]\) and \([I_2(\lambda)]\) we choose we always obtain a spectral distribution belonging to the same equivalence class \([I(\lambda)]\). We would like to understand how this can be so.

In order to proceed it will be necessary to have an idea about how the eye detects light (especially so since the perception of sound is known to work in a quite different way). It is reasonable—and indeed true—to expect that there are chemical substances in the eye with different sensitivities. Suppose for the sake of the argument that there are three such “detectors”. Each has an adsorption curve \( A_i(\lambda) \). These curves are allowed to overlap; in fact they do. Given a spectral distribution each detector then gives an output

\[
c_i = \int d\lambda \ I(\lambda) A_i(\lambda) .
\] (1.36)

Our three detectors will give us only three real numbers to parametrize the space of colours. Equation (1.35) can now be derived. According to this theory, colour space will inherit the property of being a convex cone from the space of spectral distributions. The pure states will be those equivalence classes that contain the pure spectral distributions. On the other hand the dimension of colour space will be determined by the number of detectors, and not by the nature of the pure states. This is where colour blind persons come in; they are missing one or two detectors and their experiences can be predicted by the theory. By the way, frogs apparently enjoy a colour space of four dimensions while lions make do with one.

Like any convex set, colour space is a subset of an affine space and the convex structure does not single out any natural metric. Nevertheless colour...
1.3 Colour theory

Figure 1.10. To the left, we see the MacAdam ellipses, taken from MacAdam, *Journal of the Optical Society of America* 32, p. 247 (1942). They show the points where the colour is just distinguishable from the colour at the center of the ellipse. Their size is exaggerated by a factor of ten. To the right, we see how these ellipses can be used to define the length of curves on the chromaticity diagram—the two curves shown have the same length.

space does have a natural metric. The idea is to draw surfaces around every point in colour space, determined by the requirement that colours situated on the surfaces are just distinguishable from the colour at the original point by an observer. In the chromaticity diagram the resulting curves are known as *MacAdam ellipses*. We can now introduce a metric on the chromaticity diagram which ensures that the MacAdam ellipses are circles of a standard size. This metric is called the *colour metric*, and it is curved. The distance between two colours as measured by the colour metric is a measure of how easy it is to distinguish the given colours. On the other hand this natural metric has nothing to do with the convex structure *per se*.

Let us be careful about the logic that underlies the colour metric. The colour metric is defined so that the MacAdam ellipses are circles of radius $\epsilon$, say. Evidently we would like to consider the limit when $\epsilon$ goes to zero (by introducing increasingly sensitive observers), but unfortunately there is no experimental justification for this here. We can go on to define the length of a curve in colour space as the smallest number of MacAdam ellipses that is needed to completely cover the curve. This gives us a natural notion of distance between any two points in colour space since there will be a curve between them of shortest length (and it will be uniquely defined, at least if the distance is not too large). Such a curve is called a *geodesic*. The *geodesic distance* between two points is then the length of the geodesic that connects them. This is how distances are defined in Riemannian geometry, but it is
worthwhile to observe that only the “local” distance as defined by the metric has a clear operational significance here. There are many lessons from colour theory that are of interest in quantum mechanics, not least that the structure of the convex set is determined by the nature of the detectors.

1.4 What is “distance”?

In colour space distances are used to quantify distinguishability. Although our use of distances will mostly be in a similar vein, they have many other uses too—for instance, to prove convergence for iterative algorithms. But what are they? Though we expect the reader to have her share of inborn intuition about the nature of geometry, a few indications of how this can be made more precise are in order. Let us begin by defining a distance $D(x, y)$ between two points in a vector space (or more generally, in an affine space). This is a function of the two points that obeys three axioms:

i) The distance between two points is a non-negative number $D(x, y)$ that equals zero if and only if the points coincide.

ii) It is symmetric in the sense that $D(x, y) = D(y, x)$.

iii) It satisfies the triangle inequality $D(x, y) \leq D(x, z) + D(z, y)$.

Actually both axiom ii) and axiom iii) can be relaxed—we will see what can be done without them in section 2.3—but as is often the case it is even more interesting to try to restrict the definition further, and this is the direction that we are heading in now. We want a notion of distance that meshes naturally with convex sets, and for this purpose we add a fourth axiom:

iv) It obeys $D(\lambda x, \lambda y) = \lambda D(x, y)$ for non-negative numbers $\lambda$.

