Notes in Spherical Geometry

Herman Jaramillo

January 4, 2016

Contents

0.1 Introduction

Notes in spherical geometry...

A comment on notation. In plane geometry we to a segment between the points A, and B, with the symbol $\overline{AB}$, and to its measure as $m(\overline{AB})$, or just AB for short. In these notes I will avoid this notation and refer to a segment or its length as AB . The same applies for angles. An angle $\angle ABC$ will mean also the angle and its measure. If the context requires to separate its measure from the object we will clarify.

Chapter 1

Fundamentals

The context here is that of the unit sphere

We will use the following two identities from vector algebra.

The triple product, used to find the determinant of a 3 by 3 matrix M with columns A, B, and C is given by

$$\det M = \langle A\,,\,B \times C\rangle = \langle B\,,\,C \times A\rangle = \langle C\,,\,A \times B\rangle \tag{1.1}$$

and the $\epsilon - \delta$ [1] identity

$$(A \times B) \times C = \langle A\,,\,C\rangle B - \langle B\,,\,C\rangle A. \tag{1.2}$$

1.1 Basic Definitions

We introduce the definitions of: unit sphere, point, line, segment, antipode, north pole, south pole, equatorial line, angle between two lines, triangle, etc.

Definition 1 (Unit Sphere). *The* **unit sphere** *is the set of point in* $\mathbb{R}^3$ *with unit distance. That is:*

$$S^2 = \{x \in \mathbb{R}^3 : |x| = 1\}.$$

We assume that the reader is familiar with basic linear algebra. Mainly with dot and cross products of vectors.

Definition 2 (Point). *Any point* $x \in S^2$ *is defined as a* **point**

[1] http://www.luc.edu/faculty/dslavsk/courses/phys301/classnotes/proofepsilondelta.pdf

Definition 3 (**Line**)**.** *A line l is a* **great circle** *in the unit sphere S^2. That is, a great circle is defined based on a given point ξ, known as a* **pole** *. All vectors in S^2 orthogonal to ξ are sitting in the* **equator.** *Mathematically we write*

$$l = \{x \in S^2 : \langle x \,,\, \xi \rangle = 0\}.$$

The line l is connected to the point ξ by saying that l is the polar line of ξ (the equator), or ξ is the pole of l. There are two poles, since eather ξ and $-\xi$ generate the same polar line. We could say that the "+" could correspond to the north pole and the "−" to the south pole. The line in spherical geometry is also known as **geodesic** *.*

To compute the equatorial line that goes through A and C in Figure 1.1 we find the vector

$$P = \frac{A \times B}{\|A \times B\|}.$$

Then the equatorial line is the set of all points on S^2 orthogonal to P. Observe that if A and C are **antipodes** , that is, if they are two ends of the same diameter [2], the cross product of A and B is zero, and there is no a clear definition of P. This makes sense, there are an infinite number of circles corresponding to the two antipodes. The uniqueness of a segment or of a circle through a segment represented by the points A and B is limited to the fact that A and B should not be antipodes.

Definition 4 (Segment)**.** *Given two points $A \in S^2$ and $B \in S^2$, if they are not antipodes, there is a unit great circle through them. The points divide the circle into two segments. The shorter segment is called the* **minor segment** *and the other is the* **major segment** *. A segment is noted by the two vertices AB.*

In the context of this document we do not consider intersections of planes with the sphere such that those planes do not contain the center of the sphere. All segments are used along the geodesic line. We show (see section 1.5.1) that along the geodesic line the distance between two points in the sphere is the last. In this way the geodesics in the sphere play the role of lines in the plane.

Figure 1.1 illustrates the definitions above of point, line, great circle, segment, and pole, and equatorial line.

[2] We can define antipodes as two vectors in S^2 with opposite sign

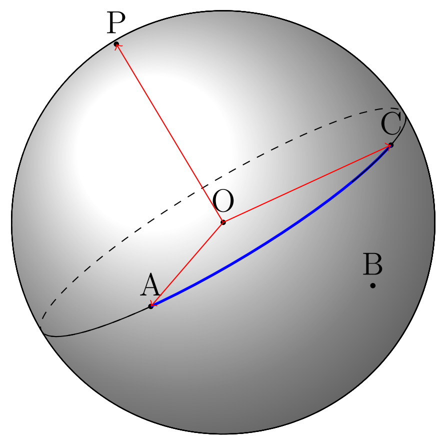


Figure 1.1: Illustration of A point “B”, a line (great circle through “A” and “C”. The great circle is computed from the two vectors “A” and “C” which cross product generates the pole “P”. All points in the great circle are orthogonal to the vector “P”. A segment is described by the blue arc between “A” and “C”.

Definition 5 (angle). *Let us assume that two segments AP and CP met at some point P. The great circles through AP and CP, are located in the intersection of two planes (one plane each) and the sphere. The dihedral angle of the planes is the* **angle between the segments.** *The angle between the segments AP and CP is noted as $\angle APC$.*

We are interested in angles $0 < \theta < \pi$. These generates non-degenerate convex objects. The angle π would correspond to a circle that divide the sphere in two hemispheres. We observe several anomalies that would be present if we let $\theta = \pi$.

- The two end points of the angle are opposite in the sphere. That is, they are antipodes. Then the angle is not uniquely defined in the sense that those two points could have legs all over the sphere. Each great circle through those two points is equally suited to be the angle.

- In a triangle with such an angle, we would have that two vertices are antipodes and they are not linearly independent which makes the triangle degenerate.

Note that the same applies for segments or sides of polygons. We do not want a segment go to π (in a unit sphere) since it generates the same indeterminacies described above for the angle.

We will use AP both for the segment and its length for short. Figure 1.2 illustrates the angle between segments $A_1 - C$ and $A_2 - C$.

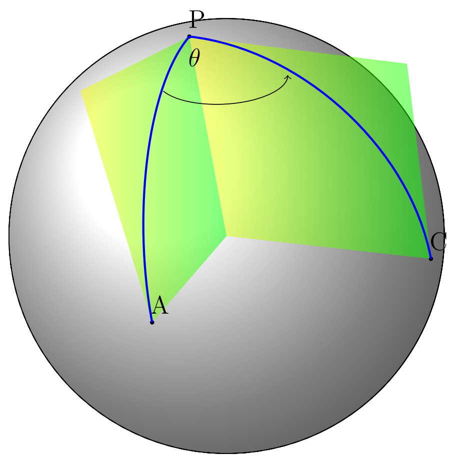


Figure 1.2: Illustration of the angle made by the segments AP and CP. The angle is measured by the dihedral angle formed by the planes containing the segments AP, and CP.

Let us see how to find the intersection of two segments.

A segment is part of a great circle and has the property that all its points are orthogonal to a given point ξ_1. If another segment intersect the first at a point P, then at P both ξ_1 and ξ_2, which is the pole for the second segment, should be orthogonal to P. The line that is orthogonal to both ξ_1 and ξ_2 has the direction $\xi_1 \times \xi_2$. The point on S^2 with this direction is

$$P = \pm \frac{\xi_1 \times \xi_2}{\|\xi_1 \times \xi_2\|}.$$

The $\pm$ is because at the antipode of P there occurs also an intersection.

We are now ready to compute measures of objects on the sphere. We start with the measure of distances and angles.

1.2 Measure of distance and angle

All segments on the sphere are pieces of a circle. Their measure is in the interval $[0, 2\pi]$. The extremes are for a point or a whole circumference (a line). Given that we can see the points of the sphere as unit vectors the measure of a segment from P to Q is computed with the formula

$$d(P, Q) = \cos^{-1}\langle P\, ,\, Q\rangle. \tag{1.3}$$

To measure the θ angle in Figure 1.2 we note that the measure of the dihedral angle is equal with the angle between the normals of the planes, which are the pole vectors, say ξ_1 and ξ_2. Since the pole vectors are given by

$$\xi_1 = \frac{A \times P}{\|A \times P\|} \qquad \xi_2 = \frac{C \times P}{\|C \times P\|}$$

then

$$\theta = \cos^{-1}\langle \xi_1\, ,\, \xi_2\rangle = \cos^{-1}\left\langle \frac{A \times P}{\|A \times P\|}\, ,\, \frac{C \times P}{\|C \times P\|} \right\rangle. \tag{1.4}$$

We now define spherical triangles and measures of the spherical angles on triangles.

Definition 6 (Spherical Triangles). *By* **spherical triangle** *we mean a triple A, B, and C such that they are not collinear. That is, no point can be an antipode of any other. In addition it is required that the segments forming the sides of the triangle are minor segments. The triangle defined here is noted as $\triangle ABC$.*

In Figure 1.3 we show that the triangle has the following elements.

(i) Three vertices : points A, B, and C.

(ii) Three segments $a = BC$, $b = AC$, and $c = BA$.

(iii) Three angles α, β, and γ

In general any figure formed by great circles in the sphere is formed twice in oposite (antipode) sides. For example the triangle in Figure 1.4 shows two triangles formed by three great circles. The triangles are and we can think

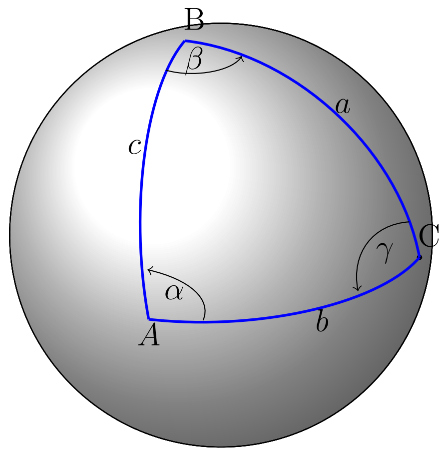


Figure 1.3: Illustration of a triangle formed by the vertices A, B , and C. The sides are noted as $a = BC, b = AC$, and $c = AB$, and the angles are $\angle BAC = \alpha, \angle ABC = \beta$, and $\angle ACB = \gamma$.

of them as mirrows of each other. When we say triangle we will ignore its spherical mirror image and refer to only one of the triangles in the figure. The same apply for any figure formed on the sphere.

We should clarify how the arc length in the triangle is computed and we use a figure to explain this. Figure 1.5 shows the dihedral angle at C in the triangle with vertices A, B, and C. This dihedral angle is called here γ. Opposite to that angle we find the side (segment) labeled as c. This angle is computed using the central angle also labeled here c. Since we are using radians and the sphere has radius 1 it is fine to call both the arc and central angle as c. A scaling of the sphere to radius other than 1 will require a rename of the arc (or the angle) to avoid confusions. You can imagine a similar figure where b and c are explained both as central angles and the other two sides of the triangle.

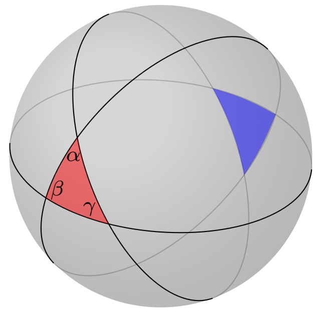


Figure 1.4: Two triangles as spherical mirror images of each other.

1.3 Dualities

Discussion: I rewrite here my post in the Mathematics StackExchange [3] website about duality:

"It is true that *'duality'* is a quite broad concept in many fields and the field of mathematics is not the exception. For example the Wikipedia website [4] shows many forms of duality in Mathematics. As mathematicians we want to abstract the meaning of a word into a unique concept that ecompasses all situations.

For duality we need to define two spaces of objects and an attribute (property) of those objects. Then establish a relation between objects of one space and the other thru this attribute. If this relation is unique we say that objects of one space are duals of objects in the other space. This is not anything else that the definition of a bijective function. Any duality in mathematics can be expressed as a bijective function between two spaces of objects. So $a \in A$ is dual of $b \in B$ if there is some relation f such that $b = f(a)$ and $a = f^{-1}(b)$ in a unique way.

Two properties should be always present in a duality:

(i) **Symmetry:** If a is dual of b, b is dual of a.

(ii) **Idempotence:** If d is dual operation then $d^2 = I$.In words, the dual of the dual is the original object.

[3] http://math.stackexchange.com/questions/364782/what-is-duality/1539669#1539669

[4] https://en.wikipedia.org/wiki/Duality_(mathematics)

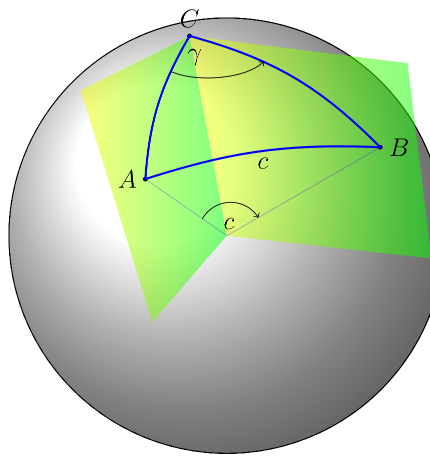


Figure 1.5: A dihedral angle γ and the central angle c which has the same measure in radians as the arc c in the triangle $\triangle ABC$.

If the two spaces of objects are the same, the function f described in the first paragraph is idemponent. That is $f^2 = I$. Otherwise we need two different functions f and f^{-1} to get back to the starting point. For example in linear algebra, in the space of square matrices, the transpose is an idempotent operator and so it is a dual operation which does not exit the space (it is closed). If a matrix is not square the transpose leaves in a different space and the transposition function is not idempotent anymore. This generalizes to functional analysis with the concept of adjoint. In spherical geometry a north pole is a dual of its equator but the objects do not leave in the same space: the first are points and the second "lines" in the sphere. But in spherical geometry triangles are dual of polar triangles and they live in the same space so the duality here is an idempotence.

Please observe that the dual in this discussion can be confused with the concept of inverse and a claritication is needed. In the language of the first paragraph it is, since we establish a bijection between objects and by definition of inverse, duality and inversion are linked together. However it is not an inverse in many respects. In the space of matrices, the transpose is a dual but not an inverse matrix and in general the adjoint is not the inverse for the more general context of functional analysis. Now, if we define a function in the space of square matrices as sending a matrix into its transpose, then the inverse of this function coincides with the concept of dual, but it is not the inverse of the matrix, so by "inverse" we need to be precise of what type of inverse we are looking for.

We then say that duality is a word attached to objects and spaces where those objects leave. We could say that two spaces are dual of each other if there is a bijective function between them. In this sense duality is an equivalence relation:

(i) **Reflexive:** Every space is dual to itself. The identity function is always a duality.

(ii) **Symmetric:** If a space A is dual to a space B, then a space B is dual to a space A. This is by definition since it exist a bijective function between the two spaces.

(iii) **Transitive:** Function composition of bijective functions.

"

There are many dualities in spherical geometry. For example the blue and red triangles in Figure 1.4 are dual of each other. In geometry the word "dual" is quite common. Going back tothe Figure 1.4, if we have an operator P that takes a triangle (red in the figure) into its spherical image (antipode triangle in blue), then the operator P applied to the antipode triangle (blue) should return back to the original triangle (red). That is $P^2 = I$. This particular duality between triangles is not very useful. Instead we will define below polar duality between triangles which will be used later on the text to prove trigonometrical identities.

1.3.1 The north and the equator

Given any point in the sphere we could think of its antipode as a dual point, since the antipode of the antipode is the original point. However that duality is not very important. Instead we consider the dual pair north and its equator. Given a north there is one and only one equator; mainly all the points on the sphere orthogonal to the north vector. Now, given an equator there should be only one north. The is a south as well. We think of the north using the right hand side rule. By imaging that we are grubbing the sphere from the equator with all fingers but the thumb. The direction where the thumb points is the north.

Figure 1.6 displays a few pairs of poles (with vector arrows pointing outside of the sphere) and their equatorial lines.

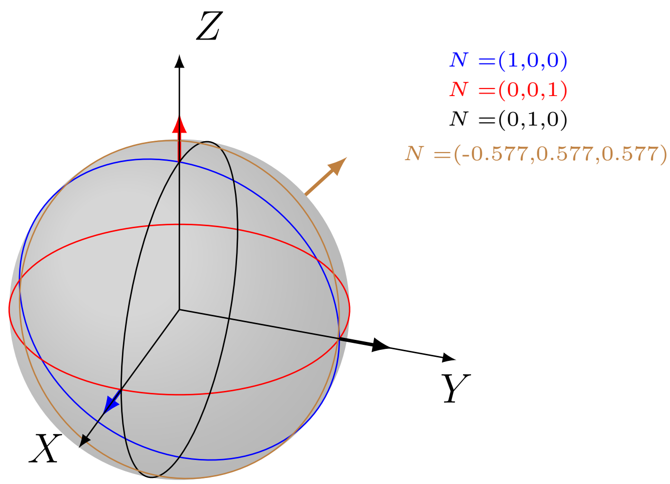


Figure 1.6: We show 4 north poles. Three on the main coordinate axis (blue, red, and black) and one on a 45 degree polar, 0 azimuthal line. The corresponding equators are color coded correspondingly with the north poles.

In which sense is a north pole dual of its equator? Recall the definition of line 3. Poles and lines are connected together. To each north pole there corresponds a unique equatorial line, and to each equator there is a unique north pole. While geometrically speaking poles and equators do not leave in the same space (one is a point, the other is a line in a sphere) they are connected in a unique way and knowing one we know the other. We do not say here that we have an operator P from one space to the other because the domains are different. In this case we can talk about two operators P and Q. The first from poles to equators and the second from equators to poles.

1.3.2 A segment and its dual: a lune

Let us assume now that we have a discrete set of points in a set A. To each point a we match an equator b in a set of equators B. The sets A and B are dual of each other in the sense of the previous section. We are not interested in arbitrary sets but rather in specific sets such as for example a segment on a sphere. This is a continuous finite path along a great circle. To each point of the segment we assign a great circle in a way that the point is a north pole and the circle is an equator. If we proceed in this way, the set of points that make the segment are mapped into a piece of surface of the sphere called a **lune** . As the segment goes from point A to point B in the sphere, the lune is a rotating great circle that starts at a given angle and ends at another given angle. Figure 1.7 illustrates the segment and its lune duality. The segment

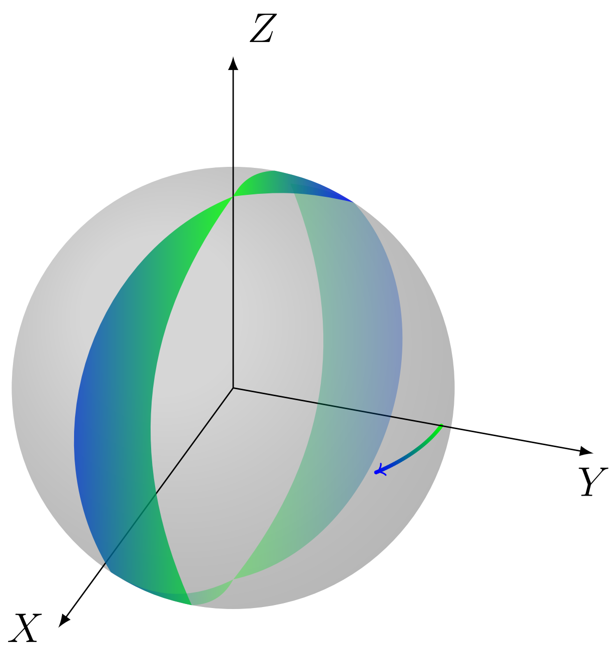


Figure 1.7: A segment and its dual a lune. Think of the firsts point of the segment as the point $(0, 1, 0)$ in the Y axis with green color. The equator corresponding to this point is the intersection of the XZ plane with the sphere, a green circle. Now if the segment is a small piece of arc rotating 30 degrees from the axis Y toward the axis X, as we move along this path, the vertical circle in the XZ plane starts rotating with the segment at the same speed toward a blue color. The shaded area on the sphere is a lune.

is transversed from the Y axis toward the X axis, the lune is created by the rotation of the vertical great circles as rotated with respect to the Z axis.

1.3.3 A triangle and its dual : a polar triangle

We divide this section into two main parts.

(i) Geometrical Insights.

(ii) Applications

- Linear Algebra
- Tensor Analysis
- Funtional Analysis

1.3.3.1 Geometrical Insights

Another important triangle is the **polar triangle** . Each side of a triangle defines a great circle which is an equator of some north pole point [5] Since the three sides of the triangle define three different planes (equators) their north poles (three of them) define a new triangle which is the polar triangle. Figure 1.8 illustrates a polar triangle (blue) corresponding to a given triangle (in red). Please note that for the side a, the north pole is located at $\alpha\prime$, and so for the other two sides.

We show that the polar triangle is a dual of the original triangle known also as the **prime** . By dual we mean that the polar triangle of the polar triangle is the prime triangle.

To better understand the duality we exploit the fact that a triangle, its polar image, and the relation between them is rotationally invariant. That is, when we want to understand properties for them we can choose a rotation (or coordinate system) that makes the problem easy. I include the following figure because : what is easier that start with a $90-90-90$ triangle with coordinates at the three unit vectors $(1,0,0),(0,1,0)$, and $(0,0,1)$ and then transform this "unit" triangle into a more general triangle while observing the effect on the polar triangle. The first thing that we note is that a $90-90-90$ triangle (equilateral) is equal to its polar triangle. This is easy to see because each vertex is the north pole of the equator passing through the other two vertices. This is shown in the green frame in Figure 1.3.3.1

We now want to deform or stretch this triangle to see how the dual (polar) triangle will change. If we keep the north pole at the angle fixed and start

[5] for each equator there are two poles but we can define the north pole by using the right hand side rule.

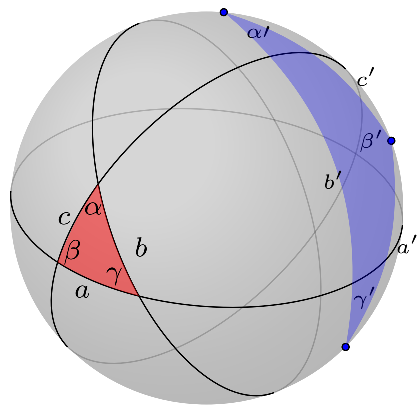


Figure 1.8: Given an spherical triangle (red) with angles α, β, and γ, and sides a, b, and c, the corresponding polar triangle is found (blue) with angles α', β', and γ', sides a', b' and c'. The sides of the polar triangle join the vertices along great circles. The vertex at α' is the north pole of the equator for the side a, the vertex at β' is the north pole for the equator having the side b, and the vertex at γ' is the north pole for the equator having the side c.

with an wider triangle with angles $-20°$ to $110°$ in the base (plane XY), we see that the dual (the prime is red, the dual is blue) narrows to an angle $20°$ to $70°$. So the angles α and α' become supplementary and the same happens for β and β'. Since $\gamma = \gamma'$ they are each $90°$ they are supplementary as well. The last case is in the bottom frame. We now allow the polar angle (and here by "polar angle" I mean the angle with respect to the Z the axis, called θ by physicist and ϕ by mathematicians) degrees and now the dual (polar) is tilted in a way that the prime angles and the dual angles are supplementary. We show a formal proof of this in Theorem 1.3.1

Again, properties such as fixed point objects, that is objects X such that if P is the map between a triangle and its polar image remain fixed, or in equations $P(X) = X$, exists and these are all equilateral $90-90-90$ triangles.

