Mathematical Geodesy
Maa-6.3230

Martin Vermeer

16th February 2015

Figure: Cluster of galaxies Abell 2218, distance 2 billion light years, acts as a gravitational lense, The geometry of space-time within the cluster is non-Euclidean.
Image credit: NASA/ESA

Course Description

Workload 7 cr

Teaching Period III

Learning Outcomes After completing the course, the student

- Is able to do simple computations on the sphere and understands the geometry of the reference ellipsoid
- Is able to solve, using tools like Matlab™, the geodetic forward and inverse problems etc. on the reference ellipsoid
- Masters the basics of global and local reference systems and is able to execute transformations
- Masters the basics of Gaussian and Riemannian surface theories and can derive the metric tensor, Christoffel symbols and curvature tensor for simple surfaces
- Masters the basic math of map projections, esp. conformal ones, and the behaviour of map scale, and is able to compute the isometric latitude.

Content Spherical trigonometry, geodetic co-ordinate computations in ellipsoidal and rectangular spatial co-ordinate systems, astronomical co-ordinates, co-ordinate system transformations, satellite orbits and computations, Gaussian and Riemannian surface theories, map projection computations.

Foreknowledge

Equivalences Replaces course Maa-6.230.

Target Group

Completion Completion in full consists of the exam, the exercise works and the calculation exercises.

Workload by Component

- Lectures 16×2 h = 32 h
- Independent study 55 h
- Exercise works 2×12 h = 24 h (independent work)
- Calculation exercises at home 16, of which must be done 12×6 h = 72 h (independent work)
- Exam 3 h
- Total 186 h

Grading The grade of the exam becomes the grade of the course, 1-5

Study Materials Lecture notes. Background material Hirvonen: Matemaattinen geodesia; Torge: Geodesy

Teaching Language Suomi

Course Staff and Contact Info Martin Vermeer, room Technopolis Innopoli 3D (Vaisalantie 8) 4.143, martin dot vermeer at aalto dot fi

Reception times: email `martin dot vermeer at aalto dot fi`, or phone 050 3574139

CEFR-taso

Lisätietoja

Contents

1. Spherical trigonometry **1**
1.1. Introduction . . . 1
1.2. Spherical excess . . . 1
1.3. The surface area of a triangle on a sphere . . . 2
1.4. A rectangular spherical triangle . . . 3
1.5. A general spherical triangle . . . 4
1.6. Deriving the formulas with the aid of vectors in space . . . 5
1.7. Polarization . . . 6
1.8. Solving the spherical triangle by the method of additaments . . . 7
1.9. Solving the spherical triangle by LEGENDRE's method . . . 8
1.10. The forward geodetic problem on the sphere . . . 8
1.11. The geodetic inverse problem on the sphere . . . 9
1.12. The half-angle cosine rule . . . 9

2. The geometry of the reference ellipsoid **11**
2.1. Introduction . . . 11
2.2. The geodesic as solution to a system of differential equations . . . 11
2.3. An invariant . . . 12
2.4. The geodetic main problem . . . 13
2.5. The geodetic inverse problem . . . 13

3. Co-ordinates on the reference ellipsoid **15**
3.1. Representations of the sphere and the ellipsoid . . . 15
3.2. Various latitude types . . . 15
3.3. Measures for the flattening . . . 16
3.4. Relationships between different types of latitude . . . 16
3.5. Co-ordinates in the meridional ellipse . . . 17
3.6. Three-dimensional rectangular co-ordinates on the reference ellipsoid . . . 17
3.7. Computing geographic co-ordinates from rectangular ones . . . 18
3.8. Meridian arc length . . . 19

4. Reference systems **21**
4.1. The GRS80 system and geometric parameters . . . 21
4.2. Gravimetric parameters . . . 22
4.3. Reference frames . . . 23
4.4. The orientation of the Earth . . . 24
4.5. Transformations between systems . . . 25

5. Using rotation matrices **29**
5.1. General . . . 29
5.2. Chaining matrices in three dimensions . . . 30
5.3. Orthogonal matrices . . . 30
5.4. Topocentric systems . . . 32
5.5. From geocentric to topocentric and back . . . 33
5.6. The geodetic main and inverse problems with rotation matrices . . . 34
5.6.1. Geodetic main problem . . . 34

5.6.2. Geodetic inverse problem . . . 35
5.6.3. Comparison with ellipsoidal surface solution . . . 35

6. Co-ordinate systems and transformations **37**
6.1. Geocentric terrestrial systems . . . 37
6.2. Conventional Terrestrial System . . . 37
6.3. Polar motion . . . 37
6.4. The Instantaneous Terrestial System . . . 38
6.5. The quasi-inertial geocentric system . . . 38

7. Reference systems and realizations **41**
7.1. Old and new reference systems; ED50 vs. WGS84/GRS80 . . . 41
7.2. WGS84 and ITRS . . . 41
7.3. Co-ordinate system realizations . . . 42
7.4. Realization of WGS84 . . . 42
7.5. Realizations of ITRS/ETRS systems . . . 42
7.6. The three-dimensional Helmert transformation . . . 43
7.7. Transformations between ITRF realizations . . . 44

8. Celestial co-ordinate systems **47**
8.1. Sidereal time . . . 47
8.2. Trigonometry on the celestial sphere . . . 49
8.3. Using rotation matrices . . . 50
8.4. On satellite orbits . . . 52
8.5. Crossing a given latitude in the inertial system . . . 52
8.6. The satellite's topocentric co-ordinates . . . 53
8.7. Crossing a given latitude in the terrestrial system . . . 54
8.8. Determining the orbit from observations . . . 54

9. The surface theory of Gauss **59**
9.1. A curve in space . . . 59
9.2. The first fundamental form (metric) . . . 60
9.3. The second fundamental form . . . 61
9.4. Principal curvatures . . . 63
9.5. A curve embedded in a surface . . . 66
9.6. The geodesic . . . 69

10. The surface theory of Riemann **73**
10.1. What is a tensor? . . . 73
10.2. The metric tensor . . . 77
10.3. The inverse metric tensor . . . 79
10.3.1. Raising or lowering sub- or superscripts of a tensor . . . 79
10.3.2. The eigenvalues and -vectors of a tensor . . . 80
10.3.3. The graphic representation of a tensor . . . 80
10.4. The Christoffel symbols . . . 80
10.5. The geodesic revisited . . . 82
10.6. The curvature tensor . . . 83
10.7. Gauss curvature and spherical excess . . . 86
10.8. The curvature in quasi-Euclidean geometry . . . 88

11. Map projections **91**
11.1. Map projections and scale . . . 91
11.1.1. On the Earth surface . . . 91
11.1.2. In the map plane . . . 92
11.1.3. The scale . . . 92

11.1.4. The TISSOT-indicatrix . . . 94
11.2. The LAMBERT projection (LCC) . . . 95
11.3. On the isometric latitude . . . 97
11.4. The MERCATOR projection . . . 98
11.5. The stereographic projection . . . 99
11.6. The GAUSS-KRÜGER projection . . . 101
11.7. Curvature of the Earth surface and scale . . . 104

12. Map projections in Finland **107**
12.1. Traditional map projections . . . 107
12.2. Modern map projections . . . 107
12.3. The triangulated affine transformation used in Finland . . . 108
12.3.1. Plane co-ordinates . . . 108

A. Isometric latitude on the ellipsoid **111**

B. Useful equations connecting the main radii of curvature **113**

C. The Christoffel symbols from the metric **115**

D. The Riemann tensor from the Christoffel symbols **117**

Chapter 1

Spherical trigonometry

1.1. Introduction

The formulas of spherical geometry are very useful in geodesy. The surface of the Earth, which to first approximation is a plane, is in second approximation (i.e., in a small, but not *so* small, area) a spherical surface. Even in case of the whole Earth, the deviation from spherical shape is only 0.3%.

The starry sky again may be treated as a precise spherical surface, the radius of which is undefined; in practical computations we often set $R = 1$.

1.2. Spherical excess

See figure 1.1. Let us assume that the radius of the sphere is 1. The "front half" of the sphere is a semi-sphere, the surface area of which is 2π. A triangle is formed between three great circles. The same great circle form, on the back surface of the sphere, a "antipode triangle" of the same size and shape.

When the surface area of the whole semi-sphere is 2π, the area of the "orange slice" bounded by two great circles will be $\frac{\alpha}{\pi} \cdot 2\pi$, where α is the angle between the great circles. We obtain

$$\begin{aligned} A_1 + A_2 &= 2\alpha \\ A_1 + A_3 &= 2\beta \\ A_1 + A_4 &= 2\gamma \end{aligned}$$

and

$$A_1 + A_2 + A_3 + A_4 = 2\pi.$$

By summing up the first three equations we obtain

$$2A_1 + A_1 + A_2 + A_3 + A_4 = 2\left(\alpha + \beta + \gamma\right)$$

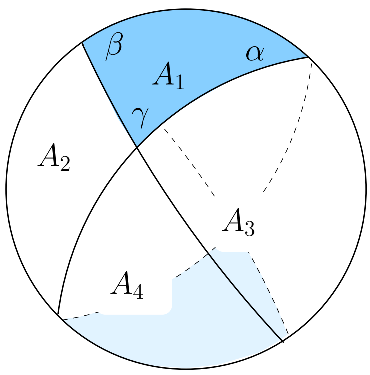


Figure 1.1.: Spherical triangles on a semi-sphere. The back side surface of the sphere has been depicted in a lighter shade with its "antipode triangles"

i.e.,

$$A_1 = \alpha + \beta + \gamma - \pi = \varepsilon,$$

where ε is called the *spherical excess.*

If the radius of the sphere is not 1 but R, we obtain

$$A_1 = \varepsilon R^2 \;\Rightarrow\; \varepsilon = \frac{A_1}{R^2}.$$

Here ε is expressed in *radians.* If ε is not in radians, we may write

$$\varepsilon\,[\text{unit}] = \frac{\rho_{\text{unit}} A_1}{R^2},$$

where ρ_{unit} is the conversion factor of the unit considered, e.g., for degrees, 57.29577951308232087721 or for gons, 63.66197723675813430801.

As we see is the spherical excess inversely proportional to R^2, i.e., *directly proportional to the total curvature* R^{-2}. It is also directly proportional to the surface area of the triangle.

This is a special case of a more general rule:

> The directional closing error of a vector which is transported in a parallel way around the closed edge of a surface is the same as the integral over the surface of the total curvature.

As a formula:

$$\varepsilon = \int_A K d\sigma,$$

where σ is the variable for surface integration, and K is the total curvature of the surface according to K.F. Gauss, which thus can vary from place to place. E.g., on the surface of an ellipsoid

$$K = \frac{1}{MN},$$

where M is the meridional curvature (in the North-South direction) and N the so-called transverse curvature in the East-West direction. Both depend on the latitude φ. In a smallish area, the internal geometry of the ellipsoidal surface does not differ noticeably from a spherical surface, the radius of which is $R = \sqrt{MN}$.

If the triangle of the surface of the sphere is *small* compared to the radius of the Earth, also the spherical excess will be small. In the limit we have $\varepsilon \to 0$ ja $\alpha + \beta + \gamma = \pi$ exactly. We say that the a plane surface (or a very small part of a spherical surface) forms a *Euclidean space*, whereas a spherical surface is *non-euclidean.*

1.3. The surface area of a triangle on a sphere

If the triangle isn't very large – i.e., just as large as geodetic triangulation triangles generally on, at most some 50 km –, we may calculate its surface area using the formula for the plane triangle:

$$A = \frac{1}{2} a \cdot h_a = \frac{1}{2} ab \sin\gamma,$$

where h_a is the height of the triangle relative to the a side, i.e., the straight distance of corner point A from side a.

Because according to the sine rule $b = a\dfrac{\sin\beta}{\sin\alpha}$, it also follows that

$$A = \frac{a^2 \sin\beta \sin\gamma}{2\sin\alpha}.$$

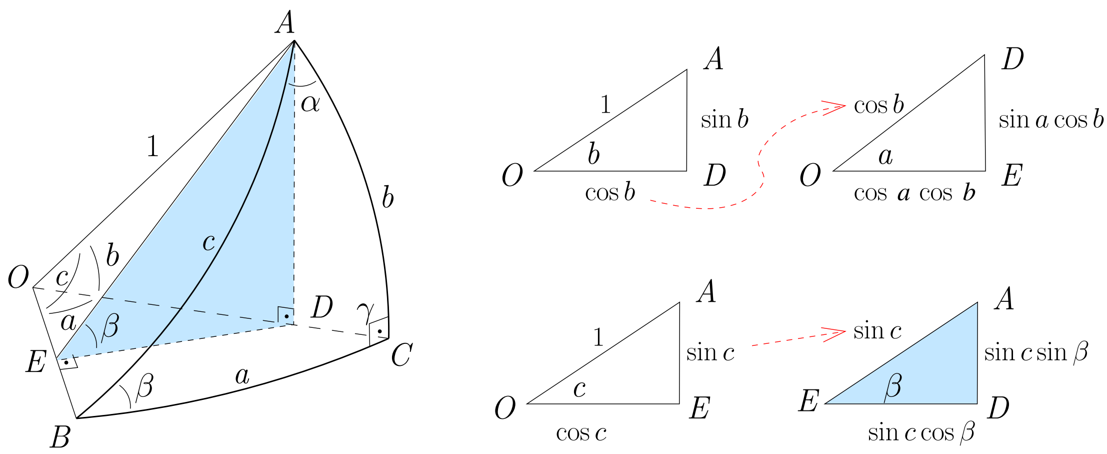


Figure 1.2.: A rectangular spherical triangle. The partial triangles are lifted out for visibility

At least for computing the spherical excess, these approximate formulas are good enough:

$$\varepsilon = \frac{A}{R^2},$$

where also the approximate value for R, e.g., $R \approx a = 6378137\,\mathrm{m}$, is completely sufficient.

1.4. A rectangular spherical triangle

This case is depicted in figure 1.2. Many simple formulas follow directly from the figure and the separately drawn plane triangles:

$$\begin{aligned} EO &= \cos c = \cos a \cos b \\ DE &= \sin c \cos \beta = \sin a \cos b \\ AD &= \sin c \sin \beta = \sin b \end{aligned} \tag{1.1}$$

By interchanging the roles of a and b (and thus of α and β) we obtain furthermore

$$\begin{aligned} \sin c \cos \alpha &= \sin b \cos a \\ \sin c \sin \alpha &= \sin a \end{aligned} \tag{1.2}$$

of which the first yields

$$\cos \alpha = \cos a \frac{\sin b}{\sin c} = \cos a \sin \beta,$$

according to the last equation in group (1.1).

By dividing the first of group (1.2) by its second, we obtain

$$\cot \alpha = \cot a \sin b$$

and the second of group (1.1) by its third, correspondingly

$$\cot \beta = \cot b \sin a.$$

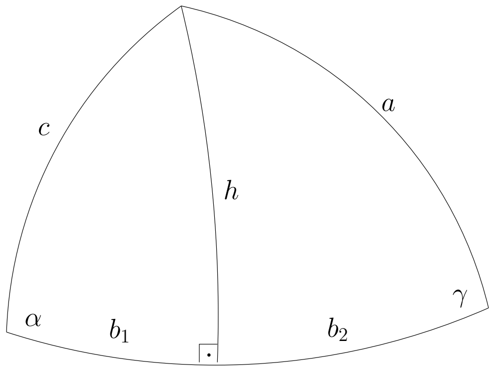


Figure 1.3.: General spherical triangle

1.5. A general spherical triangle

The formulae for a spherical triangle are obtained by dividing the triangle into two right triangles, see figure 1.3. here, the third side is $b = b_1 + b_2$.

If we apply to the sub-triangles the formulas derived above, we obtain:

$$\begin{aligned}
\cos a &= \cos h \cos b_2, \\
\sin a \cos \gamma &= \cos h \sin b_2, \\
\sin a \sin \gamma &= \sin h,
\end{aligned}$$

$$\begin{aligned}
\cos c &= \cos h \cos b_1, \\
\sin c \cos \alpha &= \cos h \sin b_1, \\
\sin c \sin \alpha &= \sin h.
\end{aligned}$$

By substituting

$$\begin{aligned}
\sin b_1 = \sin (b - b_2) &= \sin b \cos b_2 - \cos b \sin b_2, \\
\cos b_1 = \cos (b - b_2) &= \cos b \cos b_2 + \sin b \sin b_2
\end{aligned}$$

we obtain

$$\begin{aligned}
\cos c &= \cos h (\cos b \cos b_2 + \sin b \sin b_2) = \\
&= \cos b (\cos h \cos b_2) + \sin b (\cos h \sin b_2) = \\
&= \cos b \cos a + \sin b \sin a \cos \gamma, \qquad (1.3)
\end{aligned}$$

the so-called *cosine rule of spherical trigonometry*, and

$$\begin{aligned}
\sin c \cos \alpha &= \cos h (\sin b \cos b_2 - \cos b \sin b_2) = \\
&= \sin b (\cos h \cos b_2) - \cos b (\cos h \sin b_2) = \\
&= \sin b \cos a - \cos b \sin a \cos \gamma.
\end{aligned}$$

From the two "$\sin h$" formulas we obtain

$$\sin c \sin \alpha = \sin a \sin \gamma,$$

or, more generally,

$$\frac{\sin a}{\sin \alpha} = \frac{\sin b}{\sin \beta} = \frac{\sin c}{\sin \gamma}, \qquad (1.4)$$

the so-called *sine rule of spherical trigonometry.*

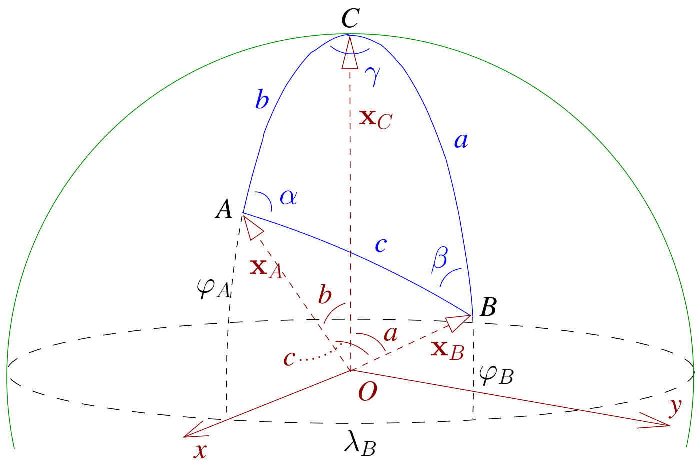


Figure 1.4.: The spherical triangle ABC, for deriving the cosine and sine rules using vectors in space

For comparison the corresponding formulas for a plane triangle:

$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$

and

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

At least in the case of the sine rule it is clear, that in the limit for a small triangle $\sin a \longrightarrow a$ etc., in other words, the spherical sine rule morphs into the one for a plane triangle. For the cosine rule this is not immediately clear.

1.6. Deriving the formulas with the aid of vectors in space

If *on the sphere* we look at a triangle consisting of two points $A\left(\varphi_A = \frac{\pi}{2} - b, \lambda_A = 0\right)$ ja $B = \left(\varphi_B = \frac{\pi}{2} - a, \lambda_B = \gamma\right)$ and a *pole* $C = \left(\varphi_C = \frac{\pi}{2}, \lambda_C = arbitrary\right)$, see figure 1.4, we may write two vectors:

$$\mathbf{x}_A = \begin{bmatrix} x_A \\ y_A \\ z_B \end{bmatrix} = \begin{bmatrix} \cos\varphi_A \\ 0 \\ \sin\varphi_A \end{bmatrix}, \mathbf{x}_B = \begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} = \begin{bmatrix} \cos\varphi_B\cos\lambda_B \\ \cos\varphi_B\sin\lambda_B \\ \sin\varphi_B \end{bmatrix}.$$

The vectors' *dot product* is

$$\begin{aligned} \cos c = \mathbf{x}_A \cdot \mathbf{x}_B &= \cos\varphi_A\cos\varphi_B\cos\lambda_B + \sin\varphi_A\sin\varphi_B = \\ &= \sin b\sin a\cos\gamma + \cos b\cos a, \end{aligned}$$

the *cosine rule* for a spherical triangle.

The *cross product* of the vectors is

$$\begin{aligned} \mathbf{x}_A \times \mathbf{x}_B &= \begin{bmatrix} -\sin\varphi_A\cos\varphi_B\sin\lambda_B \\ \sin\varphi_A\cos\varphi_B\cos\lambda_B - \cos\varphi_A\sin\varphi_B \\ \cos\varphi_A\cos\varphi_B\sin\lambda_B \end{bmatrix} = \\ &= \begin{bmatrix} -\cos a\sin b\sin\gamma \\ \cos a\sin b\cos\gamma - \sin a\cos b \\ \sin b\sin a\sin\gamma \end{bmatrix}. \end{aligned}$$

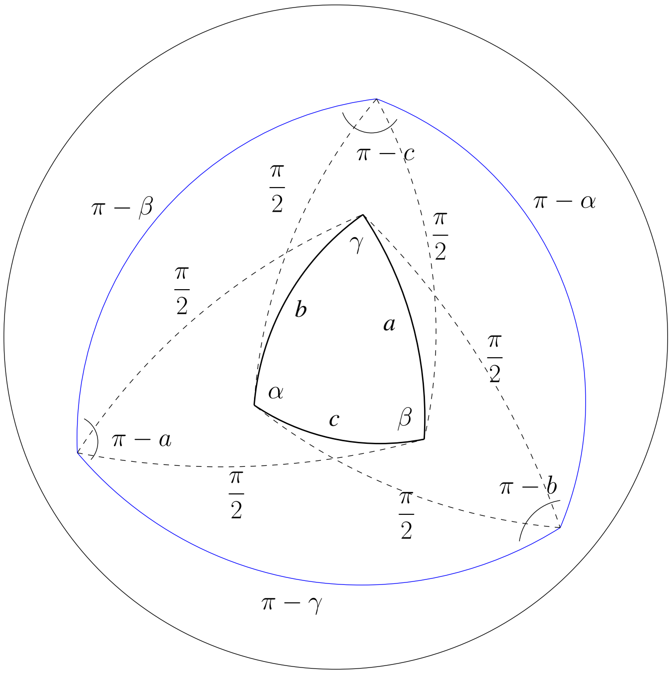


Figure 1.5.: Polarization of a spherical triangle

When the third vector is

$$\mathbf{x}_C = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

we obtain the *volume* of the parallelepiped spanned by the three vectors (i.e., twice the volume of the tetrahedron $ABCO$) as follows:

$$\mathrm{Vol}\,\{\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C\} = (\mathbf{x}_A \times \mathbf{x}_B) \cdot \mathbf{x}_C = \sin b \sin a \sin \gamma.$$

The volume contained by the three vectors does not however depend on the order of the vectors, so also

$$\mathrm{Vol}\,\{\mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C\} = \sin b \sin c \sin \alpha = \sin a \sin c \sin \beta.$$

Division yields

$$\begin{aligned} \sin a \sin \gamma &= \sin c \sin \alpha, \\ \sin b \sin \gamma &= \sin c \sin \beta, \end{aligned}$$

i.e.,

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

the *sine rule* for a spherical triangle.

1.7. Polarization

For every corner of the triangle we may define an “equator” or great circle, one “pole” of which is that corner point. If we do this, we obtain three “equators”, which themselves also form a triangle. This procedure is called *polarization*.

Because the angular distance between the two corner points on the surface of the sphere is the length of the side, the angle between the planes of two such great circles must be the same as this length. And the “polarization triangle”’s angle is 180° minus this.

The polarization method is *symmetric*: the original triangle is also the polarization of the polarization triangle. The intersection point of two edges of the polarization triangle is at a distance of 90° from both "poles", i.e., the corners of the original triangle, and the edge between them thus is the "equator" of the intersection point.

For symmetry reasons also the length of an edge of the polarization triangle equals 180° minus the corresponding angle of the original triangle.

For an arbitrary angle α we have:

$$\begin{aligned}\sin(180^\circ - \alpha) &= \sin\alpha \\ \cos(180^\circ - \alpha) &= -\cos\alpha\end{aligned}$$

Because of this we obtain of the spherical trigonometry cosine rule (1.3) the following *polarized version*:

$$-\cos\gamma = (-\cos\beta)(-\cos\alpha) + \sin\beta\sin\alpha(-\cos c)$$

or

$$\cos\gamma = -\cos\beta\cos\alpha + \sin\beta\sin\alpha\cos c,$$

a formula with which one may calculate an angle if the two other angles and the side between them are given.

1.8. Solving the spherical triangle by the method of additaments

In the additaments method we reduce a spherical triangle to a plane triangle by changing the lengths of the sides. Generally all angles and one side of a triangle are given, and the problem is to compute the other sides.

As the sides are small in comparison with the radius R of the Earth, we may write (series expansion):

$$\sin\psi = \psi - \frac{1}{6}\psi^3 + \ldots \approx \frac{s}{R}\left(1 - \frac{s^2}{6R^2}\right).$$

Now the spherical sine rule is ($s = a, b, c$):

$$\frac{a(1-\partial a)}{\sin\alpha} = \frac{b(1-\partial b)}{\sin\beta} = \frac{c(1-\partial c)}{\sin\gamma},$$

jossa $\partial s = \dfrac{s^2}{6R^2}$: $\partial a = \dfrac{a^2}{6R^2}$, $\partial b = \dfrac{b^2}{6R^2}$ ja $\partial c = \dfrac{c^2}{6R^2}$.

The method works now so, that

1. From the known side we subtract its additament ∂s;
2. The other sides are computed using the sine rule for a plane triangle and the known angle values ;
3. To the computed sides are now added *their* additaments.

The additaments are computed using the best available approximate values; if they are initially poor, we iterate.

The additament correction

$$s' = s(1 - \partial s)$$

may be changed from a combination of a multiplication and a subtraction into *a simple subtraction* by taking logarithms:

$$\ln s' = \ln s + \ln(1 - \partial s) = \ln s + (0 - \partial s) = \ln s - \partial s,$$

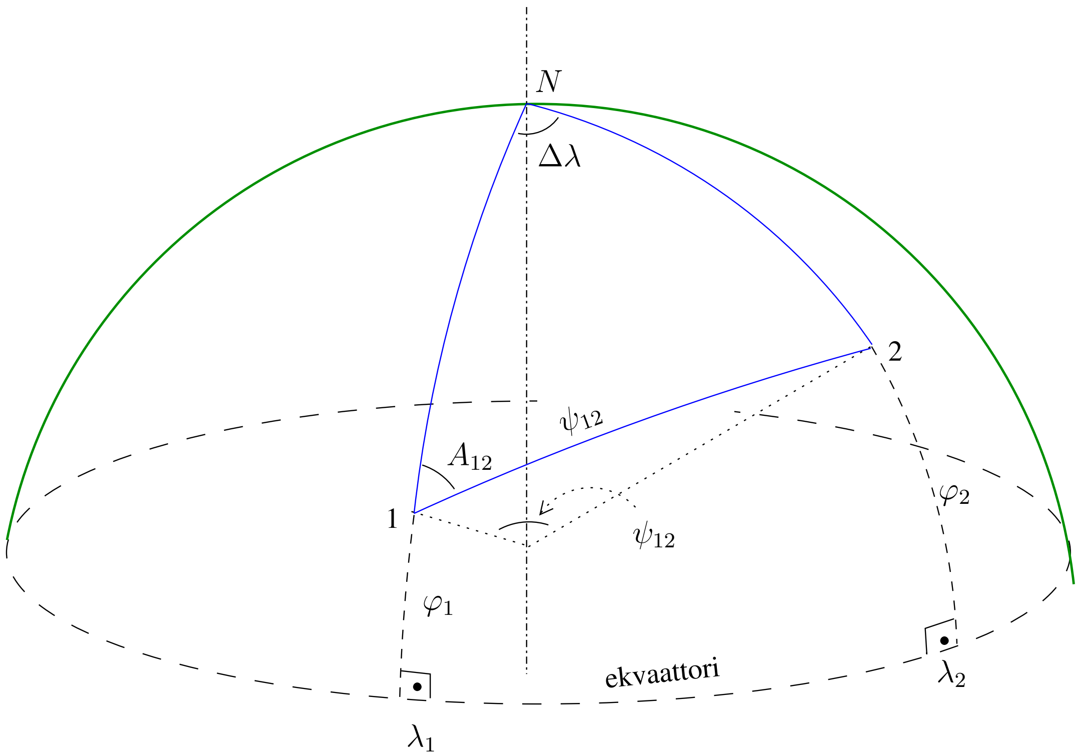


Figure 1.6.: The triangle $1-2-N$ on the globe. N is the North pole

or, in base-ten logarithms

$$^{10}\log s' = ^{10}\log s - M \cdot \partial s,$$

where $M = ^{10}\log e = 0.43429448$. In the age of logarithm tables this made the practical computations significantly easier.

1.9. Solving the spherical triangle by Legendre's method

In the Legendre method the reduction from a spherical to a plane triangle is done by changing the *angles*. If again all angles and one edge is given, we apply the following formula:

$$\frac{a}{\sin(\alpha - \varepsilon/3)} = \frac{b}{\sin(\beta - \varepsilon/3)} = \frac{c}{\sin(\gamma - \varepsilon/3)},$$

i.e., from every angle we subtract *one third of the spherical excess* ε.

It is however important to understand that the further calculations must be made *using the original angles* α, β, γ! The removal of the spherical excess is *only* done for the computation of the unknown sides of the triangle.

Nowadays these approximate methods (additaments and Legendre) are no longer used. It is easy to compute directly by computer using the spherical sine rule.

1.10. The forward geodetic problem on the sphere

The spherical trigonometry cosine and sine rules, suitably applied:

$$\sin\varphi_2 = \sin\varphi_1 \cos\psi_{12} + \cos\varphi_1 \sin\psi_{12} \cos A_{12},$$

and

$$\begin{aligned}\frac{\sin(\lambda_2-\lambda_1)}{\sin\psi_{12}} &= \frac{\sin A_{12}}{\cos\varphi_2} \Rightarrow \\ \lambda_2 &= \lambda_1 + \arcsin\left(\frac{\sin\psi_{12}\sin A_{12}}{\cos\varphi_2}\right).\end{aligned}$$

1.11. The geodetic inverse problem on the sphere

The cosine and sine rules of spherical trigonometry, suitably applied:

$$\cos\psi_{12} = \sin\varphi_1\sin\varphi_2 + \cos\varphi_1\cos\varphi_2\cos(\lambda_2-\lambda_1), \tag{1.5}$$

$$\sin A_{12} = \cos\varphi_2\frac{\sin(\lambda_2-\lambda_1)}{\sin\psi_{12}}. \tag{1.6}$$

1.12. The half-angle cosine rule

The above spherical cosine rule (1.3):

$$\cos a = \cos b\cos c + \sin b\sin c\cos\alpha,$$

is numerically ill-behaved when the triangle is very small compared to the sphere, in other words, if a, b, c are small. E.g., the triangle Helsinki-Tampere-Turku is very small compared to the globe, some 200 km/6378 km ~ 0.03. Then $\cos b\cos c \sim 0.999$, but $\sin b\sin c \sim 0.0009$! We are adding two terms of which one is approx. 1 and the other about a thousand times smaller. That is the way to lose numerical precision.

For solving this, we first remark, that

$$\begin{aligned}\cos\alpha &= 1-2\sin^2\frac{\alpha}{2};\\ \cos a &= 1-2\sin^2\frac{a}{2};\end{aligned}$$

and

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

after a few rearrangements we obtain the half-angle cosine rule for a spherical triangle, which also for small triangles is wel behaved[1]:

$$\sin^2\frac{a}{2} = \sin^2\frac{b-c}{2} + \sin b\sin c\sin^2\frac{\alpha}{2}.$$

[1] ...which, surprise, contains only sines!

Chapter 2

The geometry of the reference ellipsoid

2.1. Introduction

The reference ellipsoid is already a pretty precise description of the true figure of the Earth. The deviations of mean sea level from the GRS80 reference ellipsoid are of magnitude $\pm 100\,\mathrm{m}$.

Regrettably, the geometry of the reference ellipsoid is not as simple as that of the sphere. Nevertheless, rotational symmetry brings in at least one beautiful invariant.

The geodetic forward and reverse problems have traditionally been solved by series expansions of many terms, the coefficients of which also contain many terms. Here we rather offer numerical methods, which are conceptually simpler and easier to implement in an error-free way.

2.2. The geodesic as solution to a system of differential equations

We may write in a small rectangular triangle (dy, dx the metric east and north shift, see Fig. 2.1):

$$\begin{aligned} dx = M(\varphi)\, d\varphi &= \cos A ds, \\ dy = p(\varphi)\, d\lambda &= \sin A ds \text{ ja} \\ dA &= \sin \varphi d\lambda, \end{aligned}$$

where $p = N \cos \varphi$ is the distance from the axis of rotation, and M and N are the meridional and transversal curvatures, respectively.

The system of equations, normalized:

$$\begin{aligned} \frac{d\varphi}{ds} &= \frac{\cos A}{M(\varphi)}, \\ \frac{d\lambda}{ds} &= \frac{\sin A}{p(\varphi)}, \\ \frac{dA}{ds} &= \sin \varphi \frac{d\lambda}{ds} = \frac{\sin \varphi \sin A}{p(\varphi)}. \end{aligned} \tag{2.1}$$

Group 2.1 is valid not only on the ellipsoid of revolution; it applies to all figures of revolution. On a rotationally symmetric body we have $p(\varphi) = N(\varphi) \cos \varphi$, i.e.,

$$\begin{aligned} \frac{d\lambda}{ds} &= \frac{\sin A}{N(\varphi) \cos \varphi}, \\ \frac{dA}{ds} &= \frac{\tan \varphi \sin A}{N(\varphi)}. \end{aligned}$$

If the initial condition is given as $\varphi_1, \lambda_1, A_{12}$, we may obtain the geodesic $\varphi(s), \lambda(s), A(s)$ as a solution parametrized by arc length s. Numerically computing the solution using the

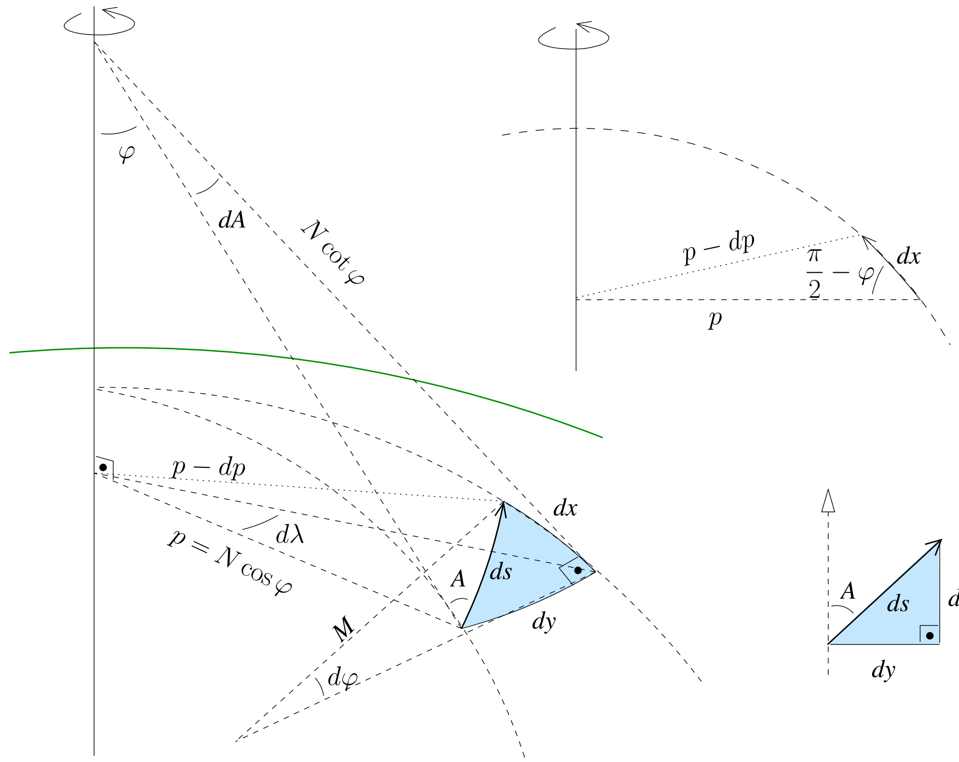


Figure 2.1.: The geometry of integrating the geodesic

MatLab software is also fairly easy thanks to the `ODE` routines ("Ordinary Differential Equation") provided.

Transformation to rectangular form is easy:

$$\begin{aligned} X(s) &= N(\varphi(s))\cos\varphi(s)\cos\lambda(s), \\ Y(s) &= N(\varphi(s))\cos\varphi(s)\sin\lambda(s), \\ Z(s) &= \frac{b^2}{a^2}N(\varphi)\sin\varphi(s). \end{aligned}$$

2.3. An invariant

Let us look closer at the quantity $p(\varphi) = N(\varphi)\cos\varphi$, the distance of a point from the rotation axis. let us compute the s derivative:

$$\frac{dp}{ds} = \frac{dp}{dx}\frac{dx}{ds} = -\sin\varphi\cos A.$$

Now, with the equation 2.1

$$\frac{dA}{ds} = \frac{\sin\varphi\sin A}{p}$$

we obtain by division

$$\frac{dp}{dA} = -\frac{\cos A}{\sin A}p.$$

Now

$$\begin{aligned}\frac{d\left(p\sin A\right)}{dA} &= \frac{dp}{dA}\sin A + p\cos A = \\ &= -\frac{\cos A}{\sin A}p\cdot\sin A + p\cos A = \\ &= p\left(-\cos A + \cos A\right) = 0.\end{aligned}$$

Result:

$$\boxed{\textit{the expression } p\sin A \textit{ is an invariant.}^1}$$

This applies *on all rotationally symmetric bodies*, i.e., also on the ellipsoid of revolution – where this is called the CLAIRAUT equation –, and of course on the plane. This invariant can be used to eliminate the differential equation in A from the system 2.1. This yields

$$\frac{d\varphi}{ds} = \frac{\cos A\left(\varphi\right)}{M\left(\varphi\right)}, \tag{2.2}$$

$$\frac{d\lambda}{ds} = \frac{\sin A\left(\varphi\right)}{p\left(\varphi\right)}, \tag{2.3}$$

where $A\left(\varphi\right)$ is obtained from the invariant formula

$$\sin A\left(\varphi\right) = \sin A_{12}\frac{p\left(\varphi_1\right)}{p\left(\varphi\right)}. \tag{2.4}$$

The shape of the object is defined by giving the function $p\left(\varphi\right)$.

2.4. The geodetic main problem

Solving the forward geodetic problem now amounts simply to substituting the also given arc length s_{12} into this solution.

Computing the solution is easiest in practice using numerical integration; the methods are found in textbooks on numerical analysis, and the routines needed in many numerical libraries.

The "classical" alternative, series expansions found in many older textbooks, are more efficient in theory but complicated.

For the ellipsoid we may specialize the equations 2.2, 2.3 ja 2.4 with the aid of the following expressions:

$$\begin{aligned}M\left(\varphi\right) &= a\left(1-e^2\right)\left(1-e^2\sin^2\varphi\right)^{-3/2}, \\ p\left(\varphi\right) &= N\left(\varphi\right)\cos\varphi = a\left(1-e^2\sin^2\varphi\right)^{-1/2}\cdot\cos\varphi.\end{aligned}$$

2.5. The geodetic inverse problem

A direct numerical method for solving the inverse problem is *iteration*.

Let us have as given φ_1, λ_1 and φ_2, λ_2. First we compute the approximate values[2] $\left(A_{12}^{(1)}, s_{12}^{(1)}\right)$, and solve the geodetic forward problem in order to compute $\left(\varphi_2^{(1)}, \lambda_2^{(1)}\right)$. Thus we obtain the *closing errors*

$$\begin{aligned}\Delta\varphi_2^{(1)} &\equiv \varphi_2^{(1)} - \varphi_2, \\ \Delta\lambda_2^{(1)} &\equiv \lambda_2^{(1)} - \lambda_2.\end{aligned}$$

[2] ... e.g., using the closed formulas of spherical trigonometry.

Next, these closing errors could be used for computing improved values $\left(A_{12}^{(2)}, s_{12}^{(2)}\right)$, and so on.

The nearly linear dependence between (A_{12}, s_{12}) and (φ_2, λ_2) can be approximated using *spherical geometry*. The formulas needed (1.5, 1.6):

$$\begin{aligned}
\cos s_{12} &= \sin\varphi_1 \sin\varphi_2 + \cos\varphi_1 \cos\varphi_2 \cos(\lambda_2 - \lambda_1), \\
\sin s_{12} \sin A_{12} &= \cos\varphi_2 \sin(\lambda_2 - \lambda_1).
\end{aligned}$$

From this, by differentiation:

$$\begin{aligned}
-\sin s_{12}\Delta s_{12} &= [\sin\varphi_1 \cos\varphi_2 - \cos\varphi_1 \sin\varphi_2 \cos(\lambda_2 - \lambda_1)]\,\Delta\varphi_2 - \\
&\quad - \cos\varphi_1 \cos\varphi_2 \sin(\lambda_2 - \lambda_1)\,\Delta\lambda_2, \\
\sin s_{12} \cos A_{12}\Delta A_{12} + \cos s_{12} \sin A_{12}\Delta s_{12} &= -\sin\varphi_2 \sin(\lambda_2 - \lambda_1)\,\Delta\varphi_2 + \\
&\quad + \cos\varphi_2 \cos(\lambda_2 - \lambda_1)\,\Delta\lambda_2.
\end{aligned}$$

So, if we write

$$A = \begin{bmatrix} \sin\varphi_1 \cos\varphi_2 - \cos\varphi_1 \sin\varphi_2 \cos(\lambda_2 - \lambda_1) & -\cos\varphi_1 \cos\varphi_2 \sin(\lambda_2 - \lambda_1) \\ -\sin\varphi_2 \sin(\lambda_2 - \lambda_1) & \cos\varphi_2 \cos(\lambda_2 - \lambda_1) \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & -\sin s_{12} \\ \sin s_{12} \cos A_{12} & \cos s_{12} \sin A_{12} \end{bmatrix},$$

we obtain as *iteration formulas*:

$$\begin{bmatrix} s_{12}^{(i+1)} \\ A_{12}^{(i+1)} \end{bmatrix} = \begin{bmatrix} s_{12}^{(i)} \\ A_{12}^{(i)} \end{bmatrix} + \begin{bmatrix} \Delta s_{12}^{(i)} \\ \Delta A_{12}^{(i)} \end{bmatrix} = \begin{bmatrix} s_{12}^{(i)} \\ A_{12}^{(i)} \end{bmatrix} + \mathrm{B}^{-1}\mathrm{A} \begin{bmatrix} \Delta\varphi_2^{(i)} \\ \Delta\lambda_2^{(i)} \end{bmatrix}.$$

Using the new values $\left(s_{12}^{(i+1)}, A_{12}^{(i+1)}\right)$ we repeat the computation of the geodetic forward problem to obtain new values $\left(\varphi_2^{(i+1)}, \lambda_2^{(i+1)}\right)$ until convergence. The matrices A, B can be recomputed if needed using better approximate values.

