Lecture notes to the courses:
Messverfahren der Erdmessung und Physikalischen Geodäsie
Modellbildung und Datenanalyse in der Erdmessung und Physikalischen Geodäsie

Physical Geodesy

Nico Sneeuw
Institute of Geodesy
Universität Stuttgart

15th June 2006

These are lecture notes in progress. Please contact me (`sneeuw@gis.uni-stuttgart.de`) for remarks, errors, suggestions, etc.

Contents

1. Introduction

1.1. Physical Geodesy

Geodesy aims at the determination of the geometrical and physical shape of the Earth and its orientation in space. The branch of geodesy that is concerned with determining the physical shape of the Earth is called *physical geodesy*. It does interact strongly with the other branches, though, as will be seen later.

Physical geodesy is different from other geomatics disciplines in that it is concerned with field quantities: the scalar potential field or the vectorial gravity and gravitational fields. These are continuous quantities, as opposed to point fields, networks, pixels, etc., which are discrete by nature.

Gravity field theory uses a number of tools from mathematics and physics:

- Newtonian gravitation theory (relativity is not required for now)
- Potential theory
- Vector calculus
- Special functions (Legendre)
- Partial differential equations
- Boundary value problems
- Signal processing

Gravity field theory is interacting with many other disciplines. A few examples may clarify the importance of physical geodesy to those disciplines. The Earth sciences disciplines are rather operating on a global scale, whereas the engineering applications are more local. This distinction is not fundamental, though.

1.2. Links to Earth sciences

Oceanography. The Earth's gravity field determines the geoid, which is the equipotential surface at mean sea level. If the oceans would be at rest—no waves, no currents, no tides—the ocean surface would coincide with the geoid. In reality it deviates by up to 1 m. The difference is called *sea surface topography*. It reflects the dynamical equilibrium in the oceans. Only large scale currents can sustain these deviations.

The sea surface itself can be accurately measured by radar altimeter satellites. If the geoid would be known up to the same accuracy, the sea surface topography and consequently the global ocean circulation could be determined. The problem is the insufficient knowledge of the marine geoid.

Geophysics. The Earth's gravity field reflects the internal mass distribution, the determination of which is one of the tasks of geophysics. By itself gravity field knowledge is insufficient to recover this distribution. A given gravity field can be produced by an infinity of mass distributions. Nevertheless, gravity is is an important constraint, which is used together with seismic and other data.

As an example, consider the gravity field over a volcanic island like Hawaii. A volcano by itself represents a geophysical anomaly already, which will have a gravitational signature. Over geologic time scales, a huge volcanic mass is piled up on the ocean sphere. This will cause a bending of the ocean floor. Geometrically speaking one would have a cone in a bowl. This bowl is likely to be filled with sediment. Moreover the mass load will be supported by buoyant forces within the mantle. This process is called *isostasy*. The gravity signal of this whole mass configuration carries clues to the density structure below the surface.

Geology. Different geological formations have different density structures and hence different gravity signals. One interesting example of this is the Chicxulub crater, partially on the Yucatan peninsula (Mexico) and partially in the Gulf of Mexico. This crater with a diameter of 180 km was caused by a meteorite impact, which occurred at the K-T boundary (cretaceous-tertiary) some 66 million years ago. This impact is thought to have caused the extinction of dinosaurs. The Chicxulub crater was discovered by careful analysis of gravity data.

Hydrology. Minute changes in the gravity field over time—after correcting for other time-variable effects like tides or atmospheric loading—can be attributed to changes in hydrological parameters: soil moisture, water table, snow load. For static gravimetry these are usually nuisance effects. Nowadays, with precise satellite techniques, hydrology is one of the main aims of spaceborne gravimetry. Despite a low spatial resolution, the results of satellite gravity missions may be used to constrain basin-scale hydrological parameters.

Glaciology and sea level. The behaviour of the Earth's ice masses is a critical indicator of global climate change and global sea level behaviour. Thus, monitoring of the melting of the Greenland and Antarctica ice caps is an important issue. The ice caps are huge mass loads, sitting on the Earth's crust, which will necessarily be depressed. Melting causes a rebound of the crust. This process is still going on since the last Ice Age, but there is also an instant effect from melting taking place right now. The change in surface ice contains a direct gravitational component and an effect, due to the uplift. Therefore, precise gravity measurements carry information on ice melting and consequently on sea level rise.

1.3. Applications in engineering

Geophysical prospecting. Since gravity contains information on the subsurface density structure, gravimetry is a standard tool in the oil and gas industry (and other mineral resources for that matter). It will always be used together with seismic profiling, test drilling and magnetometry. The advantages of gravimetry over these other techniques are:

- relatively inexpensive,
- non destructive (one can easily measure inside buildings),
- compact equipment, e.g. for borehole measurements

Gravimetry is used to localize salt domes or fractures in layers, to estimate depth, and in general to get a first idea of the subsurface structure.

Geotechnical Engineering. In order to gain knowledge about the subsurface structure, gravimetry is a valuable tool for certain geotechnical (civil) engineering projects. One can think of determining the depth-to-bedrock for the layout of a tunnel. Or making sure no subsurface voids exist below the planned building site of a nuclear power plant.

For examples, see the (micro-)gravity case histories and applications on:
`http://www.geop.ubc.ca/ubcgif/casehist/index.html`, or
`http://www.esci.keele.ac.uk/geophysics/Research/Gravity/`.

Geomatics Engineering. Most surveying observables are related to the gravity field.

i) After leveling a **theodolite** or a **total station**, its vertical axis is automatically aligned with the local gravity vector. Thus all measurements with these instruments are referenced to the gravity field—they are in a local astronomic frame. To convert them to a geodetic frame the deflection of the vertical (ξ,η) and the perturbation in azimuth (ΔA) must be known.

ii) The line of sight of a **level** is tangent to the local equipotential surface. So levelled height differences are really physical height differences. The basic quantity of physical heights are the potentials or the potential differences. To obtain precise height differences one should also use a gravimeter:

$$\Delta W = \int_A^B \boldsymbol{g} \cdot \mathrm{d}\boldsymbol{x} = \int_A^B g \, \mathrm{d}h \approx \sum_i g_i \Delta h_i \, .$$

The Δh_i are the levelled height increments. Using gravity measurements g_i along the way gives a geopotential difference, which can be transformed into a physical height difference, for instance an orthometric height difference.

iii) **GPS positioning** is a geometric techniques. The geometric GPS heights are related to physically meaningful heights through the geoid or the quasi-geoid:

$$\begin{aligned} h &= H + N = \text{orthometric height + geoid height}, \\ h &= H^{\mathrm{n}} + \zeta = \text{normal height + quasi-geoid height}. \end{aligned}$$

In geomatics engineering, GPS measurements are usually made over a certain baseline and processed in differential mode. In that case, the above two formulas become $\Delta h = \Delta N + \Delta H$, etc. The geoid difference between the baseline's endpoints must therefore be known.

iv) The basic equation of **inertial surveying** is $\ddot{\boldsymbol{x}} = \boldsymbol{a}$, which is integrated twice to provide the trajectory $\boldsymbol{x}(t)$. The equation says that the kinematic acceleration equals the specific force vector $\boldsymbol{a}$: the sum of all forces (per unit mass) acting on a proof mass). An inertial measurement unit, though, measures the sum of kinematic acceleration and gravitation. Thus the gravitational field must be corrected for, before performing the integration.

2. Gravitation

2.1. Newtonian gravitation

In 1687 Newton[1] published his *Philosophiae naturalis principia mathematica*, or *Principia* in short. The Latin title can be translated as *Mathematical principles of natural philosophy*, in which *natural philosophy* can be read as *physics*. Although Newton was definitely not the only physicist working on gravitation in that era, his name is nevertheless remembered and attached to gravity because of the *Principia*. The greatness of this work lies in the fact that Newton was able to bring empirical observations on a mathematical footing and to explain in a unifying manner many natural phenomena:

- planetary motion (in particular elliptical motion, as discovered by Kepler[2]),
- free fall, e.g. the famous apple from the tree,
- tides,
- equilibrium shape of the Earth.

Newton made fundamental observations on gravitation:

- The force between two attracting bodies is proportional to the individual masses.
- The force is inversely proportional to the square of the distance.
- The force is directed along the line connecting the two bodies.

Mathematically, the first two are translated into:

$$F_{12} = G\frac{m_1 m_2}{r_{12}^2}\,, \tag{2.1}$$

[1] Sir Isaac Newton (1642–1727).

[2] Johannes Kepler (1571–1630), German astronomer and mathematician; formulated the famous laws of planetary motion: *i)* orbits are ellipses with Sun in one of the foci, *ii)* the areas swept out by the line between Sun and planet are equal over equal time intervals (area law), and *iii)* the ratio of the cube of the semi-major axis and the square of the orbital period is constant (or $n^2a^3 = GM$).

in which G is a proportionality factor. It is called the *gravitational constant* or *Newton constant*. It has a value of $G = 6.672 \cdot 10^{-11}\,\mathrm{m^3 s^{-2} kg^{-1}}$ (or $\mathrm{N\,m^2\,kg^{-2}}$).

Remark 2.1 (mathematical model of gravitation) *Soon after the publication of the* Principia *Newton was strongly criticized for his law of gravitation, e.g. by his contemporary Huygens. Equation (2.1) implies that gravitation acts at a distance, and that it acts instantaneously. Such action is unphysical in a modern sense. For instance, in Einstein's relativity theory no interaction can be faster than the speed of light. However, Newton did not consider his formula (2.1) as some fundamental law. Instead, he saw it as a convenient mathematical description. As such, Newton's law of gravitation is still a viable model for gravitation in physical geodesy.*

Equation (2.1) is symmetric: the mass m_1 exerts a force on m_2 and m_2 exerts a force of the same magnitude but in opposite direction on m_1. From now on we will be interested in the gravitational field generated by a single test mass. For that purpose we set $m_1 := m$ and we drop the indices. The mass m_2 can be an arbitrary mass at an arbitrary location. Thus we eliminate m_2 by $a = F/m_2$. The gravitational attraction a of m becomes:

$$a = G\frac{m}{r^2}\,, \tag{2.2}$$

in which r is the distance between mass point and evaluation point. The gravitational attraction has units $\mathrm{m/s^2}$. In geodesy one often uses the unit Gal, named after Galileo[3]:

$$\begin{aligned} 1\,\mathrm{Gal} &= 10^{-2}\,\mathrm{m/s^2} = 1\,\mathrm{cm/s^2} \\ 1\,\mathrm{mGal} &= 10^{-5}\,\mathrm{m/s^2} \\ 1\,\mu\mathrm{Gal} &= 10^{-8}\,\mathrm{m/s^2}\,. \end{aligned}$$

Remark 2.2 (kinematics vs. dynamics) *The gravitational attraction is not an acceleration. It is a dynamical quantity: force per unit mass or specific force. Accelerations on the other hand are kinematic quantities.*

2.1.1. Vectorial attraction of a point mass

The gravitational attraction works along the line connecting the point masses. In this symmetrical situation the attraction at point 1 is equal in size, but opposite in direction, to the attraction at point 2: $\boldsymbol{a}_{12} = -\boldsymbol{a}_{21}$. This corresponds to Newton's law: *action* $=$ $-$*reaction*.

[3] Galileo Galilei (1564–1642).

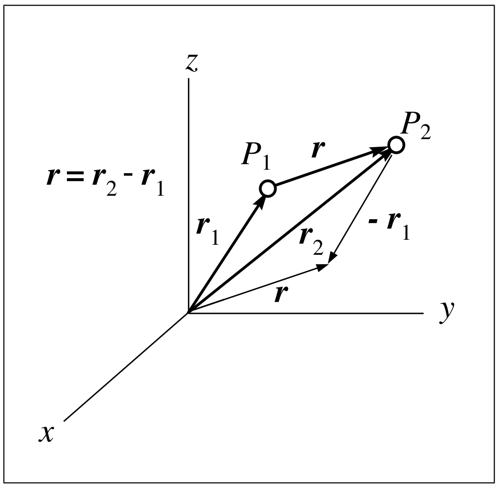


Figure 2.1: Attraction of a point mass m, located in point P_1, on P_2.

In case we have only one point mass m, located in $\boldsymbol{r}_1$, whose attraction is evaluated in point $\boldsymbol{r}_2$, this symmetry is broken. The vector $\boldsymbol{a}$ is considered to be the corresponding attraction.

$$
\begin{aligned}
\boldsymbol{r} &= \boldsymbol{r}_2 - \boldsymbol{r}_1 = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix} \quad \text{, and } r = |\boldsymbol{r}| \\
\boldsymbol{a} &= -G\frac{m}{r^2}\boldsymbol{e}_{12} \;=\; -G\frac{m}{r^2}\frac{\boldsymbol{r}}{r} \;=\; -G\frac{m}{r^3}\boldsymbol{r} \\
&= -G\frac{m}{\left[(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2\right]^{3/2}}\begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix}.
\end{aligned}
$$

2.1.2. Gravitational potential

The gravitational attraction field $\boldsymbol{a}$ is a conservative field. This means that the same amount of work has to be done to go from point A to point B, no matter which path you take. Mathematically, this is expressed by the fact that the field $\boldsymbol{a}$ is curl-free:

$$\operatorname{rot}\boldsymbol{a} = \nabla \times \boldsymbol{a} = \boldsymbol{0}\,. \tag{2.3}$$

Now from vector analysis it is known that the curl of any gradient field is always equal to zero: $\operatorname{rot}\operatorname{grad} F = \nabla \times \nabla F = \boldsymbol{0}$. Therefore, $\boldsymbol{a}$ can be written as a gradient of some scalar field. This scalar field is called gravitational potential V. The amount of energy it takes (or can be gained) to go from A to B is simply $V_B - V_A$. Instead of having to deal with a vector field (3 numbers at each point) the gravitational field is fully described by a scalar field (1 number).

The gravitational potential that would generate $\boldsymbol{a}$ can be derived by evaluating the amount of work (per unit mass) required to get to location r. We assume that the mass point is at the origin of our coordinate system. Since the integration in a conservative field is path independent we can choose our path in a convenient way. We will start at infinity, where the attraction is zero and go straight along the radial direction to our point $\boldsymbol{r}$.

$$\Delta V = \int_A^B \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{x} \quad =: \quad V = \int_\infty^r -G\frac{m}{r^2}\,\mathrm{d}r = G\frac{m}{r}\bigg|_\infty^r = G\frac{m}{r}\,. \tag{2.4}$$

The attraction is generated from the potential by the gradient operator:

$$\boldsymbol{a} = \operatorname{grad} V = \nabla V = \begin{pmatrix} \frac{\partial V}{\partial x} \\ \frac{\partial V}{\partial y} \\ \frac{\partial V}{\partial z} \end{pmatrix}.$$

That this indeed leads to the same vector field is demonstrated by performing the partial differentiations, e.g. for the x-coordinate:

$$\begin{aligned} \frac{\partial V}{\partial x} &= Gm\frac{\partial \frac{1}{r}}{\partial x} = Gm\frac{\partial \frac{1}{r}}{\partial r}\frac{\partial r}{\partial x} \\ &= Gm\left(-\frac{1}{r^2}\right)\left(\frac{2x}{2r}\right) = -G\frac{m}{r^3}x\,, \end{aligned}$$

and similarly for y and z.

2.1.3. Superposition—discrete

Gravitational formulae were derived for single point masses so far. One important property of gravitation is the so-called superposition principle. It says that the gravitational potential of a system of masses can be achieved simply by adding the potentials of single masses. In general we have:

$$V = \sum_{i=1}^{N} V_i = G\frac{m_1}{r_1} + G\frac{m_2}{r_2} + \ldots + G\frac{m_N}{r_N} = G\sum_{i=1}^{N}\frac{m_i}{r_i}\,. \tag{2.5}$$

The m_i are the single masses and the r_i are the distances between masspoints and the evaluation point. The total gravitational attraction is simply obtained by $\boldsymbol{a} = \nabla V$ again:

$$\boldsymbol{a} = \nabla V = \sum_i \nabla V_i = -G\sum_i \frac{m_i}{r_i^3}\boldsymbol{r}_i\,. \tag{2.6}$$

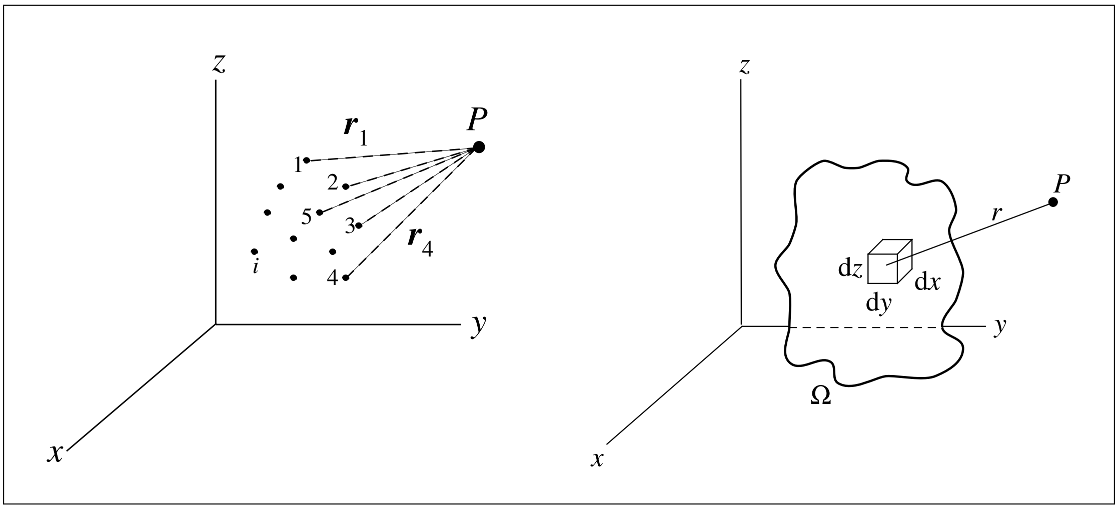


Figure 2.2.: Superposition for discrete (left) and continuous (right) mass distributions.

2.1.4. Superposition—continuous

Real world mass configurations can be thought of as systems of infinitely many and infinitely close point masses. The discrete formulation will become a continuous one.

$$N \to \infty$$
$$\sum_i \to \iiint\limits_\Omega$$
$$m_i \to \mathrm{d}m$$

The body Ω consists of mass elements $\mathrm{d}m$, that are the infinitesimal masses of infinitesimal cubes $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ with local density $\rho(x, y, z)$:

$$\mathrm{d}m(x, y, z) = \rho(x, y, z)\ \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,. \tag{2.7}$$

Integrating over all mass elements in Ω—the continuous equivalent of superposition—gives the potential generated by Ω:

$$V_P = G \iiint\limits_\Omega \frac{\mathrm{d}m}{r} = G \iiint\limits_\Omega \frac{\rho(x, y, z)}{r}\ \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,, \tag{2.8}$$

with r the distance between computation point P and mass element $\mathrm{d}m$. Again, the gravitational attraction of Ω is obtained by applying the gradient operator:

$$\boldsymbol{a} = \nabla V = -G \iiint\limits_\Omega \frac{\rho(x, y, z)}{r^3} \boldsymbol{r}\ \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,. \tag{2.9}$$

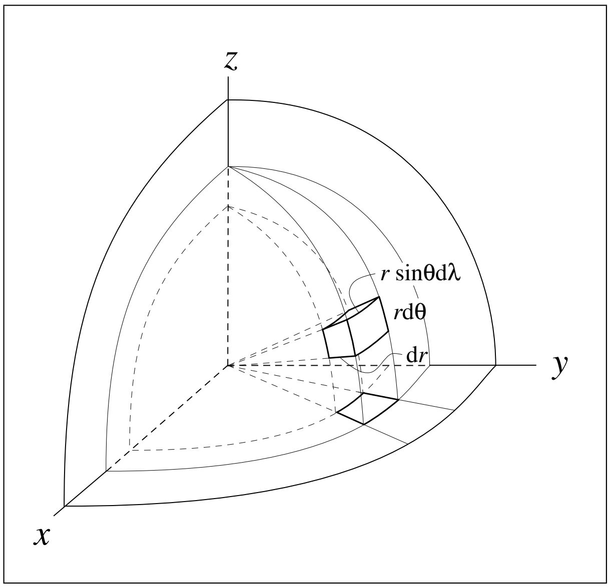


Figure 2.3: One octant of a solid sphere. The volume element has sides $\mathrm{d}r$ in radial direction, $r\,\mathrm{d}\theta$ in co-latitude direction and $r\sin\theta\,\mathrm{d}\lambda$ in longitude direction.

The potential (2.8) and the attraction (2.9) can in principle be determined using volume integrals if the density distribution within the body Ω is known. However, we can obviously not apply these integrals to the real Earth. The Earth's internal density distribution is insufficiently known. For that reason we will make use of potential theory to turn the volume integrals into surface integrals in a later chapter.

2.2. Ideal solids

Using the general formulae for potential and attraction, we will investigate the gravitational effect of some ideal solid bodies now.

2.2.1. Solid homogeneous sphere

Consider a sphere of radius R with homogeneous density $\rho(x,y,z) = \rho$. In order to evaluate the integrals (2.8) we assume a coordinate system with its origin at the centre of the sphere. Since a sphere is rotationally symmetric we can evaluate the gravitational potential at an arbitrary point. Our choice is a general point P on the positive z-axis. Thus we have for evaluation point P and mass point Q the following vectors:

$$\boldsymbol{r}_P = \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}, \; r_P = z\,,$$

$$\boldsymbol{r}_Q = \begin{pmatrix} r\sin\theta\cos\lambda \\ r\sin\theta\sin\lambda \\ r\cos\theta \end{pmatrix}, \; r_Q = r < R\,,$$
$$\boldsymbol{r}_{PQ} = \boldsymbol{r}_Q - \boldsymbol{r}_P = \begin{pmatrix} r\sin\theta\cos\lambda \\ r\sin\theta\sin\lambda \\ r\cos\theta - z \end{pmatrix}, \; r_{PQ} = \sqrt{z^2 + r^2 - 2rz\cos\theta}\,. \tag{2.10}$$

It is easier to integrate in spherical coordinates than in Cartesian[4] ones. Thus we use the radius r, co-latitude θ and longitude λ. The integration bounds become $\int\limits_{r=0}^{R}\int\limits_{\theta=0}^{\pi}\int\limits_{\lambda=0}^{2\pi}$ and the volume element $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ is replaced by $r^2\sin\theta\,\mathrm{d}\lambda\,\mathrm{d}\theta\,\mathrm{d}r$. Applying this change of coordinates to (2.8) and putting the constant density ρ outside of the integral, yields the following integration:

$$\begin{aligned} V_P &= G\rho\iiint \frac{1}{r_{PQ}}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ &= G\rho\int\limits_{r=0}^{R}\int\limits_{\theta=0}^{\pi}\int\limits_{\lambda=0}^{2\pi}\frac{1}{r_{PQ}}r^2\sin\theta\,\mathrm{d}\lambda\,\mathrm{d}\theta\,\mathrm{d}r \\ &= G\rho\int\limits_{r=0}^{R}\int\limits_{\theta=0}^{\pi}\int\limits_{\lambda=0}^{2\pi}\frac{r^2\sin\theta}{\sqrt{z^2+r^2-2rz\cos\theta}}\,\mathrm{d}\lambda\,\mathrm{d}\theta\,\mathrm{d}r \\ &= 2\pi G\rho\int\limits_{r=0}^{R}\int\limits_{\theta=0}^{\pi}\frac{r^2\sin\theta}{\sqrt{z^2+r^2-2rz\cos\theta}}\,\mathrm{d}\theta\,\mathrm{d}r\,. \end{aligned}$$

The integration over λ was trivial, since λ doesn't appear in the integrand. The integration over θ is not straightforward, though. A good trick is to change variables. Call r_{PQ} (2.10) l now. Then

$$\begin{aligned} \frac{\mathrm{d}l}{\mathrm{d}\theta} = \frac{\mathrm{d}\sqrt{z^2+r^2-2rz\cos\theta}}{\mathrm{d}\theta} = \frac{zr\sin\theta}{l} &: \frac{l}{zr}\,\mathrm{d}l = \sin\theta\,\mathrm{d}\theta \\ &: r^2\sin\theta\,\mathrm{d}\theta = \frac{rl}{z}\,\mathrm{d}l\,. \end{aligned}$$

Thus the integral becomes:

$$V_P = 2\pi G\rho\int\limits_{r=0}^{R}\int\limits_{l=l_-}^{l_+}\frac{r}{z}\,\mathrm{d}l\,\mathrm{d}r\,. \tag{2.11}$$

[4]René Descartes or Cartesius (1596–1650), French mathematician, scientist and philosopher whose work *La géométrie* (1637), includes his application of algebra to geometry from which we now have Cartesian geometry.

The integration bounds of $\int_l$ have to be determined first. We have to distinguish two cases.

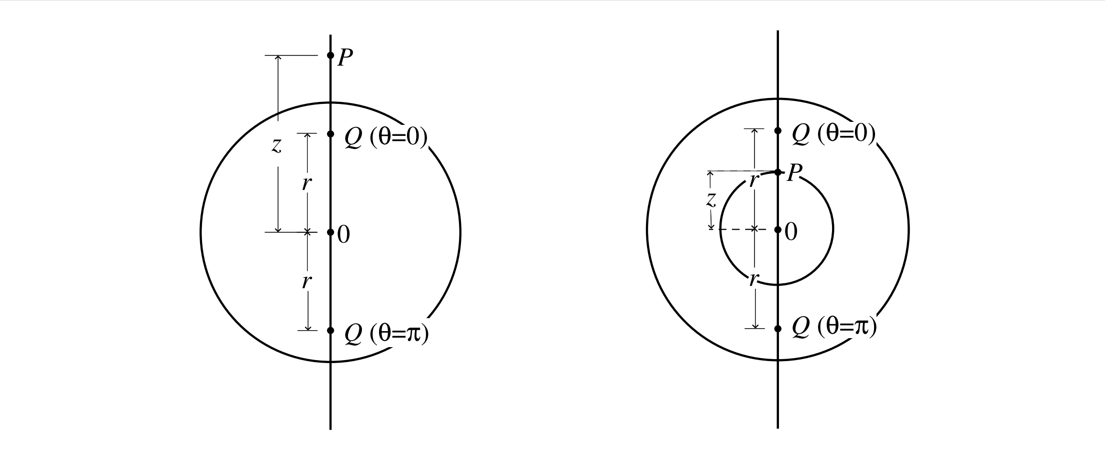


Figure 2.4.: Determining the integration limits for variable l when the evaluation point P is outside (left) or inside (right) the solid sphere.

Point P outside the sphere ($z > R$): From fig. 2.4 (left) the integration bounds for l become immediately clear:

$$\theta = 0 : l_- = z - r$$
$$\theta = \pi : l_+ = z + r$$

$$\begin{aligned} V_P &= 2\pi G\rho \int_{r=0}^{R} \int_{z-r}^{z+r} \frac{r}{z}\, \mathrm{d}l\, \mathrm{d}r \;=\; 2\pi G\rho \int_{r=0}^{R} \left[\frac{r}{z}\, l \right]_{z-r}^{z+r} \mathrm{d}r \\ &= 2\pi G\rho \int_{r=0}^{R} \frac{2r^2}{z}\, \mathrm{d}r \;=\; \frac{4}{3}\pi G\rho \frac{R^3}{z} \,. \end{aligned}$$

We chose the evaluation point P arbitrary on the z-axis. In general, we can replace z by r now because of radial symmetry. Thus we obtain:

$$V(r) = \frac{4}{3}\pi G\rho R^3 \frac{1}{r} \,. \tag{2.12}$$

Recognizing that the mass M of a sphere equals $\frac{4}{3}\pi\rho R^3$, we simply obtain $V = \frac{GM}{r}$. So the potential of a solid sphere equals that of a point mass, at least outside the sphere.

Point P within sphere ($z < R$): For this situation we must distinguish between mass points below the evaluation point ($r < z$) and mass point outside ($z < r < R$). The former configuration would be a sphere of radius z. Its potential in point $P(= [0, 0, z])$ is

$$V_P = \frac{4}{3}\pi G\rho\frac{z^3}{z} = \frac{4}{3}\pi G\rho z^2\,. \tag{2.13}$$

For the masses outside P we have the following integration bounds for l:

$$\theta = 0 : l_- = r - z\,,$$
$$\theta = \pi : l_+ = r + z\,.$$

The integration over r runs from z to R. With the same change of variables we obtain

$$\begin{aligned} V_P &= 2\pi G\rho \int_{r=z}^{R}\int_{r-z}^{r+z} \frac{r}{z}\,\mathrm{d}l\,\mathrm{d}r \;=\; 2\pi G\rho \int_{r=z}^{R} \left[\frac{r}{z}\,l\right]_{r-z}^{r+z} \mathrm{d}r \\ &= 2\pi G\rho \int_{r=z}^{R} 2r\;\mathrm{d}r \;=\; 2\pi G\rho \left[r^2\right]_z^R \;=\; 2\pi G\rho(R^2 - z^2)\,. \end{aligned}$$

The combined effect of the smaller sphere ($r < z$) and spherical shell ($z < r < R$) is:

$$V_P = \frac{4}{3}\pi G\rho z^2 + 2\pi G\rho(R^2 - z^2) = 2\pi G\rho(R^2 - \frac{1}{3}z^2)\,. \tag{2.14}$$

Again we can replace z now by r. In summary, the gravitational potential of a sphere of radius R reads

$$\text{outside:} \qquad V(r > R) = \frac{4}{3}\pi G\rho R^3\frac{1}{r}\,, \tag{2.15a}$$

$$\text{inside:} \qquad V(r < R) = 2\pi G\rho(R^2 - \frac{1}{3}r^2)\,. \tag{2.15b}$$

Naturally, at the boundary the potential will be continuous. This is verified by putting $r = R$ in both equations, yielding:

$$V(R) = \frac{4}{3}\pi G\rho R^2\,. \tag{2.16}$$

This result is visualized in fig. 2.5. Not only is the potential continuous across the surface of the sphere, it is also smooth.

Attraction. It is very easy now to find the attraction of a solid sphere. It simply is the radial derivative. Since the direction is radially towards the sphere's center (the origin) we only need to deal with the radial component:

$$\text{outside:} \quad a(r > R) = -\frac{4}{3}\pi G\rho R^3 \frac{1}{r^2}\,, \tag{2.17a}$$

$$\text{inside:} \quad a(r < R) = -\frac{4}{3}\pi G\rho r\,. \tag{2.17b}$$

Continuity at the boundary is verified by

$$a(R) = -\frac{4}{3}\pi G\rho R\,. \tag{2.18}$$

Again, the result is visualized in fig. 2.5. Although the attraction is continuous across the boundary, it is not differentiable anymore.

Exercise 2.1 *Given the gravitation* $a = 981\,\mathrm{Gal}$ *on the surface of a sphere of radius* $R = 6378\,\mathrm{km}$, *calculate the mass of the Earth* M_E *and its mean density* ρ_E.

Exercise 2.2 *Consider the Earth a homogeneous sphere with mean density* ρ_E. *Now assume a hole in the Earth through the Earth's center connecting two antipodal points on the Earth's surface. If one would jump into this hole: what type of motion arises? How long does it take to arrive at the other side of the Earth? What (and where) is the maximum speed?*

Exercise 2.3 *Try to find more general gravitational formulae for* V *and* a *for the case that the density is not constant but depends on the radial distance:* $\rho = \rho(r)$. *First, set up the integrals and then try to solve them.*

2.2.2. Spherical shell

A spherical shell is a hollow sphere with inner radius R_1 and outer radius R_2. The gravitational potential of it may be found analogous to the derivations in 2.2.1. Of course the proper integration bounds should be used. However, due to the superposition principle, we can simply consider a spherical shell to be the difference between two solid spheres. Symbolically, we could write:

$$\text{spherical shell}(R_1, R_2) = \text{sphere}(R_2) - \text{sphere}(R_1)\,. \tag{2.19}$$

By substracting equations (2.15) and (2.17) with the proper radii, one arrives at the potential and attraction of the spherical shell. One has to be careful in the area $R_1 <$

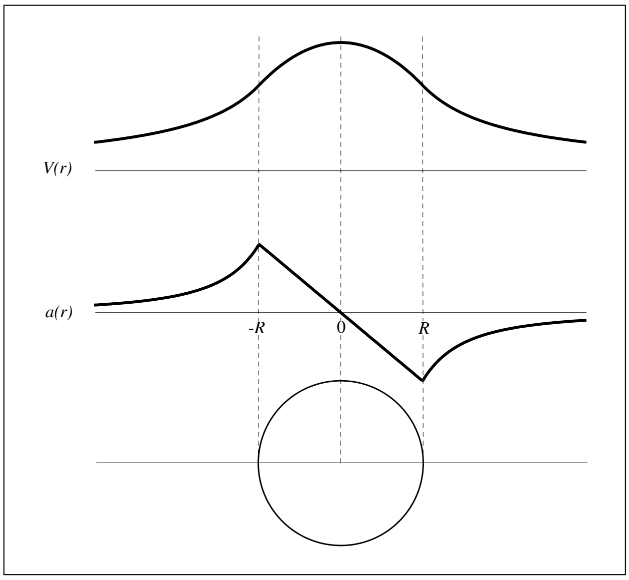


Figure 2.5: Potential V and attraction a as a function of r, due to a solid homogeneous sphere of radius R.

$r < R_2$, though. We should pick the outside formula for sphere (R_1) and the inside formula for sphere (R_2).

$$\text{outer part:} \qquad V(r > R_2) = \frac{4}{3}\pi G\rho(R_2^3 - R_1^3)\frac{1}{r}\,, \tag{2.20a}$$

$$\text{in shell:} \quad V(R_1 < r < R_2) = 2\pi G\rho(R_2^2 - \frac{1}{3}r^2) - \frac{4}{3}\pi G\rho R_1^3\frac{1}{r}\,, \tag{2.20b}$$

$$\text{inner part:} \qquad V(r < R_1) = 2\pi G\rho(R_2^2 - R_1^2)\,. \tag{2.20c}$$

Note that the potential in the inner part is constant.

Remark 2.3 *The potential outside the spherical shell with radi R_1 and R_2 and density could also have been generated by a point mass with $M = \frac{4}{3}\pi\rho(R_2^3 - R_1^3)$. But also by a solid sphere of radius R_2 and density $\rho' = \rho(R_2^3 - R_1^3)/R_2^3$. If we would not have seismic data we could never tell if the Earth was hollow or solid.*

Remark 2.4 *Remark 2.3 can be generalized. If the density structure within a sphere of radius R is purely radially dependent, the potential outside is of the form GM/r:*

$$\rho = \rho(r) : V(r > R) = \frac{GM}{r}\,.$$

Similarly, for the attraction we obtain:

$$\text{outer part:} \qquad a(r > R_2) = -\frac{4}{3}\pi G\rho(R_2^3 - R_1^3)\frac{1}{r^2}\,, \tag{2.21a}$$

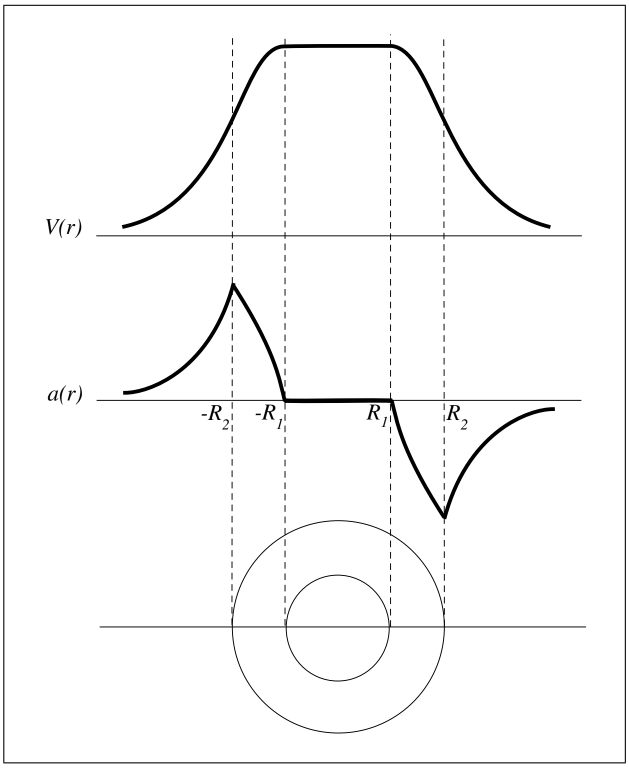


Figure 2.6: Potential V and attraction a as a function of r, due to a homogeneous spherical shell with inner radius R_1 and outer radius R_2.

$$\text{in shell:} \qquad a(R_1 < r < R_2) = -\frac{4}{3}\pi G\rho(r^3 - R_1^3)\frac{1}{r^2}\,, \tag{2.21b}$$

$$\text{inner part:} \qquad a(r < R_1) = 0\,. \tag{2.21c}$$

Since the potential is constant within the shell, the gravitational attraction vanishes there. The resulting potential and attraction are visualized in fig. 2.6.

Exercise 2.4 *Check the continuity of V and a at the boundaries $r = R_1$ and $r = R_2$.*

Exercise 2.5 *The basic structure of the Earth is radial: inner core, outer core, mantle, crust. Assume the following simplified structure:*

$$\begin{aligned} \textit{core:}\ R_\mathrm{c} &= 3500\,\mathrm{km}\ ,\ \rho_\mathrm{c} = 10\,500\,\mathrm{kg\,m^{-3}} \\ \textit{mantle:}\ R_\mathrm{m} &= 6400\,\mathrm{km}\ ,\ \rho_\mathrm{m} = 4500\,\mathrm{kg\,m^{-3}}\,. \end{aligned}$$

Write down the formulae to evaluate potential and attraction. Calculate these along a radial profile and plot them.

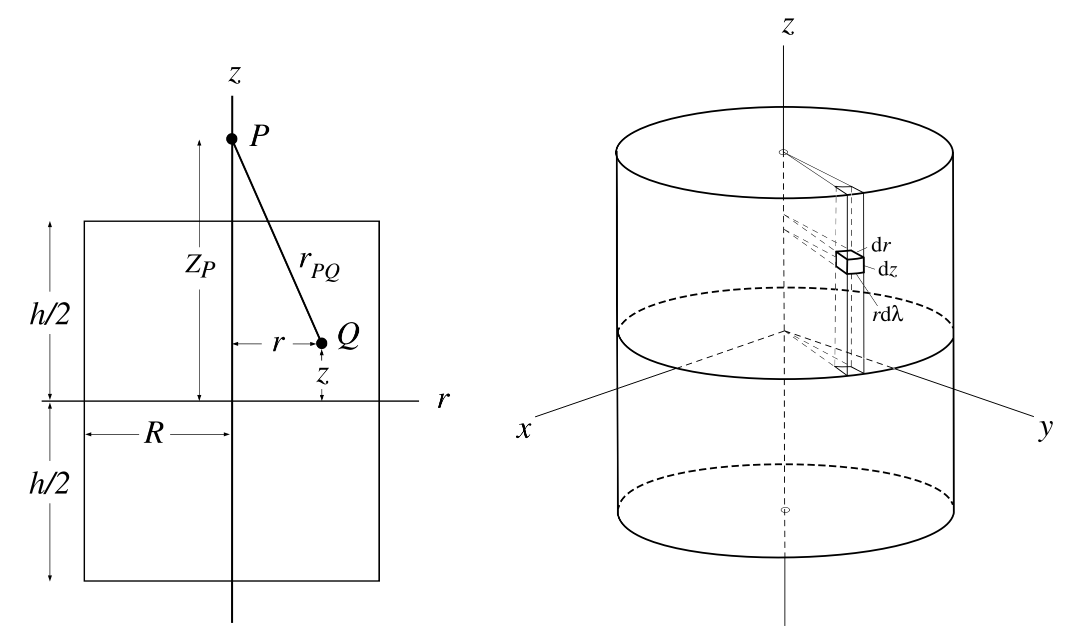


Figure 2.7.: Cylinder (left) of radius R and height h. The origin of the coordinate system is located in the center of the cylinder. The evaluation point P is located on the z-axis (symmetry axis). The volume element of the cylinder (right) has sides $\mathrm{d}r$ in radial direction, $\mathrm{d}z$ in vertical direction and $r\,\mathrm{d}\lambda$ in longitude direction.