1.3.3.2 Applications

1.3.3.2.1 Linear Algebra The concept of dual basis is of great importance in linear algebra. We now formulate the definition of a polar triangle

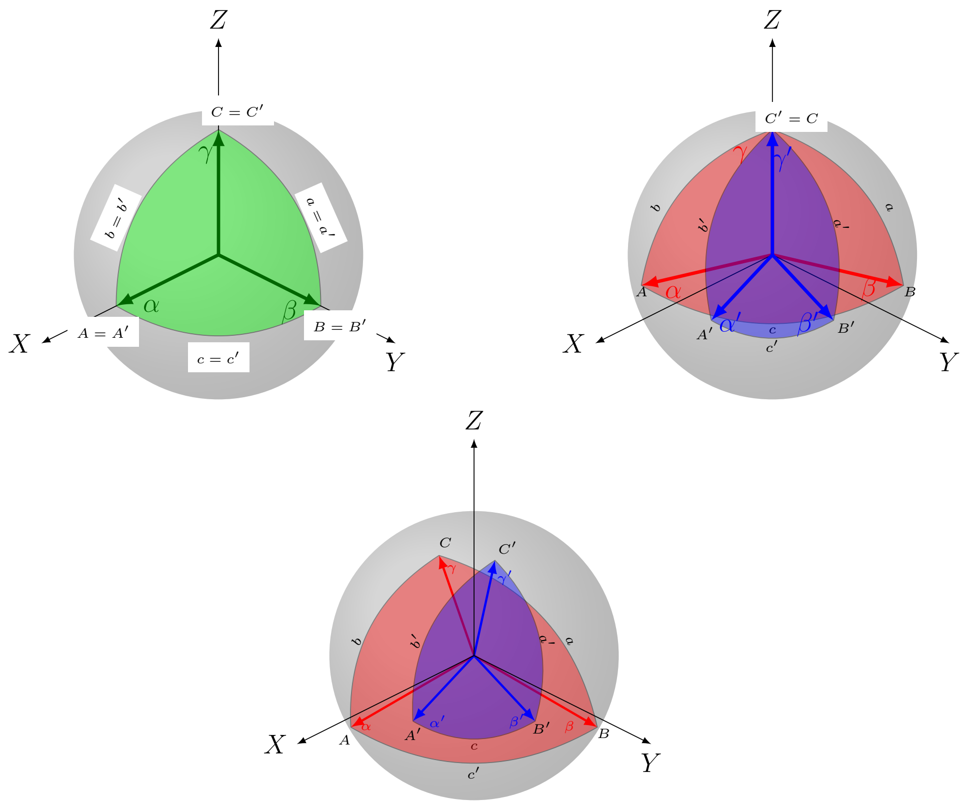


Figure 1.9: The top left figure shows a $90-90-90$ (equilateral) triangle which is equal to its polar dual. The second frame in top shows the effect of stretching the base of the triangle to angles between $-20°$ and $110°$ degrees. The north pole stays fixed but the polar triangle is not longer equal to the prime triangle. It is a shirnk version of the prime triangle in a way that the angles α, β, and γ remain complements of the arcs a', b', and c'. The frame in the bottom further stretches the angle along the polar (angle with respect to the Z axis) direction.

in terms of linear algebra:

Definition 7 (polar triangle). *Given three unit vectors A, B, and C in $\mathbb{R}^3$, linearly independent, they form a basis for $\mathbb{R}^3$. We can build a "dual" basis as follows:*

$$A' = \frac{B \times C}{\|B \times C\|} \quad B' = \frac{C \times A}{\|C \times A\|} \quad C' = \frac{A \times B}{\|A \times B\|} \tag{1.5}$$

the new basis $\{A, B, C\}$ constitutes a polar triangle.

Please observe the quotes around the word "dual" . They are dual in the sense of that the polar vectors of the polar vectors are the prime vectors. We show this in Theorem 1.3.1. They have the direction of the dual vectors in linear algebra. However the polar triangle (being in the unit sphere) has the vectors normalized. It is not hard to compute the scaling between the polar (vectors) triangle and the dual base vectors is $s_k = \|X_i \times X_j/V$ where $V = X_1 \cdot X_2 \times X_3$ is the volume of the parallelepiped formed by the three vectors X_1, X_2 and X_3 in the set $\{A, B, C\}$, and $i, j, k \in \{1, 2, 3\}$ no repeated. The scaling values s_k, $k = 1, 2, 3$ the link between a polar triangle and the dual basis of linear algebra.

1.3.3.2.2 Tensor Analysis In the field of tensors we have orthogonal coordinates which transform an space with no stretching or shear deformation, but we have general transformation which are no necessarily orthogonal. When the coordinates are orthogonal the concept of "covariant" or "contravariant" merge into one and the notation using sub-indices and super-indices is superfluos. This corresponds in spherical geometry to equilateral triangles where their polar duals are the same. In a general transformation which is not orthogonal we have two systems of coordinate axes. One is a covariant system and its (dual) is a contravariant. You will find that the axis of one system are those of a prime triangle (contravariant) while the axis of the other (the covariant) are given by the polar triangle. Here the notation such as δ_{ijk}^{pqr} with subindices and superindices is required.

1.3.3.2.3 Functional Analysis In the field of functional analysis: A space and its dual (the dual is the space of functionals –functions from the space to the field $\mathbb{R}$ or $\mathbb{C}$) are such that coordinate axes in one space (say covariant for example) map into coordinate axes in the other space (contravariant) in a way that correspond to the geometry of the angles (reducing the problem to $\mathbb{R}^3$ as a prototype) between the prime and the polar triangles respectively. Again, beware of scaling factors as explained above.

Theorem 1.3.1 (Duality between polar and prime triangles). *If one triangle is the polar triangle of another, the latter is the polar triangle of the former.*

Proof. We use the $\epsilon - \delta$ identity 1.2 together with the definition 7 to verify the dualities. To show that A, B, C are the duals of A', B', C' we need to show

$$A = \frac{B' \times C'}{\|B' \times C'\|} \quad B = \frac{C' \times A'}{\|C' \times A'\|} \quad C = \frac{A' \times B'}{\|A' \times B'\|}. \tag{1.6}$$

We show only the first of the three equations above since the other two are shown in similar form. Replacing B' and C' from equation 7 into the first equation above we see that

$$B' \times C' = (C \times A)/\|C \times A\| \times (A \times B)/\|A \times B\|$$

We now use the $\epsilon - \delta$ identity 1.2 to write

$$(C \times A) \times (A \times B) = \langle C, A \times B \rangle A - \cancelto{0}{\langle A, A \times B \rangle C}. \tag{1.7}$$

The cancellation to zero is due to the fact that A is orthogonal to $A \times B$, and so their inner product is 0. Now since the vectors A, B, and C are linearly independent, their determinant $\langle C, A \times B \rangle$ is non zero. Then

$$B' \times C' = \frac{\langle C, A \times B \rangle A}{\|C \times A\| \|A \times B\|}$$

No matter what the non-zero scale factor multiplying A here is, the expression

$$\frac{B' \times C'}{\|B' \times C'\|}$$

will cancel this scalar through normalization and so

$$\frac{B' \times C'}{\|B' \times C'\|} = A.$$

which proves the result for the first equation. The proof for the other two equations in 1.6 is similar and obtained by cyclic rotation of the symbols A, B and C. This is left to the reader. □

We now show the sum of each angle in the prime triangle with each polar corresponding arc is π, and likewise the sum of each arc in the prime triangle with an angle on the polar.

Theorem 1.3.2 (conservation sums). *the sum of arcs on the prime triangle with angles in polar triangle and vice-versa is π. That is arcs in prime are supplementary to angles in polar and arcs in polar are supplementary to angles in primary. In symbols this is:*

$$\alpha + a' = \beta + b' = \gamma + c' = a + \alpha' = b + \beta' = c + \beta' = \pi. \tag{1.8}$$

Proof. We recall the definitions of the angles and arc lengths in terms of the points A, B, C and A', B', C'. We start by the side a in the dual triangle. That is,

$$a' = \cos^{-1}\langle B', C'\rangle.$$

Now, from the definition of B' and C' in equation 1.5 we find that

$$\begin{aligned} a' &= \cos^{-1}\left\langle \frac{C \times A}{|C \times A|}, \frac{A \times B}{|A \times B|} \right\rangle \\ &= -\cos^{-1}\left\langle \frac{A \times C}{|C \times A|}, \frac{A \times B}{|A \times B|} \right\rangle \\ &= \pi - \cos^{-1}\left\langle \frac{A \times C}{|A \times C|}, \frac{A \times B}{|A \times B|} \right\rangle \\ &= \pi - \alpha. \end{aligned}$$

The other equalities are derived in similar fashion by cycling the symbols, and are left to the reader.

□

In view of the previous theorem polar triangles are also known as **suplmentary triangles** .

There are other dualities on the sphere of great importance. For example the Voronoi and Delaunay dual pairs on the sphere [6], which are out of the

[6] http://diginole.lib.fsu.edu/cgi/viewcontent.cgi?article=5571&context=etd

scope of this document, but of great usefullness in the field of computational geometry. [7] Another important set of dualities on the sphere are the Platonic solids and their duals (which happen to be Platonic solids as well), for which we devote chapter **??** of this document.

1.4 Spherical Trigonometry

It is hard to do anything in the sphere without refering to trigonometrical functions. We should cover this issue from the early stage of development.

We show here a few facts of spherical trigonometry. The Wikipedia Website [8] has much more information about spherical trigonometry. We will only deal here with the cosine and the sine laws, and a few important formulas for trigonometrical relations between triangle sides and angles.

1.4.1 The cosine and sine laws

There are laws which are extensions of the cosine and sine laws of plane geometry to spherical geometry. Let us start with the cosine law.

1.4.1.1 The cosine law

In plane geometry for a triangle $\triangle ABC$, with side lengths a, b, c, and angles α, β, and γ, we could show that

$$a^2 = b^2 - 2ab\cos\alpha + c^2, \tag{1.9}$$

where γ is the angle at the vertex A.

To find relationships between the lengths of the triangle sides and its angles, we start with equation 1.4, which we write in terms of α as

$$\cos\alpha = \left\langle \frac{A\times B}{\|A\times C\|}, \frac{A\times B}{\|A\times C\|} \right\rangle.$$

We now observe that $A\times B$ is the sin of the angle between the vectors A and B which delimit the segment c. We then have that $\|A\times B\| = \sin c$, and similarly $\|A\times C\| = \sin b$. Then we write

[7] https://en.wikipedia.org/wiki/Computational_geometry
[8] https://en.wikipedia.org/wiki/Spherical_trigonometry

$$\cos\alpha = \frac{1}{\sin b \sin c}\,\langle A\times B\ ,\ A\times C\rangle.$$

We now use the inner and cross product identities 1.1 and 1.2. Let us call $D = A\times B$, then

$$\begin{aligned}
\cos\alpha &= \frac{1}{\sin b \sin c}\,\langle A\times B\ ,\ A\times C\rangle \\
&= \frac{1}{\sin b \sin c}\,\langle D\ ,\ A\times C\rangle \\
&= \frac{1}{\sin b \sin c}\,\langle A\ ,\ C\times D\rangle \quad \text{from equation } \ 1.1 \\
&= \frac{1}{\sin b \sin c}\,\langle A\ ,\ C\times(A\times B)\rangle \\
&= \frac{1}{\sin b \sin c}\,\langle A\ ,\ (C\times A)\times B\rangle \quad \text{from associativity of cross product} \\
&= \frac{1}{\sin b \sin c}\,\langle A\ ,\langle C\ ,\ B\rangle A - \langle A\ ,\ B\rangle C\rangle \quad \text{from the } \epsilon-\delta \text{ identity } \ 1.2 \\
&= \frac{1}{\sin b \sin c}\,\langle C\ ,\ B\rangle\ \langle A\ ,\ A\rangle - \langle A\ ,\ B\rangle\langle A\ ,\ C\rangle \quad \text{distributive law of } \langle,\rangle \\
&= \frac{1}{\sin b \sin c}\,\langle C\ ,\ B\rangle\ - \langle A\ ,\ B\rangle\langle A\ ,\ C\rangle \quad \text{, since } \langle A\ ,\ A\rangle = 1.
\end{aligned}$$

We now recall from equation 1.3 that the inner product is the argument (inverse cosine) of the distance between the involved points in the segment. That is $\langle C\ ,\ B\rangle = \cos a$, and likewise $\langle A\ ,\ B\rangle = \cos c$, and $\langle A\ ,\ C\rangle = \cos b$. Then we find that

$$\cos\alpha = \frac{\cos a - \cos b \cos c}{\sin b\ \sin c}. \tag{1.10}$$

or

$$\cos a = \cos b \cos c + \cos\alpha \sin b \sin c. \tag{1.11}$$

which is the cosine law for spherical geometry. This formula does not resemble the Euclidean (plane) geometry for the cosine law provided by equation 1.9. However we will see that in the asymptotic approximation of large radius (or very small curvature) we will find how this formula approximates the cosine law equation 1.9.

For the very large radius [9] we can say that the central angles a, b, and c are very small and still the segment lengths a, b and c could be large (since the length of the segment is linearly amplified by the radius of the sphere). Then by using Taylor series approximations, we have

$$\cos a \approx 1 - \frac{a^2}{2}$$
$$\sin a \approx a - \frac{a^3}{6}.$$

So up to second order we can write equation 1.11 as

$$1 - \frac{a^2}{2} \approx 1 - \frac{1}{2}(b^2 + c^2) + bc\cos\alpha$$

That is

$$a^2 \approx b^2 + c^2 - 2ab\cos\alpha,$$

which corresponds with the cosine law 1.9 for plane geometry.

With help of equation 1.8 we can write a new version of the cosine law. This can be named the dual cosine law. That is, in a polar triangle using equation 1.10 we find

$$\cos(\pi - a) = \frac{\cos(\pi-\alpha) - \cos(\pi-\beta)\cos(\pi-\gamma)}{\sin(\pi-\beta)\sin(\pi-\gamma},$$

that is

$$\cos a = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\,\sin\gamma}.$$

In summary; we get a couple of (dual) equations that help us find the angle of a triangle given its sides or a side of a triangle given its angles.

[9]We assume that the radius of the sphere is 1 but we have not said which units we are using. For very large spheres you can use units of light years, but we do not have to go that far to make the flat approximation valid.

$$\cos\alpha = \frac{\cos a - \cos b\cos c}{\sin b\,\sin c} \quad , \quad \cos a = \frac{\cos\alpha + \cos\beta\cos\gamma}{\sin\beta\,\sin\gamma}. \tag{1.12}$$

or, separating $\cos a$ and $\cos\alpha$ from these two equations:

$$\begin{aligned} \cos a &= -\cos b\cos c + \cos\alpha\sin b\sin c \\ \cos\alpha &= -\cos\beta\cos\gamma + \cos a\sin\beta\sin\gamma. \end{aligned} \tag{1.13}$$

We will use the last equation later on the text after considering that the angle β is a right angle. That is, the equation becomes

$$\cos\alpha = \cos a\sin\gamma. \tag{1.14}$$

This equation will be fundamental to find the dihedral angle on a polyhedron in Chapter 3.

1.4.1.2 The sine law

Let us review two basic equations of trigonometry, which convert sums of trigonometrical functions on products.

From $\cos(X \pm Y) = \cos X\cos Y \mp \sin X\sin Y$, and by subtracting these two (for the plus "+ " and the minus "-") we find

$$\cos(X+Y) + \cos(X-Y) = 2\cos X\sin Y \tag{1.15}$$

$$\cos(X+Y) - \cos(X-Y) = -2\sin X\sin Y \tag{1.16}$$

Likewise, from $\sin(X \pm Y) = \sin X\cos Y \pm \sin Y\cos X$ adding and subtracting we find

$$\sin(X+Y) + \sin(X-Y) = 2\sin X\cos Y \tag{1.17}$$

$$\sin(X+Y) - \cos(X-Y) = -2\sin Y\cos X \tag{1.18}$$

Let us now write

$$\sin^2\alpha = (1-\cos\alpha)(1+\cos\alpha) = M\,P,$$

where $M = 1-\cos\alpha$ and $P = 1+\cos\alpha$. We simplify each of these expressions using what we now know.

$$
\begin{aligned}
M &= 1 - \cos\alpha \\
&= 1 - \frac{\cos a - \cos b \cos c}{\sin b \sin c} \quad \text{from equation 1.10} \\
&= \frac{\sin b \sin c - \cos a + \cos b \cos c}{\sin b \sin c} \\
&= \frac{\sin b \sin c + \cos b \cos c - \cos a}{\sin b \sin c} \\
&= \frac{\cos(b-c) - \cos a}{\sin b \sin c} \\
&= \frac{-2\sin[(b-c+a)/2]\sin[(b-c-a)/2]}{\sin b \sin c}
\end{aligned}
$$

where in the last equation we used equation 1.16 with $X = (b-c+a)/2$ and $Y = (b-c-a)/2$. Then, by using the semisum $s = (a+b+c)/2$, we can write

$$M = \frac{2\sin(s-c)\sin(s-b)}{\sin b \sin c}.$$

Similarly

$$
\begin{aligned}
P &= 1 + \cos\alpha \\
&= 1 + \frac{\cos a - \cos b \cos c}{\sin b \sin c} \\
&= \frac{\sin b \sin c - \cos b \cos c + \cos a}{\sin b \sin c} \\
&= \frac{\cos(b+c) - \cos a}{\sin b \sin c} \\
&= \frac{-2\sin(\frac{b+c+a}{2})\sin(\frac{b+c-a}{2})}{\sin b \sin c} \\
&= \frac{2\sin(s)\sin(s-a)}{\sin b \sin c}
\end{aligned}
$$

Then

$$\sin^2\alpha = PM = \frac{4\sin(s-a)\sin(s-b)\sin(s-c)\sin s}{\sin^2 b \sin^2 c}.$$

and dividing by $\sin^2 a$, we find

$$\frac{\sin^2\alpha}{\sin^2 a} = \frac{4\sin s\sin(s-a)\sin(s-b)\sin(s-c)}{(\sin a\sin b\sin c)^2}$$

If we call the expression in the right hand side as $f(a,\alpha)$, we observe that $f(a,\alpha) = f(b,\beta) = f(c,\gamma)$, and so

$$\frac{\sin\alpha}{\sin a} = \frac{\sin\beta}{\sin b} = \frac{\sin\gamma}{\sin c} \tag{1.19}$$

Of course we could compute the three quotients individually and then get to the same conclusion.

Note, again, that in the limit where the interior angles a, b, c are small we find the approximation

$$\frac{\sin\alpha}{a} \approx \frac{\sin\beta}{b} \approx \frac{\sin\gamma}{c}$$

which for plane geometry is the exact sine law.

As with the cosine law we could use the duality between the prime and polar triangle to find a dual sine law. With the help of equations 1.8 and 1.19

$$\frac{\sin(\pi-a)}{\sin\pi-\alpha} = \frac{\sin(\pi-b)}{\sin(\pi-\beta)} = \frac{\sin(\pi-c)}{\sin(\pi-\gamma)}.$$

That is

$$\frac{\sin a}{\sin\alpha} = \frac{\sin b}{\sin\beta} = \frac{\sin c}{\sin\gamma}.$$

Clearly we did not gain much, since this equation is the same as equation 1.19 by reversing the order (numerator and denominator) of the fractions.

1.4.2 Half Angle, half side formulas

Theorem 1.4.1 (**Half angle formulas).** *The half angle formulas are provided with the additional notation $2s = (a+b+c)$, and $2S = \alpha+\beta+\gamma$.*

$$\sin\frac{1}{2}\alpha = \left[\frac{\sin(s-b)\sin(s-c)}{\sin b\sin c}\right]^{1/2} \qquad \sin\frac{1}{2}a = \left[\frac{-\cos(S-\alpha)\cos(S)}{\sin B\sin C}\right]^{1/2}$$

$$\cos\frac{1}{2}\alpha = \left[\frac{\sin s\sin(s-a)}{\sin b\sin c}\right]^{1/2} \qquad \cos\frac{1}{2}a = \left[\frac{\cos(S-B)\cos(S-C)}{\sin B\sin C}\right]^{1/2}$$

$$\tan\frac{1}{2}\alpha = \left[\frac{\sin(s-b)\sin(s-c)}{\sin s\sin(s-a)}\right]^{1/2} \qquad \tan\frac{1}{2}a = \left[\frac{-\cos S\cos(S-\alpha)}{\cos(S-B)\cos(S-C}\right]^{1/2}$$

To prove the first formula we start with

$$\cos\alpha = \cos 2(\alpha/2) = \cos^2(\alpha/2) - \sin^2(\alpha/2) = 1 - 2\sin^2(\alpha/2). \qquad (1.20)$$

from which we find the half angle formula for the sine function

$$\sin\frac{1}{2}\alpha = \sqrt{\frac{1-\cos\alpha}{2}}.$$

Now, from the cosine law (equation 1.10) we see that

$$\sin\frac{1}{2}\alpha = \sqrt{\frac{\sin b\sin c - \cos a + \cos b\cos c}{2\sin b\sin c}} = \sqrt{\frac{\cos(b-c) - \cos a}{2\sin b\sin c}}$$

Now we use equation 1.11 to convert the difference of cosines into a product formula. That is, we write

$$\cos(b-c) - \cos a = -2\sin\left(\frac{b-c+a}{2}\right)\sin\left(\frac{b-c-a}{2}\right)$$

with $X = (b-c+a)/2$, and $Y = (b-c-a)/2$. Now from $s = (a+b+c)/2$ we see that $(b-c+a)/2 = s-c$, and $(b-c-a)/2 = b-s$, so we write

$$\cos(b-c) - \cos a = 2\sin(s-c)\sin(s-b).$$

and the result

$$\sin\frac{1}{2}\alpha = \left[\frac{\sin(s-b)\sin(s-c)}{\sin b\sin c}\right]^{1/2} \qquad (1.21)$$

follows.

To prove the second formula (in the first column) we use the first and third equality expressions on equation 1.20 to write

$$\cos\frac{1}{2}\alpha = \sqrt{\frac{1+\cos\alpha}{2}}.$$

We use, again, the cosine law 1.10 to write

$$\cos\frac{1}{2}\alpha = \sqrt{\frac{\sin b\sin c + \cos a - \cos b\cos c}{2\sin b\sin c}} = \sqrt{\frac{\cos a - \cos(b+c)}{2\sin b\sin c}}$$

and using equation 1.16 to convert the difference of cosine into a product of sines we find

$$\cos a - \cos(b+c) = -2\sin\left(\frac{a+b+c}{2}\right)\sin\left(\frac{a-b+c}{2}\right)$$

with $X = (a+b+c)/2$, and $Y = (a-b-c)/2$. Now from $s = (a+b+c)/2$ we see that $(a-b-c)/2 = a-s$, so we write

$$\cos\frac{1}{2}\alpha = \sqrt{\frac{\sin s\sin(s-a)}{\sin b\sin c}}.$$

The formula for $\tan(\alpha/2)$ is directly obtained from the division of the two first formulas in the first column.

To prove the first formula in the second column, we use theorem 1.3.2. That is, we can write the first formula on the first column for the polar triangle as

$$\sin\frac{1}{2}\alpha' = \left[\frac{\sin(s'-b')\sin(s'-c')}{\sin b'\sin c'}\right]^{1/2}$$

where $s' = (a'+b'+c')/2$. In terms of non-prime symbos we have

$$s' = \frac{3\pi - (\alpha+\beta+\gamma)}{2} = \frac{3\pi}{2} - S$$

Then

$$s'-b' = \frac{\pi}{2} - S + \beta \quad , \quad s'-c' = \frac{\pi}{2} - S + \gamma$$

and

$$\sin(s'-b') = \cos(S-\beta) \quad , \quad \sin(s'-c) = \cos(S-\gamma).$$

On the other hand

$$\frac{\alpha'}{2} = \frac{\pi}{2} - \frac{a}{2} \qquad \Longrightarrow \sin\frac{\alpha'}{2} = \cos\frac{1}{2}a.$$

We find

$$\cos\frac{1}{2}a = \left[\frac{\cos(S-\beta)\cos(S-\gamma)}{\sin\beta\sin\gamma}\right]^{1/2}$$

since $\sin(\pi - \beta)\sin(\pi - \gamma) = \sin\beta\sin\gamma$.