Chapter 3

Co-ordinates on the reference ellipsoid

3.1. Representations of the sphere and the ellipsoid

An implicit representation of the *circle* is

$$x^2 + y^2 - a^2 = 0,$$

where a is the radius (Pythagoras). The parametric representation is

$$\begin{aligned} x &= a \cos \beta, \\ y &= a \sin \beta. \end{aligned}$$

From this we obtain an ellipse by "squeezing" the y axis by the factor b/a, i.e.,

$$\begin{aligned} x &= a \cos \beta, \\ y &= b \sin \beta, \end{aligned}$$

from which again

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = \sin^2 \beta + \cos^2 \beta = 1$$

or

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$$

is the implicit representation.

3.2. Various latitude types

The latitude on the ellipsoid of revolution can be defined in at least three different ways. Let us consider a cross-section of the ellipsoid, itself an ellipse; the so-called *meridian ellipse*.

The figure shows the following three concepts of latitude:

1. Geographical latitude φ: the direction of the ellipsoidal normal relative to the plane of the equator;
2. Geocentric latitude ϕ (tai ψ): the angle of the line connecting the point with the origin, relative to the plane of the equator;
3. Reduced latitude β: point P is shifted straight in the y direction to the circle around the merian ellipse, to become point Q. The geocentric latitude of point Q is the reduced latitude of point P.

Reduced latitude is used only in theoretical contexts. In maps, geographical (i.e., geodetic, or sometimes, ellipsoidal) latitude is used. Geocentric latitude appears in practice only in satellite and space geodesy.

The *longitude* λ is the same whether we are considering geographical, geocentric or reduced co-ordinates.

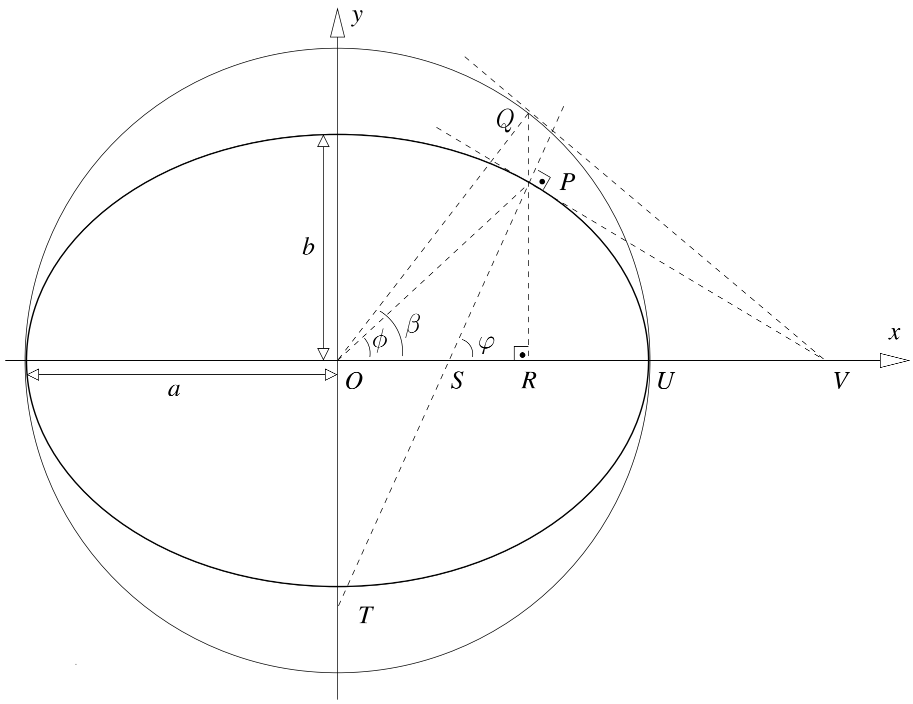


Figure 3.1.: The meridian ellipse and various types of latitude

3.3. Measures for the flattening

The flattening of the reference ellipsoid is described by a variety of measures:

1. The flattening $f = \dfrac{a-b}{a}$;
2. The first eccentricity (square) $e^2 = \dfrac{a^2-b^2}{a^2}$;
3. The second eccentricity (square) $e'^2 = \dfrac{a^2-b^2}{b^2}$.

3.4. Relationships between different types of latitude

From figure 3.1 can be seen that

$$\frac{PR}{QR} = \frac{b}{a}.$$

In the triangles ORQ and ORP we have

$$\tan\beta = \frac{QR}{OR} \text{ ja } \tan\phi = \frac{PR}{OR};$$

so

$$\frac{\tan\phi}{\tan\beta} = \frac{PR}{QR} = \frac{b}{a},$$

i.e.,

$$\tan\beta = \frac{a}{b}\tan\phi.$$

In the triangles RVQ, RVP:

$$\tan\left(\frac{\pi}{2}-\beta\right) = \frac{QR}{VR} \text{ ja } \tan\left(\frac{\pi}{2}-\varphi\right) = \frac{PR}{VR},$$

i.e.,

$$\frac{\cot\varphi}{\cot\beta} = \frac{PR}{QR} = \frac{b}{a},$$

eli

$$\tan\varphi = \frac{a}{b}\tan\beta.$$

By combining still

$$\tan\varphi = \frac{a^2}{b^2}\tan\phi.$$

3.5. Co-ordinates in the meridional ellipse

Let us compute the co-ordinates x, y of the point P on the surface of the ellipsoid as follows.

We mark the distance PT with the symbol N, the so-called *transversal radius of curvature*, i.e., the radius of curvature of the ellipsoid in the West-East direction.

Now we have

$$x = N\cos\varphi. \tag{3.1}$$

Also

$$PR = OR\tan\phi = OR\frac{b^2}{a^2}\tan\varphi = N\cos\varphi\cdot\left(1-e^2\right)\tan\varphi$$

using $OR = x = N\cos\varphi$. The end result is because $y = PR$:

$$y = N\left(1-e^2\right)\sin\varphi. \tag{3.2}$$

The equations 3.1, 3.2 represent a description of the meridian ellipse as a function of geodetic latitude φ. *Remember* that N is a function of φ too, so *not a constant*! In fact

$$\begin{aligned}
\frac{x^2}{a^2}+\frac{y^2}{b^2} &= \frac{N^2}{a^2}\cos^2\varphi + \frac{N^2\left(1-e^2\right)^2}{b^2}\sin^2\varphi = \\
&= \frac{N^2}{a^2}\cos^2\varphi + \frac{N^2}{a^2}\frac{b^2}{a^2}\sin^2\varphi = \\
&= \frac{N^2}{a^2}\left(\cos^2\varphi + \frac{b^2}{a^2}\sin^2\varphi\right) = 1;
\end{aligned}$$

the latter condition yields

$$N = \frac{a}{\sqrt{1-e^2\sin^2\varphi}},$$

using the definition $e^2 \equiv \dfrac{a^2-b^2}{a^2}$.

3.6. Three-dimensional rectangular co-ordinates on the reference ellipsoid

The above formulas are easily generalized: if x and y are co-ordinates within the meridian section, the rectangular co-ordinates are

$$\begin{aligned}
X = x\cos\lambda &= N\cos\varphi\cos\lambda, \\
Y = x\sin\lambda &= N\cos\varphi\sin\lambda, \\
Z = y &= N\left(1-e^2\right)\sin\varphi.
\end{aligned}$$

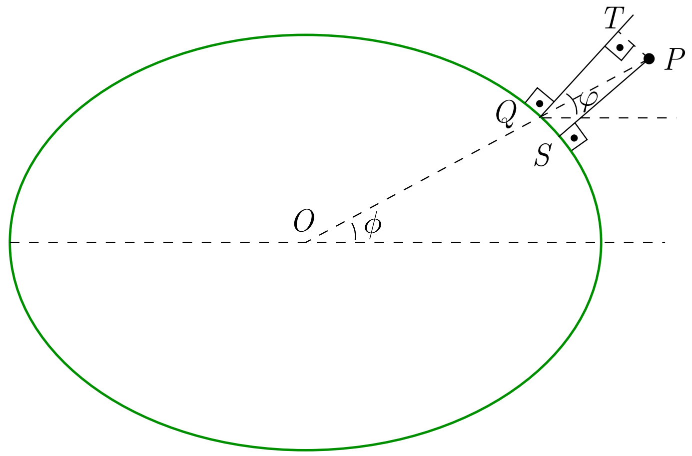


Figure 3.2.: Suorakulmaisista koordinaateista maantieteellisiin

If we look at points not on the surface of the reference ellipsoid but above or below it in space, we may write

$$\begin{aligned} X &= (N+h)\cos\varphi\cos\lambda, \\ Y &= (N+h)\cos\varphi\sin\lambda, \\ Z &= \left(N\left(1-e^2\right)+h\right)\sin\varphi. \end{aligned} \tag{3.3}$$

Here, h is the straight distance of the point from the surface of the ellipsoid ("ellipsoidal height"). This quantity is interesting because satellite positioning devices can be said to directly measure precisely this quantity (more precisely, they measure X, Y, Z and compute from these h).

3.7. Computing geographic co-ordinates from rectangular ones

This, the reverse problem from that of equations (3.3) , isn't quite simple to solve.

Closed solutions exist, but are complicated. Of course an iterative solution based directly on equations (3.3) is certainly possible and often used.

Computing the longitude λ is extremely simple:

$$\tan\lambda = \frac{Y}{X}.$$

A possible stumbling block is identifying the correct quadrant for λ.

φ and h are more complicated. See figure 3.2, where point P's rectangular co-ordinates X, Y, Z are known and the geographical φ, λ, h are to be computed.

Let us first compute the *geocentric latitude* using the formula:

$$\tan\phi = \frac{Z}{\sqrt{X^2+Y^2}}$$

and the *geometric distance* (radius) by the equation:

$$OP = \sqrt{X^2+Y^2+Z^2}.$$

If point P would *be located on the reference ellipsoid, we* could *determine its geographical latitude* φ by the following formula:

$$\tan\varphi_P = \frac{Z}{\left(1-e^2\right)\sqrt{X^2+Y^2}}$$

(proof directly from eqs. (3.3).) Now that P is located above the ellipsoid, we obtain in this way the latitude φ_Q of point Q , where Q is the intersection of the ellipsoid and the radius of P, for which (geometrically obviously) the above ratio is the same as for P.

Now we compute

$$\begin{aligned} X_Q &= N \cos \varphi_Q \cos \lambda, \\ Y_Q &= N \cos \varphi_Q \sin \lambda, \\ Z_Q &= N \left(1 - e^2\right) \sin \varphi_Q, \end{aligned}$$

from which

$$OQ = \sqrt{X_Q^2 + Y_Q^2 + Z_Q^2}$$

and thus

$$PQ = OP - OQ.$$

Additionally, in the little triangle TQP we may compute (φ_Q abbreviated to φ):

$$\angle TQP = \varphi - \phi$$

and thus

$$\begin{aligned} TP &= PQ \sin(\phi - \varphi), \\ TQ &= PQ \cos(\varphi - \phi). \end{aligned}$$

Now

$$h = PS \approx TQ$$

and

$$\varphi_P \approx \varphi_Q - \frac{TP}{PO}, \tag{3.4}$$

or perhaps a hair's width more precise [1]

$$\varphi_P \approx \varphi_Q - \frac{TP}{PO} \cos(\varphi - \phi). \tag{3.5}$$

This procedure is in practice fairly precise. If $h = 8000\,\mathrm{m}$ and $\varphi = 45°$, then $TP \approx 26\,\mathrm{m}$, between the φ_P solutions (3.4) and (3.5) there is a difference of 0.1 mm[2], which is also the order of magnitude of the error that is possibly present in the different solutions. The approximation $PS \approx TQ$ contains 0.05 mm of error.[3].

3.8. Meridian arc length

The length of a meridian arc, a quantity needed, e.g., with map projections (UTM, Gauss-Krüger) is computed by integration.

Above we already defined the quantity N, the *transverse radius of curvature*. The other radius of curvature of the Earth surface is the *meridional radius of curvature* M. If it is given as a function of latitude φ, we compute an element of path ds as follows:

$$ds = M d\varphi.$$

[1] Or then, not. Instead of PO:n one should take $M(\varphi) + h$, where M the meridional radius of curvature.

[2] Linearly $TP(1 - \cos(\phi - \varphi))$.

[3] $\frac{1}{2}\frac{TP^2}{OP}$.

Now we may calculate the length of a meridian arc as follows:

$$s(\varphi_0) = \int_0^{\varphi_0} M d\varphi. \tag{3.6}$$

On the reference ellipsoid

$$M = \frac{a\left(1 - e^2\right)}{\left(1 - e^2 \sin^2 \varphi\right)^{3/2}},$$

i.e.,

$$s(\varphi_0) = a\left(1 - e^2\right) \int_0^{\varphi_0} \left(1 - e^2 \sin^2 \varphi\right)^{-3/2} d\varphi.$$

Here, the last factor can be expanded into a series – because $e^2 \sin^2 \varphi \ll 1$ – and integrated termwise. See the literature. Of course also a numeric approach is possible, nowaday it may even be the superior alternative. MatLab offers for this purpose its `QUAD` (quadrature) routines.

Chapter 4

Reference systems

4.1. The GRS80 system and geometric parameters

Nowadays, the overwhelmingly most used global geodetic reference system is the *Geodetic Reference System 1980*, GRS80. The parameters defining it (e.g., (Heikkinen, 1981)) are given in table 4.1.

Some of the parameters are geometric (a), some are dynamic (J_2,ω). Other geometric and dynamic parameters may be derived as follows ((Moritz, 1992), following (Heiskanen and Moritz, 1967, eqs. 2-90, 2-92)):

$$\begin{aligned} J_2 &= \frac{e^2}{3}\left(1 - \frac{2}{15}\frac{me'}{q_0}\right) \Rightarrow \\ e^2 &= 3J_2 + \frac{2me'e^2}{15q_0}. \end{aligned}$$

Here(Heiskanen and Moritz, 1967, eq. 2-70))

$$m = \frac{\omega^2 a^2 b}{GM}$$

and (based on the definitions)

$$be' = ae$$

with the aid of which

$$e^2 = 3J_2 + \frac{4}{15}\frac{\omega^2 a^3}{GM}\frac{e^3}{2q_0}. \tag{4.1}$$

Furthermore we know ((Heiskanen and Moritz, 1967, eq. 2-58))

$$2q_0\left(e'\right) = \left(1 + \frac{3}{e'^2}\right)\arctan e' - \frac{3}{e'}$$

ja

$$e'\left(e\right) = \frac{e}{\sqrt{1-e^2}}.$$

Now we may compute e^2 *iteratively* using eq. (4.1) , which computes $q_0\left(e'\left(e\right)\right)$ anew in every step. The result is (table 4.2):

Quantity	Value	Explanation
a	6378137 m	semi-major axis
GM	$3986005 \cdot 10^8\,\mathrm{m^3 s^{-2}}$	Earth mass $\times G$
J_2	$108263 \cdot 10^{-8}$	Dynamic form factor
ω	$7292115 \cdot 10^{-11}\,\mathrm{rad\,s^{-1}}$	Angular rotation rate

Table 4.1.: GRS80 defining parameters

Quantity	Value	Explanation
e^2	0.00669438002290	Ensimmäinen eksentrisyys neliöön
e'^2	0.00673949677548	Toinen eksentrisyys neliöön
b	6356752.314140 m	Lyhyt akselipuolikas
$1/f$	298.257222101	Käänteinen litistyssuhde

Table 4.2.: GRS80 derived parameters

Quantity	Value (WGS84)	Remark
a	6378137 m	same
$1/f$	298.257223563	different!
b	6356752.314245 m	diff. 0.1 mm

Table 4.3.: WGS84 ellipsoidal parameters

Here we used $1-e^2 = 1-\frac{a^2-b^2}{a^2} = \frac{b^2}{a^2}$ ja $1-\frac{1}{f} = 1-\frac{a-b}{a} = \frac{b}{a}$, i.e.,

$$1-e^2 = \left(1-\frac{1}{f}\right)^2 \Rightarrow \frac{1}{f} = 1-\sqrt{1-e^2}.$$

Often we use (a, f) together to define the GRS80 reference ellipsoid.

The official reference system of the GPS system is the *World Geodetic System 1984* (WGS84), whose reference ellipsoid is *almost* identical with GRS80. However, *not exactly,* table 4.3:

Most often the difference, a bit over 0.1 mm at most, can be neglected. It is apparently due to imprecise numerics.

4.2. Gravimetric parameters

Computing geodetic parameters is done as follows ((Heiskanen and Moritz, 1967, kaavat 2-73, 2-74)):

$$\begin{aligned}\gamma_e &= \frac{GM}{ab}\left(1-m-\frac{m}{6}\frac{e'q_0'}{q_0}\right),\\ \gamma_p &= \frac{GM}{a^2}\left(1+\frac{m}{3}\frac{e'q_0'}{q_0}\right),\end{aligned}$$

where ((Heiskanen and Moritz, 1967, kaava 2-67)):

$$q_0'(e') = 3\left(1+\frac{1}{e'^2}\right)\left(1-\frac{1}{e'}\arctan e'\right)-1.$$

The solution is again obtained iteratively, yielding ($f^* \equiv \frac{\gamma_p-\gamma_e}{\gamma_e}$):

Quantity	Value	Explanation
γ_e	$9.7803267715\,\mathrm{ms}^{-2}$	Normal gravity at equator
γ_p	$9.8321863685\,\mathrm{ms}^{-2}$	Normal gravity at poles
f^*	0.00530244011229	"Gravity flattening"

As a check, we may still compute CLAIRAUT's equation in its precise form ((Heiskanen and Moritz, 1967, eq. 2-75)):

$$f + f^* = \frac{\omega^2 b}{\gamma_e}\left(1 + e'\frac{q_0'}{2q_0}\right).$$

A simple closed, beautiful formula for normal gravity on the reference ellipsoid is SOMIGLIANA-PIZZETTI's formula:

$$\gamma = \frac{a\gamma_e \cos^2\varphi + b\gamma_p \sin^2\varphi}{\sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}}.$$

4.3. Reference frames

Every twenty-four hours, the Earth rotates around its axis relative to the heavens at what is very nearly a constant rate, about what it very nearly a fixed axis. The direction of this rotation axis will serve as the z axis of both the celestial and the terrestrial frame. In order to completely define the orientation of our reference frame, we then need to *conventionally fix two longitudes*:

1. On the celestial sphere: we take for this the *vernal equinox*, where the Sun crosses the equator S-N

2. On the Earth: the International Meridian Conference in Washington DC, 1884, chose *Greenwich* as the prime meridian.

   A *bonus* of this choice, which was realized after the conference, was that at the same time was defined a single, unified *global time system*, comprising 15° broad hourly time zones, so – especially in the United States, which was expanding Westward over many time zones – the trains would run on time.

See figure 4.1.

Red denotes an ECEF (Earth-Centred, Earth-Fixed) reference frame, which co-rotates with the solid Earth, so the x axis always lies in the plane of the Greenwich meridian. This is also called a CT (Conventional Terrestrial) System. Locations on the Earth's surface are (almost) constant in this kind of frame, and can be published, e.g., on maps. However, moving vehicles, ocean water and atmospheric air masses will sense "pseudo-forces" (like the Coriolis force) due to the non-uniform motion of this reference frame

blue denotes a (quasi-)inertial system, which does not undergo any (rapid) rotations relative to the fixed stars. Also called a celestial reference frame, as the co-ordinates of the fixed stars are nearly constant in it and may be published. Also the equations of motion of, e.g., satellites or gyroscopes apply strictly, without pseudo-forces induced by non-uniform reference system motion.

A *Conventional Terrestrial System* (CTS) is defined as follows:

- the origin of the frame coincides with the centre of mass of the Earth

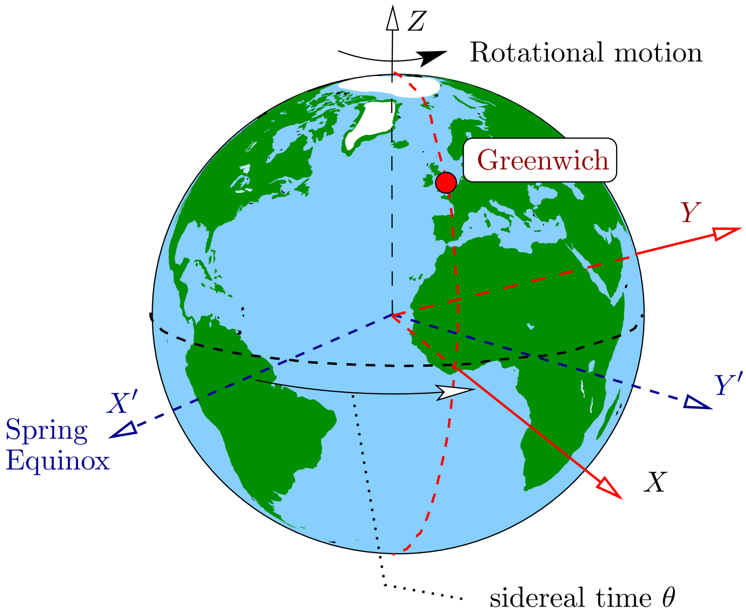


Figure 4.1.: Geocentric reference frames

- the Z axis is directed along the rotation axis of the Earth, more precisely the Conventional International Origin (CIO) , i.e., the average direction of the axis over the time span 1900-1905
- the XZ-plane is parallel to the zero meridian as defined by "Greenwich", more precisely by: earlier the BIH (Bureau International de l'Heure, International Time Bureau), today the IERS (International Earth Rotation and Reference Systems Service), based on their precise monitoring of the Earth rotation.

In figure 4.2 we see the Earth orbit or *ecliptic*, the Earth axis tilt relative to the ecliptic plane, and how this tilt causes the most impressive climating variation observable to human beings: the four seasons.

4.4. The orientation of the Earth

The orientation of the Earth's rotation axis undergoes slow changes. Relative to the stars, i.e., in inertial space, this motion consists of *precession* and *nutation*. It is caused by the gravitational torque exerted by the Sun and the Moon, which are either in (Sun) or close to (Moon) the ecliptic plane. See figure 4.3.

If we study the motion of the Earth's axis, and Earth rotation in general, relative to a reference frame connected to the solid Earth itself, we find different quantitities:

- Polar motion: this consists of an annual (forced) component and a 14-months component called the *Chandler wobble*.
- Length of Day.

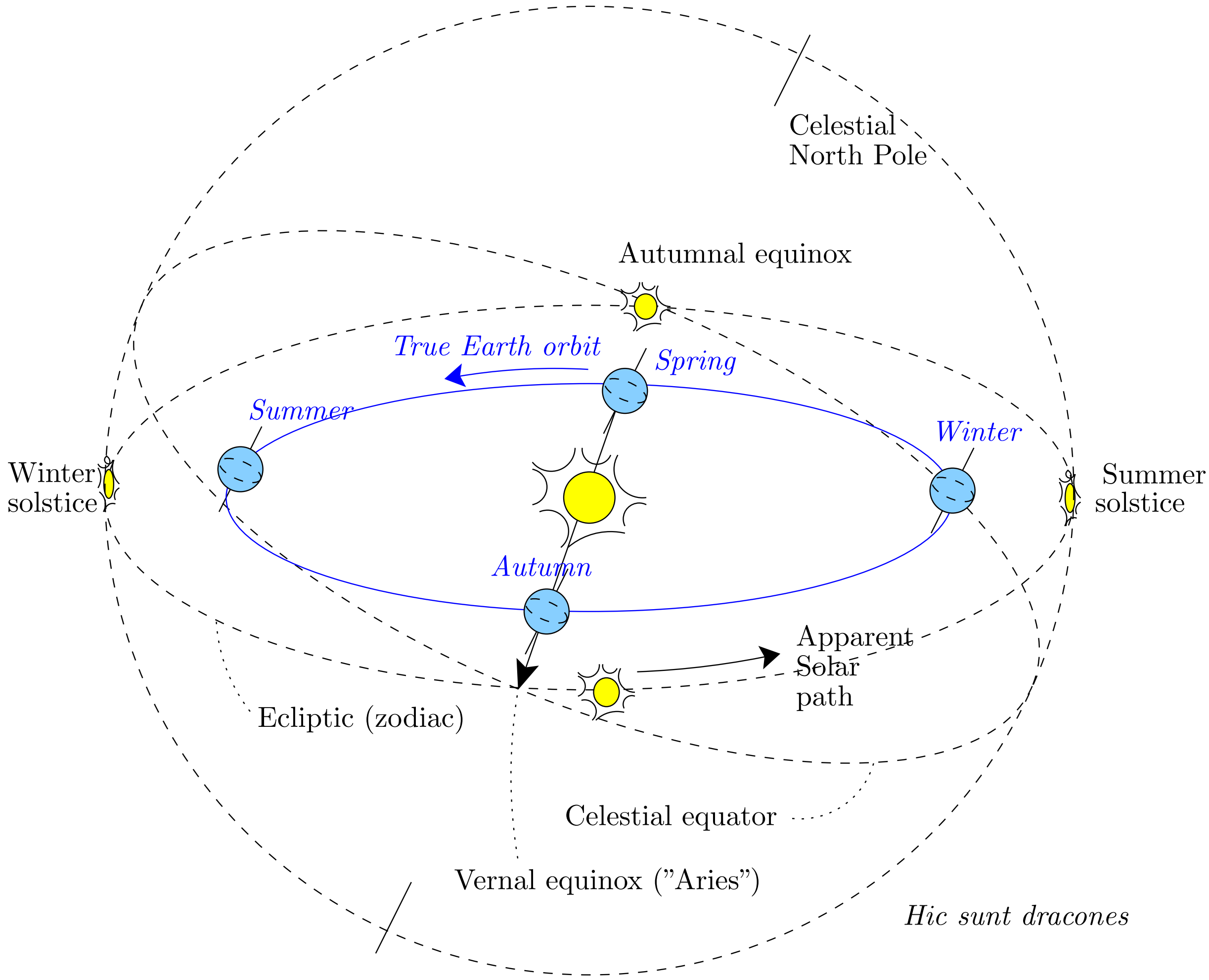


Figure 4.2.: Geometry of the Earth's orbit and rotation axis. The seasons indicated are boreal

Together these are called *Earth Orientation Parameters* (EOP). They are nowadays monitored routinely, and available after the fact from the International Earth Rotation Service over the Internet. The variation of these parameters is geophysically well understood, e.g., for the Chandler wobble it is mainly the pressure of the Earth's oceans and atmosphere that is responsible ((Gross, 2000)).

4.5. Transformations between systems

See the following diagram, which depicts only the *rotations* between the various systems:

$$\begin{array}{ccccccc} \text{Local} & \overset{\Phi,\Lambda}{\Longleftrightarrow} & \text{Conventional} & \overset{x_p, y_p}{\Longleftrightarrow} & \text{Instantaneous} & \overset{\theta_0}{\Longleftrightarrow} & \text{Real} \\ \text{Astronomical} & & \text{Terrestrial} & & \text{Terrestrial} & & \text{Astronomical} \end{array}$$

Here, Φ, Λ are local astronomical co-ordinates (direction of the plumbline), while x_p, y_p are the co-ordinates of the pole in the CIO system. θ_0 is Greenwich Apparent Sidereal Time (GAST).

In the sequel we shall show how these transformations are realized in practice.

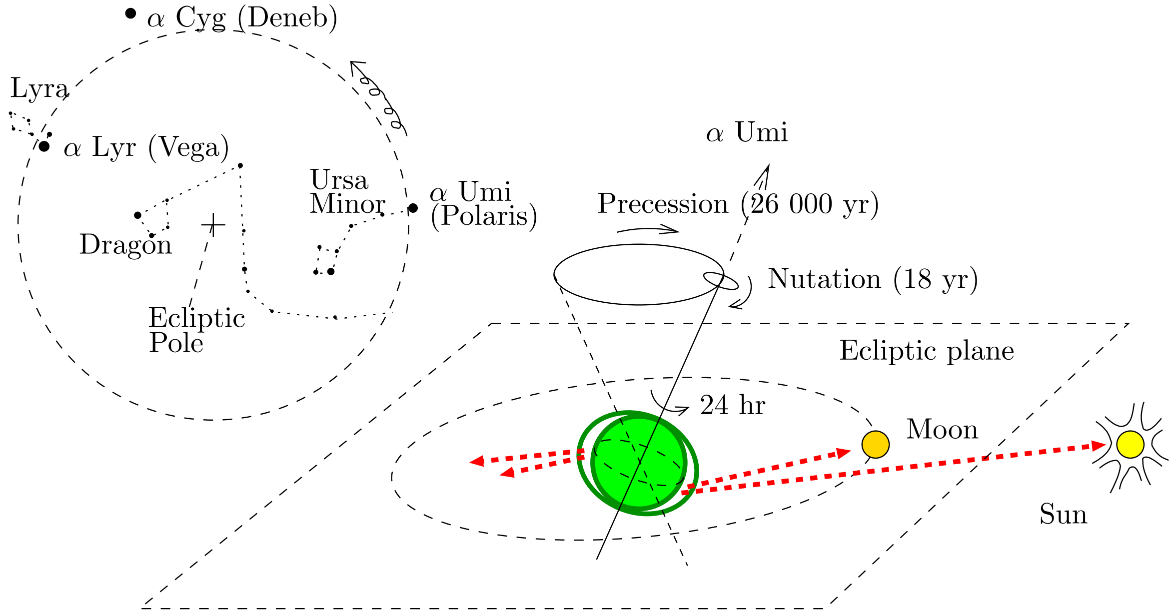


Figure 4.3.: Precession, nutation and the torques from Sun and Moon

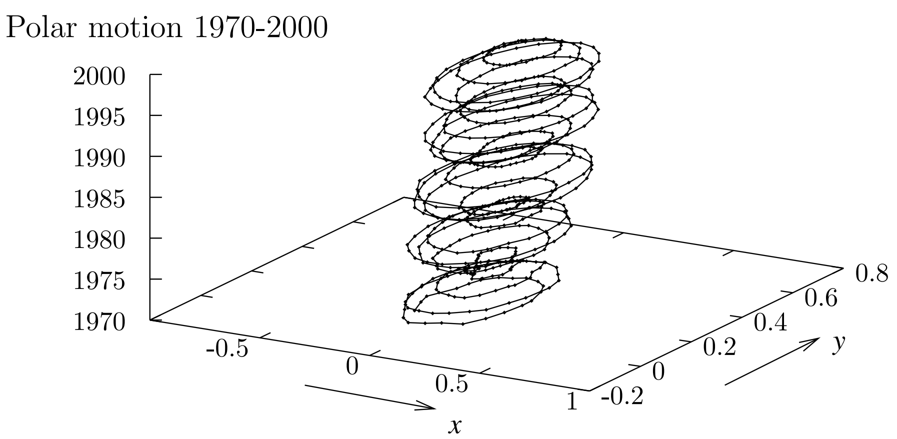


Figure 4.4.: Polar motion

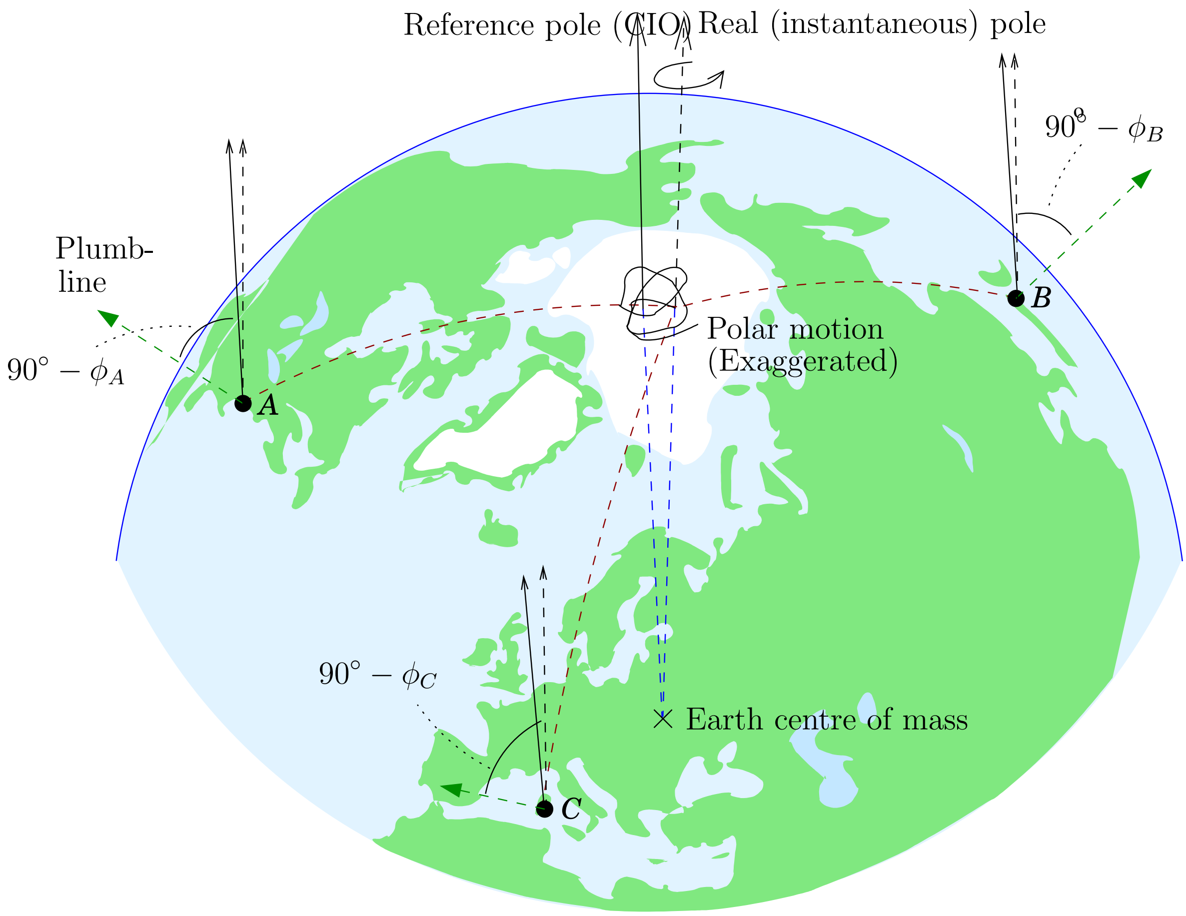


Figure 4.5.: How to monitor polar motion using latitude observatories

Chapter 5

Using rotation matrices

5.1. General

Always when we change the orientation of the axes of a co-ordinate system, we have, written in rectangular co-ordinates, a *multiplication with a rotation matrix.*

Let us investigate the matter in the(x, y) plane, figure 5.1.

The new co-ordinate

$$x'_P = OU = OR\cos\alpha,$$

where

$$OR = OS + SR = x_P + PS\tan\alpha = x_P + y_P\tan\alpha.$$

By substituting

$$x'_P = (x_P + y_P\tan\alpha)\cos\alpha = x_P\cos\alpha + y_P\sin\alpha.$$

In the same fashion

$$y'_P = OT = OV\cos\alpha,$$

where

$$OV = OQ - VQ = y_P - PQ\tan\alpha = y_P - x_P\tan\alpha,$$

where

$$y'_P = (y_P - x_P\tan\alpha)\cos\alpha = -x_P\sin\alpha + y_P\cos\alpha.$$

Summarizingly in a matrix equation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The place of the minus sign is the easiest to obtain by sketching both pairs of axes on paper, mark the angle α, and infer graphically whether the for a point on the positive x axis (i.e.: $y = 0$) the new y' co-ordinate is positive or negative: in the above case

$$y' = \begin{bmatrix} -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = -\sin\alpha \cdot x\,,$$

i.e., $y' < 0$, i.e., the minus sign is indeed in the lower left corner of the matrix.

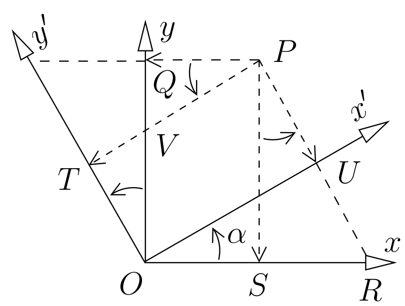


Figure 5.1.: Rotation in the plane

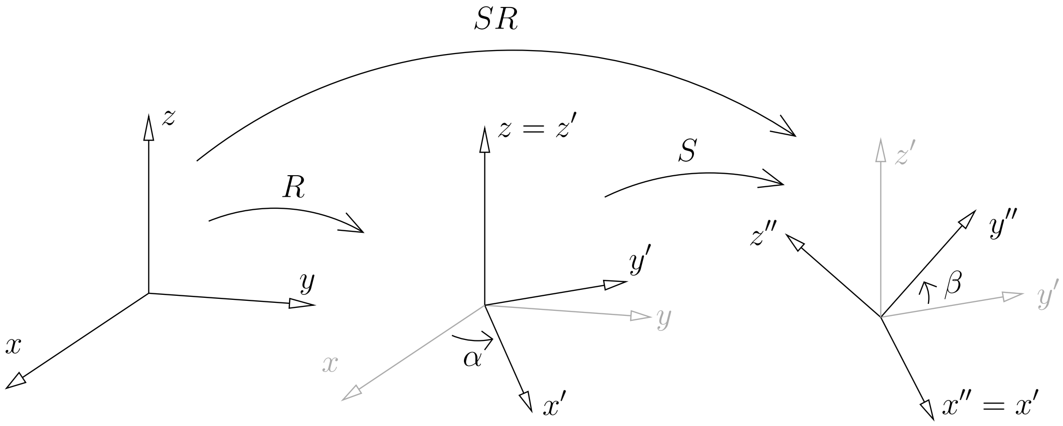


Figure 5.2.: The associativity of rotations

5.2. Chaining matrices in three dimensions

In a three-dimensional co-ordinate system we may write a two-dimensional rotation matrix as follows:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

i.e., the z co-ordinate is copied as such $z' = z$, while x and y transform into each other according to the above formula.

If there are several transformations in sequence, we obtain the final transformation by "chaining" the transformation matrices. I.e., if

$$\mathbf{r}'' = S\mathbf{r}', \ \mathbf{r}' = R\mathbf{r},$$

in which

$$R = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{bmatrix},$$

then (associativity):

$$\mathbf{r}'' = S(R\mathbf{r}) = (SR)\,\mathbf{r},$$

i.e., the matrices are multiplied with each other.

Remember that

$$RS \neq SR,$$

in other words, matrices and transformations are not commutative[1]!

See figure 5.2.

5.3. Orthogonal matrices

Rotation matrices are *orthogonal*, i.e.

$$RR^T = R^T R = I; \tag{5.1}$$

their inverse matrix is the same as the transpose.

[1] Two dimensional rotations *are* in fact commutative; they can be described also by complex numbers.

E.g.,

$$\begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}^{-1} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \left(= \begin{bmatrix} \cos(-\alpha) & \sin(-\alpha) \\ -\sin(-\alpha) & \cos(-\alpha) \end{bmatrix} \right),$$

completely understandable, because this is a rotation around the same axis, by the same amount, but *in the opposite direction.*

The above formula written in the following way:

$$\sum_{i=1}^{n} R_{ij} R_{ik} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}.$$

The columns *of a rotation matrix* are orthonormal, *their norm (length) is 1 and they are mutually orthogonal. This can be seen for the case of our example matrix:*

$$\begin{aligned} \sqrt{\cos^2\alpha + \sin^2\alpha} &= 1, \\ \cos\alpha \cdot \sin\alpha + (-\sin\alpha) \cdot \cos\alpha &= 0. \end{aligned}$$

Often we encounter other orthogonal matrices:

1. The mirror matrix for an axis, e.g.:

$$M_2 \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

   which inverts the direction, or algebraic sign, of the y co-ordinate.

2. The axes interchange matrix (permutation):

$$P_{12} \equiv \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3. Inversion *of all axes*:

$$X = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Both M and P differ from rotation matrices in this way, that their *determinant* is -1, when for rotation matrices it is $+1$. The determinant of the X matrix is $(-1)^n$, with n the number of dimensions (i.e., 3 in the above case). A determinant of -1 means that the transformation changes a right handed co-ordinate frame into a left handed one, and conversely.

If we multiply, e.g., M_2 and P_{12}, we obtain

$$M_2 P_{12} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The determinant of this is $+1$. However, it is again a rotation matrix:

$$R_3(90^\circ) = \begin{bmatrix} \cos 90^\circ & \sin 90^\circ & 0 \\ -\sin 90^\circ & \cos 90^\circ & 0 \\ 0 & 0 & 1 \end{bmatrix}!$$

All orthogonal transformations having positive determinants are rotations.