A distance function obeying this property is known as a Minkowski distance. Two important consequences follow, neither of them difficult to prove. First, any convex combination of two vectors becomes a metric straight line in the sense that

$$z = \lambda x + (1-\lambda)y \quad \Rightarrow \quad D(x, y) = D(x, z) + D(z, y) , \quad 0 \leq \lambda \leq 1 . \quad (1.37)$$

Second, if we define a unit ball with respect to a Minkowski distance we find that such a ball is always a convex set.

Let us discuss the last point in a little more detail. A Minkowski metric is naturally defined in terms of a norm on a vector space, that is a real valued function $| | x | |$ that obeys

i) $| | x | | \geq 0$ , and $| | x | | = 0 \iff x = 0$ .

ii) $| | x + y | | \leq | | x | | + | | y | |$ . \quad (1.38)

iii) $| | \lambda x | | = | \lambda | | x | |$ , $\lambda \in \mathbb{R}$ .
1.4 What is “distance”? *17*

The distance between two points $x$ and $y$ is now defined as $D(x, y) \equiv \|x - y\|$, and indeed it has the properties i-iv. The unit ball is the set of vectors $x$ such that $\|x\| \leq 1$, and it is easy to see that

$$\|x\|, \|y\| \leq 1 \implies \|\lambda x + (1 - \lambda)y\| \leq 1.$$  \hfill (1.39)

So the unit ball is convex. In fact the story can be turned around at this point—any centrally symmetric convex body can serve as the unit ball for a norm, and hence it defines a distance. (A centrally symmetric convex body $K$ has the property that, for some choice of origin, $x \in K \implies -x \in K$.) Thus the opinion that balls are round is revealed as an unfounded prejudice. It may be helpful to recall that water droplets are spherical because they minimize their surface energy. If we want to understand the growth of crystals in the same terms, we must use a notion of distance that takes into account that the surface energy depends on direction.

We need a set of norms to play with, so we define the $l_p$-norm of a vector by

$$\|x\|_p \equiv (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{\frac{1}{p}}, \quad p \geq 1.$$  \hfill (1.40)

In the limit we obtain the Chebyshev norm $\|x\|_\infty = \max_i x_i$. The proof of the triangle inequality is non-trivial and uses Hölder’s inequality

$$\sum_{i=1}^N |x_i y_i| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1,$$  \hfill (1.41)

where $p, q \geq 1$. For $p = 2$ this is the Cauchy-Schwarz inequality. If $p < 1$ there is no Hölder inequality, and the triangle inequality fails. We can easily draw a picture (namely Fig. 1.11) of the unit balls $B_p$ for a few values of $p$, and we see that they interpolate between a hypercube (for $p \to \infty$) and a cross-polytope (for $p = 1$), and that they fail to be convex for $p < 1$. We also see that in general these balls are not invariant under rotations, as expected because the components of the vector in a special basis were used in the definition. The topology induced by the $l_p$-norms is the same, regardless of $p$. The corresponding distances $D_p(x, y) \equiv \|x - y\|_p$ are known as the $l_p$-distances.

Depending on circumstances, different choices of $p$ may be particularly relevant. The case $p = 1$ is relevant if motion is confined to a rectangular grid (say, if you are a taxi driver on Manhattan). As we will see (in section 13.1) it is also of particular relevance to us. It has the slightly awkward property that the shortest path between two points is not uniquely defined. Taxi drivers know this, but may not be aware of the fact that it happens only because the unit ball is a polytope, i.e., it is convex but not strictly convex. The $l_1$-distance goes under many names: taxi cab, Kolmogorov, or variational distance.

The case $p = 2$ is consistent with Pythagoras’ theorem and is the most useful choice in everyday life; it was singled out for special attention by Riemann when he made the foundations for differential geometry. Indeed we used a $p = 2$ norm when we defined the colour metric at the end of section 1.3.
Figure 1.11. Left: points at distance 1 from the origin, using the $l_1$-norm for the vectors (the inner square), the $l_2$-norm (the circle) and the $l_\infty$-norm (the outer square). The $l_{\frac{1}{2}}$-case is shown dashed—the corresponding ball is not convex because the triangle inequality fails, so it is not a norm. Right: in three dimensions one obtains respectively an octahedron, a sphere and a cube. We illustrate the $p = 1$ case.