Likewise if we start with the second formula in the first column of the equations at the begining of this section and use it for the polar, instead of the prime triangle, we will find the first equation of the second column.

The third equation of the second column is obtained by dividing the first two equations of the second column.

Finally we show the de Delambre or Gauss analogies.

1.4.3 Delambre or Gauss Analogies

Theorem 1.4.2 (Delamber). *The following identities are known as the* **Delambre** *or* **Gauss** *analogies*

$$\frac{\sin\frac{1}{2}(\alpha+\beta)}{\cos\frac{1}{2}\gamma} = \frac{\sin\frac{1}{2}(a-b)}{\cos\frac{1}{2}c} \quad , \quad \frac{\sin\frac{1}{2}(\alpha-\beta)}{\cos\frac{1}{2}\gamma} = \frac{\sin\frac{1}{2}(a-b)}{\cos\frac{1}{2}c}$$

$$\frac{\cos\frac{1}{2}(\alpha+\beta)}{\sin\frac{1}{2}\gamma} = \frac{\cos\frac{1}{2}(a+b)}{\cos\frac{1}{2}c} \quad , \quad \frac{\cos\frac{1}{2}(\alpha-\beta)}{\sin\frac{1}{2}\gamma} = \frac{\sin\frac{1}{2}(a+b)}{\sin\frac{1}{2}c}$$

Proof. We show only the first equation and left the other to the reader, since the methodology is the same: expand the numerator and use the half angle formulas.

Expanding the numerators on the first fraction we find

$$\sin\frac{1}{2}(\alpha+\beta) = \sin\frac{1}{2}\alpha\cos\frac{1}{2}\beta + \cos\frac{1}{2}\alpha\sin\frac{1}{2}\beta$$

From the half angle formulas in Theorem 1.4.1 we see that

$$\begin{aligned}\sin\frac{1}{2}\alpha\cos\frac{1}{2}\beta &= \left[\frac{\sin(s-b)\sin(s-c)\sin s\sin(s-b)}{\sin b\,\sin c\,\sin c\,\sin a}\right]^{1/2} \\ &= \frac{\sin(s-b)}{\sin c}\left[\frac{\sin s\sin(s-c)}{\sin a\sin b}\right]^{1/2}.\end{aligned}$$

Likewise

$$\begin{aligned}\sin\frac{1}{2}\beta\cos\frac{1}{2}\alpha &= \left[\frac{\sin(s-c)\sin(s-a)\sin s\sin(s-a)}{\sin c\,\sin a\,\sin b\,\sin c}\right]^{1/2} \\ &= \frac{\sin(s-a)}{\sin c}\left[\frac{\sin s\sin(s-c)}{\sin a\sin b}\right]^{1/2}.\end{aligned}$$

Then

$$\begin{aligned}\sin\frac{1}{2}(\alpha+\beta) &= \frac{1}{\sin c}\left[\frac{\sin s\sin(s-c)}{\sin a\sin b}\right]^{1/2}[\sin(s-b)+\sin(s-a)]\\ &= 2\frac{1}{\sin c}\left[\frac{\sin s\sin(s-c)}{\sin a\sin b}\right]^{1/2}\cos\frac{1}{2}(a-b)\sin\frac{1}{2}c\end{aligned}$$

where we use the sum into product equation 1.17. Now, from the half-angle formula for $\cos\alpha/2$, we observe that the expression in square brackets is $\cos\gamma/2$. So we write

$$\sin\frac{1}{2}(\alpha+\beta) = \frac{2}{\sin c}\cos\frac{\gamma}{2}\cos\frac{1}{2}(a-b)\sin\frac{1}{2}c$$

Then

$$\begin{aligned}\frac{\sin\frac{1}{2}(\alpha+\beta)}{\cos\frac{1}{2}\gamma} &= \frac{2}{\sin c}\cos\frac{1}{2}(a-b)\sin\frac{1}{2}c\\ &= \frac{\cos\frac{1}{2}(a-b)}{\sin c/(2\sin(1/2)c)}.\end{aligned}$$

and from

$$\sin c = \sin 2(c/2) = 2\sin(c/2)\cos(c/2),$$

we see that

$$\frac{\sin\frac{1}{2}(\alpha+\beta)}{\cos\frac{\gamma}{2}} = \frac{\cos\frac{1}{2}(a-b)}{\cos\frac{c}{2}}.$$

The other three identities could be derived following similar steps.

□

The Wikipedia [10] website on spherical trigonometry shows some additional important identities which we will not consider on this document.

[10] https://en.wikipedia.org/wiki/Spherical_trigonometry

1.5 Euclid

We make a small comparison of spherical geometic with plane (Euclid) geometry. Euclidean geometry is based on the following 5 postulates:

1. A straight segment can be drawn joining two points.

2. Any striaght segment can be extended indefinitely in a straight line.

3. Given any straight segment, a circle can be drawn having the segment as a radius and one endpoint as its center.

4. All right angles are congruent.

5. Through any give point can be drawn exactly one striaghtline parallel to a given line.

The fifth postulate was not posted by Euclid in this way, but it is equivalent to the original Euclid postulate. This fifth posulate is known as the Playfair's axioma [11].

The spherical geometry is an example of a non-Euclidean geometry.

A brief history in the discovery of the non-Euclidean geometry is shown in the Wikepedia link above. We highlight a few points.

- This is attributed mainly to the Russian mathematician Nikolai Lobachevsky and the Hungarian mathematician János Bolyai. Gauss told Bolyai's father that he "had developed such a geometry several years before". Here is the Wikepedia statement about this: " *In the letter to Wolfgang (Farkas) Bolyai of March 6, 1832 Gauss claims to have worked on the problem for thirty or thirty-five years (Faber 1983, pg. 162). In his 1824 letter to Taurinus (Faber 1983, pg. 158) he claimed that he had been working on the problem for over 30 years and provided enough detail to show that he actually had worked out the details. According to Faber (1983, pg. 156) it wasn't until around 1813 that Gauss had come to accept the existence of a new geometry.* "

- There are three geometries based on three basic surfaces. These are: the plane (plane geometry), the ellipsoid (elliptical geometry, and as a

[11] https://en.wikipedia.org/wiki/Playfair's_axiom

particular case the spherical geometry which is the object of this document), and hyperbolic geometry. Hyperbolic geometry was actually the initial work developed by Lovachevsky and Bolyai independently.

- - Bernhard Riemann made a generalization of non-Euclidean geometry for any surface. Riemannian Geometry [12] deals with surfaces and their curvatures as well as measures of distances, angles, and areas on them.
  - This is the geometry used by Einstein in its Theory of Relativity, [13].
  - Riemannian geometry provides an infinite collection of non-Euclidean geometries.
  - Riemannian geometry is a branch of Differential Geometry [14], which is the mother of all geometries.

Let is review the 5 postulates of Euclid now in the context of spherical geometry.

1. In a plane, the joining of two points produces only one segment. In the sphere, we can joint to points along two different paths. This produces the minor and major segments. The case is even worse if the two points of reference are antipodes. In this case there is an infinite number of segments connecting the two points.

2. A segment can be extended infinitely by going around the unit sphere as many times as we want. However up to 2π, its length is finite.

3. We can not create circles in the sphere with arbitrary radius.

4. All right triangles are "equal". They are really congruent. This means they all have the same measure, which is 90°.

5. There are no parallel lines in the circle. Different lines intersect in at least two (antipode) points. It is interesting that while in plane geometry there is exactly one parallel line through an exterior point of a given line, in hyperbolic geometry there are an infinite number

[12]https://en.wikipedia.org/wiki/Riemannian_geometry
[13]https://en.wikipedia.org/wiki/Theory_of_relativity
[14]https://en.wikipedia.org/wiki/Differential_geometry

of lines parallel to the given line through the exterior point, while in elliptic (and so in spherical) geometry there are no parallel lines.

Mathematicians in the nineteenth century suspected that Euclid's parallel postulate was equivalent to the fact that the sum of the angles in a triangle is 180°. According to Liviu I. Nicolaescu [15] Gauss got interest on this matter and was shocked to find out that the sum of angles of a triangle fromed by three far away stars was less than 180°. Today we know that in hyperbolic geometry that is the case, while in spherical geometry the sum of the angles is larger than 180°. In plane (Euclidean) geometry the sum is exactly 180° and this is easy to prove with Euclid's fith postulate.

We now show in Appendix A that in plane geometry the shortest path between two points is along a straight line that joints the points (see example A.3.1). Likewise in spherical geometry the shortest path between two points is along the minor segment that joints the point. We show that next.

1.5.1 Geodesics of a Sphere

We offer a few explanations that help us understand in which context and why the shortest distance between two points along a spherical surface is given by the distance along a great circle.

Of course, if the points are antipodes, there are an infinite number of paths, all with the same distance (π) between them. To gain some understanding about why the shortest distance should be along great circles think about this: We know that the length of the segment is the same no matter how we rotate the sphere. Imagine that we are at the center of the sphere looking above at the segment. We could think about the first point of the segment A as located in the north pole, and the second point B toward the east. The segment can only wiggle in three dimensions, but since it is attached to the sphere, we have only two dimensions of free motion. As we scan the segment from north to east (from A to B). The two free dimensions are "west-to-east" and "north-to-south". If the segment AB is not along a great circle, there is a projection of it into the north-to-south direction, this projection is a wiggle in the path that only will make it longer. That is, if the trajectory is in a great circle, the path looks like a straight line. There is no projection in the north-to-south directions and there is no shortest path than this.

[15] http://www3.nd.edu/ ndreu/nicolaescuREU.pdf

Another argument to understand why the shortest path between two points in the sphere is along a great circle, is based on curvature.

Given two points, think of all the segments that joint the two points. The straightest one should be the shortest. By "straightest" I mean the less curved. Curvature add length, that should be clear.

Now, take any plane intersecting the two points. Why plane? because it has the least curvature of all surfaces (0 curvature).

Then if the plane is very shallow, the curvature will be larger. Hence the plane that, after intersecting both points, has the least curvature needs to be a plane that goes through the center of the sphere. This is, the arc of least curvature on the sphere. Then the shortest path is along a great circle. Figure 1.10 shows a collection of possible intersections of a surface with planes which share the same segment AB. We rotated the curves so that they all are in the same vertical plane, but each curve comes from a tilted plane. The shortest curve is that corresponding to the smallest curvature which is the intersection of the plane going through the center of the sphere. That is, Assume that the points A and B are in the sides of the sphere just in front of your eyes. The center of the sphere is the green point CC. This center is in the plane through the points A, B, and C which has a great circle when intersecting with the sphere and it is the plane you are looking at (plane having the "flat" screen of your computer) As the planes with the line AB on them move away from the center (up in the figure) the radius of curvature decreases (from r which is the radius of the sphere down to $r0 = AB/2$ which is the radius of the intersection of a horizontal plane having the points A and B.). See that the planes are being tilted because they are pinned to the line AB, so they are rotating down beyond AB, up toward the reader. The color of the dot (center) matches the color of the curve. When the plane, which started vertical rotated around the axis AB until it becomes horizontal, the intersection is a the yellow dashed circle (only half shown here) in the figure. That is, the largest it can get, being flat (no wiggling. Of course you can wind all you want and that is where there is no upper limit, but the longest of the shortest is half circle with radius $r_0 = AB/2$, that is $r_0\pi/2$).

It should be clear from this figure that the green path is the shortest (after the black which is a sphere of infinite radius and has the center down at infinity).

Still, we write next a formal proof of this fact.

A unit sphere can be parametrized in terms of its azimuthal angle ϕ with $0 \leq \phi < 2\pi$, and polar angle θ, with $0 \leq \theta \leq \pi$, with $\theta = 0$ at the north pole.

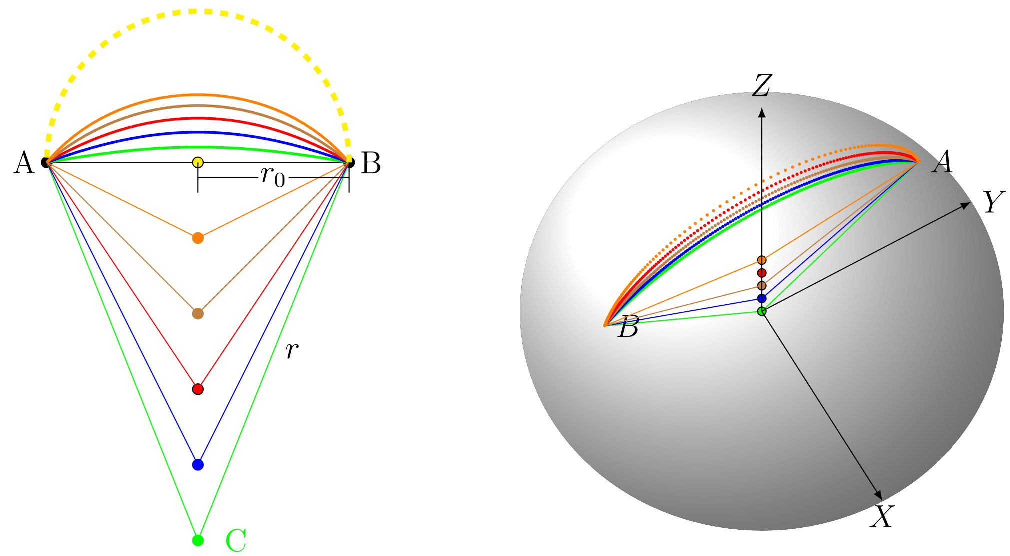


Figure 1.10: The green arc is standing along a great circle with a center at the center of the sphere labeled with the green C character. The other colors are different circles created by planes that intersect the sphere away from the center. The furthest from the center the plane is, the smaller the radius but the larger the curvature. So the segments are larger. The shortest path in a plane is the segment AB. The shortest in the sphere is the green segment. In the right figure, we used 200 points per curve, and the point density becomes greater for curves with center above the $z = 0$ plane indicating that those paths are longer.

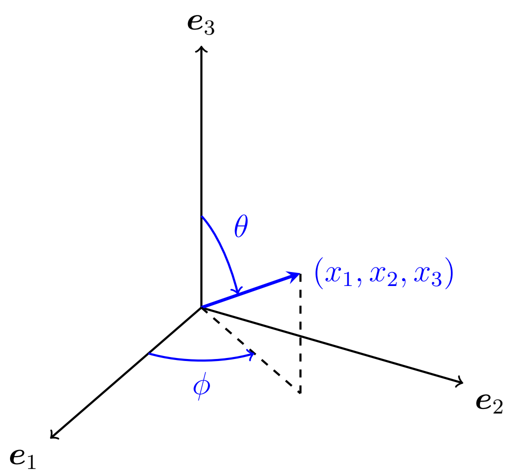
Figure 1.11 illustrates the symbols used here.
with

$$\begin{aligned} x_1 &= r \sin\theta \cos\phi \\ x_2 &= r \sin\theta \sin\phi \\ x_3 &= r \cos\theta \end{aligned} \tag{1.22}$$

We find the metric tensor. We have

$$\boldsymbol{r}_\theta = \begin{pmatrix} \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial \theta} \\ \frac{\partial x_3}{\partial \theta} \end{pmatrix} = \begin{pmatrix} r\cos\theta\cos\phi \\ r\cos\theta\sin\phi \\ -r\sin\theta \end{pmatrix} \quad , \quad \boldsymbol{r}_\phi = \begin{pmatrix} \frac{\partial x_1}{\partial \phi} \\ \frac{\partial x_2}{\partial \phi} \\ \frac{\partial x_3}{\partial \phi} \end{pmatrix} = \begin{pmatrix} -r\sin\theta\sin\phi \\ r\sin\theta\cos\phi \\ 0 \end{pmatrix} .$$

Figure 1.11: Polar–Spherical Coordinates in 3D



Then

$$\begin{aligned} E = g_{11} &= \boldsymbol{r}_\theta \cdot \boldsymbol{r}_\theta = r^2 \\ F = g_{12} &= g_{21} = \boldsymbol{r}_\theta \cdot \boldsymbol{r}_\phi = 0 \\ G = g_{22} &= \boldsymbol{r}_\phi \cdot \boldsymbol{r}_\phi = r^2 \sin^2\theta. \end{aligned} \tag{1.23}$$

and then the stretch element is given by

$$ds = \sqrt{r^2(d\theta)^2 + r^2\sin^2\theta(d\phi)^2}. \tag{1.24}$$

While by using the metric tensor we can derive the element of distance, we could also, from the Figure 1.12 use the diagonal of the spherical cube from (r, θ, ϕ), to $(r + dr, \theta + d\theta, \phi + d\phi)$ for this purpose. The differential element along the r direction is dr, the differential element along the ϕ direction is $r\,\sin\theta d\phi$, and the differential element along the θ direction is $rd\theta$. Then by the distance formula:

$$ds = \sqrt{(dr)^2 + r^2\sin^2\theta(d\phi)^2 + r^2(d\theta)^2}.$$

Now, since on the sphere $dr = 0$, this formula simplifies to

$$ds = \sqrt{r^2\sin^2\theta(d\phi)^2 + r^2(d\theta)^2}.$$

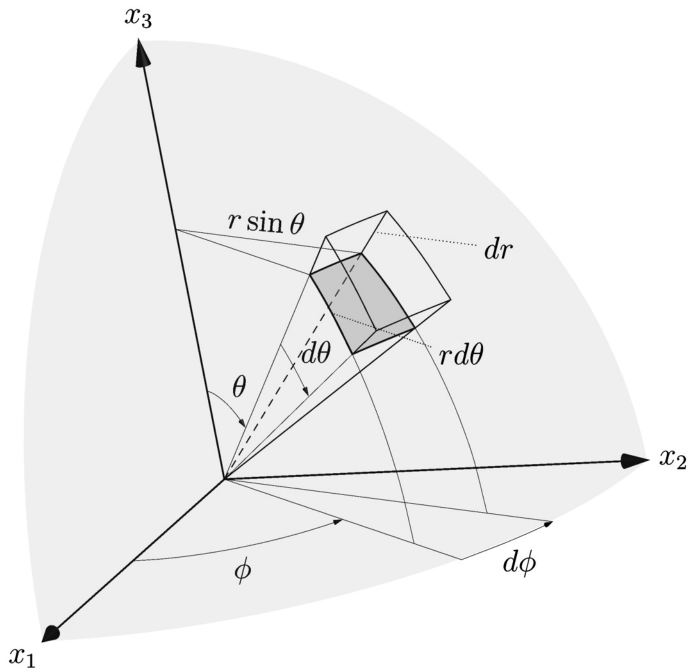


Figure 1.12: Sketch of spherical coordinates

which is the same equation 1.24 above.

We know that r is constant along the sphere can can be factor out of the integral. Then we want to consider a simpler function

$$ds = \sqrt{\sin^2\theta(d\phi)^2 + (d\theta)^2}.$$

whatever the distance we find, it should be scaled by r if we are not sitting in a unit sphere.

Now, we know that we are searching for a curve and a curve is parametrized with a single parameter t. To avoid introducing new parameters we can use either θ or ϕ as the parameter. We will see that if we choose the parameter $t = \phi$, then we could use the Beltrami identity 1.26. We note that, since the function inside the radical is a function of θ, but is not a function of ϕ, we want to use ϕ as the independent variable. Then we can write the integral as

$$S = \int_a^b \sqrt{\sin^2\theta + (\theta')^2} d\theta \tag{1.25}$$

with $\theta' = d\theta/d\phi$.

We see that the reduction of these equations for case that the operator L does not depend on the independent variable ϕ, then , from equation A.7

$$L - \theta' \frac{\partial L}{\partial \theta'} = \text{constant} \equiv C. \tag{1.26}$$

where $L = \sqrt{\sin^2\theta + (\theta')^2}$. Then find

$$\begin{aligned} L - \theta' \frac{\partial L}{\partial \theta'} &= \sqrt{\sin^2\theta + (\theta')^2} - \theta' \frac{\theta'}{\sqrt{\sin^2\theta + (\theta')^2}} \\ &= \frac{\sin^2\theta}{\sqrt{\sin^2\theta + (\theta')^2}} \\ &= C \end{aligned}$$

From here

$$\frac{\sin^4\theta}{C^2} = \sin^2\theta + (\theta')^2$$

$$\theta' = \pm\sqrt{\frac{\sin^4\theta}{C^2} - \sin^2\theta}.$$

To integrate this equation we write

$$\frac{d\theta}{d\phi} = \pm\sqrt{\frac{\sin^4\theta}{C^2} - \sin^2\theta} = \pm\sqrt{\frac{\sin^4\theta - C^2\sin^2\theta}{C^2}} = \frac{\sin\theta\sqrt{\sin^2\theta - C^2}}{C},$$

where we removed the $\pm$ in the last equation since it can be absorbed by C. Now, making a separation of variables so that all θ variables are in on side and ϕ in the other. That is,

$$d\phi = \frac{C\, d\theta}{\sin\theta\sqrt{\sin^2\theta - C^2}}.$$

and

$$\phi = \int_{\theta_a}^{\theta_b} \frac{C\,d\theta}{\sin\theta\sqrt{\sin^2\theta - C^2}} d\theta.$$

While we can find this integral using tables or computer software, as a calculus exercise, we will find it step-by-step.

Let us think of this as an indefinite integral. We find

$$\begin{aligned}
\phi &= \int \frac{C d\theta}{\sin\theta\sqrt{\sin^2\theta - C^2}} \\
&= \int \frac{C\csc^2\theta d\theta}{\csc^2\theta\sin\theta\sqrt{\sin^2\theta - C^2}} \\
&= \int \frac{C\csc^2\theta d\theta}{\sqrt{1 - C^2\csc^2\theta}} \\
&= \int \frac{C\csc^2\theta d\theta}{\sqrt{1 - C^2(1+\cot^2\theta)}}
\end{aligned}$$

Let us now use the following change of variables: $z = C\cot\theta$, so $dz = -C\csc^2\theta d\theta$, and

$$\begin{aligned}
\phi &= \int \frac{-dz}{\sqrt{1 - C^2 - z^2}} \\
&= -\int \frac{dz}{\sqrt{a^2 - z^2}} \quad \text{with} \quad a^2 = 1 - C^2
\end{aligned}$$

Now call $z = aw$, so $dz = adw$, $a > 0$ and

$$\begin{aligned}
\phi &= -\int \frac{adw}{\sqrt{a^2 - a^2w^2}} \\
&= -\int \frac{dw}{\sqrt{1 - w^2}} \qquad (1.27)
\end{aligned}$$

Call $w = \sin\alpha$, then $1 - w^2 = \cos^2\alpha$, $dw = \cos\alpha d\alpha$ $-\pi/2 \le \alpha \le \pi/2$, and so, since in the domain of α, $\sec\alpha \ge 0$,

$$\phi = -\int d\alpha = \alpha + C_2. \qquad (1.28)$$

and

We now do backward substitution.

$$\begin{aligned}
\phi &= -\alpha + C_2 \\
\phi &= -\sin^{-1} w + C_2 \\
\phi &= -\sin^{-1}(z/a) + C_2 \\
\phi &= -\sin^{-1}\left(\frac{C\cot\theta}{\sqrt{1-C^2}}\right) + C_2
\end{aligned}$$

We can take the sine function on both sides and find that

$$\sin(C_2 - \phi) = \frac{C\cot\theta}{\sqrt{1-C^2}}.$$

which can also be written as

$$\cot\theta = k\sin(\phi - C_2) \quad , \quad k = -\frac{\sqrt{1-C^2}}{C}.$$

We now multiply by $r\sin\theta$ to obtain

$$r\cos\theta = (k\sin\phi\cos C_2)r\sin\theta - (k\sin C_2\cos\phi)r\sin\theta.$$

We now use equations 1.22 to write this in terms of x_i coordinates and find, after calling $m = k\cos C_2$, and $n = k\sin C_2$:

$$mx_2 - nx_1 - x_3 = 0$$

This is the equation of a plane going through the origin perpendicular to the vector $(m, -n, -1)$. The values of m, and n are found from k and C_2, and these from the boundary conditions. That is by having the initial and final points fixed. So we know $\phi(a)$ and $\phi(b)$ with a, b the initial angle θ values in integral 1.25. Since, in addition, $x_1^2 + x_2^2 + x_3^2 = r^2$ then we find that the solution for the path lies in the intersection of a plane through the origin and the sphere. That is a great circle. Hence the great circles are the geodesics of the sphere.