(Without proof still, that *all* orthogonal transformations can be written as either a rotation around a certain axis, or a mirroring through a certain plane.)

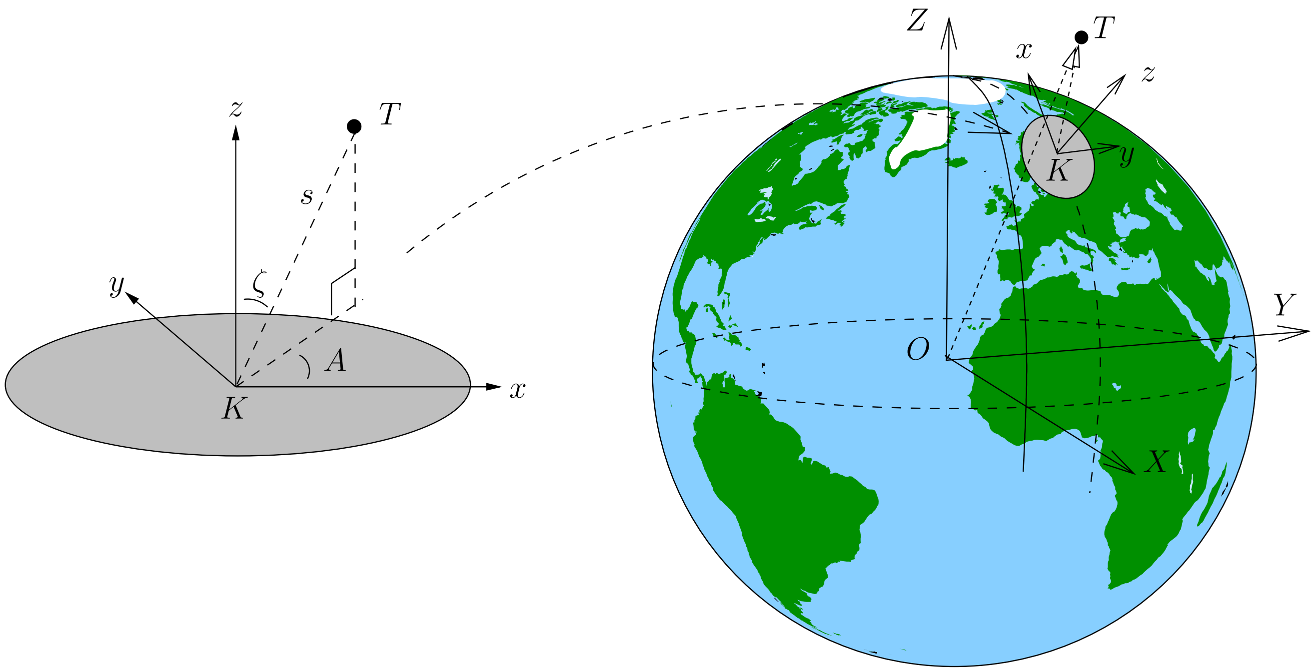


Figure 5.3.: Local astronomical co-ordinates

5.4. Topocentric systems

Also "local astronomical". Note that, whereas the geocentric system is "unique", i.e., there is only one of a certain type, there are as many local systems as there are points on Earth, i.e., an infinity of them.

The system's axes:

1. The z axis points to the local zenith, straight up.
2. The x axis points to the local North.
3. The y axis is perpendicular to both others and points East.

In Figure 5.3 the situation of the local topocentric system in the global context is depicted.

The spherical co-ordinates of the system are:

- The azimuth A
- The zenith angle ζ, alternatively the elevation angle () $\eta = 100\,\text{gon} \; - \zeta$.
- Distance s

The transformation between a point P's topocentric spherical co-ordinates and rectangular co-ordinates is:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_T = \begin{bmatrix} s\cos A \sin\zeta \\ s\sin A \sin\zeta \\ s\cos\zeta \end{bmatrix}_T .$$

The inverse transformation:

$$\begin{aligned} \zeta &= \arctan\frac{\sqrt{x^2+y^2}}{z}, \\ A &= 2\arctan\frac{y}{x+\sqrt{x^2+y^2}}. \end{aligned}$$

The latter formula is known as the *half-angle formula* and avoids the problem of finding the correct quadrant for A. The result is in the interval $(-180°, 180°]$ and negative values may be incremented by $360°$ to make them positive.

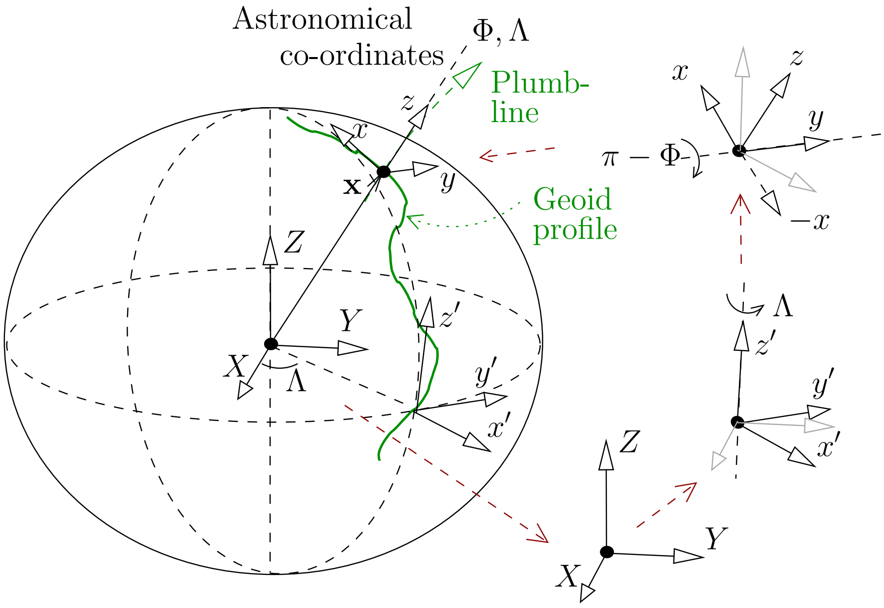


Figure 5.4.: From the geocentric to the topocentric system. The matrix R_1 mentioned in the text was left out here

5.5. From geocentric to topocentric and back

Let (X, Y, Z) be a geocentric co-ordinate system and (x, y, z) a topocentric *instrument co-ordinate system* (i.e., the x axis points to the zero direction of the instrument instead of North; in the case of a theodolite, the zero direction on the horizontal circle.)

In this case we can symbolically write:

$$\mathbf{x} = R_1 R_2 R_3 (\mathbf{X} - \mathbf{X}_0) ,$$

where the rotation matrices R_3, R_2, R_1 act in succession to transform $\mathbf{X}$ into $\mathbf{x}$. See figure 5.4. $\mathbf{X}_0$ denotes the co-ordinates of the local origin in the geocentric system.

The inverse transformation chain of this is

$$\mathbf{X} = \mathbf{X}_0 + R_3^T R_2^T R_1^T \mathbf{x},$$

as can be easily dervied by multiplying the first equation *from the left* by the matrix $R_1^T = R_1^{-1}$, then by the matrix R_2^T, and then by the matrix R_3^T, and finally by moving X_0 to the other side.

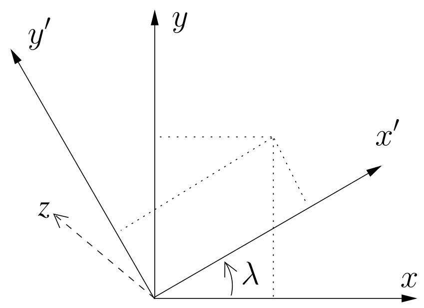
R_3 rotates the co-ordinate frame around the z axis from the Greenwich meridian to the local meridian of the observation site, rotation angle Λ:

$$R_3 = \begin{bmatrix} \cos\Lambda & +\sin\Lambda & 0 \\ -\sin\Lambda & \cos\Lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{5.2}$$

Seen from the direction of the z axis we see (figure), that

$$\begin{aligned} x' &= x\cos\Lambda + y\sin\Lambda, \\ y' &= -x\sin\Lambda + y\cos\Lambda. \end{aligned}$$

(The correct algebraic signs should always be established by the aid of a sketch! Also the directional conventions of different countries may differ.)



R_2 turns the co-ordinate frame around the y axis, in such a way that the z axis points to the North celestial pole instead of to the zenoth. The rotation angle needed for this is $\Phi - 90°$.

$$R_2 = \begin{bmatrix} \sin\Phi & 0 & -\cos\Phi \\ 0 & 1 & 0 \\ +\cos\Phi & 0 & \sin\Phi \end{bmatrix}. \tag{5.3}$$

R_1 rotates the co-ordinate frame around the new z axis or vertical axis by the amount A_0, after which the x axis points to the azimuth of the zero point of the instrument's horizontal circle:

$$R_1 = \begin{bmatrix} \cos A_0 & +\sin A_0 & 0 \\ -\sin A_0 & \cos A_0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

5.6. The geodetic main and inverse problems with rotation matrices

It is possible to solve the geodetic main and inverse problems three-dimensionally, without using the surface geometry of the ellipsoid of revolution.

The idea is based on that the three-dimensional co-ordinate of a point or points are given, e.g., in the form (φ, λ, h) relative to some reference ellipsoid; and that is given or to be computed the azimuth, elevation angle and distance of a second point as seen from the first point. In the forward problem one should compute the secont point's ellipsoidal co-ordinates (φ, λ, h).

5.6.1. Geodetic main problem

Given the ellipsoidal co-ordinates $(\varphi_A, \lambda_A, h_A)$ of point A and in point A, the azimuth A_{AB}, of another point B, the distance s_{AB}, and either the elevation η_{AB} or the zenith angle $z_{AB} \equiv 90° - \eta_{AB}$.

Now we have to compute the co-ordinates $(\varphi_B, \lambda_B, h_B)$ of point B.

As follows:

1. Transform the local A-topocentric co-ordinates of B, (A_{AB}, s_{AB}, z_{AB}) to rectangular:

$$\mathbf{x}_{AB} \equiv \begin{bmatrix} x_{AB} \\ y_{AB} \\ z_{AB} \end{bmatrix} = s_{AB} \begin{bmatrix} \cos A_{AB} \sin z_{AB} \\ \sin A_{AB} \sin z_{AB} \\ \cos z_{AB} \end{bmatrix}.$$

2. Transform, using rotation matrices, these rectangular co-ordinate differences into geocentric[2]:

$$\mathbf{X}_{AB} = R_3^T R_2^T \mathbf{x}_{AB},$$

in which R_3 and R_2 are already given, equations (5.2) ja (5.3).

[2] More precisely, into co-ordinate differences in the geocentric *orientation* – as in this case the origin is not the Earth's centre of mass!

3. Add to the result the geocentric co-ordinates of point A:

$$\mathbf{X}_B = \mathbf{X}_A + \mathbf{X}_{AB},$$

jossa

$$\mathbf{X}_A = \begin{bmatrix} (N(\varphi_A) + h_A)\cos\varphi_A \cos\lambda_A \\ (N(\varphi_A) + h_A)\cos\varphi_A \sin\lambda_A \\ \left(N(\varphi_A)\left(1-e^2\right) + h_A\right)\sin\varphi_A \end{bmatrix}.$$

4. Transform the geocentric co-ordinates of B obtained back to ellipsoidal form (3.7) in the way depicted:

$$\mathbf{X}_B \rightarrow (\varphi_B, \lambda_B, h_B).$$

5.6.2. Geodetic inverse problem

Given the ellipsoidal co-ordinates of two points $(\varphi_A, \lambda_A, h_A)$ and $(\varphi_B, \lambda_B, h_B)$. To be computed the topocentric spherical co-ordinates of point B, A_{AB}, zenith angle z_{AB} ja etäisyys s_{AB}.

1. Transform the ellipsoidal co-ordinates of A and B into geocentric co-ordinates: $\mathbf{X}_A$, $\mathbf{X}_B$.
2. Compute the relative **vector** between the points

$$\mathbf{X}_{AB} = \mathbf{X}_B - \mathbf{X}_A.$$

3. In point A, transform this vector into the topocentric rectangular system

$$\mathbf{x}_{AB} = R_2 R_3 \mathbf{X}_{AB};$$

4. transform to spherical co-ordinates by translating the formula

$$\mathbf{x}_{AB} = s_{AB} \begin{bmatrix} \cos A_{AB} \sin z_{AB} \\ \sin A_{AB} \sin z_{AB} \\ \cos z_{AB} \end{bmatrix},$$

with the familiar arc tangent and Pythagoras formulas (and using the half-angle formula to avoid quadrant problems):

$$\begin{aligned} s_{AB} &= \sqrt{x_{AB}^2 + y_{AB}^2 + z_{AB}^2}, \\ \tan A_{AB} &= \frac{y_{AB}}{x_{AB}} = 2\arctan \frac{y_{AB}}{x_{AB} + \sqrt{x_{AB}^2 + y_{AB}^2}}, \\ \tan z_{AB} &= \frac{\sqrt{x_{AB}^2 + y_{AB}^2}}{z_{AB}}. \end{aligned}$$

5.6.3. Comparison with ellipsoidal surface solution

The solution thus obtained is, concerning the azimuths, very close to the one obtained by using the geodesic between the projections of points A and B on the surface of the ellipsoid. *However, not identical.* The azimuths are so-called *normal plane azimuths*, which differ by a fraction of an arc second from the geodesic's azimuths even on a distance of 100 km. A very small difference, but not zero!

The distance is of course the straight distance in space, not the length of the geodesic. This distance is substantial.

Chapter 6

Co-ordinate systems and transformations

6.1. Geocentric terrestrial systems

Generally geocentric systems, like WGS84, are defined in the following way:

1. The origin of the co-ordinate system coincides with the centre of mass of the Earth.
2. The Z axis of the co-ordinate system points in the direction of the Earth's rotation axis, i.e., the direction of the North pole.
3. The X axis of the co-ordinate system is parallel to the Greenwich meridian.

Such "rotating along" systems are called *terrestrial*. Also ECEF (Earth Centred, Earth Fixed).

6.2. Conventional Terrestrial System

To this we must however make the following *further restrictions*:

1. As the direction of the Z axis we use the so-called CIO, i.e., the Conventional International Origin, the average place of the pole over the years 1900-1905.

   The instantaneous or true pole circles around the CIO in a quasi-periodic fashion: the polar motion. The main periods are a year and approx. 435 days (the so-called "CHANDLER wobble"), the amplitude being a few tenths of a second of arc — corresponding on the Earth's surface to a few metres.
2. Nowadays the zero meridian plane no is longer based on observation by the Greenwichin observatory, but on worldwide VLBI observations. These are co-ordinated by the International Earth Rotation Service (IERS). So, it is no longer *precisely* the meridian of the Greenwich observatory.

In this way we obtain a system co-rotating with the solid Earth, i.e., an ECEF (Earth Centered, Earth Fixed) reference system, e.g., WGS84 or ITRSxx (ITRS = International Terrestrial Reference System, xx year number; created by the IERS). Another name used is Conventional Terrestrial System (CTS).

6.3. Polar motion

The direction of the Earth's rotation axis is slightly varying over time. This *polar motion* has two components called x_P and y_P, the offset of the instantaneous pole from the CIO pole in the direction of Greenwich, and perpendicular to it in the West direction, respectively. The transformation between the instantaneous and conventional terrestrial references is done as follows:

$$\mathbf{X}_{IT} = R_Y\left(x_p\right) R_X\left(y_p\right) \mathbf{X}_{CT}.$$

Here, note that the matrix R_Y denotes a rotation by an amount x_P *about the Y axis*, i.e., the Y axis stays fixed, while the X and Z co-ordinates change. Similarly, the matrix R_X denotes a rotation y_p about the X axis, which changes only the Y and Z co-ordinates. The matrices are

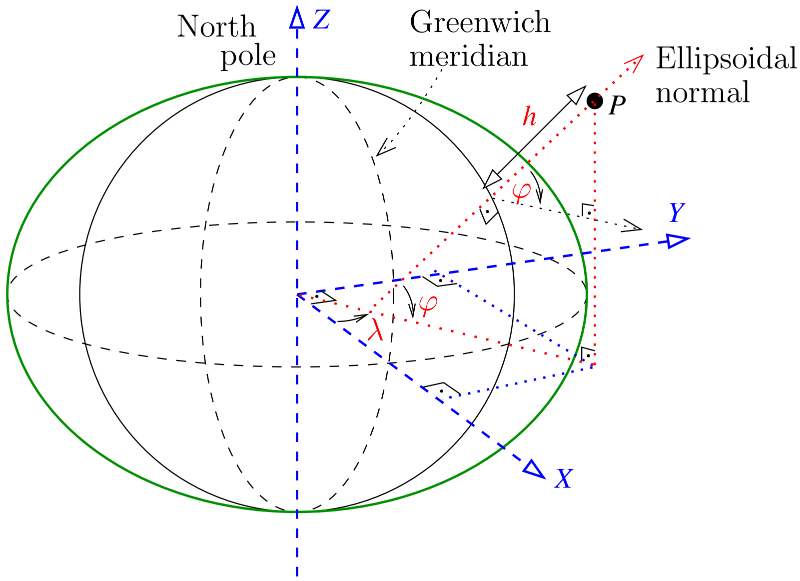


Figure 6.1.: The ellipsoid and geocentric co-ordinates

$$R_Y(x_p) = \begin{bmatrix} \cos x_P & 0 & -\sin x_P \\ 0 & 1 & 0 \\ \sin x_P & 0 & \cos x_P \end{bmatrix} \quad \text{and} \quad R_X(y_p) \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos y_P & \sin y_P \\ 0 & -\sin y_P & \cos y_P \end{bmatrix}.$$

Because the angles x_p and y_p are very small, order of magnitude second of arc, we may approximate $\sin x_p \approx x_p$ and $\cos x_p \approx 1$ (same for y_p), as well as $x_P y_P \approx 0$, obtaining

$$R_Y(x_p) R_X(y_p) \approx \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & 0 \\ x_p & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & y_p \\ 0 & -y_p & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & y_p \\ x_p & -y_p & 1 \end{bmatrix}.$$

6.4. The Instantaneous Terrestial System

If we take, instead of the conventional pole, the instantaneous pole, i.e., the direction of the Earth's rotation axis, we obtain, instead of the conventional, the so-called *instantaneous terrestrial system* (ITS). This is the system to use with, e.g., astronomical or satellite observations, because it describes the true orientation of the Earth relative to the stars.

The transformation between the conventional and the instantaneous system happens with the aid of *polar motion parameters*: if they are x_P, y_P — x_p points to Greenwich and y_p to the East from the Greenwich meridian — we obtain (see above):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{IT} = R_Y(x_p) R_X(y_p) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CT} \approx \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & y_p \\ x_p & -y_p & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CT},$$

because the angles x_p, y_p are so extremely small.

6.5. The quasi-inertial geocentric system

The quasi-inertial, also celestial, or real astronomical (RA) reference frame, drawn in blue in figure 4.1, is a *geocentric* system, like the conventional terrestrial system. It is however celestial in nature and the positions of stars are approximately constant in it. It is defined as follows:

1. The origin of the co-ordinate system again coincides with the Earth's centre of mass.
2. The Z axis of the co-ordinate system is again oriented in the direction of the Earth's rotation axis, i.e., the North Pole.
3. But*:* The X *axis is pointing to the* vernal equinox*.*

Such a reference system doesn't rotate along with the solid Earth. It is (to good approximation) *inertial.* We also refer to it as an *equatorial* co-ordinate system. In this system the direction co-ordinates are the astronomical right ascention and declination α, δ.

If the "celestial spherical co-ordinates" α, δ are known, we may compute the unit direction vector as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{RA} = \begin{bmatrix} \cos\delta\cos\alpha \\ \cos\delta\sin\alpha \\ \sin\delta \end{bmatrix}.$$

Here we must consider, however, that, while δ is given in degrees, minutes and seconds, α is given in *time units.* They must first be converted to degrees etc. One hour corresponds to 15 degrees, one minute to 15 minutes of arc, and one second to 15 seconds of arc.

Going from rectangular to spherical again requires the following formulae (note the use of the half-angle formula for α, which is precise over the whole range and avoids the problem of identifying the right quadrant):

$$\begin{aligned} \delta &= \arcsin Z \\ \alpha &= 2\arctan\frac{Y}{X + \sqrt{X^2 + Y^2}}. \end{aligned}$$

Again, negative values for α can be made positive by adding 24^{h}.

The co-ordinates, or *places*, of stars are *apparent*, a technical term meaning "as they appear at a certain point in time [1]". The places α, δ read from a celestial chart are *not* apparent. They refer to a certain point in time, e.g., 1950.0 or 2000.0. Obtaining apparent places requires a long reduction chain, taking into account precession, nutation, the variations in Earth rotation, and also the annual parallax and the possible proper motion of the star.

The apparent places of stars are found from the reference work "Apparent Places of Fundamental Stars" precalculated and tabulated according to date.

We can transform between RA and IT (instantaneous terrestrial) co-ordinates as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{RA} = R_Z(-\theta_0)\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{IT} = \begin{bmatrix} \cos(\theta_0) & -\sin(\theta_0) & 0 \\ \sin(\theta_0) & \cos(\theta_0) & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{IT}.$$

here, θ_0 on Greenwich Apparent Sidereal Time, or GAST.

[1] ... from the centre of the Earth. The fixed stars are so far way, however, that the location of the observer doesn't matter.

Chapter 7

Reference systems and realizations

7.1. Old and new reference systems; ED50 vs. WGS84/GRS80

In Finland like in many European countries, the traditional reference system is non-geocentric and based on an old reference ellipsoid, the International Ellipsoid computed by John Fillmore Hayford, and adopted by the International Union of Geodesy and Geophysics (IUGG) in 1924. European Datum 1950 (ED50) was created in 1950 by unifying the geodetic networks of the countries of Western Europe, and was computed on the Hayford ellipsoid.

The newer systems, both World Geodetic System (WGS84) and Geodetic Reference System 1980 (GRS80) are designed to be geocentric.

So, the differences can be summarized as:

1. Reference ellipsoid used: International Ellipsoid (Hayford) 1924 vs. GRS80
2. Realized by terrestrial measurements vs. based on satellite (and space geodetic) data
3. Non-geocentric (100 m level) vs. geocentric (cm level).

The figure of a *reference ellipsoid* is unambiguously fixed by two quantities: the semi-major exis or equatorial radius a, and the flattening f.

- International ellipsoid 1924: $a = 6378388\,\mathrm{m}$, $f = 1 : 297$
- GRS80: $a = 6378137\,\mathrm{m}$, $f = 1 : 298.257222101$

The reference ellipsoid of GPS's WGS84 system is in principle the same as GRS80, but due to poor numerics it ended up with

- $f = 1 : 298.257223563$. The net result is that the semi-minor axis (polar radius) of WGS84 is longer by 0.1 mm[1] compared to GRS80.

To complicate matters, as the basis of the ITRS family of co-ordinate systems, and their realizations ITRF, was chosen *the GRS80 reference ellipsoid.* The parameters of this system were already given in Tables 4.1 and 4.2.

7.2. WGS84 and ITRS

Both WGS84 and the International Terrestrial Reference System (ITRS) are *realized* by computing co-ordinates for polyhedra of points (stations) on the Earth's surface. The properties of these systems are:

- *Geocentric*, i.e., the co-ordinate origin and centre of reference ellipsoid is the Earth's centre of mass (and the Earth's mass includes oceans and atmosphere, but not the Moon!). This is realized by using observations to satellites, whose equations of motion are implicitly geocentric.

[1] Computation:

$$\Delta b = -\frac{\Delta f}{f}(a-b) = \frac{\Delta\left(1/f\right)}{\left(1/f\right)}(a-b) = \frac{0.000001462}{298.25722}\cdot(21384\,\mathrm{m}) = 0.000105\,\mathrm{m}.$$

- The *scale* derives from the SI system. This is realized by using range measurements by propagation of electromagnetic waves. The velocity of these waves in vacuum is conventionally fixed to $299\,792\,458\,\mathrm{m\,s^{-1}}$. Thus, range measurement becomes time measurement by atomic clock, which is very precise.
- *Orientation*: originally the Conventional International Origin (CIO) of the Earth axis, i.e., the mean orientation over the years 1900-1905, and the direction of the Greenwich Meridian, i.e., the plumbline of the Greenwich transit circle. Currently, as the orientation is maintained by the International Earth Rotation and Reference Systems Service (IERS) using VLBI and GPS, this is no longer the formal definition; but continuity is maintained.
- The current *definition* uses the BIH (Bureau International de l'Heure) 1984 definition of the conventional pole, and their 1984 definition of the zero meridian plane. Together, X, Y and Z form a right-handed system.

7.3. Co-ordinate system realizations

Internationally, somewhat varying terminology is in use concerning the realization of co-ordinate systems or *datums*.

- ISO: Co-ordinate reference system / co-ordinate system
- IERS: Reference system / reference frame
- Finnish: koordinaattijärjestelmä / koordinaatisto ((Anon., 2008))

The *latter* of the names is used to describe a system that was implemented in the terrain, using actual measurements, producing co-ordinate values for the stations concerned; i.e., a *realization*. Then also, a *datum* was defined, with one or more *datum points* being kept fixed at their conventional values.

The *former* refers to a more abstract definition of a co-ordinate system, involving the choice of reference ellipsoid, origin (Earth center of mass, e.g.) and axes orientation.

7.4. Realization of WGS84

Because "WGS84" is often referred to as the system in which satellite positioning derived co-ordinates are obtained, we shall elaborate a little on how this system has been actually realized over time. Our source is (Kumar and Reilly, 2006). The first version of WGS84 was released in 1987 by the US Defense Mapping Agency, currently the NGA (National Geospatial-Intelligence Agency). After that, it was updated in 1994 (G730), 1996 (G873) and 2001 (G1150).

As you will see, there are a number of problems even with the latest realization of WGS84. For this reason it is better to consider WGS84 as an approximation at best, of the reference frames of the ITRF/ETRF variety. The precision of this approximation is clearly sub-metre, so using WGS84 for metre precision level applications should be OK. See the following note: `ftp://itrf.ensg.ign.fr/pub/itrf/WGS84.TXT`.

If you want more confusion, read (Stevenson, 2008).

7.5. Realizations of ITRS/ETRS systems

All these systems are the responsibility of the international geodetic community, specifically the IERS (International Earth Rotation and Reference Systems Service). "I" stands for International, "E" for European. The "S" stands for "system", meaning the principles for creating a reference frame before actual realization. With every ITRS corresponds a number of ITRF's ("Frames"), which are *realizations*, i.e., co-ordinate solutions for networks of ground stations computed from sets of actual measurements. Same for ETRS/ETRF, which are the corresponding things for the European area, where the effect of the slow motion of the rigid part of the Eurasian tectonic plate has been corrected out in order to obtain approximately constant co-ordinates.

Data used for realizing ITRF/ETRF frames: mostly GPS, but also Very long Baseline Interferometry (VLBI) providing a strong orientation; satellite and lunar laser ranging (SLR, LLR) contributing to the right scale, and the French DORIS satellite system. Nowadays also GLONASS is used.

Currently the following realizations exist for ITRS: ITRF88, 89, 90, 91, 92, 93, 94; 96, 97; ITRF2000, ITRF2005 and ITRF2008.

The *definition* of an ITRFyy is as follows (McCarthy, 1996):

- The mean rotation of the Earth's crust in the reference frame will vanish globally (cf. for ETRF: on the Eurasian plate). Obviously then, co-ordinates of points on the Earth's surface will slowly change due to the motion of the plate that the point is on. Unfortunately at the current level of geodetic precision, it is not possible to define a global co-ordinate frame in which points are fixed.
- The Z-axis corresponds to the IERS Reference pole (IRP) which corresponds to the BIH Conventional terrestrial Pole of 1984, with an uncertainty of 0.005"
- The X-axis, or IERS Reference Meridian, similarly corresponds to the BIH zero meridian of 1984, with the same uncertainty.

Finally, note that the Precise Ephemeris which are computed by IGS (the International GPS Geodynamics Service) and distributed over the Internet, are referred to the current (newest) ITRF, and are computed using these co-ordinates for the tracking stations used.

7.6. The three-dimensional Helmert transformation

The form of the transformation, in the case of small rotation angles, is

$$\begin{bmatrix} X^{(2)} \\ Y^{(2)} \\ Z^{(2)} \end{bmatrix} = (1+m)\begin{bmatrix} 1 & e_z & -e_y \\ -e_z & 1 & e_x \\ e_y & -e_x & 1 \end{bmatrix} \cdot \begin{bmatrix} X^{(1)} \\ Y^{(1)} \\ Z^{(1)} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{7.1}$$

where $\begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$ is the translation vector of the origin, m is the scale factor correction from unity, and (e_x, e_y, e_z) are the (small) rotation angles about the respective axes. Together we thus have *seven* parameters. The superscripts (1) and (2) refer to the old and new systems, respectively.

Eq. (7.1) can be re-written and linearized as follows, using $me_x = me_y = me_z = 0$, and replacing the vector $\begin{bmatrix} X^{(1)} & Y^{(1)} & Z^{(1)} \end{bmatrix}^T$ by approximate values $\begin{bmatrix} X^0 & Y^0 & Z^0 \end{bmatrix}^T$. This is allowed as m and the e angles are all assumed small.

$$\begin{aligned} \begin{bmatrix} X^{(2)} - X^{(1)} \\ Y^{(2)} - Y^{(1)} \\ Z^{(2)} - Z^{(1)} \end{bmatrix} &\approx \begin{bmatrix} m & e_z & -e_y \\ -e_z & m & e_x \\ e_y & -e_x & m \end{bmatrix} \begin{bmatrix} X^{(1)} \\ Y^{(1)} \\ Z^{(1)} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \approx \\ &\approx \begin{bmatrix} m & e_z & -e_y \\ -e_z & m & e_x \\ e_y & -e_x & m \end{bmatrix} \begin{bmatrix} X^0 \\ Y^0 \\ Z^0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}. \end{aligned}$$

An elaborate rearranging yields

$$\begin{bmatrix} X_i^{(2)} - X_i^{(1)} \\ Y_i^{(2)} - Y_i^{(1)} \\ Z_i^{(2)} - Z_i^{(1)} \end{bmatrix} = \left[\begin{array}{c|ccc|ccc} X_i^0 & 0 & -Z_i^0 & +Y_i^0 & 1 & 0 & 0 \\ Y_i^0 & +Z_i^0 & 0 & -X_i^0 & 0 & 1 & 0 \\ Z_i^0 & -Y_i^0 & +X_i^0 & 0 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} m \\ \hline e_x \\ e_y \\ e_z \\ \hline t_x \\ t_y \\ t_z \end{array}\right]. \tag{7.2}$$

Here we have added for generality a point index i, $i = 1, \ldots, n$. The number of points is then n, the number of "observations" (available co-ordinate differences) is $3n$. The full set of these "observation equations" then becomes

$$\left[\begin{array}{c} X_1^{(2)} - X_1^{(1)} \\ Y_1^{(2)} - Y_1^{(1)} \\ Z_1^{(2)} - Z_1^{(1)} \\ \hline \vdots \\ \hline X_i^{(2)} - X_i^{(1)} \\ Y_i^{(2)} - Y_i^{(1)} \\ Z_i^{(2)} - Z_i^{(1)} \\ \hline \vdots \\ \hline X_n^{(2)} - X^{(1)} \\ Y_n^{(2)} - Y_n^{(1)} \\ Z_n^{(2)} - Z_n^{(1)} \end{array}\right] = \left[\begin{array}{c|ccc|ccc} X_1^0 & 0 & -Z_1^0 & +Y_1^0 & 1 & 0 & 0 \\ Y_1^0 & +Z_1^0 & 0 & -X_1^0 & 0 & 1 & 0 \\ Z_1^0 & -Y_1^0 & +X_1^0 & 0 & 0 & 0 & 1 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline X_i^0 & 0 & -Z_i^0 & +Y_i^0 & 1 & 0 & 0 \\ Y_i^0 & +Z_i^0 & 0 & -X_i^0 & 0 & 1 & 0 \\ Z_i^0 & -Y_i^0 & +X_i^0 & 0 & 0 & 0 & 1 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline X_n^0 & 0 & -Z_n^0 & +Y_n^0 & 1 & 0 & 0 \\ Y_n^0 & +Z_i^0 & 0 & -X_n^0 & 0 & 1 & 0 \\ Z_n^0 & -Y_n^0 & +X_n^0 & 0 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} m \\ \hline e_x \\ e_y \\ e_z \\ \hline t_x \\ t_y \\ t_z \end{array}\right]. \tag{7.3}$$

This is a set of observation equations of form $\underline{\ell} + \underline{v} = A\widehat{x}$ (but without the residuals vector $\underline{v}$ which is needed to make the equality true in the presence of observational uncertainty). There are seven unknowns $\widehat{x}$ on the right. They can be solved in the least-squares sense if we have co-ordinates (X, Y, Z) in both the old (1) and the new (2) system for at least three points, i.e., nine "observations" in the observation vector $\underline{\ell}$. In fact, two points and one co-ordinate from a third point would suffice. However, it is always good to have redundancy.

7.7. Transformations between ITRF realizations

For transformation parameters between the various ITRF realizations, see the IERS web page: `http://itrf.ensg.ign.fr/trans_para.php`. As an example, the transformation parameters *from* ITRF2008 *to* ITRF2005, at epoch 2005.0, `http://itrf.ensg.ign.fr/ITRF_solutions/2008/tp_08-05.php`:

	$T1$ mm	$T2$ mm	$T3$ mm	D ppb	$R1$ 0.001"	$R2$ 0.001"	$R3$ 0.001"
	-0.5	-0.9	-4.7	0.94	0.000	0.000	0.000
±	0.2	0.2	0.2	0.03	0.008	0.008	0.008
Rate	0.3	0.0	0.0	0.00	0.000	0.000	0.000
±	0.2	0.2	0.2	0.03	0.008	0.008	0.008

These parameters[2] are to be used as follows:

$$
\begin{aligned}
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2005} (t) &= \left\{ 1 + \begin{bmatrix} D & -R_3 & R_2 \\ R_3 & D & -R_1 \\ -R_2 & R_1 & D \end{bmatrix} \right\} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2008} (t) + \\
&+ (t - t_0) \frac{d}{dt} \begin{bmatrix} D & -R_3 & R_2 \\ R_3 & D & -R_1 \\ -R_2 & R_1 & D \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2008} (t) + \\
&+ \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} + (t - t_0) \frac{d}{dt} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \\
&= \left\{ 1 + \begin{bmatrix} 0.94 & 0 & 0 \\ 0 & 0.94 & 0 \\ -0 & 0 & 0.94 \end{bmatrix} \cdot 10^{-9} \right\} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2008} (t) + \\
&+ \begin{bmatrix} -0.5 + 0.3\,(t - 2005.0) \\ -0.9 \\ -4.7 \end{bmatrix}
\end{aligned}
$$

with the numbers given, and forgetting about the uncertainties. Here $\frac{d}{dt}$ refers to the rates, of which only that of T_1 is non-vanishing in this example.

[2]Note the change in parameter names compared to the previous. For (t_x, t_y, t_z) we now have $(T1, T2, T3)$; m is now called D; and (e_x, e_y, e_z) became $(R1, R2, R3)$.

Chapter 8

Celestial co-ordinate systems

8.1. Sidereal time

The transformation from the inertial system to the terrestrial one goes through *sidereal time.*

- Greenwich Apparent Sidereal Time, GAST, symbol θ_0
- The apparent sidereal time at the observation location, LAST, symbol θ.

$$\theta = \theta_0 + \Lambda^{\mathrm{hms}},$$

in which Λ is the *astronomical longitude* of the place of observation, converted to suitable time units.

GAST is

- the transformation angle between the inertial and the terrestrial ("co-rotating") systems, i.e.
- the angle describing the Earth's orientation in the inertial system, i.e.
- the difference in longitude between the Greenwich meridian and the vernal equinox.

Also the Greenwich apparent sidereal time is tabulated – afterwards, when the precise Earth rotation is known. GAST can be computed to one second precision based on the calendar and civil time.

If a few minutes of precision suffices, we may even tabulate GAST as a function of day of the year only. The table for the annual part of sidereal time per month is:

Month	θ_m	**Month**	θ_m	**Month**	θ_m	**Month**	θ_m
January	6 40	April	12 40	July	18 40	October	0 40
February	8 40	May	14 40	August	20 40	November	2 40
March	10 40	June	16 40	September	22 40	December	4 40

In constructing this table, the following knowledge was used: on March 21 at 12 UTC in Greenwich, the hour angle of the Sun, i.e., of the vernal equinox, i.e., sidereal time, is 0^{h}. This Greenwich sidereal time θ_0 consists of *two parts*: an annual part θ_a, and a clock time τ (UTC or Greenwich Mean Time). So we obtain the annual part of sidereal time by subtracting: $\theta_a = \theta_0 - \tau = -12^{\mathrm{h}}$, i.e., 12^{h} after adding a full turn, 24^{h}.

If on March 21, or more precisely at midnight after March 20, sidereal time is $12^{\mathrm{h}}\,00^{\mathrm{m}}$, then sidereal time for March 0 is $12^{\mathrm{h}}\,00^{\mathrm{m}} - 4 \times 20^{\mathrm{m}} = 10^{\mathrm{h}}\,40^{\mathrm{m}}$; Remember that one day corresponds to about four minutes.

We may fill out the table using the rule that $1\,\mathrm{month} \ \approx 2^{\mathrm{t}}$. (In principle we could slightly improve the table's accuracy by taking into account the varying lengths of the months. However, the cycle of leap years causes error of similar magnitude.)

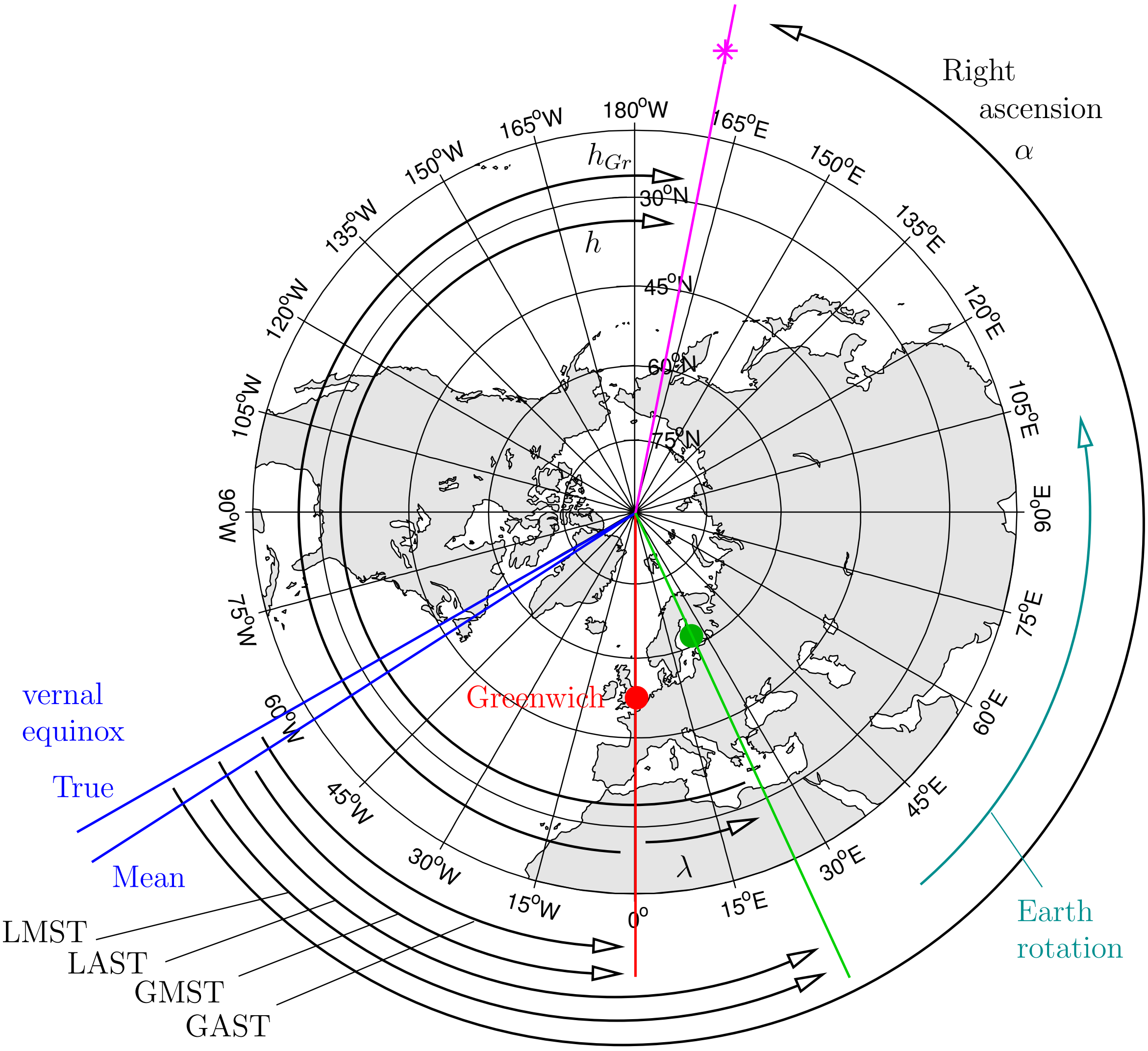


Figure 8.1.: Sidereal time

After this, local sidereal time is obtained as follows:

$$\begin{aligned}\theta &= \theta_{\mathrm{m}} + \theta_{\mathrm{d}} + \tau + \Lambda^{\mathrm{hms}} \\ &= \theta_{\mathrm{a}} + \tau + \Lambda^{\mathrm{hms}} = \\ &= \theta_0 + \Lambda^{\mathrm{hms}}\end{aligned}$$

where

θ_{m} the value taken from the above table

θ_{d} four times the day number within a month

$\theta_{\mathrm{a}} = \theta_{\mathrm{m}} + \theta_{\mathrm{d}}$ annual part of sidereal time

τ time (UTC)

Λ the longitude of the Earth station converted to hours, minute and seconds ($15° = 1^{\mathrm{h}}$, $1° = 4^{\mathrm{m}}$, $1' = 4^{\mathrm{s}}$).