2.2.3. Solid homogeneous cylinder

The gravitational attraction of a cylinder is useful for gravity reductions (Bouguer corrections), isostasy modelling and terrain modelling. Assume a configuration with the origin in the center of the cylinder and the z-axis coinciding with the symmetry axis. The cylinder has radius R and height h. Again, assume the evaluation point P on the positive z-axis. As before in 2.2.1 we switch from Cartesian to suitable coordinates. In this case that would be cylinder coordinates (r, λ, z):

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\lambda \\ r\sin\lambda \\ z \end{pmatrix} . \tag{2.22}$$

For the vector from evaluation point P to mass point Q we can write down:

$$\boldsymbol{r}_{PQ} = \boldsymbol{r}_Q - \boldsymbol{r}_P = \begin{pmatrix} r\cos\lambda \\ r\sin\lambda \\ z \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ z_P \end{pmatrix} = \begin{pmatrix} r\cos\lambda \\ r\sin\lambda \\ z - z_P \end{pmatrix} \quad , \quad r_{PQ} = \sqrt{r^2 + (z_P - z)^2}\,. \tag{2.23}$$

The volume element $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$ becomes $r\,\mathrm{d}\lambda\,\mathrm{d}r\,\mathrm{d}z$ and the the integration bounds are $\int\limits_{z=-h/2}^{h/2}\int\limits_{r=0}^{R}\int\limits_{\lambda=0}^{2\pi}$. The integration process for the potential of the cylinder turns out to be somewhat cumbersome. Therefore we integrate the attraction (2.9) directly:

$$\begin{aligned} \boldsymbol{a}_P &= G\rho \int\limits_{z=-h/2}^{h/2}\int\limits_{r=0}^{R}\int\limits_{\lambda=0}^{2\pi} \frac{1}{r_{PQ}^3} \begin{pmatrix} r\cos\lambda \\ r\sin\lambda \\ z - z_P \end{pmatrix} r\,\mathrm{d}\lambda\,\mathrm{d}r\,\mathrm{d}z \\ &= 2\pi G\rho \int\limits_{z=-h/2}^{h/2} (z - z_P) \int\limits_{r=0}^{R} \frac{1}{r_{PQ}^3} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} r\,\mathrm{d}r\,\mathrm{d}z\,. \end{aligned}$$

On the symmetry axis the attraction will have a vertical component only. So we can continue with a scalar a_P now. Again a change of variables brings us further. Calling r_{PQ} (2.23) l again gives:

$$\frac{\mathrm{d}l}{\mathrm{d}r} = \frac{\mathrm{d}\sqrt{r^2 + (z_P - z)^2}}{\mathrm{d}r} = \frac{r}{\sqrt{r^2 + (z_P - z)^2}} = \frac{r}{l} \quad =: \quad \frac{r}{l}\,\mathrm{d}r = \mathrm{d}l\,. \tag{2.24}$$

Thus the integral becomes:

$$a_P = 2\pi G\rho \int\limits_{z=-h/2}^{h/2} (z - z_P) \int\limits_{l=l_-}^{l_+} \frac{1}{l^2}\,\mathrm{d}l\,\mathrm{d}z = -2\pi G\rho \int\limits_{z=-h/2}^{h/2} (z - z_P) \left[\frac{1}{l}\right]_{l_-}^{l_+} \mathrm{d}z\,. \tag{2.25}$$

Indeed, the integrand is much easier now at the cost of more difficult integration bounds of l, which must be determined now. Analogous to 2.2.1 we could distinguish between P outside (above) and P inside the cylinder. It will be shown later, though, that the latter case can be derived from the former. So with $z_P > h/2$ we get the following bounds:

$$\begin{aligned} r = 0 &: l_- = z_P - z \\ r = R &: l_+ = \sqrt{R^2 + (z_P - z)^2}\,. \end{aligned}$$

With these bounds we arrive at:

$$\begin{aligned} a_P &= -2\pi G\rho \int\limits_{z=-h/2}^{h/2} \left(\frac{z - z_P}{\sqrt{R^2 + (z_P - z)^2}} + 1 \right) \mathrm{d}z \\ &= -2\pi G\rho \left[\sqrt{R^2 + (z_P - z)^2} + z \right]_{-h/2}^{h/2} \\ &= -2\pi G\rho \left(h + \sqrt{R^2 + (z_P - h/2)^2} - \sqrt{R^2 + (z_P + h/2)^2} \right) . \end{aligned}$$

Now that the integration over z has been performed we can use the variable z again to replace z_P and get:

$$a(z > h/2) = -2\pi G\rho \left(h + \sqrt{R^2 + (z - h/2)^2} - \sqrt{R^2 + (z + h/2)^2} \right) . \tag{2.26}$$

Recall that this formula holds outside the cylinder along the positive z-axis (symmetry axis).

Negative z-axis The corresponding attraction along the negative z-axis ($z < -h/2$) can be found by adjusting the integration bounds of l. Alternatively, we can replace z by $-z$ *and* change the overall sign.

$$\begin{aligned} a(z < -h/2) &= +2\pi G\rho \left(h + \sqrt{R^2 + (-z - h/2)^2} - \sqrt{R^2 + (-z + h/2)^2} \right) \\ &= -2\pi G\rho \left(-h + \sqrt{R^2 + (z - h/2)^2} - \sqrt{R^2 + (z + h/2)^2} \right) . \end{aligned} \tag{2.27}$$

P within cylinder First, we need to know the attraction at the top and at the base of the cylinder. Inserting $z = h/2$ in (2.26) and $z = -h/2$ in (2.27) we obtain

$$a(h/2) = -2\pi G\rho \left(h + R - \sqrt{R^2 + h^2} \right) , \tag{2.28a}$$

$$a(-h/2) = -2\pi G\rho \left(-h + \sqrt{R^2 + h^2} - R \right) . \tag{2.28b}$$

Notice that $a(h/2) = -a(-h/2)$ indeed.

In order to calculate the attraction inside the cylinder, we separate the cylinder into two cylinders exactly at the evaluation point. So the evaluation point is at the base of a cylinder of height $(h/2 - z)$ and at the top of cylinder of height $(h/2 + z)$. Replacing the heights h in (2.28) by these new heights gives:

$$\begin{aligned}
\text{base of upper cylinder} \; &: \; -2\pi G\rho \left(-({}^h\!/\!{}_2 - z) + \sqrt{R^2 + ({}^h\!/\!{}_2 - z)^2} - R \right) , \\
\text{top of lower cylinder} \; &: \; -2\pi G\rho \left(({}^h\!/\!{}_2 + z) + R - \sqrt{R^2 + ({}^h\!/\!{}_2 + z)^2} \right) , \\
=: a(-{}^h\!/\!{}_2 < z < {}^h\!/\!{}_2) \; &= \; -2\pi G\rho \left(2z + \sqrt{R^2 + (z - {}^h\!/\!{}_2)^2} - \sqrt{R^2 + (z + {}^h\!/\!{}_2)^2} \right) .
\end{aligned}$$

Summary The attraction of a cylinder of height h and radius R along its symmetry axis reads:

$$a(z) = -2\pi G\rho \left(\left\{ \begin{array}{r} h \\ 2z \\ -h \end{array} \right\} + \sqrt{R^2 + (z - {}^h\!/\!{}_2)^2} - \sqrt{R^2 + (z + {}^h\!/\!{}_2)^2} \right) \begin{array}{l} {\scriptstyle z > h/2} \\ {\scriptstyle -h/2 < z < h/2} \\ {\scriptstyle z < -h/2} \end{array} \tag{2.29}$$

This result is visualized in fig. 2.8.

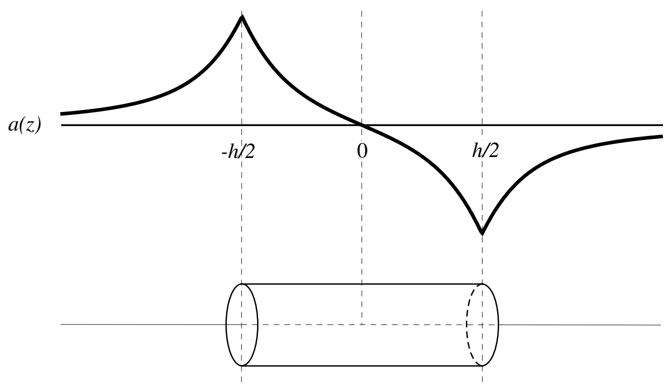


Figure 2.8.: Attraction a as a function of z, due to a solid homogeneous cylinder of radius R and height h. Note that the horizontal axis is the z-axis in this visualization.

Exercise 2.6 *Find out the formulae for a cylindrical shell, i.e. a hollow cylinder with inner radius R_1 and outer radius R_2.*

Infinite plate of thickness *h*. Using the above formulae one can easily derive the attraction of an infinite plate. If we let the radius R go to infinity, we will get:

$$\lim_{R\to\infty} a(z) = -2\pi G\rho \begin{cases} h & \text{, above} \\ 2z & \text{, within} \\ -h & \text{, below} \end{cases} . \tag{2.30}$$

This formula is remarkable. First, by taking the limits, the square root terms have vanished. Second, above and below the plate the attraction does not depend on z anymore. It is constant there. The above infinite plate formula is often used in gravity reductions, as will be seen in 4.

Exercise 2.7 *Calgary lies approximately 1000 m above sealevel. Calculate the attraction of the layer between the surface and sealevel. Think of the layer as a homogeneous plate of infinite radius with density* $\rho = 2670\,\mathrm{kg/m^3}$.

Exercise 2.8 *Simulate a volcano by a cone with top angle* $90°$*, i.e. its height equals the radius at the base. Derive the corresponding formulae for the attraction by choosing the proper coordinate system (hint:* $z = 0$ *at base) and integration bounds. Do this in particular for Mount Fuji* ($H = 3776\,\mathrm{m}$) *with* $\rho = 3300\,\mathrm{kg\,m^{-3}}$.

2.3. Tides

Sorry, it's ebb.

2.4. Summary

point mass in origin

$$V(r) = Gm\frac{1}{r} \quad , \quad a(r) = -Gm\frac{1}{r^2}$$

solid sphere of radius R and constant density ρ, centered in origin

$$V(r) = \begin{cases} \frac{4}{3}\pi G\rho R^3\frac{1}{r} \\ 2\pi G\rho(R^2 - \frac{1}{3}r^2) \end{cases} \quad , \quad a(r) = \begin{cases} -\frac{4}{3}\pi G\rho R^3\frac{1}{r^2} & \text{, outside} \\ -\frac{4}{3}\pi G\rho r & \text{, inside} \end{cases}$$

spherical shell with inner and outer radii R_1 and R_2, resp., and constant density ρ

$$V(r) = \begin{cases} \frac{4}{3}\pi G\rho(R_2^3 - R_1^3)\frac{1}{r} \\ 2\pi G\rho(R_2^2 - \frac{1}{3}r^2) - \frac{4}{3}\pi G\rho R_1^3\frac{1}{r} \\ 2\pi G\rho(R_2^2 - R_1^2) \end{cases} \quad , \quad a(r) = \begin{cases} -\frac{4}{3}\pi G\rho(R_2^3 - R_1^3)\frac{1}{r^2} & \text{, outside} \\ -\frac{4}{3}\pi G\rho(r^3 - R_1^3)\frac{1}{r^2} & \text{, in shell} \\ 0 & \text{, inside} \end{cases}$$

cylinder with height h and radius R, centered at origin, constant density ρ

$$a(z) = -2\pi G\rho\left(\left\{\begin{matrix} h \\ 2z \\ -h \end{matrix}\right\} + \sqrt{R^2 + (z - h/2)^2} - \sqrt{R^2 + (z + h/2)^2}\right) \begin{matrix} \text{, above} \\ \text{, within} \\ \text{, below} \end{matrix}$$

infinite plate of thickness h and constant density ρ

$$a(z) = -2\pi G\rho\begin{cases} h & \text{, above} \\ 2z & \text{, within} \\ -h & \text{, below} \end{cases} .$$

3. Rotation

kinematics Gravity related measurements take generally place on non-static platforms: sea-gravimetry, airborne gravimetry, satellite gravity gradiometry, inertial navigation. Even measurements on a fixed point on Earth belong to this category because of the Earth's rotation. Accelerated motion of the reference frame induces inertial accelerations, which must be taken into account in physical geodesy. The rotation of the Earth causes a centrifugal acceleration which is combined with the gravitational attraction into a new quantity: *gravity*. Other inertial accelerations are usually accounted for by correcting the gravity related measurements, e.g. the Eötvös correction. For these and other purposes we will start this chapter by investigating velocity and acceleration in a rotating frame.

dynamics One of geodesy's core areas is determining the orientation of Earth in space. This goes to the heart of the transformation between inertial and Earth-fixed reference systems. The solar and lunar gravitational fields exert a torque on the flattened Earth, resulting in changes of the polar axis. We need to elaborate on the dynamics of solid body rotation to understand how the polar axis behaves in inertial and in Earth-fixed space.

geometry Newton's laws of motion are valid in inertial space. If we have to deal with satellite techniques, for instance, the satellite's ephemeris is most probably given in inertial coordinates. Star coordinates are by default given in inertial coordinates: right ascension α and declination δ. Moreover, the law of gravitation is defined in inertial space. Therefore, after understanding the kinematics and dynamics of rotation, we will discuss the definition of inertial reference systems and their realizations. An overview will be presented relating the conventional inertial reference system to the conventional terrestrial one.

3.1. Kinematics: acceleration in a rotating frame

Let us consider the situation of motion in a *rotating* reference frame and let us associate this rotating frame with the Earth-fixed frame. The following discussion on velocities and accelerations would be valid for any rotating frame, though.

Inertial coordinates, velocities and accelerations will be denoted with the index i. Earth-fixed quantities get the index e. Now suppose that a *time-dependent* rotation matrix $R = R(\alpha(t))$, applied to the inertial vector $\boldsymbol{r}_i$, results in the Earth-fixed vector $\boldsymbol{r}_e$. We would be interested in velocities and accelerations in the rotating frame. The time derivations must be performed in the inertial frame, though.

From $R\boldsymbol{r}_i = \boldsymbol{r}_e$ we get:

$$\begin{aligned}
\boldsymbol{r}_i &= R^\mathsf{T}\boldsymbol{r}_e && \text{(3.1a)}\\
&\Downarrow \text{ time derivative} \\
\dot{\boldsymbol{r}}_i &= R^\mathsf{T}\dot{\boldsymbol{r}}_e + \dot{R}^\mathsf{T}\boldsymbol{r}_e && \text{(3.1b)}\\
&\Downarrow \text{ multiply by } R \\
R\dot{\boldsymbol{r}}_i &= \dot{\boldsymbol{r}}_e + R\dot{R}^\mathsf{T}\boldsymbol{r}_e \\
&= \dot{\boldsymbol{r}}_e + \Omega\boldsymbol{r}_e && \text{(3.1c)}
\end{aligned}$$

The matrix $\Omega = R\dot{R}^\mathsf{T}$ is called *Cartan*[1] *matrix*. It describes the rotation rate, as can be seen from the following simple 2D example with $\alpha(t) = \omega t$:

$$\begin{aligned}
R &= \begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix} \\
: \Omega &= \begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix} \omega \begin{pmatrix} -\sin\omega t & -\cos\omega t \\ \cos\omega t & -\sin\omega t \end{pmatrix} = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix}
\end{aligned}$$

It is useful to introduce Ω. In the next time differentiation step we can now distinguish between time dependent rotation matrices and time variable rotation rate. Let's pick up the previous derivation again:

$$\begin{aligned}
&\Downarrow \text{ multiply by } R^\mathsf{T} \\
\dot{\boldsymbol{r}}_i &= R^\mathsf{T}\dot{\boldsymbol{r}}_e + R^\mathsf{T}\Omega\boldsymbol{r}_e && \text{(3.1d)}\\
&\Downarrow \text{ time derivative} \\
\ddot{\boldsymbol{r}}_i &= R^\mathsf{T}\ddot{\boldsymbol{r}}_e + \dot{R}^\mathsf{T}\dot{\boldsymbol{r}}_e + \dot{R}^\mathsf{T}\Omega\boldsymbol{r}_e + R^\mathsf{T}\dot{\Omega}\boldsymbol{r}_e + R^\mathsf{T}\Omega\dot{\boldsymbol{r}}_e
\end{aligned}$$

[1] Élie Joseph Cartan (1869–1951), French mathematician.

$$= R^{\mathsf{T}}\ddot{\boldsymbol{r}}_e + 2\dot{R}^{\mathsf{T}}\dot{\boldsymbol{r}}_e + \dot{R}^{\mathsf{T}}\Omega\boldsymbol{r}_e + R^{\mathsf{T}}\dot{\Omega}\boldsymbol{r}_e \tag{3.1e}$$

$$\Downarrow \text{ multiply by } R$$

$$R\ddot{\boldsymbol{r}}_i = \ddot{\boldsymbol{r}}_e + 2\Omega\dot{\boldsymbol{r}}_e + \Omega\Omega\boldsymbol{r}_e + \dot{\Omega}\boldsymbol{r}_e$$

$$\Downarrow \text{ or the other way around}$$

$$\ddot{\boldsymbol{r}}_e = R\ddot{\boldsymbol{r}}_i - 2\Omega\dot{\boldsymbol{r}}_e - \Omega\Omega\boldsymbol{r}_e - \dot{\Omega}\boldsymbol{r}_e \tag{3.1f}$$

This equation tells us that acceleration in the rotating e-frame equals acceleration in the inertial i-frame—in the proper orientation, though—when 3 more terms are added. The additional terms are called *inertial* accelerations. Analyzing (3.1f) we can distinguish the four terms at the right hand side:

- $R\ddot{\boldsymbol{r}}_i$ is the inertial acceleration vector, expressed in the orientation of the rotating frame.
- $2\Omega\dot{\boldsymbol{r}}_e$ is the so-called *Coriolis*[2] acceleration, which is due to motion in the rotating frame.
- $\Omega\Omega\boldsymbol{r}_e$ is the *centrifugal* acceleration, determined by the position in the rotating frame.
- $\dot{\Omega}\boldsymbol{r}_e$ is sometimes referred to as *Euler*[3] acceleration or inertial acceleration of rotation. It is due to a non-constant rotation rate.

Remark 3.1 *Equation (3.1f) can be generalized to moving frames with time-variable origin. If the linear acceleration of the e-frame's origin is expressed in the i-frame with $\ddot{\boldsymbol{b}}_i$, the only change to be made to (3.1f) is $R\ddot{\boldsymbol{r}}_i \to R(\ddot{\boldsymbol{r}}_i - \ddot{\boldsymbol{b}}_i)$.*

Properties of the Cartan matrix Ω. Cartan matrices are skew-symmetric, i.e. $\Omega^{\mathsf{T}} = -\Omega$. This can be seen in the simple 2D example above already. But it also follows from the orthogonality of rotation matrices:

$$RR^{\mathsf{T}} = I \;=:\; \frac{\mathrm{d}}{\mathrm{d}t}(RR^{\mathsf{T}}) = \underbrace{\dot{R}R^{\mathsf{T}}}_{\Omega^{\mathsf{T}}} + \underbrace{R\dot{R}^{\mathsf{T}}}_{\Omega} = 0 \;=:\; \Omega^{\mathsf{T}} = -\Omega\,. \tag{3.2}$$

A second interesting property is the fact that multiplication of a vector with the Cartan matrix equals the cross product of the vector with a corresponding rotation vector:

$$\Omega\boldsymbol{r} = \boldsymbol{\omega} \times \boldsymbol{r} \tag{3.3}$$

[2] Gaspard Gustave de Coriolis (1792–1843).
[3] Leonhard Euler (1707–1783).

This property becomes clear from writing out the 3 Cartan matrices, corresponding to the three independent rotation matrices:

$$\left.\begin{array}{l} R_1(\omega_1 t) : \Omega_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\omega_1 \\ 0 & \omega_1 & 0 \end{pmatrix} \\ R_2(\omega_2 t) : \Omega_2 = \begin{pmatrix} 0 & 0 & \omega_2 \\ 0 & 0 & 0 \\ -\omega_2 & 0 & 0 \end{pmatrix} \\ R_3(\omega_3 t) : \Omega_3 = \begin{pmatrix} 0 & -\omega_3 & 0 \\ \omega_3 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array}\right\} \overset{\text{general}}{=:} \quad \Omega = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} . \tag{3.4}$$

Indeed, when a general rotation vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^{\mathsf{T}}$ is defined, we see that:

$$\begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

The skew-symmetry (3.2) of Ω is related to the fact $\boldsymbol{\omega} \times \boldsymbol{r} = -\boldsymbol{r} \times \boldsymbol{\omega}$.

Exercise 3.1 *Convince yourself that the above Cartan matrices Ω_i are correct, by doing the derivation yourself. Also verify (3.3) by writing out* LHS *and* RHS*.*

Using property (3.3), the velocity (3.1c) and acceleration (3.1f) may be recast into the perhaps more familiar form:

$$\dot{\boldsymbol{r}}_e = R\dot{\boldsymbol{r}}_i - \boldsymbol{\omega} \times \boldsymbol{r}_e \tag{3.5a}$$

$$\ddot{\boldsymbol{r}}_e = R\ddot{\boldsymbol{r}}_i - 2\boldsymbol{\omega} \times \dot{\boldsymbol{r}}_e - \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{r}_e) - \dot{\boldsymbol{\omega}} \times \boldsymbol{r}_e \tag{3.5b}$$

Inertial acceleration due to Earth rotation

Neglecting precession, nutation and polar motion, the transformation from inertial to Earth-fixed frame is given by:

$$\boldsymbol{r}_e = R_3(\text{GAST})\boldsymbol{r}_i \overset{\text{or}}{\rightarrow} \boldsymbol{r}_e = R_3(\omega t)\boldsymbol{r}_i \,. \tag{3.6}$$

The latter is allowed here, since we are only interested in the acceleration effects, due to the rotation. We are not interested in the rotation of position vectors. With great precision, one can say that the Earth's rotation rate is constant: $\dot{\omega} = 0$ The corresponding Cartan matrix and its time derivative read:

$$\Omega = \begin{pmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \dot{\Omega} = 0\,.$$

The three inertial accelerations, due to the rotation of the Earth, become:

$$\text{Coriolis:} \qquad -2\Omega\dot{\boldsymbol{r}}_e = 2\omega \begin{pmatrix} \dot{y}_e \\ -\dot{x}_e \\ 0 \end{pmatrix} \tag{3.7a}$$

$$\text{centrifugal:} \quad -\Omega\Omega\boldsymbol{r}_e = \omega^2 \begin{pmatrix} x_e \\ y_e \\ 0 \end{pmatrix} \tag{3.7b}$$

$$\text{Euler:} \qquad -\dot{\Omega}\boldsymbol{r}_e = \boldsymbol{0} \tag{3.7c}$$

The Coriolis acceleration is perpendicular to both the velocity vector and the Earth's rotation axis. It will be discussed further in 4.1. The centrifugal acceleration is perpendicular to the rotation axis and is parallel to the equator plane, cf. fig. 4.2.

Exercise 3.2 *Determine the direction and the magnitude of the Coriolis acceleration if you are driving from Calgary to Banff with* 100 km/h.

Exercise 3.3 *How large is the centrifugal acceleration in Calgary? On the equator? At the North Pole? And in which direction?*

3.2. Dynamics: precession, nutation, polar motion

Instead of *linear velocity* (or *momentum*) and *forces* we will have to deal with *angular momentum* and *torques*. Starting with the basic definition of angular momentum of a point mass, we will step by step arrive at the angular momentum of solid bodies and their tensor of inertia. In the following all vectors are assumed to be given in an inertial frame, unless otherwise indicated.

Angular momentum of a point mass The basic definition of angular momentum of a point mass is the cross product of position and velocity: $\boldsymbol{L} = m\boldsymbol{r} \times \boldsymbol{v}$. It is a vector quantity. Due to the definition the direction of the angular momentum is perpendicular to both $\boldsymbol{r}$ and $\boldsymbol{v}$.

In our case, the only motion $\boldsymbol{v}$ that exists is due to the rotation of the point mass. By substituting $\boldsymbol{v} = \boldsymbol{\omega} \times \boldsymbol{r}$ we get:

$$\begin{aligned}
\boldsymbol{L} &= m\boldsymbol{r} \times (\boldsymbol{\omega} \times \boldsymbol{r}) \qquad (3.8a)\\
&= m \begin{pmatrix} x \\ y \\ z \end{pmatrix} \times \left[\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} \right] \\
&= m \begin{pmatrix} \omega_1 y^2 - \omega_2 xy - \omega_3 xz + \omega_1 z^2 \\ \omega_2 z^2 - \omega_3 yz - \omega_1 yx + \omega_2 x^2 \\ \omega_3 x^2 - \omega_1 zx - \omega_2 zy + \omega_3 y^2 \end{pmatrix} \\
&= m \begin{pmatrix} y^2 + z^2 & -xy & -xz \\ -xy & x^2 + z^2 & -yz \\ -xz & -yz & x^2 + y^2 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \\
&= \boldsymbol{M}\boldsymbol{\omega} \,. \qquad (3.8b)
\end{aligned}$$

The matrix $\boldsymbol{M}$ is called the *tensor of inertia*. It has units of [kg m^2]. Since $\boldsymbol{M}$ is not an ordinary matrix, but a tensor, which has certain transformation properties, we will indicate it by boldface math type, just like vectors.

Compare now the angular momentum equation $\boldsymbol{L} = \boldsymbol{M}\boldsymbol{\omega}$ with the linear momentum equation $\boldsymbol{p} = m\boldsymbol{v}$, see also tbl. 3.1. It may be useful to think of m as a *mass scalar* and of $\boldsymbol{M}$ as a *mass matrix*. Since the mass m is simply a scalar, the linear momentum $\boldsymbol{p}$ will always be in the same direction as the velocity vector $\boldsymbol{v}$. The angular momentum $\boldsymbol{L}$, though, will generally be in a different direction than $\boldsymbol{\omega}$, depending on the matrix $\boldsymbol{M}$.

Exercise 3.4 *Consider yourself a point mass and compute your angular momentum, due to the Earth's rotation, in two ways:*

- *straightforward by (3.8a), and*
- *by calculating your tensor of inertia first and then applying (3.8b).*

Is $\boldsymbol{L}$ parallel to $\boldsymbol{\omega}$ in this case?

Angular momentum of systems of point masses The concept of tensor of inertia is easily generalized to systems of point masses. The total tensor of inertia is just the superposition of the individual tensors. The total angular momentum reads:

$$\boldsymbol{L} = \sum_{n=1}^{N} m_n \boldsymbol{r}_n \times \boldsymbol{v}_n = \sum_{n=1}^{N} \boldsymbol{M}_n \boldsymbol{\omega} \,. \qquad (3.9)$$

Angular momentum of a solid body We will now make the transition from a discrete to a continuous mass distribution, similar to the gravitational superposition case in 2.1.4. Symbolically:

$$\lim_{N\to\infty} \sum_{n=1}^{N} m_n \ldots = \iiint\limits_{\Omega} \ldots \,\mathrm{d}m\,.$$

Again, the angular momentum reads $\boldsymbol{L} = \boldsymbol{M}\boldsymbol{\omega}$. For a solid body, the tensor of inertia $\boldsymbol{M}$ is defined as:

$$\begin{aligned}
\boldsymbol{M} &= \iiint\limits_{\Omega} \begin{pmatrix} y^2+z^2 & -xy & -xz \\ -xy & x^2+z^2 & -yz \\ -xz & -yz & x^2+y^2 \end{pmatrix} \mathrm{d}m \\
&= \begin{pmatrix} \iiint (y^2+z^2)\,\mathrm{d}m & -\iiint xy\,\mathrm{d}m & -\iiint xz\,\mathrm{d}m \\ & \iiint (x^2+z^2)\,\mathrm{d}m & -\iiint yz\,\mathrm{d}m \\ \text{symmetric} & & \iiint (x^2+y^2)\,\mathrm{d}m \end{pmatrix}.
\end{aligned}$$

The diagonal elements of this matrix are called *moments of inertia*. The off-diagonal terms are known as *products of inertia*.

Exercise 3.5 *Show that in vector-matrix notation the tensor of inertia $\boldsymbol{M}$ can be written as: $\boldsymbol{M} = \iiint (\boldsymbol{r}^\mathsf{T}\boldsymbol{r}I - \boldsymbol{r}\boldsymbol{r}^\mathsf{T})\,\mathrm{d}m$.*

Torque If no external torques are applied to the rotating body, angular momentum is conserved. A change in angular momentum can only be effected by applying a torque $\boldsymbol{T}$:

$$\frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t} = \boldsymbol{T} = \boldsymbol{r} \times \boldsymbol{F}\,. \tag{3.10}$$

Equation (3.10) is the rotational equivalent of $\dot{\boldsymbol{p}} = \boldsymbol{F}$, see tbl. 3.1. Because of the cross-product, the change in the angular momentum vector is always perpendicular to both r and F. Try to intuitively change the axis orientation of a spinning wheel by applying a force to the axis and the axis will probably go a different way. If no torques are applied ($\boldsymbol{T} = 0$) the angular momentum will be constant, indeed.

Three cases will be distinguished in the following:

- $\boldsymbol{T}$ is constant $\longrightarrow$ precession,
  which is a secular motion of the angular momentum vector in inertial space,
- $\boldsymbol{T}$ is periodic $\longrightarrow$ nutation (or forced nutation),
  which is a periodic motion of $\boldsymbol{L}$ in inertial space,
- $\boldsymbol{T}$ is zero $\longrightarrow$ free nutation, polar motion,
  which is a motion of the rotation axis in Earth-fixed space.

Table 3.1.: Comparison between linear and rotational dynamics

	linear	**rotational** point mass	solid body	
linear momentum	$\boldsymbol{p} = m\boldsymbol{v}$	$\boldsymbol{L} = m\boldsymbol{r} \times \boldsymbol{v}$	$\boldsymbol{L} = \boldsymbol{M}\boldsymbol{\omega}$	angular momentum
force	$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = \boldsymbol{F}$	$\frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t} = \boldsymbol{r} \times \boldsymbol{F}$	$\frac{\mathrm{d}\boldsymbol{L}}{\mathrm{d}t} = \boldsymbol{T}$	torque

Precession The word precession is related to the verb to precede, indicating a steady, secular motion. In general, precession is caused by constant external torques. In the case of the Earth, precession is caused by the constant gravitational torques from Sun and Moon. The Sun's (or Moon's) gravitational pull on the nearest side of the Earth is stronger than the pull on the. At the same time the Earth is flattened. Therefore, if the Sun or Moon is not in the equatorial plane, a torque will be produced by the difference in gravitational pull on the equatorial bulges. Note that the Sun is only twice a year in the equatorial plane, namely during the equinoxes (beginning of Spring and Fall). The Moon goes twice a month through the equator plane.

Thus, the torque is produced because of three simultaneous facts:

- the Earth is not a sphere, but rather an ellipsoid,
- the equator plane is tilted with respect to the ecliptic by $23\overset{\circ}{.}5$ (the obliquity ε) and also tilted with respect to the lunar orbit,
- the Earth is a spinning body.

If any of these conditions were absent, no torque would be generated by solar or lunar gravitation and precession would not take place.

As a result of the constant (or mean) part of the lunar and solar torques, the angular momentum vector will describe a conical motion around the northern ecliptical pole (NEP) with a radius of ε. The northern celestial pole (NCP) slowly moves over an ecliptical latitude circle. It takes the angular momentum vector 25 765 years to complete one revolution around the NEP. That corresponds to $50\overset{\prime\prime}{.}3$ per year.

Nutation The word nutation is derived from the Latin for *to nod*. Nutation is a periodic (nodding) motion of the angular momentum vector in space on top of the secular precession. There are many sources of periodic torques, each with its own frequency:

- The orbital plane of the moon rotates once every 18.6 years under the influence of the Earth's flattening. The corresponding change in geometry causes also a change in the lunar gravitational torque of the same period. This effect is known as Bradley nutation.

- The sun goes through the equatorial plane twice a year, during the equinoxes. At those time the solar torque is zero. Vice versa, during the two solstices, the torque is maximum. Thus there will be a semi-annual nutation.
- The orbit of the Earth around the Sun is elliptical. The gravitational attraction of the Sun, and consequently the gravitational torque, will vary with an annual period.
- The Moon passes the equator twice per lunar revolution, which happens roughly twice per month. This gives a nutation with a fortnightly period.

3.3. Geometry: defining the inertial reference system

3.3.1. Inertial space

The word *true* must be understood in the sense that precession and nutation have not been modelled away. The word *mean* refers to the fact that nutation effects have been taken out. Both systems are still time dependent, since the precession has not been reduced yet. Thus, they are actually not inertial reference systems.

3.3.2. Transformations

Precession The following transformation describes the transition from the mean inertial reference system at epoch T_0 to the mean instantaneous one $i_0 \rightarrow \bar{\imath}$:

$$\boldsymbol{r}_{\bar{\imath}} = P\boldsymbol{r}_{i_0} = R_3(-z)R_2(\theta)R_3(-\zeta_0)\boldsymbol{r}_{i_0}\,. \tag{3.11}$$

Figure 3.1 explains which rotations need to be performed to achieve this transformation. First, a rotation around the north celestial pole at epoch T_0 (NCP_0) shifts the mean equinox at epoch T_0 ($\bar{\Upsilon}_0$) over the mean equator at T_0. This is $R_3(-\zeta_0)$. Next, the NCP_0 is shifted along the cone towards the mean pole at epoch T (NCP_T). This is a rotation $R_2(\theta)$, which also brings the mean equator at epoch T_0 is brought to the mean equator at epoch T. Finally, a last rotation around the new pole, $R_3(-z)$ brings the mean equinox at epoch T ($\bar{\Upsilon}_T$) back to the ecliptic. The required precession angles are given with a precision of $1''$ by:

$$\begin{aligned} \zeta_0 &= 2306\overset{''}{.}2181\,T + 0\overset{''}{.}301\,88\,T^2 \\ \theta &= 2004\overset{''}{.}3109\,T - 0\overset{''}{.}426\,65\,T^2 \\ z &= 2306\overset{''}{.}2181\,T + 1\overset{''}{.}094\,68\,T^2 \end{aligned}$$

The time T is counted in Julian centuries (of 36 525 days) since J2000.0, i.e. January 1, 2000, 12^{h} UT1. It is calculated from calendar date and universal time (UT1) by first converting to the so-called *Julian day* number (JD), which is a continuous count of the number of days. In the following Y, M, D are the calendar year, month and day

$$
\begin{aligned}
&\text{Julian days} && \mathrm{JD} = 367Y - \mathtt{floor}(7(Y + \mathtt{floor}((M+9)/12))/4) \\
& && \qquad + \mathtt{floor}(275M/9) + D + 1721\,014 + \mathrm{UT1}/24 - 0.5 \\
&\text{time since J2000.0 in days} && d = \mathrm{JD} - 2451\,545.0 \\
&\text{same in Julian centuries} && T = \frac{d}{36\,525}
\end{aligned}
$$

Exercise 3.6 *Verify that the equinox moves approximately $50''$ per year indeed by projecting the precession angles ζ_0, θ, z onto the ecliptic. Use $T = 0.01$, i.e. one year.*

Nutation The following transformation describes the transition from the mean instantaneous inertial reference system to the true instantaneous one $\bar{\imath} \to i$:

$$\boldsymbol{r}_i = N\boldsymbol{r}_{\bar{\imath}} = R_1(-\varepsilon - \Delta\varepsilon)R_3(-\Delta\psi)R_1(\varepsilon)\boldsymbol{r}_{\bar{\imath}}\,. \tag{3.12}$$

Again, fig. 3.1 explains the individual rotations. First, the mean equator at epoch T is rotated into the ecliptic around $\bar{\Upsilon}_T$. This rotation, $R_1(\varepsilon)$, brings the mean north pole towards the NEP. Next, a rotation $R_3(-\Delta\psi)$ lets the mean equinox slide over the ecliptic towards the true instantaneous epoch. Finally, the rotation $R_1(-\varepsilon - \Delta\varepsilon)$ brings us back to an equatorial system, to the true instantaneous equator, to be precise. The nutation angles are known as nutation in obliquity ($\Delta\varepsilon$) and nutation in (ecliptical) longitude ($\Delta\psi$). Together with the obliquity ε itself, they are given with a precision of $1''$ by:

$$
\begin{aligned}
\varepsilon &= 84\,381''.448 - 46''.8150\,T \\
\Delta\varepsilon &= 0^\circ.0026\cos(f_1) + 0^\circ.0002\cos(f_2) \\
\Delta\psi &= -0^\circ.0048\sin(f_1) - 0^\circ.0004\sin(f_2) \\
&\text{with} \\
f_1 &= 125^\circ.0 - 0^\circ.052\,95\,d \\
f_2 &= 200^\circ.9 + 1^\circ.971\,29\,d
\end{aligned}
$$

The obliquity ε is given in seconds of arc. Converted into degrees we would have $\varepsilon \approx 23^\circ.5$ indeed. On top of that it changes by some $47''$ per Julian century. The nutation angles are not exact. The above formulae only contain the two main frequencies, as expressed by the time-variable angles f_1 and f_2. The coefficients to the variable d are frequencies in units of degree/day:

$$\begin{array}{lll} f_1: & \text{frequency} = 0.052\,95\,^\circ/\text{day} & : \text{period} = 18.6\,\text{years} \\ f_2: & \text{frequency} = 1.971\,29\,^\circ/\text{day} & : \text{period} = 0.5\,\text{years} \end{array}$$

The angle f_1 describes the precession of the orbital plane of the moon, which rotates once every 18.6 years. The angle f_2 describes a half-yearly motion, caused by the fact that the solar torque is zero in the two equinoxes and maximum during the two solstices. The former has the strongest effect on nutation, when we look at the amplitudes of the sines and cosines.

GAST For the transformation from the instantaneous true inertial system i to the instantaneous Earth-fixed sytem e we only need to bring the true equinox to the Greenwich meridian. The angle between the x-axes of both systems is the Greenwich Actual Siderial Time (GAST). Thus, the following rotation is required for the transformation $i \to e$:

$$\boldsymbol{r}_e = R_3(\text{GAST})\boldsymbol{r}_e\,. \tag{3.13}$$

The angle GAST is calculated from the Greenwich Mean Siderial Time (GMST) by applying a correction for the nutation.

$$\begin{aligned} \text{GMST} &= \text{UT1} + (24\,110.548\,41 + 8640\,184.812\,866\,T + 0.093\,104\,T^2 - 6.2\,10^{-6}\,T^3)/3600 \\ &\quad + 24n \\ \text{GAST} &= \text{GMST} + (\Delta\psi\,\cos(\varepsilon + \Delta\varepsilon))/15 \end{aligned}$$

Universal time UT1 is in decimal hours and n is an arbitrary integer that makes $0 \leq \text{GMST} < 24$.

3.3.3. Conventional inertial reference system

Not only is the International Earth Rotation and Reference Systems Service (IERS) responsible for the definition and maintenance of the conventional *terrestrial* coordinate system ITRS (International Terrestrial Reference System) and its realizations ITRF. The IERS also defines the conventional *inertial* coordinate system, called ICRS (International Celestial Reference System), and maintains the corresponding realizations ICRF.