As an illustration we can use Figure 1.1. The tilted circle is the intersection of a plane with normal direction P through the center O, and the sphere. That is the great circle having the blue arc AC. We can say that A and C are the initial and the final points. See however that since our differential equation is of second order, we have a second solution and this is the path through the back which is the whole circumpherence having the arc AC minus the arc AC or, in other words, the complement of AC with respect to the circumpherence. So, the stationary solution does not always provides the shortest path.

Chapter 2

Polygons

2.1 Basics

In the previous chapter we introduced the basic definitions of point, line, segment, angle, and triangle. The triangle is the simplest of the polygons. We want to extend this to new objects such as quadrilaterals, pentagons, hexagons, and general n-polygons.

Note that since objects on the sphere come in pairs (by using the antipodes) we will always refer to segment as the minor segment unless it is explicit that we are refering to the major segment. Recall also that all segmens are taken along geodesic lines (great circles).

Definition 8 (Polygon). *Given a set of points $S = \{A_i : i = 1, \cdots, n\}$. The set of segments that join the points in the given order is known as a polygon. We want to enforce the location of all points of a polygon along the same plane and in such a way that no two edges cross inside the polygon.*

Figure 2.1 shows a pentagon formed by 5 vertices, as well as a triangle. We will study angles and areas in polygons starting with the triangle and going up to higher number of sides. In plane geometry we know that the sum of the angles in a triangle is 180°, and this is shown using the 5th postulate of a parallel line through a side of a triangle. In spherical geometry we will see that, for a triangle, the sum of the interior angles is larger than 180°. Actually, if two of the sides run through lines of constant latitude and intersect a line of constant altitude the three angles are right and they add to 270°. In plane geometry the angles in a triangle are smaller than 180°. The

Figure 2.1: A pentagon (left) and triangle (right) on a sphere. The spherical triangle surrounds a plane triangle showing that the spherical angles are larger than the triangular angles and so the sum of its angles is expected to be larger than 180°.

extreme case is degenerate corresponding to a segment of length $c = a + b$, with a, b the lengths of the other two segments. All segments are co-linear. In the case of spherical geometry an angle can approach 540° (3π in radians) as much as possible having as a limit (degenerated triangle) a great circle. The sum of angles in spherical geometry is between 180° and 540°. The proof of this is just after equation 2.6.

2.2 The lune

In the case of the sphere there are polygons of two sides, which do not exist in plane geometry. These polygons are called **lunes** and its vertices are two antipodes. We discussed lunes in Chapter 1, section 1.3.2 and extend that discussion here. A lune is limited by the intersection of two great circles. There are multiple names of lunes in the literature. For example **biangles , digons, and bigons.** Figure 2.2 shows a lune , and a lune together with its antipode (spherical reflected) antipode.

Let us now find the area of a lune. It seems obvious that this area is directly proportional to the diheral angle under it. That is, the area seems to be $A = 2\theta r^2$, where θ is the dihedral angle corresponding to the lune. If $\theta = 2\pi$ we get the area of the sphere which is $4\pi r^2$. A way to prove this is that by checking that two lunes with the same angle (due to invariance of rotation) are congruent. They should have the same area. Since the areas are additive (for adjacent or not overlapping surfaces) we have

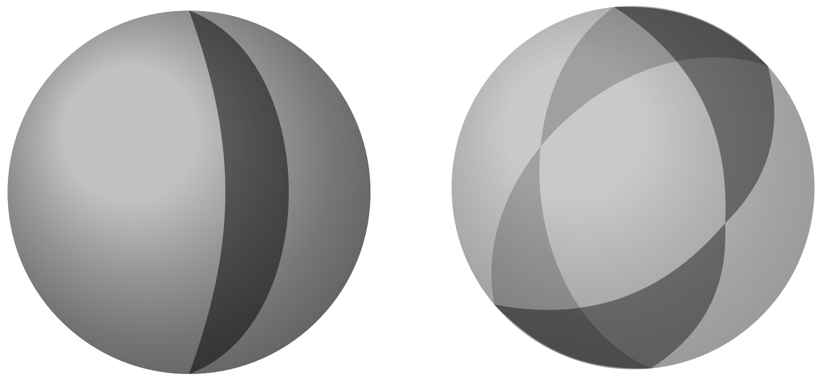


Figure 2.2: On the left frame we see a lune; on the right frame we see a lune and its antipode lune.

that $A(l_1 \cup l_2) = A(l_1) + A(l_2)$, as long as the overlapping between the two lunes l_1 and l_2 is of measure zero. In general for any $n \in \mathbb{N}$ we have that $A(\cup_{i=1}^n l_i) = \sum_{i=1}^n A(l_i)$ for a finite set of lunes l_i, $i = 1, \cdots n$. Now, for the moment think of an angle $\theta = 2\pi/n$, then $n\theta = 2\pi$, and so:

$$A(\cup_{i=1}^n l(\theta_i)) = nA(l_\theta) = A(l_{2\pi}) = 4\pi r^2,$$

where we accept that the whole sphere has area $4\pi r^2$, and $\cup_{i=1}^n l(\theta_i)$ is the union of n congruent lunes that make up a sphere. We remove the index i since congruent surfaces have equal areas.

From here

$$A(l_\theta) = \frac{4\pi r^2}{n} = 2\theta\, r^2. \tag{2.1}$$

This is good for θ of the form $2\pi/n$. Since l is additive then this should be good for any angle of the form $2\pi p/q$, with p and q in $\mathbb{Z}$ with $q \neq 0$. If $2\pi/\theta$ is irrational, then this irrational is a limit of a sequence of rational numbers, all of them providing the identical equation $A(l_\theta) = 2\theta r^2$, so in the limit this equation is valid as well, since the length operator is continuous.

Appendix B discusses some facts about functions which are either homogeneous or additive and their implication on their linear representation.

2.3 Triangles revisited

We had already defined a spherical triangle (6) and provided a few properties in section 3.2.6. Here we study the triangles from a different point of view, by considering the sum of its internal and external angles and the relation of that sum with the area. First we want to choose a convention for the size of a triangle side. We assume on these notes that the size of a triangle is smaller than half of a great circle. The case of exactly half of great circle an uninteresting and an extreme where the sphere is split in two equal parts. The case of a side longer than half of a circle creates non-convex lunes which we want to exclude.

2.3.1 Area of the triangle

We use equation 2.1 to find the area of an spherical triangle in terms of its angles.

Figure 2.3 shows a triangle (red) its anti-pode triangle (blue) and the lunes generated by those triangles.

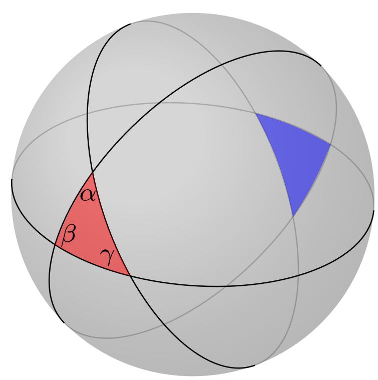


Figure 2.3: Triangles and their asscociated lunes.

We first observe that the two triangles are congruent, since the angles, by definition, are the dihedral angles defined by the planes of each side of the lune and the lunes are shared by the triangles. That each, to each vertex of the triangle there are two lunes sharing the same dihedral angle. Furthermore, since the side lengths correspond to the dihedral angles, then the sides have the same length and the triangles are congruent. If we add the area of the six lunes associated with the two triangles we see that all the sphere has been covered, but the triangles were counted each three times. That is, adding all the lunes we have

$$4\pi + 4A(\triangle) = 2L(\alpha) + 2L(\beta) + 2L(\gamma) = 4(\alpha + \beta + \gamma),$$

and from here

$$A(\triangle) = \alpha + \beta + \gamma - \pi \tag{2.2}$$

The value $E = A(\triangle)$ is known as **triangular excess** , or **spherical excess** . If, as in Euclidean geometry, we define the **exterior angle** as the suplement of the interior angle, that is, $\alpha_E = \pi - \alpha_I$, where α_I is the interior angle and α_E is the exterior angle we can rewrite the excess formula as

$$A(\triangle) = 2\pi - (\alpha_E + \beta_E + \gamma_E),$$

which shows that the excess as the difference between a whole angle 2π in a circle and the sum of the exterior angles. In plane geometry this is 0.

At this point we are curious about this formula as the radius of the sphere increases without limit. It seems to indicate that in Euclidean geometry (where the radius of the sphere is infinite) all triangles have zero area. We could say that relative to such a big sphere that is true, but if we look a bit more carefully we are ignoring here an r^2 factor. Recall that our spheres are unit spheres. Still if the area of a spherical the triangle in a sphere of radius r is given by $A(\triangle) = r^2 [\alpha + \beta + \gamma - \pi.]$, then in the limit as $r \to \infty$, the difference between $\alpha + \beta + \gamma$ and π goes to zero but not as fast as r^2 goes to infinity, right? Objects become amplified in area by r^2 and length by r so it makes sense that as $r \to \infty$ we should have infinite areas, but is there a way to pull these objects back?

To answer the question of how to obtain the plain geometry area of a triangle from spherical geometry we need to prove :

Theorem 2.3.1 (**Cagnoli's Theorem**). *Let $E = A(\triangle)$ be the triangular excess. Then*

$$\sin\frac{E}{2} = \frac{\sqrt{\sin s \sin(s-a)\sin(s-b)\sin(s-c)}}{2\cos(a/2)\cos(b/2)\cos(c/2)}$$

where s is the semi-sum $s = (a + b + c)/2$.

Proof. Recall that a, b, c are the arc lengths of the sides of the triangle corresponding to the central (dihedral) angles on each vertex of the triangle with angles α, β, and γ respectively.

From equation 2.2 we see that

$$\begin{aligned}\sin\frac{E}{2} = \sin\frac{\alpha+\beta+\gamma-\pi}{2} &= \sin\left[\frac{\alpha+\beta}{2} - \frac{\pi-\gamma}{2}\right] \\ &= \sin\frac{\alpha+\beta}{2}\sin\frac{\gamma}{2} - \cos\frac{\alpha+\beta}{2}\cos\frac{\gamma}{2}\end{aligned}$$

We now use Delambre's theorem 1.4.2. That is, for the first term

$$\sin\frac{\alpha+\beta}{2}\sin\frac{\gamma}{2} = \frac{\cos\frac{1}{2}(a-b)\cos\frac{\gamma}{2}\sin\frac{\gamma}{2}}{\cos\frac{c}{2}}.$$

Similarly

$$\cos\frac{\alpha+\beta}{2}\cos\frac{\gamma}{2} = \frac{\cos\frac{1}{2}(a+b)\cos\frac{\gamma}{2}\sin\frac{\gamma}{2}}{\cos\frac{c}{2}}.$$

Then the subtraction yields

$$\begin{aligned}\sin\frac{\alpha+\beta}{2}\sin\frac{\gamma}{2} - \cos\frac{\alpha+\beta}{2}\cos\frac{\gamma}{2} &= \frac{\cos\frac{\gamma}{2}\sin\frac{\gamma}{2}}{\cos\frac{c}{2}}\left[\cos\frac{1}{2}(a-b) - \cos\frac{1}{2}(a+b)\right] \\ &= 2\frac{\cos\frac{\gamma}{2}\sin\frac{\gamma}{2}}{\cos\frac{c}{2}}\sin\frac{a}{2}\sin\frac{b}{2}\end{aligned}$$

We now use the half angle formulas in Theorem 1.4.1. Then

$$\begin{aligned}\sin\frac{E}{2} &= \frac{2}{\cos\frac{c}{2}}\left[\frac{\sin s\sin(s-c)}{\sin a\sin b}\right]^{1/2}\left[\frac{\sin(s-a)\sin(s-b)}{\sin a\sin b}\right]^{1/2}\sin\frac{a}{2}\sin\frac{b}{2} \\ &= \frac{2\sin\frac{a}{2}\sin\frac{b}{2}\sqrt{\sin s\sin(s-a)\sin(s-b)\sin(s-c)}}{\cos\frac{c}{2}\sin a\sin b} \\ &= \frac{2\cancel{\sin\frac{a}{2}}\cancel{\sin\frac{b}{2}}\sqrt{\sin s\sin(s-a)\sin(s-b)\sin(s-c)}}{4\cancel{\sin\frac{a}{2}}\cos\frac{a}{2}\cancel{\sin\frac{b}{2}}\cos\frac{b}{2}\cos\frac{c}{2}} \\ &= \frac{\sqrt{\sin s\sin(s-a)\sin(s-b)\sin(s-c)}}{2\cos\frac{a}{2}\cos\frac{b}{2}\cos\frac{c}{2}}.\end{aligned}$$

which is Cagnoli's equation. □

We now are ready to see how the area in Euclidean geometry is an asymptotic limit (as the radius of the sphere goes to infinity) of the spherical geometry area shown above.

We write the trigonometrical functions as Taylor series:

$$\begin{aligned}\sin x &= x - \frac{x^3}{3} + \text{H.O.T.}\\ \cos x &= 1 - \frac{x^2}{2} + \text{H.O.T.}\end{aligned}$$

When $r \to \infty$ we get an asymptotic solution by retaining only the leading order terms here. That is:

$$\lim_{r\to\infty} \sin\frac{E}{2} = \lim_{r\to\infty}\frac{E}{2}. \tag{2.3}$$

[1]. Similarly for the right hand side term, as $r \to \infty$ we find

$$\lim_{r\to\infty}\frac{\sqrt{\sin s \sin(s-a)\sin(s-b)\sin(s-c)}}{2\cos\frac{a}{2}\cos\frac{b}{2}\cos\frac{c}{2}} = \lim_{r\to\infty}\frac{\sqrt{s(s-a)(s-b)(s-c)}}{2} \tag{2.4}$$

Please observe that for very large r the interior angles a, b, and c become really small and so their semi-perimeter s. That is, if $r = 1$, then a, b, and c are simultaneously lengths of the triangle sides (arc segments) and central angles in the sphere. If we want to increase $r \gg 1$ then, in order to preserve the size of the segments, we need to shrink the angles by a factor of $1/r$, so that the central angles a, b, and c shrink to zero, but the length of the segments a,b, and c remain constant. Hence there is a duality on the meaning of the symbols a, b, and c. As arguments of the sine and cosine functions, they are angles in radians, but as lengths they are the fixed lengths for $r = 1$ and they should be preserved as the sphere explodes.

From the previous two equations we find that in the limit as $r \to \infty$:

$$E = \sqrt{s(s-a)(s-b)(s-c)}. \tag{2.5}$$

[1]Note that the radius r is implicit on these equations, in addition when $r \to \infty$, $E \to 0$, since the excess will be nothing once you get from a sphere to a plane.

which is Heron's formula [2] For the area of a triangle.

Equation 2.2 has a double fold advantage. We can find the area knowing the angles, but we can also find the sum of the angles knowing the area. Since the area is never zero, we see that in a spherical triangle

$$\alpha + \beta + \gamma > \pi. \tag{2.6}$$

In addition, since the angles on a triangle are each smaller than π we have that $\alpha + \beta + \gamma < 3\pi$. As the angles α, β, and γ approach to π the triangle becomes a great circle. This shows one of the important differences between spherical and plane geometry. It can be shown that Euclid fifth postulate is equivalent to the fact that the sum of the interior angles of a triangle is π. Here we see how equation 2.6 breaks with Euclidean geometry. Equation 2.2 is attributed to Thomas Harriot [3] in 1603.

Another important concept is **spherical defect** . This is the difference betwen 2π and the sum of the triangle's arc lengths. We write $D = 2\pi - (a + b + c)$. We could show that as the triangle increases the spherical defect D decreases. On the other end, as the triangle shirnks the excess increase with the upper limit of 2π.

2.4 Area of polygons with more than 3 sides

As it is done in the case of Euclidean geometry, the area of a polygon can be found by dividing the polygon in triangles and adding up the areas of each triangle. Figure 2.4 illustrates the technique with a pentagon.

- The triangle $\triangle ABC$ has an area of $\alpha_3 + \beta + \gamma_1 - \pi$
- The triangle $\triangle ACD$ has an area of $\alpha_2 + \gamma_2 + \delta_1 - \pi$
- The triangle $\triangle ADE$ has an area of $\alpha_1 + \delta_2 + \epsilon - \pi$

Adding the three areas of the three triangles we get (reordering)

$$\alpha_1 + \alpha_2 + \alpha_3 + \beta + \gamma_1 + \gamma_2 + \delta_1 + \delta_2 + \epsilon - 3\pi.$$

[2] https://en.wikipedia.org/wiki/Heron's_formula
[3] http://elib.mi.sanu.ac.rs/files/journals/zr/14/n014p055.pdf

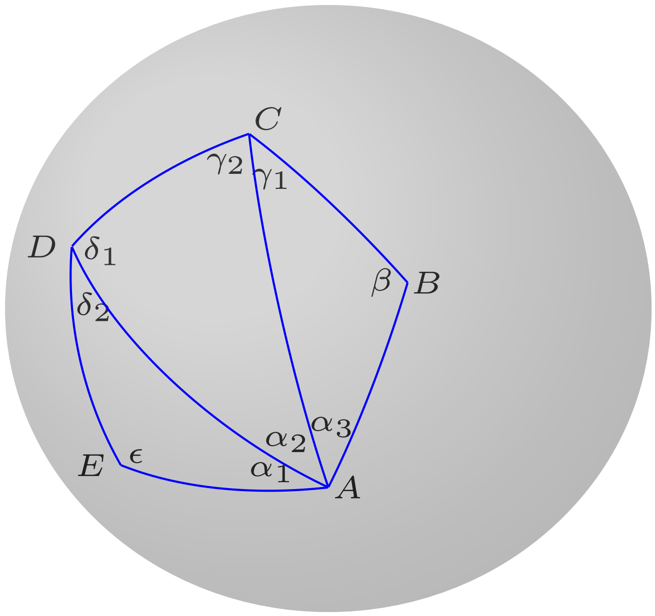


Figure 2.4: A pentagon and how the area is computed.

Or rewriting this in terms of one angle per vertex we have

$$\angle A + \angle B + \angle C + \angle D + \angle E - 3\pi.$$

Note that it does not matter if the pentagon is regular or not (we assume it is convex). In general we find that the area of a polygon of n sides is given by

$$A_p = \sum_{i=1}^{n} \alpha_i - (n-2)\pi, \tag{2.7}$$

where α_i is the measure of the angle at vertex i, $i = 1, \cdots, n$. The area of the polygon in equation 2.7 is also known as the excess of the angles of the polygon. Note that on a the plane this amount is zero since the total internal angle of a polygon is given by $(n-2)\pi$.

In terms of external angles, $\alpha_i = \pi - \beta_i$, where β_i is the exterior angle corresponding to α_i, and so

$$A_p = 2\pi - \sum_{i=1}^{n} \beta_i \tag{2.8}$$

which shows a unified formula for excess angle from any polygon on the sphere with respect to its external angles β_i, $i = 1, \cdots, n$.

The excess equation plays an important role in the proof of Euler's theorem 3.1.1 for polygons.

Chapter 3

Polyhedra

3.1 Definition and Fundamental Properties

Definition 9 (**Polyhedron**). *A polyhedron is a solid in three dimensions formed by flat polygonal faces which intersect at stright edges and sharp vertices. A convex polyhedron is, by definition, a polyhedron such that for any two points inside the polyhedron, the segment that joins the points is totally inside the polyhedron.*

An important theorem which relates the number of faces, edges, and vertices in a polyhedron is due to Euler.

Theorem 3.1.1 (**Euler's Theorem**). *Let P be a convex polyhedron with vertices V, edges E, and faces F. Then*

$$V - E + F = 2. \tag{3.1}$$

Proof. We use equation 2.8. Take all the faces of the convex polyhedron and find an sphere around them with the center of the sphere in the center of the polyhedron. We can project the faces of the polyhedron into the surface of the sphere. This projection does not change the number of faces, edges, or vertices of the poyhedral net. Think of a unit sphere (where the unit is large enough to have all vertices inside) and that we inflate the polyhedron until it becomes a spherical polyhedron. Now since the polyhedron is now sitting in a unit sphere we can add all the areas of spherical polygons making up the spherical polyhedron and find, by using equation 2.8

$$4\pi = \sum A_i = 2\pi F - \sum_i \sum_j \beta_{ij} \tag{3.2}$$

where i is an index for a face with area A_i, F is the number of faces, and β_{ij} is the j-th exterior angle corresponding to face i.

We want to compute the double sum above. The angles β_{ij} are external angles. Now we know that corresponding internal angle to β_{ij} is α_{ij} such that $\alpha_{ij} + \beta_{ij} = \pi$. At each vertex of a polygon the sum of all internal angles is 2π. We can re-order the angles in the following way. For each vertex v on a face f we can label the edges of that vertex as E_i then at each vertex v we have

$$\sum_{j=1}^{E_i} \alpha_{ij} = 2\pi.$$

On the other hand, since $\alpha_{ij} = \pi - \beta_{ij}$ we can write

$$2\pi = \sum_{j=1}^{E_i} \alpha_{ij} = \sum_{j=1}^{E_i} \pi - \beta_{ij} = E_i\pi - \sum_{j=1}^{E_i} \beta_{ij}.$$

so

$$\sum_{j=1}^{E_i} \beta_{ij} = E_i\pi - 2\pi.$$

We now add over all vertices re-labelled as i,

$$\sum_{i=1}^{V} \sum_{j=1}^{E_i} \beta_{ij} = \sum_{i=1}^{V} E_i\pi - 2V\pi.$$

Now the total number of edges is given by

$$E = \frac{\sum_{i=1}^{V} E_i}{2},$$

so we find

$$\sum_{i=1}^{V}\sum_{j=1}^{E_i}\beta_{ij} = 2(E-V)\pi$$

We now combine this equation with equation 3.2 to find

$$4\pi = 2\pi F - 2(E-V)\pi$$

That is,

$$V - E + F = 2.$$

□

Where did we use the fact that the polyhedron was convex? It could happen that the polyhedron is non-convex and still the formula above would be valid. Think of a polyhedron on a sphere. Then take a point and push it inside a small distance creating a dimple. The new polyhedron is not convex and still the formula above is valid. The problem with non-convexity is that if the solid is non-convex the projection in the sphere can produce polygons that overlap and then we would not have a polygon on a sphere but some other object.

A second theorem of fundamental importance is due to Cauchy. We A.D. Alexandrov book Convex Polyhedra [1] as a reference for the Cauchy lemma shown below. This might be no needed.

3.1.1 The Dihedral Angle

We provide the tools required to understand the computation of dihedral angles in regular polyhedra. We use the book Spherical Trigonometry [2] by I. Todhunder as a reference for the dihedral angle derivations and in particular Figure 3.1 below.