See the pretty figure 8.1. We have the following quantities:

- $GAST$ = Greenwich Apparent Sidereal Time ($= \theta_0$)

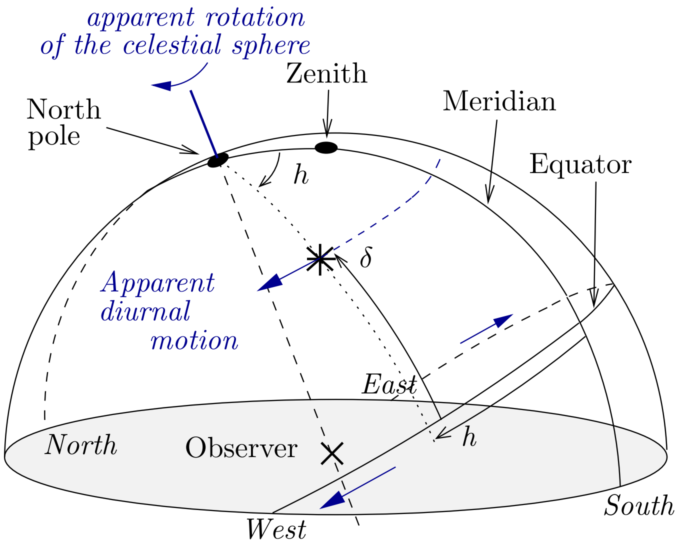


Figure 8.2.: Hour angle and other co-ordinates on the celestial sphere

- $LMST$ = Local Mean Sidereal Time $(= \overline{\theta}\ )$
- the equinox varies irregularly with time due to precession and nutation. That's why we distinguish Mean and Apparent. The difference is called the *equation of equinoxes, ee.*
- h is the hour angle
- h_{Gr} is the Greenwich hour angle
- α is the right ascension (of a celestial object)
- Λ is the longitude (of a terrestrial object).

$$\begin{aligned} h &= \theta - \alpha, \\ h_{Gr} &= \theta_0 - \alpha, \\ \theta &= \theta_0 + \Lambda, \\ \overline{\theta} &= \overline{\theta}_0 + \Lambda; \\ \theta - \overline{\theta} &= \theta_0 - \overline{\theta}_0 = ee. \end{aligned}$$

8.2. Trigonometry on the celestial sphere

On the celestial sphere we have at least two different kinds of co-ordinates: local and equatorial.

Local spherical co-ordinates are related to the local astronomical rectangular system as follows (x to the North, z to the zenith):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos A \sin\zeta \\ \sin A \sin\zeta \\ \cos\zeta \end{bmatrix} = \begin{bmatrix} \cos A \cos\eta \\ \sin A \cos\eta \\ \sin\eta \end{bmatrix},$$

where A is the azimut (from the North clockwise), ζ is the zenith angle and η is the height or elevation angle.

Equatorial co-ordinates are α, δ, right ascension and declination; their advantage is, that the co-ordinates ("places") of stars are nearly constant. The disadvantage is that there is no simple relationship to local astronomical co-ordinates.

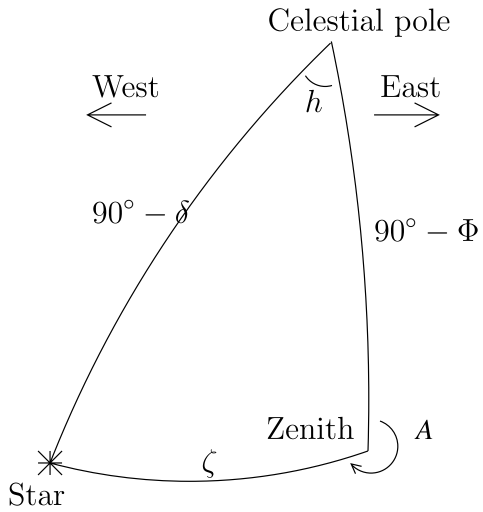


Figure 8.3.: Fundamental triangle of astronomy

An *intermediate form* of co-ordinates is: *hour angle* and declination, h, δ. The hour angle is defined as shown in figure 8.3, the angular distance around the Earth's rotation axis (or at the celestial pole) from the meridian of the location.

Equation:

$$h = \theta_0 + \Lambda - \alpha,$$

where θ_0 is Greenwich apparent sidereal time (GAST), Λ is the local (astronomical) longitude, and α is the right ascension of the star.

If the star is in the meridian, we have $h = 0$ and $\alpha = \theta_0 + \Lambda$. On this is based the use of a transit instrument: if, of the three quantities θ_0, α or Λ, two are knownthe third may be computed. According to the application we speak of astronomical position determination, time keeping of determination of the places of stars. "*One man's noise is another man's signal*".

On the celestial sphere there is a *fundamental triangle of astronomy*: it consists of the star, the celestial North pole, and the zenith. Of the angles of the triangle we mention t (North pole) and A (the zenith), of its sides, $90° - \Phi$ (pole-zenith), $90° - \delta$ (star-pole) and ζ (star-zenith).

The sine rule:

$$\frac{-\sin A}{\cos\delta} = \frac{\sin h}{\sin\zeta}.$$

The cosine rule:

$$\begin{aligned}\cos\zeta &= \sin\delta\sin\Phi + \cos\delta\cos\Phi\cos h,\\ \sin\delta &= \sin\Phi\cos\zeta + \cos\Phi\sin\zeta\cos A.\end{aligned}$$

We compute first, using the cosine rule, either δ or ζ, and then, using the sine rule, either h or A. Thus we obtain either $(A, \zeta) \leftrightarrow (h, \delta)$ in both directions.

8.3. Using rotation matrices

The various transformations between celestial co-ordinate systems can be derived also in rectangular co-ordinates by using rotation matrices.

Let the topocentric co-ordinate vector (length 1) be $\mathbf{r} = \begin{bmatrix} x & y & z \end{bmatrix}^T$. In spherical co-ordinates this is

$$\mathbf{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos A \sin \zeta \\ \sin A \sin \zeta \\ \cos \zeta \end{bmatrix}.$$

The same vector we may also write in local astronomical co-ordinates:

$$\mathbf{r}' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos \alpha \cos \delta \\ \sin \alpha \cos \delta \\ \sin \delta \end{bmatrix}.$$

The transformation between them is:

1. the direction of the x axis is changed from North to South
2. the new xz axis pair is rotated by an amount $90° - \Phi$ to the South, Φ being the astronomical latitude.
3. the new xy axis pair is rotated to the West (clockwise) from the local meridian to the vernal equinox by an amount θ, the local (apparent) sidereal time.

The matrices are:

$$\begin{aligned}
\mathrm{M}_1 &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\
\mathrm{R}_2 &= \begin{bmatrix} \cos(90° - \Phi) & 0 & \sin(90° - \Phi) \\ 0 & 1 & 0 \\ -\sin(90° - \Phi) & 0 & \cos(90° - \Phi) \end{bmatrix} = \begin{bmatrix} \sin \Phi & 0 & -\cos \Phi \\ 0 & 1 & 0 \\ \cos \Phi & 0 & \sin \Phi \end{bmatrix}, \\
\mathrm{R}_3 &= \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\end{aligned}$$

Let us compute

$$\begin{aligned}
\mathrm{R}_3 \mathrm{R}_2 \mathrm{M}_1 &= \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\sin \Phi & 0 & \cos \Phi \\ 0 & 1 & 0 \\ \cos \Phi & 0 & \sin \Phi \end{bmatrix} = \\
&= \begin{bmatrix} -\cos \theta \sin \Phi & -\sin \theta & \cos \theta \cos \Phi \\ -\sin \theta \sin \Phi & \cos \theta & \sin \theta \cos \Phi \\ \cos \Phi & 0 & \sin \Phi \end{bmatrix}.
\end{aligned}$$

After this:

$$\begin{bmatrix} \cos \alpha \cos \delta \\ \sin \alpha \cos \delta \\ \sin \delta \end{bmatrix} = \begin{bmatrix} -\cos \theta \sin \Phi & -\sin \theta & \cos \theta \cos \Phi \\ -\sin \theta \sin \Phi & \cos \theta & \sin \theta \cos \Phi \\ \cos \Phi & 0 & \sin \Phi \end{bmatrix} \begin{bmatrix} \cos A \sin \zeta \\ \sin A \sin \zeta \\ \cos \zeta \end{bmatrix},$$

where we identify immediately

$$\sin \delta = \sin \Phi \cos \zeta + \cos \Phi \sin \zeta \cos A,$$

the cosine rule in the triangle star-celestial pole-zenith.

The inverse transformation is, based on orthogonality (the transpose!)

$$\begin{bmatrix} \cos A \sin \zeta \\ \sin A \sin \zeta \\ \cos \zeta \end{bmatrix} = \begin{bmatrix} -\cos \theta \sin \Phi & -\sin \theta \sin \Phi & \cos \Phi \\ -\sin \theta & \cos \theta & 0 \\ \cos \theta \cos \Phi & \sin \theta \cos \Phi & \sin \Phi \end{bmatrix} \begin{bmatrix} \cos \alpha \cos \delta \\ \sin \alpha \cos \delta \\ \sin \delta \end{bmatrix}.$$

With these, we can do the transformation of spherical co-ordinates with the aid of three-dimensional "direction cosines".

8.4. On satellite orbits

We describe here the computation of circular satellite orbits around a spherical Earth; this is sufficient at least for enabling visual satellite observations and finding the satellites.

There are thousands of satellites orbiting the Earth, which for the most part are very small. A few hundred however are so large, generally last stages of launcher rockets, that they can be seen after dark in the light of the Sun even with the naked eye. With binoculars these can be observed easily. The heights of their orbits vary from 400 km to over 1000 km; the inclination angle of the orbital plane may vary a lot, but certain inclination values, like $56°$, $65°$, $72°$, $74°$, $81°, 90°$ and $98°$ are especially popular.

In a class of their own are the *Iridium satellites*, which have stayed in orbit after an ill-fated mobile telephone project. Every Iridium satellite has a long metal antenna which reflects sunlight in a suitable orientation extremely brightly. Predictions for the Iridium satellites are found from the World Wide Web.

When we know the satellite's equator crossing, i.e., the time of the "ascending node" of the orbit, t_0, and the right ascension ("inertial "longitude") α_0, we can compute the corrections to time and longitude for various latitudes like we describe in the sequel.

8.5. Crossing a given latitude in the inertial system

If the target latitude φ is given, we can compute the distance ν from the ascending node ("downrange-angle") in angular units as follows:

$$\sin\nu = \frac{\sin\phi}{\sin i},$$

where i is the *inclination* of the satellite orbit. From this follows again the elapsed time using Kepler's third law; the period or time of completing one orbit is:

$$P = \sqrt{\frac{4\pi^2}{GM}a^3}.$$

From this

$$\Delta\tau = \frac{\nu}{2\pi}P,$$

the flight time from the equator to latitude ϕ.

The azimut angle between satellite orbit and local meridian is obtained from the Clairaut formula (luku 2.3):

$$\cos\phi\sin A = \cos(0)\cos i,$$

because at the equator $(\phi = 0)$ $A = \frac{\pi}{2} - i$, i.e.

$$\sin A = \frac{\cos i}{\cos\phi}.$$

Now we obtain the difference in right ascension with the equator crossing using the sine rule for a spherical triangle:

$$\frac{\sin\Delta\alpha}{\sin A} = \frac{\sin\phi}{\sin i} \Rightarrow \sin\Delta\alpha = \frac{\tan\phi}{\tan i}.$$

After this, we obtain the satellite's right ascension and time when crossing latitude φ:

$$\begin{aligned}\tau &= \tau_0 + \Delta\tau,\\ \alpha &= \Omega + \Delta\alpha.\end{aligned}$$

Here, Ω is the *right ascension of the ascending node* of the satellite orbit.

Both longitude and right ascension α (ja $\Delta\alpha$) are reckoned positive to the East.

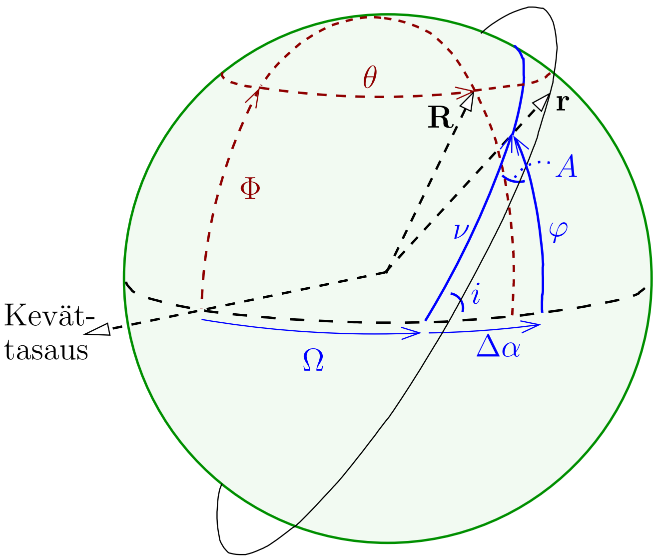


Figure 8.4.: The places of satellite and Earth station in the inertial system

8.6. The satellite's topocentric co-ordinates

Generally, the satellite orbit's *height* h is given; according to its definition, the height of the satellite from the geocentre is

$$r = \|\mathbf{r}\| = a_e + h,$$

where a_e is the equatorial radius of the Earth. In spherical approximation $a_e = R$. Then the rectangular, inertial co-ordinates of the satellite are

$$\mathbf{r} = \begin{bmatrix} r\cos\phi\cos\alpha_G \\ r\cos\phi\sin\alpha_G \\ r\sin\phi \end{bmatrix},$$

where ϕ is the satellite's geocentric latitude (or, if one wants to put it this way, the "geocentric declination" δ_G) and α_G the geocentric right ascension.

Let the geocentric latitude of the ground station be Φ and its longitude Λ; then, at moment t, the *geocentric right ascension of the ground station* is

$$\theta = \Lambda + \tau + \theta_a,$$

where τ is the time (UTC) and θ_a the annual part of sidereal time. θ is the same as the local sidereal time, which thus represents the *orientation of the local meridian* in (inertial) space.

Now the Earth station's rectangular, inertial co-ordinates are

$$\mathbf{R} = \begin{bmatrix} R\cos\Phi\cos\theta \\ R\cos\Phi\sin\theta \\ R\sin\Phi \end{bmatrix}.$$

Subtraction yields:

$$\mathbf{d} = \mathbf{r} - \mathbf{R} \equiv \begin{bmatrix} d\cos\delta_T\cos\alpha_T \\ d\cos\delta_T\sin\alpha_T \\ d\sin\delta_T \end{bmatrix},$$

from which we may solve the *topocentric co-ordinates*:

$$\tan\delta_T = \frac{d_3}{\sqrt{d_1^2 + d_2^2}},\ \tan\alpha_T = \frac{d_2}{d_1}$$

These are thus the satellite's right ascension and declination as seen against the starry sky. With the aid of a star chart and binoculars we may now wait for and observe the satellite.

We also obtain the satellite's *distance*:

$$d = \sqrt{d_1^2 + d_2^2 + d_3^2}.$$

Using the distance we may compute the satellite's visual brightness of *magnitude.* The distances are generally of order 500-1000 km and the magnitudes 2-5.

8.7. Crossing a given latitude in the terrestrial system

If we do the computation in the terrestrial system, we must have been given the *longitude* of the equator crossing λ_0 ($\lambda_0 = \Omega - \theta_0(\tau_0)$, where θ_0 is GAST at the time τ_0 of the equator crossing.)

The time difference $\Delta\tau$ from the equator crossing to latitude Φ is obtained in the same way; however, now we compute the *longitude difference* as follows:

$$\Delta\lambda = \Delta\alpha - \Delta\tau \cdot \omega,$$

where ω is the *rotation rate of the Earth,* some $0.25°$ per minute. Thus we obtain the satellite's longitude:

$$\lambda = \lambda_0 + \Delta\lambda.$$

The geocentric co-ordinates of both the ground station and the satellite can also be described in the *terrestrial system.* In this system the co-ordinates of the ground station are

$$\mathbf{R} = R \begin{bmatrix} \cos\Phi\cos\Lambda \\ \cos\Phi\sin\Lambda \\ \sin\Phi \end{bmatrix}$$

and the satellite co-ordinates

$$\mathbf{r} = r \begin{bmatrix} \cos\varphi\cos\lambda \\ \cos\varphi\sin\lambda \\ \sin\varphi \end{bmatrix}.$$

From this again we obtain the *terrestial* topocentric vector:

$$\mathbf{d} = \mathbf{r} - \mathbf{R} \equiv \begin{bmatrix} d\cos\delta_T\cos(\Lambda - h_T) \\ d\cos\delta_T\sin(\Lambda - h_T) \\ d\sin\delta_T \end{bmatrix},$$

from which we may solve the topocentric declination δ_T and *hour angle* h_T.

If we then want the right ascension for use with a celestial chart, all we need to do is subtract it from the local sidereal time:

$$\alpha_T = \theta - h_T = (\theta_0 + \Lambda) + h_T = \theta_0 + (\Lambda - h_T.).$$

In this, θ_0 is Greenwich sidereal time, and $(\Lambda - h_T)$ the "hour angle of Greenwich".

8.8. Determining the orbit from observations

If the satellite's place on the sky has been observed at two different points in time τ_1 and τ_2 – i.e., we have as given $(\alpha(\tau_1), \delta(\tau_1))$ and $(\alpha(\tau_2), \delta(\tau_2))$ – we may compute firstly the topocentric direction vectors (unit vectors):

$$\mathbf{e}_1 = \frac{\mathbf{d}(\tau_1)}{d_1(\tau_1)} = \begin{bmatrix} \cos\delta(\tau_1)\cos\alpha(\tau_1) \\ \cos\delta(\tau_1)\sin\alpha(\tau_1) \\ \sin\delta(\tau_1) \end{bmatrix}, \quad \mathbf{e}_2 = \frac{\mathbf{d}(\tau_2)}{d_2(\tau_2)} = \begin{bmatrix} \cos\delta(\tau_2)\cos\alpha(\tau_2) \\ \cos\delta(\tau_2)\sin\alpha(\tau_2) \\ \sin\delta(\tau_2) \end{bmatrix}.$$

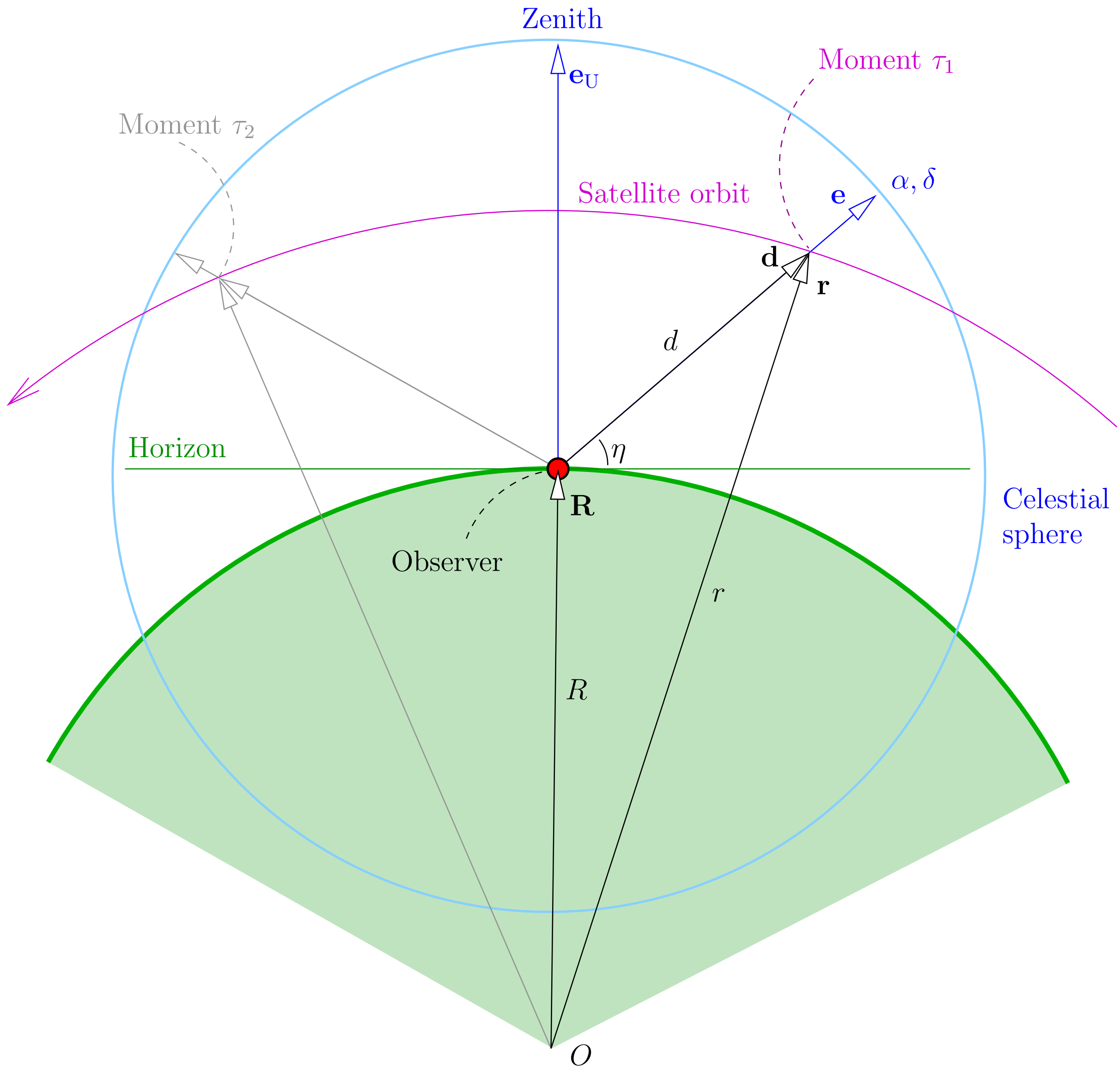


Figure 8.5.: Satellite orbit determination from two visual observations

See Figure 8.5. When also the ground station vector $\mathbf{R}$ has been computed, we may compute d_1, d_2 by the cosine rule if a suitable value[1]for the satellite height h — or equivalently, for the satellite orbit's radius $r = R + h$ – has been given:

$$
\begin{aligned}
r^2 &= R^2 + d^2 + 2Rd\cos(\angle\mathbf{e},\mathbf{R}) \Rightarrow \\
&\Rightarrow d^2 + 2Rd\sin\eta + R^2 - r^2 = 0 \Rightarrow \\
\Rightarrow \quad d^{(1,2)} &= \frac{-2R\sin\eta \pm \sqrt{4R^2\sin^2\eta - 4\left(R^2 - r^2\right)}}{2} = \\
&= -R\sin\eta \pm \sqrt{r^2 - R^2\left(1 - \sin^2\eta\right)},
\end{aligned}
$$

in which $\sin\eta = \cos(\angle\mathbf{e},\mathbf{R}) = \langle\mathbf{e}\cdot\mathbf{e}_U\rangle$ is the projection of the satellite's direction vector onto the local vertical, and

$$
\mathbf{e}_U = \frac{\mathbf{R}}{R} = \begin{bmatrix} \cos\Phi\cos\theta \\ \cos\Phi\sin\theta \\ \sin\Phi \end{bmatrix},
$$

the "zenith vector" of the observer, a unit vector pointing straight upward. The value $\sin\eta$, the sine of the elevation angle, may be calculated directly as the dot product $\langle\mathbf{e}\cdot\mathbf{e}_U\rangle$ when both

[1] An experienced observer can guess a satellite's height from the observed motion in the sky surprisingly precisely.

vectors are given in rectangular form.

Only the positive solution makes physical sense:

$$d = \sqrt{r^2 - R^2\left(1-\sin^2\eta\right)} - R\sin\eta.$$

Thus we obtain d_1, d_2 and thus $\mathbf{d}_1 = d_1\mathbf{e}_1$, $\mathbf{d}_2 = d_2\mathbf{e}_2$. For the satellite velocity we obtain

$$v = \frac{\|\mathbf{d}_2 - \mathbf{d}_1\|}{\tau_2 - \tau_1}.$$

We know however what the velocity for a circular orbit *should* be at height h: according to KEPLER's third law

$$P = \sqrt{\frac{4\pi^2}{GM}\left(R+h\right)^3} \Rightarrow v_k = \frac{2\pi\left(R+h\right)}{P} = \sqrt{\frac{GM}{R+h}}.$$

Let us prepare the following little table:

Height (km)	500	750	1000	1500
Speed (m/s)	7612.609	7477.921	7350.139	7113.071

We see that the satellite's linear speed of flight decreases only slowly with height. Therefore we may use the "observed velocity" v for correcting the height h according to the following formula:

$$h' = h\frac{v_k}{v},$$

where v_k is the velocity according to KEPLER (from the table for the value h), v the calculated velocity, and h' the improved value for the satellite height. This process converges already in one step to almost the correct height.

Note *that the height thus obtained is, in the case of an elliptical orbit, only (approximately) the* "overflight height"! The real height will vary along the orbit. For the same reason, the height obtained will not be good enough to calculate the period P (or equivalently: the semi-major axis a) from. The real period can be inferred only of the satellite has been observed for at least a couple of successive days.

The computed vector for the change in position between the two moments of observation, $\mathbf{d}_2 - \mathbf{d}_1 = \mathbf{r}_2 - \mathbf{r}_1$ also tells about the inclination of the orbit, by calculating its vectorial product with, e.g., the satellite's geocentric location vector $\mathbf{r}_1$, as follows:

$$\left(\mathbf{d}_2 - \mathbf{d}_1\right) \times \mathbf{r}_1 = \|\mathbf{d}_2 - \mathbf{d}_1\| \, \|\mathbf{r}_1\| \cos i,$$

from which i may be solved. And if i and some "footpoint" (φ, λ) of the satellite is known, also Ω and (with P) τ_0 may be computed. Then we can already generate predictions!

When the height of the satellite (and its approximate period) as well as the inclination are known, we can also compute the rapid precessional rate of the ascending node[2]:

$$\begin{aligned}
\Omega\left(\tau\right) &= \Omega\left(\tau_0\right) - \left(\tau - \tau_0\right)\frac{d\Omega}{d\tau} = \\
&= \Omega\left(\tau_0\right) - \left(\tau - \tau_0\right) \cdot \frac{3}{2}\sqrt{\frac{GM}{a^3}}\left(\frac{a_e}{a}\right)^2 J_2 \cos i,
\end{aligned}$$

[2]The formula applies for circular orbits.

in which J_2 is the *dynamic flattening of the Earth*, value $J_2 = 1082.62 \cdot 10^{-6}$. One of the first achievements of the satellite era was the precise determination of J_2[3]. As a numerical formula:

$$\frac{d\Omega}{d\tau} = -6.52927 \cdot 10^{24} \frac{\cos i}{a^{3.5}} \quad \left[\mathrm{m}^{3.5}\,\mathrm{degrees}/\mathrm{day}\right]$$

(note the unit!) Thus we obtain the following table (unit degree/days):

Ht./Incl.	500	750	1000	1500
0°	-7.651	-6.752	-5.985	-4.758
56°	-4.278	-3.776	-3.347	-2.661
65°	-3.233	-2.854	-2.529	-2.011
74°	-2.109	-1.861	-1.650	-1.311
81°	-1.197	-1.056	-0.936	-0.744
90°	0	0	0	0
0.9856	97°.401	98°.394	99°.478	101°.955

In the table the last row is special. The value $0.9856\,\mathrm{degree}/\mathrm{day}$ is *the Sun's apparent angular velocity* relative to the background of the stars. If the precessional velocity of the satellite's orbital plane is set to this value, the satellite will always fly over the same area at the same local solar time. In this way we achieve a so-called heliostationary orbit, also known as "no-shadow" orbit, the advantages of which are continuous light on the solar cells, and in remote sensing, always the same elevation angle of the Sun when imaging the Earth surface. Given is the inclination angle which, for each value of the mean height for each column of the table, gives precisely such a sun-stationary orbit. The precession rate of the satellite's orbital plane *relative to the sun* is now

$$q \equiv \frac{d\Omega}{d\tau} - 0.9856\,\mathrm{degree}/\mathrm{day}$$

E.g., for a satellite whose height is 500 km and inclination 56°:

$$q = -4.278 - 0.9856 = -5.2636\,\mathrm{degree}/\mathrm{day},$$

and the period of one cycle is $360/q = 68\,\mathrm{days} \approx 2.2\,\mathrm{months}$. This is the time span after which the satellite appears again, e.g., in the evening sky of the same latitude.

[3] In fact, this motion of the ascending node is so rapid, that without considering it it is not possible to generate usable orbit predictions.

Chapter 9

The surface theory of Gauss

Carl Friedrich GAUSS[1] was among the first to develop the theory of curved surfaces. The theory he developed assumed still, that the two-dimensional surface is inside three-dimensional space – many derivations are then simple and nevertheless the theory applies as such to the curved surface of the Earth: note that Gauss was a geodesist who measured and calculated the geodetic networks of Hannover and Brunswick using the method of least squares.

Let a curved surface S be given in three-dimensional space $\mathbb{R}^3$. The surface is *parametrized* by the parameters (u, v). Example: the surface of the Earth, parametrization (φ, λ).

In three-dimensional space, we may describe point positions by creating an orthonormal triad of base vectors

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e_3}\}.$$

On this base, a point, or location vector, is

$$\mathbf{x} = x^1\mathbf{e}_1 + x^2\mathbf{e}_2 + x^3\mathbf{e}_3 = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3.$$

We will now often choose to write $\mathbf{x}$ by its *representation* on this basis:

$$\mathbf{x} = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

These three parameters thus form a *parametrization* of $\mathbb{R}^3$.

9.1. A curve in space

A curve running through space C can be parametrized by a parameter s. Then, the points on the curve are

$$\mathbf{x}(s).$$

If it holds for the parameter, that

$$ds^2 = dx^2 + dy^2 + dz^2,$$

we say that C is *parametrized according to distance* (or arc length).

Examples:

1. The numbers on a measurement tape form a parametrization by distance.
2. If you drive along a road, the trip meter readings form a parametrization of the road according to distance.
3. Along Mannerheimintie, the *kkj* co-ordinate x constitutes a parametrization, however *not* according to distance (because the direction of the road is variable).

[1] Carl Friedrich Gauss, (1777 – 1855), was also among the first to speculate on the possibility of non-Euclidean geometry. Others were János Bolyai and Nikolai I. Lobachevsky.

Let us assume in the sequel, that the parametrization used is unambiguous and differentiable (and thus continuous).

The *tangent* of the curve C is obtained by differentiation:

$$\mathbf{t}(s) = \frac{d\mathbf{x}(s)}{ds} = \mathbf{x}_s;$$

the length of the tangent is

$$\|\mathbf{t}\| = \sqrt{\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 + \left(\frac{dz}{ds}\right)^2} = \sqrt{\frac{dx^2 + dy^2 + dz^2}{ds^2}} = \sqrt{1} = 1.$$

This applies *only* if s is a parametrization by distance.

An arbitrary parametrization t can always be converted into a parametrization by distance in the following way:

$$\begin{aligned} s(t) &= \int_0^t \frac{ds}{d\tau} d\tau = \int_0^t \frac{\sqrt{dx^2 + dy^2 + dz^2}}{d\tau} d\tau = \\ &= \int_0^t \sqrt{\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2} d\tau, \end{aligned} \tag{9.1}$$

i.e., in differential form

$$ds = dt\sqrt{\left(\frac{dx(t)}{dt}\right)^2 + \left(\frac{dy(t)}{dt}\right)^2 + \left(\frac{dz(t)}{dt}\right)^2},$$

from which $s(t)$ can always be computed by integration 9.1.

Another differentiation yields the *curvature vector*:

$$\mathbf{k}(s) = \frac{d\mathbf{t}(s)}{ds} = \frac{d^2}{ds^2}\mathbf{x}(s).$$

9.2. The first fundamental form (metric)

The GAUSS *first fundamental form*:

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2. \tag{9.2}$$

ds is a *path element* within the surface.

Later we shall see that this fundamental form is the same as the *metric* of the surface under consideration, and an alternative way of writing is

$$ds^2 = g_{11}du^2 + g_{12}dudv + g_{21}dvdu + g_{22}dv^2.$$

If the point $\mathbf{x}$ is on the surface S, we may take the *derivative* of its co-ordinates:

$$\mathbf{x}_u = \frac{\partial \mathbf{x}}{\partial u} = \begin{bmatrix} \partial x/\partial u \\ \partial y/\partial u \\ \partial z/\partial u \end{bmatrix}, \ \mathbf{x}_v = \frac{\partial \mathbf{x}}{\partial v} = \begin{bmatrix} \partial x/\partial v \\ \partial y/\partial v \\ \partial z/\partial v \end{bmatrix}.$$

These vectors are called the *tangent vectors* of surface S and parametrization (u, v).

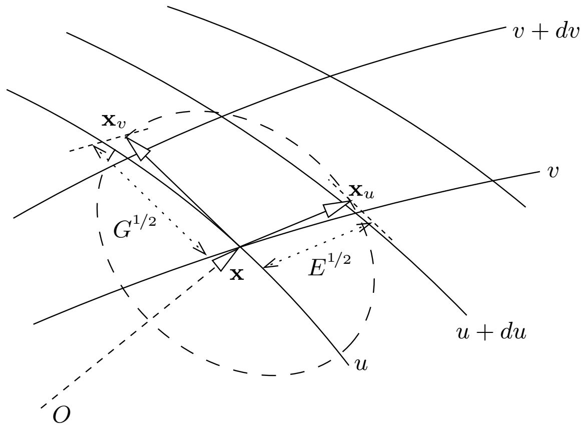


Figure 9.1.: The Gauss first fundamental form

From these we we obtain, as "dot products" of two vectors in space, the *elements of the fundamental form*:

$$
\begin{aligned}
E &= \langle \mathbf{x}_u \cdot \mathbf{x}_u \rangle\,, \\
F &= \langle \mathbf{x}_u \cdot \mathbf{x}_v \rangle\,, \\
G &= \langle \mathbf{x}_v \cdot \mathbf{x}_v \rangle\,.
\end{aligned}
$$

In figure 9.1 we have depicted the Gauss first fundamental form and the tangent vectors $\mathbf{x}_u, \mathbf{x}_v$. It is easy to show (chain rule in three dimensions (x, y, z)), that

$$
\begin{aligned}
ds^2 &= dx^2 + dy^2 + dz^2 = \\
&= \left[\left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2\right] du^2 + \\
+2 & \left[\left(\frac{\partial x}{\partial u}\frac{\partial x}{\partial v}\right) + \left(\frac{\partial y}{\partial u}\frac{\partial y}{\partial v}\right) + \left(\frac{\partial z}{\partial u}\frac{\partial z}{\partial v}\right)\right] dudv + \\
+ & \left[\left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2\right] dv^2 = \\
&= \langle \mathbf{x}_u \cdot \mathbf{x}_u \rangle\, du^2 + 2\,\langle \mathbf{x}_u \cdot \mathbf{x}_v \rangle\, dudv + \langle \mathbf{x}_v \cdot \mathbf{x}_v \rangle\, dv^2,
\end{aligned}
$$

from which (9.2) follows directly.

We see, e.g., that in the direction of the v curves ($dv = 0$):

$$ds^2 = E du^2,$$

in other words, E represents the *metric distance* between two successive curves $(u, u+1)$. Similarly G represents the distance between two successive v curves. The closer the curves are to each other, the larger $\mathbf{x}_u$ or $\mathbf{x}_v$ and also the larger E or G. F again represents the *angle* between the u and v curves: it vanishes if the angle is straight.

9.3. The second fundamental form

The *normal* on a surface is the vector which is orthogonal to every curve running within the surface, also the u and v curves. We write

$$\mathbf{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}.$$

Apparently

$$\langle \mathbf{n} \cdot \mathbf{x}_u \rangle = \langle \mathbf{n} \cdot \mathbf{x}_v \rangle = 0. \tag{9.3}$$

Let us also require

$$\|\mathbf{n}\| = 1,$$

or

$$n_x^2 + n_y^2 + n_z^2 = 1.$$

We differentiate $\mathbf{x}$ a second time:

$$\begin{aligned} \mathbf{x}_{uu} &= \frac{\partial^2 \mathbf{x}}{\partial u^2}, \\ \mathbf{x}_{uv} &= \frac{\partial \mathbf{x}}{\partial u \partial v}, \\ \mathbf{x}_{vv} &= \frac{\partial \mathbf{x}}{\partial v^2}. \end{aligned}$$

Now, Gauss's *second fundamental form* is

$$edu^2 + 2f dudv + gdv^2,$$

where

$$\begin{aligned} e &= \langle \mathbf{n} \cdot \mathbf{x}_{uu} \rangle, \\ f &= \langle \mathbf{n} \cdot \mathbf{x}_{uv} \rangle, \\ g &= \langle \mathbf{n} \cdot \mathbf{x}_{vv} \rangle. \end{aligned} \tag{9.4}$$

Based on condition (9.3) we also have

$$\begin{aligned} 0 = \frac{\partial}{\partial u} \langle \mathbf{n} \cdot \mathbf{x}_u \rangle = \langle \mathbf{n}_u \cdot \mathbf{x}_u \rangle + \langle \mathbf{n} \cdot \mathbf{x}_{uu} \rangle &\Rightarrow e = -\mathbf{n}_u \cdot \mathbf{x}_u, \\ 0 = \frac{\partial}{\partial v} \langle \mathbf{n} \cdot \mathbf{x}_u \rangle = \langle \mathbf{n}_v \cdot \mathbf{x}_u \rangle + \langle \mathbf{n} \cdot \mathbf{x}_{uv} \rangle &\Rightarrow f = -\mathbf{n}_u \cdot \mathbf{x}_v, \\ 0 = \frac{\partial}{\partial u} \langle \mathbf{n} \cdot \mathbf{x}_v \rangle = \langle \mathbf{n}_u \cdot \mathbf{x}_v \rangle + \langle \mathbf{n} \cdot \mathbf{x}_{uv} \rangle &\Rightarrow f = -\mathbf{n}_v \cdot \mathbf{x}_u, \\ 0 = \frac{\partial}{\partial v} \langle \mathbf{n} \cdot \mathbf{x}_v \rangle = \langle \mathbf{n}_v \cdot \mathbf{x}_v \rangle + \langle \mathbf{n} \cdot \mathbf{x}_{vv} \rangle &\Rightarrow g = -\mathbf{n}_v \cdot \mathbf{x}_v. \end{aligned} \tag{9.5}$$

See figure 9.2. When moving from location to location, the normal vector's direction changes: when we move from point $\mathbf{x}$ (u, v) to point $\mathbf{x}'$ $(u + du, v)$**,** the normal changes $\mathbf{n} \to \mathbf{n}' = \mathbf{n} + d\mathbf{n}$'. In the same way, when moving from point $\mathbf{x}$ to $\mathbf{x}''$ $(u, v + dv)$, the normal changes $\mathbf{n} \to \mathbf{n}'' = \mathbf{n} + d\mathbf{n}''$.

As a formula:

$$d\mathbf{n} = \frac{\partial \mathbf{n}}{\partial u} du + \frac{\partial \mathbf{n}}{\partial v} dv = \mathbf{n}_u du + \mathbf{n}_v dv.$$

The norm of the normal vector, i.e., its length, is always 1, like we saw above; that's why the vector can change only in two directions, either in the direction of tangent vector $\mathbf{x}_u$, or in the direction of tangent vector $\mathbf{x}_v$.

Let us separate them by *projection*:

$$\begin{aligned} \langle d\mathbf{n} \cdot \mathbf{x}_u \rangle &= \langle \mathbf{n}_u \cdot \mathbf{x}_u \rangle du + \langle \mathbf{n}_v \cdot \mathbf{x}_u \rangle dv, \\ \langle d\mathbf{n} \cdot \mathbf{x}_v \rangle &= \langle \mathbf{n}_u \cdot \mathbf{x}_v \rangle du + \langle \mathbf{n}_v \cdot \mathbf{x}_v \rangle dv, \end{aligned}$$

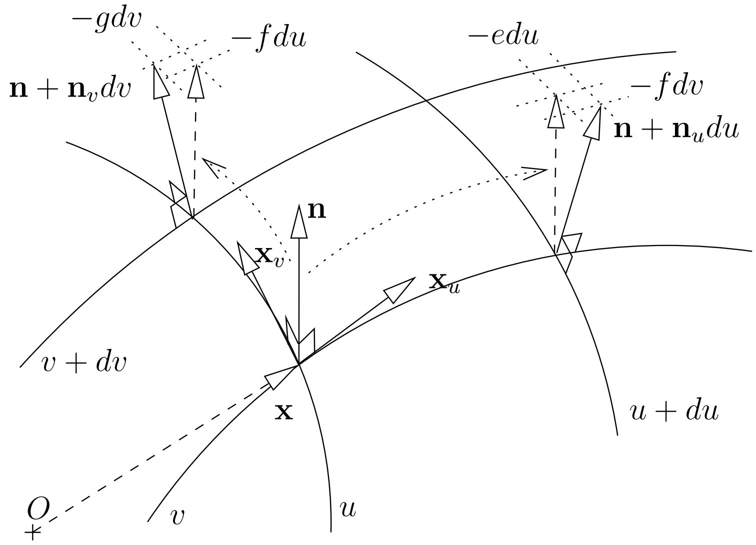


Figure 9.2.: Explaining geometrically the GAUSS second fundamental form

in which we may directly identify the elements of the second fundamental form e, f, g with the aid of (9.5):

$$
\begin{aligned}
-\langle d\mathbf{n}\cdot\mathbf{x}_u\rangle &= edu + fdv,\\
-\langle d\mathbf{n}\cdot\mathbf{x}_v\rangle &= fdu + gdv.
\end{aligned}
$$

Contrary to the first fundamental form, the second fundamental form has no corresponding object in the Riemann surface theory. It exists only for surfaces *embedded* in a surrounding (Euclidean) space.