The idea is that once we have some coordinates to describe colour space then the MacAdam ellipse surrounding a point is given by a quadratic form in the coordinates. The interesting thing—that did not escape Riemann—is the ease with which this “infinitesimal” notion of distance can be converted into the notion of geodesic distance between arbitrary points. (A similar generalization based on other $l_p$-distances exists and is called Finslerian geometry, as opposed to the Riemannian geometry based on $p = 2$.)

Riemann began by defining what we now call differentiable manifolds of arbitrary dimension\footnote{Riemann lectured on the hypotheses which lie at the foundations of geometry in 1854, in order to be admitted as a Dozent at Göttingen. As Riemann says, only two instances of continuous manifolds were known from everyday life at the time: the space of locations of physical objects, and the space of colours. In spite of this he gave an essentially complete sketch of the foundations of modern geometry. For a more detailed account see (for instance) Murray and Rice (1993) [?]; a very readable albeit old-fashioned account is by our Founding Father: Schrödinger (1950) [?]. For beginners the definitions in this section can become bewildering; if so our advice is to ignore them, and look at some examples of curved spaces first.}; for our purposes here let us just say that this is something that locally looks like $\mathbb{R}^n$ in the sense that it has open sets, continuous functions, and differentiable functions; one can set up a one-to-one correspondence between the points in some open set and $n$ numbers $\theta^i$, called coordinates, that belong to some open set in $\mathbb{R}^n$. There exists a tangent space $T_q$ at every point $q$ in the manifold; intuitively we can think of the manifold as some curved surface in space and of a tangent space as a flat plane touching the surface in some point. By definition the tangent space $T_q$ is the vector space whose elements are tangent vectors at $q$, and a tangent vector at a point of a differentiable manifold is defined as the tangent vector of a smooth curve passing through the point. Intuitively, it is a little arrow sitting at the point. Formally, it is a contravariant vector (with index upstairs). Each tangent vector $V^i$ gives rise to a directional derivative $\sum_i V^i \partial_i$ acting on the functions on the space; in differential geometry it has therefore become customary to think of a tangent vector as a derivative operator. In particular we can take the derivatives in...
1.4 What is “distance”? 

the directions of the coordinate lines, and any directional derivative can be expressed as a linear combination of these. Therefore, given any coordinate system \(\theta^i\) the derivatives \(\partial_i\) with respect to the coordinates form a basis for the tangent space—not necessarily the most convenient basis one can think of, but one that certainly exists. To sum up a tangent vector is written as

\[
V = \sum_i V^i \partial_i ,
\]

(1.42)

where \(V\) is the vector itself and \(V^i\) are the components of the vector in the coordinate basis spanned by the basis vectors \(\partial_i\).

It is perhaps as well to emphasize that the tangent space \(T_q\) at a point \(q\) bears no \textit{a priori} relation to the tangent space \(T_{q'}\) at a different point \(q'\), so that tangent vectors at different points cannot be compared unless additional structure is introduced. Such an additional structure is known as “parallel transport” or “covariant derivatives”, and will be discussed in section 3.2.

At every point \(q\) of the manifold there is also a \textit{cotangent space} \(T^*_q\), the vector space of linear maps from \(T_q\) to the real numbers. Its elements are called \textit{covariant vectors} and they have indices downstairs. Given a coordinate basis for \(T_q\) there is a natural basis for the cotangent space consisting of \(n\) covariant vectors \(d\theta^i\) defined by

\[
d\theta^i(\partial_j) = \delta^i_j ,
\]

(1.43)

with the Kronecker delta appearing on the right hand side. The tangent vector \(\partial_i\) points in the coordinate direction, while \(d\theta^i\) gives the level curves of the coordinate function. A general element of the cotangent space is also known as a \textit{one-form}. It can be expanded as \(U = U_i d\theta^i\), so that covariant vectors have indices downstairs. The linear map of a tangent vector \(V\) is given by

\[
U(V) = U_i d\theta^i(V^j \partial_j) = U_i V^j d\theta^i(\partial_j) = U_i V^i .
\]

(1.44)

From now on the \textit{Einstein summation convention} is in force, which means that if an index appears twice in the same term then summation over that index is implied. A natural next step is to introduce a scalar product in the
tangent space, and indeed in every tangent space. (One at each point of the manifold.) We can do this by specifying the scalar products of the basis vectors $\partial_i$. When this is done we have in fact defined a Riemannian metric tensor on the manifold, whose components in the coordinate basis are given by

$$g_{ij} = \langle \partial_i, \partial_j \rangle .$$

(1.45)

It is understood that this has been done at every point $q$, so the components of the metric tensor are really functions of the coordinates. The metric $g_{ij}$ is assumed to have an inverse $g^{ij}$. Once we have the metric it can be used to raise and lower indices in a standard way ($V_i = g_{ij}V^j$). Otherwise expressed it provides a canonical isomorphism between the tangent and cotangent spaces.