We assume that each face of the polyhedron is a regular polygon, and that the number of edges leaving each vertex is the same along the polyhedra set

[1] http://www.springer.com/cda/content/document/cda_downloaddocument/9783540231585-c1.pdf?SGWID=0-0-45-140315-p37887978

[2] http://www.gutenberg.org/files/19770/19770-pdf.pdf

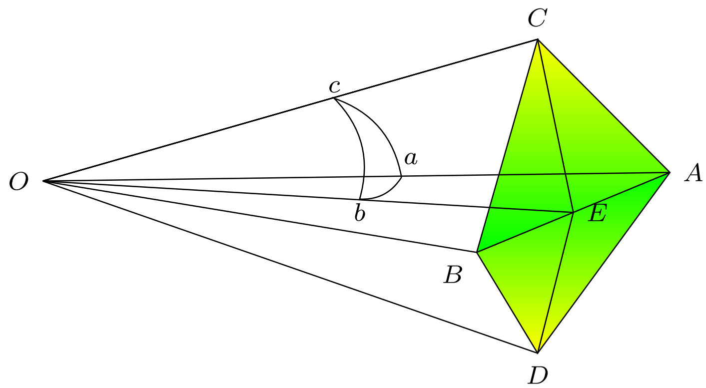


Figure 3.1: Computing the dihedral angle.

of vertices. Given two faces of the polyhedron, let us choose the segment AB to be the common edge between the faces with the midpoint at E. We name the centers of the faces as C and D. We joint the centers of the two faces with the midpoint E. Then CE and DE are orthogonal to the common edge AB. The dihedral angle is the angle $\angle CED$. Since C, E, and D are in the same plane we can trace right angles at C and D in the direction of the closing (dihedral) angle $\angle CED$. This lines meet at a point O. That is the lines CO and DO are perpendicular to CE and DE. We now think of an sphere centered at O and that meets the segments OA, OC, and OE at the points a, b, and c respectively on the spherical triangle $\triangle abc$. Since the dihedral angle CEA in the prism $OCEA$ is a right angle, we have that the triangle $\triangle abc$ is a right spherical triangle with the angle at b being a right angle. Let us now assume that m is the number of sides in each face of the polyhedron, and n the number of edges meeting at each vertex. We find then that half of the central angle of each face has the size of $\angle ACE$ or $\angle ace$, which is $2\pi/2m = \pi/m$. Let us now focus on the vertex A. This is a vertex of the polyhedron. Since C and D in the center of the two adjacent faces, the vertex A is surrounded by $2n$ angles such as the angle CAE, all of them congruent. That is, $\angle CAE = \angle cae = 2\pi/n = \pi/n$.

We now use equation 1.14. That is,

$$\cos \angle cae = \cos cOe \, \sin ace.$$

Since the dihedral angle we are searching for is $\delta = \angle CED$ and $\angle ECO$ is a right angle, then $\angle cOe = 90 - \delta/2$. Then, in radians,

$$\cos\frac{\pi}{n} = \cos\left(\frac{\pi}{2} - \frac{\delta}{2}\right)\sin\frac{\pi}{m}.$$

Then the dihedral angle δ can be found from the formula:

$$\sin\frac{\delta}{2} = \frac{\cos(\pi/n)}{\sin(\pi/m)}. \tag{3.3}$$

This shows that the dihedral angle is the same as long as the faces of the polyhedron are regular and the number of angles formed at each vertex (here m) is the the same. We can write this as a theorem as follows:

Theorem 3.1.2. *In a polyhedron with all regular and congruent faces the following two statements are equivalent:*

(i) *The number of edges at each vertex is the same.*

(ii) *All dihedral angles are congruent.*

Proof.

(1) We already proved that "(i) $\Longrightarrow$ (ii) ".

(2) The proof that "(ii) $\Longrightarrow$ (i) " is based on equation 3.3 . We derived the value of $\sin\delta/2$ as a function of the number of edges m that leave a given vertex. If the angle δ is the same on each case then m is constant, since n is assumed to be constant by hypothesis.

□

3.2 Platonic Solids

Before showing the construction of Platonic solids we proof the fundamental theorem.

Theorem 3.2.1. *There are only 5 regular polyhedra.*

Proof. We assume that the faces are regular polygons and at each vertex there is an equal number of converging faces. The proof of this theorem is based on Euler's equation 3.1. Let us assume that m is the number of edges in each regular polygonal face and n is the number of internal angles at each vertex. The total number of internal angles k can be written in several ways. That is,

$$k = mF = nV = 2E$$

where F and V are the number of faces and vertices respectively. With the addition of Euler's formula $V - E + F = 2$ and we can solve the three linear equations above for V, E, F in terms of m and n. That is,

$$\begin{aligned} V &= \frac{4m}{2(m+n) - mn} \\ E &= \frac{2mn}{2(m+n) - mn} \\ F &= \frac{4n}{2(m+n) - mn} \end{aligned} \tag{3.4}$$

We know that V, E, and F should be positive numbers. Hence $2(m+n) > mn$, and so

$$\frac{1}{m} + \frac{1}{n} > \frac{1}{2}.$$

We know that $n \geq 3$ so $1/n \leq 1/3$, and $1/m > 1/2 - 1/3 = 1/6$. Then $m < 6$. Since m is the number of sides of each regular polygon it should then be $3, 4$, or 5. That is, the faces of a regular polyhedron should be either triangles, squares, or pentagons. We can try m, n in equations 3.4 above. Without starting trying there are 9 combinations of (m, n) pairs. Not all of them produce integer numbers. For example, the combinations $(m, n) = (5, 4)$, $(m, n) = (4, 5)$, $(m, n) = (5, 5)$, and $(m, n) = (4, 4)$ violate the $2(m+n) > mn$ constraint. That leaves us with the remaining 5 combinations shown in the table 3.1

□

Table 3.1: Platonic solids and their graph data.

m	n	V	E	F	Solid
3	3	4	6	4	Tetrahedron
4	3	8	12	6	Hexahedron(cube)
3	4	6	12	8	Octahedron
5	3	20	30	12	Dodecahedron
3	5	12	30	20	Icosahedron

By reviewing the proof above we see that we did not use regularity other than the number of sides on each face is the same, and the number of edges arriving at each vertex is the same. This means that we found five cathegories of solids. Each solid could be stretched or squeeze without breaking it and still the relation between edges, faces, and vertices is preserved but the regularity or equality of the sides and angles violated. The fact that indeed extist exactly 5 Platonic (regular) solids is shown by construction below.

We show a construction of the five Platonic solids following known methods. For a different approach to the construction of the Platonic solids the reader can check my notes [3] on Platonic solids.

We can use equation 3.3 in combination with table 3.1 to find the exact dihedral angle for each of the 5 Platonic solids.

By definition a polygon is regular if is equilateral and equiangle. The same definition is extended to regular polyhedra.

Definition 10 (**Regular Polyhedron**)**.** *A* **regular polyhedron** *satisfies two conditions:*

(i) *All faces are regular polygons*

(ii) *The dihedral angles (between two adjacent faces) are all congruent.*

We will not compute explicit dihedral angles for all the Platonic solids, we use theorem 3.1.2 and count the number of edges at each vertex which is much easier than compute dihedral angles.

We compute the vertices of each regular polyhedra (which are the Platonoic solid) and verify that they are such that the polyhedron is indeed regular.

[3] file:///home/herman/Technical/LandMark/PlatonicSolids/input.pdf

Figure 3.2: On the left an equilateral triangle divided into 4 equilateral triangles of side a. Folding the three outside equilateral triangles to join their tips produces the figure in the right which is a tetrahedron seen from above.

3.2.1 The Tetrahedron

Before we relate a regular tetrahedron with a sphere, we describe some basic properties. A tetrahedron is formed by three equilateral triangles. Figure 3.2 shows how to fold the tetrahedron from an equilateral triangle of size $2a$ into a tetrahedron of size a.

We have then that $m = 3$ and $n = 3$. So the dihedral angle of the tetrahedron is given by

$$\delta = 2 \arcsin\left(\frac{\cos(\pi/3)}{\sin(\pi/3)}\right) = \arcsin\left(\frac{\sqrt{3}}{3}\right) \approx 70.7288^\circ.$$

We now divide the construction of the tetrahedron on a sphere in three parts.

3.2.1.1 The base

The base of a tetrahedron is an equilateral triangle of side size a. From this triangle we want to find the height (or altitude) [4] h and the center. From elementary geometry the three altitudes intersect at the same point or ar concurrent. Also, for an equilateral triangle the three altitudes also are concurrent at the point where the perpendicular bisectors meet. That point

[4] perpendicular from one vertex to the opposite side

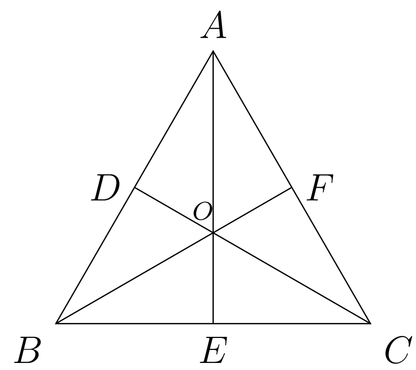


Figure 3.3: Base of a tetrahedron. The computations on the main text use this figure as reference.

will be equidistant to each vertex of the triangle, so the center of the triangle is the center of a circle having the three points of the triangle. Figure 3.3 shows the base of the tetrahedron with the three altitudes drawn. The sides of the triangle are $AB = AC = BC = a$. The altitude is $h = AE$ found by using the pythagoras theorem

$$h = \sqrt{AC^2 - EC^2} = \sqrt{a^2 - (a/2)^2} = \frac{\sqrt{3}}{2}a.$$

The angle $\angle OCE$ is equal to $30°$, so since $\sin(\angle OCE) = 1/2$, we have that

$$\frac{OE}{OC} = \frac{1}{2},$$

but since the triangles $\triangle AFO$ and $\triangle CFO$ are congruent then $OE + OC = OE + OA = h$, so

$$OE + 2OE = h,$$

from which $OE = h/3$. That is, the center is a third of the altitude from the base.

3.2.1.2 The altitude

The computation of the altitude of the tetrahedron is based on Figure 3.4 We

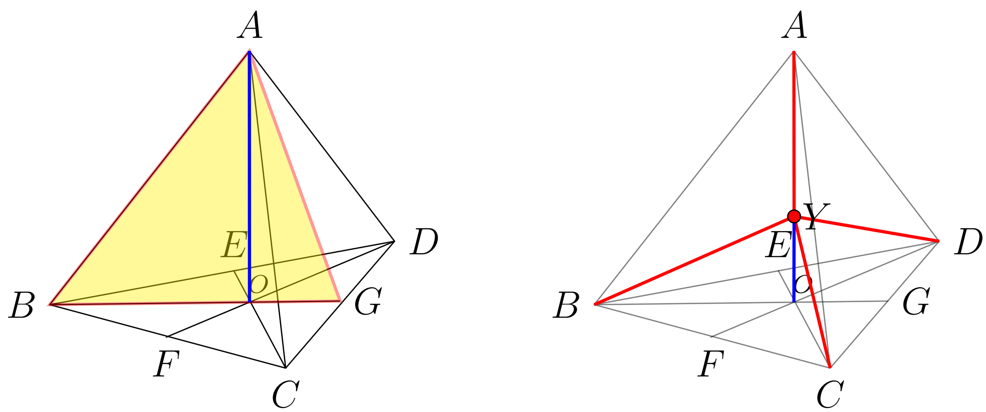


Figure 3.4: On the left frame the height of the tetrahedron is the blue segment $H = AO$. It is computed from the yellow triangle using Pythagoras theorem. On the right frame we add the point Y wich is the center of the tetrahedrom.

should focus in the vertical yellow triangle $\triangle ABG$. The altitude is $H = AO$. Since $OB = 2h/3$, we have from Pythagoras theorem

$$H = AO = \sqrt{AB^2 - OB^2} = \sqrt{a^2 - 4h^2/9} = \sqrt{a^2 - 4(3a^2/4)/9} = \frac{\sqrt{6}}{3}a$$

3.2.1.3 The circumscribed sphere

Any point in altitude AO is equidistant from the points B, C, and D. The center of the sphere is in a point Y such that $YA = YB = YC = YD$. This is illustrated in the right frame of Figure 3.4. There all red segments are congruent. The triangles $\triangle YOC$, $\triangle YOD$ and $\triangle YOB$ are all congruent since they are all right triangles sharing the same side YO and with the same length of hyptotenuse (red segments). We have then that $YD = AY$, $OY = AO - AY = H - AY$. Now

$$AY = YC = \sqrt{OY^2 + OC^2} = \sqrt{OY^2 + 4h^2/9}.$$

so

$$OY = H - \sqrt{OY^2 + 4h^2/9},$$

That is

$$\cancel{OY^2} - 2(OY)H + H^2 = \cancel{OY^2} + 4h^2/9,$$

so

$$OY = \frac{H^2 - 4h^2/9}{2H} = a\frac{2/3 - 1/3}{2\sqrt{6}/3} = a\frac{1/\cancel{3}}{2\sqrt{6}/\cancel{3}} = a\frac{\sqrt{6}}{12} = \frac{H}{4}.$$

The radius of the circumscribed sphere is

$$R = AO - OY = H - \frac{H}{4} = \frac{3H}{4} = \frac{\sqrt{6}}{4}a.$$

So we have that the center is located at 1/4-th of the altitude with respect to the base triangle, and the radius is given by $R = a\sqrt{6}/4$. We use captital R because small r is used for the inscribed sphere. We will not study the inscribed sphere on this document since by definition the spherical polyhedra have their vertices lying on the sphere. The inscribed sphere is tangent to the faces (walls) of a polyhedron and the vertices are inside the polyhedron.

3.2.1.4 Coordinates of the tetrahedron

We studied the geometry of the tetrahedron and the distribution of the vertices with respect to its center. Since we want to relate the tetrahedron to a sphere we want to move the vertices such that the center of the tetrahedron is at the origin $O = (0,0,0)$. Remapping the vertices we now, referring to right frame of Figure 3.4, want say that $Y = O = (0,0,0)$, and that $A = (0,0,AY) = (0,0,r) = (0,0,a\sqrt{6}/4)$. The other three vertices of the tetrahedron are down in a horizontal plane $z = -OY = -H/4 = -a\sqrt{6}/12$. The (x,y) components of those three vertices are in a circle of radius $R_0 = OC = 2h/3 = a\sqrt{3}/3$. They could be rotated with respect to the z axis and the tetrahedron still be preserved as such. We pick $\theta_i = i(2\pi/3)$, and $i = 0,1,2$, so the three vertices are located at the points

$$X_i = (0,0,-a\sqrt{6}/12) + R_0(\cos\theta_i, \sin\theta_i, 0),$$

or explicitly

$$\begin{aligned} T_1 &= X_0 = a(\sqrt{3}/3, 0, -\sqrt{6}/12) \\ T_2 &= X_1 = a(-\sqrt{3}/6, 1/2, -\sqrt{6}/12) \\ T_3 &= X_2 = a(-\sqrt{3}/6, -1/2, -\sqrt{6}/12) \\ N &= (0,0,\sqrt{3}). \end{aligned} \tag{3.5}$$

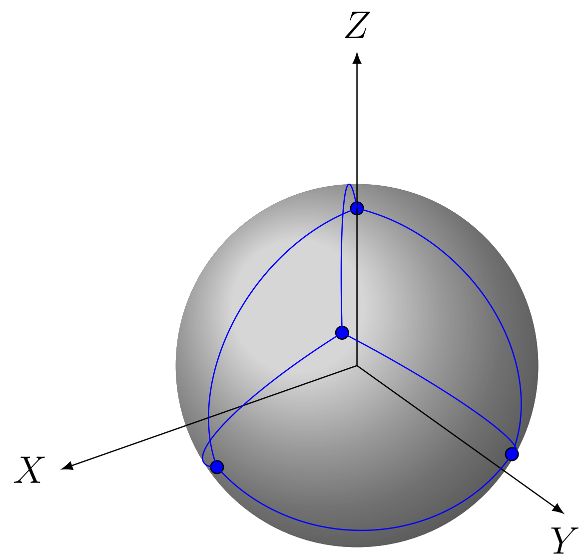


Figure 3.5: Tetrahedron in the sphere. We see the 4 equilateral triangles each occupying each one fourth of the area of the sphere.

Figure 3.5 shows an sketch of the tetrahedron in the sphere. We observe that the tetrahedron divides the sphere in 4 congruent regions, each an equilateral triangle.

3.2.1.5 The altenating cube

It is a well known fact that a tetrahedron can be constructed from a cube by tracing diagonals. This can be found, for example, in the Wikipedia website for the tetrahedron.

We want to relate the tetrahedron obtained by cutting through diagonals with the tetrahedron built above. For this purpose we need to a rigid rotation. We show below the rotations that map one tetrahedron into the other. My document Notes about 3x3 Rotation Matrices

Coordinate rotations are easily performed with matrix product. We will perform two rotations.

For the first rotation we use a matrix P [5] around the Y axis, with an angle of $\theta = \arccos(\sqrt{3}/3) \approx 54°$. This is,

[5] here "P" stands for pitch from navigation coordinate system.

$$P = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{3} & 0 & -\frac{\sqrt{2}}{\sqrt{3}} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{3} & 0 & \frac{\sqrt{3}}{3} \end{pmatrix}$$

The second is a rotation about the Z axis and we denote it by Y [6] axis by an angle $\alpha = \pi/4 = 45°$. This is the matrix

$$Y = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We use the T_i, $i = 1, 2, 3$ as shown in equation 3.5, together with the he apex $N = A = (0, 0, a\sqrt{6}/4)$. It can be verified that

$$\begin{aligned} YPT_1 &= (1,1,1) \quad , \quad YPT_2 = (-1,1,-1) \\ YPT_3 &= (1,-1,-1) \quad , \quad YPA = (-1,-1,1) \end{aligned}$$

Figure 3.6 illustrates the original tetrahedron in blue, with $a = 2\sqrt{2}$ as well as the rotated version in red. The red tetrahedron with vertices $(1, 1, 1)$, $(-1, 1, -1)$, $(1, -1, -1)$, and $(-1, -1, 1)$ are found by starting at the cube vertex $(1, 1, 1)$ and choosing every other vertex so that there are no two consecutive vertices. This is called the altenating cube.

3.2.2 The Hexahedron (cube)

The hexahedron is the simplest and perhaps most common of the platonic solids. We can collocate all the points of a cube in a unit sphere. All the points would be of the form $R(\pm1, \pm1, \pm1)$, for a total of 8 points, with $R = \sqrt{3}/3$.

Figure 3.7 shows a regular cube on the left and a curved cube on the sphere dividing the sphere in 6 congruent curved squares.

We already know that the dihedral angles of the regular hexahedron are all 90°. We verify this using equation 3.3. That is, in the cube $m = 4, n = 3$ and so

[6] Here "Y" stands for yaw which is a term used in navigation.

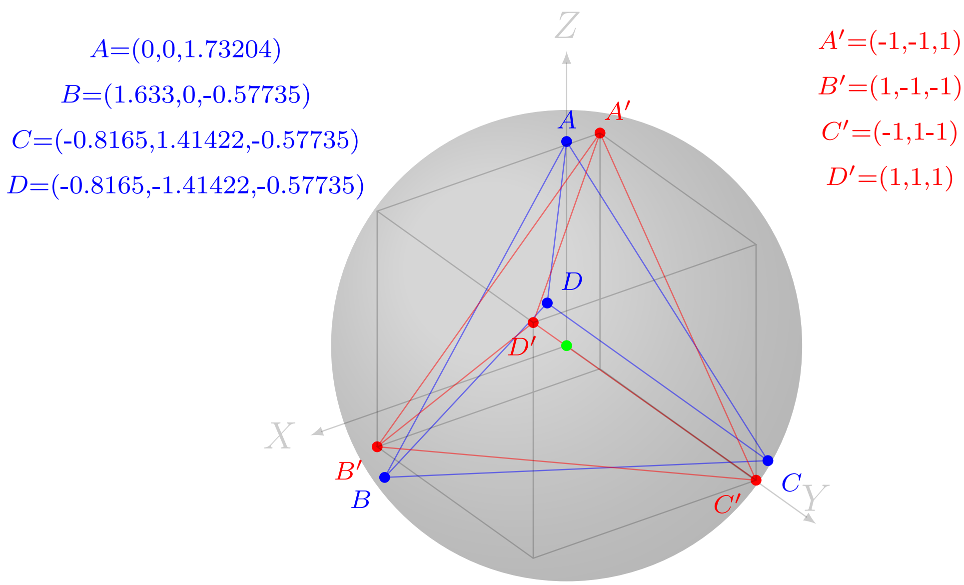


Figure 3.6: An original tetrahedron in blue and its rotated version in red. The red tetrahedron is sitting in vertices of the cube in such a way that there are no two consecutive cube vertices.

$$\sin\frac{\delta}{2} = \frac{\sin(\pi/3)}{\cos(\pi/4)} = \frac{1/2}{\sqrt{2}/2} = \frac{\sqrt{2}}{2}.$$

That is $\delta/2 = 45°$, and $\delta = 90°$.

3.2.3 The octahedron

The octahedron can be seen as two tetrahedra (pyramids) with a square base joined by the base. The base is easy to understand since it is a square of side a. This provides 4 of the six vertices of the octahedron. To find the remaining two vertices we need to find the altitude of the tetrahedra. Figure 3.8 shows half of the octahedron. We use the yellow triangle to find the altitude NB as follows. In the triangle $\triangle NCD$ since the angle at C is right, and the distance CD is half of the side, that is $CD = a/2$ we have that

$$CN = \sqrt{DN^2 - CD^2} = \sqrt{a^2 - \frac{a^2}{4}} = \frac{a\sqrt{3}}{2}.$$

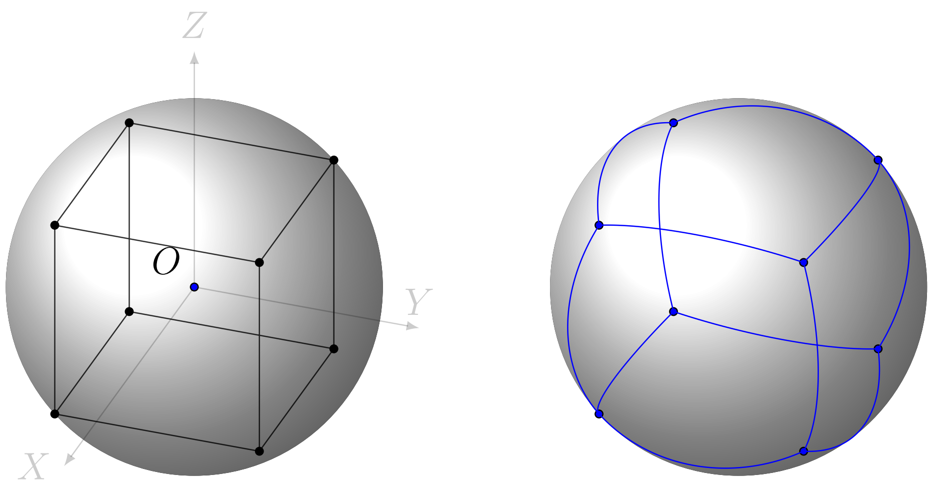


Figure 3.7: On the left a cube inscribed in a sphere. On the right the points are joined by arcs of great circles forming a curved cube.

Now we use the Pythagoras theorem on the yellow triangle. That is,

$$NB = \sqrt{CN^2 - BC^2} = \sqrt{\frac{3a^2}{4} - \frac{a^2}{4}} = \frac{a\sqrt{2}}{2}.$$

So the north and south poles would be located at $\pm a\sqrt{2}/2$.

If the four vertices in the base of the top pyramid are $(\pm 1, 1, 0)$ and $(1, \pm 1, 0)$, then octagon side is $a = \sqrt{2}$ and so the north and south poles are at $(0, 0, \pm 1)$. Hence the six vertices of the octahedron are:

$$\{(\pm 1, 1, 0), (1, \pm 1, 0), (0, 0, \pm 1\}.$$

Figure 3.9 shows on the left an octahedron inscribed in a sphere. On the right the front (or top if you wish) side of the octahedron with 4 equilateral triangles with angles $90°, 90°, 90°$.