Often we can also find a “tensorial” notation:

$$
\begin{aligned}
\beta_{11} &= e,\\
\beta_{12}=\beta_{21} &= f,\\
\beta_{22} &= g.
\end{aligned}
$$

9.4. Principal curvatures

The GAUSS second fundamental form describes in a way the curvature of a surface in space, by depicting how the direction of the normal vector changes, when we travel either in the u or in the v co-ordinate curve direction. Unfortunately this is not enough for an *absolute* characterization of the curvature, because the parametrization (u, v) is

1. an arbitrary choice, and
2. not metrically scaled.

The latter means that if the direction of the normal vector $\mathbf{n}$ changes by an amount $d\mathbf{n}$ when we travel a distance du along the v *co-ordinate curve, we still don't know* how many metres *the distance* du corresponds to. *If it is a long distance, then the same change in the normal vector* $d\mathbf{n}$ means only a small curvature of the surface; if it is a short distance, then the same change $d\mathbf{n}$ corresponds to a large surface curvature.

The fact that the parameter curves $u =$ constant and $v =$ constant generally not are perpendicular to each other, makes this problem even trickier.

Write the first and second fundamental forms in matrix form:

$$
H = \begin{bmatrix} E & F \\ F & G \end{bmatrix}, \; B = \begin{bmatrix} e & f \\ f & g \end{bmatrix}.
$$

Form the matrix:

$$C \equiv H^{-1}B = \begin{bmatrix} G & -F \\ -F & E \end{bmatrix} \frac{1}{EG-F^2} \begin{bmatrix} e & f \\ f & g \end{bmatrix} = .$$
$$= \frac{1}{EG-F^2} \begin{bmatrix} Ge-Ff & Gf-Fg \\ Ef-Fe & Eg-Ff \end{bmatrix}$$

This is called the *shape operator*, and the above, the WEINGARTEN[2] equations, see `http://en.wikipedia.org/wiki/Differential_geometry_of_surfaces#Shape_operator`

We can say that multiplication by the inverse of the H matrix, i.e., the first fundamental form, which describes the *length of a distance element*, performs a *metric scaling* of the B *matrix*[3].

Principal curvatures:

The matrix C has *two eigenvalues*: the values $\kappa_{1,2}$ for which

$$(C - \kappa I)\,\mathrm{x} = 0 \tag{9.6}$$

for suitable value pairs $\mathrm{x} = \begin{bmatrix} du & dv \end{bmatrix}^T$. The solutions $\kappa_{1,2}$ are called the *principal curvatures* of surface S. They are *invariant* with respect to the chosen parametrization (u, v). The corresponding value pairs $\mathrm{x}_1 = \begin{bmatrix} du_1 & dv_1 \end{bmatrix}^T$ and $\mathrm{x}_2 = \begin{bmatrix} du_2 & dv_2 \end{bmatrix}^T$ define the local *principal directions of curvature* on the surface.

Other invariants:

1. The product $\kappa_1\kappa_2 = \det C = \dfrac{\det B}{\det H} = \dfrac{eg - f^2}{EG - F^2}$ is the *Gaussian curvature.*
2. The half-sum $\dfrac{1}{2}(\kappa_1 + \kappa_2) = \dfrac{1}{2}(C_{11} + C_{22}) = \dfrac{1}{2}\dfrac{eG + gE - 2fF}{EG - F^2}$ is the *mean curvature.*

Principal directions of curvature:

We may also study the *eigenvectors* of C, which are called *principal directions of curvature.* For this, write

$$0 = H(C - \kappa_1 I)\,\mathrm{x}_1 = (B - \kappa_1 H)\,\mathrm{x}_1,$$
$$0 = H(C - \kappa_2 I)\,\mathrm{x}_2 = (B - \kappa_2 H)\,\mathrm{x}_2.$$

Multiply the first from the left with x_2^T, and the second with x_1^T and transposing:

$$0 = \mathrm{x}_2^T B \mathrm{x}_1 - \kappa_1 \mathrm{x}_2^T H \mathrm{x}_1,$$
$$0 = \left(\mathrm{x}_1^T B \mathrm{x}_2 - \kappa_2 \mathrm{x}_1^T H \mathrm{x}_2\right)^T = \mathrm{x}_2^T B \mathrm{x}_1 - \kappa_2 \mathrm{x}_2^T H \mathrm{x}_1,$$

as, both B and H being symmetric matrices, we have

$$\left(\mathrm{x}_1^T B \mathrm{x}_2\right)^T = \mathrm{x}_2^T B \mathrm{x}_1 \text{ and}$$
$$\left(\mathrm{x}_1^T H \mathrm{x}_2\right)^T = \mathrm{x}_2^T H \mathrm{x}_1.$$

[2] Julius WEINGARTEN, German mathematician 1836-1910

[3] A more technical description: B is a covariant tensor β_{ij}, and H the covariant metric tensor g_{ij}. H^{-1} again corresponds to the contravariant tensor g^{ij}. C is now the "mixed tensor" $\beta^i_k = g^{ij}\beta_{jk}$, whose *tensorial* eigenvalue problem is

$$\left(\beta^i_j - \kappa\delta^i_j\right) x^j = 0,$$

the same equation as (9.6).
See chapter 10.

Subtraction of the two yields

$$(\kappa_2 - \kappa_1)\, \mathrm{x}_2^T H \mathrm{x}_1 = 0.$$

This shows that *if the principal radii of curvature are different*[4], then the expression $x_2^T H x_1$ vanishes. This expression[5] can be interpreted as an *inner product*: in fact, in Cartesian plane co-ordinates in the tangent plane, the matrix $H = I$, and we have $\mathrm{x}_2^T \mathrm{x}_1 = 0$, or $\mathrm{x}_1 \perp \mathrm{x}_2$ in the Euclidean sense.

The principal directions of curvature are mutually perpendicular.

This is just a special case of self-adjoint (symmetric) operators having mutually orthogonal eigenvectors, e.g., the eigenfunctions of Sturm-Liouville theory (`http://en.wikipedia.org/wiki/Sturm%E2%80%93Liouville_theory`).

Example:

The co-ordinates of a point on the surface of the ellipsoidal Earth are

$$\mathbf{x} = \begin{bmatrix} N(\varphi)\cos\varphi\cos\lambda \\ N(\varphi)\cos\varphi\sin\lambda \\ N(\varphi)\left(1-e^2\right)\sin\varphi \end{bmatrix}.$$

From this

$$\mathbf{x}_\varphi = \frac{\partial \mathbf{x}}{\partial \varphi} = \begin{bmatrix} \cos\lambda \dfrac{d}{d\varphi}\left(N(\varphi)\cos\varphi\right) \\ \sin\lambda \dfrac{d}{d\varphi}\left(N(\varphi)\cos\varphi\right) \\ \left(1-e^2\right)\dfrac{d}{d\varphi}\left(N(\varphi)\sin\varphi\right) \end{bmatrix};$$

Using the formulas derived in the Appendix B we obtain:

$$\mathbf{x}_\varphi = M(\varphi) \begin{bmatrix} -\sin\varphi\cos\lambda \\ -\sin\varphi\sin\lambda \\ +\cos\varphi \end{bmatrix}.$$

$$\mathbf{x}_\lambda = \frac{\partial \mathbf{x}}{\partial \lambda} = N(\varphi) \begin{bmatrix} -\cos\varphi\sin\lambda \\ +\cos\varphi\cos\lambda \\ 0 \end{bmatrix}.$$

The surface normal is obtained as the vectorial product, normalized:

$$\mathbf{n} = \frac{\langle \mathbf{x}_\varphi \times \mathbf{x}_\lambda \rangle}{\| \mathbf{x}_\varphi \times \mathbf{x}_\lambda \|},$$

where

$$\begin{aligned} \langle \mathbf{x}_\varphi \times \mathbf{x}_\lambda \rangle &= NM \begin{bmatrix} -\cos^2\varphi\cos\lambda \\ -\cos^2\varphi\sin\lambda \\ -\sin\varphi\cos\varphi\cos^2\lambda - \sin\varphi\cos\varphi\sin^2\lambda \end{bmatrix} = \\ &= -NM \begin{bmatrix} \cos^2\varphi\cos\lambda \\ \cos^2\varphi\sin\lambda \\ \sin\varphi\cos\varphi \end{bmatrix} = -NM\cos^2\varphi \begin{bmatrix} \cos\lambda \\ \sin\lambda \\ \tan\varphi \end{bmatrix}, \end{aligned}$$

the norm of which is

$$\| \mathbf{x}_\varphi \times \mathbf{x}_\lambda \| = NM\cos^2\varphi\sqrt{1+\tan^2\varphi} = NM\cos\varphi.$$

[4] And if they are not, any linear combination of x_1 and x_2 will again be an eigenvector, and we can always choose two that are mutually perpendicular.

[5] In index notation: $g_{ij} x_2^i x_1^j$.

Thus

$$\mathbf{n} = -\begin{bmatrix} \cos\varphi\cos\lambda \\ \cos\varphi\sin\lambda \\ \sin\varphi \end{bmatrix},$$

not a surprising result...

Let us compute the first fundamental form:

$$\begin{aligned} E &= \langle \mathbf{x}_\varphi \cdot \mathbf{x}_\varphi \rangle = M^2 \left(\sin^2\varphi \left(\sin^2\lambda + \cos^2\lambda \right) + \cos^2\varphi \right) = M^2, \\ F &= \langle \mathbf{x}_\varphi \cdot \mathbf{x}_\lambda \rangle = 0, \\ G &= \langle \mathbf{x}_\lambda \cdot \mathbf{x}_\lambda \rangle = N^2 \cos^2\varphi = p^2. \end{aligned}$$

We calculate for use in calculating the second fundamental form

$$\mathbf{n}_\varphi = \begin{bmatrix} +\sin\varphi\cos\lambda \\ +\sin\varphi\sin\lambda \\ -\cos\varphi \end{bmatrix};\ \mathbf{n}_\lambda = \begin{bmatrix} +\cos\varphi\sin\lambda \\ -\cos\varphi\cos\lambda \\ 0 \end{bmatrix}$$

and thus (equations 9.5)

$$\begin{aligned} e &= -\langle \mathbf{n}_\varphi \cdot \mathbf{x}_\varphi \rangle = +M, \\ f &= -\langle \mathbf{n}_\varphi \cdot \mathbf{x}_\lambda \rangle = -\langle \mathbf{n}_\lambda \cdot \mathbf{x}_\varphi \rangle = 0, \\ g &= -\langle \mathbf{n}_\lambda \cdot \mathbf{x}_\lambda \rangle = +N\cos^2\varphi. \end{aligned}$$

In other words:

$$\begin{aligned} H &= \begin{bmatrix} E & F \\ F & G \end{bmatrix} = \begin{bmatrix} M^2 & 0 \\ 0 & N^2\cos^2\varphi \end{bmatrix}, \\ B &= \begin{bmatrix} e & f \\ f & g \end{bmatrix} = \begin{bmatrix} M & 0 \\ 0 & N\cos^2\varphi \end{bmatrix}, \text{ and} \\ C &= H^{-1}B = \begin{bmatrix} \frac{1}{M} & 0 \\ 0 & \frac{1}{N} \end{bmatrix}. \end{aligned}$$

The principal curvatures $\kappa_{1,2}$ are C's characteristic values or *eigenvalues*, solutions of the eigenvalue problem

$$(C - \kappa I)\,\mathrm{x} = 0;$$

we obtain the values by solving

$$\det(C - \kappa I) = 0 \ \Rightarrow\ \det \begin{bmatrix} \frac{1}{M} - \kappa & 0 \\ 0 & \frac{1}{N} - \kappa \end{bmatrix} = 0 \ \Rightarrow$$

$$\Rightarrow \left(\frac{1}{M} - \kappa \right) \left(\frac{1}{N} - \kappa \right) = 0 \ \Rightarrow\ \kappa_1 = \frac{1}{M},\ \kappa_2 = \frac{1}{N}.$$

As could be expected...

9.5. A curve embedded in a surface

If a curve C runs inside a curved surface, we may study other interesting things.

The tangent vector

If we call[6] $t^i = \begin{bmatrix} t^1 \\ t^2 \end{bmatrix} = \begin{bmatrix} du/ds \\ dv/ds \end{bmatrix}$ "the $\mathbf{t}$ vector's components in the (u, v) co-ordinate system", this is $\mathbf{t} = \frac{d\mathbf{x}}{ds} = \frac{d\mathbf{x}}{du}\frac{du}{ds} + \frac{d\mathbf{x}}{dv}\frac{dv}{ds} = t^1\mathbf{x}_u + t^2\mathbf{x}_v$. In other words, the tangent to the curve is also one of the tangents to the surface, and lies inside the tangent plane.

The curvature vector

$$
\begin{aligned}
\mathbf{k} = \frac{d\mathbf{t}}{ds} &= \frac{d}{ds}\left(t^1\mathbf{x}_u + t^2\mathbf{x}_v\right) = \\
&= \mathbf{x}_u \frac{dt^1}{ds} + \mathbf{x}_v \frac{dt^2}{ds} + \\
&+ \mathbf{x}_{uu}\left(t^1\right)^2 + 2\mathbf{x}_{uv}t^1t^2 + \mathbf{x}_{vv}\left(t^2\right)^2.
\end{aligned}
$$

In other words, the "curvature vector's components in the (u, v) co-ordinate system" ontain other things besides just the derivatives of the component values dt^1/ds ja dt^2/ds.

We write

$$
\begin{aligned}
\mathbf{x}_{uu} &= \Gamma^1_{11}\mathbf{x}_u + \Gamma^2_{11}\mathbf{x}_v + \langle \mathbf{x}_{uu} \cdot \mathbf{n} \rangle \mathbf{n}, \\
\mathbf{x}_{uv} &= \Gamma^1_{12}\mathbf{x}_u + \Gamma^2_{12}\mathbf{x}_v + \langle \mathbf{x}_{uv} \cdot \mathbf{n} \rangle \mathbf{n}, \\
\mathbf{x}_{vv} &= \Gamma^1_{22}\mathbf{x}_u + \Gamma^2_{22}\mathbf{x}_v + \langle \mathbf{x}_{vv} \cdot \mathbf{n} \rangle \mathbf{n};
\end{aligned} \tag{9.7}
$$

i.e., we develop the three-dimensional vectors on the base $(\mathbf{x}_u, \mathbf{x}_v, \mathbf{n})$ of the space.

Here appear — or naturally arise — the Γ symbols or "Christoffel symbols" which we discuss later (Chapter 10). They illustrate the reality that *differentiating a vector* (also known as "parallel transport", see Chapter 10) in curvilinear co-ordinates on a curved surface is not a trivial thing.

The third term on the right hand side however represents, according to definition (9.4) , the *elements of the second fundamental form* e, f, g.

We obtain

$$
\begin{aligned}
\mathbf{k} &= \left(\frac{dt^1}{ds} + \Gamma^1_{11}\left(t^1\right)^2 + 2\Gamma^1_{12}t^1t^2 + \Gamma^1_{22}\left(t^2\right)^2\right)\mathbf{x}_u + \\
&+ \left(\frac{dt^2}{ds} + \Gamma^2_{11}\left(t^1\right)^2 + 2\Gamma^2_{12}t^1t^2 + \Gamma^2_{22}\left(t^2\right)^2\right)\mathbf{x}_v + \\
&+ \left(e\left(t^1\right)^2 + 2ft^1t^2 + g\left(t^2\right)^2\right)\mathbf{n}.
\end{aligned}
$$

Here, the first two terms represent the internal curvature $\mathbf{k}_{int}$ of the curve C, the curvature inside surface S; the latter term represents the *exterior curvature* $\mathbf{k}_{ext}$, the curvature of the curve "along with" the itself curved surface.

The internal curvature again has two components in "(u, v) co-ordinates", which can be read from the above equation. We write

$$
\mathbf{k}_{int} = k^1\mathbf{x}_u + k^2\mathbf{x}_v,
$$

[6] Don't get nervous about the use of a superscript (upper index). It's just a writing convention, the good sense of which we will argue later on

where

$$k^i = \begin{bmatrix} k^1 \\ k^2 \end{bmatrix} = \begin{bmatrix} dt^1/ds + \Gamma^1_{11}\left(t^1\right)^2 + 2\Gamma^1_{12}t^1t^2 + \Gamma^1_{22}\left(t^2\right)^2 \\ dt^2/ds + \Gamma^2_{11}\left(t^1\right)^2 + 2\Gamma^2_{12}t^1t^2 + \Gamma^2_{22}\left(t^2\right)^2 \end{bmatrix} =$$

$$= \begin{bmatrix} dt^1/ds + \sum_{i=1}^{2}\sum_{j=1}^{2}\Gamma^1_{ij}t^it^j \\ dt^2/ds + \sum_{i=1}^{2}\sum_{j=1}^{2}\Gamma^2_{ij}t^it^j \end{bmatrix},$$

where we have "economized" the formulas by using summation signs. The external curvature again is

$$\mathbf{k}_{ext} = \langle \mathbf{k} \cdot \mathbf{n} \rangle \, \mathbf{n},$$

where

$$\langle \mathbf{k} \cdot \mathbf{n} \rangle = e\left(t^1\right)^2 + 2ft^1t^2 + g\left(t^2\right)^2.$$

Example:

a latitude circle on the Earth surface.

$$\mathbf{x} = \begin{bmatrix} N(\varphi)\cos\varphi\cos\lambda \\ N(\varphi)\cos\varphi\sin\lambda \\ N(\varphi)\left(1-e^2\right)\sin\varphi \end{bmatrix};$$

$$\mathbf{t} = \frac{d\mathbf{x}}{ds} = \frac{d\mathbf{x}}{d\lambda}\frac{d\lambda}{ds} = \begin{bmatrix} -N\cos\varphi\sin\lambda \\ +N\cos\varphi\cos\lambda \\ 0 \end{bmatrix}\frac{1}{N\cos\varphi} = \begin{bmatrix} -\sin\lambda \\ +\cos\lambda \\ 0 \end{bmatrix};$$

$$\mathbf{k} = \frac{d\mathbf{t}}{ds} = \frac{d\mathbf{t}}{d\lambda}\frac{d\lambda}{ds} = \begin{bmatrix} -\cos\lambda \\ -\sin\lambda \\ 0 \end{bmatrix}\frac{1}{N\cos\varphi}.$$

External curvature:

$$\mathbf{n} = \begin{bmatrix} \cos\varphi\cos\lambda \\ \cos\varphi\sin\lambda \\ \sin\varphi \end{bmatrix},$$

so

$$\mathbf{k}_{ext} = \langle \mathbf{k} \cdot \mathbf{n} \rangle \, \mathbf{n} = -\frac{1}{N}\left(\cos^2\lambda + \sin^2\lambda\right)\mathbf{n} = -\frac{\mathbf{n}}{N}.$$

Because the vector $\mathbf{n}$ is a unit vector, we may infer, that the exterior curvature of the curve is precisely the inverse of the transverse radius of curvature, and directed inward. This is the same as the curvature of the surface taken in the direction of the curve.

The internal curvature is

$$\mathbf{k}_{int} = \mathbf{k} - \mathbf{k}_{ext} = \frac{1}{N}\begin{bmatrix} -\frac{\cos\lambda}{\cos\varphi} + \cos\varphi\cos\lambda \\ -\frac{\sin\lambda}{\cos\varphi} + \cos\varphi\sin\lambda \\ +\sin\varphi \end{bmatrix} = \frac{\tan\varphi}{N}\begin{bmatrix} -\cos\lambda\sin\varphi \\ -\sin\lambda\sin\varphi \\ +\cos\varphi \end{bmatrix}.$$

We obtain for the length of this vector

$$\|\mathbf{k}_{int}\| = \frac{\tan\varphi}{N}.$$

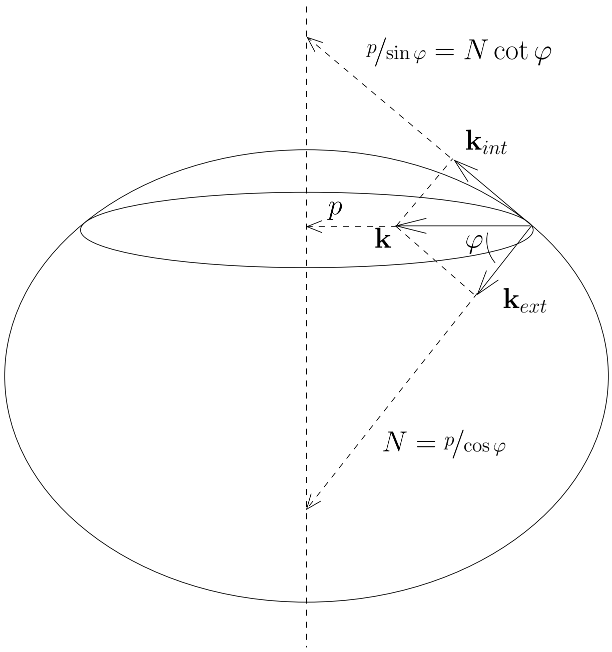


Figure 9.3.: The curvatures and radii of curvature of a latitude circle

In Figure 9.3 we see the curvature vector $\mathbf{k}$, length $k = \dfrac{1}{N\cos\varphi} = \dfrac{1}{p(\varphi)}$; its internal part $\mathbf{k}_{int}$, length $k\sin\varphi = \dfrac{1}{p(\varphi)}\sin\varphi = \dfrac{\tan\varphi}{N(\varphi)}$; and its external part $\mathbf{k}_{ext}$, length $k\cos\varphi = \dfrac{1}{p(\varphi)}\cos\varphi = \dfrac{1}{N(\varphi)}$. In the figure one also sees how the distance from the Earth's rotation axis is in every direction ($\mathbf{k}, \mathbf{k}_{int}$ ja $\mathbf{k}_{ext}$) the inverse of the curvature. This is also intuitively clear from rotational symmetry.

9.6. The geodesic

Internally, in surface co-ordinates

We obtain the formula for the geodesic by requiring $\mathbf{k}_{int} = k^i = 0$, i.e., the curve *has no interior curvature* (the exterior curvature cannot be eliminated as the curve is on a curved surface):

$$\frac{dt^i}{ds} + \sum_{j=1}^{2}\sum_{k=1}^{2}\Gamma^i_{jk}t^j t^k = 0. \tag{9.8}$$

This approach is developed further later on in chapter 10.5.

Externally, using vectors in space

An alternative, three-dimensional ("exterior") form is obtained by observing, that the geodesic is *only externally curved*, i.e., by writing

$$\frac{d\mathbf{t}}{ds} = \mathbf{k}_{ext} = \left(e\left(t^1\right)^2 + 2ft^1t^2 + g\left(t^2\right)^2\right)\mathbf{n} = \left(\begin{bmatrix} t^1 & t^2 \end{bmatrix}\begin{bmatrix} e & f \\ f & g \end{bmatrix}\begin{bmatrix} t^1 \\ t^2 \end{bmatrix}\right)\mathbf{n}.$$

Because

$$\begin{aligned}
\langle \mathbf{t}\cdot\mathbf{x}_u\rangle &= t^1\langle \mathbf{x}_u\cdot\mathbf{x}_u\rangle + t^2\langle \mathbf{x}_u\cdot\mathbf{x}_v\rangle = Et^1 + Ft^2,\\
\langle \mathbf{t}\cdot\mathbf{x}_v\rangle &= t^1\langle \mathbf{x}_v\cdot\mathbf{x}_u\rangle + t^2\langle \mathbf{x}_v\cdot\mathbf{x}_v\rangle = Ft^1 + Gt^2,
\end{aligned}$$

it follows that

$$\begin{bmatrix} t^1 \\ t^2 \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{x}_u \rangle \\ \langle \mathbf{t} \cdot \mathbf{x}_v \rangle \end{bmatrix}.$$

If we define

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \equiv \begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{x}_u \rangle \\ \langle \mathbf{t} \cdot \mathbf{x}_v \rangle \end{bmatrix}$$

we may write

$$\frac{d\mathbf{t}}{ds} = \left(\begin{array}{r} \begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{x}_u \rangle & \langle \mathbf{t} \cdot \mathbf{x}_v \rangle \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \cdot \\ \cdot \begin{bmatrix} e & f \\ f & g \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{x}_u \rangle \\ \langle \mathbf{t} \cdot \mathbf{x}_v \rangle \end{bmatrix} \end{array} \right) \mathbf{n}.$$

Here

$$\begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} \begin{bmatrix} e & f \\ f & g \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1} = H^{-1}BH^{-1}$$

following the earlier used notation[7].

Example: on the ellipsoidal surface of the Earth, we have

$$\begin{aligned} H^{-1}BH^{-1} &= \begin{bmatrix} M^{-2} & 0 \\ 0 & N^{-2}\cos^{-2}\varphi \end{bmatrix} \begin{bmatrix} M & \\ & N\cos^2\varphi \end{bmatrix} \begin{bmatrix} M^{-2} & 0 \\ 0 & N^{-2}\cos^{-2}\varphi \end{bmatrix} = \\ &= \begin{bmatrix} M^{-3} & 0 \\ 0 & N^{-3}\cos^{-2}\varphi \end{bmatrix} = \begin{bmatrix} 1/M^3 & 0 \\ 0 & \cos\varphi/p^3 \end{bmatrix} \end{aligned}$$

and

$$\begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{x}_u \rangle \\ \langle \mathbf{t} \cdot \mathbf{x}_v \rangle \end{bmatrix} = \begin{bmatrix} M \langle \mathbf{t} \cdot \mathbf{e}_N \rangle \\ p \langle \mathbf{t} \cdot \mathbf{e}_E \rangle \end{bmatrix},$$

from which we obtain

$$\begin{aligned} \frac{d\mathbf{t}}{ds} &= \begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{e}_N \rangle & \langle \mathbf{t} \cdot \mathbf{e}_E \rangle \end{bmatrix} \begin{bmatrix} 1/M & 0 \\ 0 & \cos\varphi/p \end{bmatrix} \begin{bmatrix} \langle \mathbf{t} \cdot \mathbf{e}_N \rangle \\ \langle \mathbf{t} \cdot \mathbf{e}_E \rangle \end{bmatrix} \mathbf{n} = \\ &= \left\{ \frac{1}{M} \langle \mathbf{t} \cdot \mathbf{e}_N \rangle^2 + \frac{1}{N} \langle \mathbf{t} \cdot \mathbf{e}_E \rangle^2 \right\} \mathbf{n}. \end{aligned}$$

Here, the unit vectors

$$\mathbf{e}_N = \frac{\mathbf{x}_u}{\|\mathbf{x}_u\|} = \begin{bmatrix} -\sin\varphi\cos\lambda \\ -\sin\varphi\sin\lambda \\ \cos\varphi \end{bmatrix}, \ \mathbf{e}_E = \frac{\mathbf{x}_v}{\|\mathbf{x}_v\|} = \begin{bmatrix} -\cos\varphi\sin\lambda \\ +\cos\varphi\cos\lambda \\ 0 \end{bmatrix}.$$

The approach is geometrically intuitive. The expression in wave brackets $\frac{1}{M} \langle \mathbf{t} \cdot \mathbf{e}_N \rangle^2 + \frac{1}{N} \langle \mathbf{t} \cdot \mathbf{e}_E \rangle^2 = \frac{1}{M}\cos^2 A + \frac{1}{N}\sin^2 A$ is precisely the *curvature of the surface in the direction of the curve.*

In addition we have the equations $\frac{d\mathbf{x}}{ds} = \mathbf{t}$, altogether $2 \times 3 = 6$ ordinary differential equations.

The method of space vectors has both *advantages* and *disadvantages* compared to the surface co-ordinate method (e.g., equations (2.1)).

Advantage: in surface co-ordinates (φ, λ) there are inevitably two *poles*, singularities where the curvature of the latitude circles goes to infinity and numerical methods can be ill-behaved. This will not happen in rectangular space co-ordinates.

[7] In index notation: $g^{ij}\beta_{jk}g^{kl}$, see chapter 10. A logical notation for this would be β^{il}.

Disadvantages:

1. More equations, more computational effort.
2. At every point, we must compute M and N and for this $\varphi = \arctan \dfrac{Z}{(1-e^2)\,p}$, where $p = \sqrt{X^2+Y^2}$.
3. The "roll-in" and "roll-out" of the computation of the geodesic presupposes the transformation of (φ, λ) to (X, Y, Z) co-ordinates in the starting point, and back in the end point.

Chapter 10

The surface theory of Riemann

An important theoretical frame in which curved surfaces and curves are often described, is RIEMANN's[1] surface theory. *In this theory, we study the curved surface* intrinsically, *i.e., without taking into account that the curved surface of the Earth is "embedded" in a complete three-dimensional (Euclidean) space.*

This makes the surface theory of Riemann useful also in situations, where this embedding higher-dimensional space doesn't necessarily even exist. For example, in General Relativity we describe spacetime (x, y, z, t) as a curved manifold according to Riemann's theory, the curvature parameters of which relate to the densities of mass and current through EINSTEIN[2]'s field equations. The gravitational field is the expression of this curvature.

10.1. What is a tensor?

In physics, more than in geodesy, we often encounter the concept of "tensor". What is a tensor?

Vectors

First of all, a *vector*. Initially we shall look at the situation in two-dimensional Euclidean space, i.e., the *plane*.

A pair of co-ordinate differences between two adjacent points is a vector:

$$\mathbf{v} = v^i = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \tag{10.1}$$

is a vector, written with a superscript.

In the other co-ordinate system we write this vector as

$$\mathbf{v} = v^{i'} = \begin{bmatrix} \Delta x' \\ \Delta y' \end{bmatrix}.$$

Note that here we have both a symbolic notation ($\mathbf{v}$) and a component notation (v^i, $i = 1, 2$). The components depend on the chosen co-ordinate system ($\Delta x' \neq \Delta x$, $\Delta y' \neq \Delta y$, although always $\Delta x^2 + \Delta y^2 = \left(\Delta x'\right)^2 + \left(\Delta y'\right)^2 =$ constant — an *invariant*), but the symbolic notation does not depend on this. A vector is always the same thing, even if its co-ordinates transform with the co-ordinate system. A vector is always the same "arrow in space".

There exists the following transformation formula for changing the components of a vector:

$$v^{i'} = \sum_i \alpha_i^{i'} v^i, \tag{10.2}$$

[1] Georg Friedrich Bernhard RIEMANN (1826 – 1866), German mathematician.

[2] Albert EINSTEIN, 1879 – 1955, German-born theoretical physicist, discoverer of relativity theory, and icon of science.

or in "matrix language"

$$\mathbf{v}' = A\mathbf{v},$$

where

$$\mathbf{v} = v^i = \begin{bmatrix} v^1 \\ v^2 \end{bmatrix} \left(= \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right)$$

is the column matrix formed by the components, and

$$A = \alpha_i^{i'} = \begin{bmatrix} \alpha_1^1 & \alpha_2^1 \\ \alpha_1^2 & \alpha_2^2 \end{bmatrix} \left(= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \right)$$

is the component matrix of the transformation operator.

Every quantity that transforms according to equation (10.2) is called a *vector*. Familiar vector quantities are velocity, acceleration, force, ... what they have in common is, that they can be graphically presented as arrows.

Tensors

A *tensor* is just a square object — a matrix — which has the same transformation property from on co-ordinate system to another, but *for every index*:

$$T^{i'j'} = \sum_{i,j} \alpha_i^{i'} \alpha_j^{j'} T^{ij}.$$

In "matrix language" this is[3]

$$T' = ATA^T,$$

where A is the same as defined above.

We have already encountered many tensors in geodetic theory:

1. The inertial tensor of the Earth.
2. The gravity gradient or MARUSSI *tensor*
$$M = \begin{bmatrix} \frac{\partial \mathbf{g}}{\partial x} \\ \frac{\partial \mathbf{g}}{\partial y} \\ \frac{\partial \mathbf{g}}{\partial z} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 W}{\partial x^2} & \frac{\partial^2 W}{\partial x \partial y} & \frac{\partial^2 W}{\partial x \partial y} \\ \frac{\partial^2 W}{\partial y \partial x} & \frac{\partial^2 W}{\partial y^2} & \frac{\partial^2 W}{\partial x \partial z} \\ \frac{\partial^2 W}{\partial z \partial x} & \frac{\partial^2 W}{\partial z \partial y} & \frac{\partial^2 W}{\partial z^2} \end{bmatrix}.$$
3. The variance "matrix" is really a tensor: $\mathrm{Var}\,(\mathrm{x}) = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$, where σ_x and σ_y are the mean errors of the co-ordinates, i.e., the components of x, and σ_{xy} is the covariance.
4. In the earlier discussed surface theory of GAUSS, the first, $H = \begin{bmatrix} E & F \\ F & G \end{bmatrix}$, and the second, $B = \begin{bmatrix} e & f \\ f & g \end{bmatrix}$ fundamental form, as well as $C = H^{-1}B$;
5. Also the elements of the map projection fundamental form (to be discussed later) $\widetilde{E}, \widetilde{F}, \widetilde{G}$ constitute a tensor $\widetilde{H} = \begin{bmatrix} \widetilde{E} & \widetilde{F} \\ \widetilde{F} & \widetilde{G} \end{bmatrix}$, a metric tensor of sorts. The object $H^{-1}\widetilde{H} = \begin{bmatrix} \frac{\widetilde{E}}{M^2} & \frac{\widetilde{F}}{Mp} \\ \frac{\widetilde{F}}{Mp} & \frac{\widetilde{G}}{p^2} \end{bmatrix}$ may again be called the *scale tensor.*

[3] Verify that the matrix equation produces the same index summations as the index equation!

The geometric representation of a tensor and invariants

In the same way that the geometric depiction of a *vector* is an *arrow*, is the geometric depiction of a *tensor* an *ellipse* (in two dimensions) or an *ellipsoid* (in three dimensions).The lengths of the principal axes of the ellipse/ellipsoid depict the eigenvalues[4] of the tensor; the directions of the principal axes again are the directions of the tensor's eigenvectors.

In Euclidean space and in rectangular co-ordinates, tensors T^{ij} are typically *symmetric*; therefore, to different eigenvalues $\lambda_i, \lambda_j, i \neq j$ belong eigenvectors $\mathbf{x}_i, \mathbf{x}_j$ that are mutually perpendicular, as proven in mathematics textbooks.

In an n-dimensional space, a tensor T has n independent invariants. *The eigenvalues* λ_i, $i = 1, \ldots, n$ are of course invariants. So are their *sum* and *product.*

$$\sum_i \lambda_i = \sum_i T^{ii},$$

on the two-dimensional plane

$$\lambda_1 + \lambda_2 = T^{11} + T^{22},$$

the sum of the diagonal elements or *trace*[5]; ja

$$\prod_i \lambda_i = \det(T),$$

on the two-dimensional plane again

$$\lambda_1 \lambda_2 = \det(T) = T^{11}T^{22} - T^{12}T^{21},$$

the *determinant* of the tensor.

The trace of the variance matrix $\sigma_x^2 + \sigma_y^2$ we already know as the *point variance* σ_P^2; it was chosen precisely because it is an invariant, independent of the direction of the xy axes. Also the trace of the gravity gradient tensor is known:

$$\sum_i M_{ii} = \frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} \equiv \Delta W,$$

the LAPLACE operator.

Tensors in general co-ordinates

And what about *non-Euclidean spaces, with non-rectangular co-ordinate systems*? Well, in that case the difference between sub- and superscripts becomes meaningful, and the reason we write superscripts may finally be appreciated.

A *contravariant vector* transforms as follows:

$$v^{i'} = \sum_i \alpha_i^{i'} v^i \tag{10.3}$$

and a *covariant vector* as follows:

$$v_{i'} = \sum_i \alpha_{i'}^i v_i. \tag{10.4}$$

(These look very similar, but are not the same!)

[4]More precisely, the lengths of the semi-axes are the square roots of the eigenvalues λ_i.
[5]germ. *Spur.*

Where the "prototype" of a contravariant vector is the co-ordinate differences if two (nearby) points, as above:

$$v^i = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix},$$

is the prototype of a covariant vector the *gradient operator*:

$$v_j = \frac{\partial V}{\partial x^j} = \begin{bmatrix} \partial V/\partial x \\ \partial V/\partial y \end{bmatrix}. \tag{10.5}$$

$V(x, y)$ is some scalar field in space.

If we take for a model vector

$$v^i = \begin{bmatrix} dx \\ dy \end{bmatrix},$$

we obtain

$$v^{i'} = \begin{bmatrix} dx' \\ dy' \end{bmatrix} = \begin{bmatrix} \partial x'/\partial x & \partial x'/\partial y \\ \partial y'/\partial x & \partial y'/\partial y \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix};$$

also with the chain rule

$$\begin{bmatrix} \partial V/\partial x' \\ \partial V/\partial y' \end{bmatrix} = \begin{bmatrix} \partial x/\partial x' & \partial x/\partial y' \\ \partial y/\partial x' & \partial y/\partial y' \end{bmatrix} \begin{bmatrix} \partial V/\partial x \\ \partial V/\partial y \end{bmatrix}.$$

The coefficient matrices $\dfrac{\partial x^{i'}}{\partial x^i} = \alpha_i^{i'}$ and $\dfrac{\partial x^i}{\partial x^{i'}} = \alpha_{i'}^i$ are apparently *each other's inverse matrices.*

So, if the matrix form of the covariant transformation equation's (10.4) transformation parameters $\alpha_{i'}^i$ is A (i row and i' column index), then the matrix of the contravariant transformation parameters (equation 10.3) $\alpha_i^{i'}$ must be A^{-1} (i' row and i column index). From this circumstance, the names "covariant" and "contravariant" derive.

A tensor may have both super- and subscripts. Under a co-ordinate transformation, each index changes according to its "character". There may even be more than two indices.

"Trivial" tensors

1. If we compute the gradient of a contravariant vector x^i we obtain

$$\frac{\partial x^i}{\partial x^j} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \equiv \delta_j^i,$$

the so-called KRONECKER[6] delta, in practice a unit matrix. It is a tensor:

$$\delta_{j'}^{i'} = \alpha_{j'}^j \alpha_i^{i'} \delta_j^i = \sum_{i,j} \alpha_i^{i'} \delta_j^i \alpha_{j'}^j$$

or in matrix language

$$\begin{array}{ccccccc} i' \downarrow & & i' \downarrow & i \downarrow & j \downarrow & & \\ [I'] & = & [A^{-1}] & [I] & [A] & = A^{-1}A = I, \\ j' \rightarrow & & i \rightarrow & j \rightarrow & j' \rightarrow & & \end{array}$$

because if the matrix form of $\alpha_{j'}^j$ is A, then that of $\alpha_i^{i'}$ – or correspondingly, that of $\alpha_j^{j'}$ is A^{-1}, as shown above.

[6]Leopold Kronecker (1823 – 1891), German mathematician

2. The "corkscrew tensor" of Tullio LEVI-CIVITA[7] in three dimensions:

$$\epsilon_{ijk} = \begin{cases} 0 & \text{if, of } ijk\text{, two are the same} \\ 1 & \text{if } ijk \text{ is an even permutation of the numbers } (123) \\ -1 & \text{if } ijk \text{ is an odd permutation of the numbers } (123) \end{cases}$$

That is, $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, $\epsilon_{132} = \epsilon_{321} = \epsilon_{213} = -1$, all others $= 0$.

Ks. `http://folk.uio.no/patricg/teaching/a112/levi-civita/`.

10.2. The metric tensor

The metric tensor, or the *metric*, describes the PYTHAGORAS theorem in a curved space. It is identical to the earlier discussed Gaussian First Fundamental Form. In the ordinary plane ($\mathbb{R}^2$) we may choose rectangular co-ordinates (x, y), after which we may write the distance s between to points $1, 2$ as:

$$s^2 = \Delta x^2 + \Delta y^2,$$

where $\Delta x = x_2 - x_1, \Delta y = y_2 - y_1$ are the co-ordinate differences between the points. This equation applies throughout the plane. Its differential version looks the same:

$$ds^2 = dx^2 + dy^2.$$

This is now written into the following form:

$$ds^2 = g_{ij} dx^i dx^j, \tag{10.6}$$

where

$$g_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ ja } dx^i = dx^j = \begin{bmatrix} dx \\ dy \end{bmatrix}.$$

This is called the *metric* of the rectangular co-ordinate system in the Euclidean plane. In equation (10.6) it is assumed that we sum over the indices i and j; we call this the EINSTEIN *summation convention*. Always when in an equation we have a subscript and a superscript with the same name, we sum over it. I.e., in this case

$$ds^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} g_{ij} dx^i dx^j.$$

An alternative way of writing using matrix notation is a *quadratic form*:

$$\begin{aligned} ds^2 &= \langle \mathrm{dx} \cdot H \mathrm{dx} \rangle = \mathrm{dx}^T H \mathrm{dx} = \\ &= \overbrace{\begin{bmatrix} dx^1 & dx^2 \end{bmatrix}}^{\mathrm{dx}^T} \overbrace{\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}}^{H} \overbrace{\begin{bmatrix} dx^1 \\ dx^2 \end{bmatrix}}^{\mathrm{dx}}. \end{aligned}$$

If the surface is not curved, one may always find a co-ordinate system which is *everywhere* rectangular and both co-ordinates "scaled" correctly such that to a co-ordinate difference of 1 m corresponds also a location difference of 1 m. Then the g_{ij} matrix or *metric tensor* has the form of the unit matrix, like above.