Riemann went on to show that one can always define coordinates on the manifold in such a way that the metric at any given point is diagonal and has vanishing first derivatives. In effect—provided that the metric tensor is a positive definite matrix, which we assume—the metric gives a 2-norm on the tangent space at that special point. Riemann also showed that in general it is not possible to find coordinates so that the metric takes this form everywhere; the obstruction that may make this impossible is measured by a quantity called the Riemann curvature tensor. It is a linear function of the second derivatives of the metric (and will make its appearance in section 3.2). The space is said to be flat if and only if the Riemann tensor vanishes, which is if and only if coordinates can be found so that the metric takes the same diagonal form everywhere. The 2-norm was singled out by Riemann precisely because his grandiose generalization of geometry to the case of arbitrary differentiable manifolds works much better if $p = 2$.

With a metric tensor at hand we can define the length of an arbitrary curve $x^i = x^i(t)$ in the manifold as the integral

$$\int ds = \int \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}} dt$$

(1.46)

along the curve. The shortest curve between two points is called a geodesic, and we are in a position to define the geodesic distance between the two points just as we did at the end of section 1.3. The geodesic distance obeys the axioms that we laid down for distance functions, so in this sense the metric tensor defines a distance. Moreover, at least as long as two points are reasonably close, the shortest path between them is unique.

One of the hallmarks of differential geometry is the ease with which the tensor formalism handles coordinate changes. Suppose we change to new coordinates $x^i = x^i(x)$. Provided that these functions are invertible the new coordinates are just as good as the old ones. More generally, the functions may be invertible only for some values of the original coordinates, in which case we have a pair of partially overlapping coordinate patches. It is elementary that

$$\partial_i = \frac{\partial x^j}{\partial x^i} \partial_j .$$

(1.47)
1.4 What is “distance”?

Since the vector $V$ itself is not affected by the coordinate change—which is after all just some equivalent new description—eq. (1.42) implies that its components must change according to

$$V^i \partial_i = V^i \partial_i \quad \Rightarrow \quad V^i(x') = \frac{\partial x'^i}{\partial x^j} V^j(x) .$$

(1.48)

In the same way we can derive how the components of the metric change when the coordinate system changes, using the fact that the scalar product of two vectors is a scalar quantity that does not depend on the coordinates:

$$g_{i'j'} U^{i'} V^{j'} = g_{ij} U^i V^j \quad \Rightarrow \quad g_{i'j'} = \frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j} g_{kl} .$$

(1.49)

We see that the components of a tensor, in some basis, depend on that particular and arbitrary basis. This is why they are often regarded with feelings bordering on contempt by professionals, who insist on using “coordinate free methods” and think that “coordinate systems do not matter”. But in practice few things are more useful than a well chosen coordinate system. And the tensor formalism is tailor made to construct scalar quantities invariant under coordinate changes.

In particular the formalism provides invariant measures that can be used to define lengths, areas, volumes, and so on, in a way that is independent of the choice of coordinate system. This is because the square root of the determinant of the metric tensor, $\sqrt{g}$, transforms in a special way under coordinate transformations:

$$\sqrt{g'}(x') = \left( \det \frac{\partial x'}{\partial x} \right)^{-1} \sqrt{g}(x) .$$

(1.50)

The integral of a scalar function $f'(x') = f(x)$, over some manifold $\mathcal{M}$, then behaves as

$$I = \int_{\mathcal{M}} f'(x') \sqrt{g'}(x') \, d^n x' = \int_{\mathcal{M}} f(x) \sqrt{g}(x) \, d^n x$$

(1.51)

—the transformation of $\sqrt{g}$ compensates for the transformation of $d^n x$, so that the measure $\sqrt{g} d^n x$ is invariant. A submanifold can always be locally defined via equations of the general form $x = x(x')$, where $x'$ are intrinsic coordinates on the submanifold and $x$ are coordinates on the embedding space in which it sits. In this way eq. (1.49) can be used to define an induced metric on the submanifold, and hence an invariant measure as well. Eq. (1.46) is in fact an example of this construction—and it is good to know that the geodesic distance between two points is independent of the coordinate system.