We could rotate the octagon $\pi/4$ with respect to z axis and obtain instead the following eight vertices.

$$\{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1\}.$$

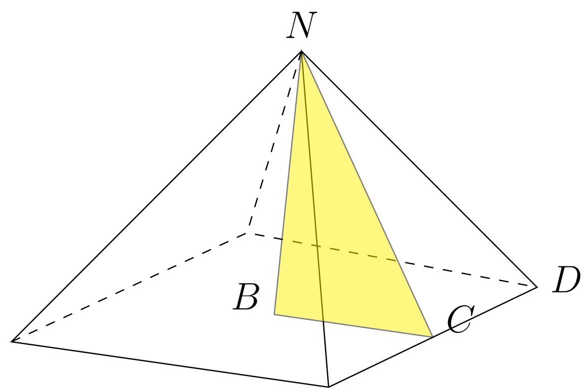


Figure 3.8: We see half of the octahedron. The yellow triangle help us find the altitude NB by using Pythagoras theorem.

From the number of edges per face and number of edges at each vertex we see that in the octahedron $m = 3, n = 4$. So, from equation 3.3

$$\sin\frac{\delta}{2} = \frac{\cos(\pi/4)}{\sin(\pi/3)} = \frac{\sqrt{2}/2}{\sqrt{3}/2} = \frac{\sqrt{2}}{\sqrt{3}}.$$

That is $\delta/2 \approx 54.74°$, and $\delta \approx 109.45122$.

3.2.4 The Dodecahedron

The dodecahedron and the icosahedron are the most complex of the Platonic solids due to the high number of faces. Euclid's method to build a dodecahedron is to start with a cube and collocate at each side of the cube a "roof" shaped structure. Figure 3.10 illustrates

The construction of the dodecahedron here is based on Euclid's Book XIII [7].

Euclid's method. We observe that each plane face is a pentagon with two of its sides in the triangular face of the roof-like shape and the other tree sides in the trapezoidal of an adjacent roof-like shape. The common edge between the two polygons is a side of the cube. For example see that, after

[7] http://aleph0.clarku.edu/ djoyce/elements/bookXIII/propXIII17.html

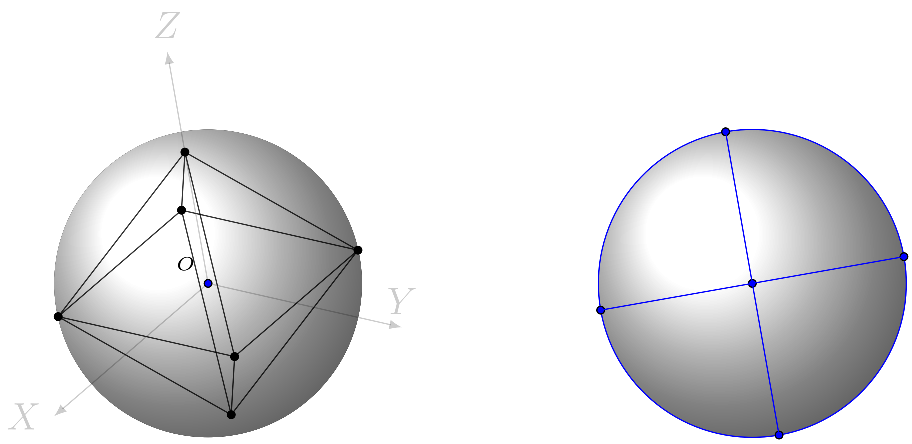


Figure 3.9: On the right we see an octahedron inscribed in a sphere. On the right we see half of the octahedron (top or front) with four equilateral triangles with angles of $90°, 90°, 90°$.

joining the blue triangle with the blue trapezoidal we should get a regular pentagon.

We need to find the vertices of the dodecahedron, and impose the following two conditions:

(i) The pentagons are regular. That is, their sides and angles are all congruent.

(ii) Pentagonal faces are flat. That is we need to verify that the triangle of a roof-like structure with a trapezium of an adjacent roof-like structure (see the blue shapes in Figure 3.10) are in the same plane.

Let us assume that the cube has the collection of eight vertices $(\pm 1, \pm 1, \pm 1)$. Here, the notation $(\pm a, \pm b, \pm c)$ indicates all the 8 combinations of the form (a, b, c) with either a "+", or a " $-$ " sign in front of each coordinate. These vertices also belong to the dodecahedron but we need to find a few (12) more. We see first that all the points are sitting in sphere of radius $\sqrt{3}$. This adds an extra constraint to the rest of the points. Their size should be $\sqrt{3}$.

The diagonals of the pentagon are given (see my notes [8] in Platonic solids)

[8] file:///home/herman/Technical/LandMark/PlatonicSolids/input.pdf

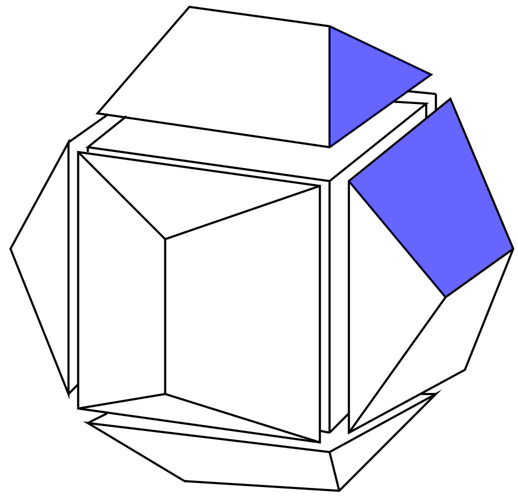


Figure 3.10: Euclid's method to build a dodecahedron. A cube is surrounded by four 'roof-like" shapes sitting on each side. Each triangle with the roof-like shape together with a trapezoid from an adjacent roof-shape produced a pentagon. The blue shaded polygons show this.

by the side times the Golden ratio [9] $\phi = (1 + \sqrt{5})/2$. That is,

$$d = \phi a \tag{3.6}$$

but d is the base of each triangle in the roof-like structure. Since $d = 2$, we have that the side of the dodecahedron is

$$a = \frac{2}{\phi}.$$

Initially we want to find the top two vertices of the blue plane in Figure 3.10. Let us focus on the top roof-like structure which we show in Figure 3.11. Consider a bisector plane through the roof-like structure going through the top side BC and the middle line on the base NM as shown in the figure. Also, trace perpendiculars from BO and CP from the top to the base of the structure. We have the following measures as functions of the Golden ration ϕ:

[9] https://en.wikipedia.org/wiki/Golden_ratio

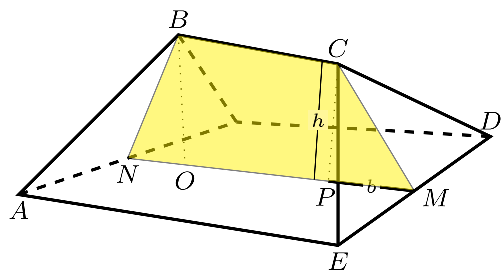


Figure 3.11: Roof-like structure. The main purpose of this is to find the two offsets h (for altitude) abd b for a base offset. These offsets measure the distance to the center cube on each of the sides where the roof-like structure is collocated.

$$\begin{aligned}
BC &= OP = a = \frac{2}{\phi} \\
MN &= 2 \\
ON &= PM = \frac{2-a}{2} = \frac{2-2/\phi}{2} = \frac{\phi-1}{\phi} \\
CM &= \sqrt{(CD)^2-(DM)^2} = \sqrt{a^2-1} = \sqrt{\frac{4}{\phi^2}-1} = \frac{\sqrt{4-\phi^2}}{\phi} \\
PC &= \sqrt{(CM)^2-(PM)^2} = \sqrt{\frac{4-\phi^2}{\phi^2}-\frac{(\phi-1)^2}{\phi^2}} = \frac{\sqrt{3-2\phi^2+2\phi}}{\phi}
\end{aligned}$$

Now, since $3-2\phi^2+2\phi = 1$ [10] we find that the altitude of the roof-like structure is $PC = h = \frac{1}{\phi}$, we name $b = ON = PM = (\phi-1)/\phi = 1-1/\phi$ which is another important element to compute the other 12 vertices of the dodecahedron. Both h and b are offsets from the cube in orthogonal directions from each other. We rewrite them here for clarity of exposition since they will be used repeatedly in the computations below.

[10]This could be verified by direct evaluation or by using the equation $\phi^2-\phi-1=0$, which is used sometimes to define the Golden ratio. That is, $3-2\phi^2+2\phi = 1-2(\phi^2-\phi-1) = 1$.

$$\begin{aligned} h &= \frac{1}{\phi} \\ b &= \frac{\phi - 1}{\phi} = 1 - \frac{1}{\phi} = \frac{1}{\phi^2}, \end{aligned}$$

since $\phi = 1 + 1/\phi$.

We have all the elements to complete the other 12 vertices of the dodecahedron. We do this in 6 couples as follows

(i) **Top and bottom vertices:** We find the vertices B and C and their reflections with respect to the XY plane at the bottom. The vertex at B has an x coordinate of $-1 + ON = -1 + b = -1 + 1 - 1/\phi = -1/\phi$. The y coordinate is $y = 0$, and the z coordinate is $1 + h = 1 + 1/\phi$. The vertex C is the vertex B plus the coordinate $(a, 0, 0) = (2/\phi, 0, 0)$. That is, the vertex at C is $(1/\phi, 0, 1 + 1/\phi)$. The bottom vertices are a reflection where the sign of z is changed. That is the top and bottom vertices are:

$$(\pm 1/\phi,\ 0\,, \pm(1 + 1/\phi))$$

It can be verified that all these points sit in a sphere of radius $\sqrt{3}$.

(ii) **Left and right vertices :** All these vertices have the same z coordiante $z = 0$. The left front vertices share the same x coordinate $x = -1 - h = -1 - 1/\phi$. The y coordinate of the left front vertex is $1 - b = 1 - 1 + 1/\phi = 1/\phi$. The back coordinate has $y = -1 + b = -1/\phi$. The right vertices are a reflection with respect to the YZ plane (change sign of x) of these two. That is, in summary the left and right vertices are given by the combinations:

$$(\pm(-1 - 1/\phi),\ \pm 1/\phi\,, 0)$$

Again, all these points are collocated in a sphere of radius $\sqrt{3}$.

(iii) **The front and back vertices :** The front top and bottom x coordinates are $x = 0$, the front top and bottom y coordiantes are located at the same y coordinate $y = 1 + h = 1 + 1/\phi$. The front top vertex is located at $z = 1 - b = 1 - (1 - 1/\phi) = 1/\phi$, while the front bottom vertex is $-1 + b = -1 + (1 - 1/\phi) = -1/\phi$. The back vertices are a reflection of these by switching the sign of y. That is the front and back vertices are given by

$$(0, \pm(1 + 1/\phi), \pm 1/\phi)$$

These four points are located in a sphere of radius $\sqrt{3}$.

This completes the 20 vertices of the dedecahedron.

Finally we need to verify that the faces of the triangles and the trapezoids are sitting in the same plane for each of the pentagons. To verify this we find the normal vectors to a triangle and a contiguous trapezoid. For example in Figure 3.10 we find the normal to the two blue polygons. Normals are easy constructed by taking cross products. For example the blue triangle can be seen as the triangle $\triangle CED$ in Figure 3.11 Consider the points (vectors) top-right C, cube top-front-right E and cube point D top-back-right. That is,

$$\begin{aligned} C &= (1/\phi, 0, 1 + 1/\phi) \\ E &= (1, 1, 1) \\ D &= (1, -1, 1) \end{aligned}$$

Two vectors involved are

$$\begin{aligned} \overrightarrow{CE} &= E - C = \left(1 - \frac{1}{\phi}, 1, -\frac{1}{\phi}\right) \\ \overrightarrow{CD} &= D - C = (1 - 1/\phi, -1, -1/\phi) \end{aligned}$$

Then

$$n_{CED} = \overrightarrow{CE} \times \overrightarrow{CD} = -2\left(\frac{1}{\phi}, 0, \frac{\phi - 1}{\phi}\right) = \frac{-2}{\phi}(1, 0, \phi - 1).$$

Likewise for the blue trapezoid we consider the three points: E front-top-right, F back,top-right, and the center-right-front on the roof-like structure G.

$$\begin{aligned} E &= (1,1,1) \\ F &= (1,-1,1) \\ G &= (1+1/\phi\,,\ 1/\phi\,,\ 0) \end{aligned}$$

We then build the cross product

Two vectors involved are

$$\begin{aligned} \overrightarrow{EF} &= F - E = (0,-2,0) \\ \overrightarrow{EG} &= G - E = \left(\frac{1}{\phi}, \frac{1}{\phi} - 1, -1\right) \end{aligned}$$

$$n_{EFG} = \overrightarrow{EF} \times \overrightarrow{EG} = \left(2, 0, \frac{2}{\phi}\right) = 2\left(1, 0, \frac{1}{\phi}\right).$$

and since $1/\phi = \phi - 1$ we see that the vectors n_{CED} and n_{EFG} are colinear. That is the blue polygons are sitting at the same plane. Due to symmetries all the other polygons satisfy the same requirement and then the dodecahedron is built.

Figure 3.12 shows a dodecahedron inscribed in a sphere and a dodecahedron as a spherical polyhedron.

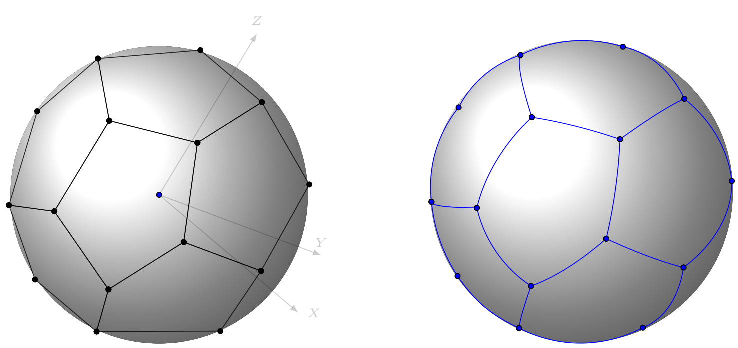


Figure 3.12: On the left we see an dodecahedron inscribed in a sphere. On the right we see the corresponding spherical (curved) pentagons for the plot on the left.

3.2.4.1 Radii of circles and spheres related to the tetrahedron

Think of a sphere of radius R_i. At some point on the sphere we have a tangent plane and in this plane we draw a regular pentagon kissing the sphere at point O'. Figure 3.13 shows an example where a pentagon is tangent to a sphere

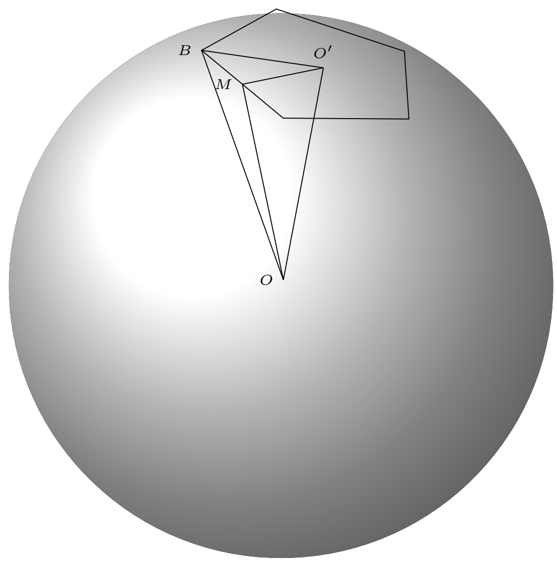


Figure 3.13: A face of a regular polygon (a pentagon) tangent to a sphere at point O'.

at the point O'. The radius of the sphere is the distance $r = OO'$. Let us assume that the side of the pentagon is a. M is a midpoint along one side of the pentagon. The triangles $\triangle O'MB$, $\triangle MO'O$, and $\triangle BO'O$ are all right triangles, with the right angle at the center letter. By applying three times the Pythagoras theorem we have that

$$\begin{aligned}(O'B)^2 &= (BM)^2 + (O'M)^2 = \left(\frac{a}{2}\right)^2 + (O'M)^2 \\ (OM)^2 &= (OO')^2 + (O'M)^2 \\ (OB)^2 &= (OO') + (O'B)^2\end{aligned}$$

where a is the length of the side of the pentagon.

There are four radii here. Two radii of the circle in the plane having the pentagon and two radii in the sphere. The polyhedron is inscribed in the sphere with a radius R_o, and has an inscribed sphere of radius R_i.

The pentagon is inscribed in a circle of radius r_o and has a circle inscribed of radius r_i. The radius r_i is also the apothem of the pentagon.

From the figure we identify:

$$R_o = OB \quad , \quad R_i = OO' \quad , \quad r_o = O'B \quad , \quad r_i = O'M.$$

We can write the three equations above in terms of the radii of the circle and sphere. That is

$$\begin{aligned} r_o^2 &= \left(\frac{a}{2}\right)^2 + r_i^2 \\ (OM)^2 &= R_i^2 + r_i^2 \\ R_o^2 &= R_i^2 + r_o^2 \end{aligned} \tag{3.7}$$

My notes [11] on Platonic solids show some relations on a pentagon. The central angle of the pentagon $\angle MO'B$ is $2\pi/10 = \pi/5$, and I showed that

$$\cos\frac{\pi}{5} = \frac{\phi}{2},$$

where ϕ is the Golden ratio. Then we have that

$$\frac{r_i}{r_o} = \frac{\phi}{2}. \tag{3.8}$$

Given a point O' on an inscribed sphere with radius R_i, we would like to build a Platonic solid such that this point is a point of tangency between the platonic solid and the sphere. If we can find the side length of the Platonic solid a (also a side length of the polygon) we would be able to build the polyhedron.

We need one more relation to link the equations above to the radius of the sphere having the vertices of the tetrahedron. We can join the diagonals

[11] file:///home/herman/Technical/LandMark/PlatonicSolids/input.pdf

of the pentagonal faces of the tetrahedron to build a cube. See, for example, Figure 3.10

Let us focus in the blue pentagon. The diagonal separates the pentagon into a triangle and a trapezoid. The other 11 diagonals (chosen here. Note that you could form up to 5 cubes by selecting different sets of diagonals) form a cube. The vertices of the cube lie in a sphere which is the same sphere circumscribing the dodecahedron. So, to find the radius of that sphere is just half of the diagonal of the cube.

Equation 3.6 shows that the diagonal of the pentagon is related to its side length a by $d = \phi a$, where $\phi = (1+\sqrt{5})/2$ is the golden ratio.

Now, the diagonal of the cube is

$$D = \sqrt{d^2 + d^2 + d^2} = \sqrt{3}d = \sqrt{3}\phi a. \tag{3.9}$$

So the radius of the circumsphere is

$$R_o = \frac{D}{2} = \frac{\sqrt{3}\,\phi a}{2} = \frac{\sqrt{3}(1+\sqrt{5})a}{4}. \tag{3.10}$$

but $\sin\pi/10 = (a/2)/r_o$. On the other hand

$$\sin\pi/10 = \sqrt{1-\cos^2(\pi/10)} = \sqrt{1-\frac{\phi^2}{4}} = \frac{1}{2}\sqrt{4-\phi^2}$$

We find then that

$$r_0 = \frac{a}{2}\frac{2}{\sqrt{4-\phi^2}} = \frac{a}{\sqrt{4-\phi^2}} \tag{3.11}$$

and we rewrite equation 3.7 as

$$\frac{3\phi^2 a^2}{4} = R_i^2 + \left(\frac{a}{2}\right)^2 \frac{4}{4-\phi^2} = R_i^2 + \frac{a^2}{4-\phi^2}.$$

Then

$$a^2\left(\frac{3\phi^2}{4} - \frac{1}{4-\phi^2}\right) = a^2\left(\frac{3\phi^2(4-\phi^2)-4}{4(4-\phi^2)}\right) = R_i^2,$$

and from here

$$a = \frac{2R_i\sqrt{4-\phi^2}}{\sqrt{3\phi^2(4-\phi^2)-4}} = \frac{2\sqrt{5-\sqrt{5}}}{\sqrt{3\sqrt{5}+7}}R_i. \tag{3.12}$$

The apothem $O'M = r_i$, from equation 3.8 is

$$r_i = \frac{\phi r_0}{2} \tag{3.13}$$

In summary: Provided a sphere with radius R_i, we can locate a tangent plane at any point of this sphere and find a pentagon on that plane as follows:

(i) Find the side length a using equation 3.12.

(ii) draw a circle of radius r_o found in equation 3.11. This is the cirumcircle around the pentagon.

(iii) draw a circle of radius r_i found in equation 3.13. This is the incircle inside the pentagon.

(iv) Pick any point in the circumcircle and raw a tangent to the incircle which intersects the circumcircle at two points (the starting point and another point) the second point is the start of a new tangent which will define the next point until the end of the loop when the last tangent intersects back the first point that we started with. The sets of tangets is the pentagon.

(v) The radius of the circumsphere with the vertices of the polyhedron is R_0 and can be found using equation 3.10.

(vi) In a similar way as the contruction of the polygon we can construct the polyhedra by tracing tangent planes to the sphere of radius R_i starting at two points of the pentagon (in the circumsphere of radius R_o) and finding the intersection of the plane with the (outer) circumsphere. This intersection is a small circle (does not go through the center). Two points are already defined for this new pentagon. These are the points in common between two neighboar pentagons. The other three can be constructed by tracing tangents between the inner and outer circles for the pentagon as in step (iv) above.

The implementation of steps (iv) and (vi) is done as follows. How do we find a line that intersects a circle in one point (tangent), given that we know the circle (centered at the origin) and a point to start. We define a set of lines all starting at the same point and having a set of directions. The point of intersection is such that the direction of the line is orthogonal to the direction of the point in the sphere (in a sphere a tangent is normal to the radius). At the same time the line at the sphere intersect at the tangent point. We can then use dot product and solve linear equations this way. The idea is extended to 3D with planes starting at two points and finding the third point (which defines the plane) such that the normal to the plane is parallel to the radius of the sphere. When ever that happens we identify the point of tangency in the sphere and from that point on we can build the pentagon given that we already know two points. This could be built using the method in (iv). The steps (iv) and (vi) run in a cycle until the polyhedron is completed. Twelve faces are constructed and twelve points of tangency are located at the center of those faces. These points of tangency are the vertices of equilateral triangles which from the icosahedron. This is one of the dualities discussed in section 3.2.6

Finally we know that since the faces are pentagons and each vertex has three edges we find that $m = 5, n = 3$ and using equation 3.3 we find

$$\sin\frac{\delta}{2} = \frac{\sin(\pi/3)}{\cos(\pi/5)} = \frac{1/2}{\sqrt{2}/2} = \frac{\sqrt{2}}{2}.$$

That is $\delta \approx 116.5650°$.

3.2.5 The Icosahedron

The construction of the icosahedron shown here is based on the method discovered by Piero della Francesca [12] in the fifteen century AD.

The icosahedron [13] has: twelve vertices, 20 faces, and 30 edges. Piero della Francesca thought about collocating the vertices in the faces of a cube. It is interesting that Euclid's did not catch (or at least it is not in his Elements books) this observation, since as we saw in the construction of the dodecahedron the design is built on top of a cube. It is also interesting that

[12] http://www.georgehart.com/virtual-polyhedra/piero.html

[13] https://en.wikipedia.org/wiki/Icosahedron

both, the dodecahedron and the icosahedron, use the cube as a reference and, as we will show, the Golden ratio plays a central roll in where the vertices are located with respect to the cubic surface.