If the surface is curved, we can find a unit matrix only in *some* places. E.g., on the surface of the Earth, only within a strip in the vicinity of the equator. It won't be possible in a unified manner over the whole surface of the Earth.

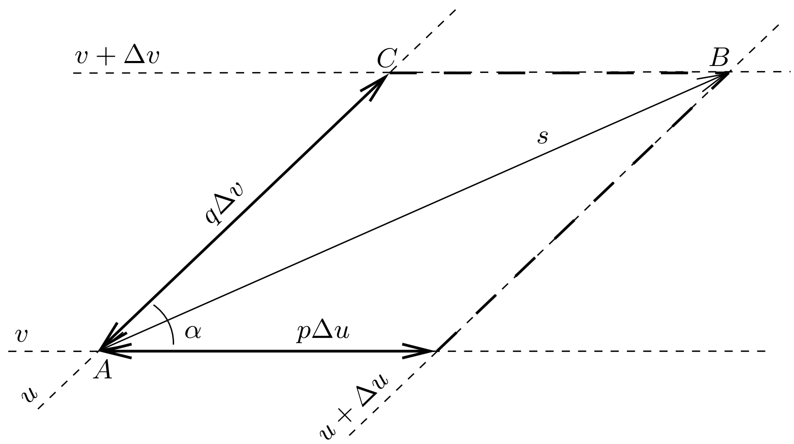


Figure 10.1.: Skewed metric in the plane

Nevertheless we an choose also on a non-curved surface a co-ordinate system that isn't rectangular but *skewed*, and where the co-ordinates are scaled arbitrarily. See figure 10.1.

In this case we find with the aid of the cosine rule:

$$s^2 = p^2\Delta u^2 + q^2\Delta v^2 + 2p\Delta u q\Delta v \cos\alpha,$$

or, differentially

$$ds^2 = p^2 du^2 + q^2 dv^2 + 2pq\cos\alpha du dv,$$

or, in index notation

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

where

$$g_{ij} = \begin{bmatrix} p^2 & pq\cos\alpha \\ pq\cos\alpha & q^2 \end{bmatrix}, \; dx^i = \begin{bmatrix} du \\ dv \end{bmatrix}.$$

The matrix representation:

$$ds^2 = \overbrace{\begin{bmatrix} du & dv \end{bmatrix}}^{dx^i} \overbrace{\begin{bmatrix} p^2 & pq\cos\alpha \\ pq\cos\alpha & q^2 \end{bmatrix}}^{g_{ij},\ \downarrow i,\ \to j} \overbrace{\begin{bmatrix} du \\ dv \end{bmatrix}}^{dx^j}.$$

On a *curved* surface g_{ij} will depend on place, $g_{ij}\left(x^i\right)$, where x^i is a "vector" of parameters describing the surface, $x^i = \begin{bmatrix} u & v \end{bmatrix}^T$. Also on a non-curved surface can g_{ij} depend on place, e.g., when choosing curvilinear, e.g., polar, co-ordinates.

Examples:

1. The surface of a spherical Earth:

$$ds^2 = R^2 d\varphi^2 + R^2\cos^2\varphi d\lambda^2,$$

i.e.,

$$g_{ij}(\varphi, \lambda) = \begin{bmatrix} R^2 & 0 \\ 0 & R^2\cos^2\varphi \end{bmatrix}, \; dx^i = \begin{bmatrix} d\varphi \\ d\lambda \end{bmatrix},$$

from which in "matrix language"

$$ds^2 = \overbrace{\begin{bmatrix} d\varphi & d\lambda \end{bmatrix}}^{dx^i} \overbrace{\begin{bmatrix} R^2 & 0 \\ 0 & R^2\cos^2\varphi \end{bmatrix}}^{g_{ij},\ \downarrow i,\ \to j} \overbrace{\begin{bmatrix} d\varphi \\ d\lambda \end{bmatrix}}^{dx^j}.$$

[7] Tullio LEVI-CIVITA (1873 – 1941), Italian mathematician.

2. Polar co-ordinates (ρ, θ) in the plane:

$$ds^2 = d\rho^2 + \rho^2 d\theta^2,$$

i.e.,

$$g_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & \rho^2 \end{bmatrix}, \; dx^i = \begin{bmatrix} d\rho \\ d\theta \end{bmatrix},$$

from which

$$ds^2 = \overbrace{\begin{bmatrix} d\rho & d\theta \end{bmatrix}}^{dx^i} \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & \rho^2 \end{bmatrix}}^{g_{ij},\, \downarrow i,\, \rightarrow j} \overbrace{\begin{bmatrix} d\rho \\ d\theta \end{bmatrix}}^{dx^j}.$$

3. In three dimensions *in the air space using "aviation co-ordinates"*: nautical miles North dN, nautical miles East dE, feet above sea level dH; in metres

$$ds^2 = (1852)^2\, dN^2 + (1852)^2\, dE^2 + (0.3048)^2\, dH^2,$$

eli

$$g_{ij} = \begin{bmatrix} 1852^2 & 0 & 0 \\ 0 & 1852^2 & 0 \\ 0 & 0 & 0.3048^2 \end{bmatrix} dx^i = \begin{bmatrix} dN \\ dE \\ dH \end{bmatrix},$$

ja

$$ds^2 = \overbrace{\begin{bmatrix} dN & dE & dH \end{bmatrix}}^{dx^i} \overbrace{\begin{bmatrix} 1852^2 & 0 & 0 \\ 0 & 1852^2 & 0 \\ 0 & 0 & 0.3048^2 \end{bmatrix}}^{g_{ij},\, \downarrow i,\, \rightarrow j} \overbrace{\begin{bmatrix} dN \\ dE \\ dH \end{bmatrix}}^{dx^j}.$$

10.3. The inverse metric tensor

The inverse tensor of the metric tensor g_{ij} is written g^{ij}. As a matrix, it is the inverse matrix of g_{ij}:

$$g^{ij} = (g_{ij})^{-1},$$

or, in index notation

$$g^{ij} g_{jk} = \delta^i_k.$$

This is nothing but the *definition* of the inverse matrix:

$$H^{-1} H = I,$$

where I is the unit matrix, the matrix representation of the KRONECKER tensor δ^i_k. Written out:

$$\overbrace{\begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}}^{g^{ij},\, \downarrow i,\, \rightarrow j} \overbrace{\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}}^{g_{jk},\, \downarrow j,\, \rightarrow k} = \overbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}^{\delta^i_k,\, \downarrow i,\, \rightarrow k}.$$

10.3.1. Raising or lowering sub- or superscripts of a tensor

The sub- or superscript of an arbitrary tensor can be "raised" or "lowered" by multiplying with the metric tensor g_{ij} or inverse tensor g^{ij}:

$$T^i_j = g^{ik} T_{kj} = g_{jk} T^{ki}.$$

All forms T_{ij}, T^{ij}, T^i_j designate the same tensor, written in different ways. As a special case, $\delta^i_j = g_{jk} g^{ki} = g^{ik} g_{kj}$, i.e., the KRONECKER delta tensor is a "mixed form" of the metric tensor and could be written g^i_j. The delta style of writing has established itself, however.

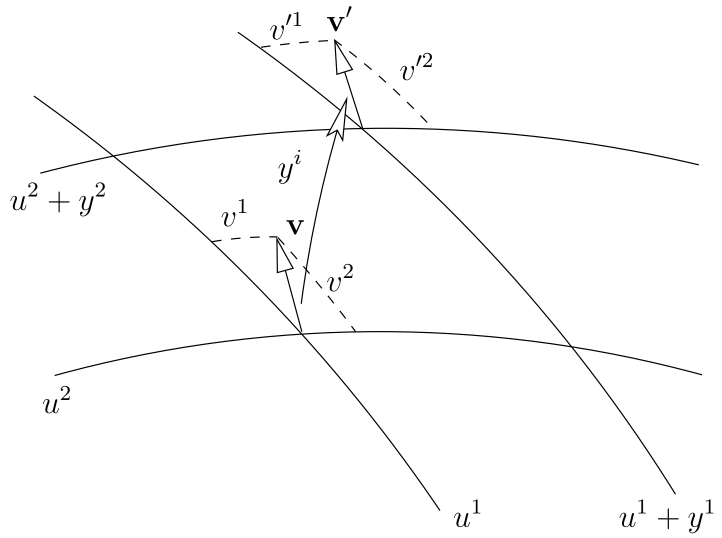


Figure 10.2.: CHRISTOFFEL symbols and parallel transport

10.3.2. The eigenvalues and -vectors of a tensor

The eigenvalue problem for a square tensor has the following form:

$$\left(T^{ij} - \lambda g^{ij}\right) x_j = 0,$$
$$\left(T_{ij} - \lambda g_{ij}\right) x^j = 0,$$
$$\left(T^i_j - \lambda \delta^i_j\right) x^j = \left(T^i_j - \lambda \delta^i_j\right) x_i = 0.$$

All three forms are equivalent, which is easily proved. For eigenvectors we have $x_i = g_{ij}x^j$. If the tensor T^{ij} (or T^i_j, or T_{ij}) is symmetric (meaning $T^{ij} = T^{ji}$ etc.), then the eigenvalues λ are real and the eigenvectors mutually orthogonal: if x^i, y^i are different eigenvectors, then

$$g_{ij}x^i y^j = 0.$$

There are as many eigenvalues as there are dimensions in the space, i.e., in the plane $\mathbb{R}^2$ two.

10.3.3. The graphic representation of a tensor

The quadratic form

$$T_{ij}x^i x^j = 1$$

defines an ellipsoid (in $\mathbb{R}^2$ an ellipse) that may be considered the graphic of T_{ij}. E.g., the inertial ellipsoid, the variance ellipsoid.

10.4. The Christoffel symbols

The metric tensor isn't yet the same thing as curvature. It isn't even the same thing as the curvature of the co-ordinate curves; to study this, we need the *Christoffel*[8] *symbols*.

The Christoffel symbols describe what happens to a *vector* when it is *transported parallelly* along the surface; more precisely, what happens to its components.

See figure 10.2. When a vector $\mathbf{v}$, the components of which are v^i, is transported parallelly from one point to another over a distance y^i, its components will change by amounts $\Delta v^i = v'^i - v^i$,

[8] Elwin Bruno CHRISTOFFEL (1829–1900), German mathematician-physicist

although for the "vector itself" $\mathbf{v}' = \mathbf{v}$. When both the transported vector and the distance of transport are *small*, we may assume that the change depends *linearly* on both the vector and the direction of transport, in the following way:

$$\Delta v^i = \Gamma^i_{jk} v^j y^k.$$

The Christoffel symbols Γ^i_{jk} do not form a tensor *like the metric tensor does. They describe the "curvilinearity" of a curvilinear co-ordinate system, i.e., a property of the co-ordinate system. A co-ordinate transformation that removes – at least locally if not everywhere – the curvature of the co-ordinate curves, also makes the elements of* Γ^i_{jk} vanish in that point (for tensors, such a local "transforming away" will never succeed!).

E.g., on the Earth surface in the (φ, λ) co-ordinate system on the equator the co-ordinate curves are locally non-curved and later we shall show, theat there indeed all Γ^i_{jk} vanish. Another example is from the general theory of relativity, where the components of *acceleration* are CHRISTOFFEL symbols: acceleration can be transformed away by moving to a "falling along" reference system. In EINSTEIN's falling elevator the people inside the elevator accelerate with respect the Earth surface but are weightless (i.e., the acceleration vanishes) in a reference system connected to the elevator.

The Christoffel symbols can be computed from the metric tensor; the equation is (complicated proof; see appendix C):

$$\Gamma^i_{jk} = \frac{1}{2} g^{i\ell} \left(\frac{\partial g_{k\ell}}{\partial x^j} + \frac{\partial g_{\ell j}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^\ell} \right). \tag{10.7}$$

$(g^{ij} \equiv (g_{ij})^{-1}.)$ From this is can be seen, that the Christoffel symbols are, like, the first derivatives of the metric. In a straight-lined co-ordinate system on a non-curved surface g_{ij} is a constant and thus all Γ^i_{jk} vanish. Also, because g_{ij} is symmetric $(g_{ij} = g_{ji})$, we obtain

$$\Gamma^i_{jk} = \Gamma^i_{kj}.$$

The Christoffel symbols are useful when writing the *equation of the geodesic*[9]in this formalism (cf. equation (9.8)):

$$\frac{d}{ds} t^i + \Gamma^i_{jk} t^j t^k = 0.$$

Here, t^i is the *tangent vector* of the geodesic $t^i = {}^{dx^i}\!/_{ds}$.

Examples:

1. On the surface of the spherical Earth (φ, λ). We use a notation where φ and λ symbolize the index values 1 and 2 (for these symbols EINSTEIN's summation convention thus fails to work!):

$$\frac{\partial g_{k\ell}}{\partial \varphi} = \begin{bmatrix} 0 & 0 \\ 0 & -R^2 \sin 2\varphi \end{bmatrix}, \quad \frac{\partial g_{k\ell}}{\partial \lambda} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

so only

$$\frac{\partial g_{\lambda\lambda}}{\partial \varphi} = \frac{\partial g_{22}}{\partial \varphi} = -R^2 \sin 2\varphi$$

[9] In fact we write

$$\frac{Dt^i}{ds} \equiv \frac{dt^i}{ds} + \Gamma^i_{jk} t^j t^k,$$

the so-called *absolute* or *covariant derivative* which is a tensor; and

$$\frac{Dt^i}{ds} = 0$$

is then the equation for the geodesic, which thus applies in curvilinear co-ordinates.

is non-vanishing. Then, remembering that

$$g^{i\ell} = (g_{i\ell})^{-1} = \begin{bmatrix} R^2 & 0 \\ 0 & R^2\cos^2\varphi \end{bmatrix}^{-1} = \begin{bmatrix} R^{-2} & 0 \\ 0 & R^{-2}\cos^{-2}\varphi \end{bmatrix},$$

the only non-zero elements are:

$$\Gamma^{\varphi}_{\lambda\lambda} = \frac{1}{2}(g_{\varphi\varphi})^{-1}\left(-\frac{\partial g_{\lambda\lambda}}{\partial\varphi}\right) == -\frac{1}{2}R^{-2}\cdot R^2\sin 2\varphi = +\frac{1}{2}\sin 2\varphi = \sin\varphi\cos\varphi,$$

$$\Gamma^{\lambda}_{\varphi\lambda} = \frac{1}{2}(g_{\lambda\lambda})^{-1}\left(\frac{\partial g_{\lambda\lambda}}{\partial\varphi}\right) = \frac{1}{2}\left(R^{-2}\cos^{-2}\varphi\right)\left(-R^2\sin 2\varphi\right) = -\frac{\sin 2\varphi}{2\cos^2\varphi} = -\tan\varphi,$$

$$\Gamma^{\lambda}_{\lambda\varphi} = \frac{1}{2}(g_{\lambda\lambda})^{-1}\left(\frac{\partial g_{\lambda\lambda}}{\partial\varphi}\right) = -\tan\varphi.$$

Note *that at the equator* $\varphi = 0$ all $\Gamma^i_{jk} = 0$.

2. Polar co-ordinates in the plane (ρ, θ):

$$\frac{\partial g_{k\ell}}{\partial\rho} = \begin{bmatrix} 0 & 0 \\ 0 & 2\rho \end{bmatrix}, \quad \frac{\partial g_{k\ell}}{\partial\theta} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

yielding

$$\frac{\partial g_{22}}{\partial\rho} = \frac{\partial g_{\theta\theta}}{\partial\rho} = 2\rho,$$

so

$$\begin{aligned} \Gamma^{\rho}_{\theta\theta} &= \frac{1}{2}(g_{\rho\rho})^{-1}\left(-\frac{\partial g_{\theta\theta}}{\partial\rho}\right) = \frac{1}{2}\cdot -2\rho = -\rho, \\ \Gamma^{\theta}_{\rho\theta} = \Gamma^{\theta}_{\theta\rho} &= \frac{1}{2}(g_{\theta\theta})^{-1}\left(+\frac{\partial g_{\theta\theta}}{\partial\rho}\right) = \frac{1}{2}\cdot\frac{1}{\rho^2}\cdot 2\rho = +\frac{1}{\rho}. \end{aligned}$$

10.5. The geodesic revisited

Here are alternative formulas for integrating the geodesic on the sphere (generalization to the *ellipsoid of revolution* is complicated but possible):

$$\begin{aligned} \frac{d\xi}{ds} + \eta^2\sin\varphi\cos\varphi &= 0, \\ \frac{d\eta}{ds} - 2\eta\xi\tan\varphi &= 0. \end{aligned}$$

Here we have already used the formulas derived above for the elements of Γ^i_{jk}.

Simultaneous integration of the differential equations would give

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} = \begin{bmatrix} \dfrac{\cos A}{R} \\ \dfrac{\sin A}{R\cos\varphi} \end{bmatrix},$$

as a function of s, from which A and φ may be computed.

To this set we add the definition equations of the tangent

$$\begin{aligned} \frac{d\varphi}{ds} = \xi &= t^1, \\ \frac{d\lambda}{ds} = \eta &= t^2, \end{aligned}$$

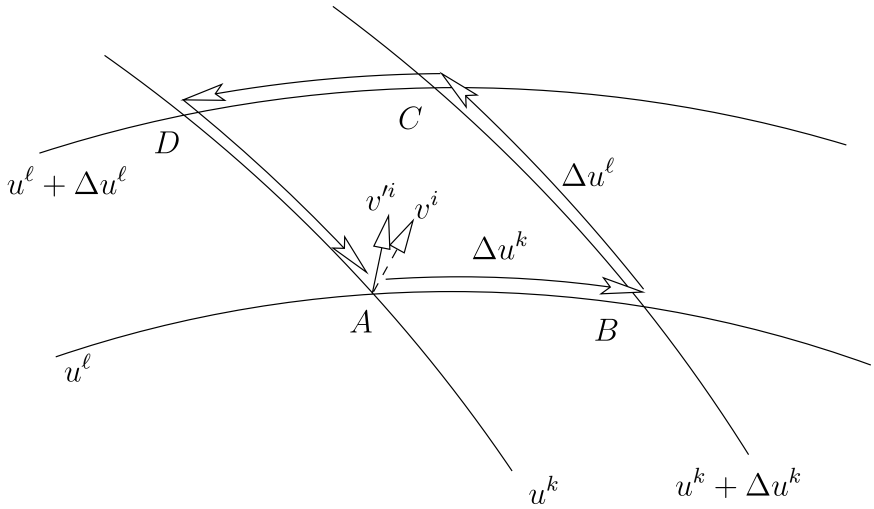


Figure 10.3.: The curvature tensor and parallel transport around a closed grid path

In this way also λ comes along.

Perhaps this approach seems overly complicated; its major theoretical advantage is, that it will work for *all* curvilinear co-ordinate systems on the surface considered, also, e.g., for stereographic map projection co-ordinates used in the polar areas, as long as we first manage to write down the *metric tensor* of the co-ordinate curves.

The *length* of the tangent vector $\begin{bmatrix} \xi & \eta \end{bmatrix}^T$ is computed as follows:

$$ds^2 = g_{ij}t^i t^j = R^2 \left(\frac{\cos A}{R}\right)^2 + R^2 \cos^2\varphi \left(\frac{\sin A}{R\cos\varphi}\right)^2 = 1,$$

if we remember that

$$g_{ij} = \begin{bmatrix} R^2 & 0 \\ 0 & R^2\cos^2\varphi \end{bmatrix}$$

on the surface of the sphere, The lengths of the tangent vectors have always to be 1; this requirement is fulfilled if s is a parametrization of the curve "by distance".

10.6. The curvature tensor

The *curvature* is described again in a slightly more complicated way by means of *parallel transport around a closed path* — a small rectangle[10].

As seen from figure 10.3 we obtain the change $\Delta v^i \equiv \left(v'\right)^i - v^i$ of the vector v^i, which depends on

1. the orientation of the closed rectangular path, the side indices k and ℓ;
2. the size of the rectangle, measured in curvilinear co-ordinates: let the length of the u^k side beΔu^k and the length of the u^ℓ side, Δu^ℓ;
3. the transported vector v^i itself.

In the following way:

$$\Delta v^i = R^i_{jk\ell} v^j \Delta u^k \Delta u^\ell. \tag{10.8}$$

Here, the creature $R^i_{jk\ell}$ is called the RIEMANN *curvature tensor*. In two-dimensional space (i.e., on a surface) it has $2^4 = 16$ elements.

[10]More precisely, a parallellogram

We can compute RIEMANN from the CHRISTOFFELS[11]:

$$R^i_{jk\ell} = \frac{\partial}{\partial x^k}\Gamma^i_{j\ell} - \frac{\partial}{\partial x^\ell}\Gamma^i_{jk} + \Gamma^i_{km}\Gamma^m_{j\ell} - \Gamma^i_{\ell m}\Gamma^m_{jk}. \tag{10.9}$$

From this we spot immediately the following antisymmetry:

$$R^i_{jk\ell} = -R^i_{j\ell k}.$$

There exist many more such symmetries; the number of independent components is actually small.

Esimerkkejä:

1. Surface of a spherical Earth (φ, λ):

   Based on the antisymmetry property we may say that only elements for which $k \neq l$ can differ from zero. Additionally the CHRISTOFFEL symbols depend only on φ. Thus we obtain the auxiliary terms:

$$\begin{aligned} \frac{\partial}{\partial\varphi}\Gamma^\varphi_{\lambda\lambda} &= \frac{\partial}{\partial\varphi}\left(\frac{1}{2}\sin 2\varphi\right) = +\cos 2\varphi, \\ \frac{\partial}{\partial\varphi}\Gamma^\lambda_{\varphi\lambda} = \frac{\partial}{\partial\varphi}\Gamma^\lambda_{\lambda\varphi} &= \frac{\partial}{\partial\varphi}\left(-\tan\varphi\right) = -\frac{1}{\cos^2\varphi}, \end{aligned}$$

   the others zero.

   The terms $\Gamma^i_{km}\Gamma^m_{jl}$ are obtained in the following way:

$$\begin{aligned} \Gamma^\varphi_{\lambda\lambda}\Gamma^\lambda_{\varphi\lambda} = \Gamma^\lambda_{\lambda\varphi}\Gamma^\varphi_{\lambda\lambda} = \Gamma^\varphi_{\lambda\lambda}\Gamma^\lambda_{\lambda\varphi} &= -\sin^2\varphi, \\ \Gamma^\lambda_{\varphi\lambda}\Gamma^\lambda_{\lambda\varphi} = \Gamma^\lambda_{\varphi\lambda}\Gamma^\lambda_{\varphi\lambda} &= +\tan^2\varphi. \end{aligned}$$

   By combining

$$\begin{aligned} R^\varphi_{\varphi\varphi\lambda} = -R^\varphi_{\varphi\lambda\varphi} &= \frac{\partial}{\partial\varphi}\Gamma^\varphi_{\varphi\lambda} - \frac{\partial}{\partial\lambda}\Gamma^\varphi_{\varphi\varphi} + \Gamma^\varphi_{\varphi m}\Gamma^m_{\varphi\lambda} - \Gamma^\varphi_{\lambda m}\Gamma^m_{\varphi\varphi} = 0; \\ R^\lambda_{\varphi\varphi\lambda} = -R^\lambda_{\varphi\lambda\varphi} &= \frac{\partial}{\partial\varphi}\Gamma^\lambda_{\varphi\lambda} - \frac{\partial}{\partial\lambda}\Gamma^\lambda_{\varphi\varphi} + \Gamma^\lambda_{\varphi m}\Gamma^m_{\varphi\lambda} - \Gamma^\lambda_{\lambda m}\Gamma^m_{\varphi\varphi} = -\frac{1}{\cos^2\varphi} + \tan^2\varphi = -1; \\ R^\varphi_{\lambda\varphi\lambda} = -R^\varphi_{\lambda\lambda\varphi} &= \frac{\partial}{\partial\varphi}\Gamma^\varphi_{\lambda\lambda} - \frac{\partial}{\partial\lambda}\Gamma^\varphi_{\lambda\varphi} + \Gamma^\varphi_{\varphi m}\Gamma^m_{\lambda\lambda} - \Gamma^\varphi_{\lambda m}\Gamma^m_{\lambda\varphi} = \cos 2\varphi + \sin^2\varphi = \cos^2\varphi; \\ R^\lambda_{\lambda\varphi\lambda} = -R^\lambda_{\lambda\lambda\varphi} &= \frac{\partial}{\partial\varphi}\Gamma^\lambda_{\lambda\lambda} - \frac{\partial}{\partial\lambda}\Gamma^\lambda_{\lambda\varphi} + \Gamma^\lambda_{\varphi m}\Gamma^m_{\lambda\lambda} - \Gamma^\lambda_{\lambda m}\Gamma^m_{\lambda\varphi} = 0. \end{aligned}$$

2. Polar co-ordinates (ρ, θ) plane:

$$\begin{aligned} \frac{\partial}{\partial\rho}\Gamma^\rho_{\theta\theta} &= -1, \\ \frac{\partial}{\partial\rho}\Gamma^\theta_{\rho\theta} = \frac{\partial}{\partial\rho}\Gamma^\theta_{\theta\rho} &= -\frac{1}{\rho^2}, \end{aligned}$$

   the others again vanish. Then:

$$\begin{aligned} \Gamma^\rho_{\theta\theta}\Gamma^\theta_{\rho\theta} = \Gamma^\theta_{\theta\rho}\Gamma^\rho_{\theta\theta} = \Gamma^\rho_{\theta\theta}\Gamma^\theta_{\theta\rho} &= -1, \\ \Gamma^\theta_{\rho\theta}\Gamma^\theta_{\theta\rho} = \Gamma^\theta_{\rho\theta}\Gamma^\theta_{\rho\theta} &= +\frac{1}{\rho^2}. \end{aligned}$$

[11] The derivation is found in appendix D.

After this:

$$\begin{aligned}
R^{\rho}_{\ \rho\rho\theta} = -R^{\theta}_{\ \rho\theta\rho} &= 0;\\
R^{\theta}_{\ \rho\rho\theta} = -R^{\theta}_{\ \rho\theta\rho} &= \frac{\partial}{\partial\rho}\Gamma^{\theta}_{\ \rho\theta} - \frac{\partial}{\partial\theta}\Gamma^{\theta}_{\ \rho\rho} + \Gamma^{\theta}_{\ \rho m}\Gamma^{m}_{\ \rho\theta} - \Gamma^{\theta}_{\ \theta m}\Gamma^{m}_{\ \rho\rho} = -\frac{1}{\rho^2} + \frac{1}{\rho^2} = 0;\\
R^{\rho}_{\ \theta\rho\theta} = -R^{\rho}_{\ \theta\theta\rho} &= \frac{\partial}{\partial\rho}\Gamma^{\rho}_{\ \theta\theta} - \frac{\partial}{\partial\theta}\Gamma^{\rho}_{\ \theta\rho} + \Gamma^{\rho}_{\ \rho m}\Gamma^{m}_{\ \theta\theta} - \Gamma^{\rho}_{\ \theta m}\Gamma^{m}_{\ \theta\rho} = -1 + 1 = 0;\\
R^{\theta}_{\ \theta\rho\theta} = -R^{\theta}_{\ \theta\theta\rho} &= 0.
\end{aligned}$$

So *the whole* RIEMANN *tensor vanishes*, as it should, because the surface is not curved.

From the RIEMANN tensor we obtain the smaller RICCI[12]-*tensor* in the following way:

$$R_{jk} = R^{i}_{\ jik} = \sum_{i=1}^{2} R^{i}_{\ jik}. \tag{10.10}$$

This is a symmetric tensor, $R_{ij} = R_{ji}$.

Examples

1. Let us continue with the computation of R_{ij} for the case of a spherical surface:

$$\begin{aligned}
R_{\varphi\varphi} &= R^{\varphi}_{\ \varphi\varphi\varphi} + R^{\lambda}_{\ \varphi\lambda\varphi} = +1;\\
R_{\varphi\lambda} &= R^{\varphi}_{\ \varphi\varphi\lambda} + R^{\lambda}_{\ \varphi\lambda\lambda} = 0 + 0 = 0;\\
R_{\lambda\lambda} &= R^{\lambda}_{\ \lambda\lambda\lambda} + R^{\varphi}_{\ \lambda\varphi\lambda} = \cos^2\varphi.
\end{aligned}$$

2. For polar co-ordinates in the plane, all $R_{ij} = 0$.

We may continue this process to obtain the *curvature scalar*[13]:

$$R = g^{ij}R_{ji} = \sum_{j=1}^{2} (g_{ij})^{-1} R_{ji}. \tag{10.11}$$

For the example case:

1. Spherical surface:

$$R^{i}_{k} = g^{ij}R_{jk} = \begin{bmatrix} \dfrac{1}{R^2} & 0 \\ 0 & \dfrac{1}{R^2\cos^2\varphi} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & +\cos^2\varphi \end{bmatrix} = \begin{bmatrix} R^{-2} & 0 \\ 0 & R^{-2} \end{bmatrix},$$

i.e.[14]

$$R = \sum_i R^{i}_{i} = \frac{1}{R^2} + \frac{\cos^2\varphi}{R^2\cos^2\varphi} = \frac{2}{R^2}.$$

2. In polar co-ordinates $R_{ij} = 0$, i.e., $R = 0$.

More generally we have (without proof):

$$R = 2K = 2\kappa_1\kappa_2 = \frac{2}{R_1R_2}, \tag{10.12}$$

twice the GAUSS total curvature K, the inverse of the product of the two principal radii of curvature.

[12] Gregorio RICCI-CURBASTRO (1853 – 1925), Italian mathematician, inventor of tensor calculus

[13] This is in fact the "trace" of R_{ij}, more precisely, of $R^{i}_{j} \equiv g^{ik}R_{kj}$, written R^{i}_{i}, see above. This also serves as an example of how in general curvilinear co-ordinates an index may be "raised" from covariant to contravariant, or "lowered", using the metric tensor g_{ij} or its inverse g^{ij}. In rectangular co-ordinates these are unit matrices and the difference between super- and subscripts is without consequence.

[14] Here we use the symbol R both for the radius of the Earth and for the curvature scalar. We hope it doesn't cause confusion.

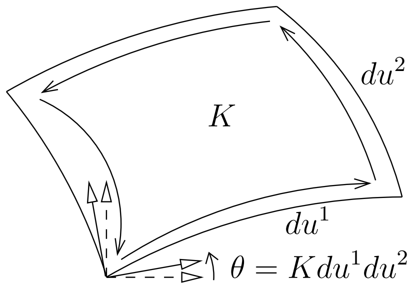


Figure 10.4.: The curvature of a two-dimensional surface may be characterized by just one parameter: θ.

10.7. Gauss curvature and spherical excess

Contrary to the eigenvalues of the GAUSS second fundamental form $\beta^i_k = g^{ij}\beta_{jk}$ (chapter 9.4) are the eigenvalues of the tensor R^i_j *unrelated* to the principal curvatures of the surface[15] κ_1 and κ_2. In the *two-dimensional case* both the R_{ij} tensor and the R^i_{jkl} have *only one essentially independent element*, which is related to the GAUSS total curvature $K = \kappa_1\kappa_2$.

This is not hard to show:

1. in the equation (10.8) $\Delta u\Delta v$ describe the sides of a small diamond shape, around which the vector v^i is transported in a parallel fashion. In two dimensions there are only two choices for the sides in the u^k, u^ℓ directions: $\Delta u^k\Delta u^\ell$ and $\Delta u^\ell\Delta u^k$. One represents clockwise transport, the other, counterclockwise. The corresponding elements $R^i_{jk\ell} = -R^i_{j\ell k}$ thus are essentially the same.
2. In the same equation v^j and Δv^i are mutually perpendicular. Δv^i represents a small *rotation* of vector v^j. On a surface, in two dimensions, the rotation is described by *one angle*. Because the angle between two parallelly transported vectors doesn't change, we may infer that this rotation angle is the same for all vectors v^i. Again we find only one independent parameter. See figure 10.4.

   In other words, when give a $\Delta u^1\Delta u^2$ diamond shape, the expression $R^i_{j12}\Delta u^1\Delta u^2 = -R^i_{j21}\Delta u^2\Delta u^1$ is a 2×2 rotation matrix $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, containing only one free parameter.

However, for higher dimensionalities, the number of independent elements in the tensors of RIEMANN and RICCI is larger. This case is interesting because of General Relativity (four dimensions!) though not for geodesy.

We can remark here, that the *spherical excess* already discussed in section 1.2 is a special case of the change in direction of a vector that is parallelly transported around: the spherical excess is the small change of direction of a vector transported around a closed *triangle*!

When we transport a vector around a larger surface area, it is the same as if we had transported it successively around all of the little patches making up the surface area. This is depicted in figure 10.5. In every little patch du^1du^2, area dS, the change in direction of the vector amounts to KdS, where K is the GAUSS total curvature at the patch location.

By generalization one obtains from this the following integral equation (GAUSS-BONNET[16] for the triangle, *Theorema elegantissimum*):

$$\varepsilon = \iint_\Delta K(\varphi, \lambda)\, dS,$$

[15] The *radii* of curvature exist only in the three-dimensional space surrounding the Earth surface, in which it is embedded. R on the other hand is an *intrinsic* property of the surface.

[16] Pierre Ossian BONNET (1819 – 1892), French mathematician

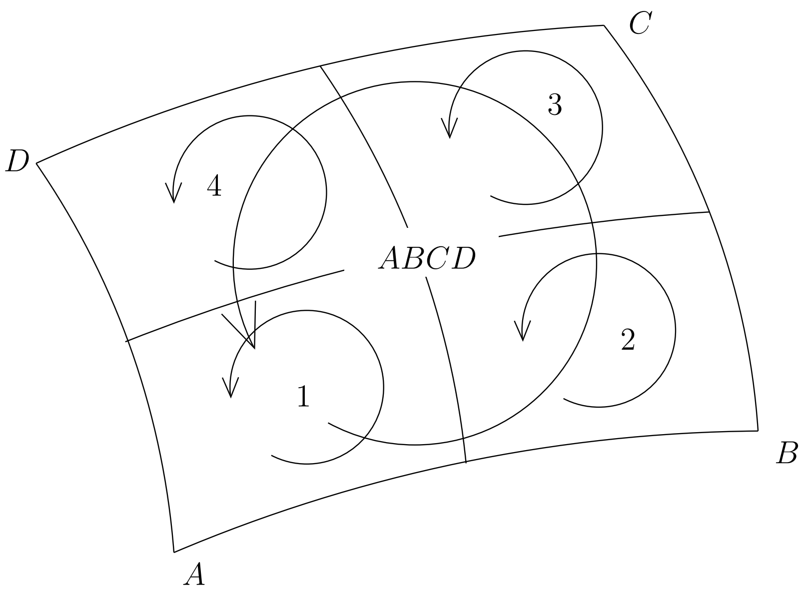


Figure 10.5.: Transport of a vector around a larger surface area. $\theta_{ABCD} = \theta_1 + \theta_2 + \theta_3 + \theta_4$.

where K is a function of place.

On the reference ellipsoid $K = (MN)^{-1}$ and

$$\begin{aligned} \varepsilon = \iint_{\Delta} \frac{1}{MN} dS &= \iint_{\Delta} \frac{1}{MN} MN \cos\varphi d\varphi d\lambda = \\ &= \iint_{\Delta} d\sigma = \sigma_{\Delta}, \end{aligned}$$

which is the surface area of the corresponding triangle, but on the *unit sphere*.[17].

From this follows that

> on the reference ellipsoid the spherical excess depends only on the *directions* of the normals in the corner points (φ_i, λ_i), $i = 1, 2, 3$, not on the shape of the ellipsoidal surface. *Shorter:* spherical excesses are computed on the sphere, even while all other computations are done on the ellipsoid.

It suffices that the geographical co-ordinates of the corner points, which describe the ellipsoidal normal, are known

A fast geometric way of computing the spherical excess of a triangle is the following:

Let the corner points

$$\mathrm{x}_i = \begin{bmatrix} \cos\varphi_i \cos\lambda_i \\ \cos\varphi_i \sin\lambda_i \\ \sin\varphi_i \end{bmatrix}, \ i = 1, 2, 3;$$

we use the *polarization method* (1.7) in three dimensions, in order to find the "poles" of the triangle's sides:

$$\mathrm{y}_1 = \frac{\langle \mathbf{x}_2 \times \mathbf{x}_3 \rangle}{\|\mathbf{x}_2 \times \mathbf{x}_3\|}, \ \mathbf{y}_2 = \frac{\langle \mathbf{x}_3 \times \mathbf{x}_1 \rangle}{\|\mathbf{x}_3 \times \mathbf{x}_1\|}, \ \mathbf{y}_3 = \frac{\langle \mathbf{x}_1 \times \mathbf{x}_2 \rangle}{\|\mathbf{x}_1 \times \mathbf{x}_2\|}.$$

[17] This is not *quite exactly* true... the triangle symbols under the integral sign $d\varphi d\lambda$ and under the integral sign $d\sigma$ are not exactly corresponding. On the reference elipsoid the triangle consists of geodesics, on the unit sphere, of great circle segments; however, a geodesic on the ellipsoid *does not map* to a great circle!

This may be suspected already from the observation, that a long geodesic around the ellipsoid does not generally close upon itself, while a great circle around a sphere always does.

Furthermore, the *angles* of an ellipsoidal and a spherical triangle are not individually equal; their *sums* (and with that, their spherical excesses) however are.

Now the inter-pole distances correspond to the angles of the spherical triangle (more precisely, π minus those angles):

$$\begin{aligned}\alpha_1 &= \pi - 2\arctan\frac{\|\mathbf{y}_2-\mathbf{y}_3\|}{\|\mathbf{y}_2+\mathbf{y}_3\|},\\ \alpha_2 &= \pi - 2\arctan\frac{\|\mathbf{y}_1-\mathbf{y}_3\|}{\|\mathbf{y}_1+\mathbf{y}_3\|},\\ \alpha_3 &= \pi - 2\arctan\frac{\|\mathrm{y}_1-\mathbf{y}_2\|}{\|\mathrm{y}_1+\mathbf{y}_2\|},\end{aligned}$$

numerically strong equations[18]. Here α_i is the angle at corner point i. The spherical excess is now

$$\varepsilon = \sum_{i=1}^{3}\alpha_i - \pi.$$

10.8. The curvature in quasi-Euclidean geometry

We mention without proof that in a RIEMANN space we may *transform* curvilinear co-ordinated always in such a way, that in a certain point P

1. the metric tensor is the unit matrix,

$$g_{ij} = g^{ij} = \begin{cases} 1 \text{ if } i=j \\ 0 \text{ otherwise,}\end{cases}$$

2. the metric tensor is (locally) *stationary*, i.e.

$$\left.\frac{\partial g_{ij}}{\partial x^k}\right|_{x^k=x^k_P} = 0.$$

In this case we speak of a *quasi-Euclidean neighbourhood* around P.

In this case the CHRISTOFFEL symbols (equation 10.7) are

$$\begin{aligned}\Gamma^i_{jk} &= \frac{1}{2}g^{i\ell}\left(\frac{\partial g_{k\ell}}{\partial x^j}+\frac{\partial g_{\ell j}}{\partial x^k}-\frac{\partial g_{jk}}{\partial x^\ell}\right) = \\ &= \frac{1}{2}\left(\frac{\partial g_{ki}}{\partial x^j}+\frac{\partial g_{ij}}{\partial x^k}-\frac{\partial g_{jk}}{\partial x^i}\right)\approx 0\end{aligned}$$

based on the above assumption of stationarity.

Of the RIEMANN curvature tensor (equation 10.9) the last two terms vanish:

$$\begin{aligned}R^i_{jk\ell} &= \frac{\partial}{\partial x^k}\Gamma^i_{j\ell} - \frac{\partial}{\partial x^\ell}\Gamma^i_{jk} = \\ &= \frac{1}{2}\frac{\partial}{\partial x^k}\left(\frac{\partial g_{\ell i}}{\partial x^j}+\frac{\partial g_{ij}}{\partial x^\ell}-\frac{\partial g_{j\ell}}{\partial x^i}\right) - \\ &- \frac{1}{2}\frac{\partial}{\partial x^\ell}\left(\frac{\partial g_{ki}}{\partial x^j}+\frac{\partial g_{ij}}{\partial x^k}-\frac{\partial g_{jk}}{\partial x^i}\right) = \\ &= \frac{1}{2}\frac{\partial}{\partial x^j}\left(\frac{\partial g_{\ell i}}{\partial x^k}-\frac{\partial g_{ki}}{\partial x^\ell}\right) - \frac{1}{2}\frac{\partial}{\partial x^i}\left(\frac{\partial g_{j\ell}}{\partial x^k}-\frac{\partial g_{jk}}{\partial x^\ell}\right).\end{aligned}$$

[18] In some programming languages, we have for $\arctan\left({}^{x}\!/_{y}\right)$ the symmetrical alternative form $\operatorname{atan2}(x,y)$, where the zerodivide problem does not occur.