Since this is not a textbook on differential geometry we leave these matters here, except that we want to draw attention to some possible ambiguities. First there is an ambiguity of notation. The metric is often presented in terms of the squared line element,

$$ds^2 = g_{ij} dx^i dx^j .$$

(1.52)
The ambiguity is this: in modern notation $dx^i$ denotes a basis vector in cotangent space, and $ds^2$ is a linear operator acting on the tensor product $T \otimes T$. There is also an old fashioned way of reading the formula, which regards $ds^2$ as the length squared of that tangent vector whose components (at the point with coordinates $x$) are $dx^i$. A modern mathematician would be appalled by this, rewrite it as $g_x(ds, ds)$, and change the label $ds$ for the tangent vector to, say, $A$. But a liberal reader will be able to read eq. (1.52) in both ways. The old fashioned notation has the advantage that we can regard $ds$ as the distance between two “nearby” points given by the coordinates $x$ and $x + dx$; their distance is equal to $ds$ plus terms of order higher than two in the coordinate differences. We then see that there are ambiguities present in the notion of distance too. To take the sphere as an example, we can define a distance function by means of geodesic distance. But we can also define the distance between two points as the length of a chord connecting the two points, and the latter definition is consistent with our axioms for distance functions. Moreover both definitions are consistent with the metric, in the sense that the distances between two nearby points will agree to second order. However, in this book we will usually regard it as understood that once we have a metric we are going to use the geodesic distance to measure the distance between two arbitrary points.

### 1.5 Probability and statistics

The reader has probably surmised that our interest in convex sets has to do with their use in statistics. It is not our intention to explain the notion of probability, not even to the extent that we tried to explain colour. We are quite happy with the Kolmogorov axioms, that define probability as a suitably normalized positive measure on some set $\Omega$. If the set of points is finite, this is simply a finite set of positive numbers adding up to one. Now there are many viewpoints on what the meaning of it all may be, in terms of frequencies, propensities, and degrees of reasonable beliefs. We do not have to take a position on these matters here because the geometry of probability
distributions is invariant under changes of interpretation. We do need to fix some terminology however, and will proceed to do so.

Consider an experiment that can yield $N$ possible outcomes, or in mathematical terms a random variable $X$ that can take $N$ possible values $x_i$ belonging to a sample space $\Omega$, which in this case is a discrete set of points. The probabilities for the respective outcomes are

$$P(X = x_i) = p_i.$$  \hspace{1cm} (1.53)

For many purposes the actual outcomes can be ignored. The interest centers on the probability distribution $P(X)$ considered as the set of $N$ real numbers $p^i$ such that

$$p^i \geq 0, \quad \sum_{i=1}^{N} p^i = 1.$$  \hspace{1cm} (1.54)

(We will sometimes be a little ambiguous about whether the index should be up or down—although they should be upstairs according to the rules of differential geometry.) Now look at the space of all possible probability distributions for the given random variable. This is a simplex with the $p^i$ playing the role of barycentric coordinates; a convex set of the simplest possible kind. The pure states are those for which the outcome is certain, so that one of the $p^i$ is equal to one. The pure states sit at the corners of the simplex and hence they form a zero dimensional subset of its boundary. In fact the space of pure states is isomorphic to the sample space. As long as we keep to the case of a finite number of outcomes—the multinomial probability distribution as it is known in probability theory—nothing could be simpler.

Except that, as a subset of an $n$ dimensional vector space, an $n$ dimensional simplex is a bit awkward to describe using Cartesian coordinates. Frequently it is more convenient to regard it as a subset of an $N = n + 1$ dimensional vector space instead, and use the unrestricted $p^i$ to label the axes. Then we can use the $l_p$-norms to define distances. The meaning of this will be discussed in chapter 2; meanwhile we observe that the probability simplex lies somewhat askew in the vector space, and we find it convenient to adjust the definition a little. From now on we set

$$D_p(P, Q) \equiv ||P - Q||_p \equiv \left(\frac{1}{2} \sum_{i=1}^{N} |p_i - q_i|^p \right)^{\frac{1}{p}}, \quad 1 \leq p .$$  \hspace{1cm} (1.55)