Figure 3.14 shows an sketch of Piero della Francesca idea:

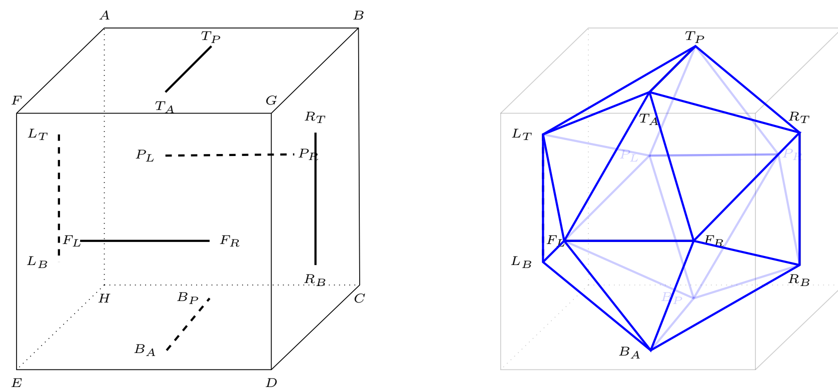


Figure 3.14: On the left we show a cube and a segment on each face of the cube. Each end point of the segment will be a point of the icosahedron. On the right is the picture of the icosahedron built by joining the points sitting on the cube in a prescribed order.

On the left frame of the figure we display a cube with vertices from A to H. To each face of the cube we attach an abbreviation. The abbreviations are T, B, L, R, A, P for top, bottom, left, right, anterior, and posterior faces respectively. This provides the nomenclature for the 12 vertices as shown in the figure.

We have the follow observations:

(i) All segments are centered at each face.

(ii) All segments are congruent (equal length).

(iii) The direction of a segment is such that it is orthogonal to to each of the segments in adjacent faces.

(iv) Segment in parallel faces are parallel.

(v) Segments have equal offsets from the edges of the cube.

The finding of these offsets is interesting. Here the Golden ratio, which was part of the analysis of the dodecahedron, plays an important roll as well. It is easy to see that if we joint any vertex with the 5 closest neighbors we get a 20 face polyhedron. This is a good start for the regular icosahedron but we need to guarantee that:

(i) All triangles formed in this way are equilateral.

(ii) The angle between any two adjacent triangular faces (dihedral angle) is the same.

(iii) All vertices are at the same distance from the center. That is, all vertices lie in a sphere with some radius R that we should find as a function of the side length a.

Using these constraints we will determine the exact position of each of the twelve vertices of the icosahedron. Since all triangles are equilateral then

$$d(T_P, R_T) = d(T_P, T_A) = a.$$

where $d(x, y)$ means distance between point x and point y. Now, the coordinates of those points are

$$T_P = (0, -a/2, 1) \quad , \quad R_T = (1, 0, a/2) \quad , \quad T_A = (0, a/2, 1).$$

with a the size of any edge. Then

$$d(T_P, R_T) = \sqrt{1 + a^2/4 + (1 - a/2)^2} = a$$

From which we find

$$1 + \frac{a^2}{4} + 1 + \frac{a^2}{4} - a = a^2 \quad \Longrightarrow \quad \frac{a^2}{2} + a - 2 = 0$$

or

$$a^2 + 2a - 4 = 0.$$

That is, by choosing the positive solution,

$$a = -1 + \sqrt{5}$$

There is no difference if we had use any other face of the cube to do the previous computation due to the symmetry of the problem. That is, all sides have the same length $a = -1 + \sqrt{5}$. This proofs the first statment above under item (i).

It is interesting to observe that

$$\frac{2}{a} = \frac{2}{-1 + \sqrt{5}} = \frac{1 + \sqrt{5}}{2} = \phi$$

where ϕ is the Golden ratio. This means that the ratio of the side of the cube (2) with respect to the side of the icosahedron $(-1+\sqrt{5})$ is the Golden ratio. Then the tree rectangles $R_1 = \{T_P, T_A, B_A, B_P\}$, $R_2 = \{L_T, L_B, R_T, R_B\}$, and $R_3 = \{F_L, F_R, P_L, P_R\}$ are Golden rectangles where the ratio between the largest and smaller side is the golden ratio. Figure 3.15 shows the icosahedron in Figure 3.14 with the three Golden rectangles R_1 (green), R_2 (yellow), and R_3 (red) included.

We will list all the vertices of the icosahedron in Figure 3.14 in terms of the Golden ratio. Look down from T_A to the pentagon with vertices L_T, P_L, P_R, R_T, and T_P. That is half of the icosahedron with the six vertices on the upper part. Sit now in the lowest point B_P and look up to the nearest 5 vertices above B_A, F_L, P_L, P_T, and R_B. These completes the 12 vertices of the icosahedron that we list with their coordinates as follows;

$$T_A = (0, -\frac{1}{\phi}, 1)$$

$$L_T = (-1, 0, \frac{1}{\phi})\ ,\ p_L = (-\frac{1}{\phi}, -1, 0)\ ,\ P_R = (\frac{1}{\phi}, 1, 0)\ R_T = (1, 0, \frac{1}{\phi})\ ,\ T_A = (0, -\frac{1}{\phi}, 1)$$

$$B_A = (0, -\frac{1}{\phi}, -1)\ ,\ F_L = (-\frac{1}{\phi}, -1, 0)\ ,\ P_L = (-\frac{1}{\phi}, 1, 0)\ ,\ P_R = (\frac{1}{\phi}, 1, 0)\ ,\ R_B = (1, 0, -\frac{1}{\phi})$$

$$B_P = (0, -\frac{1}{\phi}, -1)$$

We observe that all points have the components ± 1, $\pm 1/\phi$, and 0. That is, all points have a radius

$$R = \sqrt{1 + \frac{1}{\phi^2}} = \sqrt{\frac{1 + \phi^2}{\phi^2}}.$$

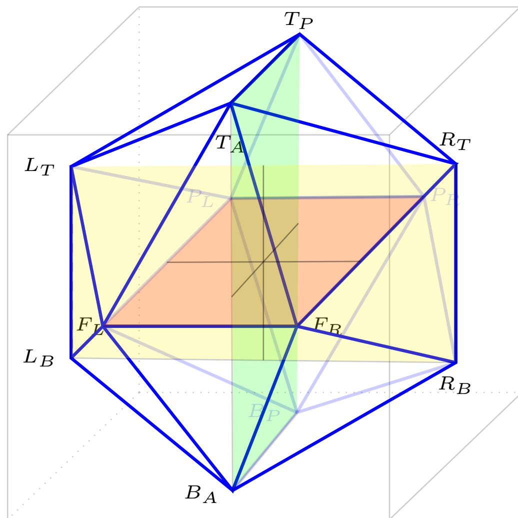


Figure 3.15: The icosahedron and its three Golden rectangles: R_1 vertical in the YZ plane with green color, vertical in the XZ plane with yellow color, and horizontal with red color.

Then all points are sitting on a sphere.

Finally, we observe that an icosahedron having triangular faces and each vertex connected to 5 edges has a pair $m = 3, n = 5$, and so

$$\cos\frac{\delta}{2} = \frac{\cos\pi/5}{\sin\pi/3}$$

from which $\delta \approx 138.1897$.

3.2.6 Dualities

This section explains the dualities between Platonic solids.

Appendix A

Calculus of Variations

A.1 Introduction

The subject of Calculus of Variations [1] started when Johann Bernoulli, in June 1696, posted a challenge to his colleges about finding the shortest time curve under a constant gravity field. The word Brachistochrone [2] from the ancient Greek means shortest time. What is interesting about the subject of calculus of variations is that the variables are not just real variables in an interval but functionals. A functional is a function from a vector space into the real numbers. Many solutions were presented including those of the Bernoulli brothers (Jacob and Johann), Newton, L'Hôpital, and Leibniz. Euler (1744) put the problem into a general setting, followed by Lagrange and Legendre (1786). The Wikipedia link above presents a brief history of the calculus of variations subject.

According to Andrej Chearkaev and Elena Cherkaev Lecture Notes [3] the brachistochrone problem was considered before by Galileo in 1638. He concluded that the straight line is not the fastest path between two points but he made an error concluding that an optimal trajectory is part of a circle. The actual trajectory which we will find later is known as a cycloid. It is interesting that if the velocity gradient (instead of the time derivative of the velocity which is the gravity) with depth (or height) is constant, the trajectory is a piece of circle as predicted by Galileo. For this, the reader can check

[1] https://en.wikipedia.org/wiki/Calculus_of_variations

[2] https://en.wikipedia.org/wiki/Brachistochrone_curve

[3] http://www.math.utah.edu/ cherk/teach/5710-03/print10-19.pdf

Slotnick Lessons in seismic computing [4]

The problem about the minimum path could be casted into an integral

$$t_{A\to B} = \int_A^B dt = \int_A^B \frac{ds}{v}, \tag{A.1}$$

where t is time, A and B are two fixed points in the two dimensional $\mathbb{R}^2$ space, v is the velocity (speed), as a function of (x, y), and s is the arc length. [5]

Figure A.1 sketches a path in $\mathbb{R}^2$ under a constant gravitational field g pointing along the $-y$ direction.

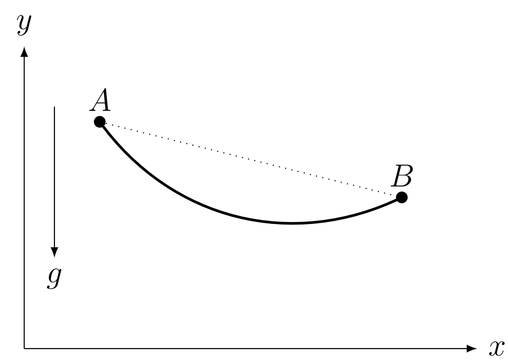


Figure A.1: The arc between the points A and B represents the shortest time path under the constant gravitational field g pointing down.

From differential geometry we have the arc length differential can be written as

$$ds = \sqrt{dx^2 + dy^2} = \sqrt{1 + y'}\,dx,$$

with

$$y' = \frac{dy}{dx}.$$

After Newton introduced his laws of mechanical motion which comes from integrating several time the second law (from force to acceleration, to speed,

[4] https://books.google.com/books?id=tLM_AAAAIAAJ&source=gbs_book_other_versions

[5] If the velocity v is a wavespeed such an electromagnetic wave for seismic wave, then the equation above is a ray tracing equation and from it we can follow Fermat's principle and Snell's law.

to distance), some other approaches surfaced. These approaches were based on conservation principles. One of them was the conservation of energy. We find the speed from this principle. That is, the total energy is the kinetic plus potential energy. At the beginning we assume that the particle is at rest at point $A = (x_A, y_A)$. Then we have that

$$\frac{mv^2}{2} + mgy = mgy_A,$$

from which

$$v = \sqrt{2g(y_A - y)}.$$

[6] Now we can return to the original equation A.1 to find

$$t_{A\to B} = \int_{x_A}^{x_B} \left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2} dx. \tag{A.2}$$

We now observe that the function $t_{A\to B}$ that we want to minimize is a function of x, y, y', where y and y' are functions (possible paths) from the interval $[x_A, x_B]$ to the interval y_A, y_B. So our variables are functionals and a new calculus needed to be created to deal with these problems. This was the origin of the calculus of variations.

A.2 The abstract problem in two dimensions

We then write a prototype of the problem we want solve in two dimensions. This is

$$J(y) = \int_{x_A}^{x_B} L[x, y(x), y'(x)]dx \tag{A.3}$$

where x_A, x_B are constants, and $y(x)$ is a twice continuous differentiable function, and L (for Lagrangian) is a twice continuous differentiable function of its arguments.

[6] Of course this equation could have been obtained by integrating the Newton second law $F_y = md^2y/dx^2 = mg$, or better just $g = d^2y/dx^2$.

The function J is known as the cost functional or objective functional that we intend to minimize. The minimization algorithm for the cost function studied here provides a differential equation known as the Euler-Lagrange Equation [7] which is the topic of the next section.

A.3 The Euler-Lagrange Equation

We assume that there is a solution y_0 for equation A.3. We can consider any differentiable function $\eta(x)$ such that $\eta(x_A) = \eta(x_B) = 0$, so that adding this function to the solution will not move the extremes. Then any linear combination

$$y(x, \epsilon) = y_0(x) + \epsilon\eta(x)$$

also satisfies the boundary conditions $y(x_A, \epsilon) = y_A$, $y(x_B, \epsilon) = y_B$.

Figure A.3 shows a minimum time path in green color and neighbor paths (blue) added from a given smooth function $\eta(x)$. The fact that ϵ is a vari-

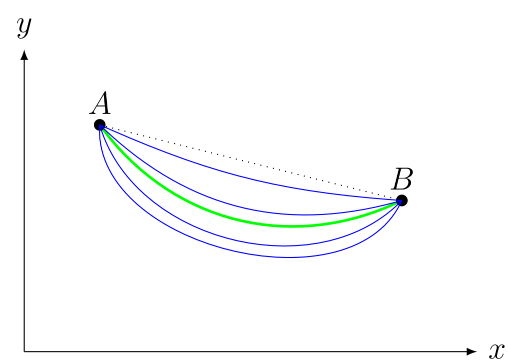


Figure A.2: The green arc between the points A and B represents the shortest path that solves the optimization integral A.3. The blue neighbour paths are variations around the shortest path.

able indicates that the cost function should be stationary at $\epsilon = 0$ which corresponds to the point where the least value of the cost function $J(y, \epsilon)$ is obtained. We can then assert that

[7] https://en.wikipedia.org/wiki/Euler%E2%80%93Lagrange_equation

$$\left.\frac{dJ}{d\epsilon}\right|_{\epsilon=0} = \frac{d}{d\epsilon}\int_{x_A}^{x_B} L[y(x,\epsilon), y'(x,\epsilon), x]dx$$
$$= \int_{x_A}^{x_B}\left[\frac{\partial L}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial L}{\partial y'}\frac{\partial y'}{\partial \epsilon}\right]dx = 0. \qquad \text{(A.4)}$$

Now,

$$\frac{\partial y(x,\epsilon)}{\partial \epsilon} = \eta(x),$$

and

$$\frac{\partial y'(x,\epsilon)}{\partial \epsilon} = \frac{\partial}{\partial \epsilon}\left[\frac{\partial y(x,\epsilon)}{\partial x}\right] = \frac{d}{dx}\left[\frac{\partial y(x,\epsilon)}{\partial \epsilon}\right] = \frac{d\eta(x)}{dx},$$

where we reversed the differential operators due to the second differential continuity of the function y, with respect to the parameter ϵ. We now replace these results into equation A.4 to find

$$\left.\frac{dJ}{d\epsilon}\right|_{\epsilon=0} = \int_{x_A}^{x_B}\left[\frac{\partial L}{\partial y}\eta(x) + \frac{\partial L}{\partial y'}\frac{d\eta(x)}{dx}\right]dx = 0. \qquad \text{(A.5)}$$

Let us perform integration by parts on the second term of the integration. That is

$$\int_{x_A}^{x_B}\left[\frac{\partial L}{\partial y'}\frac{d\eta(x)}{dx}\right]dx = \left.\frac{\partial L}{\partial y'}\eta(x)\right|_{x_A}^{x_B} - \int_{x_A}^{x_B}\eta(x)\frac{d}{dx}\frac{dL}{dy'} = -\int_{x_A}^{x_B}\eta(x)\frac{d}{dx}\frac{dL}{dy'}$$

since $\eta(x_A) = \eta_(x_B) = 0$. Then equation A.5 turns out to be

$$\int_{x_A}^{x_B}\left[\frac{\partial L}{\partial y} - \frac{d}{dx}\frac{\partial L}{\partial y'}\right]\eta(x)dx = 0.$$

Since $\eta(x)$ is arbitrary (up to the fact that it should be differentiable and vanishes at x_A and x_B) we can assert that

$$\frac{\partial L}{\partial y} - \frac{d}{dx}\frac{\partial L}{\partial y'} = 0. \qquad \text{(A.6)}$$

This is known as the Euler-Lagrange equation due to the fact that Euler first derived it using geometrical insights and Lagrange then derived it using calculus.

We found then that a necessary condition to solve the optimization integral problem A.3 is given by the differential equation A.6.

As a first application, before solving the brachistocrone problem, we show that the shortest path between two points in a plane [8] is given by a straight line.

Example A.3.1. *A straight line is the shortest path between two points on a plane.*

A.3.0.0.1 Solution: Given that an element of distance is given by $ds = \sqrt{dx^2 + dy^2}$ (assuming a plane in 2D), and so

$$\frac{ds}{dx} = \sqrt{1 + y'^2},$$

then we can write the integral

$$S_{A \to B} = \int_A^B \sqrt{1 + y'^2} dx$$

where we identify the L operator as

$$L(y, y', x) = \sqrt{1 + y'^2}.$$

Then we insert this L operator into the differential equation A.6 to find

$$-\frac{d}{dx} \frac{y'}{\sqrt{1 + y'^2}} = 0$$

which imples that

$$\frac{y'}{\sqrt{1 + y'^2}} = \text{constant} \equiv C.$$

from which we find

$$y'^2 = C^2(1 + y'^2)$$

or,

$$y' = \pm \frac{C}{\sqrt{1 - C^2}} = C_1.$$

[8] When we say plane we can assume without loss of generality that the plane is $\mathbb{R}^2$.

This first order differential equation has the easy solution

$$y = C_1 x + C_2.$$

We now should find the values of C_1 and C_2 from the boundary conditions. That is, we know that $y(x_A) = y_A$, and $y(x_B) = y_B$, that is

$$\begin{aligned} y_A &= C_1 x_A + C_2 \\ y_B &= C_1 x_B + C_2 \end{aligned}$$

from which, after subtraction

$$C_1 = \frac{y_A - y_B}{x_A - x_B}$$

which is the slope of the line through the points A and B, and then

$$C_2 = y_A - C_1 x_A = y_A - \frac{y_A - y_B}{x_A - x_B} x_A$$

Then

$$y = \frac{y_A - y_B}{x_A - x_B}(x - x_A) + y_A.$$

This is the line that joints the points (x_A, y_A) with (x_B, y_B).

However we showed that this is a necessary condition to have a stationary point. We did not show that it is sufficient that the path is straight to guarantee a unique shortest path. Euler-Lagrange equations are necessary conditions, no sufficient conditions. In general it is hard to prove the sufficient conditions. For example, if the cost function $J(y)$ is convex then we can guarantee a unique minimum but to prove the convexity requires some extra work. Similarly we could go to higher order derivatives of the cost function $J(y)$, and evaluate the second derivative as we do in calculus to find maxima/minima points, but again this requires extra work since the cost function $J(y)$ is not a simple function, but a functional of some functions with integrals and derivatives.

For the case of a path in a plane we could use simple arguments.

(i) There is no maximum. You give me a path, and I can always go extra far to find a larger path.

(ii) The set of lengths is bounded below. All paths are lengths greater or equal than 0, then there is a lower bound and hence an infimum.

(iii) Even if there is an infimum which is unique, do this imply that the path is unique? This is not necessarily true if we are working on a unit sphere. Think of two antipodes, the number of paths between them is infinite and all of them have the same length π. In the plane we can use Euclid's postulate that given two points there is only one line through them.

(iv) Let us call $\ell = d(A, B)$ where $d(A, B)$ is the distance along a straight segment that joints A and B. We prove that the ℓ is a lower bound for the set of lengths of all continuous paths between between A and B, and since the ℓ is in that set, then ℓ is the infimum and minimum at the same time. That is, ℓ is the shortest distance between A and B in the plane.

Assume that a continuously differentiable path

$$\gamma : [0, 1] \to \mathbb{R}^2$$

with $\gamma(0) = A$ and $\gamma(1) = B$. Without loss of generality we can consider A to be the origin and B a point sitting in the positive side of x axis. Even more, since the length of the curve does not change with a new parametrization we can stretch the $[0, 1]$ interval to $[0, B]$, and say that γ is now a function of $f(x)$ in the interval $[0, B]$, and its length is given by

$$S = \int_0^B \sqrt{1 + f'^2(x)}dx$$

Now, since $f'^2(x) \geq 0$, then $\sqrt{1 + f'^2(x)} \geq 1$, so

$$S = \int_0^B \sqrt{1 + f'^2(x)}dx \geq \int_0^B dx = B = \ell.$$

So the length of any other segment S is larger than ℓ, from which we found that ℓ is a lower bound, but since the length of the straight segment is ℓ, it is a minimum and the shortest distance between A and B.

Could we had use this argument from the beginning without recurring to the Euler-Lagrange formula? Yes, we could, but the good thing about the Euler-Lagrange formula is that it exhibits an explicit solution that we can then test as we did here.

A.4 The Brachistochrone

A.4.1 The Beltrami Identity

We can reduce the Euler-Lagrange second order differential equation A.6 to a first order equation under a specific case. This happens when the integrand L does not depend explicitly on x. The resulting equation was derived for first time by the Italian mathematician Eugenio Beltrami .

Since $L = L(y, y', x)$ we can use the chain rule to find that

$$\frac{dL}{dx} = \frac{\partial L}{\partial x} + \frac{\partial L}{\partial y}y' + \frac{\partial L}{\partial y'}\frac{dy'}{dx}.$$

Further

$$\frac{d}{dx}\left(y'\frac{dL}{dy'}\right) = \frac{dy'}{dx}\frac{\partial L}{\partial y'} + y'\frac{d}{dx}\left(\frac{\partial L}{\partial y'}\right).$$

From here we can write the second term of the Euler-Lagrange equation as

$$\begin{aligned}
\frac{d}{dx}\left(\frac{\partial L}{\partial y'}\right) &= \frac{1}{y'}\left[\frac{d}{dx}\left(y'\frac{\partial L}{\partial y'}\right) - \frac{dy'}{dx}\frac{\partial L}{\partial y'}\right] \\
&= \frac{1}{y'}\left[\frac{d}{dx}\left(y'\frac{\partial L}{\partial y'}\right) - \frac{dL}{dx} + \cancelto{0}{\frac{\partial L}{\partial x}} + \frac{\partial L}{\partial y}y'\right],
\end{aligned}$$

then from here bringing the first term of the Euler-Lagrange equation to the front, and multiplying both sides by $-y'$,

$$y'\left[\frac{\partial L}{\partial y} - \frac{d}{dx}\left(\frac{\partial L}{\partial y'}\right)\right] = \frac{d}{dx}\left(L - y'\frac{\partial f}{\partial y'}\right)$$

and from the Euler-Lagrange equations we find that the right hand side (assuming $y' \neq 0$) vanishes. That is

$$\frac{d}{dx}\left(L - y'\frac{\partial f}{\partial y'}\right) = 0.$$

or in other words

$$L - y'\frac{\partial L}{\partial y'} = \text{constant} \equiv C. \tag{A.7}$$

This is Beltrami identity. We use this identity directly to solve the Brachistochrone problem, which is a first order partial differential equation on L.

We now find the solution of the brachistochrone problem posted by Johan Bernoulli.