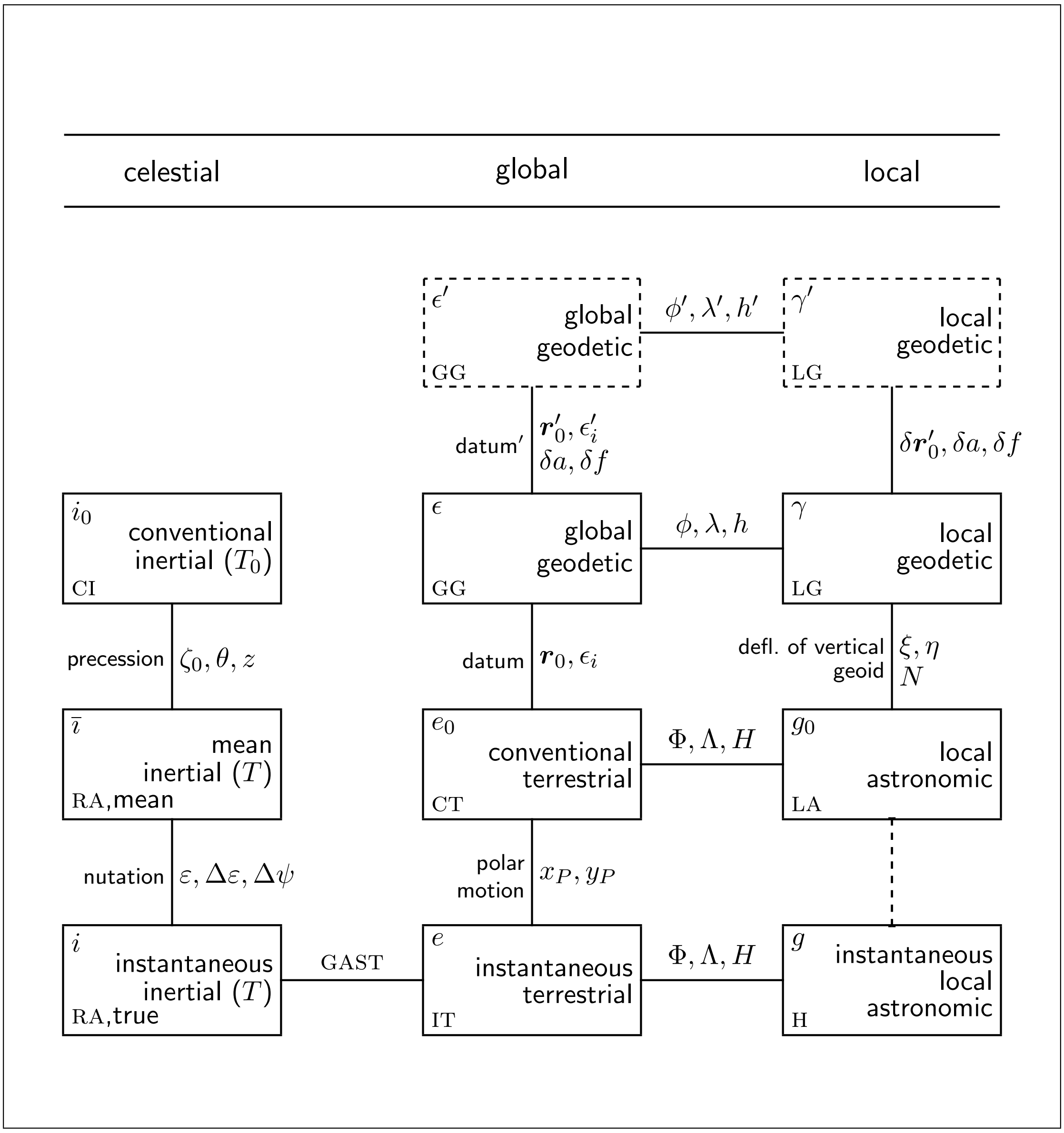
system The ICRS constitutes a set of prescriptions, models and conventions to define at any time a triad of inertial axes.

–	origin:	barycentre of the solar system ($\neq$ Sun's centre of mass),
–	orientation:	mean equator and mean equinox $\bar{\Upsilon}_0$ at epoch J2000.0,
–	time system:	barycentric dynamic time TDB,
–	time evolution:	formulae for P and N.

frame A coordinate system like the ICRS is a set of rules. It is not a collection of points and coordinates yet. It has to materialize first. The International Celestial Reference Frame (ICRF) is realized by the coordinates of over 600 that have been observed by Very Long Baseline Interferometry (VLBI). The position of the quasars, which are extragalactic radio sources, is determined by their right ascension α and declination δ.

Classically, star coordinates have been measured in the optical waveband. This has resulted in a series of fundamental catalogues, e.g. FK5. Due to atmospheric refraction, these coordinates cannot compete with VLBI-derived coordinates. However, in the early nineties, the astrometry satellite HIPPARCOS collected the coordinates of over 100 000 stars with a precision better than 1 milliarcsecond. The HIPPARCOS catalogue constitutes the primary realization of an inertial frame at optical wavelengths. It has been aligned with the ICRF.

3.3.4. Overview



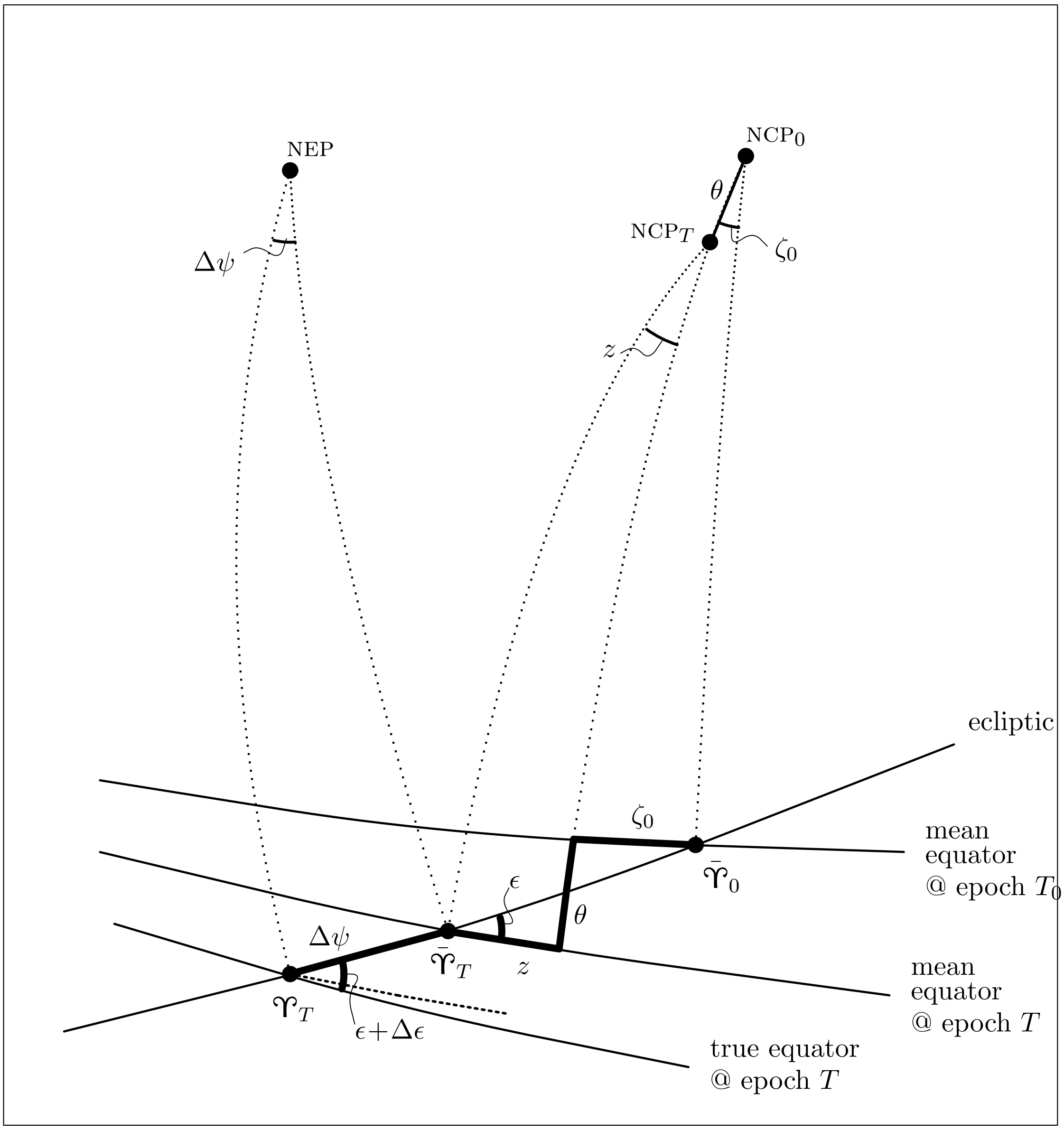


Figure 3.1.: Motion of the true and mean equinox along the ecliptic under the influence of precession and nutation. This graph visualizes the rotation matrices P and N of 3.3.2. Note that the drawing is incorrect or misleading to the extent that *i)* The precession and nutation angles are grossly exaggerated compared to the obliquity ϵ, and *ii)* NCP$_0$ and NCP$_T$ should be on an ecliptical latitude circle $90° - \epsilon$. That means that they should be on a curve *parallel* to the ecliptic, around NEP.

4. Gravity and Gravimetry

4.1. Gravity attraction and potential

Suppose we are doing gravitational measurements at a fixed location on the surface of the Earth. So $\dot{\boldsymbol{r}}_e = 0$ and the Coriolis acceleration in (3.7) vanishes. The only remaining term is the centrifugal acceleration $\boldsymbol{a}_\mathrm{c}$, specified in the e-frame by: $\boldsymbol{a}_\mathrm{c} = \omega^2 (x_e,\ y_e,\ 0)^\mathsf{T}$. Since this acceleration is always present, it is usually added to the gravitational attraction. The sum is called *gravity*:

$$\boxed{\text{gravity} = \text{gravitational attraction} + \text{centrifugal acceleration}}$$

$$\boldsymbol{g} = \boldsymbol{a} + \boldsymbol{a}_\mathrm{c}\,.$$

The gravitational attraction field was seen to be curl-free ($\nabla \times \boldsymbol{a} = \boldsymbol{0}$) in chapter 2. If the curl of the centrifugal acceleration is zero as well, the gravity field would be curl-free, too.

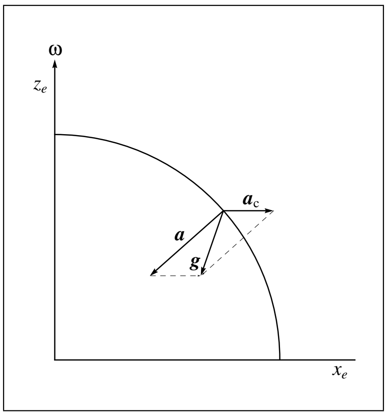


Figure 4.1: Gravity is the sum of gravitational attraction and centrifugal acceleration. Note that $\boldsymbol{a}_c$ is hugely exaggerated. The centrifugal acceleration vector is about 3 orders of magnitude smaller than the gravitational attraction.

Applying the curl operator ($\nabla\times$) to the centrifugal acceleration field $\boldsymbol{a}_\mathrm{c} = \omega^2 (x_e,\ y_e,\ 0)^\mathsf{T}$ obviously yields a zero vector. In other words, the centrifugal acceleration is conservative. Therefore, a corresponding centrifugal potential must exist. Indeed, it is easy to see that this must be V_c, defined as follows:

$$V_\mathrm{c} = \frac{1}{2}\omega^2(x_e^2 + y_e^2) =: \boldsymbol{a}_\mathrm{c} = \nabla V_\mathrm{c} = \omega^2 \begin{pmatrix} x_e \\ y_e \\ 0 \end{pmatrix} . \tag{4.1}$$

Correspondingly, a *gravity potential* is defined

gravity potential = gravitational potential + centrifugal potential

$$W = V + V_\mathrm{c} \,.$$

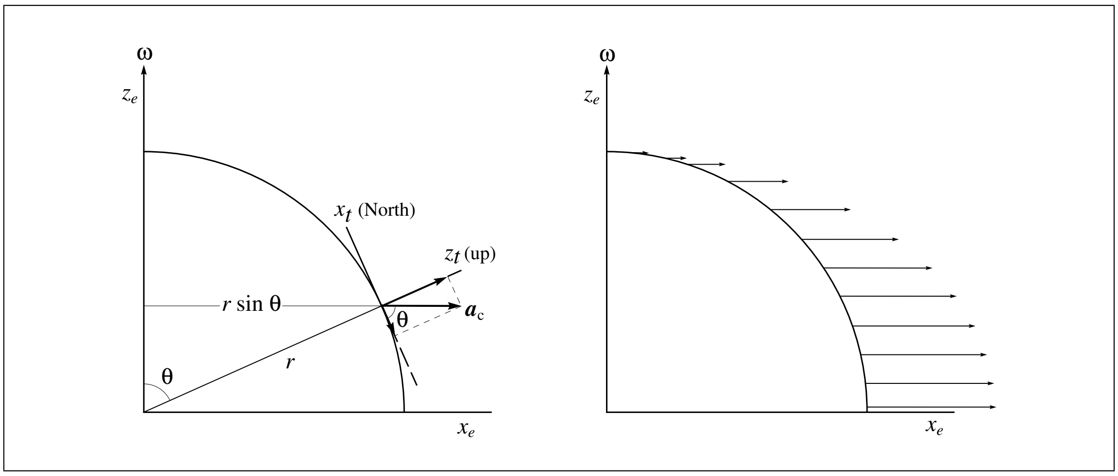


Figure 4.2.: Centrifugal acceleration in Earth-fixed and in topocentric frames.

Centrifugal acceleration in the local frame. Since geodetic observations are usually made in a local frame, it makes sense to express the centrifugal acceleration in the following topocentric frame (t-frame):

- x-axis tangent to the local meridian, pointing North,
- y-axis tangent to spherical latitude circle, pointing East, and
- z-axis complementary in left-handed sense, point up.

Note that this is is a *left-handed* frame. Since it is defined on a sphere, the t-frame can be considered as a spherical approximation of the local astronomic g-frame. Vectors in the Earth-fixed frame are transformed into this frame by the sequence (see fig. 4.3):

$$\boldsymbol{r}_t = P_1 R_2(\theta) R_3(\lambda) \boldsymbol{r}_e = \begin{pmatrix} -\cos\theta\cos\lambda & -\cos\theta\sin\lambda & \sin\theta \\ -\sin\lambda & \cos\lambda & 0 \\ \sin\theta\cos\lambda & \sin\theta\sin\lambda & \cos\theta \end{pmatrix} \boldsymbol{r}_e \,, \tag{4.2}$$

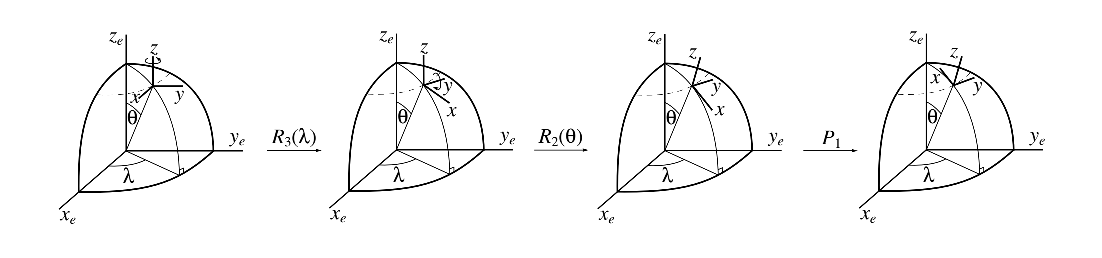


Figure 4.3.: From Earth-fixed global to local topocentric frame.

in which λ is the longitude and θ the co-latitude. The mirroring matrix $P_1 = \mathrm{diag}(-1, 1, 1)$ is required to go from a right-handed into a left-handed frame. Note that we did not include a translation vector to go from geocenter to topocenter. We are only interested in directions here. Applying the transformation now to the centrifugal acceleration vector in the e-frame yields:

$$\boldsymbol{a}_{\mathrm{c},t} = P_1 R_2(\theta) R_3(\lambda) r\omega^2 \begin{pmatrix} \sin\theta\cos\lambda \\ \sin\theta\sin\lambda \\ 0 \end{pmatrix} = r\omega^2 \begin{pmatrix} -\cos\theta\sin\theta \\ 0 \\ \sin^2\theta \end{pmatrix} = r\omega^2 \sin\theta \begin{pmatrix} -\cos\theta \\ 0 \\ \sin\theta \end{pmatrix} . \tag{4.3}$$

The centrifugal acceleration in the local frame shows no East-West component. On the Northern hemisphere the centrifugal acceleration has a South pointing component. For gravity purposes, the vertical component $r\omega^2 \sin^2\theta$ is the most important. It is always pointing up (thus reducing the gravitational attraction). It reaches its maximum at the equator and is zero at the poles.

Remark 4.1 *This same result could have been obtained by writing the centrifugal potential in spherical coordinates: $V_\mathrm{c} = \frac{1}{2}\omega^2 r^2 \sin^2\theta$ and applying the gradient operator in spherical coordinates, corresponding to the local North-East-Up frame:*

$$\nabla_t V_\mathrm{c} = \left(-\frac{1}{r}\frac{\partial}{\partial\theta},\ \frac{1}{r\sin\theta}\frac{\partial}{\partial\lambda},\ \frac{\partial}{\partial r} \right)^\mathsf{T} \left(\frac{1}{2}\omega^2 r^2 \sin^2\theta \right) .$$

Exercise 4.1 *Calculate the centrifugal potential and the zenith angle of the centrifugal acceleration in Calgary ($\theta = 39°$). What is the centrifugal effect on a gravity measurement?*

Exercise 4.2 *Space agencies prefer launch sites close to the equator. Calculate the weight reduction of a 10-ton rocket at the equator relative to a Calgary launch site.*

The Eötvös correction

As soon as gravity measurements are performed on a moving platform the Coriolis acceleration plays a role. In the e-frame it is given by (3.7a) as $\boldsymbol{a}_{\text{Cor.},e} = 2\omega(\dot{y}_e,\ -\dot{x}_e,\ 0)^{\mathsf{T}}$. Again, in order to investigate its effect on local measurements, it makes sense to transform and evaluate the acceleration in a local frame. Let us first write the velocity in spherical coordinates:

$$\boldsymbol{r}_e = r \begin{pmatrix} \sin\theta\cos\lambda \\ \sin\theta\sin\lambda \\ \cos\theta \end{pmatrix} \tag{4.4a}$$

$$\begin{aligned}\dot{\boldsymbol{r}}_e &= r \begin{pmatrix} \cos\theta\cos\lambda \\ \cos\theta\sin\lambda \\ -\sin\theta \end{pmatrix} \dot{\theta} + r \begin{pmatrix} -\sin\theta\sin\lambda \\ \sin\theta\cos\lambda \\ 0 \end{pmatrix} \dot{\lambda} \\ &= \begin{pmatrix} -\cos\theta\cos\lambda v_{\mathrm{N}} - \sin\lambda v_{\mathrm{E}} \\ -\cos\theta\sin\lambda v_{\mathrm{N}} + \cos\lambda v_{\mathrm{E}} \\ \sin\theta v_{\mathrm{N}} \end{pmatrix}\end{aligned} \tag{4.4b}$$

It is assumed here that there is no radial velocity, i.e. $\dot{r} = 0$. The quantities v_{N} and v_{E} are the velocities in North and East directions, respectively, given by:

$$v_{\mathrm{N}} = -r\dot{\theta} \quad \text{and} \quad v_{\mathrm{E}} = r\sin\theta\dot{\lambda}\,.$$

Now the Coriolis acceleration becomes:

$$\boldsymbol{a}_{\text{Cor.},e} = 2\omega \begin{pmatrix} -\cos\theta\sin\lambda v_{\mathrm{N}} + \cos\lambda v_{\mathrm{E}} \\ \cos\theta\cos\lambda v_{\mathrm{N}} + \sin\lambda v_{\mathrm{E}} \\ 0 \end{pmatrix}.$$

Although we use North and East velocities, this acceleration is still in the Earth-fixed frame. Similar to the previous transformation of the centrifugal acceleration, the Coriolis acceleration is transformed into the local frame according to (4.2):

$$\boldsymbol{a}_{\text{Cor.},t} = P_1 R_2(\theta) R_3(\lambda) \boldsymbol{a}_{\text{Cor.},e} = 2\omega \begin{pmatrix} -\cos\theta v_{\mathrm{E}} \\ \cos\theta v_{\mathrm{N}} \\ \sin\theta v_{\mathrm{E}} \end{pmatrix}. \tag{4.5}$$

The most important result of this derivation is, that horizontal motion—to be precise the velocity component in East-West direction—causes a vertical acceleration. This effect can be interpreted as a secondary centrifugal effect. Moving in East-direction the actual rotation would be faster than the Earth's: $\omega' = \omega + \mathrm{d}\omega$ with $\mathrm{d}\omega = v_{\mathrm{E}}/(r\sin\theta)$. This would give the following modification in the vertical centrifugal acceleration:

$$
\begin{aligned}
a'_{\mathrm{c}} &= \omega'^2 r \sin^2\theta \;=\; (\omega + \,\mathrm{d}\omega)^2 r \sin^2\theta \\
&\approx \omega^2 r \sin^2\theta + 2\omega\,\mathrm{d}\omega r \sin^2\theta \;=\; a_{\mathrm{c}} + 2\omega \frac{v_{\mathrm{E}}}{r\sin\theta} r \sin^2\theta \\
&= a_{\mathrm{c}} + 2\omega v_{\mathrm{E}} \sin\theta\,.
\end{aligned}
$$

The additional term $2\omega v_{\mathrm{E}} \sin\theta$ is indeed the vertical component of the Coriolis acceleration. This effect must be accounted for when doing gravity measurements on a moving platform. The reduction of the vertical Coriolis effect is called *Eötvös*[1] *correction*.

As an example suppose a ship sails in East-West direction at a speed of 11 knots ($\approx 20\,\mathrm{km/h}$) at co-latitude $\theta = 60°$. The Eötvös correction becomes $-70 \cdot 10^{-5}\,\mathrm{m/s^2} = -70\,\mathrm{mGal}$, which is significant. A North- or Southbound ship is not affected by this effect.

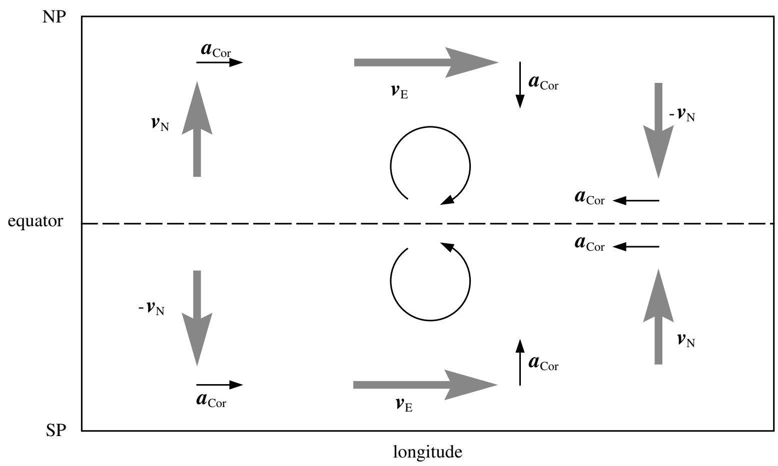


Figure 4.4.: Horizontal components of the Coriolis acceleration.

Remark 4.2 *The horizontal components of the Coriolis acceleration are familiar from weather patterns and ocean circulation. At the northern hemisphere, a velocity in East direction produces an acceleration in South direction; a North velocity produces an acceleration in East direction, and so on.*

Remark 4.3 *The Coriolis acceleration can be interpreted in terms of angular momentum conservation. Consider a mass of air sitting on the surface of the Earth. Because of Earth*

[1] Lóránd (Roland) Eötvös (1848–1919), Hungarian experimental physicist, widely known for his gravitational research using a torsion balance.

rotation it has a certain angular momentum. When it travels North, the mass would get closer to the spin axis, which would imply a reduced moment of inertia and hence reduced angular momentum. Because of the conservation of angular momentum the air mass needs to accelerate in East direction.

Exercise 4.3 *Suppose you are doing airborne gravimetry. You are flying with a constant 400 km/h in Eastward direction. Calculate the horizontal Coriolis acceleration. How large is the Eötvös correction? How accurately do you need to determine your velocity to have the Eötvös correction precise down to 1 mGal? Do the same calculations for a Northbound flight-path.*

4.2. Gravimetry

Gravimetry is the measurement of gravity. Historically, only the measurement of the length of the gravity vector is meant. However, more recent techniques allow vector gravimetry, i.e. they give the direction of the gravity vector as well. In a wider sense, indirect measurements of gravity, such as the recovery of gravity information from satellite orbit perturbations, are sometimes referred to as gravimetry too.

In this chapter we will deal with the basic principles of measuring gravity. We will develop the mathematics behind these principles and discuss the technological aspects of their implementation. Error analyses will demonstrate the capabilities and limitations of the principles. This chapter is only an introduction to gravimetry. The interested reader is referred to (Torge, 1989).

Gravity has the dimension of an acceleration with the corresponding SI unit $\mathrm{m/s^2}$. However, the unit that is most commonly used in gravimetry, is the Gal ($1\,\mathrm{Gal} = 0.01\,\mathrm{m/s^2}$), which is named after Galileo Galilei because of his pioneering work in dynamics and gravitational research. Since we are usually dealing with small differences in gravity between points and with high accuracies—up to 9 significant digits—the most commonly used unit is rather the mGal. Thus, gravity at the Earth's surface is around 981 000 mGal.

4.2.1. Gravimetric measurement principles: pendulum

Huygens[2] developed the mathematics of using a pendulum for time keeping and for gravity measurements in his book *Horologium Oscillatorium* (1673).

[2] Christiaan Huygens (1629–1695), Dutch mathematician, astronomer and physicist, member of the Académie Française.

Mathematical pendulum. A mathematical pendulum is a fictitious pendulum. It is described by a point mass m on a massless string of length l, that can swing without friction around its suspension point or *pivot*. The motion of the mass is constrained to a circular arc around the equilibrium point. The coordinate along this path is $s = l\phi$ with ϕ being the off-axis deflection angle (at the pivot) of the string. See fig. 4.5 (left) for details.

Gravity g tries to pull down the mass. If the mass is not in equilibrium there will be a tangential component $g \sin\phi$ directed towards the equilibrium. Differentiating $s = l\phi$ twice produces the acceleration. Thus, the equation of motion becomes:

$$\ddot{s} = l\ddot{\phi} = -g\sin\phi \approx -g\phi \quad \Leftrightarrow \quad \ddot{\phi} + \frac{g}{l}\phi \approx 0\,, \tag{4.6}$$

in which the small angle approximation has been used. This is the equation of an harmonic oscillator. Its solution is $\phi(t) = a\cos\omega t + b\sin\omega t$. But more important is the frequency ω itself. It is the basic gravity measurement:

$$\omega = \frac{2\pi}{T} = \sqrt{\frac{g}{l}} \quad =: \quad \boxed{g = l\omega^2 = l\left(\frac{2\pi}{T}\right)^2}\,. \tag{4.7}$$

Thus, measuring the swing period T and the length l allows a determination of g. Note that this would be an *absolute* gravity determination.

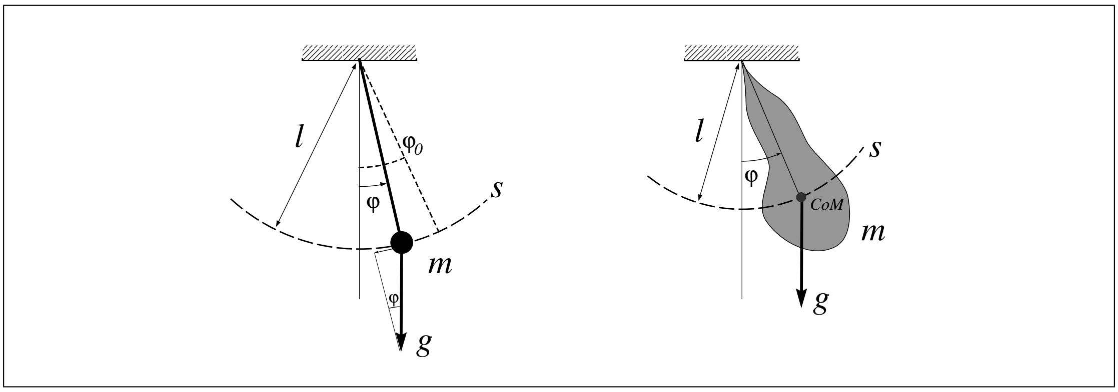


Figure 4.5.: Mathematical (left) and physical pendulum (right)

Remark 4.4 *Both the mass m and the initial deflection ϕ_0 do not show up in this equation. The latter will be present, though, if we do not neglect the non-linearity in (4.6). The non-linear equation $\ddot{\phi} + \frac{g}{l}\sin\phi = 0$ is solved by elliptical integrals. The swing period becomes:*

$$T = 2\pi\sqrt{\frac{l}{g}}\left(1 + \frac{\phi_0^2}{16} + \ldots\right)\,.$$

For an initial deflection of 1° *(1.7 cm for a string of 1 m) the relative effect* $\mathrm{d}T/T$ *is* $2 \cdot 10^{-5}$.

If we differentiate (4.7) we get the following error analysis:

$$\text{absolute:}\quad \mathrm{d}g = -\frac{2}{T}\left(\frac{2\pi}{T}\right)^2 l\,\mathrm{d}T + \left(\frac{2\pi}{T}\right)^2 \mathrm{d}l\,, \tag{4.8a}$$

$$\text{relative:}\quad \frac{\mathrm{d}g}{g} = -2\frac{\mathrm{d}T}{T} + \frac{\mathrm{d}l}{l}\,. \tag{4.8b}$$

As an example let's assume a string of $l = 1\,\mathrm{m}$, which would correspond to a swing period of approximately $T = 2\pi\sqrt{1/g} \approx 2\,\mathrm{s}$. Suppose furthermore that we want to measure gravity with an absolute accuracy of 1mGal, i.e. with a relative accuracy of $\mathrm{d}g/g = 10^{-6}$. Both terms at the right side should be consistent with this. Thus the timing accuracy must be $1\,\mu\mathrm{s}$ and the string length must be known down to $1\,\mu\mathrm{m}$.

The timing accuracy can be relaxed by measuring a number of periods. For instance a measurement over 100 periods would reduce the timing requirement by a factor 10.

Physical pendulum. As written before, a mathematical pendulum is just a fictitious pendulum. In reality a string will not be without mass and the mass will not be a point mass. Instead of accelerations, acting on a point mass, we have to consider the torques, acting on the center of mass of an extended object. So instead of $F = m\ddot{r}$ we get the torque $T = M\ddot{\phi}$ with M the moment of inertia, see also tbl. 3.1. T is the length of the torque vector $\boldsymbol{T}$ here, not to be confused with the swing period T. In general, a torque is a vector quantity (and the moment of inertia would become a tensor of inertia). Here we consider the scalar case $T = |\boldsymbol{T}| = |\boldsymbol{r} \times \boldsymbol{F}| = Fr\sin\alpha$ with α the angle between both vectors.

Drehmoment

Trägheitsmoment

In case of the physical pendulum the torque comes from the gravity force, acting on the center of mass. With ϕ again the angle of deflection, cf. fig. 4.5 (right), we get the equation of motion:

$$M\ddot{\phi} = -mgl\sin\phi \quad =: \quad \ddot{\phi} + \frac{mgl}{M}\phi \approx 0\,, \tag{4.9}$$

in which the small angle approximation was used again. The distance l is from pivot to center of mass. Thus also the physical pendulum is an harmonic oscillator. From the frequency or from the swing duration we get gravity:

$$\omega = \frac{2\pi}{T} = \sqrt{\frac{mgl}{M}} \quad =: \quad \boxed{g = \frac{M}{ml}\omega^2 = \frac{M}{ml}\left(\frac{2\pi}{T}\right)^2}\,. \tag{4.10}$$

The T in (4.10) is a swing period again. Notice that, because of the ratio M/m, the swing period would be independent of the mass again, although it would depend on the mass distribution. The moment of inertia of a point mass is $M = ml^2$. Inserting that in (4.10) would bring us back to the mathematical pendulum.

The problem with a physical pendulum is, that is very difficult to determine the moment of inertia M and the center of mass—and therefore l—accurately. Thus the physical pendulum is basically a *relative gravimeter*. It can be used in two ways:

i) purely as relative gravimeter. The ratio between the gravity at two different locations is:

$$\frac{g_1}{g_2} = \left(\frac{T_2}{T_1}\right)^2 ,$$

which does not contain the physical parameters anymore. By measuring the periods and with one known gravity point, one can determine gravity at all other locations.

ii) as a relative gravimeter turned absolute. This is done by *calibration*. By measuring gravity at two or more locations with known gravity values, preferably with a large difference between them, one can determine the *calibration* or *gravimeter* constant M/ml.

Remark 4.5 *The above example of a pure relative gravimeter reveals already the similarity between gravity networks and levelled height networks. Only differences can be measured. You need to know a point with known gravity (height) value. This is your datum point. The second example, in which the gravimeter factor has to be determined, is equivalent to a further datum parameter: a scale factor.*

Reversionspendel

Actually, a third possibility exists that makes the physical pendulum into a true absolute gravimeter. This is the so-called *reverse pendulum*, whose design goes back to Bohnenberger[3]. The pendulum has two pivots, aligned with the center of mass and one on each side of it. The distances between the pivot points and the center of mass are l_1 and l_2 respectively. The swing period can be tuned by movable masses on the pendulum. They are tuned in such a way that the oscillating motion around both pivots have exactly the same period T.

It can be shown that as a result the unknown moment of inertia is eliminated and that the instrument becomes an absolute gravimeter again. Using the parallel axis theorem we have $M = M_0 + ml^2$ in which M_0 would be the moment of inertia around the center

[3] Johann Gottlieb Friedrich Bohnenberger (1765–1831), German astronomer and geodesist, professor at Tübingen University, father of Swabian geodesy, inventor of Cardanic gyro, probably author of first textbook on higher geodesy.

of mass. According to the above condition of equal T the moments of inertia around the two pivots are $M_0 + ml_1^2$ and $M_0 + ml_2^2$, respectively, giving:

$$\begin{aligned}
&\omega^2 = \frac{mgl_1}{M_0 + ml_1^2} = \frac{mgl_2}{M_0 + ml_2^2} \\
\Leftrightarrow \quad & (M_0 + ml_2^2)l_1 = (M_0 + ml_1^2)l_2 \\
\Leftrightarrow \quad & (ml_1)l_2^2 - (M_0 + ml_1^2)l_2 + M_0 l_1 = 0 \\
\Leftrightarrow \quad & l_2^2 - \left(\tfrac{M_0}{ml_1} + l_1\right) l_2 + \tfrac{M_0}{m} = 0 \\
\Leftrightarrow \quad & (l_2 - l_1)(l_2 - \tfrac{M_0}{ml_1}) = 0 \\
\Leftrightarrow \quad & l_2 = \begin{cases} l_1 & \leftarrow \text{trivial solution} \\ \dfrac{M_0}{ml_1} & : M_0 = ml_1 l_2 \end{cases}
\end{aligned}$$

With this result, the moment of inertia around pivot 2, which is $M_0 + ml_2^2$, can be recast into $ml_1l_2 + ml_2^2 = m(l_1 + l_2)l_2$. Inserting this into (4.10) eliminates the moments of inertia and leads to:

$$g = (l_1 + l_2)\left(\frac{2\pi}{T}\right)^2, \tag{4.11}$$

Thus, if the distance between the pivots is defined and measured accurately, we have an absolute gravimeter again. Gravimeters based on the principle of a physical pendulum were in use until the middle of the 20$^{\text{th}}$ century. Their accuracy was better than 1 mGal.

4.2.2. Gravimetric measurement principles: spring

If we attach a mass to a string, the gravity force results in an elongation of the spring. Thus, measuring the elongation gives gravity.

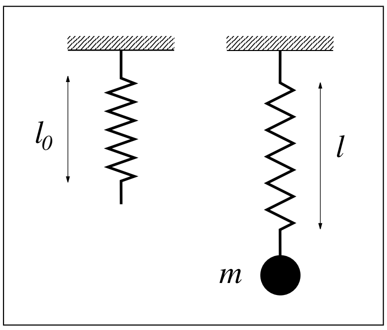


Figure 4.6: Spring without (left) and with mass attached (right).

Static spring. Consider a vertically suspended spring as in fig. 4.6. Suppose that without the mass, the length of the spring is l_0. After attaching the mass, the length becomes elongated to a length of l. According to Hooke's law[4] we have:

$$k(l - l_0) = mg \quad =: \quad \boxed{g = \frac{k}{m}\Delta l}\,, \tag{4.12}$$

with k the spring constant. The spring constant is also a physical parameter that is difficult to determine with sufficient accuracy. Thus a spring gravimeter is fundamentally a relative gravimeter, too. Again we can use it in two ways:

i) purely as a relative gravimeter by measuring ratios:

$$\frac{g_1}{g_2} = \frac{\Delta l_1}{\Delta l_2}\,,$$

ii) or determine the gravimeter factor (k/m in this case) by calibration over known gravity points. In particular since the non-stretched length l_0 might be difficult to determine, this is the usual way of using spring gravimeters:

$$\Delta g_{1,2} = g_2 - g_1 = \frac{k}{m}(\Delta l_2 - \Delta l_1) = \frac{k}{m}(l_2 - l_1)\,.$$

If we write this as $\Delta g = \kappa \Delta l$ then a differentiation gives us again an error analysis:

$$\text{absolute:} \quad \mathrm{d}\Delta g = \Delta l\,\mathrm{d}\kappa + \kappa\,\mathrm{d}\Delta l\,, \tag{4.13a}$$

$$\text{relative:} \quad \frac{\mathrm{d}\Delta g}{g} = \frac{\mathrm{d}\kappa}{\kappa} + \frac{\mathrm{d}\Delta l}{\Delta l}\,. \tag{4.13b}$$

So for an absolute accuracy of 1 mGal the gravimeter constant κ must be calibrated and the elongation must be measured with a relative precision of 10^{-6}.

Remark 4.6 *The dynamic motion of a spring is by itself irrelevant to gravimetry. The equation of motion $m\ddot{l} + kl = 0$, which is also an harmonic oscillator, does not contain gravity. However, the equation does show that the gravimeter constant $\kappa = k/m$ is the square of the frequency of the oscillation. Thus, weak and sensitive springs would show long-period oscillations, whereas stiffer springs would oscillate faster.*

Astatic spring. The above accuracy requirements can never be achieved with a simple static, vertically suspended spring. LaCoste[5] developed the concept of a zero-length

[4]Robert Hooke (1635–1703), physicist, surveyor, architect, cartographer.
[5]Lucien J.B. LaCoste (1908–1995).

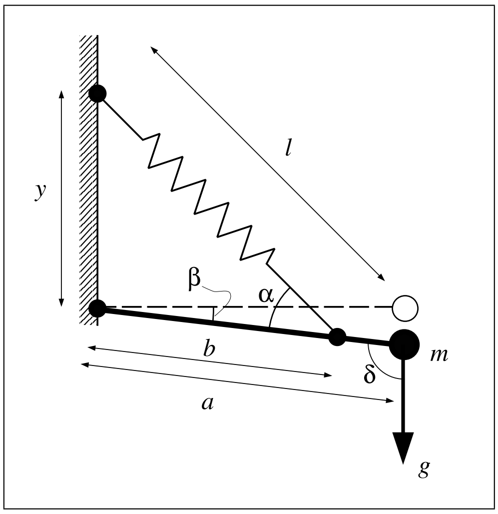


Figure 4.7: Astatized spring: LaCoste-Romberg design

spring as a graduate student at the University of Texas in 1932. His faculty advisor, Dr. Arnold Romberg, had given LaCoste the task to design a seismometer sensitive to low frequencies. After the successful design, involving the zero-length spring, they founded a company, *LaCoste-Romberg*, manufacturer of gravimeters and seismographs. They used the zero-length spring in the design of fig. 4.7, a so-called *astatic* or *astatized* spring.

A bar with a mass m attached can rotate around a pivot. A spring keeps the bar from going down. The angle between bar direction and gravity vector is δ. In equilibrium the torque exerted by gravity equals the spring torque:

$$\begin{aligned} &\text{gravity torque:} & mga\sin\delta &= mga\cos\beta\,, \\ &\text{spring torque:} & k(l-l_0)b\sin\alpha &= k(l-l_0)b\frac{y}{l}\cos\beta\,. \end{aligned}$$

The latter step is due to the sine law $\sin(\alpha)/y = \sin(\beta + \frac{1}{2}\pi)/l$. Thus, equilibrium is attained for:

$$mga\cos\beta = k(l-l_0)b\frac{y}{l}\cos\beta \quad : \quad \boxed{g = \frac{k}{m}\frac{b}{a}\left(1 - \frac{l_0}{l}\right)y}\,. \tag{4.14}$$

A *zero-length spring* has $l_0 = 0$, which means that its length is zero if it is not under tension. In case of fig. 4.6, the left spring would shrink to zero. Zero-length springs are realized by twisting the wire when winding it during the production of the spring. Such springs are used in the above configuration. What is remarkable about this configuration is that equilibrium is independent of β or l. Thus if all parameters are tuned such that we have equilibrium, we would be able to move the arm and still have equilibrium.