The extra factor of $1/2$ ensures that the edge lengths of the simplex equal 1, and also has the pleasant consequence that all the $l_p$-distances agree when $N = 2$. However, it is a little tricky to see what the $l_p$-balls look like inside the probability simplex. The case $p = 1$, which is actually important to us, is illustrated in Fig. 1.5: we are looking at the intersection of a cross-polytope

---

6 The reader may consult the book by von Mises [?] for one position, and the book by Jaynes [?] for another. Jaynes regards probability as quantifying the degree to which a proposition is plausible, and finds that $\sqrt{\frac{1}{p}}$ has a status equally fundamental as that of $p_i$. 

Figure 1.14. For $N = 2$ we show why all the $l_0$-distances agree when the definition (1.55) is used. For $N = 3$ the $l_1$-distance gives hexagonal “spheres”, arising as the intersection of the simplex with an octahedron. For $N = 4$ the same construction gives an Archimedean solid known as the cub-octahedron.

with the probability simplex. The result is a convex body with $N(N - 1)$ corners. For $N = 2$ it is a hexagon, for $N = 3$ a cuboctahedron, and so on.

The $l_1$-distance has the interesting property that probability distributions with orthogonal support—meaning that the product $p_i q_i$ vanishes for each value of $i$—are at maximal distance from each other. One can use this observation to show, without too much effort, that the ratio of the radii of the in- and outspheres for the $l_1$-ball obeys

$$\frac{r_{\text{in}}}{R_{\text{out}}} = \sqrt{\frac{2}{N}} \text{ if } N \text{ is even}, \quad \frac{r_{\text{in}}}{R_{\text{out}}} = \sqrt{\frac{2N}{N^2 - 1}} \text{ if } N \text{ is odd}. \quad (1.56)$$

Hence, although some corners have been “chopped off”, the body is only marginally more spherical than is the cross-polytope. Another way to say the same thing is that, with our normalization, $||p||_1 \leq ||p||_2 \leq R_{\text{out}} ||p||_1 / r_{\text{in}}$.

We end with some further definitions, that will put more strain on the notation. Suppose we have two random variables $X$ and $Y$ with $N$ and $M$ outcomes and described by the distributions $P_1$ and $P_2$, respectively. Then there is a joint probability distribution $P_{12}$ of the joint probabilities,

$$P_{12}(X = x_i, Y = y_j) = p_{ij}^{12}. \quad (1.57)$$

This is a set of $NM$ non-negative numbers summing to one. Note that it is not implied that $p_{ij}^{12} = p_i^1 p_j^2$; if this does happen the two random variables are said to be independent, otherwise they are correlated. More generally $K$ random variables are said to be independent if

$$p_{i_1 \ldots i_K} = p_1^{i_1} p_2^{i_2} \ldots p_K^{i_K}, \quad (1.58)$$

and we may write this schematically as $P_{12 \ldots K} = P_1 P_2 \ldots P_K$. A marginal distribution is obtained by summing over all possible outcomes for those random variables that we are not interested in. Thus a first order distribution, describing a single random variable, can be obtained as a marginal of a second order
distribution, describing two random variables jointly, by
\[ p_i^j = \sum_j p_{12}^{ij} . \tag{1.59} \]

There are also special probability distributions that deserve special names. Thus the \textit{uniform} distribution for a random variable with \( N \) outcomes is denoted by \( Q_{(N)} \) and the distributions where one outcome is certain are collectively denoted by \( Q_{(1)} \). The notation can be extended to include
\[ Q_{(M)} = (\frac{1}{M}, \frac{1}{M}, \ldots, 1, 0, \ldots, 0) , \tag{1.60} \]
with \( M \leq N \) and possibly with the components permuted.

With these preliminaries out of the way, we will devote chapter 2 to the study of the convex sets that arise in classical statistics, and the geometries that can be defined on them—in itself, a preparation for the quantum case.

\textbf{Problem 1.1} \textit{Helly’s theorem} states that if we have \( N \geq n + 1 \) convex sets in \( \mathbb{R}^n \) and if for every \( n + 1 \) of these convex sets we find that they share a point, then there is a point that belongs to all of the \( N \) convex sets. Show that this statement is false if the sets are not assumed to be convex.

\textbf{Problem 1.2} Compute the inradius and the outradius of a simplex, that is prove eq. (1.24).