If we now derive the RICCI tensor (equation 10.10):

$$
\begin{aligned}
R_{j\ell} &= R^i_{ji\ell} = -\frac{1}{2}\sum_i \frac{\partial^2 g_{j\ell}}{(\partial x^i)^2} + \\
&+ \frac{1}{2}\sum_i \left(\frac{\partial^2 g_{\ell i}}{\partial x^i \partial x^j} - \frac{\partial^2 g_{ii}}{\partial x^\ell \partial x^j} + \frac{\partial^2 g_{ji}}{\partial x^\ell \partial x^i} \right) = \\
&= -\frac{1}{2}\Delta g_{j\ell} + \frac{1}{2}\sum_i \left(\frac{\partial^2 g_{\ell i}}{\partial x^i \partial x^j} + \frac{\partial^2 g_{ij}}{\partial x^i \partial x^\ell} - \frac{\partial^2 g_{ii}}{\partial x^\ell \partial x^j} \right).
\end{aligned}
$$

This already looks a lot more symmetric. The symbol Δ here means the LAPLACE operator

$$
\Delta = \sum_i \frac{\partial^2}{(\partial x^i)^2}.
$$

Finally we still derive the curvature scalar

$$
\begin{aligned}
R &= g^{jk} R_{kj} = \sum_j R_{jj} = \\
&= -\frac{1}{2}\sum_j \Delta g_{jj} + \sum_i \sum_j \frac{\partial^2 g_{ij}}{\partial x^i \partial x^j} - \frac{1}{2}\sum_i \sum_j \frac{\partial^2 g_{ii}}{(\partial x^j)^2} = \\
&= -\frac{1}{2}\sum_j \Delta g_{jj} + \sum_i \sum_j \frac{\partial^2 g_{ij}}{\partial x^i \partial x^j} - \frac{1}{2}\sum_i \Delta g_{ii} = \\
&= -\sum_i \Delta g_{ii} + \sum_i \sum_j \frac{\partial^2 g_{ij}}{\partial x^i \partial x^j}.
\end{aligned}
$$

In the special case that the co-ordinate curves are (everywhere, not just in point P) orthogonal, *we obtain* $g_{ij} = 0$ if $i \neq j$ with its derivatives of place, i.e.

$$
\begin{aligned}
R &= -\sum_i \Delta g_{ii} + \sum_i \frac{\partial^2 g_{ii}}{(\partial x^i)^2} = \\
&= \frac{\partial^2 g_{11}}{(\partial x^1)^2} + \frac{\partial^2 g_{22}}{(\partial x^2)^2} - \left[\frac{\partial^2 g_{11}}{(\partial x^1)^2} + \frac{\partial^2 g_{11}}{(\partial x^2)^2} + \frac{\partial^2 g_{22}}{(\partial x^1)^2} + \frac{\partial^2 g_{22}}{(\partial x^2)^2} \right] = \\
&= -\left[\frac{\partial^2 g_{11}}{(\partial x^2)^2} + \frac{\partial^2 g_{22}}{(\partial x^1)^2} \right]. \qquad (10.13)
\end{aligned}
$$

This equation will be of use in the sequel.

Chapter 11

Map projections

Map projections are needed because the depiction of the curved Earth surface on a plane is not possible without error at least for larger areas. In this chapter we discuss the deformations introduced by a map projection in terms of its *scale error*, using the tools developed in the previous chapters.

11.1. Map projections and scale

11.1.1. On the Earth surface

On the surface of the Earth a distance element dS may consist of an element of latitude $d\varphi$ and an element of longitude $d\lambda$. These correspond to linear distances $M(\varphi)\,d\varphi$ and $p(\varphi)\,d\lambda$, respectively.

According to PYTHAGORAS, the length of the diagonal of the postage stamp $(d\varphi, d\lambda)$ is

$$dS^2 = M^2 d\varphi^2 + p^2 d\lambda^2. \tag{11.1}$$

Now dS^2 defines a *metric*,

$$dS^2 = \sum_{i,j} g_{ij} dx^i dx^j = \mathbf{x}^T H \mathbf{x},$$

where

$$g_{ij} = \begin{bmatrix} M^2 & 0 \\ 0 & p^2 \end{bmatrix} = \begin{bmatrix} E & F \\ F & G \end{bmatrix} = H \tag{11.2}$$

and

$$\mathbf{x} = \begin{bmatrix} dx^1 \\ dx^2 \end{bmatrix} = \begin{bmatrix} d\varphi \\ d\lambda \end{bmatrix}.$$

Here we have suitably defined two ways of writing, the index notation g_{ij}, dx^i and the matrix-vector notation $H, \mathbf{x}$. *The elements of the matrix are also* the elements of the Gauss First Fundamental Form on the Earth surface E, F, G. We see that $E = M^2, F = 0$ ja $G = p^2$.

For the azimuth A again we obtain the following equation:

$$\tan A = \frac{pd\lambda}{Md\varphi}, \tag{11.3}$$

from which, with the above,

$$dS \sin A = pd\lambda, \tag{11.4}$$
$$dS \cos A = Md\varphi. \tag{11.5}$$

11.1.2. In the map plane

When we project this little rectangle into the map plane, we obtain the sides dx and dy, and according to PYTHAGORAS the diagonal is

$$ds^2 = dx^2 + dy^2.$$

Now we may calculate an element of distance *in the map plane* as follows:

$$\begin{aligned} dx &= \frac{\partial x}{\partial \varphi} d\varphi + \frac{\partial x}{\partial \lambda} d\lambda, \\ dy &= \frac{\partial y}{\partial \varphi} d\varphi + \frac{\partial y}{\partial \lambda} d\lambda, \end{aligned}$$

i.e.,

$$\begin{aligned} ds^2 &= \left(\frac{\partial x}{\partial \varphi} d\varphi + \frac{\partial x}{\partial \lambda} d\lambda \right)^2 + \left(\frac{\partial y}{\partial \varphi} d\varphi + \frac{\partial y}{\partial \lambda} d\lambda \right)^2 = \\ &= \widetilde{E} d\varphi^2 + 2\widetilde{F} d\varphi d\lambda + \widetilde{G} d\lambda^2, \end{aligned} \tag{11.6}$$

where

$$\begin{aligned} \widetilde{E} &= \left(\frac{\partial x}{\partial \varphi} \right)^2 + \left(\frac{\partial y}{\partial \varphi} \right)^2, \\ \widetilde{F} &= \frac{\partial x}{\partial \varphi} \frac{\partial x}{\partial \lambda} + \frac{\partial y}{\partial \varphi} \frac{\partial y}{\partial \lambda}, \\ \widetilde{G} &= \left(\frac{\partial x}{\partial \lambda} \right)^2 + \left(\frac{\partial y}{\partial \lambda} \right)^2. \end{aligned}$$

$\widetilde{E}, \widetilde{F}$ ja $\widetilde{G}$ are the GAUSS First Fundamental Form *in the map plane.* If we interpret the element of distance in the map plane ds^2 as a *metric* of the Earth surface, then we obtain also here a metric tensor,

$$\widetilde{g_{ij}} = \begin{bmatrix} \widetilde{E} & \widetilde{F} \\ \widetilde{F} & \widetilde{G} \end{bmatrix} \equiv \widetilde{H}. \tag{11.7}$$

The corresponding metric is

$$ds^2 = \sum_{i,j} \widetilde{g_{ij}} dx^i dx^j = \mathbf{x}^T \widetilde{H} \mathbf{x},$$

where $dx^1 = d\varphi$ and $dx^2 = d\lambda$, , i.e., $dx^i = dx^j = \mathbf{x} = \begin{bmatrix} d\varphi & d\lambda \end{bmatrix}^T$. Also here we see as alternatives the index notation and the matrix-vector notation, which describe the same thing.

11.1.3. The scale

The *scale* is now the ratio

$$m = \frac{ds}{dS},$$

which apparently depends on the *direction* of the distance element, i.e., the azimuth A.

Let us write the *eigenvalue problem*:

$$\begin{aligned} m^2 &= \frac{ds^2}{dS^2} = \frac{\sum_{i,j} \widetilde{g_{ij}} dx^i dx^j}{\sum_{i,j} g_{ij} dx^i dx^j} \Rightarrow \\ &\Rightarrow \sum_{i,j} \widetilde{g_{ij}} dx^i dx^j - m^2 \sum_{i,j} g_{ij} dx^i dx^j = 0. \end{aligned} \tag{11.8}$$

The same equation in "matrix language", if $\mathbf{x} = \begin{bmatrix} dx^1 & dx^2 \end{bmatrix}^T = \begin{bmatrix} d\varphi & d\lambda \end{bmatrix}^T$:

$$\mathbf{x}^T \left(\widetilde{H} - m^2 H \right) \mathbf{x} = 0.$$

Equations (11.4, 11.5) The read:

$$\begin{aligned} dS \sin A &= p d\lambda, \\ dS \cos A &= M d\varphi. \end{aligned}$$

Let us now look at all vectors that are of form

$$\mathbf{x} = \begin{bmatrix} d\varphi \\ d\lambda \end{bmatrix} = \begin{bmatrix} \dfrac{\cos A}{M} \\ \dfrac{\sin A}{p} \end{bmatrix}. \tag{11.9}$$

The lengths of these vectors on the Earth surface are:

$$\begin{aligned} dS^2 &= M^2 d\varphi^2 + p^2 d\lambda^2 = \\ &= M^2 \left(\frac{\cos A}{M} \right)^2 + p^2 \left(\frac{\sin A}{p} \right)^2 = \\ &= \cos^2 A + \sin^2 A = 1. \end{aligned}$$

So: on the surface of the Earth these vectors form a *circle of unit radius.*

Substitute (11.9) into equation (11.8), using (11.6):

$$\begin{aligned} \widetilde{E} \left(\frac{\cos A}{M} \right)^2 + \widetilde{G} \left(\frac{\sin A}{p} \right)^2 + 2\widetilde{F} \frac{\cos A}{M} \frac{\sin A}{p} - \\ -m^2 \left[M^2 \left(\frac{\cos A}{M} \right)^2 + p^2 \left(\frac{\sin A}{p} \right)^2 \right] = 0 \end{aligned}$$

or, after cleaning up,

$$\begin{aligned} m^2 &= \frac{\widetilde{E}}{M^2} \cos^2 A + \frac{\widetilde{G}}{p^2} \sin^2 A + 2 \frac{\widetilde{F}}{Mp} \sin A \cos A = \\ &= \frac{1}{2} \left(\frac{\widetilde{E}}{M^2} + \frac{\widetilde{G}}{p^2} \right) + \frac{1}{2} \left(\frac{\widetilde{E}}{M^2} - \frac{\widetilde{G}}{p^2} \right) \cos 2A + \frac{\widetilde{F}}{Mp} \sin 2A. \end{aligned} \tag{11.10}$$

From this we obtain the *stationary values*:

$$0 = \frac{d}{dA} m^2 = \left(\frac{\widetilde{G}}{p^2} - \frac{\widetilde{E}}{M^2} \right) \sin 2A + \frac{\widetilde{F}}{Mp} \cos 2A,$$

in other words

$$\tan 2A = \frac{\left(\frac{\widetilde{E}}{M^2} - \frac{\widetilde{G}}{p^2} \right)}{\frac{\widetilde{F}}{Mp}} = \frac{\widetilde{E} p^2 - \widetilde{G} M^2}{\widetilde{F} M p}.$$

This yields two maximum and two minumum values, which all four are at a distance of 90° from each other[1].

These eigenvalues are obtained by writing the equation (11.10) as follows:

$$\begin{bmatrix} \cos A & \sin A \end{bmatrix} \begin{bmatrix} \dfrac{\widetilde{E}}{M^2} - m^2 & \dfrac{\widetilde{F}}{Mp} \\ \dfrac{\widetilde{F}}{Mp} & \dfrac{\widetilde{G}}{p^2} - m^2 \end{bmatrix} \begin{bmatrix} \cos A \\ \sin A \end{bmatrix} = 0;$$

[1] If $F = 0$, we obtain the condition $\sin 2A = 0$, which is fulfilled when $A = k \cdot 90^\circ$, $k = 0, 1, 2, 3$.

This presupposes that the *determinant* of the matrix in the middle *vanishes*:

$$\begin{aligned}
0 &= \det\left(H^{-1}\widetilde{H} - m^2 I\right) = \\
&= \left(\frac{\widetilde{E}}{M^2} - m^2\right)\left(\frac{\widetilde{G}}{p^2} - m^2\right) - \frac{\widetilde{F}^2}{M^2p^2} = \\
&= m^4 + \left(-\frac{\widetilde{E}}{M^2} - \frac{\widetilde{G}}{p^2}\right)m^2 + \frac{1}{M^2p^2}\left(\widetilde{E}\widetilde{G} - \widetilde{F}^2\right).
\end{aligned}$$

From this

$$m_{1,2}^2 = \frac{\left(\frac{\widetilde{E}}{M^2} + \frac{\widetilde{G}}{p^2}\right) \pm \sqrt{\left(\frac{\widetilde{E}}{M^2} + \frac{\widetilde{G}}{p^2}\right)^2 - 4\frac{1}{M^2p^2}\left(\widetilde{E}\widetilde{G} - \widetilde{F}^2\right)}}{2}.$$

These two solutions are called the *principal scale factors* m_1, m_2.

If *the* $\widetilde{H}$ *matrix is a* diagonal matrix:

$$\widetilde{H} = \begin{bmatrix} \widetilde{E} & 0 \\ 0 & \widetilde{G} \end{bmatrix},$$

we obtain

$$\begin{aligned}
\det\left(H^{-1}\widetilde{H} - m^2 I\right) &= \left(\frac{\widetilde{E}}{M^2} - m^2\right)\left(\frac{\widetilde{G}}{p^2} - m^2\right) = 0 \Rightarrow \\
m_{1,2}^2 &= \frac{\widetilde{E}}{M^2}, \frac{\widetilde{G}}{p^2}.
\end{aligned}$$

$m_1 = \sqrt{\frac{\widetilde{E}}{M^2}}$ is the *meridional scale factor*, $m_2 = \sqrt{\frac{\widetilde{G}}{p^2}}$ is the *scale factor in the direction of the parallel*. In intermediate directions A (azimuth) the magnification factor is then

$$m = \sqrt{m_1^2 \cos^2 A + m_2^2 \sin^2 A}.$$

11.1.4. The Tissot-indicatrix

The matrix ("scale tensor")

$$H^{-1}\widetilde{H} = \begin{bmatrix} \dfrac{\widetilde{E}}{M^2} & \dfrac{\widetilde{F}}{Mp} \\ \dfrac{\widetilde{F}}{Mp} & \dfrac{\widetilde{G}}{p^2} \end{bmatrix} = g^{ik}\widetilde{g}_{kj} = \widetilde{g}^i_j$$

is often visualized as an ellipse on the Earth surface (or, correspondingly, in the map plane). The eigenvalues of the matrix are m_1^2 and m_2^2; the axes of the visualizing ellipse are m_1 in the meridional direction and m_2 in the direction of the parallel. This ellipse is called the *indicatrix of* TISSOT[2]. See subsection 10.3.3.

Of the many map projections used, we need to mention especially those that are *conformal*, i.e., the scale in a point is the same in all directions:

$$m_1 = m_2 = m,$$

[2]Nicolas Auguste TISSOT (1824-1897), French cartographer

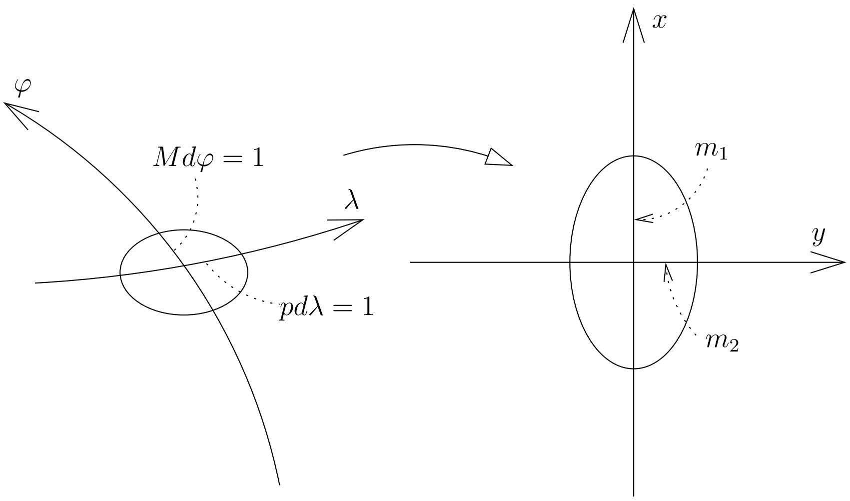


Figure 11.1.: The TISSOT indicatrix

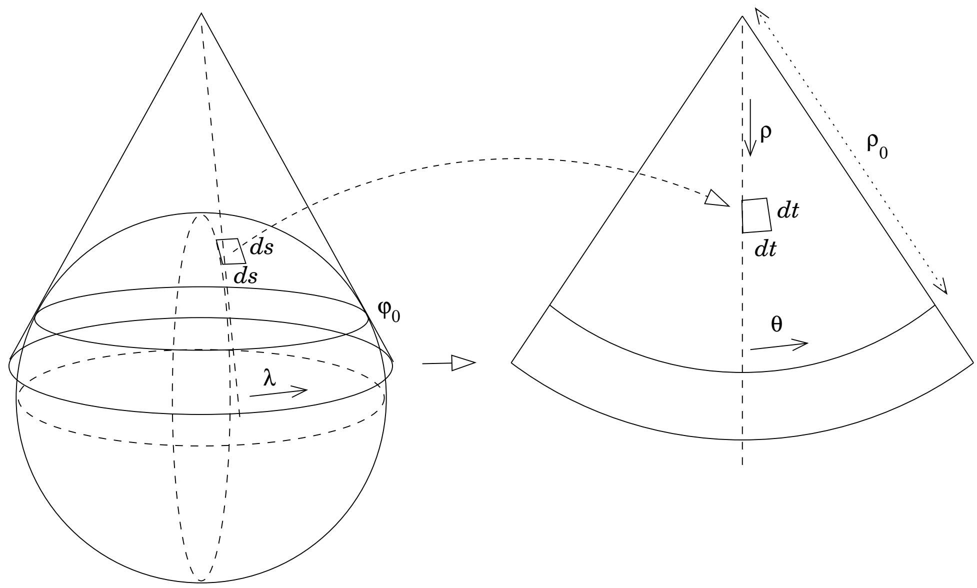


Figure 11.2.: LAMBERT projection

The TISSOT indicatrix is a *circle*. With a conformal projection, small circles, squares and local angles and length ratios are mapped true. Surface areas are distorted, however.

Conformal projections have the useful property, that the projection formulas in one co-ordinate direction can be derived when the formulas for the other co-ordinate direction are given.

A sometimes useful property is, that surface areas are mapped true [3], even though the shapes of small circles or squares are distorted. This requirement is $\det\left(H^{-1}\widetilde{H}\right) = m_1^2 m_2^2 =$ constant.

11.2. The Lambert projection (LCC)

The LAMBERT[4] projection (LCC, LAMBERT Conformal Conical) is a conformal conical projection. The scale for all latitude circles is constant; it is however different for different latitudes, and attains its maximum value for a certain latitude φ_0 in the middle of the area mapped. Generally this value is $m > 1$; in that case we have *two standard parallels* for which the scale $m = 1$.

[3] E.g. when depicting the surface density of population or some other phenomenon.
[4] Johann Heinrich LAMBERT (1728 – 1777) Swiss methematician, physicist and astronomer.

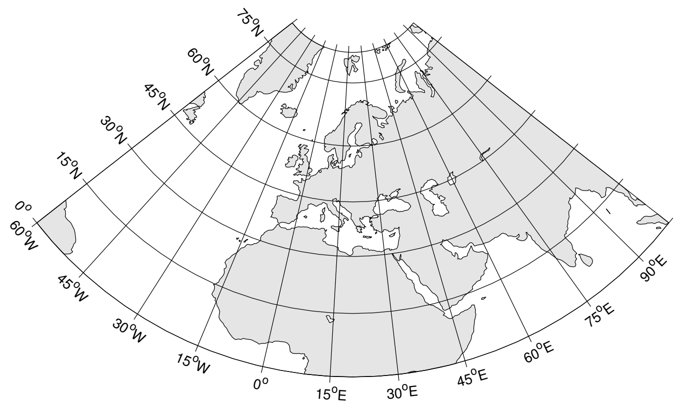


Figure 11.3.: Example of a LAMBERT projektion (software used: `m_map` (`http://www2.ocgy.ubc.ca/~rich/map.html`))

The images of the meridians in the map are straight lines, which intersect in one point, which is also the common centre of all latitude circle images.

Polar co-ordinates in the map plane are

$$\begin{aligned} x &= \rho \sin\theta, \\ y &= \rho_0 - \rho\cos\theta, \end{aligned}$$

where $\rho = \rho_0$ describes the latitude circle $\varphi = \varphi_0$.

Also

$$\theta = n\lambda. \tag{11.11}$$

Let ds be the distance element on the Earth surface and dt the corresponding element in the map plane; then along the latitude circle

$$\begin{aligned} ds &= p(\varphi)\, d\lambda \\ dt &= n\rho d\lambda \end{aligned}$$

and by dividing

$$\frac{dt}{ds} = \frac{n\rho}{p}.$$

Based on conformality this will also be valid along the meridian.

Now we have $ds = M(\varphi)\, d\varphi$, i.e.

$$\frac{dt}{d\varphi} = \frac{dt}{ds}\frac{ds}{d\varphi} = \frac{n\rho}{p}M.$$

If we reckon ρ positive toward the South, i.e., $d\rho = -dt$, we obtain:

$$\frac{d\rho}{d\varphi} = -n\rho(\varphi)\frac{M(\varphi)}{p(\varphi)}. \tag{11.12}$$

Using equations (11.11, 11.12) we may compute (θ, ρ) if given (φ, λ). However, let us first move ρ to the left hand side:

$$\frac{d}{d\varphi}\ln\rho = -n\frac{M}{p} \;\Rightarrow\; \rho = \rho_0 \exp\left\{-n\int_{\varphi_0}^{\varphi}\frac{M(\varphi')}{p(\varphi')}d\varphi'\right\}.$$

If we define

$$\psi(\varphi) \equiv \int_0^{\varphi}\frac{M(\varphi')}{p(\varphi')}d\varphi',$$

the so-called *isometric latitude*, we obtain

$$\rho = \rho_0 \exp\{-n(\psi(\varphi) - \psi(\varphi_0))\} = \rho_0 \frac{\exp\{-n\psi(\varphi)\}}{\exp\{-n\psi(\varphi_0)\}}. \tag{11.13}$$

Furthermore we must choose a value n. We do this so, that the *scale is stationary at the reference latitude* φ_0:

$$\frac{dm}{d\varphi} = 0 \text{ jos } \varphi = \varphi_0.$$

The scale is

$$m \equiv \frac{d\rho}{ds} = -\frac{dt}{ds} = -\frac{n\rho}{p},$$

i.e.,

$$\begin{aligned}\frac{dm}{d\varphi} &= \frac{n}{p}\frac{d\rho}{d\varphi} - \rho\frac{n}{p^2}\frac{dp}{d\varphi} = \\ &= -\frac{n^2\rho}{p^2}M + \frac{n\rho}{p^2}M\sin\varphi = -\frac{n\rho}{p^2}M(n - \sin\varphi),\end{aligned}$$

(using 11.12 and Appendix B) which ought to vanish for $\varphi = \varphi_0$. It does by choosing

$$n = \sin\varphi_0.$$

As an initial condition we must still choose ρ_0; it yields for the scale at the reference latitude

$$m(\varphi_0) = n\frac{\rho_0}{p(\varphi_0)}.$$

Alternatively one may choose a value φ_1 where $m(\varphi_1) = 1$ (a standard latitude): then

$$n\frac{\rho_1}{p(\varphi_1)} = 1,$$

from which $\rho_1 \equiv \rho(\varphi_1)$ follows. Now $\rho(\varphi)$ is obtained by integrating (11.12) from a starting value, either ρ_0 or ρ_1, with the integration interval being either $[\varphi_0, \varphi]$ or $[\varphi_1, \varphi]$.

The inverse operation is easy for $\theta \rightarrow \lambda$; solving in reverse the differential equation (11.12) (computing φ when ρ is given) can be done as follows:

1. invert analytically equation (11.13):

$$\psi(\varphi) = -\frac{\psi(\varphi_0)}{n}(\ln\rho - \ln\rho_0);$$

2. perform the inverse computation $\psi \rightarrow \varphi$ (see below).

11.3. On the isometric latitude

The isometric latitude,

$$\psi = \int_0^{\varphi} \frac{M(\varphi)}{p(\varphi)} d\varphi,$$

can be computed numerically (quadrature; the `QUAD` routines of Matlab). However, for the spherical case a closed solution exists

$$\psi = \ln\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right), \tag{11.14}$$

easily proven by differentiating using the chain rule.

Even for the ellipsoid of revolution a closed solution exists:

$$\psi = \ln \left[\tan \left(\frac{\pi}{4} + \frac{\varphi}{2} \right) \left(\frac{1 + e \sin \varphi}{1 - e \sin \varphi} \right)^{\frac{e}{2}} \right]. \tag{11.15}$$

See appendix A.

Reverse operation: if ψ is given and φ to be computed, we may take the equation

$$\frac{d\psi}{d\varphi} = \frac{M(\varphi)}{p(\varphi)}$$

and turn it upside down:

$$\frac{d\varphi}{d\psi} = \frac{p(\varphi)}{M(\varphi)}.$$

This equation has the general form

$$\frac{dy}{dt} = f(y, t)$$

and may be numerically solved using Matlab's `ODE` routines.

An alternative way for the sphere is to analytically invert equation (11.14):

$$\varphi = 2 \left(\arctan \exp \psi - \frac{\pi}{4} \right).$$

This is also useful as the first iteration step for the ellipsoidal case:

$$\varphi^{(0)} = 2 \left(\arctan \exp \psi - \frac{\pi}{4} \right),$$

after which

$$\varphi^{(i+1)} = 2 \left\{ \arctan \left[\left(\frac{1 - e \sin \varphi^{(i)}}{1 + e \sin \varphi^{(i)}} \right)^{\frac{e}{2}} \exp \psi \right] - \frac{\pi}{4} \right\}.$$

This converges rapidly.

11.4. The Mercator projection

The classical MERCATOR projection is obtained as a limiting case of LAMBERT by choosing the limit $n \to 0$, $\rho_0 \to \infty$, but nevertheless $n\rho_0 = 1$, and also $\varphi_0 = 0$. Then

$$\begin{aligned} \rho &= \rho_0 \exp \{ -n (\psi(\varphi) - \psi(\varphi_0)) \} \approx \\ &\approx \rho_0 - n\rho_0 (\psi(\varphi) - \psi(\varphi_0)) = \\ &= \rho_0 - \psi(\varphi). \end{aligned}$$

Let us choose $y \equiv -(\rho - \rho_0)$ and $x \equiv \lambda$ and we have the *projection formulas of* MERCATOR

$$\begin{aligned} x &= \lambda, \\ y &= \psi(\varphi) = \int_0^{\varphi} \frac{M(\varphi')}{p(\varphi')} d\varphi'. \end{aligned}$$

Here we see the isometric latitude in its most simple glory: in the case of spherical geometry

$$y = \ln \tan \left(\frac{\pi}{4} + \frac{\varphi}{2} \right).$$

MERCATOR is not a "lamp projection": There is no projection centre from which the "light" emanates.

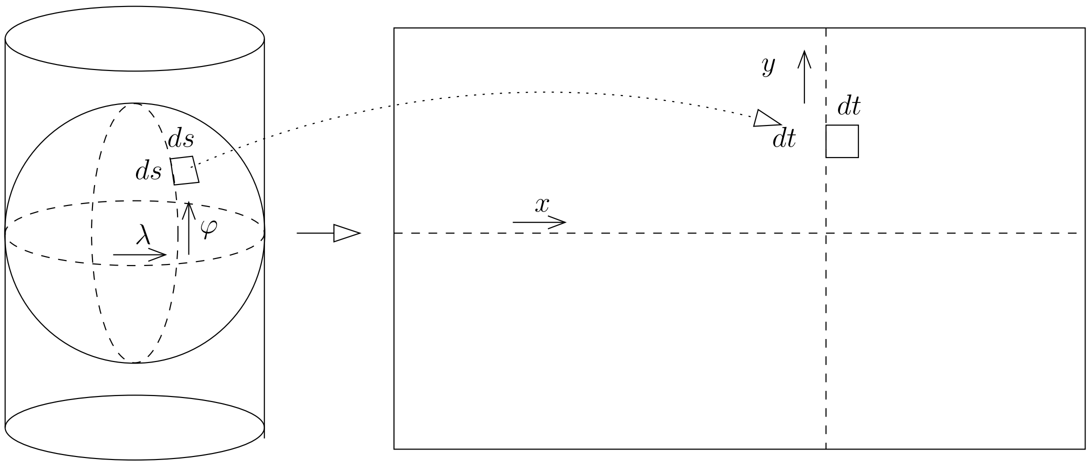


Figure 11.4.: The MERCATOR projection

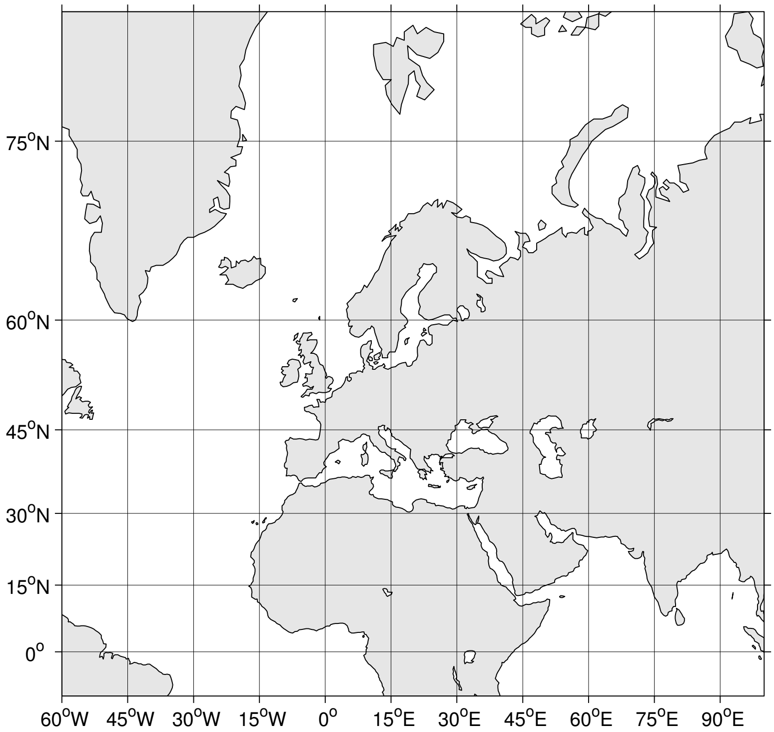


Figure 11.5.: An example of the MERCATOR projection

11.5. The stereographic projection

This so-called azimuthal projection is also conformal and also a limiting case of the LAMBERT projection.

Let us choose in the equation (11.13) $n = 1$ and choose $\varphi_0 = 0$ (the equator) and ρ_0 correspondingly, i.e., $\rho_0 \equiv \rho(\varphi_0)$. In that case

$$\rho = \rho_0 \exp\{-\psi(\varphi)\}.$$

Unfortunately in the limit $\varphi \to \dfrac{\pi}{2}$ the function $\psi(\varphi)$ diverges; we may define

$$\begin{aligned} \rho = \rho_0 \exp\{-\psi(\varphi)\} &\qquad \text{if } \varphi < \frac{\pi}{2}, \\ \rho = 0 &\qquad \text{if } \varphi = \frac{\pi}{2}, \end{aligned}$$

after which $\rho(\varphi)$ is continuous.

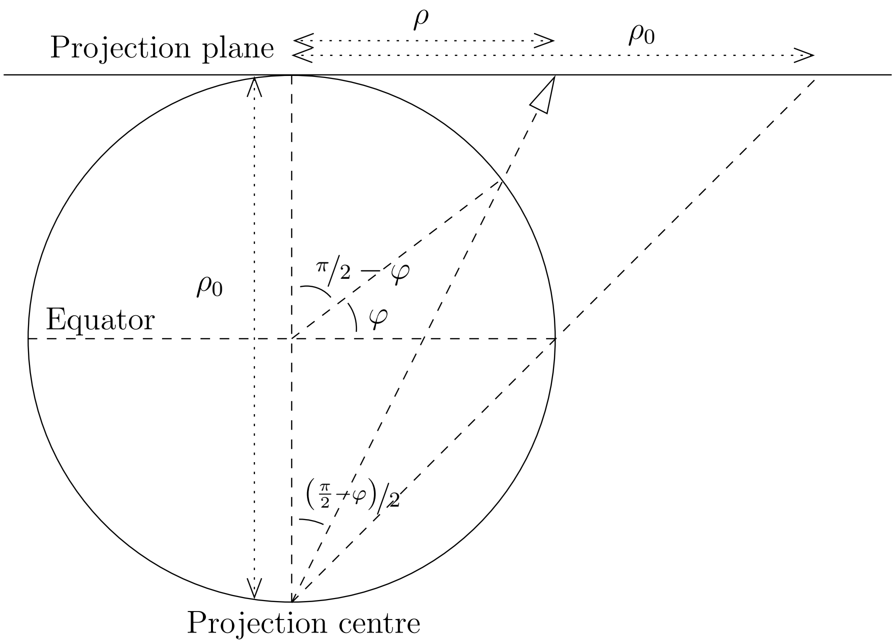


Figure 11.6.: The stereographic projection

In the spherical case we obtain

$$\begin{aligned}\rho &= {}^{\rho_0}\!/\!_{\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)} = \\ &= \rho_0 \cot\left(\frac{\pi}{4}+\frac{\varphi}{2}\right) = \\ &= \rho_0 \tan\left(\frac{\pi}{4}-\frac{\varphi}{2}\right) = \\ &= \rho_0 \tan\left[\frac{1}{2}\left(\frac{\pi}{2}-\varphi\right)\right].\end{aligned}$$

This case is depicted in figure 11.6. ρ_0 is the distance from the central point ("South pole") to the projection plane. In this case the projection is truly a "lamp projection"... but unfortunately *only* in the case of spherical geometry.

Let us derive the *scale* of the stereographical projection:

$$\frac{d\rho}{dx} = \frac{d\rho}{d\psi}\frac{d\psi}{d\varphi}\frac{d\varphi}{dx} = -\rho_0 \exp\{-\psi\} \cdot \frac{M}{p} \cdot \frac{1}{M} = -\frac{\rho}{p} = -\rho_0 \frac{\exp\{-\psi(\varphi)\}}{p(\varphi)}.$$

The value is negative because x grows to the North and ρ to the South. The co-ordinate x is the metric "northing".

By using equation 11.15 we obtain

$$\frac{d\rho}{dx} = -\frac{\rho_0}{N\cos(\varphi)} \cot\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\left(\frac{1+e\sin\varphi}{1-e\sin\varphi}\right)^{-\frac{e}{2}},$$

or close to the pole

$$\frac{d\rho}{dx} \approx -\frac{\rho_0}{2}\left(\frac{1+e}{1-e}\right)^{-\frac{e}{2}}.$$

In case of the Earth where (GRS80) $e = 0.08181919104281097693$, we obtain for the last term in e 0.99331307907268199009. If we want unity for the scale at the pole, we must set

$$\rho_0 = 2.01346387371353381594.$$

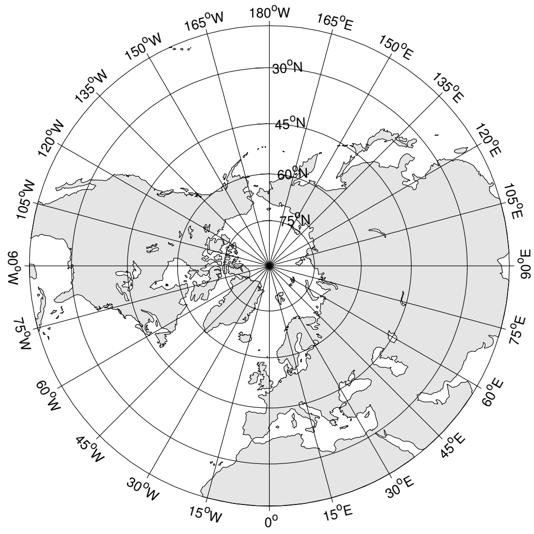


Figure 11.7.: An example of the stereographic projection

Reference latitude	Latitude interval	Number of support points	Computation point $(\varphi, \Delta\lambda)$	Number of terms
$1°$	$1° - 20°$	20	$(19°.333, 10°.0)$	12
$61°$	$61° - 80°$	20	$(79°.333, 20°.0)$	16
$70°$	$61° - 80°$	20	$(79°.333, 20°.0)$	15
$70°$	$61° - 80°$	39	$(79°.333, 20°.0)$	15
$65°$	$61° - 70°$	19	$(60°.333, 10°.0)$	14

Table 11.1.: Results of Gauss-Krüger test computations. "Number of terms": the number of terms in the polynomial expansion needed to achieve a change under ± 1 mm in the computed Gauss-Krüger co-ordinates

11.6. The Gauss-Krüger projection

A good practical example of a map projection is the Gauss-Krüger projection, in use in Finland. It is a transversal cylindrical projection which is *conformal.*

Like has been traditionally the habit, we will present projection formulas in the form of series expansions. But, differently from tradition, we will determine the coefficients of the expansion *numerically.*

Let us proceed in the following way. First we choose a suitable *starting projection*, e.g., an ordinary Mercator, the equations of which are simple. So, we map the surface of the Earth ellipsoid onto the map plane of an ordinary Mercator:

$$\begin{aligned} v &= \lambda - \lambda_0, \\ u &= \int_0^{\varphi} \frac{M(\varphi')}{p(\varphi')} d\varphi'. \end{aligned}$$

Next we construct, in the Mercator *plane, an* analytical mapping

$$u + iv \rightarrow x + iy,$$

one property of which is, that x is of true length along the central meridian $\lambda - \lambda_0 = y = v = 0$.

$$dx = M(\varphi) d\varphi,$$

i.e.,

$$x = \int_0^{\varphi} M\left(\varphi'\right) d\varphi'. \tag{11.16}$$

Furthermore we have on the central meridian:

$$y = 0.$$

Now we have a *boundary value problem*: the sought for complex map co-ordinate $\mathbf{z} \equiv x + iy$ is given as a function of the starting projection's map co-ordinate $\mathbf{w} \equiv u + iv$ on the edge $y = v = 0$ i.e., the real axis. The problem is to determine the function *in the whole complex plane*. See figure 11.8.

Intermezzo. In complex analysis we talk of *analytical functions*. An analytical function is such a mapping

$$\mathbf{z} = f\left(\mathbf{w}\right)$$

which is differentiable. Not just once, but infinitely many times. In this case, the CAUCHY-RIEMANN conditions apply:

$$\frac{\partial x}{\partial u} = \frac{\partial y}{\partial v}, \ \frac{\partial y}{\partial u} = -\frac{\partial x}{\partial v},$$

if $\mathbf{z} = x + iy$, $\mathbf{w} = u + iv$.

This means that the small vector $\begin{bmatrix} du & dv \end{bmatrix}^T$ is mapped to a vector $\begin{bmatrix} dx & dy \end{bmatrix}^T$ according to the following equation:

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix} = K \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix},$$

where $a = \frac{\partial x}{\partial u} = \frac{\partial y}{\partial v}$ and $b = \frac{\partial y}{\partial u} = -\frac{\partial x}{\partial v}$, precisely as the CAUCHY-RIEMANN conditions require.

From this it is seen that the mapping of local vectors $(du, dv) \rightarrow (dx, dy)$ is a *scaling and rotation*; i.e., we have a *conformal mapping*.

Almost all familiar functions are analytical in the complex plane: the exponent, the logarithm, trigonometric and hyperbolic functions, and especially *power series expansions*. The sum and product of two analytical functions is again analytic.

From the CAUCHY-RIEMANN conditions we still derive the LAPLACE *equations*:

$$\frac{\partial^2 x}{\partial u^2} + \frac{\partial^2 x}{\partial v^2} = 0, \ \frac{\partial^2 y}{\partial u^2} + \frac{\partial^2 y}{\partial v^2} = 0.$$

Let us try, as a general solution, a *series expansion*:

$$\mathbf{z} = a_0 + a_1\mathbf{w} + a_2\mathbf{w}^2 + a_3\mathbf{w}^3 + \ldots = \sum_{k=0}^{\infty} a_k \mathbf{w}^k. \tag{11.17}$$

Here we define for the sake of generality

$$\begin{aligned} \mathbf{z} &\equiv (x - x_0) + iy, \\ \mathbf{w} &\equiv (u - u_0) + iv. \end{aligned}$$

The values

$$\begin{aligned} x_0 &\equiv \int_0^{\varphi_0} M\left(\varphi\right) d\varphi, \\ u_0 &\equiv \int_0^{\varphi_0} \frac{M\left(\varphi\right)}{p\left(\varphi\right)} d\varphi \end{aligned}$$

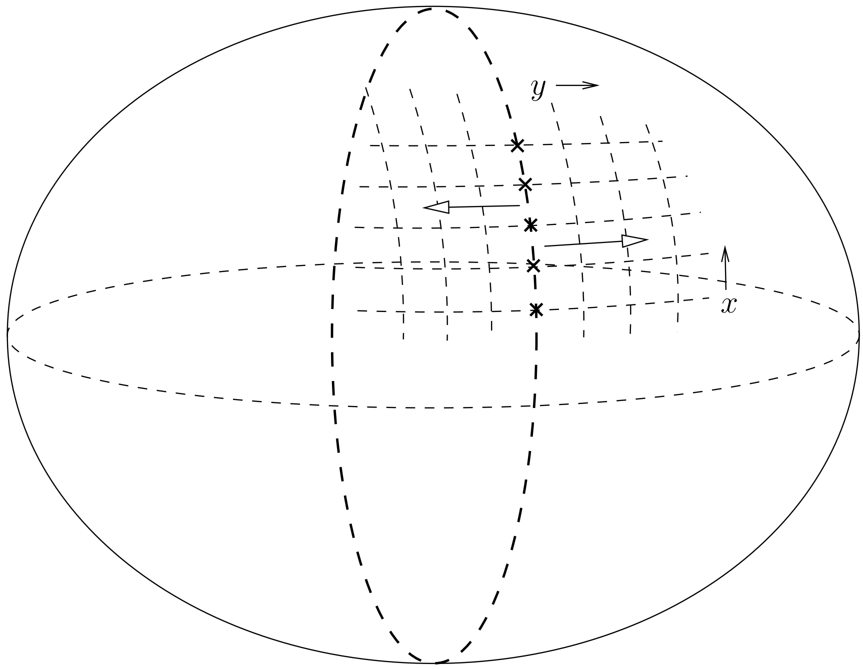


Figure 11.8.: The GAUSS-KRÜGER projection as a boundary value problem. The boundary values at the central meridian are marked with crosses, the directions of integration with arrows

have been chosen to be compatible with a suitable *reference latitude* φ_0. These series expansions are thus meant to be used only *in a relatively small area*, not the whole projection zone. Thus the number of terms needed also remains smaller.