The differential form of (4.14) reads:

$$\mathrm{d}g = \underbrace{\frac{k}{m}\frac{b}{a}\frac{l_0}{l}\frac{y}{l}}_{\kappa}\,\mathrm{d}l\,.$$

This is Hooke's law again, in differential disguise. Instead of the ordinary spring constant k/m we now have the multiple fractions expression, denoted as κ. So κ is the effective spring constant. Remember that the spring constant is the square of the oscillation frequency. Thus all parameters of this system can be tuned to produce a required frequency ω. This is very practical for building seismographs. With a zero-length spring one would get an infinite period T.

For gravimetry, the sensitivity is of importance. Inverting the above differential form yields:

$$\mathrm{d}l = \underbrace{\frac{m}{k}\frac{a}{b}\frac{l}{l_0}\frac{l}{y}}_{\kappa^{-1}}\,\mathrm{d}g\,.$$

Again, we can tune all parameters to produce a certain $\mathrm{d}l$ for a given change in gravity $\mathrm{d}g$. The astatic configuration with zero-length spring would be infinitely sensitive. Since this is unwanted we could choose a spring with nearly zero-length. What is done in practice is to tilt the left wall on which the lever arm and the spring are attached about a small angle. This slightly changes the equilibrium condition.

The existence of so many physical parameters in (4.14) implies that an astatic spring gravimeter is a relative gravimeter. Gravimeters of this type can achieve accuracies down to $10\,\mu$Gal. For stationary gravimeters, that are used for tidal observations, even $1\,\mu$Gal can be achieved. The performance of these very precise instruments depend on the material properties of the spring. At these accuracy levels springs are not fully elastic, but exhibit a creep rate. This causes a drift in the measurements. LaCoste-Romberg gravimeters use a metallic spring. Scintrex and Worden gravimeters use quartz springs. Advantages and disadvantages are listed in tbl. 4.1.

Table 4.1.: Pros and cons of metallic and quartz springs

	metallic	quartz
thermal expansion	high, protection required	low and linear
magnetic influence	yes, shielding needed	no
weight	high	low
drift rate	low	high

4.2.3. Gravimetric measurement principles: free fall

According to legend, Galileo Galilei dropped masses from the leaning tower of Pisa. Although he wanted to demonstrate the equivalence principle, this experiment shows already free fall as a gravimetric principle. Galileo probably never performed this experiment. Stevin[6], on the other hand, actually did. Twenty years before Galileo's supposed experiments, Stevin dropped two lead spheres, one of them 10 times as large as the other, from 30 feet height and noticed that they dropped exactly simultaneously. As he expressed, this was a heavy blow against Aristotle's mechanics, which was the prevailing mechanics at that time.

Drop principle (free fall). The equation of motion of a falling proof mass in a gravity field g reads:

$$\ddot{z} = g\,. \tag{4.15}$$

This ordinary differential equation is integrated twice to produce:

$$z(t) = \frac{1}{2} g t^2 + v_0 t + z_0\,, \tag{4.16}$$

with the initial height z_0 and the initial velocity v_0 as integration constants. For the simple case of $z_0 = v_0 = 0$ we simply have:

$$g = \frac{2z}{t^2}\,. \tag{4.17}$$

Thus gravity is determined from measuring the time it takes a proof mass to fall a certain vertical distance z. The free fall principle yields *absolute* gravity. The differential form of (4.17) provides us with an error analysis:

$$\text{absolute:} \quad \mathrm{d}g = \frac{2}{t^2}\,\mathrm{d}z - \frac{2}{t}\frac{2z}{t^2}\,\mathrm{d}t\,, \tag{4.18a}$$

$$\text{relative:} \quad \frac{\mathrm{d}g}{g} = \frac{\mathrm{d}z}{z} - 2\frac{\mathrm{d}t}{t}\,. \tag{4.18b}$$

As an example, assume a drop time of 1 s. The drop length is nearly 5 m. If our aim is an absolute measurement precision of 1 mGal, the relative precision is $\mathrm{d}g/g = 10^{-6}$. The absolute precision of the timing must then be 0.5 μs. The drop length must be known within 5 μm.

These requirements are met in reality by laser interferometry. The free-fall concept is realized by dropping a prism in a vacuum chamber. An incoming laser beam is reflected

[6] Simon Stevin (1548–1620), Dutch mathematician, physicist and engineer, early proponent of Copernican cosmology.

by the prism. Comparing the incoming and outgoing beams gives a interference pattern that changes under the changing height and speed of the prism. A length measurement is performed by counting the fringes of the interference pattern. Time measurement is performed by an atomic clock or hydrogen maser. Commercial free-fall gravimeters perform better than the above example. They achieve accuracies in the 1–10 μGal range, i.e. down to 10^{-9} relative precision. At this accuracy level one must make corrections for many disturbing gravitational effects, e.g.:

- tides: direct tidal attraction, Earth tides, ocean loading;
- pole tide: polar motion causes a time-variable centrifugal acceleration;
- air pressure, which is a measure of the column of atmospheric mass above the gravimeter;
- gravity gradient: over a drop length of 1 m gravity changes already 0.3 mGal, which is orders of magnitude larger than the indicated accuracy;
- changes in groundwater level.

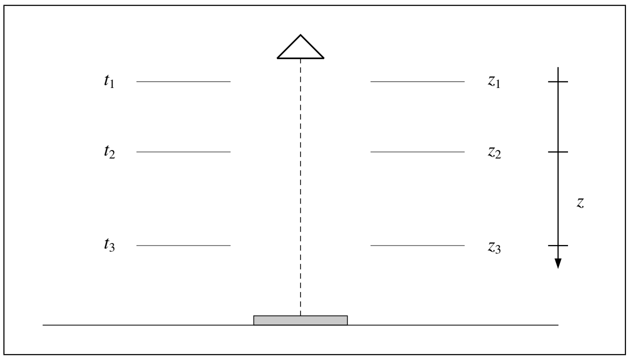


Figure 4.8: Principle of a free fall gravimeter with timing at three fixed heights.

In practice one cannot start timing exactly when $z_0 = v_0 = 0$. Instead we start timing at a given point on the drop trajectory. In this case z_0 and v_0 are unknowns too, so at least 3 measurement pairs (t_i, z_i) must be known. In case we have exactly 3 measurements, as in fig. 4.8, the initial state parameters are eliminated by:

$$g = 2\frac{(z_3 - z_1)(t_2 - t_1) - (z_2 - z_1)(t_3 - t_1)}{(t_3 - t_1)(t_2 - t_1)(t_3 - t_2)} .$$

In reality, more measurements are made during a drop experiment to provide an overdetermined problem. This is also necessary to account for the existing vertical gravity gradient, which is roughly 0.3 mGal/m, but which must be modelled as an unknown, too.

Launch principle (rise and fall). Instead of just dropping a mass one can launch it vertically and measure the time it takes to fall back. The same equation of motion applies. This situation is more symmetrical, which allows to cancel the residual air drag. Suppose we have two measurements levels z_1 and z_2. When it's going up, the mass first crosses level 1 and then level 2. On its way down it first goes through z_2 and then through z_1. Given the known height difference $\Delta z = z_2 - z1$ between the two levels, see fig. 4.9, and measuring the time differences at both levels, we can determine gravity by:

$$g = \frac{8\Delta z}{(\Delta t_1)^2 - (\Delta t_2)^2}\,. \tag{4.19}$$

In practice one would measure at more levels to obtain an overdetermined situation.

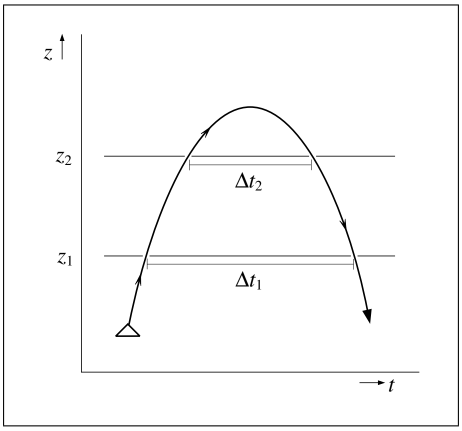


Figure 4.9: Principle of a rise-and-fall gravimeter with timing at two fixed heights.

4.3. Gravity networks

A network of gravity points, obtained from *relative gravimetry*, is similar to height networks. They have one degree of freedom, requiring one datum point with a given gravity value. Moreover, if the gravimeter factor (spring scale factor) is unknown, at least one further datum point with known gravity value has to be given. In a height network this problem would occur if the scale on the levelling rod is unknown.

A distinction between gravity and height networks is the fact that relative gravimeters display a drift behaviour. This doesn't require further datum points. It only requires that at least one of the points in the network is measured twice in order to determine the drift constant—if modelled linear in time. Sticking to the analogy of height networks a drift-type situation might happen if the levelling rod would expand over time due to temperature influences. This analogy is somewhat far-fetched, though.

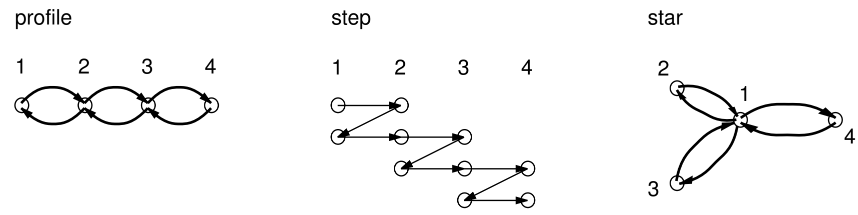


Figure 4.10.: Gravity observation procedures: profile method (1-2-3-4-3-2-1), step method (1-2-1-2-3-2-3-4-3-4) and the star method (1-2-1-3-1-4-1).

4.3.1. Gravity observation procedures

Three basic methods are used in gravimetry, cf. fig. 4.10. They are the

i) the *profile method*. Each point (except the end points) is observed twice. There is a wide variety of time differences between measurements of the same point.

ii) the *step method*. Each point (except the end points) is visited three times. The revisit time differences are short. The latter aspect is advantageous in case the drift is non linear.

iii) the *star method*. The measurements of the central point are used for drift determination. All other points are *loose ends* in a networks sense. Gross errors in them cannot be detected.

The step method is most suitable for precise and reliable networks and for calibration purposes. In reality a mixture of these methods can be used.

4.3.2. Relative gravity observation equation

Let us denote $y_n(t_k)$ the gravity observation on point n at time k and assume that it has been corrected for tides and other effects that are easily modelled. Because we are dealing with relative gravimetry the observation is not equal to gravity at point n (g_n), but has an unknown bias b. Let us furthermore assume that the drift is linear in time: dt_k. One should use this assumption with care. Drift can be non-linear. Also jumps, so-called tares, occur. Thus the observation represents:

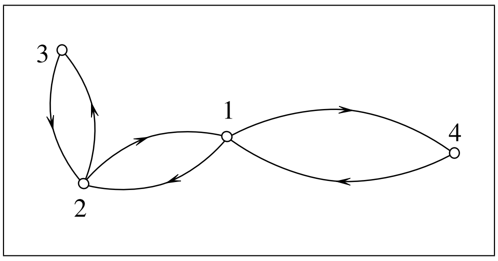
$$y_n(t_k) = g_n + b + dt_k + \epsilon \,,$$

with ϵ the measurement noise. The bias is eliminated by subtracting, for instance, the first measurement at the first point. Thus we get the basic observation equation:

$$\Delta y_n(t_k) = y_n(t_k) - y_1(t_1) + \epsilon$$

$$= g_n - g_1 + d(t_k - t_1) + \epsilon$$
$$= \Delta g_n + d\Delta t_k + \epsilon\,. \tag{4.20}$$

Exercise 4.4 *Given the following small network. Suppose points are measured in the order 1-2-3-2-1-4-1. Write down the linear observation model $y = Ax$.*



Remember that $y_1(t_1)$ does not appear as a separate measurement, since it is subtracted from all other measurements already.

$$\begin{pmatrix} \Delta y_2(t_2) \\ \Delta y_3(t_3) \\ \Delta y_2(t_4) \\ \Delta y_1(t_5) \\ \Delta y_4(t_6) \\ \Delta y_1(t_7) \end{pmatrix} = \begin{pmatrix} 1\ 0\ 0\ \Delta t_2 \\ 0\ 1\ 0\ \Delta t_3 \\ 1\ 0\ 0\ \Delta t_4 \\ 0\ 0\ 0\ \Delta t_5 \\ 0\ 0\ 1\ \Delta t_6 \\ 0\ 0\ 0\ \Delta t_7 \end{pmatrix} \begin{pmatrix} \Delta g_2 \\ \Delta g_3 \\ \Delta g_4 \\ d \end{pmatrix}.$$

The gravimeter constant, as supplied by the manufacturer of the gravimeter, is sufficient for many applications. It was assumed in (4.20) that the observations y already incorporate this constant. For precise measurements or for checking the stability of the gravimeter, a calibration should be performed. This entails the measurement at two or more gravity stations with *known* values. After drift determination, the measured values can be compared to the known gravity values—in a least-squares sense if a network is measured—and the correct gravimeter constant determined.

5. Elements from potential theory

The law of gravitation allows to calculate a body's gravitational potential and attraction if its density distribution and shape are given. In most real life situations, the density distribution is unknown, though. In general it cannot be determined from the exterior potential or attraction field, which was demonstrated by the formulae for point mass, solid sphere and spherical shell already. Moreover, in physical geodesy even the shape of the body—the geoid—must be considered unknown.

The next question is, whether the exterior field can be determined from the function (the potential or its derivatives) on the surface. This is a boundary value problem.

Potential theory is the branch of mathematical physics that deals with potentials and boundary value problems. Its tools are vector calculus, partial differential equations, integral equations and several theorems and identities of Gauss (divergence), Stokes (rotation) and Green. Potential theory describes the behaviour of potentials of any type. Thus it finds applications in such diverse disciplines as electro-magnetics, hydrodynamics and gravitational theory.

Here we will be concerned with gravitational potentials and the corresponding boundary value problems. After some initial vector calculus rules, this chapter treats Gauss's divergence theorem with some applications. Subesequently the classical boundary value problems are discussed.

5.1. Some vector calculus rules

First, let us write down the basic operators, both using the nabla operator (∇) and the operator names. Note that the definitions at the right side are in Cartesian coordinates.

$$\begin{array}{llll}
\text{gradient:} & \nabla f = \operatorname{grad} f & = & \begin{pmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \\ \dfrac{\partial f}{\partial z} \end{pmatrix} \\
\text{divergence:} & \nabla \cdot \boldsymbol{v} = \operatorname{div} \boldsymbol{v} & = & \dfrac{\partial v_x}{\partial x} + \dfrac{\partial v_y}{\partial y} + \dfrac{\partial v_z}{\partial z} \\
\text{curl, rotation:} & \nabla \times \boldsymbol{v} = \operatorname{rot} \boldsymbol{v} & = & \begin{pmatrix} \dfrac{\partial v_z}{\partial y} - \dfrac{\partial v_y}{\partial z} \\ \dfrac{\partial v_x}{\partial z} - \dfrac{\partial v_z}{\partial x} \\ \dfrac{\partial v_y}{\partial x} - \dfrac{\partial v_x}{\partial y} \end{pmatrix} \\
\text{Laplace:} & \Delta f = \operatorname{del} f & = & \\
 & = \nabla \cdot \nabla f = \operatorname{div} \operatorname{grad} f & = & \dfrac{\partial^2 f}{\partial x^2} + \dfrac{\partial^2 f}{\partial y^2} + \dfrac{\partial^2 f}{\partial z^2}
\end{array}$$

Basic properties:

$$\nabla \times \nabla f = \mathbf{0} \tag{5.1a}$$

$$\nabla \cdot (\nabla \times \boldsymbol{v}) = 0 \tag{5.1b}$$

Chain-rule type rules:

$$\nabla(fg) = f\nabla g + g\nabla f \tag{5.1c}$$

$$\nabla \cdot (f\boldsymbol{v}) = \boldsymbol{v} \cdot \nabla f + f\nabla \cdot \boldsymbol{v} \tag{5.1d}$$

$$\nabla \times (f\boldsymbol{v}) = \nabla f \times \boldsymbol{v} + f\nabla \times \boldsymbol{v} \tag{5.1e}$$

$$\nabla \cdot (\boldsymbol{u} \times \boldsymbol{v}) = \boldsymbol{v} \cdot (\nabla \times \boldsymbol{u}) - \boldsymbol{u} \cdot (\nabla \times \boldsymbol{v}) \tag{5.1f}$$

Some rules for functions of $\boldsymbol{r}$ and $r = |\boldsymbol{r}|$:

$$\nabla r^n = n r^{n-2} \boldsymbol{r} \tag{5.1g}$$

$$\nabla \cdot r^n \boldsymbol{r} = (n+3) r^n \tag{5.1h}$$

$$\nabla \times r^n \boldsymbol{r} = \mathbf{0} \tag{5.1i}$$

$$\Delta r^n = \nabla \cdot (n r^{n-2} \boldsymbol{r}) = n(n+1) r^{n-2} \tag{5.1j}$$

Exercise 5.1 *Try to prove some of the above identities. For instance, prove (5.1j) from (5.1d) and (5.1g).*

5.2. Divergence—Gauss

Vector flow. Our treatment of the Gauss[1] divergence theorem begins with the concept of vector flow through a surface. Vector flow is, loosely speaking, the amount of vectors going through a certain surface—one could think of water flowing through a section of a river. The *amount of vectors* is quantified by taking the scalar product of the vector field and the surface normal.

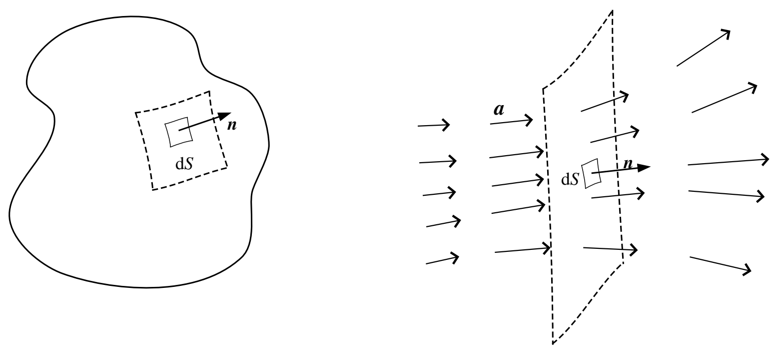


Figure 5.1.: Vectorflow through a surface S and the infinitesimal surface S.

Now consider the vector field $\boldsymbol{a}$ and a closed surface S with normal vector $\boldsymbol{n}$. Take an infinitesimal part of this surface $\mathrm{d}S$. The vector flow through it is $\boldsymbol{a} \cdot \boldsymbol{n} \, \mathrm{d}S = \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S}$. In order to determine the vector flow through the whole surface, we just integrate:

$$\text{total vector flow through } S = \iint_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S} \,. \tag{5.2}$$

More specifically, let us take a spherical surface, with radius r. The normal vector is

[1]Carl Friedrich Gauss (1777–1855), German mathematician, physicist, geodesist.

$\boldsymbol{r}/r$. In that case the vectorial surface element becomes:

$$\mathrm{d}\boldsymbol{S} = \frac{\boldsymbol{r}}{r} r^2 \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\lambda = \boldsymbol{r} r \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\lambda \,.$$

Suppose our vector field $\boldsymbol{a}$ is generated by a point mass m inside the surface, at the centre of the sphere. Then the (infinitesimal) vector flow of (5.2) becomes:

$$\boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S} = -\frac{Gm}{r^3} \boldsymbol{r} \cdot \boldsymbol{r} r \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\lambda = -Gm \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\lambda \,,$$

$$=\text{:total vector flow through sphere: } \iint\limits_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S} = -4\pi G m \,, \tag{5.3}$$

which is a constant, independent of the radius of the sphere. This is easily explained by the fact that the vector field decreases with r^{-2} and the area of S increases with r^2. In fact, this explanation is so general that (5.3) holds for the vector flow through any closed surface which contains the point mass.

In case the point mass lies outside the surface, the vector flow through it becomes zero. Infinitesimal flows $\boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S}$ at one side of the surface are cancelled at the corresponding surface element at the opposite side of the surface.

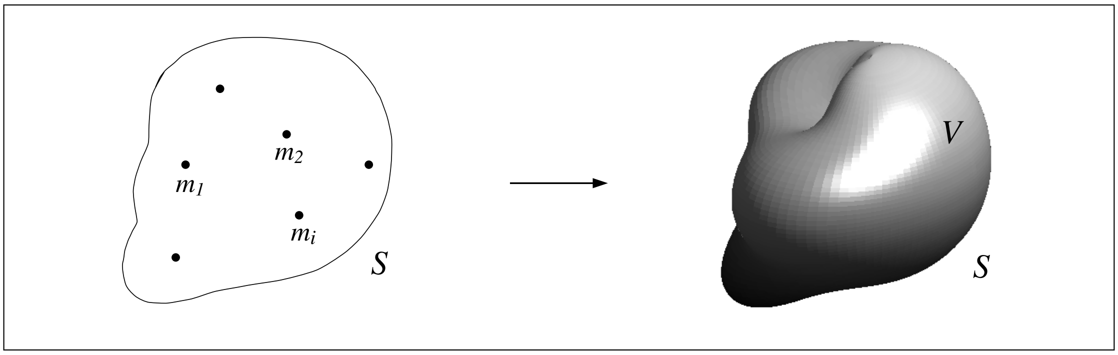


Figure 5.2.: From mass point configuration to solid body V.

From discrete to continuous. Consider a set of point masses m_i and a closed surface S. The vector flow through S is:

$$\iint\limits_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S} = -4\pi G \sum_i m_i \,,$$

in which the summation runs over all point masses inside the surface only. In a next step we make the transition from discrete to continuous. Integrating over all infinitesimal masses $\mathrm{d}m$ within a body V gives us:

$$\iint\limits_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S} = -4\pi G \iiint\limits_V \mathrm{d}m = -4\pi G \iiint\limits_V \rho \, \mathrm{d}V = -4\pi G M \,. \tag{5.4}$$

It was tacitly assumed here that the surface S, through which the flow is measured, is the surface of the body V. This is not a necessity, though. Equation (5.4) gives us a first idea of the boundary value problem. It says that we can determine the total mass of a body, without knowing its density structure, but by considering the vector field on the surface only.

From finite to infinitesimal and back again. Let us go back to (5.4) now in the formulation $\iint_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S} = -4\pi G \iiint_V \rho \, \mathrm{d}V$ and evaluate it per unit volume, i.e. divide by V. If we let the volume go to zero we get the following result:

$$\lim_{V \to 0} \frac{\iint_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S}}{V} = \lim_{V \to 0} \frac{-4\pi G \iiint_V \rho \, \mathrm{d}V}{V} = -4\pi G \rho \,. \tag{5.5}$$

The left side of this equations is exactly the definition of divergence of the vector field. Thus we get the important result:

$$\operatorname{div} \boldsymbol{a} = \begin{cases} -4\pi G\rho \text{ (inside)} & \to \text{Poisson} \\ \quad 0 \qquad \text{(outside)} & \to \text{Laplace} \end{cases} \,. \tag{5.6}$$

The divergence operator indicates the sources and sinks in a vector field. The Poisson[2] equation thus tells us that mass density—a source of gravitation—is a sink in the gravitational attraction field, mathematically speaking. The Laplace[3] equation is just a special case of Poisson by setting $\rho = 0$ outside the masses.

Exercise 5.2 *Determine the divergence of the gravitational attraction field of a point mass $\frac{GM}{r}\boldsymbol{r}$—at least outside the point mass—using the appropriate properties in 5.1.*

Finally, integrating the left side of (5.5), using the definition of divergence gives us Gauss's divergence theorem:

$$\boxed{\iiint\limits_V \operatorname{div} \boldsymbol{a} \, \mathrm{d}V = \iint\limits_S \boldsymbol{a} \cdot \mathrm{d}\boldsymbol{S}} \,. \tag{5.7}$$

[2]Simon Denis Poisson (1781–1840).
[3]Pierre Simon Laplace (1749–1827), French mathematician and astronomer.

Loosely speaking, Gauss's divergence theorem says that the net vector flow generated (or disappearing) in the total body V can be assessed by looking at the vectors coming out of (or going into) the boundary S.

Laplace fields and harmonic functions

A vector field $\boldsymbol{a}$ that has no divergence ($\nabla\cdot\boldsymbol{a} = 0$) and which is conservative ($\nabla\times\boldsymbol{a} = \mathbf{0}$) is called a *Laplace field*. Because it is conservative we will be able to write $\boldsymbol{a}$ as the gradient of a potential. Thus,

$$\left.\begin{array}{ll} \text{zero curl:} & \nabla\times\boldsymbol{a} = \mathbf{0} : \boldsymbol{a} = \nabla\Phi \\ \text{zero divergence:} & \nabla\cdot\boldsymbol{a} = 0 \end{array}\right\} =: \nabla\cdot\nabla\Phi = \Delta\Phi = 0 \tag{5.8}$$

A function Φ that satisfies the Laplace equation $\Delta\Phi = 0$ is called *harmonic*. Therefore, saying that a potential Φ is harmonic is equivalent to saying that its gradient field $\nabla\Phi$ is a Laplace field

Exercise 5.3 *Is the centrifugal acceleration a Laplace field? What about the gravity field?*

Exercise 5.4 *Given the vector field:*

$$\boldsymbol{v} = \begin{pmatrix} ax + cyz^2 \\ by + cxz^2 \\ 2cxyz + d \end{pmatrix}.$$

Determine a, b, c, d such that $\boldsymbol{v}$ is a Laplace field.

5.3. Special cases and applications

Gradient field

Equation (5.7) becomes especially meaningful if $\boldsymbol{a}$ is a gradient field. Inserting $\boldsymbol{a} = \nabla U$ in (5.6) leads to a modified Poisson equation:

$$\boxed{\Delta U = \nabla\cdot\nabla U = \operatorname{div}\operatorname{grad} U = \begin{cases} -4\pi G\rho \text{ (inside)} & \rightarrow \text{Poissson} \\ \quad 0 \quad \text{(outside)} & \rightarrow \text{Laplace} \end{cases}}. \tag{5.9}$$

Then:

$$
\begin{aligned}
& \iiint\limits_V \nabla \cdot \nabla U \; \mathrm{d}V = \iint\limits_S \nabla U \cdot \mathrm{d}\boldsymbol{S} \,, \\
\Leftarrow: \quad & \iiint\limits_V \Delta U \; \mathrm{d}V = \iint\limits_S \nabla U \cdot \boldsymbol{n} \; \mathrm{d}S \,, \\
\Leftarrow: \quad & -4\pi G \iiint\limits_V \rho \; \mathrm{d}V = \iint\limits_S \frac{\partial U}{\partial n} \, \mathrm{d}S \,.
\end{aligned} \tag{5.10}
$$

The last step is due to Poisson's equation. The left side of (5.10) is the total mass of the body V. This total mass can be calculated by integrating the normal derivative over the surface S. In particular, when S would be an equipotential surface, the surface normal is opposite to the gravity direction. So for that particular case:

$$
S = \text{equipotential surface} : M = \frac{1}{4\pi G} \iint\limits_S g \, \mathrm{d}S \,. \tag{5.11}
$$

Application in geophysical prospecting

Relation (5.11) becomes especially interesting if we're considering disturbing masses and their disturbing gravity effect δg. Suppose the body V has a homogeneous mass distribution and correspondingly a homogeneous gravity field. Now consider a disturbing body buried in V, which has a density contrast $\delta\rho$ and a corresponding disturbing potential δU. Using (5.10) and (5.11) we may immediately write:

$$
\delta M = \frac{1}{4\pi G} \iint\limits_{S_0} \delta g \, \mathrm{d}S \,. \tag{5.12}
$$

The interesting point is that the surface integration only needs to be performed over an area S_0 close to the disturbing body, since it will have a limited gravitational influence only. This method allows to estimate the total mass of a buried disturbing body by using surface gravity measurements. It does not give the location (depth) or shape of it, though.

Constant vector times potential

Define $\boldsymbol{a}$ as $\Phi\boldsymbol{c}$, i.e. a potential field times a constant vector field $\boldsymbol{c}$. Its divergence is with (5.1d):

$$
\nabla \cdot \boldsymbol{a} = \nabla\Phi \cdot \boldsymbol{c} + \Phi\nabla \cdot \boldsymbol{c} = \nabla\Phi \cdot \boldsymbol{c} \,,
$$

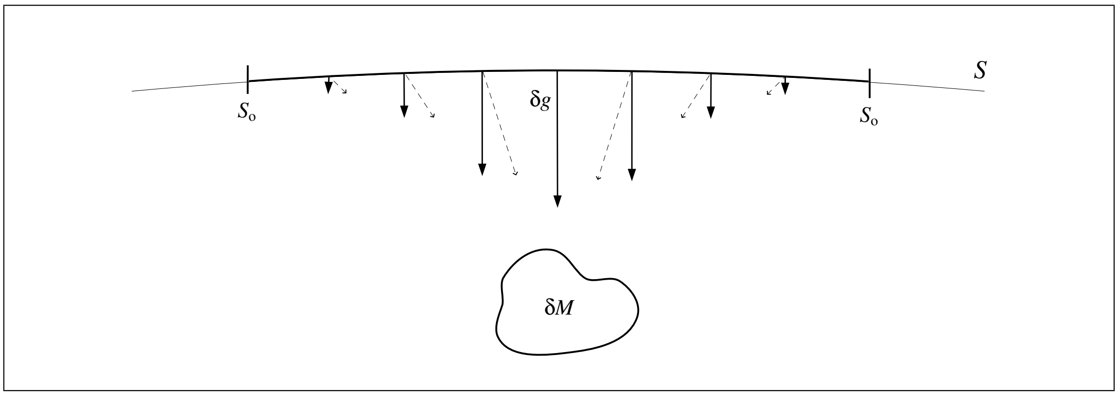


Figure 5.3.: Mass disturbance in a region S_0.

in which the second term at the right vanishes because a constant vector field has no divergence. Inserting into the divergence theorem, and taking $\boldsymbol{c}$ out of the integration yields:

$$\boldsymbol{c} \cdot \iiint_V \nabla\Phi \, \mathrm{d}V = \boldsymbol{c} \cdot \iint_S \Phi \, \mathrm{d}\boldsymbol{S} \, .$$

This relation must hold for any constant vector field $\boldsymbol{c}$. So we can conclude the following special case of the divergence theorem:

$$\iiint_V \nabla\Phi \, \mathrm{d}V = \iint_S \Phi \, \mathrm{d}\boldsymbol{S} \, . \tag{5.13}$$

Green's identities 1 and 2

Now use the following vector field: $\boldsymbol{a} = \Psi\nabla\Phi$. Using (5.1d) again gives:

$$\nabla \cdot \boldsymbol{a} = \nabla\Psi \cdot \nabla\Phi + \Psi\Delta\Phi \, .$$

Inserted into the divergence theorem, this vector field gives us the 1^{st} Green[4] identity:

$$\iiint_V (\nabla\Psi \cdot \nabla\Phi + \Psi\Delta\Phi) \, \mathrm{d}V = \iint_S \Psi\nabla\Phi \, \mathrm{d}\boldsymbol{S} = \iint_S \Psi \frac{\partial\Phi}{\partial n} \, \mathrm{d}S \, . \tag{5.14a}$$

This identity is not symmetric in Ψ and Φ. So let's repeat the exercise with $\boldsymbol{a} = \Phi\nabla\Psi$ and subtract the result from (5.14a). This results in Green's 2^{nd} identity:

$$\iiint_V (\Psi\Delta\Phi - \Phi\Delta\Psi) \, \mathrm{d}V = \iint_S \left(\Psi \frac{\partial\Phi}{\partial n} - \Phi \frac{\partial\Psi}{\partial n} \right) \mathrm{d}S \, . \tag{5.14b}$$

[4]George Green (1793–1841), British mathematician, physicist and miller.

Green's 1$^{\text{st}}$ identity will be used to prove the existence of solutions to the boundary value problems of the next section.

5.4. Boundary value problems

The starting point of this chapter was the question, whether we can determine the gravitational field in outer space without knowing the density structure of the Earth, but with the knowledge of the potential on the boundary. Take for instance a satellite, whose orbit has to be determined. In order to calculate the gravitational force we need to know the gravitational field in outer space.

The above problem is a boundary value problem (BVP). In general one tries to determine a function in a spatial domain from:

- its value on the boundary,
- its spatial behaviour, described by a partial differential equation (PDE).

In our particular case of gravitational fields we have two different partial differential equations: the Poisson equation, leading to an interior BVP and the Laplace equation, leading to the exterior BVP. We will be only concerned with the exterior problem, i.e. with the Laplace equation.

Classically, three BVPs are defined. Determine Φ in outer space from $\Delta\Phi = 0$ with one of the three following boundary conditions:

1$^{\text{st}}$ BVP	2$^{\text{nd}}$ BVP	3$^{\text{rd}}$ BVP
Dirichlet	**Neumann**	**Robin**
Φ	$\dfrac{\partial \Phi}{\partial n}$	$\Phi + c\dfrac{\partial \Phi}{\partial n}$

As a further condition we must require a regular behaviour for $r \to \infty$. The regularity condition:

$$\lim_{r\to\infty} \Phi(r) = 0\,, \tag{5.15}$$

may also be considered a boundary condition, if we think of a sphere with infinite radius as a second boundary.

Geodetic examples of the boundary functions are geopotential values (Dirichlet[5]), gravity disturbances (Neumann[6]) and gravity anomaly (Robin[7]). However, in a geodetic

[5]Peter Gustav Lejeune Dirichlet (1805-1859), German-French mathematician.
[6]Franz Ernst Neumann (1798–1895), German mathematician.
[7](Victor) Gustave Robin (1855–1897), lecturer in mathematical physics at the Sorbonne University, Paris.

context, the boundary S itself, i.e. the geoid, must be considered an unknown. This complication leads to the so-called geodetic boundary value problem. Another variation is the mixed BVP. One example is the altimetry-gravimetry BVP, which uses mixed boundary information: geoid ($\sim$ Dirichlet) over the oceans and gravity anomalies ($\sim$ Robin) on land.

Existence and uniqueness

The first step in mathematical literature after posing the BVP is to prove *existence* and *uniqueness* of the solution Φ. Existence will be the topic of next chapter, in which solutions are found in Cartesian and spherical coordinates.

Uniqueness is an issue that can be resolved with Green's 1st identity (5.14a) for Dirichlet's and Neumann's problem. Assume the BVP is not unique and we are able to find two solutions: Φ_1 and Φ_2. Their difference is called U. Insert U now into (5.14a), both for Φ and Ψ. We get:

$$\iiint\limits_V (\nabla U \cdot \nabla U + U \Delta U)\,\mathrm{d}V = \iint\limits_S U \frac{\partial U}{\partial n}\,\mathrm{d}S\,.$$

Since $\Delta U = \Delta(\Phi_1 - \Phi_2) = \Delta\Phi_1 - \Delta\Phi_2 = 0$ this leads to:

$$\iiint\limits_V \nabla U \cdot \nabla U\,\mathrm{d}V = \iint\limits_S U \frac{\partial U}{\partial n}\,\mathrm{d}S\,.$$

Both solutions Φ_1 and Φ_2 were obtained with the same boundary condition:

$$\begin{aligned} \Phi_1|_S &= \Phi_2|_S &&=: \quad U|_S = 0 \text{ (Dirichlet)} \\ \left.\frac{\partial \Phi_1}{\partial n}\right|_S &= \left.\frac{\partial \Phi_2}{\partial n}\right|_S &&=: \quad \left.\frac{\partial U}{\partial n}\right|_S = 0 \text{ (Neumann)} \end{aligned}$$

So either U or its normal derivative is zero on the boundary. As a result:

$$\iiint\limits_V |\nabla U|^2\,\mathrm{d}V = 0\,.$$

Since the integrand is always positive, ∇U must be zero, which can only happen if U is a constant.

- For the Dirichlet problem we immediately know that this constant is zero, since U is zero on the boundary already. Thus $\Phi_1 = \Phi_2$ and the Dirichlet problem has a unique solution.

- For the Neumann problem we are left with U being an arbitrary constant. So $\Phi_1 = \Phi_2 + c$ and the solution is unique up to a constant. This makes sense: only knowing the derivatives through boundary condition and PDE gives no absolute value.

6. Solving Laplace's equation

In this chapter we will solve the Laplace equation $\Delta\Phi = 0$ in Cartesian and in spherical coordinates. The former solution will be useful for planar, regional applications. The latter solution is a global solution. Both of them will lead to series developments in terms of orthogonal base functions: Fourier[1] series and spherical harmonic series respectively.

The solution of the Laplace equation is the most important step in solving the boundary value problem. The second step will be to express the boundary function in the same series development and determine the series' coefficients.

6.1. Cartesian coordinates

Consider the planar situation with x and y as horizontal coordinates and z the vertical one, $z < 0$ meaning the Earth interior and $z > 0$ being outer space. Our task is to solve $\Delta\Phi(x, y, z) = 0$ for $z > 0$. A very important first step in the solution strategy is *separation of variables*, which means:

$$\Delta\Phi(x, y, z) = \Delta f(x)g(y)h(z) = 0\,. \tag{6.1}$$

Applying the chain rule gives the following result:

$$\frac{\partial^2 f(x)}{\partial x^2}g(y)h(z) + f(x)\frac{\partial^2 g(y)}{\partial y^2}h(z) + f(x)g(y)\frac{\partial^2 h(z)}{\partial z^2} = 0\,.$$

For brevity each second derivative is indicated with a double prime. Due to the separation of variables, it is clear to which variable the differentiation is to be performed. The above equation is recast in a simpler form, after which we can divide by $\Phi = fgh$ itself:

$$\begin{aligned} & f''gh + fg''h + fgh'' = 0 \\ \overset{\cdot\frac{1}{fgh}}{\Longleftarrow}: \quad & \underbrace{\frac{f''}{f}}_{-n^2} + \underbrace{\frac{g''}{g}}_{-m^2} + \underbrace{\frac{h''}{h}}_{n^2+m^2} = 0\,. \end{aligned} \tag{6.2}$$

[1] Jean Baptiste Joseph Fourier (1768–1830), French mathematician.

The first summand of the left side only depends on x, the g-part only depends on y, etc. In order to be constant—zero in this case—each constituent at the left must be a constant by itself. It will turn out that the choice of constants below (6.2) is advantageous.

