Physical Geodesy

$\mathbf{g}$

N

Mass excess

Mass deficit

$-N$

Martin Vermeer[1]

June 21, 2019

[1]martin.vermeer@aalto.fi

Preface

This book aims to present an overview over the current state of the study of the Earth's gravity field and those parts of geophysics closely related to it, such as especially geodynamics, the study of the changing Earth. It grew out of two decades of teaching in Helsinki's two universities: Helsinki University of Technology — today absorbed into Aalto University — and the University of Helsinki. As such, it presents a somewhat Fennoscandian perspective on a very global subject. Also the author's own research, on gravimetric geoid determination, helped shape the presentation. While there exist excellent textbooks on all parts of what is presented here, he may still hope that this text will find a niche to fill.

Helsinki, June 21, 2019,

Martin Vermeer

Acknowledgements

Thanks are due to the many students and colleagues, both in academia and at the Finnish Geodetic Institute, giving useful comments and corrections over the course of many years of lecturing this course both at the University of Helsinki and at the Helsinki University of Technology, today Aalto University.

Special thanks are due to the foreign students at Aalto University, who forced me during recent years to provide an English version of this text. The translation work prompted a basic revision also to the Finnish text, which was long overdue as parts were written in the 1990s before the author had had the benefit of pedagogical training. Thanks are thus also due to the Aalto University's pedagogical training programme.

Contents

Chapters

Contents

Contents

4.   The normal gravity field   81

5.   Anomalous quantities of the gravity field   103

6.   Geophysical reductions   121

11. Gravimetric measurement devices   275

12. The geoid, mean sea level, sea-surface topography   301

13. Satellite altimetry and satellite gravity missions   319

14. Tides, atmosphere and Earth crustal movements   351

15. Earth gravity field research   365

A. Field theory and vector calculus — core knowledge   369

B. Function spaces   385

C. Why does FFT work?   399

D. Helmert condensation   403

E. The Laplace equation in spherical co-ordinates   411

Bibliography   417

Index   433

List of Tables

List of Figures

List of Figures

List of Figures

Acronyms

Acronyms

Fundamentals of the theory of gravitation

1

1.1 General

In this chapter we discuss the foundations of Newton's theory of gravitation. Intuitively, the theory of gravitation is easiest to understand as "action at a distance", latin *actio ad distans*, where the force between two masses is proportional to the masses themselves, and inversely proportional to the square of the distance between them. This is the form of Newton's general law of gravitation familiar to all.

There exists an alternative but equivalent presentation, *field theory*, which describes gravitation as a phenomenon propagating through space, a *field*. The propagation is described in the *field equations*. The field approach isn't quite as intuitive, but is a powerful theoretical tool[1].

In this chapter we acquaint ourselves with the central concept of field theory, the *gravitational potential*. We also get to know the potential fields of the theoretically interesting single and double *mass density layers*. Of the practical

---

[1]There is also a philosophical side to this. To many, e.g., to Leibniz, the idea of a force that jumps from object to object was an abomination. Many tried to explain gravitation — and also electromagnetism etc. — by a "world aether". It wasn't until the advent of relativity theory, that the understanding gained ground that a physical theory doesn't have to satisfy our preconceptions about what constitutes a "sensible explanation" — as long as it describes the physical phenomena correctly.

and theoretical applications of these may be mentioned the Bouguer plate and Helmert condensation. In the following we will discuss their properties in detail. Mass density layers are also used in deriving the theorems of Green. We will learn about important *integral theorems* like the theorems of Gauss and Green, with the aid of which we may infer the whole potential field in space from field values given only on a certain surface. Other similar examples are the Chasles theorem and the solution to Dirichlet's problem.

In chapter 3 we apply these fundamentals of potential theory to derive a spectral representation of the Earth's gravitational field, a so-called *spherical-harmonic expansion*.

Here, in the beginning of the text, we derive an extensive set of mathematical equations, such as well-known integral equations. This is an unfortunately necessary preliminary work. These equations, however, are no end in themselves and they are not worth committing to memory. Try rather to understand their logic, and how historically these various results have been arrived at. Try as well to acquire an intuition, a fingertip feeling, about the nature of this theory.