The meridian conditions (11.16) now form a *set of observation equations*:

$$x_i = a_0 + a_1 u_i + a_2 u_i^2 + a_3 u_i^3 + \ldots$$

from which the coefficients a_j can be solved, if a sufficient number of support points (u_i, x_i) is given (the crosses in figure 11.8).

Applying the found coefficients a_j to equation (11.17) yields the solution for the whole complex plane. This solution may be "squeezed" into the following, more computationally suitable, form (so-called CLENSHAW summation):

$$\mathbf{z} = a_0 + \mathbf{w}\left(a_1 + \mathbf{w}\left(a_2 + \mathbf{w}\left(a_3 + \mathbf{w}\left(\cdots + \mathbf{w}a_n\right)\right)\right)\right).$$

The computing sequence is multiplication-addition-multiplication-addition. . . the powers of1 $\mathbf{w}$ need not be computed separately. *Remember* that also the intermediate results are complex numbers!

Fortunately complex numbers belong to the popular programming languages either as a fixed part (Matlab, Fortran) or a standard library (C++).

From the same equation we obtain also easily the scale and the meridian convergence. At said, a complex analytical map is a *scaling* plus a *rotation*. When the equation of the map is (11.17), its differential version is

$$\Delta\mathbf{z} = \frac{d\mathbf{z}}{d\mathbf{w}}\Delta\mathbf{w}, \tag{11.18}$$

where

$$\frac{d\mathbf{z}}{d\mathbf{w}} = a_1 + 2a_2\mathbf{w} + 3a_3\mathbf{w}^2 + 4a_4\mathbf{w}^4 + \cdots = \sum_{k=1}^{\infty} k a_k \mathbf{w}^{\mathbf{k-1}}.$$

Which also can easily we written into the CLENSHAW form.

Now we may write the complex equation (11.18) in real-valued matrix form

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} \mathfrak{Re}\left\{\dfrac{d\mathbf{z}}{d\mathbf{w}}\right\} & -\mathfrak{Im}\left\{\dfrac{d\mathbf{z}}{d\mathbf{w}}\right\} \\ \mathfrak{Im}\left\{\dfrac{d\mathbf{z}}{d\mathbf{w}}\right\} & \mathfrak{Re}\left\{\dfrac{d\mathbf{z}}{d\mathbf{w}}\right\} \end{bmatrix} \begin{bmatrix} du \\ dv \end{bmatrix},$$

where $d\mathbf{x} = \begin{bmatrix} dx & dy \end{bmatrix}^{\mathrm{T}}$, $d\mathbf{w} = \begin{bmatrix} du & dv \end{bmatrix}^{\mathrm{T}}$,

$$\mathfrak{Re}\left\{\frac{d\mathbf{z}}{d\mathbf{w}}\right\} = \frac{\partial x}{\partial u} = \frac{\partial y}{\partial w} = K\cos\theta,$$

the *scale*; and

$$\mathfrak{Im}\left\{\frac{d\mathbf{z}}{d\mathbf{w}}\right\} = \frac{\partial y}{\partial u} = -\frac{\partial x}{\partial v} = K\sin\theta,$$

from which θ, the *meridian convergence*, may be resolved.

The choice of the classical MERCATOR as a starting projection is not the only alternative. Probably the number of terms of the series expansion (11.17) could be reduced by using a LAMBERT projection for the reference latitude φ_0. Whether the saving achievable is worth the trouble, is a so-called Good Question.

11.7. Curvature of the Earth surface and scale

Recall the equation 10.13 derived above:

$$R = -\left[\frac{\partial^2 g_{11}}{(\partial x^2)^2} + \frac{\partial^2 g_{22}}{(\partial x^1)^2}\right].$$

Because according to equation 10.12 we have $R = 2K$, we obtain

$$K = -\frac{1}{2}\left[\frac{\partial^2 g_{11}}{(\partial x^2)^2} + \frac{\partial^2 g_{22}}{(\partial x^1)^2}\right].$$

Let us now take a rectangular co-ordinate system *in the map plane* (x, y) and transfer its co-ordinate lines back to the curved surface of the Earth, forming a curvilinear co-ordinate system (ξ, η). At the origin or central meridian in the map plane the metric of this co-ordinate system g_{ij} is, at least for ordinary map projections, *quasi-Euclidean*, in other words, this metric is the unit matrix and it is stationary at the origin. In this case the theory in the previous chapter (section 10.8) applies.

In the case of a *conformal projection* the form of this metric is

$$g_{ij} = m^{-2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

where m is the scale of the map projection, which thus depends on location x^i.

Derive

$$K = -\frac{1}{2}\left(\frac{\partial^2 g_{\xi\xi}}{\partial \eta^2} + \frac{\partial^2 g_{\eta\eta}}{\partial \xi^2}\right) = -\frac{1}{2}\Delta\left(m^{-2}\right).$$

Now

$$\begin{aligned} \Delta\left(m^{-2}\right) &= \left(\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}\right) m^{-2}(\xi, \eta) = \\ &= -2\frac{\partial}{\partial \xi}\left(m^{-3}\frac{\partial m}{\partial \xi}\right) - 2\frac{\partial}{\partial \eta}\left(m^{-3}\frac{\partial m}{\partial \eta}\right) = \\ &= -2\left(\frac{\partial m^{-3}}{\partial \xi}\cdot\frac{\partial m}{\partial \xi} + m^{-3}\frac{\partial^2 m}{\partial \xi^2}\right) - 2\left(\frac{\partial m^{-3}}{\partial \eta}\cdot\frac{\partial m}{\partial \eta} + m^{-3}\frac{\partial^2 m}{\partial \eta^2}\right) \\ &\approx -2\Delta m, \end{aligned}$$

if we consider the stationarity of m (and thus of m^{-3}) and $m \approx 1$.

So
$$K = \Delta m.$$

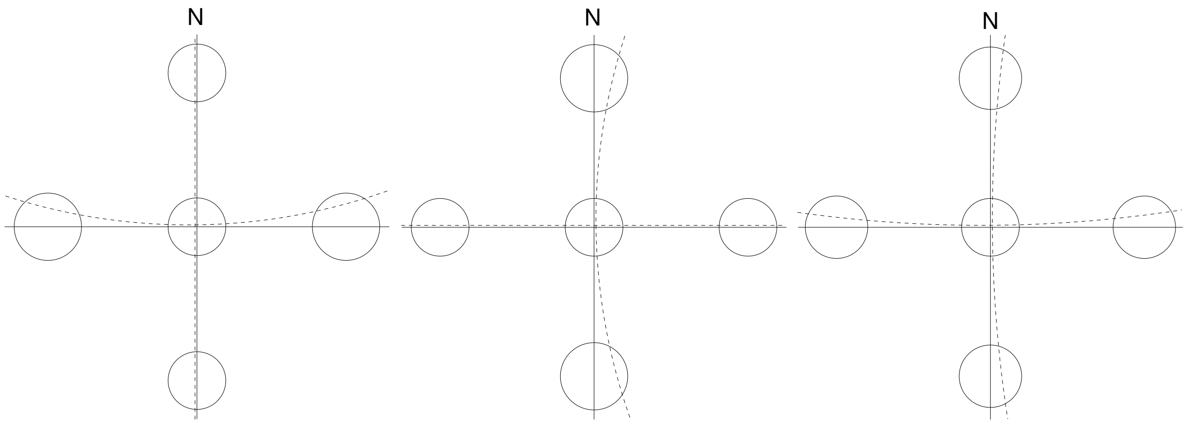
So, there is a simple relationship between the GAUSS curvature radius and the second derivative of the scale (more precisely, the LAPLACE operator Δ) As a consequence, there also is a relationship between the scale's second derivatives in the North-South and East-West directions: if, e.g.,
$$\frac{\partial^2 m}{\partial \xi^2} = 0 \Rightarrow \frac{\partial^2 m}{\partial \eta^2} = K$$
etc. In the below table we give some examples – remember that the area considered is always a neighbourhood of the central point or central meridian or standard parallel, where $m \approx 1$ and stationary!

Projection	$\partial^2 m/\partial x^2$	$\partial^2 m/\partial y^2$
MERCATOR	K	0
LAMBERT conical	K	0
Oblique stereographic	$\frac{K}{2}$	$\frac{K}{2}$
GAUSS-KRÜGER, UTM	0	K

In the table we used again, instead of (ξ, η),(x, y), i.e., we substituted
$$\frac{\partial^2 m}{\partial \xi^2} \longrightarrow \frac{\partial^2 m}{\partial x^2} \quad \text{and} \quad \frac{\partial^2 m}{\partial \eta^2} \longrightarrow \frac{\partial^2 m}{\partial y^2},$$
which is allowed based on quasi-Euclidicity.

The scale of MERCATOR and LAMBERT projections is a constant in the direction of the y co-ordinate, in other words, from map West to map East. In transversal cylindrical projections the scale again is constant in the direction of the x co-ordinate, i.e., the direction of the central meridian. The oblique stereographic projection again is symmetric and the scale behaves in the same way in all compass directions from the central point.

Thus, we may use the second derivatives of place of the scale for classifying map projections. E.g., LAMBERT is most suited for countries extending in the West-East direction (Estonia), whereas GAUSS-KRÜGER again is best for countries extending in the North-South direction (New Zealand). The conformal azimuthal projection called oblique stereographic is suitable for "square" countries (The Netherlands).



(a) Gauss-Krüger (b) Mercator, Lambert (c) Stereographic

Figure 11.9.: The classification of map projections as "vertical", "horizontal" and "square" projections

Chapter 12

Map projections in Finland

12.1. Traditional map projections

In Finland, the traditional map projection has been Gauß-Krüger with zone width 3°. System name: *kkj* ("National Map Grid Co-ordinate System") created in 1970 (Parm, 1988). Following characteristics:

- Based on International (Hayford) reference ellipsoid of 1924; datum was taken from the European datum of 1950 by keeping fixed the triangulation point Simpsiö (nr. 90), at
  - ▷ latitude and longitude values from the ED50 European adjustment, and
  - ▷ geoidal undulation from the Bomford astro-geodetic geoid (Bomford, 1963).
- Map plane co-ordinates were obtained using Gauß-Krüger for central meridians of $19°, 21°, 24°, 27°, 30°$; for small-scale all-Finland maps, 27° is used (the *ykj* system).
- These co-ordinates (x, y) were further transformed in the plane using a four-parameter similarity ("Helmert") transformation in order to achieve agreement with the pre-existing provisional *vvj* ("Old State System", also "Helsinki System") co-ordinates, cf. (Ollikainen, 1993).

  Equation:

$$\begin{bmatrix} x \\ y \end{bmatrix}_{kkj} = \begin{bmatrix} 1.00000075 & -0.00000439 \\ 0.00000439 & 1.00000075 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_{ED50} + \begin{bmatrix} -61.571\,\mathrm{m} \\ 95.693\,\mathrm{m} \end{bmatrix}$$

12.2. Modern map projections

The modern Finnish system is quite different:

- Based on the GRS80 reference ellipsoid of 1980; datum is called EUREF-FIN, created by keeping fixed four stations fixed to their ITRF96 values at epoch 1997.0: the permanent GPS stations Metsähovi, Vaasa, Joensuu and Sodankylä (Ollikainen et al., 2000). Then, a *transformation* (Boucher and Altamimi, 1995) was applied to obtain co-ordinates in ETRF89. Thus, the datum is correctly described as ETRF89, but the *epoch* remains 1997.0, as no correction for individual station motion (mostly, glacial isostatic adjustment) was made in the transformation.

Equation:

$$\mathbf{X}^E(t_C) = \mathbf{X}^I_{yy}(t_C) + \mathbf{T}_{yy} + \begin{bmatrix} 0 & -\dot{R}_3 & \dot{R}_2 \\ \dot{R}_3 & 0 & -\dot{R}_1 \\ -\dot{R}_2 & \dot{R}_1 & 0 \end{bmatrix}_{yy} \mathbf{X}^I_{yy}(t_C) \cdot (t_C - 1989.0)$$

with t_C observations central epoch, $yy = (19)96$. The values $\mathbf{T}_{96}$ and $\dot{R}_{i,96}$ are tabulated in (Boucher and Altamimi,) Tables 3 and 4.

- For small-scale and topographic maps, the UTM projection is used with a central meridian of 27° (zone 35) *for the whole country*, producing the ETRS-TM35FIN plane co-ordinate system. This also defines the map sheet division. However, on maps in parts of Finland where another central meridian would be more appropriate (like zone 34, central meridian 21°), the corresponding co-ordinate grid is also printed on the map, in purple (Anon., 2003).

- For large scale maps, such as used for planning and cadastral work, the Gauß-Krüger projection continues to be used (but based on the above reference ellipsoid and datum), with a central meridian interval of only one degree: ETRS-GKn, where n designates the central meridian longitude. This avoids the problem of significant scale distortions.

12.3. The triangulated affine transformation used in Finland

12.3.1. Plane co-ordinates

The National Land Survey offers a facility to convert *kkj* co-ordinates to the new ETRS89-TM35FIN system of projection co-ordinates. The method is described in the publication

(Anon., 2003), where it is proposed to use for the plane co-ordinate transformation between the projection co-ordinates of ETRS-89 and the *ykj* co-ordinate system, a *triangle-wise affine transformation.*

Inside each triangle we may write the affine transformation can be written like

$$\begin{aligned} x^{(2)} &= \Delta x + a_1 x^{(1)} + a_2 y^{(1)} \\ y^{(2)} &= \Delta y + b_1 x^{(1)} + b_2 y^{(1)} \end{aligned}$$

where $\left(x^{(1)}, y^{(1)}\right)$ are the point co-ordinates in ETRS-GK27, and $\left(x^{(2)}, y^{(2)}\right)$ are the co-ordinates of the same point in *ykj*. This transformation formula has six parameters: Δx, Δy, a_1, a_2, b_1 ja b_2. If, in the three corners of the triangle, are given both $\left(x^{(1)}, y^{(1)}\right)$ and $\left(x^{(2)}, y^{(2)}\right)$, we can solve for these uniquely.

The transformation formula obtained is inside the triangles linear and continuous across the edges, but not differentiable: the scale is discontinuous across triangle edges. Because the mapping is not conformal either, the scale will also be dependent upon the direction considered.

A useful property of triangulation is, that it can be locally "patched": if better data is available in the local area – a denser point set, whose co-ordinate pairs $\left(x^{(i)}, y^{(i)}\right)$, $i = 1, 2$ are known – then we can take away only the triangles of that area and replace them by a larger number of smaller triangle, inside which the transformation will become more precise. This is precisely the procedure that local players, like municipalities, can use to advantage.

Write these equations in vector form:

$$\begin{bmatrix} x^{(2)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix}.$$

Most often the co-ordinates in the (1)and (2) datums are close to each other, i.e., $\begin{bmatrix} \Delta x & \Delta y \end{bmatrix}^{\mathrm{T}}$ are small. In that case we may write the *shifts*

$$\begin{aligned} \delta x &\equiv x^{(2)} - x^{(1)} = \Delta x + (a_1 - 1)\, x^{(1)} + a_2 y^{(1)}, \\ \delta y &\equiv y^{(2)} - y^{(1)} = \Delta y + b_1 x^{(1)} + (b_2 - 1)\, y^{(1)}. \end{aligned}$$

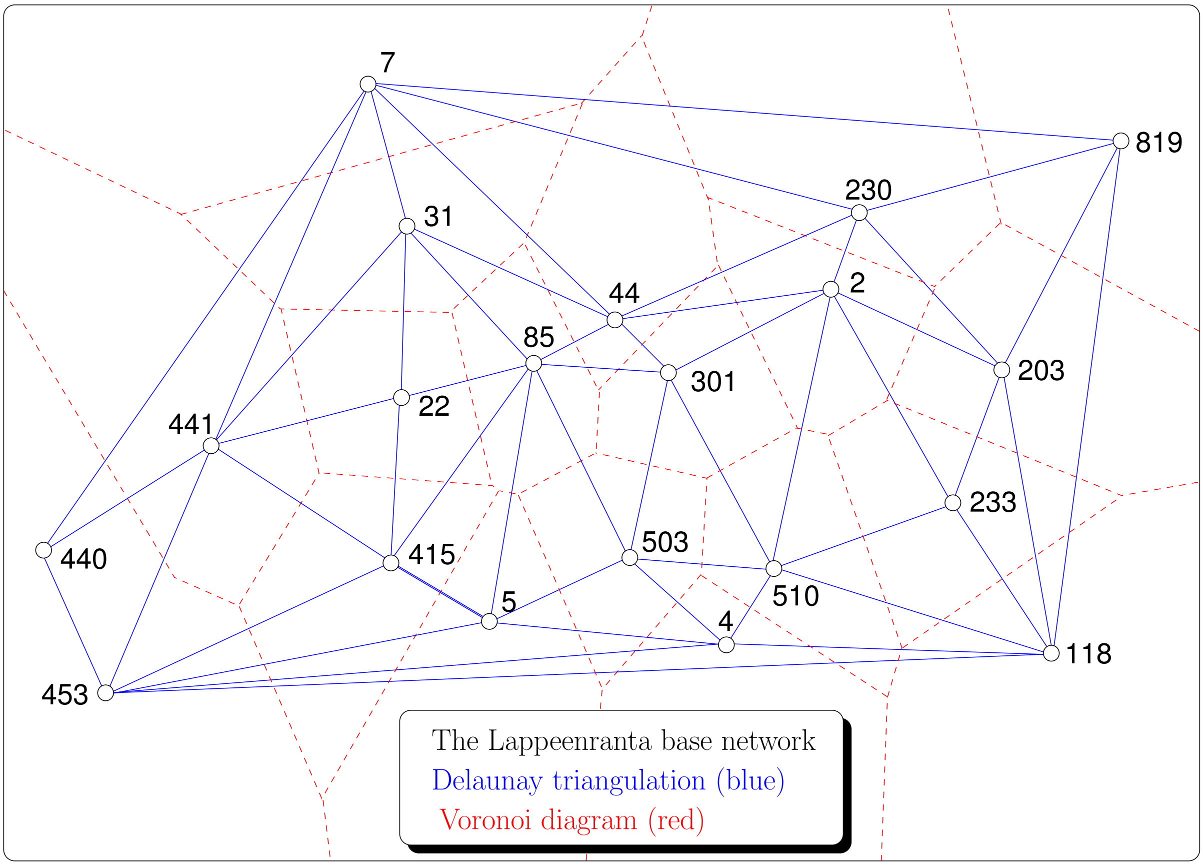


Figure 12.1.: Lappeenranta densification of the national triangular grid

If we now define

$$\Delta \mathbf{x} \equiv \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \equiv \begin{bmatrix} a_1 - 1 & a_2 \\ b_1 & b_2 - 1 \end{bmatrix},$$

we obtain the short form

$$\delta \mathbf{x} = \Delta \mathbf{x} + \mathbf{A}\mathbf{x}^{(1)}.$$

Also in this generally, if the co-ordinates are close together, the elements of $\mathbf{A}$ will be numerically small. Let there be a triangle ABC. Then we have given the shift vectors of the corners

$$\begin{aligned} \delta \mathbf{x}_A &= \Delta \mathbf{x} + \mathbf{A}\mathbf{x}_A^{(1)}, \\ \delta \mathbf{x}_B &= \Delta \mathbf{x} + \mathbf{A}\mathbf{x}_B^{(1)}, \\ \delta \mathbf{x}_C &= \Delta \mathbf{x} + \mathbf{A}\mathbf{x}_C^{(1)}. \end{aligned}$$

Write this out in components, with $\Delta \mathbf{x}, \mathbf{A}$ on the right hand side:

$$\begin{aligned} \delta x_A &= \Delta x + a_{11} x_A^{(1)} + a_{12} y_A^{(1)} \\ \delta y_A &= \Delta y + a_{21} x_A^{(1)} + a_{22} y_A^{(1)} \\ \delta x_B &= \Delta x + a_{11} x_B^{(1)} + a_{12} y_B^{(1)} \\ \delta y_B &= \Delta y + a_{12} x_B^{(1)} + a_{22} y_B^{(1)} \\ \delta x_C &= \Delta x + a_{11} x_C^{(1)} + a_{12} y_C^{(1)} \\ \delta y_C &= \Delta y + a_{21} x_C^{(1)} + a_{22} y_C^{(1)} \end{aligned}$$

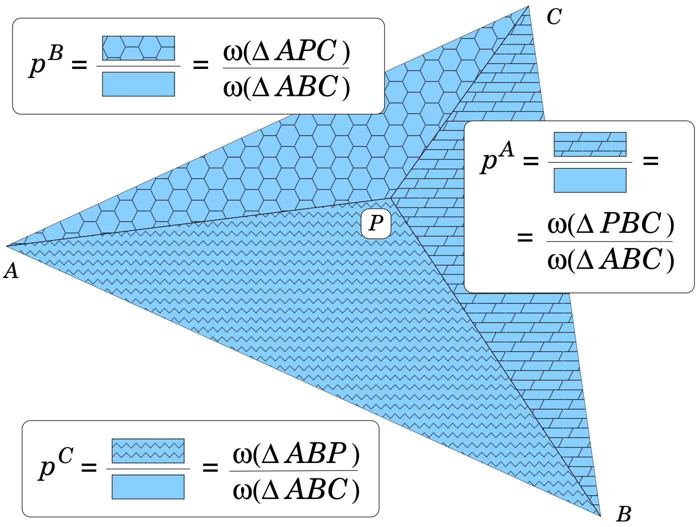


Figure 12.2.: Computing barycentric co-ordinates as the ratio of the surface areas of two triangles

or in matrix form

$$\begin{bmatrix} \delta x_A \\ \delta y_A \\ \delta x_B \\ \delta y_B \\ \delta x_C \\ \delta y_C \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_A^{(1)} & 0 & y_A^{(1)} & 0 \\ 0 & 1 & 0 & x_A^{(1)} & 0 & y_A^{(1)} \\ 1 & 0 & x_B^{(1)} & 0 & y_B^{(1)} & 0 \\ 0 & 1 & 0 & x_B^{(1)} & 0 & y_B^{(1)} \\ 1 & 0 & x_C^{(1)} & 0 & y_C^{(1)} & 0 \\ 0 & 1 & 0 & x_C^{(1)} & 0 & y_C^{(1)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix},$$

from which they can all be solved.

Let us write the coordinates (x, y) as follows:

$$\begin{aligned} x &= p^A x_A + p^B x_B + p^C x_C, \\ y &= p^A y_A + p^B y_B + p^C y_{C,} \end{aligned}$$

with the further condition $p^A + p^B + p^C = 1$. Then also

$$\begin{aligned} \delta x &= p^A \delta x_A + p^B \delta x_B + p^C \delta x_C, && (12.1) \\ \delta y &= p^A \delta y_A + p^B \delta y_B + p^C \delta y_{C.} && (12.2) \end{aligned}$$

The set of three numbers $\left(p^A, p^B, p^C\right)$ is called the *barycentric co-ordinates* of point P See figure 12.2.

They can be found as follows (geometrically $p^A = \dfrac{\omega\left(\Delta BCP\right)}{\omega\left(\Delta ABC\right)}$ etc., where ω is the surface area of the triangle) using determinants:

$$p^A = \frac{\begin{vmatrix} x_B & x_C & x \\ y_B & y_C & y \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{vmatrix}}, \; p^B = \frac{\begin{vmatrix} x_C & x_A & x \\ y_C & y_A & y \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{vmatrix}}, \; p^C = \frac{\begin{vmatrix} x_A & x_B & x \\ y_A & y_B & y \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{vmatrix}}.$$

These equations are very suitable for coding.

Appendix A

Isometric latitude on the ellipsoid

We follow the presentation from the book (Grossman, 1964).

The starting formula is

$$\psi(\varphi) = \int_0^{\varphi} \frac{M(\varphi')}{p(\varphi')} d\varphi'.$$

As a differential equation

$$d\psi = \frac{M}{p} d\varphi = \frac{1-e^2}{\left(1-e^2\sin^2\varphi\right)\cos\varphi} d\varphi.$$

The integrand is decomposed into partial fractions:

$$\begin{aligned}
\frac{1-e^2}{\left(1-e^2\sin^2\varphi\right)\cos\varphi} &= \frac{1-e^2\sin^2\varphi - e^2\cos^2\varphi}{\left(1-e^2\sin^2\varphi\right)\cos\varphi} = \\
&= \frac{1}{\cos\varphi} - \frac{e^2\cos\varphi}{1-e^2\sin^2\varphi} = \\
&= \frac{1}{\cos\varphi} - \frac{e^2\cos\varphi\,(1-e\sin\varphi) + e^2\cos\varphi\,(1+e\sin\varphi)}{2\,(1+e\sin\varphi)\,(1-e\sin\varphi)} = \\
&= \frac{1}{\cos\varphi} + \frac{e}{2}\left(-\frac{e\cos\varphi}{1+e\sin\varphi} - \frac{e\cos\varphi}{1-e\sin\varphi}\right).
\end{aligned}$$

The integral of the first term is

$$\psi = \int_0^{\varphi} \frac{1}{\cos\varphi'} d\varphi' = \ln\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right).$$

Proof by using the chain rule:

$$\begin{aligned}
\frac{d\psi}{d\varphi} &= \frac{d\ln\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)}{d\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)} \frac{d\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)}{d\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)} \frac{d\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)}{d\varphi} = \\
&= \frac{1}{\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)} \frac{1}{\cos^2\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)} \frac{1}{2} = \\
&= \frac{1}{2\sin\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\cos\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)} = \frac{1}{\sin\left(\frac{\pi}{2}+\varphi\right)} = \frac{1}{\cos\varphi}.
\end{aligned}$$

This is *the full solution* in the case that $e = 0$ (solution for the sphere).

In the case of the ellipsoid the second integral

$$\int \left(-\frac{e\cos\varphi}{1+e\sin\varphi}\right) d\varphi = \int \frac{f'(\varphi)}{f(\varphi)} d\varphi = \ln f(\varphi) = \ln(1+e\sin\varphi),$$

where we designate $f \equiv 1 + e\sin\varphi$. In the same way

$$\int \left(-\frac{e\cos\varphi}{1-e\sin\varphi}\right) d\varphi = -\ln(1-e\sin\varphi),$$

and the end result is

$$\begin{aligned}
\psi &= \ln\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)+\frac{e}{2}\left(\ln\left(1+e\sin\varphi\right)-\ln\left(1-e\sin\varphi\right)\right)= \\
&= \ln\left(\tan\left(\frac{\pi}{4}+\frac{\varphi}{2}\right)\left(\frac{1+e\sin\varphi}{1-e\sin\varphi}\right)^{\frac{e}{2}}\right).
\end{aligned}$$

Appendix B

Useful equations connecting the main radii of curvature

When we have given the principal radii of curvature of the ellipsoid of revolution:

$$\begin{aligned} N(\varphi) &= a\left(1-e^2\sin^2\varphi\right)^{-1/2}, \\ M(\varphi) &= a\left(1-e^2\right)\left(1-e^2\sin^2\varphi\right)^{-3/2}, \end{aligned}$$

we can calculate by brute-force derivation:

$$\begin{aligned} \frac{d}{d\varphi}\left(N(\varphi)\cos\varphi\right) &= -M(\varphi)\sin\varphi, \\ \frac{d}{d\varphi}\left(N(\varphi)\sin\varphi\right) &= +\frac{M(\varphi)}{1-e^2}\cos\varphi. \end{aligned}$$

Furthermore

$$\begin{aligned} \frac{d}{d\varphi}M^2(\varphi) &= a^2\left(1-e^2\right)^2\frac{d}{d\varphi}\left(1-e^2\sin^2\varphi\right)^{-3} = \\ &= 3\frac{e^2M^2N^2}{a^2}\sin 2\varphi, \\ \frac{d}{d\varphi}\left(N^2(\varphi)\cos^2\varphi\right) &= 2N\frac{d}{d\varphi}\left(N(\varphi)\cos\varphi\right) = \\ &= -MN\sin 2\varphi. \end{aligned}$$

Appendix C

The Christoffel symbols from the metric

Let us start from the definition of the metric in Chapter 9.2:

$$g_{ij} = \left\langle \frac{\partial \mathbf{x}}{\partial u^i} \cdot \frac{\partial \mathbf{x}}{\partial u^j} \right\rangle = \left[\begin{array}{cc} \left\langle \frac{\partial \mathbf{x}}{\partial u^1} \cdot \frac{\partial \mathbf{x}}{\partial u^1} \right\rangle & \left\langle \frac{\partial \mathbf{x}}{\partial u^1} \cdot \frac{\partial \mathbf{x}}{\partial u^2} \right\rangle \\ \left\langle \frac{\partial \mathbf{x}}{\partial u^2} \cdot \frac{\partial \mathbf{x}}{\partial u^1} \right\rangle & \left\langle \frac{\partial \mathbf{x}}{\partial u^2} \cdot \frac{\partial \mathbf{x}}{\partial u^2} \right\rangle \end{array} \right],$$

where $u^i = \left(u^1, u^2\right)$ is the parametrization ("co-ordinate frame") of a curved surface in space (or more generally a sub-space). Differentiate:

$$\begin{aligned} \frac{\partial}{\partial u^i} g_{jk} &= \frac{\partial}{\partial u^i} \left\langle \frac{\partial \mathbf{x}}{\partial u^j} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle = \\ &= \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle + \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^k} \cdot \frac{\partial \mathbf{x}}{\partial u^j} \right\rangle . \end{aligned} \tag{C.1}$$

Correspondingly, by interchanging indices:

$$\frac{\partial}{\partial u^j} g_{ki} = \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^j \partial u^k} \cdot \frac{\partial \mathbf{x}}{\partial u^i} \right\rangle + \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^j \partial u^i} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle \tag{C.2}$$

$$\frac{\partial}{\partial u^k} g_{ij} = \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^k \partial u^i} \cdot \frac{\partial \mathbf{x}}{\partial u^j} \right\rangle + \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^k \partial u^j} \cdot \frac{\partial \mathbf{x}}{\partial u^i} \right\rangle \tag{C.3}$$

Compute equation (C.1) plus equation (C.2) minus equation (C.3):

$$\frac{\partial}{\partial u^i} g_{jk} + \frac{\partial}{\partial u^j} g_{ki} - \frac{\partial}{\partial u^k} g_{ij} = 2 \left\langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle . \tag{C.4}$$

Let's write the second derivatives of $\mathbf{x}$ on the local base $\left(\frac{\partial \mathbf{x}}{\partial u^1}, \frac{\partial \mathbf{x}}{\partial u^2}, \mathbf{n} \right)$ as we did in equation (9.7), even if in a slightly different notation:

$$\frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j} \equiv \Gamma^{\ell}_{ij} \frac{\partial \mathbf{x}}{\partial u^{\ell}} + \beta_{ij} \mathbf{n}, \tag{C.5}$$

which implicitly defines the Γ symbols. Substitution into equation (C.4) yields

$$\Gamma^{\ell}_{ij} \left\langle \frac{\partial \mathbf{x}}{\partial u^{\ell}} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle + \beta_{ij} \left\langle \mathbf{n} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle = \frac{1}{2} \left(\frac{\partial}{\partial u^i} g_{jk} + \frac{\partial}{\partial u^j} g_{ki} - \frac{\partial}{\partial u^k} g_{ij} \right).$$

Here we recognise

$$\left\langle \frac{\partial \mathbf{x}}{\partial u^{\ell}} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle = g_{\ell k} \text{ ja } \left\langle \mathbf{n} \cdot \frac{\partial \mathbf{x}}{\partial u^k} \right\rangle = 0,$$

or

$$\begin{aligned} \Gamma^{\ell}_{ij} g_{\ell k} &= \frac{1}{2} \left(\frac{\partial}{\partial u^i} g_{jk} + \frac{\partial}{\partial u^j} g_{ki} - \frac{\partial}{\partial u^k} g_{ij} \right) \Rightarrow \\ \Rightarrow \Gamma^{\ell}_{ij} &= \frac{1}{2} g^{\ell k} \left(\frac{\partial}{\partial u^i} g_{jk} + \frac{\partial}{\partial u^j} g_{ki} - \frac{\partial}{\partial u^k} g_{ij} \right), \end{aligned}$$

equation (10.7). Here, $g^{ij} g_{jk} = \delta^i_k$, i.e., g^{ij} is the inverse matrix of g_{ij}.

Appendix D

The Riemann tensor from the Christoffel symbols

The RIEMANN tensor equation is derived with the age of parallel transport of a vector around a closed, small co-ordinate rectangle. Of the spatial vector $\mathbf{v}$ is transported parallelly inside a u^i-parametrized surface S , the following holds

$$\frac{\partial \mathbf{v}}{\partial u^i} = 0.$$

Let us write $\mathbf{v}$ on the basis of the tangent vectors:

$$\mathbf{v} = v^i \frac{\partial \mathbf{x}}{\partial u^i}.$$

Then

$$\begin{aligned} 0 = \frac{\partial \mathbf{v}}{\partial u^j} &= \frac{\partial v^i}{\partial u^j}\frac{\partial \mathbf{x}}{\partial u^i} + v^i \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j} = \\ &= \frac{\partial v^i}{\partial u^j}\frac{\partial \mathbf{x}}{\partial u^i} + \Gamma^i_{jk} v^k \frac{\partial \mathbf{x}}{\partial u^i} + \beta_{ij}\mathbf{n}, \end{aligned}$$

using (C.5). Here, the $\mathbf{v}$ derivative consists of two parts: the “interior” part, $\frac{\partial v^i}{\partial u^j}\frac{\partial \mathbf{x}}{\partial u^i} + \Gamma^i_{jk} v^k \frac{\partial \mathbf{x}}{\partial u^i}$, embedded in the surface, and the “exterior” part, $\beta_{ij}\mathbf{n}$, aperpendicular to the surface. When the surface S has been given, we can only zero the internal part, i.e.

$$\frac{\partial v^i}{\partial u^j} + \Gamma^i_{jk} v^k = 0 \tag{D.1}$$

describes the parallel transport of the vector v^i within the surface.

Let us now consider a small rectangle $ABCD$, side lengths Δu^k and Δu^ℓ (see Fig. 10.3), along co-ordinate curves. The sides AB and CD are on opposite sides, the running co-ordinate being u^k. Similarly BC and AD are opposite, the running co-ordinate being u^ℓ.

Derive the change in v^iover the distance AB :

$$\Delta_{AB} v^i = \frac{\partial v^i}{\partial u^k}\Delta u^k = -\Gamma^i_{km} v^m \Delta u^k.$$

In the same way

$$\Delta_{CD} v^i = +\Gamma^i_{km} v^m \Delta u^k.$$

For the side BC we obtain

$$\Delta_{BC} v^i = \frac{\partial v^i}{\partial u^\ell}\Delta u^\ell = -\Gamma^i_{\ell m} v^m \Delta u^\ell$$

and

$$\Delta_{DA} v^i = +\Gamma^i_{\ell m} v^m \Delta u^\ell.$$

sum together these four terms:

$$\begin{aligned}
\Delta_{ABCD}v^i &= \left\{\left(\Gamma^i_{km}v^m\right)_{CD} - \left(\Gamma^i_{km}v^m\right)_{AB}\right\}\Delta u^k - \left\{\left(\Gamma^i_{\ell m}v^m\right)_{DA} - \left(\Gamma^i_{\ell m}v^m\right)_{BC}\right\}\Delta u^\ell = \\
&= \left\{\frac{\partial}{\partial u^\ell}\left(\Gamma^i_{km}v^m\right)\Delta u^\ell\right\}\Delta u^k - \left\{\frac{\partial}{\partial u^k}\left(\Gamma^i_{\ell m}v^m\right)\Delta u^k\right\}\Delta u^\ell = \\
&= \left\{\left(\frac{\partial\Gamma^i_{km}}{\partial u^\ell} - \frac{\partial\Gamma^i_{\ell m}}{\partial u^k}\right)v^m + \Gamma^i_{km}\frac{\partial v^m}{\partial u^\ell} - \Gamma^i_{\ell m}\frac{\partial v^m}{\partial u^k}\right\}\Delta u^\ell \Delta u^k.
\end{aligned}$$

Equation (D.1) gives

$$\frac{\partial v^m}{\partial u^\ell} = -\Gamma^m_{\ell h}v^h, \ \frac{\partial v^m}{\partial u^k} = -\Gamma^m_{kh}v^h;$$

substituting:

$$\begin{aligned}
\Delta_{ABCD}v^i &= \left\{\left(\frac{\partial\Gamma^i_{km}}{\partial u^l} - \frac{\partial\Gamma^i_{lm}}{\partial u^k}\right)v^m + \left(\Gamma^i_{lm}\Gamma^m_{kh} - \Gamma^i_{km}\Gamma^m_{lh}\right)v^h\right\}\Delta u^\ell \Delta u^k = \\
&= \left(\frac{\partial\Gamma^i_{kj}}{\partial u^\ell} - \frac{\partial\Gamma^i_{\ell j}}{\partial u^k} + \Gamma^i_{\ell m}\Gamma^m_{kj} - \Gamma^i_{km}\Gamma^m_{\ell j}\right)v^j \Delta u^\ell \Delta u^k,
\end{aligned}$$

where we have changed the names of the indices $m \to j$ (in the first two terms) ja $h \to j$ (in the last two terms).

Here we see in ready form the RIEMANN *curvature tensor*:

$$R^i_{j\ell k} = \frac{\partial\Gamma^i_{kj}}{\partial u^l} - \frac{\partial\Gamma^i_{\ell j}}{\partial u^k} + \Gamma^i_{\ell m}\Gamma^m_{kj} - \Gamma^i_{km}\Gamma^m_{\ell j},$$

aapart from the names of the indices and interchanges of type $\Gamma^i_{jk} = \Gamma^i_{kj}$, just what already was given in eq. (10.9).

Bibliography

Anon. (2003). JHS154. ETRS89 -järjestelmään liittyvät karttaprojektiot, tasokoordinaatistot ja karttalehtijako (Map projections, plane co-ordinates and map sheet division in relation to the ETRS89 system). Web site, Finnish National Land Survey. URL: `http://www.jhs-suositukset.fi/suomi/jhs154`, accessed August 30, 2005.

Anon. (2008). JHS153. ETRS89-jr̈jestelmn̈ mukaiset koordinaatit Suomessa. Web site, Finnish National Land Survey. URL: `http://www.jhs-suositukset.fi/suomi/jhs153`, accessed Oct. 19, 2010.

Bomford, G. (1963). Report of Study Group No. 10. The Geoid in Europe and Connected Countries. In *Traveaux de l'Association Internationale de Géodésie*, number 22.

Boucher, C. and Altamimi, Z. *Specifications for reference frame fixing in the analysis of a EUREF GPS campaign.* Version 4, 24-10-2008, `http://etrs89.ensg.ign.fr/memo-V7.pdf`.

Boucher, C. and Altamimi, Z. (1995). Specifications for Reference Frame Fixing in the Analysis of a EUREF GPS Campaign. In *VI Meeting of the TWG in Bern, March 9 and 10, Veröff. der Bayerischen Kommission für die Internationale Erdmessung*, number 56, pages 265–267, München.

Gross, R. S. (2000). The excitation of the Chandler wobble. *Geophys. Res. Lett.*, 27:2329–2332.

Grossman, W. (1964). *Geodätische Rechnungen und Abbildungen.* Verlag Konrad Wittwer, Stuttgart.

Heikkinen, M. (1981). Solving the shape of the Earth by using digital density models. Report 81:2, Finnish Geodetic Institute, Helsinki.

Heiskanen, W. A. and Moritz, H. (1967). *Physical Geodesy.* W.H. Freeman and Company, San Francisco.

Kumar, M. and Reilly, J. P. (2006). Is definition of WGS 84 correct. URL: http://mycoordinates.org/is-definition-of-wgs-84-correct/.

McCarthy, D. D. (1996). IERS Conventions (1996). (IERS Technical Note ; 21). Technical report, Central Bureau of IERS - Observatoire de Paris. URL: http://www.iers.org/nn_11216/IERS/EN/Publications/TechnicalNotes/tn21.html.

Moritz, H. (1992). Geodetic Reference System 1980. In Tscherning, C. C., editor, *Geodesist's Handbook 1992*, pages 187–192. International Association of Geodesy, Copenhagen.

Ollikainen, M. (1993). GPS-koordinaattien muuntaminen Kartastokoordinaateiksi (Transforming GPS co-ordinates to the National Grid Co-ordinate System). Tiedote 8, Finnish Geodetic Institute, Helsinki.

Ollikainen, M., Koivula, H., and Poutanen, M. (2000). The densification of the EUREF network in Finland. Publication 129, Finnish Geodetic Institute, Kirkkonummi.

Parm, T. (1988). Kansallisen koordinaattijärjestelmän luominen Suomessa (The creation of a national co-ordinate system in Finland). *Maanmittaus*, (1).

Stevenson, P. (2008). The Truth About WGS84. URL: http://www.ashgps.com/out3/General