The separation of variables leads to a separation of the partial differential equation into three ordinary differential equations (ODE). They are respectively:

$$\frac{f''}{f} = -n^2 \ : \ f'' + n^2 f = 0 \tag{6.3a}$$

$$\frac{g''}{g} = -m^2 \ : \ g'' + m^2 g = 0 \tag{6.3b}$$

$$\frac{h''}{h} = n^2 + m^2 \ : \ h'' - (n^2 + m^2)h = 0 \tag{6.3c}$$

The solution of these ODEs is elementary. Since they are 2nd order ODEs, each gives rise to two basic solutions:

$$\begin{aligned} f_1(x) &= \cos nx & \text{and} \quad f_2(x) &= \sin nx \\ g_1(y) &= \cos my & \text{and} \quad g_2(y) &= \sin my \\ h_1(z) &= \mathrm{e}^{-\sqrt{n^2+m^2}z} & \text{and} \quad h_2(z) &= \mathrm{e}^{\sqrt{n^2+m^2}z} \end{aligned}$$

Of course, each of these solutions can be multiplied with a constant (or amplitude). The general solution $\Phi(x, y, z)$ will be the product of f, g and h. However, for each n and m we get a new solution. So we have to superpose all solutions in all (=8) possible combinations for every n and m:

$$\begin{aligned} \Phi(x, y, z) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} & (a_{nm} \cos nx \cos my + b_{nm} \cos nx \sin my + \\ & c_{nm} \sin nx \cos my + d_{nm} \sin nx \sin my)\, \mathrm{e}^{-\sqrt{n^2+m^2}z} + \\ & (e_{nm} \cos nx \cos my + f_{nm} \cos nx \sin my + \\ & g_{nm} \sin nx \cos my + h_{nm} \sin nx \sin my)\, \mathrm{e}^{\sqrt{n^2+m^2}z} \end{aligned} \tag{6.4}$$

Remark 6.1 *Had we chosen complex exponentials* e^{inx} *and* e^{imy} *instead of the sines and cosines as base functions, we would have obtained the compact expression:*

$$\Phi(x, y, z) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} A_{nm}\, \mathrm{e}^{\pm\sqrt{n^2+m^2}z + i(nx+my)} .$$

Remark 6.2 *Equation (6.4) has been formulated somewhat imprecise. For* $n = m = 0$ *only one coefficient (*a_{00} *or* e_{00}*) can exist; all others vanish. For* $n = 0$ or $m = 0$ *a number of terms has to vanish, too. (Which ones?)*

Exercise 6.1 *Check the solutions by inserting them back into the differential equations.*

The solution of the Laplace equation is not a solution of the BVP yet. It only gives us the behaviour of the potential Φ in outer space in terms of *base functions*. For the horizontal domain the base functions are sines and cosines. Thus the potential is a Fourier series in the horizontal domain. The n and m are the wave-numbers. The higher they are, the shorter the wavelengths.

The vertical base functions are called *upward continuation*, since they describe the vertical behaviour. They either have a damping (with the minus sign) or an amplifying effect. This effect apparently depends on n and m. The higher $\sqrt{n^2+m^2}$, i.e. the shorter the wavelength, the stronger the damping or amplifying effect. Thus the upward continuation either works as a low-pass filter (with the minus sign) or as a high-pass filter.

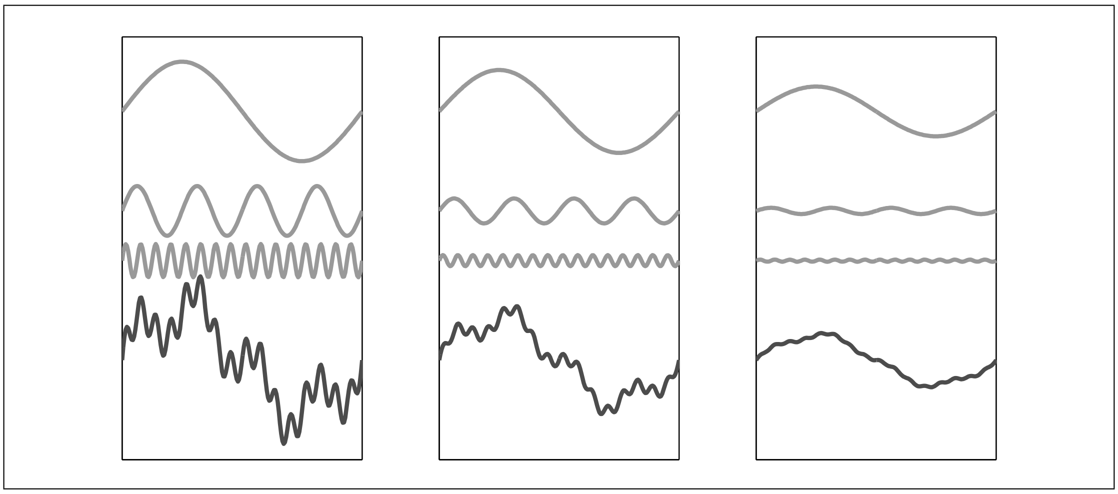


Figure 6.1.: Smoothing effect of upward continuation. The dark solid line (left panel) is composed of three harmonics of different frequencies. Upward continuation (middle panel and particularly the right panel) means low pass filtering: the higher the frequency, the stronger the damping.

6.1.1. Solution of Dirichlet and Neumann BVPs in x, y, z

Dirichlet. Given:

i) the general solution (6.4),

ii) the regularity condition (5.15) and

iii) the potential on the boundary $z = 0$: $\Phi(x, y, z\!=\!0)$, solve for Φ. The regularity condition demands already that we discard all contributions with the amplifying upward continuation. So half of (6.4), i.e. all terms with $e_{nm}, \ldots, h_{nm}$, are removed.

Next, suppose that the boundary function is developed into a 2D Fourier series:

$$\Phi(x, y, z=0) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} (p_{nm} \cos nx \cos my + q_{nm} \cos nx \sin my + \\ r_{nm} \sin nx \cos my + s_{nm} \sin nx \sin my) \ .$$

The full solution to the BVP is simply obtained now by *comparison* of these known spectral coefficients with the unknown coefficients from the general solution. In this case the full solution of the Neumann problem exists of:

$$\begin{aligned} &a_{nm} = p_{nm} \, , \quad b_{nm} = q_{nm} \, , \quad c_{nm} = r_{nm} \, , \quad d_{nm} = s_{nm} \, , \\ &e_{nm} = f_{nm} = g_{nm} = h_{nm} = 0 \, , \end{aligned} \tag{6.5}$$

and straightforward substitution of these coefficients into (6.4). To be precise, the regularity condition demands $a_{0,0}$ to be zero, too.

Boundary value at height $z = z_0$. A variant of this BVP is the case with the boundary function given at certain height. Strictly speaking, this is not a boundary anymore. Nevertheless, we can think of it as such. The same formalism applies. This situation occurs for instance in airborne gravimetry, where the gravity field is sampled at a (constant) flying altitude. Setting $z = z_0$ in (6.4) and using the same coefficients as above, we get the following comparison:

$$a_{nm} \mathrm{e}^{-\sqrt{n^2+m^2} z_0} = p_{nm} \, , \quad b_{nm} \mathrm{e}^{-\sqrt{n^2+m^2} z_0} = q_{nm} \, , \text{ etc.}$$

After solving for a_{nm}, b_{nm}, and so on, and substitution into the general solution again, we obtain:

$$\Phi(x, y, z) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} (p_{nm} \cos nx \cos my + \ldots) \, \mathrm{e}^{-\sqrt{n^2+m^2}(z - z_0)} \ .$$

Exercise 6.2 *Given the boundary function $\Phi(x, y, z=0) = 3 \cos 2x (1 - 4 \sin 6y)$, what is the full solution $\Phi(x, y, z)$ in outer space? And what, if this function was given at $z = 10$?*

Neumann. In case the vertical derivative is given as the boundary function we can proceed as before. The regularity condition again demands the vanishing of all amplification terms. Again we develop the boundary function into a 2D Fourier series:

$$\left.\frac{\partial\Phi(x,y,z)}{\partial z}\right|_{z=0} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left(p_{nm}\cos nx\cos my + q_{nm}\cos nx\sin my + \right.$$
$$\left. r_{nm}\sin nx\cos my + s_{nm}\sin nx\sin my\right) .$$

Obviously, the Fourier coefficients have a different meaning here. Now these coefficients have to be compared to the coefficients of the vertical derivative of the general solution (6.4):

$$-\sqrt{n^2+m^2}a_{nm} = p_{nm} \,, \quad -\sqrt{n^2+m^2}b_{nm} = q_{nm} \text{ , etc.}$$

After solving for a_{nm}, b_{nm}, and so on, and substitution into the general solution again, we obtain:

$$\Phi(x,y,z) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} -\left(\frac{p_{nm}}{\sqrt{n^2+m^2}}\cos nx\cos my + \dots\right)\mathrm{e}^{-\sqrt{n^2+m^2}z} .$$

Exercise 6.3 *Assume that the boundary function of exercise 6.2 is the vertical derivative $\partial\Phi/\partial z$ instead of Φ itself and solve these Neumann problems.*

6.2. Spherical coordinates

Using the same strategy as in 6.1 we can solve Laplace's equation and the first and second BVP globally. Fourier series cannot be applied on the sphere, so the solution of the Laplace equation will provide a new set of base functions. We will assume a spherical Earth and we will use spherical coordinates r, θ and λ. The strategy consists of the following steps:

i) write down Laplace's equation in spherical coordinates,
ii) separation of variables,
iii) solution of 3 separate ODEs,
iv) combining all possible solutions into a series development (superposition),
v) apply the regularity condition and discard conflicting solutions,
vi) develop boundary functions (Dirichlet and Neumann) into series,
vii) compare coefficients,
viii) write down full solution.

The Laplace equation in spherical coordinates reads:

$$\Delta\Phi = \frac{\partial^2\Phi}{\partial r^2} + \frac{2}{r}\frac{\partial\Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2} + \frac{\cot\theta}{r^2}\frac{\partial\Phi}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\lambda^2} = 0 \,.$$

After multiplication with r^2 we get the simpler form:

$$\Leftarrow: \boxed{r^2\frac{\partial^2\Phi}{\partial r^2} + 2r\frac{\partial\Phi}{\partial r} + \frac{\partial^2\Phi}{\partial\theta^2} + \cot\theta\frac{\partial\Phi}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2\Phi}{\partial\lambda^2} = 0}\,. \tag{6.6}$$

$$\Leftarrow: r^2\frac{\partial^2\Phi}{\partial r^2} + 2r\frac{\partial\Phi}{\partial r} + \Delta_S\Phi = 0$$

The latter form makes use of Δ_S, the surface Laplace operator or Beltrami[2] operator. Thus we can separate the Laplace operator in a radial and surface part. This idea is followed in the subsequent separation of variables, too. First we will treat the radial component by putting $\Phi(r,\theta,\lambda) = f(r)Y(\theta,\lambda)$. After that, the angular components are treated by setting $Y(\theta,\lambda) = g(\theta)h(\lambda)$. Applying the Laplace operator to $\Phi = fY$, omitting the arguments and using primes to abbreviate the derivatives, we can write:

$$\begin{aligned} & r^2 f''Y + 2rf'Y + f\Delta_S Y = 0 \\ \overset{\frac{1}{fY}}{\Leftarrow}: \quad & \underbrace{r^2\frac{f''}{f} + 2r\frac{f'}{f}}_{l(l+1)} + \underbrace{\frac{\Delta_S Y}{Y}}_{-l(l+1)} = 0 \end{aligned} \tag{6.7}$$

The first part only depends on r, whereas the second part solely depends on the angular coordinates. In order to yield zero for all r, θ, λ we must conclude that both parts are constant. It will turn out that $l(l+1)$ is a good choice. The radial equation becomes:

$$r^2 f'' + 2rf' - l(l+1)f = 0\,,$$

which can be readily solved by trying the function r^n. Insertion gives $n^2+n-l(l+1) = 0$, which results in n either l or $-(l+1)$. Thus we get the following two solutions (radial base functions):

$$f_1(r) = r^{-(l+1)} \quad \text{and} \quad f_2(r) = r^l\,. \tag{6.8a}$$

Next, we turn again to the surface Laplace operator and separate Y now into $g(\theta)h(\lambda)$:

$$\begin{aligned} & \Delta_S Y + l(l+1)Y = 0 \\ \Leftarrow: \quad & \frac{\partial^2 Y}{\partial\theta^2} + \cot\theta\frac{\partial Y}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\lambda^2} + l(l+1)Y = 0 \\ \Leftarrow: \quad & g''h + \cot\theta g'h + \frac{1}{\sin^2\theta}gh'' + l(l+1)gh = 0 \\ \overset{\cdot\frac{\sin^2\theta}{gh}}{\Leftarrow}: \quad & \underbrace{\sin^2\theta\frac{g''}{g} + \sin^2\theta\cot\theta\frac{g'}{g} + l(l+1)\sin^2\theta}_{m^2} + \underbrace{\frac{h''}{h}}_{-m^2} = 0 \end{aligned}$$

[2]Eugenio Beltrami (1835–1900), Italian mathematician.

Following the same reasoning again, the left part only depends on θ and the right part only on λ. The latter yields the ODE of an harmonic oscillator again, leading to the known solutions:

$$h'' + m^2 h = 0 \quad : \quad h_1(\lambda) = \cos m\lambda \quad \text{and} \quad h_2(\lambda) = \sin m\lambda \tag{6.8b}$$

The ODE of the θ-part is somewhat more elaborate. It is called the characteristic differential equation for the associated Legendre[3] functions. After division by $\sin^2\theta$ it reads:

$$\begin{aligned} & g'' + \cot\theta g' + \left(l(l+1) - \frac{m^2}{\sin^2\theta} \right) g = 0 \\ =: \; & g_1(\theta) = P_{lm}(\cos\theta) \quad \text{and} \quad g_2(\theta) = Q_{lm}(\cos\theta) \end{aligned} \tag{6.8c}$$

The functions $P_{lm}(\cos\theta)$ are called the associated Legendre functions of the 1st kind, the functions $Q_{lm}(\cos\theta)$ those of the 2nd kind. In 6.3.2 it will be explained, why the argument is $\cos\theta$ rather than θ itself. The functions $Q_{lm}(\cos\theta)$ are infinite at the poles, which is why they are discarded right away.

Solid and surface spherical harmonics. We have four base functions now, satisfying Laplace's equation:

$$\left\{ \begin{matrix} r^{-(l+1)} \\ r^l \end{matrix} \right\} P_{lm}(\cos\theta) \left\{ \begin{matrix} \cos m\lambda \\ \sin m\lambda \end{matrix} \right\} .$$

They are called *solid spherical harmonics*. Harmonics because they solve Laplace's equation, spherical because they have spherical coordinates as argument, and solid because they are 3D functions, spanning the whole outer space. If we leave out the radial part, we get the so-called *surface spherical harmonics*:

$$Y_{lm}(\theta, \lambda) = P_{lm}(\cos\theta) \left\{ \begin{matrix} \cos m\lambda \\ \sin m\lambda \end{matrix} \right\} .$$

The indices l and m of these functions have roles similar to the wave-numbers n and m in the Fourier series (6.4):

- l is the spherical harmonic *degree*,
- m is the spherical harmonic *order*, also known as the longitudinal wave-number, which becomes clear from (6.8b).

[3] Adrien Marie Legendre (1752–1833), French mathematician, (co-)inventor of the method of least-squares.

As we will see later, the degree l must always be larger than or equal to m: $l \geq m$. The full, general solution is attained now by adding all possible combinations of base functions, each multiplied by a constant, over all possible l and m.

$$\Phi(r,\theta,\lambda) = \sum_{l=0}^{\infty}\sum_{m=0}^{l} P_{lm}(\cos\theta)\,(a_{lm}\cos m\lambda + b_{lm}\sin m\lambda)\,r^{-(l+1)}$$
$$+P_{lm}(\cos\theta)\,(c_{lm}\cos m\lambda + d_{lm}\sin m\lambda)\,r^{l}\,. \tag{6.9}$$

If we compare this solution to (6.4) we see both similarities and differences:

i) In both cases we have base functions in all three space directions: two horizontal/surface directions and a vertical/radial direction.

ii) The functions $r^{-(l+1)}$ and r^l are now the radial continuation functions. Although they are different from the exponentials in the Cartesian case, they display the same behaviour: a continuous decay or amplification, that depends on the index. In the spherical case it only depends on l.

iii) For the Cartesian solution we had cosines and sines for both horizontal direction. In case of the spherical solution only in longitude direction.

iv) In latitude direction, the Legendre functions takes over the role of cosine/sines. We cannot attach a certain wave number to them, though.

6.2.1. Solution of Dirichlet and Neumann BVPs in r, θ, λ

Having solved the Laplace equation, it is very ease now to solve the 1st and 2nd BVP. In both cases the regularity condition

$$\lim_{r\to\infty}\Phi(r,\theta,\lambda) = 0$$

demands that the terms with the amplifying radial continuation (r^l) disappear. So $c_{lm} = d_{lm} = 0$ for all l, m combinations.

Dirichlet. The next step is to recognize that our boundary function is given at the boundary $r = R$. So obviously it must follow the general expression:

$$\Phi(R,\theta,\lambda) = \sum_{l=0}^{\infty}\sum_{m=0}^{l} P_{lm}(\cos\theta)\,(a_{lm}\cos m\lambda + b_{lm}\sin m\lambda)\,R^{-(l+1)}\,.$$

Next we develop our actual boundary function into surface spherical harmonics:

$$\Phi(R,\theta,\lambda) = \sum_{l=0}^{\infty}\sum_{m=0}^{l} P_{lm}(\cos\theta)\,(u_{lm}\cos m\lambda + v_{lm}\sin m\lambda)\ ,$$

in which u_{lm} and v_{lm} are known coefficients now. The solution comes from a simple comparison between these two series:

$$u_{lm} = a_{lm} R^{-(l+1)} \quad \text{and} \quad v_{lm} = b_{lm} R^{-(l+1)} \,.$$

Solving for a and b and inserting these spherical harmonic coefficients back into the general (6.9) yields:

$$\Phi(r, \theta, \lambda) = \sum_{l=0}^{\infty} \sum_{m=0}^{l} P_{lm}(\cos\theta) \left(u_{lm} \cos m\lambda + v_{lm} \sin m\lambda \right) \left(\frac{R}{r} \right)^{l+1} \,. \tag{6.10}$$

This equation says that if we know the function Φ on the boundary, we immediately know it everywhere in outer space because of the *upward continuation* term $(R/r)^{l+1}$. Since $r > R$ this becomes a damping (or low-pass filtering) effect. The higher the degree, the stronger the damping.

Neumann. Now our boundary function is the radial derivative of Φ:

$$\left.\frac{\partial \Phi(r, \theta, \lambda)}{\partial r}\right|_{r=R} = \sum_{l=0}^{\infty} \sum_{m=0}^{l} -(l+1) P_{lm}(\cos\theta) \left(a_{lm} \cos m\lambda + b_{lm} \sin m\lambda \right) R^{-(l+2)} \,.$$

The actual boundary function is developed into surface spherical harmonics again, with coefficients u_{lm} and v_{lm}. A comparison between known (u, v) and general (a, b) coefficients gives:

$$u_{lm} = -(l+1) a_{lm} R^{-(l+2)} \quad \text{and} \quad v_{lm} = -(l+1) b_{lm} R^{-(l+2)} \,.$$

Solving for a and b and inserting these spherical harmonic coefficients back into the general (6.9) yields:

$$\Phi(r, \theta, \lambda) = -R \sum_{l=0}^{\infty} \sum_{m=0}^{l} P_{lm}(\cos\theta) \left(\frac{u_{lm}}{l+1} \cos m\lambda + \frac{v_{lm}}{l+1} \sin m\lambda \right) \left(\frac{R}{r} \right)^{l+1} \,, \tag{6.11}$$

which solves Neumann's problem on the sphere.

Notation convention. The Earth's gravitational potential is usually indicated with the symbol V. The coefficients a_{lm} and b_{lm} will have the same dimension as the potential itself. It is customary, though, to make use of dimensionless coefficients C_{lm} and S_{lm}. The conventional way of expressing the potential becomes:

$$V(r, \theta, \lambda) = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r} \right)^{l+1} \sum_{m=0}^{l} P_{lm}(\cos\theta) \left(C_{lm} \cos m\lambda + S_{lm} \sin m\lambda \right) \,. \tag{6.12}$$

Indeed, the dimensioning is performed by the constant factor GM/R. Since the upward continuation only depends on degree l, it can be evaluated before the m-summation.

6.3. Properties of spherical harmonics

Surface spherical harmonics can be categorized according to the way they divide the Earth. Cosines and sines of wave number m will have $2m$ regularly spaced zeroes. Something similar cannot be said of a Legendre function $P_{lm}(\cos\theta)$. It exhibits $(l-m)$ zero crossings in a pattern that is close to equi-angular, but not fully regular. Nevertheless we can say that the sign changes of any $Y_{lm}(\theta,\lambda)$ in both directions divide the Earth in a chequer board pattern of $(l-m+1)\times 2m$ tiles, see fig. 6.2. We get the following classification:

- $m=0$: *zonal* spherical harmonics. When $m=0$, the sine-part vanishes, so that coefficients $S_{l,0}$ do not exist. Moreover, $\cos 0\lambda = 1$, so that no variation occurs in longitude. With $P_{l,0}$ we will get $(l+1)$ latitude bands, called *zones*.
- $l=m$: *sectorial* spherical harmonics. There are $2l$ sign changes in longitude direction. In latitude direction there will be zero. This does not mean that P_{ll} is constant, though. Anyway the Earth is divided in longitude bands called *sectors*.
- $l\neq m$ and $m\neq 0$: *tesseral* spherical harmonics. For all other cases we do get a pattern of *tiles* with alternating sign. After the Latin word for tiles, these functions are called tesseral.

6.3.1. Orthogonal and orthonormal base functions

Orthogonality is a key property of the base functions that solve Laplace's equations. It is the link between synthesis and analysis, i.e. it allows the forward and inverse transformation between a function and its spectrum.

Orthogonality is a common concept for vectors and matrices. Think of eigenvalue decomposition or of QR decomposition. We start with a simple example from linear algebra and extend the concept of orthogonality to functions.

From vectors to functions. Take the eigenvalue decomposition of a square symmetric matrix: $A = Q\Lambda Q^{\mathsf{T}}$. The columns of Q are the orthogonal eigenvectors $\boldsymbol{q}_i$. It is common to normalize them. So the scalar products between two eigenvectors becomes:

$$\boldsymbol{q}_i \cdot \boldsymbol{q}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i=j \\ 0 \text{ if } i\neq j \end{cases} .$$

The δ_{ij} at the right is called Kronecker delta function, which is 1 if its indices are equal and 0 otherwise. It is the discrete counterpart of the Dirac δ-function. Note that in case the eigenvectors are merely orthogonal, we would get something like $q_i\delta_{ij}$ at the right, in which q_i is the length of $\boldsymbol{q}_i$.

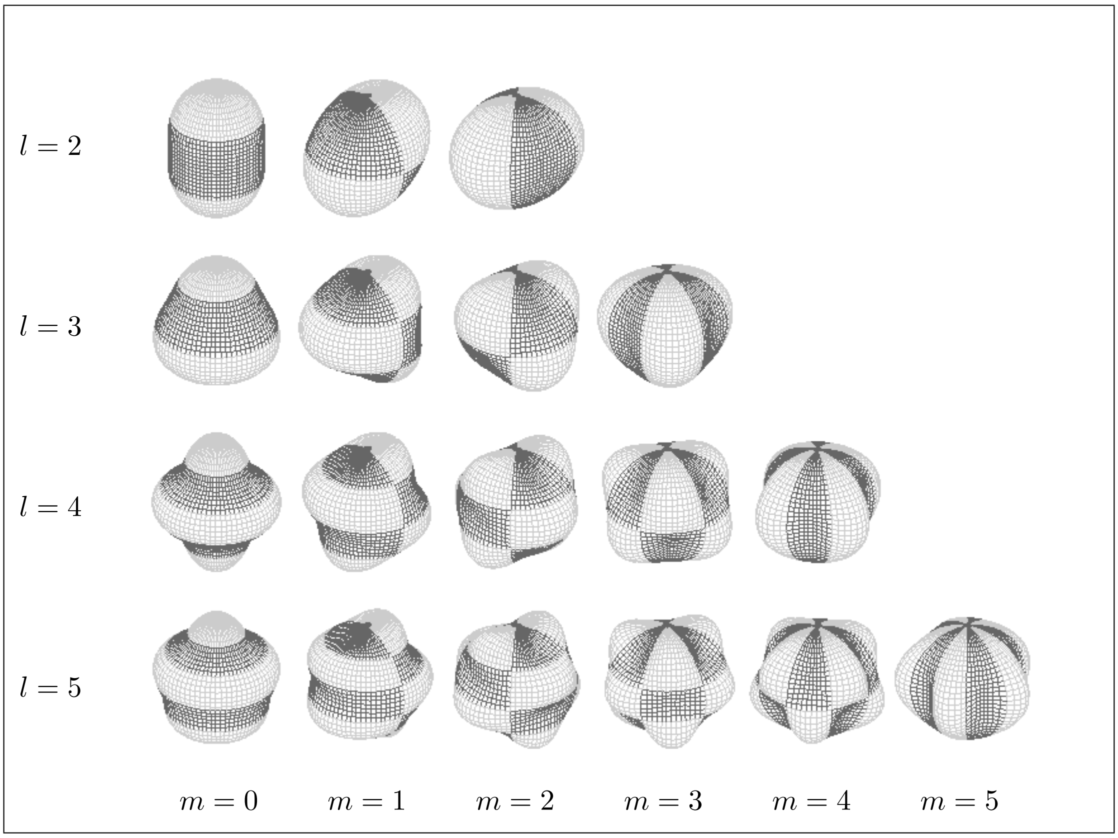


Figure 6.2.: Surface spherical harmonics up to degree and order 5.

Remark 6.3 *Realizing that the scalar product could have been written as* ${\boldsymbol{q}_i}^{\mathsf{T}}\boldsymbol{q}_j$*, the above equation is nothing else than stating the orthonormality* $Q^{\mathsf{T}}Q = I$*.*

In index notation, the above equation becomes:

$$\sum_{n=1}^{N} (q_i)_n (q_j)_n = \delta_{ij}\,,$$

in which $(q_i)_n$ stands for the n^{th} element of vector $\boldsymbol{q}_i$. A slightly different—and non-conventional—way of writing would be:

$$\sum_{n=1}^{N} q_i(n) q_j(n) \Delta n = \delta_{ij} \quad \text{with } \Delta n = 1\,.$$

The transition from discrete to continuous is made if we suppose that the length of the vectors becomes infinite ($N \to \infty$) and that the step size Δn simultaneously becomes infinitely small ($\mathrm{d}n$). At the same time we change the variable n into x now, which has a more continuous flavour.

$$\sum_{n=1}^{N} q_i(n) q_j(n) \Delta n \to \int q_i(x) q_j(x) \, \mathrm{d}x = \delta_{ij} \, .$$

Instead of orthonormal vectors we can now speak of orthonormal functions $q_i(x)$. So assessing the orthogonality of two functions means to multiply them, integrate the product and see if the result is zero or not.

Fourier. Let us see now if the base functions of Fourier series, i.e. sines and cosines are orthogonal or even orthonormal.

$$\frac{1}{2\pi} \int_0^{2\pi} \cos m\lambda \cos k\lambda \, \mathrm{d}\lambda = \frac{1}{2}(1 + \delta_{m,0}) \delta_{mk} \tag{6.13a}$$

$$\frac{1}{2\pi} \int_0^{2\pi} \cos m\lambda \sin k\lambda \, \mathrm{d}\lambda = 0 \tag{6.13b}$$

$$\frac{1}{2\pi} \int_0^{2\pi} \sin m\lambda \sin k\lambda \, \mathrm{d}\lambda = \frac{1}{2}(1 - \delta_{m,0}) \delta_{mk} \tag{6.13c}$$

Indeed, the cosines and sines are orthogonal. They are even nearly orthonormal: the norm—the right sides without δ_{mk}—is always one half, except for the special case $m = 0$, when the cosines become one and the sines zero.

Now consider the following 1D Fourier series:

$$f(x) = \sum_n a_n \cos nx + b_n \sin nx \, ,$$

and see what happens if we multiply it with a base function and evaluate the integral of that product:

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} f(x) \cos mx \, \mathrm{d}x &= \frac{1}{2\pi} \int_0^{2\pi} \left(\sum_n a_n \cos nx + b_n \sin nx \right) \cos mx \, \mathrm{d}x \\ &= \sum_n a_n \frac{1}{2\pi} \int_0^{2\pi} \cos nx \cos mx \, \mathrm{d}x + \; b_n \frac{1}{2\pi} \int_0^{2\pi} \sin nx \cos mx \, \mathrm{d}x \\ &= \sum_n a_n \frac{1}{2}(1 + \delta_{m,0}) \delta_{nm} \\ &= a_m \frac{1 + \delta_{m,0}}{2} \, . \end{aligned}$$

So the orthogonality—in particular the Kronecker symbol δ_{nm}—*filters out* exactly the right spectral coefficient a_m. In analogy to the assessment of orthogonality, this gives a recipe to perform spectral analysis:

- multiply the given function with a base function,
- integrate over the full domain,
- let orthogonality filter out the corresponding coefficient.

This demonstrates the statement at the beginning of this section, that orthogonality is a key property of systems of base functions.

$$\boxed{\begin{array}{rl} \text{synthesis:} & f(x) = \sum\limits_n a_n \cos nx + b_n \sin nx \\ \text{analysis:} & \left.\begin{array}{c} a_n \\ b_n \end{array}\right\} = \dfrac{1}{(1+\delta_{n,0})\pi} \displaystyle\int_0^{2\pi} f(x) \left\{\begin{array}{c} \cos nx \\ \sin nx \end{array}\right\} \mathrm{d}x \end{array}} . \tag{6.14}$$

Legendre—orthogonal. After this Fourier intermezzo we turn to the orthogonality of Legendre functions $P_{lm}(\cos\theta)$. It turns out that they are orthogonal indeed, but not orthonormal yet. The orthogonality is assessed now between two functions of the same order m:

$$\frac{1}{2}\int\limits_0^{\pi} P_{lm}(\cos\theta) P_{nm}(\cos\theta) \sin\theta \,\mathrm{d}\theta = \frac{1}{2l+1}\frac{(l+m)!}{(l-m)!}\delta_{ln} . \tag{6.15}$$

The procedure is visualize in fig. 6.3 for Legendre polynomials ($m = 0$): multiply two functions of the same order m and integrate over its domain, i.e. determine the grey area.

The fraction in front of the integral in (6.15) is due to $\int_0^{\pi} \sin\theta \,\mathrm{d}\theta = \int_{-1}^{1} \mathrm{d}t = 2$. So the *length* of a Legendre function depends on its degree and order. With this knowledge we're in a position to evaluate the orthogonality of surface spherical harmonics:

$$\begin{aligned} \frac{1}{4\pi}\iint\limits_{\sigma} Y_{lm}(\theta,\lambda) Y_{nk}(\theta,\lambda)\,\mathrm{d}\sigma \;=& \\ = \frac{1}{2}\int\limits_0^{\pi} P_{lm}(\cos\theta) P_{nk}(\cos\theta) &\left[\frac{1}{2\pi}\int\limits_0^{2\pi} \left\{\begin{array}{c} \cos m\lambda \cos k\lambda \\ \cos m\lambda \sin k\lambda \\ \sin m\lambda \cos k\lambda \\ \sin m\lambda \sin k\lambda \end{array}\right\} \mathrm{d}\lambda\right] \sin\theta\,\mathrm{d}\theta \\ = \frac{1}{2}\int\limits_0^{\pi} P_{lm}(\cos\theta) P_{nm}(\cos\theta) &\sin\theta\,\mathrm{d}\theta \frac{1}{2}(1+\delta_{m,0})\delta_{mk} \end{aligned}$$

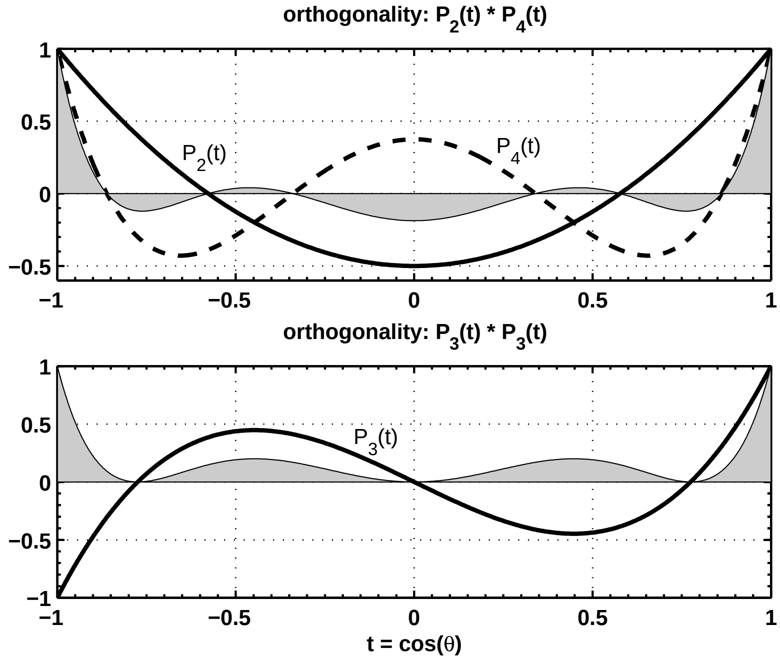


Figure 6.3.: Graphical demonstration of the concept of orthogonal functions.

$$
\begin{aligned}
&= \frac{1}{2}(1+\delta_{m,0})\frac{1}{2l+1}\frac{(l+m)!}{(l-m)!}\delta_{ln}\delta_{mk} && (6.16)\\
&= N_{lm}^{-2}\delta_{ln}\delta_{mk}\,.
\end{aligned}
$$

To be precise we must emphasize that the above result is not valid for cosine-sine combinations. Moreover we should have ruled out the case $m = 0$ in case of sine-sine orthogonality. In any case, the *length* of a surface spherical harmonic function is N_{lm}^{-1}.

Legendre—orthonormal. If we now multiply all base functions $Y_{lm}(\theta,\lambda)$ with N_{lm} they must become orthonormal. Thus N_{lm} is the called normalization factor. Normalized functions are indicated with an overbar.

$$N_{lm} = \sqrt{(2-\delta_{m,0})(2l+1)\frac{(l-m)!}{(l+m)!}} \tag{6.17a}$$

$$\bar{Y}_{lm}(\theta,\lambda) = N_{lm}Y_{lm}(\theta,\lambda) \tag{6.17b}$$

$$\bar{P}_{lm}(\cos\theta) = N_{lm}P_{lm}(\cos\theta) \tag{6.17c}$$

With this normalization we get the following results for the orthonormality:

$$\frac{1}{4\pi}\iint_\sigma \bar{Y}_{lm}(\theta,\lambda)\bar{Y}_{nk}(\theta,\lambda)\,\mathrm{d}\sigma \;=\; \delta_{ln}\delta_{mk} \tag{6.18a}$$

$$\frac{1}{2}\int_0^\pi \bar{P}_{lm}(\cos\theta)\bar{P}_{nm}(\cos\theta)\sin\theta\,\mathrm{d}\theta \;=\; (2-\delta_{m,0})\delta_{ln} \tag{6.18b}$$

Using these orthonormal base functions, the synthesis and analysis formulae of spherical harmonic computations read:

$$\boxed{\begin{array}{rrcl} \text{synthesis:} & f(\theta,\lambda) & = & \displaystyle\sum_l\sum_m \bar{P}_{lm}(\cos\theta)(\bar{a}_{lm}\cos m\lambda+\bar{b}_{lm}\sin m\lambda) \\ \text{analysis:} & \left.\begin{array}{c}\bar{a}_{lm}\\ \bar{b}_{lm}\end{array}\right\} & = & \displaystyle\frac{1}{4\pi}\iint_\sigma f(\theta,\lambda)\bar{Y}_{lm}(\theta,\lambda)\,\mathrm{d}\sigma \end{array}}\;. \tag{6.19}$$

Note that spherical harmonic coefficients are normalized by the inverse of the normalization factor: $\bar{a}_{lm}=N_{lm}^{-1}a_{lm}$, such that $a_{lm}P_{lm}(\cos\theta)=\bar{a}_{lm}\bar{P}_{lm}(\cos\theta)$, etc.

Exercise 6.4 *Verify the analysis formula by inserting the synthesis formula into it and applying the orthonormality property.*

6.3.2. Calculating Legendre polynomials and Legendre functions

Analytical recipe. Zonal Legendre functions of order 0 are called Legendre polynomials. Indeed they are polynomials in $t=\cos\theta$:

$$P_l(t)=P_{l,0}(\cos\theta)\,.$$

Legendre polynomials are calculated by the following differentiation process:

$$P_l(t)=\frac{1}{2^l\,l!}\frac{\mathrm{d}^l(t^2-1)^l}{\mathrm{d}t^l} \qquad \text{(Rodrigues)}\,. \tag{6.20a}$$

So basically, the quantity $(t^2-1)^l$ is differentiated l times. The result will be a polynomial of maximum degree $2l-l=l$, with only even or odd powers.

In a next step, the Legendre polynomials are differentiated m times by the following formula:

$$P_{lm}(t)=(1-t^2)^{m/2}\frac{\mathrm{d}^m P_l(t)}{\mathrm{d}t^m} \qquad \text{(Ferrers)}\,. \tag{6.20b}$$

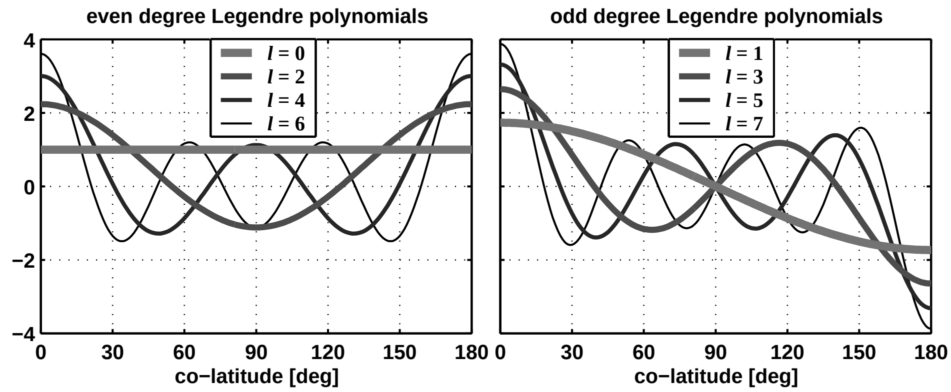


Figure 6.4.: Normalized Legendre polynomials for even and odd degrees

Thus we have a polynomial of order $(l-m)$, which is multiplied by a sort of modulation factor $(1-t^2)^{m/2}$. Replacing $t=\cos\theta$ again, we see that this factor is $\sin^m\theta$. For m odd, we do not have a polynomial anymore. Thus we speak of Legendre function. Trivial examples are:

$$P_0(t) = 1 \quad \text{and} \quad P_1(t) = \frac{1}{2}\frac{\mathrm{d}(t^2-1)}{\mathrm{d}t} = t = \cos\theta\,.$$

Exercise 6.5 *Calculate $P_{2,1}(\cos\theta)$.*

From (6.20a): $$P_2(t) = \frac{1}{8}\frac{\mathrm{d}^2(t^4-2t^2+1)}{\mathrm{d}^2t} = \frac{1}{2}(3t^2-1)\,,$$

Next, with (6.20b): $$P_{2,1}(t) = \sqrt{1-t^2}\frac{1}{2}\frac{\mathrm{d}(3t^2-1)}{\mathrm{d}t} = 3t\sqrt{1-t^2}\,,$$

Finally, with $N_{2,1} = \sqrt{2\cdot 5\cdot \frac{1}{3\cdot 2\cdot 1}}$: $$\bar{P}_{2,1}(t) = \sqrt{15}\,t\sqrt{1-t^2}\,.$$

From these formulae and from the example the following properties of Legendre functions become apparent:

i) For $m > l$ all $P_{lm}(\cos\theta)$ vanish.

ii) Legendre polynomials have either only even or only odd powers of t. Thus the $P_l(t)$ are either even or odd: $P_l(-t) = (-1)^l P_l(t)$.

iii) Since the effect of $(1-t^2)^{m/2}$ is only a sort of amplitude modulation, this symmetry statement can be generalized. Legendre functions are either even or odd, depending on the parity of $(l-m)$: $P_{lm}(-t) = (-1)^{l-m}P_{lm}(t)$.

iv) Odd Legendre functions must obviously assume the value 0 at the equator $(t=0)$.

v) Each Legendre function has $(l-m)$ zeroes on $t \in [-1;1]$, i.e. from pole to pole.

vi) The modulation factor $(1-t^2)^{m/2}$ becomes narrower around the equator for increasing order m. So does $P_{lm}(t)$.

vii) Sectorial Legendre functions $P_{mm}(\cos\theta)$ are simply a constant times the modulation factor $\sin^m\theta$.

These properties are reflected in fig. 6.4 and fig. 6.5. The functions in tbl. 6.1 have been derived from the formulae of Rodrigues[4] and Ferrers[5].