From equation A.2 we see that

$$L = \left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2},$$

Note that L does not depend explicitly on x, then from the Beltrami identity A.7 we have

$$\left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2} - y'\frac{\partial}{\partial y'}\left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2} = C.$$

Now

$$\frac{\partial}{\partial y'}\left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2} = \frac{y'}{\sqrt{2g(y_A - y)(1 + y'^2)}}$$

then

$$\begin{aligned}
\left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2} - \frac{y'^2}{\sqrt{2g(y_A - y)(1 + y'^2)}} &= \left[\frac{1 + y'^2}{2g(y_A - y)}\right]^{1/2}\left(1 - \frac{y'^2}{1 + y'^2}\right) \\
&= \left[\frac{1}{2g(y_A - y)(1 + y'^2)}\right]^{1/2} \\
&= C
\end{aligned}$$

We make a change of variables, $z = y_A - y$ (the vertical drop) and then $z' = -y'$, then we square the function and find

$$\frac{1}{2gz(1+z'^2)} = C^2,$$

or

$$z(1+z'^2) = \frac{1}{2gC^2}.$$

We find the value of C using some initial conditions. We can say that the position (x, y) is a function of time, that is $(x(t), y(t))$, and at time $t = 0$ we are at the initial position. This initial conditions are described by two parameters. The height $z = z_0$, and the angle $z' = \tan\theta_0$, that the tangent of the trajectory makes with the x axis at time $t = 0$. With these two values C could be computed. Let us call $2A = 1/(2gC^2)$, and write

$$z + zz'^2 = 2A,$$

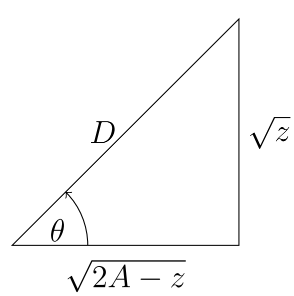
or

$$z' = \sqrt{\frac{2A - z}{z}}$$

with $2A \geq z \geq 0$. As it is, this equation does not seem simple to solve. We will consider a new parameter by defining $\tan\theta = dz/dx$ which shows the slope of the tangent line to the curve with an inclination angle θ. We have then

$$\tan\theta = \sqrt{\frac{2A - z}{z}}.$$

We can think of a right triangle with sides $\sqrt{2A - z}$, $\sqrt{z}$ as shown in the figure below, and from Pythagoras theorem,



$$D = \sqrt{2A - z + z} = \sqrt{2A}.$$

This is interesting, since we eliminated the variable z, and then we can write

$$\sin\theta = \sqrt{\frac{z}{2A}} \qquad \cos\theta = \sqrt{\frac{2A - z}{2A}},$$

and

$$\cos 2\theta = \cos^2\theta - \sin^2\theta = \frac{2A - 2z}{2A} = \frac{A - z}{A}$$

from which

$$A - z = A\cos 2\theta$$

and so

$$z = A(1 - \cos 2\theta).$$

So, a simple geometric trick solved the differential equation for us without having in terms of the parameter θ (the angle of the tangent of the curve). However we want to know the answer as a function of x, or at least x as a function of θ . For that we need to perform a few more computations. We first recognize that $dx/dz = \cot\theta$, and

$$\frac{dz}{d\theta} = 2A\sin 2\theta = 4A\sin\theta\cos\theta,$$

and from the chain rule

$$\frac{dx}{d\theta} = \frac{dx}{dz}\frac{dz}{d\theta} = 4\cot\theta A\sin\theta\cos\theta = 4A\cos^2\theta.$$

and since $\cos^2\theta = \cos 2\theta + \sin^2\theta = 1 + \cos 2\theta - \cos^2\theta$, then

$$\frac{dx}{d\theta} = 2A(1 + \cos 2\theta)$$

We can integrate this equation to find $x = x(\theta)$. That is

$$x(\theta) = 2A(\theta + \sin 2\theta/2) + c$$

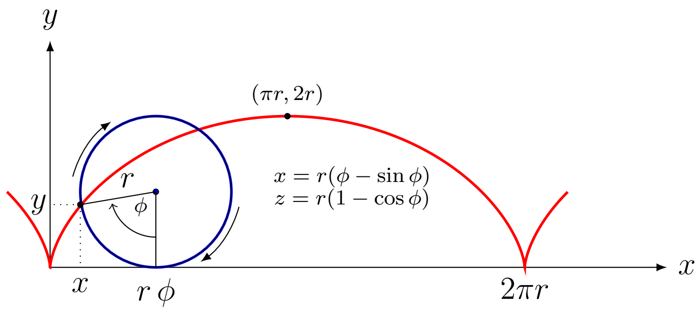


Figure A.3: The cycloid is generated by a fixed point attached to the edge of a rolling disk. The cycloid in red is an upside down brachistochrone.

where c is a constant. We then found the couple of parametric equations

$$\begin{aligned} x(\theta) &= 2A\theta + A\sin 2\theta + c \\ z(\theta) &= A - A\cos 2\theta \end{aligned}$$

We observe that if the term $2A\theta$ is removed the equation corresponds to that of a circle with center at $(c, 0)$, and radius A. First let us assume initial conditions to find c. If at $\theta = 0$ we assume the starting point in the origin $(0, 0)$, we have that $c = 0$. Note that c is just a horizontal translation of the path. We then can assume that $A = r = 1/(4gC^2)$ for a better notion and rewrite with these initial condition

$$\begin{aligned} x(\phi) &= r\phi + r\sin\phi \\ z(\phi) &= r - r\cos\phi \end{aligned}$$

with $\phi = 2\theta$.

Figure A.3 shows the graphic of this equation which is known as the cycloid. [9]

In this equation we take reverse the sign of ϕ and so instead of $\sin\phi$ we have $-\sin\phi$.

The interpretation of the equation according to the graphic is as follows. A wheel is moving to the right at a constant angular speed $d\theta/dt$, where At

[9] https://en.wikipedia.org/wiki/Cycloid

a given time the wheel's center is located at a distance $r\phi$, and at a height $y = r$. Now think of a pebble that is trapped in a wheel of a car. That pebble is represented by the black dot in the figure. The location of the pebble relative to the center of the wheel is shifted to the left by $-r\cos\theta$ and to the up by $-r\sin\theta$. That explains that the path followed by the pebble obeys the equations on the figure.

Observe that, since we reversed the z coordinate, the cycloid is the upside down and the trajectory of the brachistochrone is convex up. Christiaan Huygens showed in 1659 that the cycloid is a tautochrone, that is, the period of oscillation, or time for an object to descent without friction through the brachistochrone is constant regardless of the initial high used. This property is interesting on the construction of pendulum clocks. The circular pendulus clocks preserve the period (harmonic motion) for very small angles of oscillation. The cycloid does not present this problem and inspired Huygens to build a clock based on a cyloid pendulum. Huygens failed his attempts due to manufacturing difficulties.

Again, we will not prove that this solution is an absolute minimum. It is a stationary solution for the calculus of variations, it is not a maximum since we can make paths as long as we want and this make times as large as we want, so there is not a maximum path. We know that there is an absolute minimum because any path has a positive length (on time and distance), so 0 is a lower bound and we assume that the set of distances (and times) fill a continuum interval in the real numbers. We do not prove this. Is this a local minimum?

The history of the cycloid is quite interesting and the Wikepidia site listed above makes a brief introduction with a good number of references for further reading.

The calculus of variations and the Euler-Lagrange equations can be extended to consider several independent variables as well as variable end points but we will not consider such extensions here. However we include one more section which is of great importance. We showed already two problems where minimization occurs. One for the length of a path and the other for the travel time along a path. We know generalize the idea for more general surfaces, other than the plane.

A.5 The shortest path on parametrized surfaces

Before we do the mathematical analysis for this section let us emphasize again about the importance of understanding what we are finding.

We can not guarantee that the path found is the shortest path and neither that this is unique. If on a sphere, the shortest path is along a great circle (and we show this below), then if the two (start and end) points A and B are antipodes we would have an infinte number of paths all with length πr through them that satisfy the shortest criterion. If the points A and B are not antipodes, then the minimum is not unique. There are two segments along the sphere that join A and B. A short and long segments (they add to 2π, so if one is length a the other is length $2\pi - a$. So the minimum is no unique. Here is another simple example of a geodesic which is not unique and that the measure of the path is different. That is, it could be a local minimum but not the absolute minimum.

Assume that you start with a rectangle with a diagonal drawn between to opposite vertices as shown in the right frame of Figure A.5 sketches the rectangle as well as the diagonal in red. If we join the two vertical ends of the rectangle we can form a cylinder. The diagonal path becomes a helix. Since the diagonal is straight (and the straight segment is the shortest between two points in the plane) it is easy to think that the helix is a line of shortest distance along the surface of the cylinder, but if you observe the green line has shorter length.

It can be shown (see Boonman and Chitsakul [10]) that the geodesics of the cylinder are helices but we should check that the points A and B are not sitting along the same side of the rectangle or on the corners.

After providing the disclaimer on the efficacy of the method, let us develop the mathematical machinery of this. We consider surfaces $x = x(u, v), y = y(u, v)$, and $z = z(u, v)$ which have first order derivatives with respect to the parameters (u, v).

A path in the surface can be parametrized throught the parameter u, v as functions of a real variable t. That is, we can say $u = u(t)$, $v = v(t)$, so that we are constraining the possible values of u and v through t.

The distance differential element along the surface is given by $ds =$

[10]http://research.utar.edu.my/CMS/ICMSA2010/ICMSA2010_Proceedings/files/applied_maths/AM-Boonnam.pdf

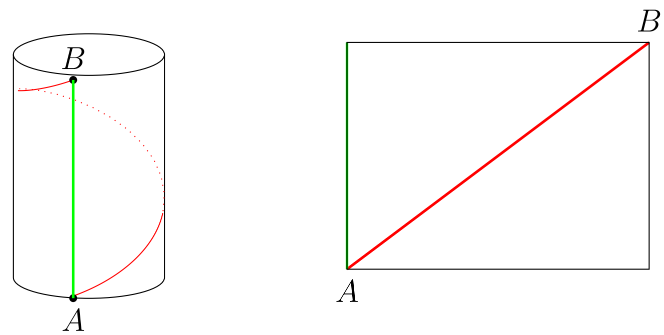


Figure A.4: A geodesic on a cylinder is a helix. The shortest path along a complete cycle is not along the helix but vertically along the edge as shown in the green color.

$\sqrt{(dx)^2 + (dy)^2 + (dz)^2}$. Without loss of generality we can consider $t \in [a, b]$. The total distance can be found by integration using the formula

$$D = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt \tag{A.8}$$

It is interesting to observe that if t is time, then the vector $\boldsymbol{v} = d\boldsymbol{r}/dt = (dx/dt, dy/dy, dz/dt)$ is the velocity vector, and the integrand above is the instantaneous speed, which indicates that the length of the segment is the integral of its local speed times time, which makes sense from the physical point of view.

From the chain rule we find

$$\frac{dx}{dt} = \frac{\partial x}{\partial u}\frac{du}{dt} + \frac{\partial x}{\partial v}\frac{dv}{dt}$$

and similarly

$$\frac{dy}{dt} = \frac{\partial y}{\partial u}\frac{du}{dt} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial t} \quad , \quad \frac{dz}{dt} = \frac{\partial z}{\partial u}\frac{du}{dt} + \frac{\partial z}{\partial v}\frac{dv}{dt}$$

We can write these identities in matrix form as follows:

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{pmatrix} \begin{pmatrix} \frac{du}{dt} \\ \frac{dv}{dt} \end{pmatrix}$$

We call $\boldsymbol{r}$ the tangent vector along the curve, $\boldsymbol{r}_u$ the tangent along the u parameter (generalized coordinate) and $\boldsymbol{r}_v$ the tangent along the generalized coordinate v. Then rewrite the matrix equation in a more compact form:

$$\boldsymbol{r} = \begin{pmatrix} \boldsymbol{r}_u & \boldsymbol{r}_v \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix}$$

where $\dot{u} = du/dt$, and $\dot{v} = dv/dt$.

The matrix $R = (\boldsymbol{r}_u, \boldsymbol{r}_v)$ plays a very important role in differential geometry. We will see how.

In the integrand we observe that we need to find the L_2 norm of the vector at the left of the equation above. That is, we need to find a dot product $\langle (\dot{x}, \dot{y}, \dot{z}), (\dot{x}, \dot{y}, \dot{z}) \rangle$. This product is, from the matrix

$$\begin{aligned} \left(\frac{ds}{dt}\right)^2 &= \begin{pmatrix} \dot{u} & \dot{v} \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_u \\ \boldsymbol{r}_v \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_u & \boldsymbol{r}_v \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} \\ &= \begin{pmatrix} \dot{u} & \dot{v} \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_u \cdot \boldsymbol{r}_u & \boldsymbol{r}_u \cdot \boldsymbol{r}_v \\ \boldsymbol{r}_v \cdot \boldsymbol{r}_u & \boldsymbol{r}_v \cdot \boldsymbol{r}_v \end{pmatrix} \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} \end{aligned}$$

The matrix of dot products of tangent vectors along the coordinates is known as the Metric Tensor [11] of the differential geometry . It is common to find the notation $g_{ij} = \boldsymbol{r}_i \boldsymbol{r}_j$, where i and j can take over the coordinates u or v. For example $u = 1$ and $v = 2$.

The reason of the name "tensor" for the matrix g_{ij} is because it is what is needed to make any parametrization preserve physical invariant measures, such as for example arc length, angles, areas, etc. These facts are well

[11] https://en.wikipedia.org/wiki/Metric_tensor

explained in the Wikipedia page above and were first observed by Ricci-Cubastro and Levi-Civita in 1900. The quadratic form $g_{ij}x_ix_j$ is called the First Fundamental Form [12] of the differential geometry. We can look at it as a weight inner product where the weight is given by the g_{ij} matrix (assuming is positive definite). In particular when instead of x and y we have the element dx_i, $(i = 1, \cdots, 3)$ we have that (using Einstein's repeated index notation) $ds^2 = g_{ij}dx_idx_j$. This is also commonly written as

$$(ds)^2 = E(du)^2 + 2Fdudv + G(dv)^2, \tag{A.9}$$

where $E = g_{11}$, $F = g_{12}$ and $G = g_{22}$. The square root of the non-diagonal terms are called scale factors $h_u = \sqrt{g_{11}}, h_v = \sqrt{g_{22}}$, which measures the stretching of the surface along the coordinates u and v. The cross term g_{12}, measures the skewness. If the coordinates are orthogonal there are not cross terms and $g_{12} = g_{21} = 0$.

We can set $t = u$, and from equation A.9 we find

$$\frac{ds}{du} = \sqrt{E + 2Fv' + Gv'^2}$$

In the integration A.8 we call (renaming y for v, and x for u)

$$L[u, v, v'] = \sqrt{E + 2Fv' + Gdv'^2}.$$

We want to use Euler-Lagrange A.6 equation to find the stationary points of this operator. First we note that

$$\begin{aligned} \frac{dL}{dv} &= \frac{1}{2\sqrt{E + 2Fv' + Gv'^2}} \left(\frac{\partial E}{\partial v} + 2\frac{\partial F}{\partial v}v' + \frac{\partial G}{\partial v}v'^2 \right) \\ \frac{dL}{dv'} &= \frac{1}{2\sqrt{E + 2Fv' + Gv'^2}}(2F + 2Gv'). \end{aligned}$$

Hence the Euler-Lagrange equation becomes

$$\frac{\left(\frac{\partial E}{\partial v} + 2\frac{\partial F}{\partial v}v' + \frac{\partial G}{\partial v}v'^2\right)}{2\sqrt{E + 2Fv' + Gv'^2}} - \frac{d}{du}\frac{2F + 2Gv'}{2\sqrt{E + 2Fv' + Gv'^2}} = 0. \tag{A.10}$$

which provides stationary points for the length between two fixed points under the metric defined by g_{ij} in the given surface.

[12] https://en.wikipedia.org/wiki/First_fundamental_form

Appendix B

Homogeneity, additivity and linearity

We can find an operator (function) which could be homogenous but not linear. A **median filter** is an example. Take a function such that for finite set A of real numbers with an odd cardinal it chooses the number m such that half of the numbers $x \in A$, satisfy $m \leq x$, and for the other half $m \geq x$. A different way to say this is that if we sort the elements of A in ascending (or descending) order into a string, the element m is the center of the string. We see that $f(cA) = cf(A)$. However f is not linear. That is, $f(A+B) \neq f(A) + f(B)$. More especifically, let us define

$$\begin{aligned} f : \mathbb{R}^3 &\to \mathbb{R} \\ (x_1, x_2, x_3) &\mapsto y \end{aligned}$$

where if x_1, x_2, x_3 are put in ascending order y will be in the middle Then $f[c(x_1, x_2, x_3]) = cy = cf(x_1, x_2, x_3)$. Now let us assume two points $\boldsymbol{x} = (x_1, x_2, x_3) = (-1, 0, 100)$ and $\boldsymbol{z} = (z_1, z_2, z_3) = (1, 1, -100)$. Then $f(\boldsymbol{x}) = 0$, $f(\boldsymbol{z}) = 1$, and $f(\boldsymbol{x}+\boldsymbol{z}) = f(0, 1, 0) = 0$. So $f(\boldsymbol{x}+\boldsymbol{z}) = 0 \neq 1 = f(\boldsymbol{x}) + f(\boldsymbol{z})$.

We then showed that in general a function could be homogeneous but not necessarily linear. In functions of one variable in the real space; that is, functions

$$f : A \subset \mathbb{R} \to \mathbb{R}$$

however, homogeneity implies linearity. Here is the simple proof. Pick any x and y in $\mathbb{R}$ (both non-zero, if one is zero there is no much to do). Then find

c such that $y = cx$ ($c = y/x$). So

$$\begin{aligned} f(x+y) = f(x+cx) = f[(1+c)x] = (1+c)f(x) &= f(x) + c(f(x) \\ = f(x) + f(cx) &= f(x) + f(y). \end{aligned}$$

We want to know if the opposite is true. That is, if we have additivity, for $f : A \subset \mathbb{R} \to \mathbb{R}$, do we have homogeneity? and so linearity? can we find a function in $f : A \subset \mathbb{R} \to \mathbb{R}$ which has additivity but no homogeneity?

If such a function f is additive, is it homogeneous? We start with the simplest case of a natural scalar. That is, let $\alpha = n$ a natural number so

$$f(\alpha x) = f(nx) = f[(n-1)x+x] = f[(n-1)x]+f(x) = f[(n-2)x+x)]+f(x)$$

and following an inductive alogorithm we see that

$$f(\alpha x) = \alpha f(x).$$

with $\alpha = n$ a natural number. If $\alpha = 0$, then we have $f(\alpha x) = f(0) = f(0+0) = f(0) + f(0) = 2f(0)$. The only way $f(0) = 2f(0)$ is if $f(0) = 0$. so the property counts for $\alpha = 0$. Now, let us assume that $\alpha = -1$. Then using $f(0) = 0$ we find

$$f(0) = f(x - x) = f(x) + f(-x) = 0.$$

so $f(-x) = -f(x)$. We see that homogeneity is achived for $\alpha = -1$. The next natural step is to extend the property to all integers $\mathbb{Z}$.

Now, let us assume that $\alpha = -n$ with n a natural number then

$$f(\alpha x) = f(-nx) = f[n(-x)] = nf(-x) = -nf(x).$$

Hence homogeneity is verified for any α in the integers $\mathbb{Z}$. How about $\alpha = 1/q$ with q an integer.

$$q\alpha = 1,$$

and then

$$f(x) = f(q\alpha x) = qf(\alpha x),$$

which implies

$$f(\alpha x) = \frac{1}{q}f(x) = \alpha f(x).$$

We verified homogeneity for $\alpha = 1/q$ with q integer. Now if $\alpha = p/q$ with p and q integers, $q \neq 0$, that is when p is rational then

$$f(\alpha x) = f\left(\frac{p}{q}x\right) = f\left(p\frac{x}{q}\right) = pf\left(\frac{x}{q}\right) = \frac{p}{q}f(x) = \alpha f(x),$$

so it works for $\alpha \in \mathbb{Q}$.

What we claim is that it could happen that $f(\alpha x) \neq \alpha f(x)$ for α irrational, since the way to obtain an irrational number through rationals is with an infinite series (or sequence) but this breaks the finite additivity required by linear operators. We are looking for the counter-example along these lines. If $f(x)$ is continuous, then assume α is irrational and $\alpha = \lim_{n\to\infty} \alpha_n$ with $\alpha_n \in \mathbb{Q}$, then due to the continuity of f we can take the limit outside of the function argument in the following lines:

$$f(\alpha x) = f(\lim_{n\to\infty} \alpha_n x) = \lim \alpha_n f(x) = \alpha f(x)$$

Now,

$$f(x) = f(x1) = xf(1).$$

So if $f(1) = a$, we have an explicit definition

$$f(x) = ax$$

for all **continuous** functions that satisfy the additifity condtions $f(x+y) = f(x) + f(y)$ on the one dimensional real space.

This [1] link sketches a proof using the fundamental theorem of calculus. The counter example (if it exist on one dimensional real functions) has to be a discontinuous function at any irrational number.

We will show an example that is additive but non-linear. That is, there is at least an element of the domain x and a scalar α such that $f(\alpha x) \neq \alpha f(x)$.

The idea comes from the axiom of choice and Hamel's basis. In fact this link indicates that [2] " The first to realize that it is possible using choice to construct a non-linear additive function was Hamel in 1905 ("Eine Basis aller Zahlen und die unstetigen Losungen der Functionalgleichung: $f(x+y) =$

[1]Execercise 2 of https://www.math.ualberta.ca/ xinweiyu/217.1.13f/217-20130913.pdf

[2]http://mathoverflow.net/questions/57426/are-there-any-non-linear-solutions-of-cauchys-equation-fxy-fxfy-wit

$f(x) + f(y)$"), Math Ann 60 459-462); indeed, a Hamel basis of $\mathbb{R}$ over $\mathbb{Q}$ allows us to provide examples."

The idea is to think on the reals in the field of the rational numbers, which we call $\mathbb{R}_{\mathbb{Q}}$. That is, the axiom of choice guarantees the existence of a Hamel basis $B = \{b_i\}_{i \in I}$ and I is a subset of the integer numbers, such that every real x can be uniquely written in the form $x = \sum_{j \in J} \lambda_j b_j$, for some $j \in J \subset I$, with λ_i rational. Here J is a finite set.

(i) The set cardinality of set B is infinity (no shown here). Pick f such that $f(b_k) = 0$ for one $b_k \in B$, $b_k \neq 1$, $b_k \neq 0$, and $f(b_j)$ is anything you want for $j \neq k$.

(ii) Force additivity by defining

$$f(x) = \sum_{i \in K} \lambda_i f(b_i),$$

where $x = \sum \lambda_i b_i$ (all sums are finite sums).

Check additivity, if $x = \sum_{i \in J} \lambda_i \alpha_i$, $y = \sum_{i \in K} \gamma_i \alpha_i$, then pick $L = J \cup K$. and

$$x + y = \sum_{i \in L} (\lambda_i + \gamma_i) b_i,$$

where some λ_i or γ_i could be zero (those 0 do not belong to the intersection of $J \cap K$).

$$\begin{aligned} f(x+y) &= \sum_{i \in L} (\lambda_i + \gamma_i) f(b_i) \\ &= \sum_{i \in L} \lambda_i f(b_i) + \sum_{i \in L} \gamma_i f(b_i) \\ &= \sum_{i \in J} \lambda_i f(b_i) + \sum_{i \in K} \gamma_i f(b_i) \\ &= f(x) + f(y), \end{aligned}$$

where in the previous to the last step we did not include those element with 0 coefficients (non in the intersection).

(iii) We show that f is not homogeneous on the element b_k. We know $f(b_k) = 0$. Now, if f is homogeneous in b_k then $f(b_k) = f(b_k 1) = b_k f(1) \neq 0$. This contradicts that $f(b_k) = 0$. So f can not be homogeneous.

The following link [3] provides a proof that the graph of f (that is, the elements $(x, f(x) \in \mathbb{R}^2)$ is dense in the plane $\mathbb{R}^2$. Then no matter where in the plane there is always an element of $(x, f(x))$ as near as you want to you. This is an amazing fact.

[3] https://www.math.ualberta.ca/ xinweiyu/217.1.13f/217-20130913.pdf

Index