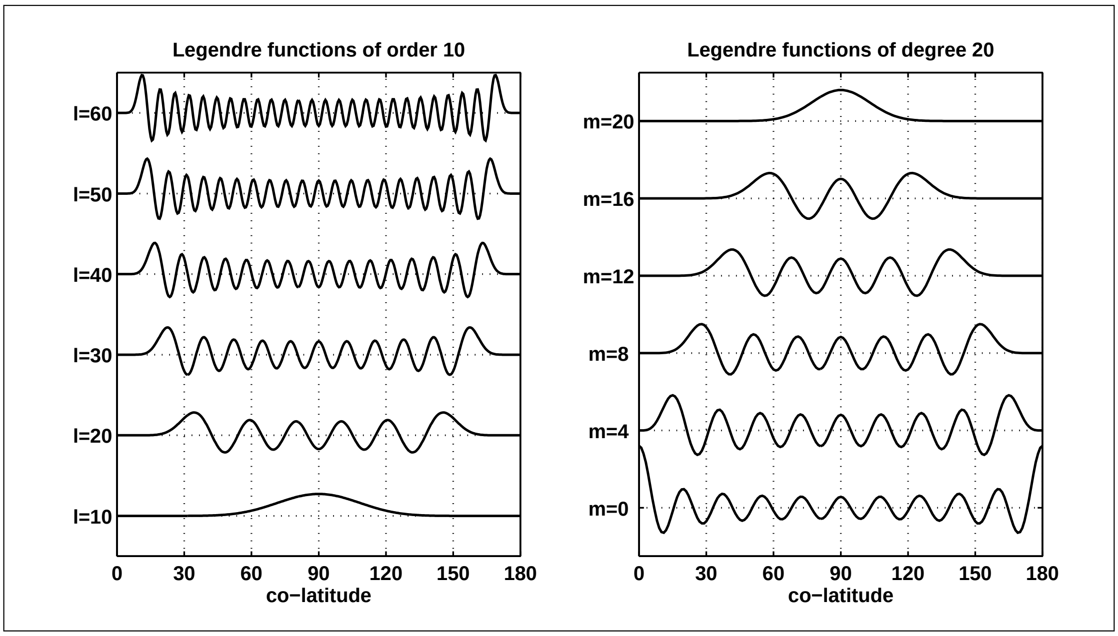


Figure 6.5.: Legendre functions $\bar{P}_{l,10}(\cos\theta)$ (left panel) and $\bar{P}_{20,m}(\cos\theta)$ (right panel).

Numerical recipe. Numerically it is more stable to calculate Legendre functions from the recursive relations below. These recursions are defined in terms of normalized Legendre functions. The strategy to calculate a certain $\bar{P}_{lm}(\cos\theta)$is to use the sectorial recursion to arrive at $\bar{P}_{mm}(\cos\theta)$. Then use the second recursion to increase the degree. For the first off-sectorial step, one must obviously assume $\bar{P}_{m-1,m}(\cos\theta)$ to be zero.

$$\bar{P}_{00}(\cos\theta) = 1 \tag{6.21a}$$

$$\bar{P}_{mm}(\cos\theta) = W_{mm}\sin\theta\,\bar{P}_{m-1,m-1}(\cos\theta) \tag{6.21b}$$

[4](Benjamin) Olinde Rodrigues (1794-1851), French mathematician, socialist and banker.

[5]Norman Macleod Ferrers (1830–...), English mathematician, professor in Cambridge, editor of the complete works of Green.

Table 6.1.: All Legendre functions and their (squared) normalization factor up till degree 3.

l	m	$P_{lm}(t)$	$P_{lm}(\cos\theta)$	N^2_{lm}
0	0	1	1	1
1	0	t	$\cos\theta$	3
	1	$\sqrt{1-t^2}$	$\sin\theta$	3
2	0	$\frac{1}{2}(3t^2-1)$	$\frac{1}{4}(1+3\cos 2\theta)$	5
	1	$3t\sqrt{1-t^2}$	$\frac{3}{2}\sin 2\theta$	$\frac{5}{3}$
	2	$3(1-t^2)$	$\frac{3}{2}(1-\cos 2\theta)$	$\frac{5}{12}$
3	0	$\frac{1}{2}t(5t^2-3)$	$\frac{1}{8}(3\cos\theta+5\cos 3\theta)$	7
	1	$\frac{3}{2}(5t^2-1)\sqrt{1-t^2}$	$\frac{3}{8}(\sin\theta+5\sin 3\theta)$	$\frac{7}{6}$
	2	$15(1-t^2)t$	$\frac{15}{4}(\cos\theta-\cos 3\theta)$	$\frac{7}{60}$
	3	$15(1-t^2)^{3/2}$	$\frac{15}{4}(3\sin\theta-\sin 3\theta)$	$\frac{7}{360}$

$$\bar{P}_{lm}(\cos\theta) = W_{lm}\left[\cos\theta\bar{P}_{l-1,m}(\cos\theta) - W^{-1}_{l-1,m}\bar{P}_{l-2,m}(\cos\theta)\right] \tag{6.21c}$$

with

$$W_{11} = \sqrt{3}\ ,\ W_{mm} = \sqrt{\frac{2m+1}{2m}}\ ,\ W_{lm} = \sqrt{\frac{(2l+1)(2l-1)}{(l+m)(l-m)}}$$

Many more recursive relations exist that step through the l, m-domain in a different way. The given set appears to be stable for a sufficiently large range of degrees and orders.

6.3.3. The addition theorem

One important formula is the *addition theorem* of spherical harmonics. Consider a Legendre polynomial in $\cos\psi_{PQ}$, in which the ψ_{PQ} is the spherical distance between points P and Q. The addition separates the composite angle argument into contributions from the points P and Q individually. This theorem will be used later on.

$$\begin{aligned} P_l(\cos\psi_{PQ}) &= \frac{1}{2l+1}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta_P)\bar{P}_{lm}(\cos\theta_Q)\left[\cos m\lambda_P\cos m\lambda_Q + \sin m\lambda_P\sin m\lambda_Q\right] \\ &= \frac{1}{2l+1}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta_P)\bar{P}_{lm}(\cos\theta_Q)\cos m(\lambda_P-\lambda_Q)\,. \end{aligned} \tag{6.22}$$

Note that at the at the left-hand side (LHS) we have a non-normalized polynomial. At the right-hand side (RHS) the Legendre functions are normalized.

Exercise 6.6 *Prove that the addition theorem for degree 1 gives the formula for calculating spherical distances:*

$$\cos\psi_{PQ} = \cos\theta_P \cos\theta_Q + \sin\theta_P \sin\theta_Q \cos\Delta\lambda_{PQ}\,. \tag{6.23}$$

Make use of tbl. 6.1.

6.4. Physical meaning of spherical harmonic coefficients

Spherical harmonic coefficients and mass distribution. The SH coefficients have a physical meaning. They are related to the internal mass distribution. In order to derive this relation we need to express the potential $V(R,\theta,\lambda)$ at the surface as a volume integral. Recall here Green's 2nd identity (5.14b), which is a consequence of Gauss's divergence theorem:

$$\iiint\limits_V (\Psi\Delta\Phi - \Phi\Delta\Psi)\,\mathrm{d}V = \iint\limits_S \left(\Psi\frac{\partial\Phi}{\partial n} - \Phi\frac{\partial\Psi}{\partial n}\right)\mathrm{d}S\,.$$

Now take a spherical surface S of radius R, such that $\partial/\partial n = \partial/\partial r$ and use the following potential functions:

$$\begin{aligned} \Phi &= r^l \bar{Y}_{lm}(\theta,\lambda) \qquad (\Delta\Phi = 0) \\ \Psi &= V(r,\theta,\lambda) \qquad (\Delta\Psi = -4\pi G\rho) \end{aligned}$$

Choosing the solid spherical harmonics $r^l Y_{lm}(\theta,\lambda)$ for Φ makes sense, since we treat the interior mass distribution. And since we are inside the masses, we have to use Poisson's equation (5.6) to describe V. Inserting these potentials into the volume integral at the LHS results in:

$$\text{LHS:}\quad \iiint\limits_V r^l \bar{Y}_{lm}(\theta,\lambda) 4\pi G\rho\,\mathrm{d}V = 4\pi G \iiint\limits_V r^l \bar{Y}_{lm}(\theta,\lambda)\rho\,\mathrm{d}V\,.$$

The volume V should not be mixed up with the potential V. The surface integral at the RHS results in:

$$\begin{aligned} \text{RHS:}\quad & \iint\limits_S \left(V l r^{l-1}\bar{Y}_{lm}(\theta,\lambda) - r^l \bar{Y}_{lm}(\theta,\lambda)\frac{\partial V}{\partial r}\right)_{r=R} \mathrm{d}S \\ &= R^l \iint\limits_S \left(\frac{l}{R}V(R,\theta,\lambda) - \left.\frac{\partial V}{\partial r}\right|_{r=R}\right)\bar{Y}_{lm}(\theta,\lambda)\,\mathrm{d}S\,. \end{aligned}$$

Note that all quantities are evaluated at $r = R$. Now if we develop the potential V in a SH series, as in (6.12), the bracketed expression in the integrand of the surface integral becomes:

$$\frac{l}{R}V(R,\theta,\lambda) - \left.\frac{\partial V}{\partial r}\right|_{r=R} = \frac{GM}{R^2}\sum_{l,m}(2l+1)\bar{P}_{lm}(\cos\theta)(\bar{C}_{lm}\cos m\lambda + \bar{S}_{lm}\sin m\lambda)\,.$$

If we insert this back into the surface integral we can let the orthonormality (6.18a) do its work. Since we are on a sphere of radius R, we get the slightly revised version:

$$\frac{1}{4\pi R^2}\iint\limits_S \bar{Y}_{lm}(\theta,\lambda)\bar{Y}_{nk}(\theta,\lambda)\,\mathrm{d}S = \delta_{ln}\delta_{mk}\,,$$

such that the RHS reduces to:

$$\text{RHS:}\quad \ldots = R^l 4\pi R^2\frac{GM}{R^2}(2l+1)\left\{\begin{matrix}\bar{C}_{lm}\\ \bar{S}_{lm}\end{matrix}\right\}.$$

So finally, if we equate LHS and RHS we get:

$$\left.\begin{matrix}\bar{C}_{lm}\\ \bar{S}_{lm}\end{matrix}\right\} = \frac{1}{2l+1}\frac{1}{MR^l}\iiint\limits_V r^l\bar{P}_{lm}(\cos\theta)\left\{\begin{matrix}\cos m\lambda\\ \sin m\lambda\end{matrix}\right\}\rho\,\mathrm{d}V\,. \tag{6.24a}$$

If we use non-normalized base functions and coefficients we would have:

$$\left.\begin{matrix}C_{lm}\\ S_{lm}\end{matrix}\right\} = (2-\delta_{m,0})\frac{(l-m)!}{(l+m)!}\frac{1}{MR^l}\iiint\limits_V r^l P_{lm}(\cos\theta)\left\{\begin{matrix}\cos m\lambda\\ \sin m\lambda\end{matrix}\right\}\rho\,\mathrm{d}V\,. \tag{6.24b}$$

Solid spherical harmonics in Cartesian coordinates. With this result, we are able to define single SH coefficients in terms of internal density distribution. But first we have to express the solid spherical harmonics in Cartesian coordinates:

$$\begin{aligned}
l=0:\ & r^0P_{0,0}(\cos\theta) &&= 1\\[1ex]
l=1:\ & r^1P_{1,0}(\cos\theta) &&= r\cos\theta &&= z\\
& r^1P_{1,1}(\cos\theta)\cos\lambda &&= r\sin\theta\cos\lambda &&= x\\
& r^1P_{1,1}(\cos\theta)\sin\lambda &&= r\sin\theta\sin\lambda &&= y\\[1ex]
l=2:\ & r^2P_{2,0}(\cos\theta) &&= r^2\tfrac{1}{2}(3\cos^2\theta-1) &&= \tfrac{1}{2}(2z^2-x^2-y^2)\\
& r^2P_{2,1}(\cos\theta)\cos\lambda &&= r^2 3\sin\theta\cos\theta\cos\lambda &&= 3xz\\
& r^2P_{2,1}(\cos\theta)\sin\lambda &&= r^2 3\sin\theta\cos\theta\sin\lambda &&= 3yz\\
& r^2P_{2,2}(\cos\theta)\cos 2\lambda &&= r^2 3\sin^2\theta\cos 2\lambda &&= 3(x^2-y^2)\\
& r^2P_{2,2}(\cos\theta)\sin 2\lambda &&= r^2 3\sin^2\theta\sin 2\lambda &&= 6xy
\end{aligned}$$

Inserting the Cartesian solid spherical harmonics into (6.24b) gives us:

For $l = 0$:

$$C_{0,0} = \frac{1}{M} \iiint \mathrm{d}m = 1$$

So the coefficient $C_{0,0}$ represents the total mass. But since all coefficients are divided by M, $C_{0,0}$ is one by definition.

For $l = 1$:

$$\begin{aligned}
C_{1,0} &= \frac{1}{MR} \iiint z \, \mathrm{d}m = \frac{1}{R} z_M \\
C_{1,1} &= \frac{1}{MR} \iiint x \, \mathrm{d}m = \frac{1}{R} x_M \\
S_{1,1} &= \frac{1}{MR} \iiint y \, \mathrm{d}m = \frac{1}{R} y_M
\end{aligned}$$

The coefficients of degree 1 represent the coordinates of the Earth's centre of mass $\boldsymbol{r}_M = (x_M, y_M, z_M)$ in our chosen coordinate system. Vice versa, if the origin coincides with the centre of mass, we consequently have $C_{1,0} = C_{1,1} = S_{1,1} = 0$.

For $l = 2$: (multiplied by MR^2)

$$\begin{aligned}
MR^2 C_{2,0} &= \tfrac{1}{2} \iiint (2z^2 - x^2 - y^2) \, \mathrm{d}m &&= \tfrac{1}{2} Q_{zz} &&= \tfrac{1}{2}(I_{xx} + I_{yy}) - I_{zz} \\
MR^2 C_{2,1} &= \iiint xz \, \mathrm{d}m &&= \tfrac{1}{3} Q_{xz} &&= -I_{xz} \\
MR^2 S_{2,1} &= \iiint yz \, \mathrm{d}m &&= \tfrac{1}{3} Q_{yz} &&= -I_{yz} \\
MR^2 C_{2,2} &= \tfrac{1}{4} \int \iiint (x^2 - y^2) \, \mathrm{d}m &&= \tfrac{1}{12}(Q_{xx} - Q_{yy}) &&= \tfrac{1}{4}(I_{yy} - I_{xx}) \\
MR^2 S_{2,2} &= \tfrac{1}{2} \iiint xy \, \mathrm{d}m &&= \tfrac{1}{6} Q_{xy} &&= -\tfrac{1}{2} I_{xy}
\end{aligned}$$

As can be seen, the coefficients of degree 2 are related to the mass moments. In the above formulae they are expressed in terms of both *quadrupole moments* Q and moments (and products) of inertia I. If we assume the vector $\boldsymbol{r}$ to be a column vector, the tensor of quadrupole moments $\mathbf{Q}$ is defined as:

$$\mathbf{Q} = \iiint \begin{pmatrix} 2x^2 - y^2 - z^2 & 3xy & 3xz \\ 3xy & 2y^2 - x^2 - z^2 & 3yz \\ 3xz & 3yz & 2z^2 - x^2 - y^2 \end{pmatrix} \mathrm{d}m = \iiint (3\boldsymbol{r}\boldsymbol{r}^T - \boldsymbol{r}^T\boldsymbol{r}I)\,\mathrm{d}m\,,$$

or in index notation:

$$Q_{ij} = \int (3x_i x_j - r^2\delta_{ij})\,\mathrm{d}m\,.$$

The tensor of inertia $\mathbf{I}$ is defined as:

$$\mathbf{I} = \iiint \begin{pmatrix} y^2 + z^2 & -xy & -xz \\ -xy & x^2 + z^2 & -yz \\ -xz & -yz & x^2 + y^2 \end{pmatrix} \mathrm{d}m = \iiint (\boldsymbol{r}^T\boldsymbol{r}I - \boldsymbol{r}\boldsymbol{r}^T)\,\mathrm{d}m\,,$$

or in index notation:

$$I_{ij} = \int (r^2\delta_{ij} - x_i x_j)\,\mathrm{d}m\,.$$

Quadrupole and inertia tensor are related by:

$$\mathbf{Q} = \mathrm{trace}(\mathbf{I})I - 3\mathbf{I}\ ,$$

or in index notation:

$$Q_{ij} = I_{ii}\delta_{ij} - 3I_{ij}\ .$$

We see that the coefficients $C_{2,1}$, $S_{2,1}$ and $S_{2,2}$ are proportional to the products of inertia I_{xz}, I_{yz} and I_{xy}, respectively. These are the off-diagonal terms. Only if they are zero, the coordinate axes are aligned with the principal axes of inertia. Vice versa, defining a coordinate system implies that one assigns values to $C_{2,1}$, $S_{2,1}$ and $S_{2,2}$. For instance, putting the conventional z-axis through the CIO-pole, means already that $C_{2,1}$ and $S_{2,1}$ will be very small, but unequal to zero. The choice of Greenwich as prime meridian defines the value of $S_{2,2}$.

Remark 6.4 *The coefficients $C_{0,0}$, $C_{1,0}$, $C_{1,1}$, $S_{1,1}$, $C_{2,1}$, $S_{2,1}$ and $S_{2,2}$ define the 7-parameter datum of the coordinate system.*

The coefficient $C_{2,0}$ is proportional to Q_{zz}, which can also be expressed as linear combination of the moments of inertia (i.e. the diagonal terms of $\mathbf{I}$. Thus $C_{2,0}$ expresses the flattening of the Earth. For the real Earth we have $C_{2,0} = -0.001\,082\,63$.

6.5. Tides revisited

Sorry, still ebb.

7. The normal field

Geodetic observables depend on the geometry ($\boldsymbol{r}$) and the gravity field (W) of the Earth. In general the functional relation will be nonlinear:

$$f = f(\boldsymbol{r}, W)\,.$$

The standard procedure is to develop the observable into a Taylor[1] series and to truncate after the linear term. As a result we obtain a linear observation equation. In order to perform this linearization we need a proper approximation of both geometry and gravity field of the Earth. Approximation by a sphere would be too inaccurate. The equatorial radius of the Earth is some 21.5 km larger than its polar radius. A rotationally symmetric ellipsoid is accurate enough, though. The geoid, which represents the physical shape of the Earth, doesn't deviate more than 100 m from the ellipsoid.

The potential and the gravity field that are consistent with such an ellipsoid are called *normal potential* and *normal gravity*. Thus, the *normal field* is an ellipsoidal approximation to the real gravity field. For the actual gravity potential we have the following linearization:

$$W = U + T \tag{7.1}$$

$$\begin{aligned} \text{with:}\quad W &= \text{full gravity potential} \\ U &= \text{normal potential } (W_0) \\ T &= \text{disturbing potential } (\delta W) \end{aligned}$$

Physical geodesy is a global discipline by nature. Therefore one has to make sure that the same normal field is used by everybody everywhere. This has been strongly advocated by the International Association of Geodesy (IAG) over the past century. As a result a number of commonly accepted so-called Geodetic Reference Systems (GRS) have evolved over the years: GRS30, GRS67 and the current GRS80, see tbl. 7.1. The parameters of the latter have been adopted by many global and regional coordinate systems and datums. For instance the WGS84 uses—with some insignificant changes—the GRS80 parameters.

[1] Brook Taylor (1685–1731).

Table 7.1.: Basic parameters of normal fields

		GRS30	GRS67	GRS80
a	[m]	6378 388	6378 160	6378 137
f^{-1}		297	298.247 167	298.257 222
GM_0	$[10^{14}\,\mathrm{m}^3\mathrm{s}^{-2}]$	3.986 329	3.986 030	3.986 005
ω	$[10^{-5}\,\mathrm{rad\,s}^{-1}]$	7.292 1151	7.292 115 1467	7.292 115

7.1. Normal potential

The geometry of the normal field, i.e. the ellipsoid is determined by two parameters for size and shape. We will choose the semi-major axis (a) and the flattening (f). The description of the physical field, i.e. the normal gravity potential, requires two further parameters. The strength is given by the geocentric gravitational constant (GM_0). And since we're dealing with gravity the Earth rotation rate (ω) must be involved, too.

This basic set of 4 parameters defines the normal field fully. See tbl. 7.1 for some examples. But vice versa, any set of 4 independent parameters will do. For GRS80 one uses the dynamical form factor J_2 (to be explained later on) instead of the geometric form factor f. The set a, J_2, GM_0, ω are called the *defining constants*.

The normal potential is defined to have the following properties:

- it is rotationally symmetric (zonal),
- it has equatorial symmetry,
- it is constant on the ellipsoid.

The latter property is the most fundamental. It defines the rotating Earth ellipsoid to be an *equipotential surface* or *level surface*.

This set of properties provides an algorithm to derive the normal potential and gravity formulae. Starting point is the SH development of a potential like (6.12), together with the centrifugal potential from 4.1. Together they give the normal potential U:

$$U(r,\theta,\lambda) = \frac{GM_0}{R}\sum_{l=0}^{\infty}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l} P_{lm}(\cos\theta)\,(c_{lm}\cos m\lambda + s_{lm}\sin m\lambda) + \frac{1}{2}\omega^2 r^2 \sin^2\theta\,.$$

Since we only want to represent the normal potential on and outside the ellipsoid, its mass distribution is irrelevant. For the following development it will be useful to assume all masses to be contained in a sphere of radius a. With the first property, rotational symmetry, we get the following simplification:

$$U(r,\theta) = \frac{GM_0}{a} \sum_{l=0}^{\infty} \left(\frac{a}{r}\right)^{l+1} c_{l,0} P_l(\cos\theta) + \frac{1}{2}\omega^2 r^2 \sin^2\theta\,. \tag{7.2}$$

The next property—equatorial symmetry—reduces the series to even degrees only. Actually only terms up to degree 8 are required. We thus get:

$$U(r,\theta) = \frac{GM_0}{a} \sum_{l=0,[2]}^{8} \left(\frac{a}{r}\right)^{l+1} c_{l,0} P_l(\cos\theta) + \frac{1}{2}\omega^2 r^2 \sin^2\theta\,. \tag{7.3}$$

For a better understanding—and easier calculus—we will continue now with only the degree 0 and 2 terms. The purpose is to derive an approximation, linear in f and $c_{2,0}$. We must keep in mind, though, that the actual development should run to degree 8. With the Legendre polynomial P_2 written out, and the centrifugal potential within brackets, we get:

$$U(r,\theta) = \frac{GM_0}{a}\left[\left(\frac{a}{r}\right)c_{0,0} + \left(\frac{a}{r}\right)^3 c_{2,0}\frac{1}{2}(3\cos^2\theta - 1) + \frac{1}{2}\frac{\omega^2 r^2 a}{GM_0}\sin^2\theta\right]. \tag{7.4}$$

In order to impose the main requirement—that of an equipotential ellipsoid—we must evaluate (7.4) on the ellipsoid. Thus we need an expression for the radius of the ellipsoid as a function of θ. The exact form is:

$$r(\theta) = \frac{ab}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}} = \frac{a}{\sqrt{1 + e'^2\cos^2\theta}}\,,$$

which is easily verified by inserting $x = r\sin\theta\cos\lambda$, etc. in the equation of the ellipsoid:

$$\frac{x^2+y^2}{a^2} + \frac{z^2}{b^2} = 1\,.$$

Since our goal is a formulation, linear in f, we expand (a/r) in a binomial series:

$$\begin{aligned}
(1+x)^q &= 1 + qx + \frac{q(q-1)}{2!}x^2 + \dots \\
:\ \sqrt{1+x} &= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \dots \\
:\ \sqrt{1+e'^2\cos^2\theta} &= 1 + \frac{1}{2}e'^2\cos^2\theta - \dots \\
:\ \frac{a}{r} &= 1 + f\cos^2\theta + \mathcal{O}(f^2)\,.
\end{aligned} \tag{7.5a}$$

The last step was due to the fact that

$$e'^2 = \frac{a^2 - b^2}{b^2} = \underbrace{\frac{a-b}{a}}_{f} \frac{a+b}{b}\frac{a}{b} = 2f + \mathcal{O}(f^2)\,.$$

Similarly, we can derive:

$$\left(\frac{a}{r}\right)^2 = 1 + 2f\cos^2\theta + \mathcal{O}(f^2)\,. \tag{7.5b}$$

The coefficient $c_{0,0} \equiv 1$. If we would insert (7.5a) into (7.4) we recognize that U depends on 3 small quantities, all of the same order of magnitude:

$$f \approx 0.003\,, \tag{7.6a}$$

$$c_{2,0} \approx 0.001\,, \tag{7.6b}$$

$$m = \frac{\omega^2 a^3}{GM_0} \approx 0.003\,. \tag{7.6c}$$

The quantity m is the relative strength of the centrifugal acceleration (at the equator) compared to gravitation. Now we will insert (7.5a) into (7.4) indeed, making use of f, $c_{2,0}$ and m. We will neglect all terms that are quadratic in these quantities, i.e. f^2, $fc_{2,0}$, fm, $c_{2,0}^2$ and $c_{2,0}m$. We then get

$$\begin{aligned} U &= \frac{GM_0}{a}\left[(1 + f\cos^2\theta) + c_{2,0}\frac{1}{2}(3\cos^2\theta - 1) + \frac{1}{2}m\sin^2\theta\right]\,, \\ &= \frac{GM_0}{a}\left[1 - \frac{1}{2}c_{2,0} + \frac{1}{2}m + \left(f + \frac{3}{2}c_{2,0} - \frac{1}{2}m\right)\cos^2\theta\right]\,. \end{aligned} \tag{7.7}$$

This normal potential still depends on θ, which contradicts the requirement of a constant potential. The latitude dependence only disappears if the following condition between f, $c_{2,0}$ and m holds:

$$\boxed{f + \tfrac{3}{2}c_{2,0} = \tfrac{1}{2}m}\,, \tag{7.8}$$

which means that the three quantities (7.6) cannot be independent. Using (7.8), we can eliminate one of the three small quantities. The constant normal potential value U_0 on the ellipsoid can be written as:

$$U_0 = \frac{GM_0}{a}\left(1 - \frac{1}{2}c_{2,0} + \frac{1}{2}m\right) = \frac{GM_0}{a}\left(1 + \frac{1}{3}f + \frac{1}{3}m\right) = \frac{GM_0}{a}\left(1 + f + c_{2,0}\right)\,. \tag{7.9}$$

Outside the ellipsoid one has to make use of (7.4) and apply the condition (7.8) to eliminate one of the three small quantities, e.g. $c_{2,0}$:

$$U(r,\theta) = \frac{GM_0}{a}\left[\frac{a}{r} + \left(\frac{a}{r}\right)^3\left(\frac{1}{2}m - f\right)\left(\cos^2\theta - \frac{1}{3}\right) + \frac{1}{2}\left(\frac{r}{a}\right)^2 m\sin^2\theta\right] . \tag{7.10}$$

Keep in mind that this is a linear approximation. For precise calculations one should revert to (7.3).

7.2. Normal gravity

Within the linear approximation of the previous section we can define normal gravity as the negative of the *radial* derivative of the normal potential. From (7.10) we get the following normal gravity outside the ellipsoid:

$$\gamma(r,\theta) = -\frac{\partial U}{\partial r} = \frac{GM_0}{a}\left[\frac{a}{r^2} + \frac{3}{r}\left(\frac{a}{r}\right)^3\left(\frac{1}{2}m - f\right)\left(\cos^2\theta - \frac{1}{3}\right) - \frac{r}{a^2}m\sin^2\theta\right] . \tag{7.11}$$

On the surface of the ellipsoid, using the same approximations as in the previous section, we will have:

$$\gamma(\theta) = \frac{GM_0}{a^2}\left[1 + m + (f - \frac{5}{2}m)\sin^2\theta\right] . \tag{7.12}$$

We cannot have a constant normal gravity on the surface of the ellipsoid simultaneously with a constant normal potential. Thus the θ-dependency remains. If we evaluate (7.12) on the equator and on the pole, we get the values:

$$\text{equator:} \qquad \gamma_a = \frac{GM_0}{a^2}(1 + f - \tfrac{3}{2}m)\,, \tag{7.13a}$$

$$\text{poles:} \qquad \gamma_b = \frac{GM_0}{a^2}(1 + m)\,. \tag{7.13b}$$

Note that $\gamma_b > \gamma_a$ since the pole is closer to the Earth's center of mass. Similar to the geometric flattening $f = (a - b)/a$ we now define the *gravity flattening*:

$$f^* = \frac{\gamma_b - \gamma_a}{\gamma_a}\,. \tag{7.14}$$

Numerically f^* is approximately 0.005, i.e. the same size as the other three small quantities. If we now insert the normal gravity on equator and pole (7.13) into this gravity flattening formula we end up with:

$$f^* = \frac{\frac{5}{2}m - f}{1 + f - \frac{3}{2}m} \approx \frac{5}{2}m - f\,,$$

leading to the remarkable result:

$$\boxed{f^* + f = \tfrac{5}{2}m} \,. \tag{7.15}$$

This result is known as the *theorem of Clairaut*[2]. It is remarkable, because it relates a dynamic quantity (f^*) to a geometric quantity (f) and the Earth's rotation (through m) in such a simple way.

Although (7.11) describes normal gravity outside the ellipsoid, it would be more practical to be able to upward continue the normal gravity value on the ellipsoid, i.e. to have a formula like $\gamma(h,\theta) = \gamma(\theta)g(h)$, in which $g(h)$ is some function of the height of above the ellipsoid. This is achieved by a Taylor series:

$$\gamma(h) = \gamma(h{=}0) + \left.\frac{\partial\gamma}{\partial h}\right|_{h=0} h + \frac{1}{2}\left.\frac{\partial^2\gamma}{\partial h^2}\right|_{h=0} h^2 \ldots .$$

After some derivation one ends up with the following upward continuation:

$$\gamma(h,\theta) = \gamma(\theta)\left[1 - \frac{2}{a}(1 + f + m - 2f\cos^2\theta)h + \frac{3}{a^2}h^2\right] . \tag{7.16}$$

7.3. Adopted normal gravity

Until (7.4) the development of the normal field was strictly valid. Starting with (7.4) approximations were introduced, such that we ended up with expressions in which all quadratic terms in f, m and $c_{2,0}$ were neglected. Even worse: for the upward continuation of the normal gravity a Taylor expansion was introduced.

Below, the exact analytical and precise numerical formulae are given. Formulae for an exact upward continuation do exist, but are not treated here.

7.3.1. Formulae

The theory of the equipotential ellipsoid was first given by Pizzetti[3] in 1894. It was further elaborated by Somigliana[4] in 1929. The following formula for normal gravity is generally valid. It is called the Somigliana-Pizzetti normal gravity formula:

$$\gamma(\phi) = \frac{a\gamma_a\cos^2\phi + b\gamma_b\sin^2\phi}{\sqrt{a^2\cos^2\phi + b^2\sin^2\phi}} = \gamma_a\frac{1 + k\sin^2\phi}{\sqrt{1 - e^2\sin^2\phi}} , \tag{7.17}$$

[2] Alexis Claude Clairaut (1713–1765).
[3] Paolo Pizzetti (1860-1918), Italian geodesist.
[4] Carlo Somigliana (1860-1955), Italian mathematical physicist.

with

$$e^2 = \frac{a^2 - b^2}{a^2} \quad \text{and} \quad k = \frac{b\gamma_b - a\gamma_a}{a\gamma_a}\,.$$

The variable ϕ is the geodetic latitude. In case of GRS80 the constants $a, b, \gamma_a, \gamma_b, e^2$ and k can be taken from the list in the following section.

For GRS80 the following series expansion is used:

$$\begin{aligned}\gamma(\phi) = \gamma_a \Big(1 \;&+\; 0.005\,279\,0414 \sin^2\phi \\ &+\; 0.000\,023\,2718 \sin^4\phi \\ &+\; 0.000\,000\,1262 \sin^6\phi \\ &+\; 0.000\,000\,0007 \sin^8\phi\Big)\,. \end{aligned} \tag{7.18a}$$

It has a relative error of 10^{-10}, corresponding to 10^{-4} mGal. For most applications, the following formula, which has an accuracy of 0.1 mGal, will be sufficient:

$$\gamma(\phi) = 9.780\,327 \Big(1 + 0.005\,3024 \sin^2\phi - 0.000\,0058 \sin^2 2\phi\Big) \quad [\mathrm{m\,s^{-2}}]\,. \tag{7.18b}$$

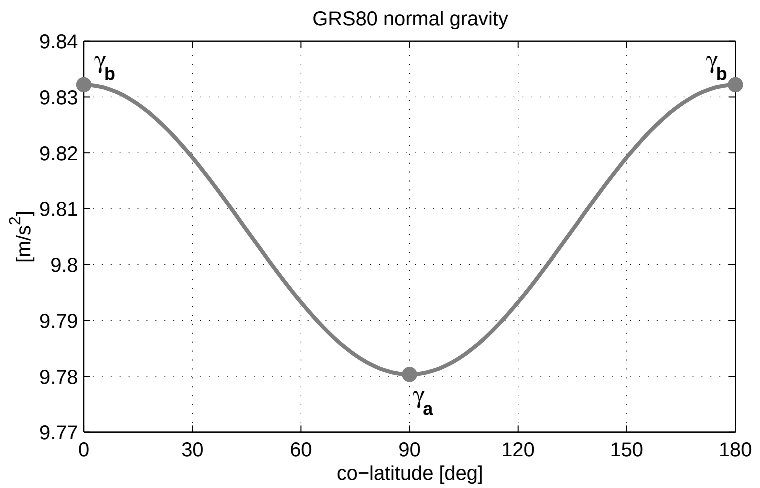


Figure 7.1.: GRS80 normal gravity.

Conversion between GRS30, GRS67 and GRS80. For converting gravity anomalies from the International Gravity Formula (1930) to the Gravity Formula 1980 we can use:

$$\gamma_{1980} - \gamma_{1930} = \Big(-16.3 + 13.7 \sin^2\phi\Big) \quad [\mathrm{mGal}]\,, \tag{7.19a}$$

where the main part comes from a change of the Potsdam reference value by -14 mGal. For the conversion from the Gravity Formula 1967 to the Gravity Formula 1980, a more accurate formula, corresponding to the precise expansion given above, is:

$$\gamma_{1980} - \gamma_{1967} = \left(0.8316 + 0.0782 \sin^2 \phi - 0.0007 \sin^4 \phi\right) \quad [\text{mGal}] \,. \tag{7.19b}$$

7.3.2. GRS80 constants

Defining constants (exact)

$$\begin{aligned}
a &= 6378\,137\,\text{m} && \text{semi-major axis}\\
GM_0 &= 3.986\,005 \cdot 10^{14}\,\text{m}^3\,\text{s}^{-2} && \text{geocentric gravitational constant}\\
J_2 &= 0.001\,082\,63 && \text{dynamic form factor}\\
\omega &= 7.292\,115 \cdot 10^{-5}\,\text{rad}\,\text{s}^{-1} && \text{angular velocity}
\end{aligned}$$

The following derived constants are accurate to the number of decimal places given. In case of doubt or in those cases where a higher accuracy is required, these quantities are to be computed from the defining constants.

Derived geometric constants

$$\begin{aligned}
b &= 6356\,752.3141\,\text{m} && \text{semi-minor axis}\\
E &= 521\,854.0097\,\text{m} && \text{linear eccentricity}\\
c &= 6399\,593.6259\,\text{m} && \text{polar radius of curvature}\\
e^2 &= 0.006\,694\,380\,022\,90 && \text{first eccentricity } (e)\\
e'^2 &= 0.006\,739\,496\,775\,48 && \text{second eccentricity } (e')\\
f &= 0.003\,352\,810\,681\,18 && \text{flattening}\\
f^{-1} &= 298.257\,222\,101 && \text{reciprocal flattening}\\
Q &= 10\,001\,965.7293\,\text{m} && \text{meridian quadrant}\\
R_1 &= 6371\,008.7714\,\text{m} && \text{mean radius } R_1 = (2a+b)/3\\
R_2 &= 6371\,007.1810\,\text{m} && \text{radius of sphere of same surface}\\
R_3 &= 6371\,000.7900\,\text{m} && \text{radius of sphere of same volume}
\end{aligned}$$

Derived physical constants

$$
\begin{aligned}
U_0 &= 62\,636\,860.850\,\mathrm{m^2\,s^{-2}} && \text{normal potential on ellipsoid} \\
J_4 &= -0.000\,002\,370\,912\,22 && \text{spherical} \\
J_6 &= 0.000\,000\,006\,083\,47 && \quad\text{harmonic} \\
J_8 &= -0.000\,000\,000\,014\,27 && \qquad\text{coefficients} \\
m &= 0.003\,449\,786\,003\,08 && m = \omega^2 a^2 b / GM_0 \\
\gamma_a &= 9.780\,326\,7715\,\mathrm{m\,s^{-2}} && \text{normal gravity at equator} \\
\gamma_b &= 9.832\,186\,3685\,\mathrm{m\,s^{-2}} && \text{normal gravity at poles} \\
\gamma_m &= 9.797\,644\,656\,\mathrm{m\,s^{-2}} && \text{average of normal gravity over ellipsoid} \\
\gamma_{45} &= 9.806\,199\,203\,\mathrm{m\,s^{-2}} && \text{normal gravity} \quad \phi = 45^\circ \\
f^* &= 0.005\,302\,440\,112 && f^* = (\gamma_b - \gamma_a)/\gamma_a \\
k &= 0.001\,931\,851\,353 && k = (b\gamma_b - a\gamma_a)/a\gamma_a
\end{aligned}
$$

8. Linear model of physical geodesy

In this chapter the linear observation model of physical geodesy is derived. The objective is to link observables (read: the boundary function) to the unknown potential and its derivatives. The following observables are discussed:

- potential (or geopotential numbers),
- astronomic latitude,
- astronomic longitude and
- gravity.

During the linearization process the concepts of *disturbance* and *anomalies* are introduced.

8.1. Two-step linearization

Linearizing a geodetic observable, related to gravity, is different from, say, linearizing a length observation in geomatics network theory. The observable is a functional of two classes of parameters: geometric ones ($\boldsymbol{r}$, position) and gravity-related ones (W, gravity potential). Both groups have approximate values and increments:

$$\text{observable:} \qquad y = y(\boldsymbol{r}, W) \text{ , with } \begin{cases} \boldsymbol{r} = \boldsymbol{r}_0 + \Delta \boldsymbol{r} \\ W = U + T \end{cases} . \tag{8.1}$$

The approximate location $\boldsymbol{r}_0$ will be identified in the Stokes approach with the ellipsoid, cf. 9. The approximate potential will be the corresponding normal field. For the following discussion this is not necessarily the case. Having two groups of parameters, the linearization process will be done in two steps:

1. linearize W alone: $\qquad y(\boldsymbol{r}, W) = y(\boldsymbol{r}, U) + \delta y + \mathcal{O}(T^2)$
2. linearize $\boldsymbol{r}$ too: $\qquad y(\boldsymbol{r}, W) = y(\boldsymbol{r}_0, U) + \sum_{i=1}^{3} \left.\frac{\partial y}{\partial r_i}\right|_{\boldsymbol{r}_0, U} \Delta r_i + \delta y + \mathcal{O}(\Delta \boldsymbol{r}^2, T^2)$

The partial derivatives have to be evaluated with the approximate potential field and at the approximate location. Note that we could have written the summation part as

$\nabla y \cdot \Delta \boldsymbol{r}$, too. Neglecting second order terms we get the following linearized quantities:

$$\text{Disturbance of } y: \qquad \delta y = y(\boldsymbol{r}, W) - y(\boldsymbol{r}, U)\,, \tag{8.2a}$$

$$\text{Anomaly of } y: \qquad \Delta y = y(\boldsymbol{r}, W) - y(\boldsymbol{r}_0, U)\,. \tag{8.2b}$$

In words, a disturbing quantity is the *observed quantity minus the computed one at the same location*. An anomalous quantity is *observed minus computed at an approximate location*. Disturbance and anomaly are obviously related by:

$$\Delta y = \delta y + \sum_i \left.\frac{\partial y}{\partial r_i}\right|_{\boldsymbol{r}_0, U} \Delta r_i\,.$$

Classically, physically geodesy is concerned with anomalies only. Since the geoid has to be considered an unknown, so must every observation location. Only the approximate geometry is known, leading to the anomalous quantities. This situation has changed, however, with the advent of global positioning by satellites. These systems allow to determine the location of a gravity related observable. Thus one can define disturbance quantities.

8.2. Disturbing potential and gravity

Disturbing potential. The disturbing potential is defined as $T = W - U$. For both W and U we will insert spherical harmonic series. From (6.12) and from 7:

$$W = \frac{GM}{r} + \frac{GM}{R} \sum_{l=2}^{\infty} \sum_{m=0}^{l} \left(\frac{R}{r}\right)^{l+1} \bar{P}_{lm}(\cos\theta)\,(\bar{C}_{lm}\cos m\lambda + \bar{S}_{lm}\sin m\lambda)$$

$$-$$

$$U = \frac{GM_0}{r} + \frac{GM_0}{R} \sum_{l=2}^{\infty} \sum_{m=0}^{l} \left(\frac{R}{r}\right)^{l+1} \bar{P}_{lm}(\cos\theta)\,(\bar{c}_{lm}\cos m\lambda + \bar{s}_{lm}\sin m\lambda)$$

$$=$$

$$T = \frac{\delta GM}{r} + \frac{GM_0}{R} \sum_{l=2}^{\infty} \sum_{m=0}^{l} \left(\frac{R}{r}\right)^{l+1} \bar{P}_{lm}(\cos\theta)\,(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda) \tag{8.3}$$

After taking the difference between first and second line we actually introduced an error by using the factor GM_0/R in front of the summations. However, GM_0 approximates the true GM very well ($\delta GM/GM \approx 10^{-8}$) so that it is safe to use GM_0 for $l \geq 2$.

In the above formulation, the approximate potential can be any spherical harmonic series. The linearization in the spectral domain reads:

$$\Delta\bar{C}_{lm} = \bar{C}_{lm} - \bar{c}_{lm}\,,$$

$$\Delta\bar{S}_{lm} = \bar{S}_{lm} - \bar{s}_{lm}\,.$$

In the spectral domain the distinction between disturbance and anomalous quantities makes no sense, since we cannot differentiate spectral coefficients towards position coordinates. Therefore we can use Δ to denote the spectrum of an anomalous quantity.

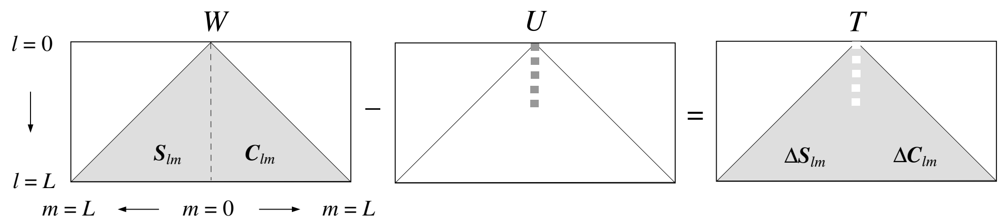


Figure 8.1.: Subtracting the normal potential U from the full gravity potential W in the spectral domain yields the disturbing potential T.

Conventionally the *normal field* is used for U. The spherical harmonic series above reduces to a series in J_2, J_4, J_6, J_8 only. With $J_l = -c_{l,0}$ and $\bar{c}_{l,0} = N_{l,0}^{-1}$ we arrive at:

$$\bar{c}_{l,0} = \frac{-J_l}{\sqrt{2l+1}}\,,$$

leading to:

$$\boxed{\begin{aligned} \Delta\bar{C}_{l,0} &= \bar{C}_{l,0} + \frac{J_l}{\sqrt{2l+1}} && , l = 2,4,6,8 \\ \Delta\bar{C}_{lm} &= \bar{C}_{lm} && , \text{all other } l,m \\ \Delta\bar{S}_{lm} &= \bar{S}_{lm} && \end{aligned}}\;. \tag{8.4}$$

The procedure of subtracting the normal field spectrum from the full spectrum is visualized in fig. 8.1.

Gravity disturbance. The scalar gravity disturbance is defined as:

$$\delta g = g(\boldsymbol{r}) - \gamma(\boldsymbol{r}) \quad \text{or} \quad \delta g = g_P - \gamma_P\,. \tag{8.5}$$

The vectorial version would be something like $\boldsymbol{\delta g}_P = \boldsymbol{g}_P - \boldsymbol{\gamma}_P$, which is visualized in fig. 8.2. However, before doing the subtraction both vectors $\boldsymbol{g}$ and $\boldsymbol{\gamma}$ must be in the same coordinate system. This is usually not the case. Normal gravity is usually expressed

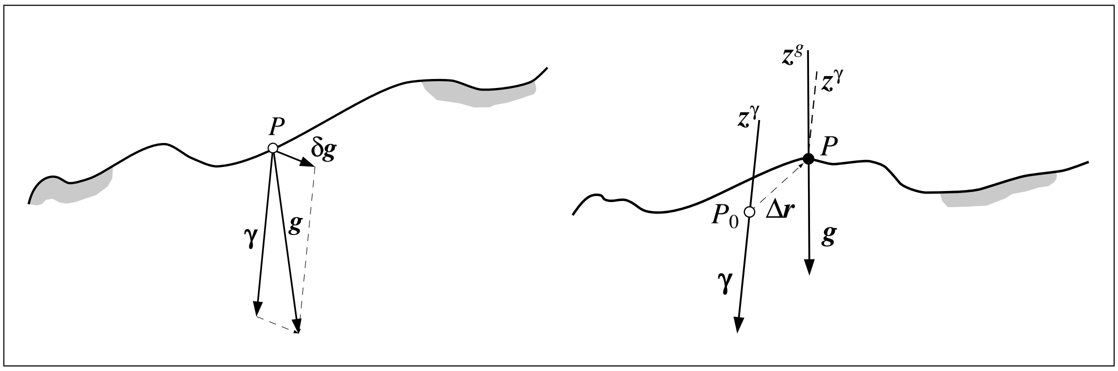


Figure 8.2.: Definition of gravity disturbance (left) and gravity anomaly (right).

in the local geodetic γ-frame, whereas $\boldsymbol{g}$ is usually expressed in the local astronomic g-frame. The coordinate system is indicated by a superindex.

$$\boldsymbol{g}^g = \begin{pmatrix} 0 \\ 0 \\ -g \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma}^\gamma = \begin{pmatrix} 0 \\ 0 \\ -\gamma \end{pmatrix}.$$

Before subtraction one of these vectors should be expressed in the coordinate system of the other. This is achieved through the following detour over global coordinate systems:

$$\boldsymbol{r}^\gamma = P_1 R_2(\tfrac{1}{2}\pi - \phi) R_3(\lambda) \boldsymbol{r}^\epsilon \tag{8.6a}$$

$$\boldsymbol{r}^g = P_1 R_2(\tfrac{1}{2}\pi - \Phi) R_3(\Lambda) \boldsymbol{r}^e \tag{8.6b}$$

$$\boldsymbol{r}^\epsilon = R_1(\epsilon_1) R_2(\epsilon_2) R_3(\epsilon_3) \boldsymbol{r}^e \tag{8.6c}$$

in which the e-frame denotes the conventional terrestrial system and the ϵ-frame denotes the global geodetic one. The angles ϵ_i represent a orientation difference between these two global systems. It is assumed that their origins coincide. Combination of these transformations gives us the transformation between the two local frames:

$$\boldsymbol{r}^\gamma = P_1 R_2(\tfrac{1}{2}\pi - \phi) R_3(\lambda) R_1(\epsilon_1) R_2(\epsilon_2) R_3(\epsilon_3) R_3(-\Lambda) R_2(\Phi - \tfrac{1}{2}\pi) P_1 \boldsymbol{r}^g$$

- neglect ϵ_i
- lot of calculus
- small angle approximation

$$= \begin{pmatrix} 1 & (\Lambda - \lambda)\sin\phi & (\Phi - \phi) \\ -(\Lambda - \lambda)\sin\phi & 1 & (\Lambda - \lambda)\cos\phi \\ -(\Phi - \phi) & -(\Lambda - \lambda)\cos\phi & 1 \end{pmatrix} \boldsymbol{r}^g \tag{8.7a}$$

$$= \begin{pmatrix} 1 & \delta\Lambda\sin\phi & \delta\Phi \\ -\delta\Lambda\sin\phi & 1 & \delta\Lambda\cos\phi \\ -\delta\Phi & -\delta\Lambda\cos\phi & 1 \end{pmatrix} \boldsymbol{r}^g \tag{8.7b}$$

$$= \begin{pmatrix} 1 & \psi & \xi \\ -\psi & 1 & \eta \\ -\xi & -\eta & 1 \end{pmatrix} \boldsymbol{r}^g\,. \tag{8.7c}$$

Between (8.7a) and (8.7b) we made use of the following definitions:

$$\text{Latitude disturbance:} \qquad \delta\Phi = \Phi_P - \phi_P\,, \tag{8.8a}$$

$$\text{Longitude disturbance:} \qquad \delta\Lambda = \Lambda_P - \lambda_P\,. \tag{8.8b}$$

The matrix element $\psi = \delta\Lambda\sin\phi = \eta\tan\phi$ is actually one of the contributions to the azimuth disturbance (see below). The step from (8.7b) to (8.7c) made use of the definition of the *deflection of the vertical*:

$$\text{Deflection in N-S:} \qquad \xi = \delta\Phi\,, \tag{8.9a}$$

$$\text{Deflection in E-W:} \qquad \eta = \delta\Lambda\cos\phi\,. \tag{8.9b}$$

Remark 8.1 *Remind that the orientation difference between the two global frames e and ϵ, represented by the angular datum parameters ϵ_i, has been neglected. The corresponding full transformation would become somewhat more elaborate. The equations are still manageable, though, since ϵ_i are small angles.*

We know our gravity vector $\boldsymbol{g}$ in the g-frame. Using (8.7c) it is easily transformed into the γ-frame now:

$$\boldsymbol{g}^\gamma = -g \begin{pmatrix} \xi \\ \eta \\ 1 \end{pmatrix}.$$

Finally, we are able to subtract the normal gravity vector from the gravity vector to get the *vector gravity disturbance*, see also fig. 8.3:

$$\delta\boldsymbol{g}^\gamma = \boldsymbol{g}^\gamma - \boldsymbol{\gamma}^\gamma = \begin{pmatrix} -g\xi \\ -g\eta \\ -g \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ -\gamma \end{pmatrix} = \begin{pmatrix} -g\xi \\ -g\eta \\ -\delta g \end{pmatrix} = \begin{pmatrix} -\gamma\xi \\ -\gamma\eta \\ -\delta g \end{pmatrix}. \tag{8.10}$$

The latter change from g into γ is allowed because the deflection of the vertical is such a small quantity that the precision of the quantity by which it is multiplied doesn't matter.

The gravity vector is the gradient of the gravity potential. Correspondingly, the normal gravity vector is the gradient of the normal potential. Thus we have:

$$\delta\boldsymbol{g} = \nabla W - \nabla U = \nabla(W - U) = \nabla T\,,$$

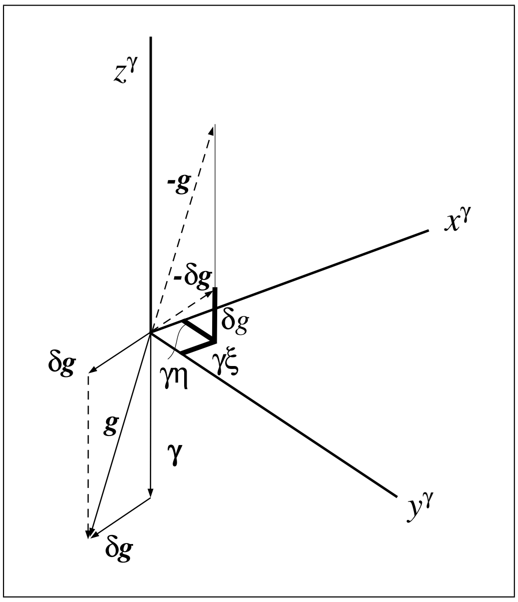


Figure 8.3: The gravity disturbance $\delta\boldsymbol{g}$ projected into the local geodetic frame is decomposed into deflections of the vertical $(\gamma\xi, \gamma\eta)$ and scalar gravity disturbance δg.

i.e. the gravity disturbance vector is the gradient of the disturbing potential. We can write the gradient for instance in local Cartesian or in spherical coordinates. Written out in components:

$$\boxed{\text{Local Cartesian: } \left\{\begin{aligned} \frac{\partial T}{\partial x} &= -\gamma\xi \\ \frac{\partial T}{\partial y} &= -\gamma\eta \\ \frac{\partial T}{\partial z} &= -\delta g \end{aligned}\right. \qquad \text{Spherical: } \left\{\begin{aligned} \frac{1}{r}\frac{\partial T}{\partial \phi} &= -\gamma\delta\Phi \\ \frac{1}{r\cos\phi}\frac{\partial T}{\partial \lambda} &= -\gamma\delta\Lambda\cos\phi \\ \frac{\partial T}{\partial r} &= -\delta g \end{aligned}\right.} \tag{8.11}$$

The RHS of each of these 6 equations represent the observable, the LHS the unknowns (derivatives of T).

Zenith and azimuth disturbances. Without going into too much detail the derivation of zenith and azimuth disturbances is straightforward. We can write any position vector $\boldsymbol{r}^\gamma$ in geodetic azimuth (α) and zenith (z). Similarly, any $\boldsymbol{r}^g$ can be written in astronomic azimuth (A) and zenith (Z). The transformation between both is known from (8.7c):

$$\begin{pmatrix} \sin z\cos\alpha \\ \sin z\sin\alpha \\ \cos z \end{pmatrix} = \begin{pmatrix} 1 & \psi & \xi \\ -\psi & 1 & \eta \\ -\xi & -\eta & 1 \end{pmatrix} \begin{pmatrix} \sin Z\cos A \\ \sin Z\sin A \\ \cos Z \end{pmatrix}.$$

After some manipulation one can derive:

$$\begin{aligned} &\text{Azimuth disturbance:} && \delta A = A_P - \alpha_P = \psi + \cot z(\xi \sin\alpha - \eta\cos\alpha)\,, && (8.12a)\\ &\text{Zenith disturbance:} && \delta Z = Z_P - z_P = -\xi\cos\alpha - \eta\sin\alpha\,. && (8.12b) \end{aligned}$$

8.3. Anomalous potential and gravity

Potential and geopotential numbers. We can immediately write down the anomalous potential according to (8.2b):

$$\begin{aligned} \Delta W_P = W_P - U_{P_0} &= \delta W_P + \sum_{i=1}^{3} \left.\frac{\partial W}{\partial r_i}\right|_{\boldsymbol{r}_0,U} \Delta r_i \\ &= T_P + \sum_{i=1}^{3} \left.\frac{\partial U}{\partial r_i}\right|_{P_0} \Delta r_i\,. \end{aligned}$$

The only problem is that the potential is not an observable quantity. The only thing one can observe are potential differences. In particular the (negative of) the difference between point P and the height datum point 0 is denoted as *geopotential number*:

$$C_P = W_0 - W_P = \int_0^P g\,\mathrm{d}H\,, \tag{8.13}$$

which can be obtained by a combination of levelling and gravimetry. In order to be able to use the above ΔW_P we consider the negative of the geopotential number $-C_P = W_P - W_0$ with the corresponding anomaly

$$\boxed{-\Delta C_P = \Delta W_P - \Delta W_0 = -\Delta W_0 + \sum_{i=1}^{3} \left.\frac{\partial U}{\partial r_i}\right|_{P_0} \Delta r_i + T_P}\,, \tag{8.14a}$$

in which

$$\Delta W_0 = \sum_{i=1}^{3} \left.\frac{\partial U}{\partial r_i}\right|_{P_0} \Delta r_i + T_0$$

not only accounts for the disturbing potential at the datum point, but also for the fact that the datum point may not be located at the geoid at all. It is defined in an operational way, after all.

In preparation of the final linear model, we will transform (8.14a) into a matrix-vector oriented structure:

$$-\Delta C_P = -\Delta W_0 + \begin{pmatrix} U_x & U_y & U_z \end{pmatrix}_{P_0} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} + T_P\,. \tag{8.14b}$$

The notation U_x means the partial derivative $\dfrac{\partial U}{\partial x}$, etc.

Gravity anomaly. The left hand side of the vector gravity anomaly, according to (8.2b) reads:

$$\Delta \boldsymbol{g} = \boldsymbol{g}_P - \boldsymbol{\gamma}_{P_0}\,,$$

which was visualized in fig. 8.2. Similar to the situation for the gravity disturbance we have to be careful with the coordinate systems in which these vectors are given. Gravity $\boldsymbol{g}_P$ is given in the g-frame at the location of P. Normal gravity $\boldsymbol{\gamma}_P$ is in the γ-frame, but now at the approximate location P_0. Similar to (8.7b) we can rotate between these frames. The translation will be dealt with at the RHS of the gravity anomaly below.

$$\boldsymbol{r}^\gamma = \begin{pmatrix} 1 & \Delta\Lambda\sin\phi & \Delta\Phi \\ -\Delta\Lambda\sin\phi & 1 & \Delta\Lambda\cos\phi \\ -\Delta\Phi & -\Delta\Lambda\cos\phi & 1 \end{pmatrix} \boldsymbol{r}^g\,.$$

Instead of the latitude and longitude disturbances, we have to use their anomalous counterpart in the rotation matrix.

$$\text{Latitude anomaly:} \qquad \Delta\Phi = \Phi_P - \phi_{P_0}\,, \tag{8.15a}$$

$$\text{Longitude anomaly:} \qquad \Delta\Lambda = \Lambda_P - \lambda_{P_0}\,. \tag{8.15b}$$

Inserting $\boldsymbol{g}^g = (0, 0, -g)$ the vector gravity anomaly becomes:

$$\boxed{\Delta\boldsymbol{g} = \begin{pmatrix} -g\Delta\Phi \\ -g\Delta\Lambda\cos\phi \\ -g \end{pmatrix}_P - \begin{pmatrix} 0 \\ 0 \\ -\gamma \end{pmatrix}_{P_0} = \begin{pmatrix} -\gamma\Delta\Phi \\ -\gamma\Delta\Lambda\cos\phi \\ -\Delta g \end{pmatrix}}\,. \tag{8.16a}$$

The last line contains the scalar gravity anomaly. Also note that g has been replaced by γ again as a multiplier to the latitude and longitude anomalies.

Equation (8.16) only represents the LHS of the anomaly equation, i.e. the part describing the observable. The RHS which describes the linear dependence on the unknowns reads:

$$\Delta\boldsymbol{g}_P = \delta\boldsymbol{g}_P + \sum_{i=1}^{3} \left.\frac{\partial \boldsymbol{g}}{\partial r_i}\right|_{\boldsymbol{r}_0, U} \Delta r_i$$

$$= \nabla T + \sum_{i=1}^{3} \frac{\partial \boldsymbol{\gamma}}{\partial r_i}\bigg|_{P_0} \Delta r_i$$

$$= \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} + \begin{pmatrix} U_{xx} & U_{xy} & U_{xz} \\ U_{yx} & U_{yy} & U_{yz} \\ U_{zx} & U_{zy} & U_{zz} \end{pmatrix}_{P_0} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} . \tag{8.16b}$$

The coefficient matrix is known as the Marussi[1]-matrix. It is a symmetric matrix containing the partial derivatives of $\boldsymbol{\gamma}$, which consists of partial derivatives of U already. So the Marussi matrix is composed of the second partial derivatives of U in all x, y, z-combinations.

The linear model. Let us combine the geopotential anomaly (8.14b) and the vector gravity anomaly (8.16) into one big vector-matrix system:

$$\begin{pmatrix} -\Delta C \\ -\gamma \Delta \Phi \\ -\gamma \cos\phi \Delta \Lambda \\ -\Delta g \end{pmatrix} = \begin{pmatrix} -\Delta W_0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} U_x & U_y & U_z \\ U_{xx} & U_{xy} & U_{xz} \\ U_{yx} & U_{yy} & U_{yz} \\ U_{zx} & U_{zy} & U_{zz} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} + \begin{pmatrix} T \\ T_x \\ T_y \\ T_z \end{pmatrix} .$$

Changing some signs and bringing over the γ's to the RHS finally gives us:

$$\boxed{\begin{pmatrix} \Delta C \\ \Delta \Phi \\ \Delta \Lambda \\ \Delta g \end{pmatrix} = \begin{pmatrix} \Delta W_0 \\ 0 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} U_x & U_y & U_z \\ U_{xx}/\gamma & U_{xy}/\gamma & U_{xz}/\gamma \\ U_{yx}/\beta & U_{yy}/\beta & U_{yz}/\beta \\ U_{zx} & U_{zy} & U_{zz} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} - \begin{pmatrix} T \\ T_x/\gamma \\ T_y/\beta \\ T_z \end{pmatrix}} . \tag{8.17}$$

The variable β is used in the 3^{rd} row as an abbreviation for $\gamma \cos\phi$. Equation (8.17) is the linear observation model for the anomalous geopotential number, latitude, longitude and gravity. The model can be extended with anomalous azimuth, zenith angle and other functionals of the gravity potential.

With (8.17) we have linked the observable boundary functions to the unknown potential and its derivatives. Together with the Laplace equation $\Delta T = 0$ and the regularity condition $\lim_{r\to\infty} T = 0$ we have achieved a description of the Geodetic Boundary Value Problem (GBVP). The GBVP distinguishes itself from other boundary value problems because the boundary itself is unknown. The geoid, which serves as the boundary

[1] Antonio Marussi (1908–1984), Italian geodesist.

function, is an equipotential surface and is therefore intimately linked to the field that we want to solve.

Remark 8.2 *In the rotation between g- and γ-frame we have again neglected the orientation parameters ϵ_i. If they are included in the linear model, they would show up in certain combinations in the vector below ΔW_0.*

The linear model has four observables $\{\Delta C, \Delta\Phi, \Delta\Lambda, \Delta g\}$ at the LHS. At the RHS we have the following unknowns:

- The vertical datum unknown ΔW_0. In principle there is one unknown per datum zone.
- The geometrical unknowns $\Delta \boldsymbol{r}$: three unknowns per point.
- The potential and its derivatives: four unknowns per point.

At first sight this situation seems hopelessly underdetermined. However, the disturbing potential T is a field quantity with a certain level of smoothness. A finite number of parameters, e.g. $\Delta\bar{C}_{lm}, \Delta\bar{S}_{lm}$ as in (8.3), will suffice to represent T and its derivatives. Due to the smoothness of the potential field this number is limited.

8.4. Gravity reductions

The solution of the Boundary Value Problems requires the function to be known on the boundary, cf. 5. In the next chapter we will identify the *geoid* as our boundary. Thus, for geoid determination we need to know the gravity field at the geoid. The surface gravity field g_{P_s} has to be reduced down to the geoid g_P, as visualized in fig. 8.4.

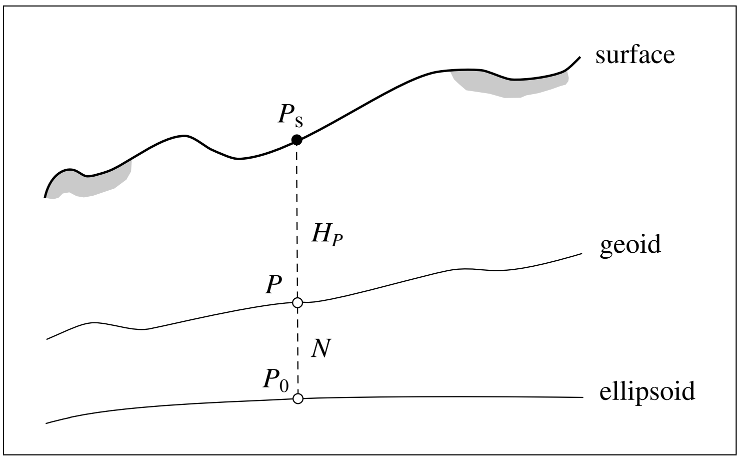


Figure 8.4: Reduction of gravity from surface point P_s down to the geoid point P. The ellipsoid is the set of approximate points P_0.

Over most of the land masses, the geoid is inside the Earth. When performing gravity reductions we need to make assumptions about the internal density structure of the Earth, at least about its upper parts. Geophysical relevance is not a criterion, though. Our main aim will be to obtain a reduced gravity field that is smooth. From a geodetic standpoint this is a sensible requirement for geoid computation. But consequently, the following sections may be irrelevant from a geophysical point of view.

We will only be concerned with reductions of gravity in this section. Reduction of the potential W_P doesn't make sense, since in any point P on the geoid we have a constant potential W_0. Reductions of further gravity related observables $(\Phi, \Lambda, A, Z, \ldots)$ will not be pursued here.

8.4.1. Free air reduction

The simplest assumption for gravity reductions would be to neglect all the topographic masses between surface point P_s and its footpoint on the geoid P. Gravity has to be downward continued along the plumbline through P_s and P over a distance H_P, the orthometric height of P_s. This type of reduction is known as *free air reduction*. In linear approximation we have:

$$g_P^{\mathrm{FA}} = g_{P_\mathrm{s}} - \left.\frac{\partial g}{\partial h}\right|_P H_P \,,$$

in which FA stands for the free-air gradient. The gravity gradient depends on the location and on the actual gravity field. In order to have a uniform definition of free-air gravity, one usually takes a fixed value. In this case FA comes from the simplification:

$$g \approx \frac{GM}{r^2} \quad : \quad \frac{\partial g}{\partial r} = -2\frac{GM}{r^3} = -2\frac{g}{r} \,.$$

Inserting numbers, we get:

$$\boxed{\begin{aligned} g_P^{\mathrm{FA}} &= g_{P_\mathrm{s}} + \mathrm{FA} \cdot H_P \\ &\text{with} \\ \mathrm{FA} &= 0.3086\,{}^{\mathrm{mGal}}\!/_{\mathrm{m}} \end{aligned}} \,. \tag{8.18}$$

The gradient is expressed in ${}^{\mathrm{mGal}}\!/_{\mathrm{m}}$. Thus heights must be expressed in m and gravity values in mGal. Note that FA is positive. If we go down, towards the center of the Earth, gravity will increase.

In calculating the free-air gravity field we have neglected the topography, in particular its gravitational attraction. This will show up in the reduced gravity field g^{FA} as a large correlation with the original topography. This is definitely unwanted for geoid computations. The topography must be considered in a different way.

8.4.2. Bouguer reduction

During the famous French arc measurement expedition to Peru, Bouguer[2] noticed the correlation between gravity and topography of the Andes. The reduction of gravity for topography is named after him: *Bouguer reduction*.

Bouguer plate. In its simplest form we approximate the topography surrounding point P_s by an infinite plate of thickness H_P: a Bouguer plate. This does not imply that we model a whole region by such a plate. On the contrary, in each terrain point we consider a new plate. The Bouguer plate can be considered a cylinder of height H_P and infinite radius. Its attraction was derived in (2.30):

$$g(\text{Bouguer plate}) = 2\pi G\rho H_P$$

$$: \quad \textsc{bo} = -\frac{\partial g(\text{B.p.})}{\partial h} = -2\pi G\rho = -0.1119\,{}^{\text{mGal}}\!/\!_{\text{m}}\,.$$

The latter number was derived with the conventional crustal density $\rho = 2670\,{}^{\text{kg}}\!/\!_{\text{m}^3}$. The Bouguer gradient BO, which is also expressed in ${}^{\text{mGal}}\!/\!_{\text{m}}$, is negative because gravity will become less if we remove the plate.

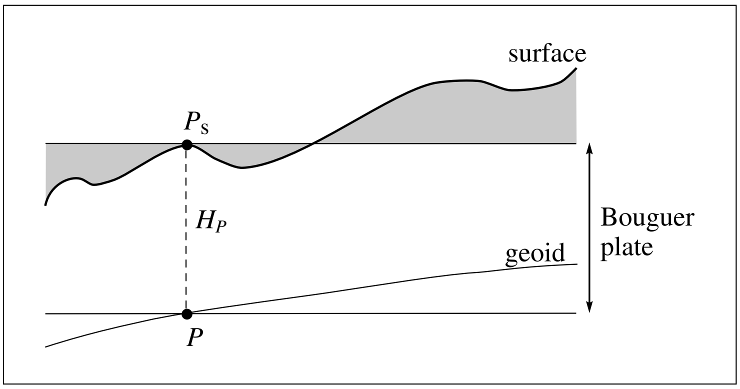


Figure 8.5: Bouguer plate: modeling the topography point-wise by an infinite slab of thickness H.

The procedure now consists of:

$$\begin{aligned}
\text{surface gravity:}&\quad g_{P_\mathrm{s}}\\
\text{remove plate:}&\quad +\textsc{bo}\cdot H_p\\
\text{go down to the geoid:}&\quad +\textsc{fa}\cdot H_P
\end{aligned}$$

[2] Pierre Bouguer (1698–1758).

$$
:\quad \boxed{\begin{aligned} g_P^{\mathrm{BO}} &= g_{P_\mathrm{s}} + (\textsc{bo} + \textsc{fa})\, H_P \\ &= g_{P_\mathrm{s}} + 0.1967 H_P \end{aligned}} \,. \tag{8.19}
$$

These steps will account for a large reduction in correlation between Bouguer gravity and original topography. Nevertheless they are still a rough approximation for two reasons:

i) The real topography surrounding a given point P_s is approximated as a plate.

ii) A constant density was assumed.

The latter point is interesting from a geophysical point of view. Variations in the Bouguer gravity field indicate variations in the underlying density structure.

Exercise 8.1 *Calculate free-air and Bouguer corrections to the gravity in Calgary, which has a height of roughly 1000 m.*

Topographical corrections. The former point can be remedied by taking into account a terrain correction TC. The problem with the Bouguer plate is the following. All topographic massed around point P_s that are higher than H_P are not accounted for by the Bouguer plate. Nevertheless they exert an upward, i.e. negative, gravitational attraction. To correct for this effect requires a positive quantity. On the other hand, the topography around point P_s below H_P has been overcompensated with the negative BO gradient. To correct for that also requires a positive quantity. So the Bouguer reduction makes a systematic error.

The effect of the terrain is calculated from knowledge of the terrain H_Q, e.g. from a digital terrain model (DTM):

$$
\textsc{tc} = \int_x \int_y \int_{z=H_P}^{H} \frac{z - H_P}{r^3} \rho(x, y, z)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \,, \tag{8.20}
$$

with r the distance from computation point P to all terrain points. For terrain points above P the same crustal density as in BO should be used. For terrain points below P its negative value should be used, since these areas represent mass deficiencies. Numerous approaches to an efficient calculation of TC from (8.20) exist. Note that this equation also allows the use of non-homogeneous density distributions, if they were known.

Correcting for the terrain effect yields the complete or refined Bouguer gravity field, or the terrain-corrected gravity field:

$$
\boxed{g_P^{\mathrm{TC}} = g_{P_\mathrm{s}} + (\textsc{bo} + \textsc{fa})\, H_P + \textsc{tc}} \,. \tag{8.21}
$$

The terrain correction further reduces the correlation with topography and thus smoothens the gravity field.

8.4.3. Isostasy

On the aforementioned Peru-expedition Bouguer noticed that the deflections of the vertical (ξ, η) were smaller than expected based on calculations of the attraction of the Andes Mountains, see fig. 8.6. The same phenomenon was observed with more accuracy in the survey of India (19^{th} century) under the guidance of Everest[3]. Similarly, Helmert[4] calculated geoid undulations of up to $\pm 500\,\mathrm{m}$, based on topography, whereas reality shows variations of up to $\pm 100\,\mathrm{m}$ only.

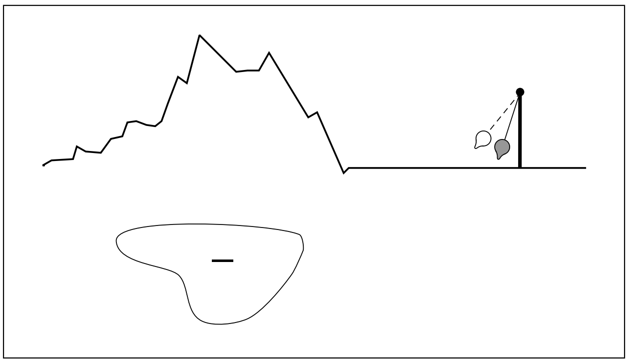


Figure 8.6: The deflection of the plumb bob is smaller than what one would expect based on the visible topography. Consequently, negative masses, that is a volume of lower density, must exist below the topography.

From these observations it was asserted that a certain compensation below the topography must exist, with a negative density contrast. This led to the concept of *isostasy* which assumes equilibrium within each column of the Earth down to a certain *level of compensation*. The equilibrium condition reads:

$$\text{isostasy:} \qquad \int_{-T}^{H} \rho \, \mathrm{d}z \,.$$

Since the interior density structure is hardly known, two competing models developed more or less simultaneously. Pratt[5] proposed a constant compensation depth T_0. Consequently the density will be variable across columns. Airy[6] on the other hand assumed a constant density, which implies a variable T.

[3]Sir G. Everest (1790–1866), Surveyor General of India (1830–1843).
[4]Friedrich Robert Helmert (1843–1917), German geodesist.
[5]John Henry Pratt (1811–1871), Cambridge trained mathematician, Archdeacon of Calcutta.
[6]Sir George Biddell Airy (1801–1892), Professor of Astronomy at Cambridge University, Astronomer Royal, president of the Royal Society.

Pratt-Hayford: constant T_0. Pratt's isostasy model was further developed for geodetic use by Hayford[7]. The model assumes a constant T_0. The corresponding density in absence of topography would be ρ_0. Thus the isostasy condition for a given column i reads:

$$\text{land:} \qquad \rho_i(T_0 + H_i) = \rho_0 T_0 \tag{8.22a}$$

$$\text{ocean:} \quad \rho_i(T_0 - d_i) + \rho_\mathrm{w} d_i = \rho_0 T_0 \tag{8.22b}$$

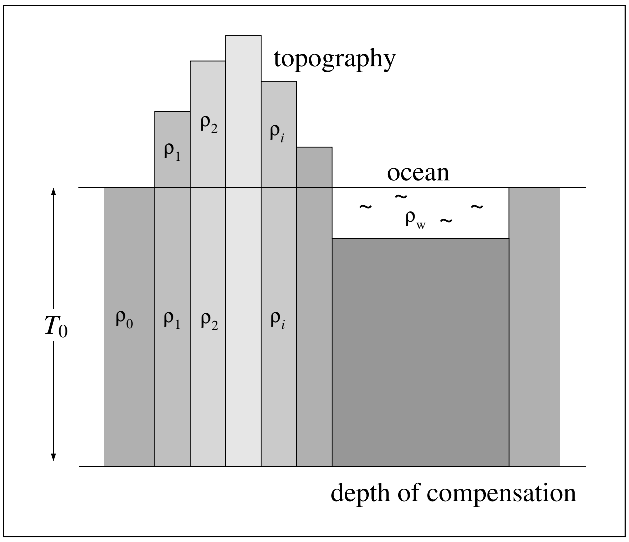


Figure 8.7: Pratt-Hayford model: constant depth of compensation T_0 and variable column density ρ_i.

In the ocean column we have salt water density $\rho_\mathrm{w} = 1030\,\mathrm{kg/m^3}$. For geodetic purposes, our only aim is to smoothen the gravity field. Any combination (T_i, ρ_i) that fulfils that aim will do. But also geophysically the Pratt-Hayford model has some merit. Based on assumptions of density Hayford calculated a compensation depth of $T_0 = 113.7\,\mathrm{km}$ from deflections of the vertical over the United States. This depth corresponds to lithospheric thickness estimates.

Airy-Heiskanen: constant ρ_c. Airy's model was developed for geodetic purposes by Heiskanen[8]. The Airy-Heiskanen model is similar to a floating iceberg. Instead of ice we now have crustal material (ρ_c) and the denser sea-water becomes mantle material (ρ_m). If there is an elevation (mountain) above the surface, there must be a corresponding *root* sticking into the mantle. Since crustal material is lighter than the mantle there will be

[7] John Fillmore Hayford (1868–1925), chief of the division of geodesy of the US Coast and Geodetic Survey.

[8] Veiko Aleksanteri Heiskanen (1895–1971), director of the Finnish Geodetic Institute, founder of the Department of Geodetic Science at the Ohio State University.

an upward buoyant force that balances the downward gravity force of the mountains. Thus a mountain is more or less floating on the mantle.

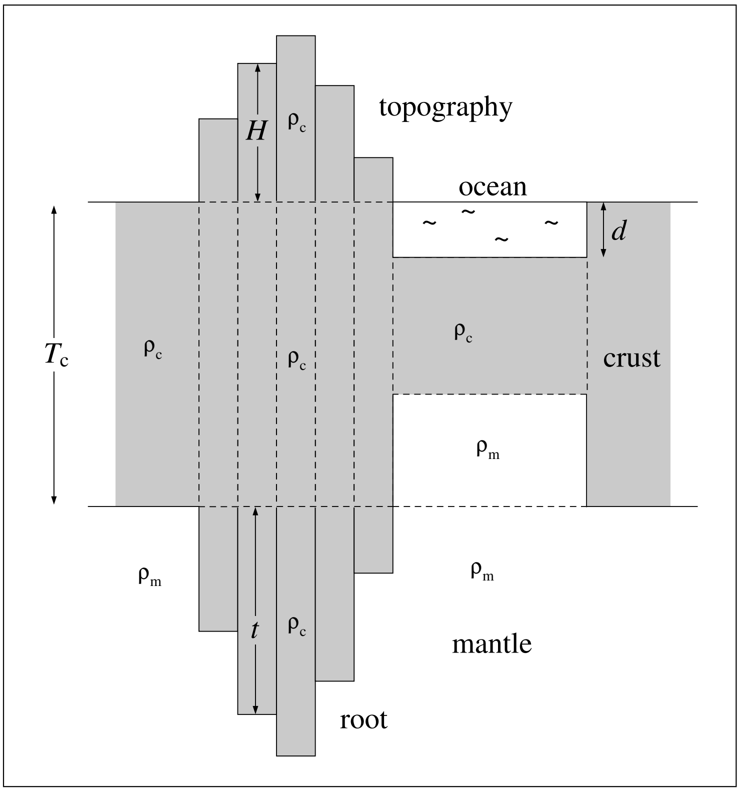


Figure 8.8: Airy-Heiskanen model: variable depth of compensation, the so-called Moho, and constant root density ρ_c.

The comparison with an iceberg is flawed, though. Between surface of the Earth and mantle—in absence of topography—there is a crust with a thickness of some 30 km. This crust separates mountain and root, but does not play a role in the isostasy condition.

A similar mechanism will take place underneath oceans. The lighter sea water will induce a negative root, i.e. a thinner crust below the oceans.

$$\text{land:} \quad (\rho_\mathrm{m} - \rho_\mathrm{c})t_i = \rho_\mathrm{c} H_i \tag{8.23a}$$

$$\text{ocean:} \quad (\rho_\mathrm{m} - \rho_\mathrm{c})t_i = (\rho_\mathrm{c} - \rho_\mathrm{w}) d_i \tag{8.23b}$$

In the Airy-Heiskanen model the root is linearly dependent on the topography:

$$\begin{aligned} \text{land:} \quad & t_i = \frac{\rho_\mathrm{c}}{\rho_\mathrm{m} - \rho_\mathrm{c}} H_i = 4.45 H_i \\ \text{ocean:} \quad & t_i = \frac{\rho_\mathrm{c} - \rho_\mathrm{w}}{\rho_\mathrm{m} - \rho_\mathrm{c}} d_i = 2.73 d_i \end{aligned}$$

The following densities were used: $\rho_c = 2670\,\mathrm{kg/m^3}$, $\rho_m = 3270\,\mathrm{kg/m^3}$, $\rho_w = 1030\,\mathrm{kg/m^3}$.

Now think of all columns with crustal material as the crust itself. It has variable thickness. Thicker over continents, especially over mountainous terrain, and thinner under oceans. The surface that separates crust from mantle is called *Moho*, after the geophysicist Mohorovičić[9]. It is a mirror of the topography.

Remark 8.3 *The deepest trenches have depths down to 10 km. This depth would imply that the ocean depth plus the anti-root would be larger than the crustal thickness itself. However, trenches are locations where oceanic plates move under continental plates and dip into the mantle. These regions are not in static equilibrium. The topographic and bathymetric features must be explained from dynamics.*

Regional or flexural isostasy. It is unrealistic to expect that a surface mass load like a mountain will only influence the column directly underneath. It will more likely bend the crust in a region around it. This is the idea of regional isostasy as developed by Vening-Meinesz[10]. Geophysically such isostatic models are more relevant. For the geodetic purpose of smoothing the gravity field it is not necessary to pursue this subject.

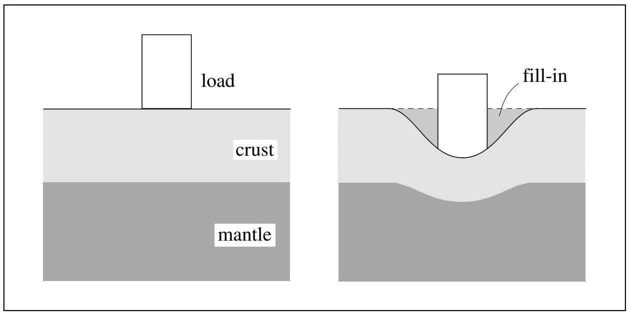


Figure 8.9: Vening-Meinesz model: regional compensation.

Isostatic gravity reductions. Using one—or a combination—of the isostatic models, we know the geometry and the density distribution of the compensating masses. Applying equations similar to those for the topographic corrections (8.21) we can correct for the gravity effect of isostasy. This will further smoothen the gravity field. For instance, we would:

- remove masses down to level of compensation, i.e. calculate the corresponding gravity effect,
- perform a free-air reduction down to the geoid,

[9] Andrija Mohorovičić (1857-1936), Croatian scientist in the field of seismology and meteorology.

[10] Felix Andries Vening-Meinesz (1887–1966), Dutch geophysicist and geodesist, developed a submarine gravimeter, explained low degree gravity field by convection cells in mantle.

- restore a homogeneous layer with, depending on the model, ρ_0 or ρ_m.

9. Geoid determination

This chapter is concerned with the actual application of the linear model (8.17) for the purpose of geoid computation. The so-called Stokes[1] approach will be followed, in which following assumptions are made:

- The geoid is the boundary. Consequently all surface data have to be reduced to the boundary first.
- Spherical approximation of the normal potential. This means we will set $U = {}^{GM}\!/_{r}$.
- Constant radius approximation. The coefficient matrix will not be evaluated on the ellipsoid, but on the sphere instead.

The resulting equations will be solved in the spectral domain (Fourier and SH) and in the spatial domain.

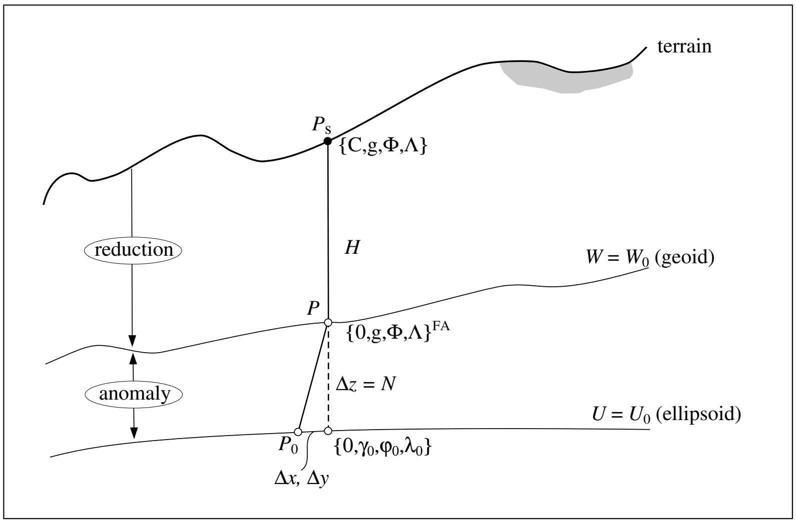


Figure 9.1: Stokes approach: the geoid is the boundary (set of all points P), the ellipsoid is the set of all approximate points P_0.

[1]George Gabriel Stokes (1819–1903), Irish mathematician, active in diverse areas as hydrodynamics, optics and geodesy; published on geoid determination in 1849.

9.1. The Stokes approach

Data reduction. If we assume the geoid to be our boundary, all data must be reduced to the geoid first. Reducing the geopotential numbers is trivial: they become zero. Reduction of the gravity can be done in several ways as outlined in 8.4. In the Stokes approach we will make use of the simplest option, namely free-air reductions. Astronomic latitude and longitude will have to be reduced, too. Since plumb lines are curved, ξ and η will change with height. Let us just assume that we have them at the geoid. Reducing a geopotential number C_{P_s} to the geoid will give $C_P = 0$ since the datum point and the reduced point are on the same equipotential surface, i.e. the geoid.

Equation (8.18) gives us free-air gravity. We need gravity anomalies, though. Thus we have to subtract normal gravity on the corresponding ellipsoid point.

$$\Delta g^{\mathrm{FA}} = g_P^{\mathrm{FA}} - \gamma_{P_0} \,. \tag{9.1}$$

Bouguer gravity anomalies, terrain corrected gravity anomalies and isostatically reduced gravity anomalies are defined in the same way.

The geoid is our boundary, consisting of all points P. The ellipsoid is the set of all approximate points P_0. They are separated by the vector $\Delta\boldsymbol{r}$. We can identify the vertical component Δz with the geoid height now:

$$\boxed{\Delta z = N} \,. \tag{9.2}$$

Spherical approximation. Geoid determination means that we invert the linear observation model (8.17) for at least Δz. Analytical inversion of the four by four coefficient—or design—matrix is very difficult, even if we have the analytical formulae for the normal potential and its 1st and 2nd derivatives. Our strategy will be to simplify the design matrix.

The first simplification will be *spherical approximation*, which means:

i) to assume the normal potential to be spherical: $U = {}^{GM}\!/_{r}$,

ii) to assume the the z-axis of the local coordinate system to be purely radial:

$$\frac{\partial}{\partial x} = \frac{1}{r}\frac{\partial}{\partial \phi}\,, \quad \frac{\partial}{\partial y} = \frac{1}{r\cos\phi}\frac{\partial}{\partial \lambda}\,, \quad \frac{\partial}{\partial z} = \frac{\partial}{\partial r}\,.$$

For the first and second derivatives of U we get:

- $U_x = U_y = 0\,, \quad U_z = -\dfrac{GM}{r^2} = -\gamma$
- $U_{xy} = U_{xz} = U_{yz} = 0$
- $U_{zz} = 2\dfrac{GM}{r^3} = 2\dfrac{\gamma}{r}\,, \quad U_{xx} = U_{yy} = -\dfrac{GM}{r^3} = -\dfrac{\gamma}{r}$

Although the derivation of U_{xx} and U_{yy} would be straightforward, it is easier to use the Laplace equation, knowing U_{zz}. Since there is no preferred direction in the horizontal plane U_{xx} and U_{yy} must be equal.

Constant radius approximation. The design matrix, although simplified, still has to be evaluated on the ellipsoid. The second simplification will therefore be to evaluate it on a sphere of radius R. This is the *constant radius approximation*: $r = |\boldsymbol{r}_0| = R$.

Applying both approximations now to the linear observation model (8.17) leads to:

$$\begin{pmatrix} \Delta W_0 \\ \Delta\Phi \\ \Delta\Lambda \\ \Delta g \end{pmatrix} = \begin{pmatrix} 0 & 0 & -\gamma \\ \dfrac{1}{R} & 0 & 0 \\ 0 & \dfrac{1}{R\cos\phi} & 0 \\ 0 & 0 & -\dfrac{2\gamma}{R} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} + \begin{pmatrix} T \\ -\dfrac{1}{R\gamma}T_\phi \\ -\dfrac{1}{R\gamma\cos^2\phi}T_\lambda \\ -T_r \end{pmatrix}. \tag{9.3}$$

For convenience the unknown ΔW_0 was moved to the vector of observables on the LHS. Although it is not an observable, we will treat it as such in the following discussion. As a result of the spherical and the constant radius approximation, the design matrix of (9.3) is far simpler than the one in (8.17).

Bruns's equation. We are now in a position to obtain the relation between the quantities ΔW_0 and Δg on the one hand and the geoid and T_r on the other hand. As a first step we write N instead of Δz. Solving N from the 1st row yields:

$$\boxed{N = \frac{T - \Delta W_0}{\gamma}}\,, \tag{9.4}$$

which is known as the *Bruns*[2] *equation*. Both N and T are unknowns at this point, but Bruns's equation says that if we know one we know the other (up to ΔW_0). Thus they

[2]Ernst Heinrich Bruns (1848–1919), German geodesist, astronomer and mathematician.

count as one unknown only. The equation is remarkable, since T is a function, evaluated on the ellipsoid whereas N is a geometrical description of a surface above the ellipsoid.

Fundamental equation of geodesy. Inserting (9.4) into the 4$^{\text{th}}$ row of (9.3) gives after some rearrangement:

$$\boxed{\Delta g - \frac{2}{R}\Delta W_0 = -\left(\frac{2}{R}T + \frac{\partial T}{\partial r}\right)} \,. \tag{9.5a}$$

This is equation is known as the *fundamental equation of physical geodesy*. It tells us how the observable at the LHS is related to the unknown disturbing potential T at the RHS by the differential operator $-(\frac{2}{R} + \frac{\partial}{\partial r})$. As such it represents the boundary function of our Boundary Value Problem. Since the boundary function contains both T and its radial derivative we can identify it as a 3$^{\text{rd}}$ BVP. If we can solve this Robin-type BVP, i.e. if we can express T as a function of gravity anomalies (and ΔW_0), we only have to substitute it back into Bruns's equation to obtain the geoid.

Flat Earth approximation. For regional geoid determination it may be sufficient to deal with the region under consideration in a planar (or flat Earth) approximation. This means that $R \to \infty$ and thus:

$$\Delta g = -\frac{\partial T}{\partial r} = \delta g \,. \tag{9.5b}$$

So in planar approximation the gravity anomaly and disturbance are equivalent. The flat Earth version of the fundamental equation represents a 2$^{\text{nd}}$ BVP or Neumann problem. Note that $\partial/\partial r$ in local Cartesian coordinates means $\partial/\partial z$.

9.2. Spectral domain solutions

Both Bruns's equation and the fundamental equation are more commonly known with the vertical datum unknown set to zero. We will take the following equations as the starting point for solutions in the spectral domain.

$$\boxed{\begin{array}{ll} \text{Bruns:} & N = \dfrac{T}{\gamma} \\ \text{fundamental/spherical:} & \Delta g = -\left(\dfrac{2}{R}T + \dfrac{\partial T}{\partial r}\right) \\ \text{fundamental/planar:} & \Delta g = -\dfrac{\partial T}{\partial z} \end{array}}$$

9.2.1. Local: Fourier

The solution of (9.5b) has been presented already in 6.1.1 for the general potential function Φ. It will be solved here once again in more detail for T as follows:

i) Write down the general Fourier series for T.

ii) Derive from that the expression for $\partial T/\partial z$ at zero height.

iii) Develop the given boundary function $\Delta g(x, y)$ in a Fourier series.

iv) Compare the Fourier spectra of steps *ii)* and *iii)*.

The latter comparison will provide the solution in the spectral domain.

$$i) \quad T(x,y,z) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} (a_{nm}\cos nx\cos my + b_{nm}\cos nx\sin my + c_{nm}\sin nx\cos my + d_{nm}\sin nx\sin my)\,\mathrm{e}^{-\sqrt{n^2+m^2}z}$$

$$ii) \quad -\left.\frac{\partial T}{\partial z}\right|_{z=0} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} (a_{nm}\cos nx\cos my + b_{nm}\cos nx\sin my + c_{nm}\sin nx\cos my + d_{nm}\sin nx\sin my)\,\sqrt{n^2+m^2}$$

$$iii) \quad \Delta g = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} (p_{nm}\cos nx\cos my + q_{nm}\cos nx\sin my + r_{nm}\sin nx\cos my + s_{nm}\sin nx\sin my)$$

$$iv) \quad : \quad \boxed{a_{nm} = \frac{p_{nm}}{\sqrt{n^2+m^2}} \quad , \quad b_{nm} = \text{ etc.}} \tag{9.6}$$

Now that we have solved for the spectral coefficients of T, the only thing that remains to be done is to apply Bruns's equation to obtain the geoid N. The spectral transfer $(\gamma\sqrt{n^2+m^2})^{-1}$ is a low-pass filter. The higher the wavenumbers n and m, the stronger the damping effect of the spectral transfer. Thus the geoid will be a smoother function than the gravity anomaly field.

Remark 9.1 *The transition from gravity anomaly to disturbing potential (9.6) is just a multiplication in the spectral domain. A corresponding convolution integral must exist therefore in the spatial domain.*

In reality the situation is a bit more complicated. The discrete Fourier transform in step *iii)* presupposes a periodic function Δg. The flat Earth approximation deals with local geoid calculations, though, in which case the data will not be periodic at all. Therefore we have to manipulate the data such that the error of the Fourier transform is minimized. This is usually done by:

- subtracting long-wavelength gravity anomalies given by a global geopotential model. After calculating a residual geoid from the residual gravity anomalies one has to add the corresponding long-wavelength geoid from the global geopotential model (*remove-restore technique*).
- letting the values at the edge of the data field smoothly go to zero by some window function (*tapering*).
- putting a ring of additional zeroes around the data (*zero-padding*).

9.2.2. Global: spherical harmonics

Inverting (9.5a), which is a Robin problem, has been treated more or less in 6.2 already. The solution in the spherical harmonic domain will be treated here in somewhat more detail. The following steps are involved:

i) Write down the general SH series for T, i.e. (8.3).

ii) Write down the SH series for $2T/R$ at zero height.

iii) Derive the expression for $\partial T/\partial r$ at zero height.

iv) Add up the previous two steps to obtain the SH series of $-(2T/R + \partial T/\partial r)$.

v) Develop the given boundary function $\Delta g(\theta, \lambda)$ in a SH series.

vi) Compare the spectra of steps *iv)* and *v)*.

The latter comparison gives us the required spectral solution. We exclude the zero-order term ${}^{\delta GM}/_{r}$ from our discussion.

i)
$$T(r,\theta,\lambda) = \frac{GM_0}{R}\sum_{l=2}^{\infty}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\left(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda\right)\left(\frac{R}{r}\right)^{l+1}$$

ii)
$$\frac{2}{R}T_{|r=R} = \frac{GM_0}{R^2}\sum_{l=2}^{\infty}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\,2\,\left(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda\right)$$

iii)
$$\frac{\partial T}{\partial r}_{|r=R} = \frac{GM_0}{R^2}\sum_{l=2}^{\infty}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\,(-l-1)\left(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda\right)$$

iv)
$$-\left(\frac{2}{R}T + \frac{\partial T}{\partial r}\right)_R = \frac{GM_0}{R^2}\sum_{l=2}^{\infty}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\,(l-1)\left(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda\right)$$

v)
$$\Delta g(\theta,\lambda) = \sum_{l=2}^{\infty}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\left(g^{\mathrm{c}}_{lm}\cos m\lambda + g^{\mathrm{s}}_{lm}\sin m\lambda\right)$$

vi)
$$\boxed{\Delta\bar{C}_{lm} = \frac{g^{\mathrm{c}}_{lm}}{\gamma(l-1)} \quad , \quad \Delta\bar{S}_{lm} = \frac{g^{\mathrm{s}}_{lm}}{\gamma(l-1)}} \,. \tag{9.7}$$

Since the solution in the spherical harmonic domain is of global nature there is no need for remove-restore procedures.

Degree 1 singularity. The spectral transfer $\frac{1}{l-1}$ is typical for the Stokes problem. It does mean, though, that we cannot solve the degree $l = 1$ contribution of the geoid from gravity data. The derivation above started the SH series at $l = 2$, so that the problem didn't arise. In general, though, we must assume a SH development of global gravity data starts at degree 0. The degree 0 coefficient $g^{\mathrm{c}}_{0,0}$, which is the global average is intertwined with the problems of ΔW_0 and δGM. In 6.4 the meaning of the degree 1 terms $C_{1,0}$, $C_{1,1}$ and $S_{1,1}$ was explained: they represent the coordinates of the Earth's center of mass. If we put the origin of our global coordinate system in this mass center, the degree 1 coefficients will become zero. In that respect there is no further need for determining them from gravity. Thus the $l = 1$ singularity of the Stokes problem shouldn't worry us.

9.3. Stokes integration

Similar to the Fourier domain solution (9.6), the SH solution (9.7) is a multiplication of the gravity spectrum with the spectral transfer of the boundary operator (9.5a). In the spatial domain, i.e. on the sphere, this corresponds to a convolution. The corresponding integral will be derived now by the following steps:

i) write down the SH coefficients explicitly using global spherical harmonic analysis,
ii) insert the spectral solution into a SH series of T,
iii) interchange integration and summation,
iv) apply the addition theorem of spherical harmonics, and
v) apply Bruns's equation to obtain an expression for the geoid.

The SH coefficients g^{c}_{lm} and g^{s}_{lm} are obtained from a spherical harmonic analysis of $\Delta g(\theta, \lambda)$, cf. (6.19). Combined with the spectral transfer from (9.7) we have:

i)
$$\left.\begin{matrix} \Delta\bar{C}_{lm} \\ \Delta\bar{S}_{lm} \end{matrix}\right\} = \frac{1}{\gamma(l-1)}\frac{1}{4\pi}\iint\limits_{\sigma} \Delta g(\theta,\lambda)\bar{P}_{lm}(\cos\theta) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \mathrm{d}\sigma \,.$$

This is inserted into the series development of the disturbing potential T. We need to distinguish between the evaluation point P and the data points Q, over which we integrate:

$$\textit{ii)} \qquad T_P = \frac{GM_0}{R\gamma}\sum_{l=2}^{\infty}\frac{1}{l-1}\sum_{m=0}^{l}\left[\left(\frac{1}{4\pi}\iint_\sigma \Delta g_Q\,\bar{P}_{lm}(\cos\theta_Q)\cos m\lambda_Q\,\mathrm{d}\sigma\right)\cos m\lambda_P \right.$$
$$\left. + \left(\frac{1}{4\pi}\iint_\sigma \Delta g_Q\,\bar{P}_{lm}(\cos\theta_Q)\sin m\lambda_Q\,\mathrm{d}\sigma\right)\sin m\lambda_P\right]\bar{P}_{lm}(\cos\theta_P)$$

The next step is to interchange summation and integration. Also, $\frac{GM_0}{R\gamma}$ is replaced by R.

$$\textit{iii)} \qquad T_P = \frac{R}{4\pi}\iint_\sigma \sum_{l=2}^{\infty}\frac{1}{l-1}\left[\sum_{m=0}^{l}(\cos m\lambda_P\cos m\lambda_Q \right.$$
$$\left. + \sin m\lambda_P\sin m\lambda_Q)\bar{P}_{lm}(\cos\theta_P)\bar{P}_{lm}(\cos\theta_Q)\right]\Delta g_Q\,\mathrm{d}\sigma\,.$$

Now compare the summation in square brackets to the addition theorem (6.22). We see that application of the addition theorem greatly reduces the length of the previous equation.

$$\textit{iv)} \qquad T_P = \frac{R}{4\pi}\iint_\sigma \underbrace{\sum_{l=2}^{\infty}\frac{2l+1}{l-1}P_l(\cos\psi_{PQ})}_{St(\psi_{PQ})}\Delta g_Q\,\mathrm{d}\sigma$$
$$= \frac{R}{4\pi}\iint_\sigma St(\psi_{PQ})\Delta g_Q\,\mathrm{d}\sigma\,.$$

Applying Bruns's equation finally gives:

$$\textit{v)} \qquad \boxed{N_P = \frac{R}{4\pi\gamma}\int_{\lambda=0}^{2\pi}\int_{\theta=0}^{\pi} St(\psi_{PQ})\Delta g_Q\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\lambda}\,. \tag{9.8}$$

This is the Stokes integral. It is the solution—or the inverse—of the fundamental equation of physical geodesy in the spatial domain. The fundamental equation applies a differential operator to T to produce Δg. Equation (9.8) is an integral operator, acting on Δg, that produces T (or N).

fundamental equation	Stokes integral
$T \to \Delta g$	$\Delta g \to T$
$\Delta g = -\left(\frac{2}{R} + \frac{\partial}{\partial r}\right) T$	$T_P = \frac{R}{4\pi} \iint_\sigma St(\psi_{PQ}) \Delta g_Q \,\mathrm{d}\sigma$

We can interpret the Stokes integral in different ways. The Stokes function is most easily understood as a weight function. To calculate the geoid height in P we have to

- consider all gravity anomalies over the Earth,
- calculate the spherical distance ψ_{PQ} using (6.23),
- multiply the Δg_Q with the Stokes function, and
- integrate all of this over the sphere.

The left panel of fig. 9.3 demonstrates the integration principle. In terms of signal processing theory the Stokes integral is a convolution on the sphere. In order to get the output signal N, we have to convolve the input signal Δg with the convolution kernel $St(\psi)$. Compare this to the continuous 1D convolution:

$$f(x) = \int h(y - x) g(y) \,\mathrm{d}y \,,$$

in which the convolution kernel h also depends on the distance between x and y only.

The Stokes function. From step *iv)* we see the definition of the Stokes function in terms of a Legendre polynomial series. Stokes himself found a closed analytical expression.

$$\text{series:} \quad St(\psi) = \sum_{l=2}^{\infty} \frac{2l+1}{l-1} P_l(\cos\psi) \tag{9.9a}$$

$$\text{closed:} \quad St(\psi) = \frac{1}{s} - 6s + 1 - 5\cos\psi - 3\cos\psi \ln(s + s^2) \tag{9.9b}$$

$$s = \sin \tfrac{1}{2}\psi$$

Figure 9.2 displays the Stokes function over its domain $\psi \in [0; \pi]$ together with successive approximations, in which the series expression was cut off at degrees 2, 3, 4 and 100. Note that $St(\psi = 0) = \infty$. This is due to the fact that we sum up $P_l(1) = 1$ from degree 2 to infinity. In the analytical formula this becomes clear from the $1/s$-term. The impossibility of having infinite weight for a gravity anomaly at the computation point will be discussed in the next section.

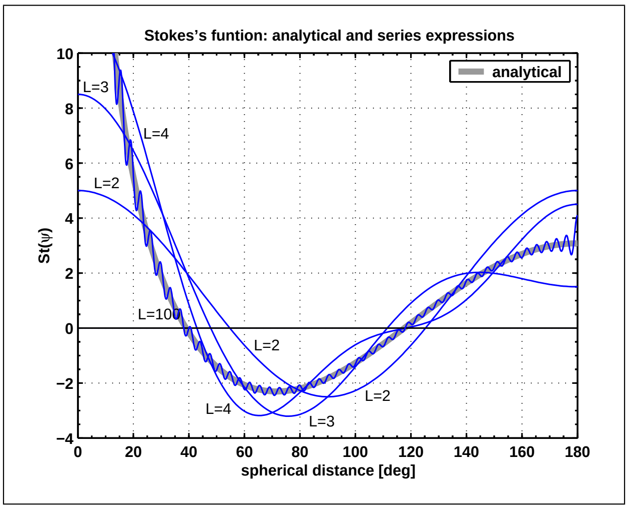


Figure 9.2: The Stokes function $St(\psi)$ and approximations with increasing maximum degree L.

Remark 9.2 *Except for two zero-crossings, the Stokes functions does not vanish. The Stokes integration, as defined in this section, therefore requires global data coverage and integration over the whole sphere. In the next section, though, the Stokes integral will be used for regional geoid determinations as well.*

9.4. Practical aspects of geoid calculation

9.4.1. Discretization

Gravity data is given in discrete points. If they are given on a grid, the data usually represents averages over grid cells. Suppose we have data $\Delta\bar{g}_{ij} = \Delta g(\theta_i, \lambda_j)$ on an equi-angular grid with block size $\Delta\theta \times \Delta\lambda$. The overbar denotes a gravity block average. The Stokes integral is discretized into:

$$N_P = \frac{R}{4\pi\gamma} \sum_i \sum_j St(\psi_{PQ_{ij}}) \Delta\bar{g}_{Q_{ij}} \sin\theta_i \Delta\theta\Delta\lambda\,. \tag{9.10}$$

The spherical distance (6.23) depends on longitude difference, but because of meridian convergence not on co-latitude difference. Equation (9.10) is strictly speaking only a convolution in longitude direction, which may be evaluated by FFT. The latitude summation must be evaluated by straightforward numerical integration. However, for regional geoid computations, in which the latitude differences do not become too large, a 2D convolution can be applied without significant loss of accuracy.

Exercise 9.1 *Calculate roughly the contribution of the Alps to the geoid in Calgary.*

Assume that the Alps are a square block with λ running from $7°$–$13°$ and ϕ from $45°$–$47°$ with an average gravity anomaly of 50 mGal.

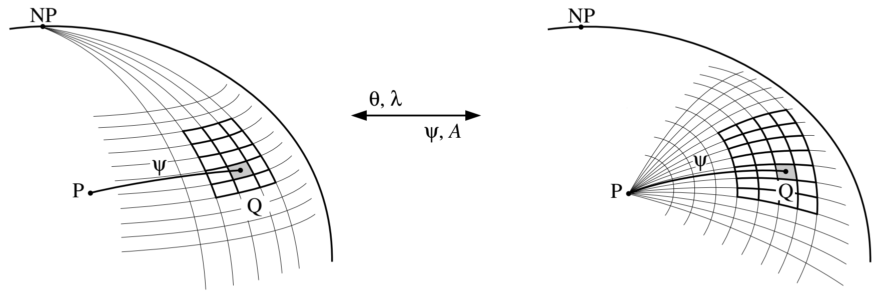


Figure 9.3.: Stokes integration on a θ, λ-grid (left) and on a ψ, A-grid (right).

9.4.2. Singularity at $\psi = 0$

In case the evaluation point P and the data point Q coincide, we would have to assign infinite weight to Δg_Q. The geoid cannot be infinite, though. The solution of this paradox comes from a change to polar coordinates $(\theta, \lambda) \rightarrow (\psi, A)$.

$$N_P = \frac{R}{4\pi\gamma} \int\limits_{\psi=0}^{\pi} \int\limits_{A=0}^{2\pi} St(\psi_{PQ}) \Delta g_Q \sin\psi \, \mathrm{d}\psi \, \mathrm{d}A \,. \tag{9.11}$$

The form of this equation is not much different from (9.8). Seen from the North pole, θ and λ are polar coordinates too. Since the Stokes function only depends on spherical distance, we can integrate over the azimuth:

$$\begin{aligned} N_P &= \frac{R}{2\gamma} \int\limits_{\psi=0}^{\pi} St(\psi) \left[\frac{1}{2\pi} \int\limits_{A=0}^{2\pi} \Delta g_Q \, \mathrm{d}A \right] \sin\psi \, \mathrm{d}\psi \\ &= \frac{R}{2\gamma} \int\limits_{\psi=0}^{\pi} St(\psi) \, \overline{\Delta g} \sin\psi \, \mathrm{d}\psi \\ &= \frac{R}{\gamma} \int\limits_{\psi=0}^{\pi} F(\psi) \, \overline{\Delta g} \, \mathrm{d}\psi \,. \end{aligned} \tag{9.12}$$

The right panel of fig. 9.3 demonstrates the integration principle using a polar grid. The quantity $\overline{\Delta g}$ is a ring integral. It is the average over all gravity values, which are a given distance ψ away from P. The function

$$F(\psi) = \frac{1}{2} St(\psi) \sin\psi$$

is presented in fig. 9.4. It is also a weight function, but now for the ring averages $\overline{\Delta g}$. Because of the multiplication with $\sin\psi$ the singularity at zero distance has become finite. This is explained by considering the area of the integrated rings for decreasing spherical distance:

$$\text{infinitesimal ring area} = \pi(\psi + \mathrm{d}\psi)^2 - \pi\psi^2 \approx 2\pi\psi\,\mathrm{d}\psi\,.$$

So for the smallest ring, which is a point, the area becomes zero, thus compensating the singularity of $St(0)$.

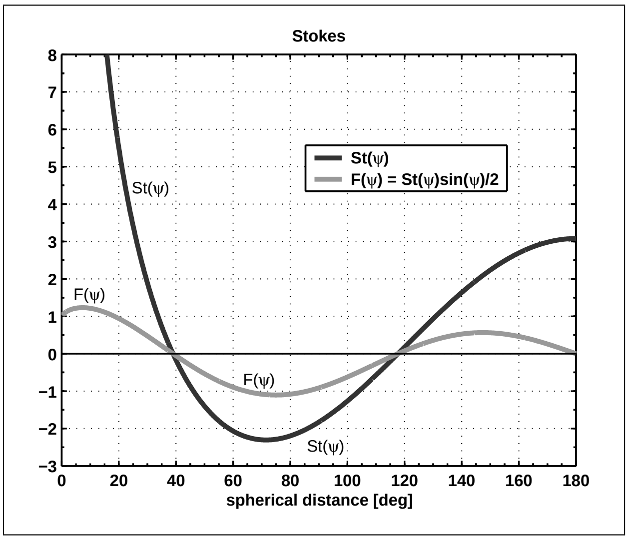


Figure 9.4: The Stokes function $St(\psi)$ together with $F(\psi)$.

Polar grid. Suppose we have a polar grid with cell sizes of $\Delta\psi$ by ΔA around our evaluation point P. This is not very practicable, since for each new P we would need to change our data grid accordingly. Let us concentrate on the innermost cell, which is a small circular cell around the evaluation point. Its geoid contribution is:

$$\delta N = \frac{R}{\gamma} \int_{\psi=0}^{\Delta\psi} F(\psi)\,\overline{\Delta g}\,\mathrm{d}\psi \approx \frac{R}{\gamma}\,\overline{\Delta g}\,\Delta\psi\,.$$

The latter step is allowed if we assume $F(\psi) = 1$ over the whole innermost cell. In most practical situations we have square grids of size $\Delta\theta \times \Delta\lambda$ or $\Delta x \times \Delta y$. So we must first find the radius $\Delta\psi$ that gives the same cell area:

$$\pi \Delta\psi^2 = \sin\theta \Delta\theta \Delta\lambda = \frac{\Delta x}{R} \frac{\Delta y}{R} .$$

So the geoid contribution of a square grid point to the geoid in the same point becomes:

$$\delta N = \frac{1}{\gamma} \Delta\bar{g} \sqrt{\frac{\Delta x \Delta y}{\pi}} = \frac{R}{\gamma} \Delta\bar{g} \sqrt{\frac{\sin\theta \Delta\theta \Delta\lambda}{\pi}} . \tag{9.13}$$

Exercise 9.2 *What is the geoid contribution of a 50 mGal gravity anomaly, representing the average over a $3' \times 5'$ block at our latitude? Use the cell's mid-point as the evaluation point.*

Remark 9.3 *In the pre-computer era the polar grid method was practised for graphical calculation of geoid heights. It was known as the* template method*. Circular templates were put on a gravity map. For each sector of the template an average gravity value was estimated visually and multiplied by a tabulated $F(\psi)$ for that sector. Summed over all sectors one obtained the geoid value for point P. For the next evaluation point the template was moved over the map and the whole exercise started again.*

9.4.3. Combination method

Except for two zero crossings, the Stokes function has non-zero values over the full sphere. In principle Stokes integration is a global process. We see in fig. 9.2, though, that $St(\psi)$ gives reasonably large values only over the first 20° or 30°. This indicates that it may be sufficient to integrate gravity data over an area of smaller spherical radius only. To reduce truncation errors we apply a remove-restore procedure again, see also fig. 9.5. The basic formula for this *combination method* reads:

$$\boxed{N = N_0 + N_1 + N_2 + \epsilon_N} , \tag{9.14}$$

with:
N_0 = constant, zero-order term
N_1 = contribution from global geopotential model
N_2 = Stokes integration over limited domain
ϵ_N = truncation error

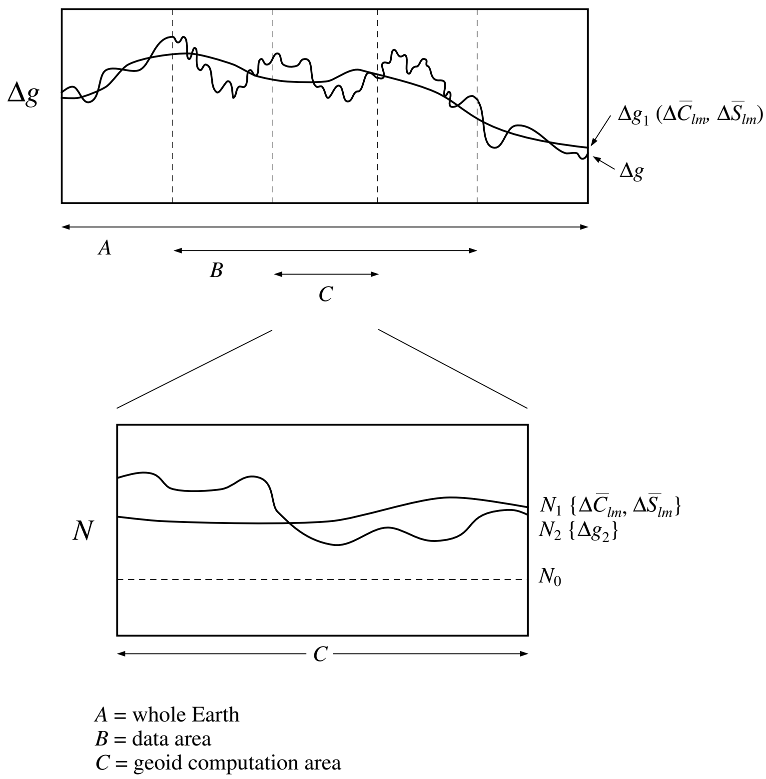


Figure 9.5.: Regional geoid computation with remove-restore technique.

The zero-order term has been neglected somewhat in the previous sections. It is related to the unknown δGM, the height datum unknown ΔW_0 and the global average gravity anomaly $g^{\mathrm{c}}_{0,0} = \frac{1}{4\pi}\iint \Delta g \,\mathrm{d}\sigma$. Only two of these three quantities are independent. Without derivation we define the constant N_0 as:

$$N_0 = -\frac{R}{\gamma} g^{\mathrm{c}}_{0,0} + \frac{\Delta W_0}{\gamma} \overset{\text{or}}{=} \frac{1}{\gamma}\left(\frac{\delta GM}{R} - \Delta W_0\right) . \tag{9.14a}$$

The geoid contribution from the global geopotential model is calculated by a spherical harmonic synthesis up to the maximum degree L of the model:

$$N_1 = R \sum_{l=2}^{L} \sum_{m=0}^{l} \bar{P}_{lm}(\cos\theta) \left(\Delta\bar{C}_{lm} \cos m\lambda + \Delta\bar{S}_{lm} \sin m\lambda\right) . \tag{9.14b}$$

For the Stokes integration we first have to remove the part from the gravity anomalies that corresponds to N_1. Otherwise that part would be accounted for twice. We need the following steps:

$$
\begin{aligned}
\Delta g_1 &= \gamma \sum_{l=2}^{L} \sum_{m=0}^{l} \bar{P}_{lm}(\cos\theta)\,(l-1)\,\left(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda\right) \,, \\
\Delta g_2 &= \Delta g - \Delta g_1 \,, \\
N_2 &= \frac{R}{4\pi\gamma} \iint_{\sigma_0} St(\psi)\Delta g_2 \,\mathrm{d}\sigma \,. \qquad (9.14c)
\end{aligned}
$$

Note that the domain of integration is σ_0 which is a limited portion of the sphere only.

The truncation error ϵ_N depends on the maximum degree of the global geopotential model and the size of the domain of integration. We have:

$$
\epsilon_N = \frac{R}{4\pi\gamma} \iint_{\sigma/\sigma_0} St(\psi)\Delta g_2 \,\mathrm{d}\sigma \,, \qquad (9.14d)
$$

in which σ/σ_0 denotes the surface of the sphere outside the integration area. The higher the maximum degree L gets, the better the global model represents Δg. As a result Δg_2 and consequently the truncation error are reduced by increasing L. Also, the larger the domain of integration σ_0 becomes, the smaller ϵ_N will be.

Truncation effects usually show up at the edges of the geoid calculation area. As a further countermeasure to suppress ϵ_N one usually uses an evaluation area smaller than the integration area σ_0 in which gravity data is available.

9.4.4. Indirect effects

In 8.4 the reduction of gravity from surface point P to geoid point P_0 was discussed. Errors—less or more severe—are introduced in almost every step:

- $\dfrac{\partial g}{\partial h} \approx \dfrac{\partial g}{\partial r} \approx \dfrac{\partial \gamma}{\partial r}$ (for FA).
- a fixed crustal density of 2670 kg/m³ (for BO).
- a limited-resolution digital terrain model (for TC).
- heights H in a given height system, which may or may not refer to the geoid (datum unknown).
- type of isostatic compensation.

But even if gravity reductions could be done perfectly an error would be made for a simple reason. Removing topographic masses implies that we change the potential field, too. The equipotential surface with potential W_0 (the geoid) has changed its shape. The new surface with potential W_0 is the *co-geoid*. Going from geoid to co-geoid is called the *primary indirect effect*. Now the reduced gravity field does not correspond to the new mass configuration anymore: we have to reduce further to the co-geoid. This is called the *secondary indirect effect*.

So in the Stokes approach we will not determine geoid heights N, but co-geoid heights N_c instead. In order to correct for the indirect effects we use the following recipe:

i) Direct effect: perform the usual gravity reductions.

ii) Primary indirect effect: calculate the change in potential δW_{top}, due to removing the topography.

iii) Calculate the separation between geoid and co-geoid: $\delta N = \delta W_{\mathrm{top}}/\gamma$.

iv) Secondary indirect effect: reduce gravity further using $\delta g_{\mathrm{top}} = 2\frac{\gamma}{r}\delta N$ (FA-type).

v) Calculate the co-geoid height N_c using any of the methods described in this chapter.

vi) Correct for the indirect effect: $N = N_c + \delta N$.

A. Reference Textbooks

For these lecture notes, extensive use has been made of existing reference material by Rummel, Schwarz, Torge, Wahr and others. The following list of reference works, which is certainly non-exhaustive, is recommended for study along with these lecture notes.

References

Blakely RJ (1995). *Potential Theory in Gravity & Magnetic Applications*, Cambridge University Press

Heiskanen W & H Moritz (1967). *Physical Geodesy*, WH Freeman and Co., San Francisco (reprinted by TU Graz)

Hofmann-Wellenhof B & H Moritz (2005). *Physical Geodesy*, Springer Verlag

Rummel R (2004). *Erdmessung, Teil 1,2&3*, Institut für Astronomische und Physikalische Geodäsie, TU Munich

Schwarz K-P (1999). *Fundamentals of Geodesy*, UCGE Report No. 10014, Department of Geomatics Engineering, University of Calgary

Torge W (1989). *Gravimetry*, De Gruyter

Torge W (2001). *Geodesy*, 3rd edition, De Gruyter

Vaníček P & EJ Krakiwsky (1982,1986). *Geodesy: The Concepts*, Elsevier

Wahr J (1996). *Geodesy and Gravity*, Samizdat Press, `samizdat.mines.edu`

B. The Greek alphabet

α	A	alpha
β	B	beta
γ	Γ	gamma
δ	Δ	delta
ϵ, ε	E	epsilon
ζ	Z	zeta
η	H	eta
θ, ϑ	Θ	theta
ι	I	iota
κ	K	kappa
λ	Λ	lambda
μ	M	mu
ν	N	nu
ξ	Ξ	ksi
o	O	o*micron*
π, ϖ	Π	pi
ρ, ϱ	P	rho
σ, ς	Σ	sigma
τ	T	tau
υ	Υ	upsilon
ϕ, φ	Φ	phi
χ	X	chi
ψ	Ψ	psi
ω	Ω	o*mega*