1.2   Gravitation between two masses

We start the investigation of the Earth's gravity field suitably with Isaac Newton[2]'s general law of gravitation:

$$F = G\frac{m_1 m_2}{\ell^2}. \tag{1.1}$$

F is the attractional force between bodies 1 and 2, m_1 and m_2 are the masses of the bodies, and ℓ is the distance between them. We assume the masses to be points. The constant G, the *universal gravitational constant*, has the value

$$G = 6.6726 \cdot 10^{-11} \mathrm{m}^3 \mathrm{kg}^{-1} \mathrm{s}^{-2}.$$

---

[2]Sir Isaac Newton PRS (1642 – 1727) was an English universal genius who derived the mathematical underpinning of astronomy, and much of geophysics, in his main work "*Philosophiæ Naturalis Principia Mathematica*", "Mathematical Foundations of Physics".
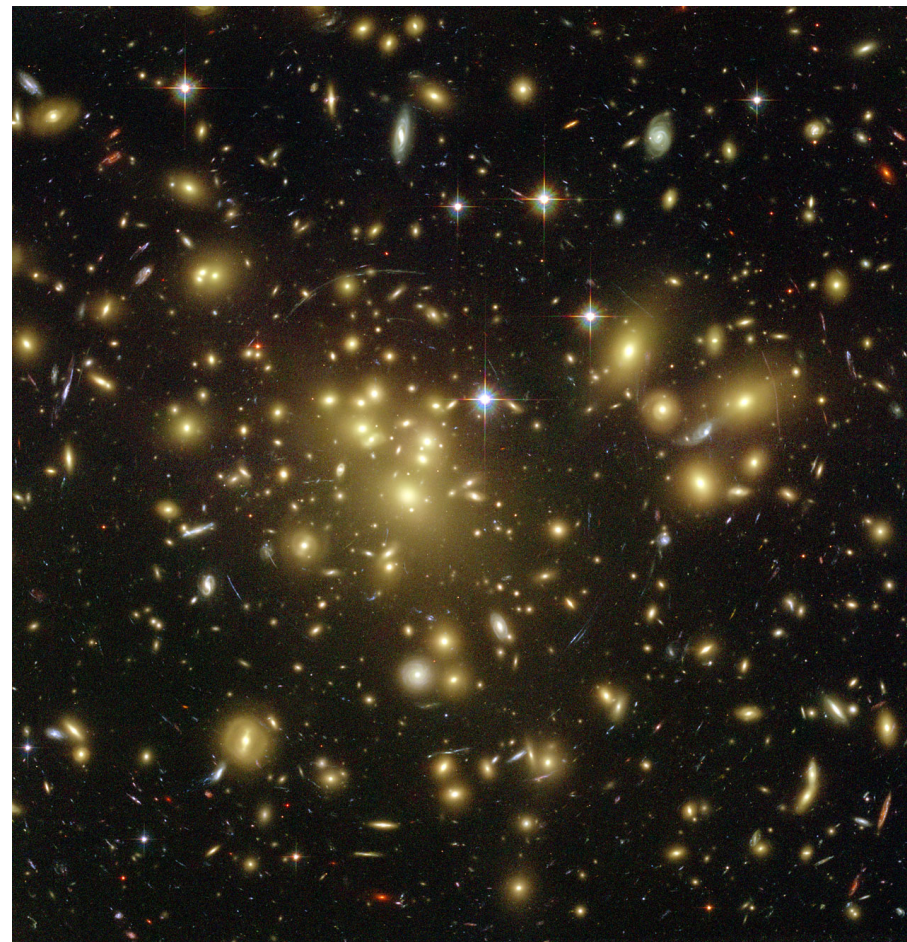
Figure 1.1. Gravitation is universal. A gravitational lens imaged by the Hubble Space Telescope, the cluster of galaxies Abell 1689 at a distance of 2.2 billion light years. Benitez et al. (2003).
Credits: NASA, N. Benitez (JHU), T. Broadhurst (The Hebrew University), H. Ford (JHU), M. Clampin (STSCI), G. Hartig (STSCI), G. Illingworth (UCO/Lick Observatory), the ACS Science Team and ESA.

kiertoheiluri

The value of G was determined for the first time by Henry Cavendish[3] using a sensitive torsion balance (Cavendish, 1798).

Let us call the small body, the test mass, e.g., a satellite, m, and the large mass, the planet or the Sun, M. Then, $m_1 = M$ may be called the *attracting* mass, and $m_2 = m$ the *attracted* mass, and we obtain

$$F = G\frac{mM}{\ell^2}.$$

According to Newton's law of motion

$$F = ma,$$

where a is the gravitational acceleration of body m. From this follows

$$a = G\frac{M}{\ell^2}.$$

From this equation, the quantity $m = m_2$ has vanished. This is the famous observation by Galileo, that *all bodies fall equally fast*[4], irrespective of their mass. This is also known as Einstein[5]'s principle of equivalence.

Both the force F and the acceleration a have the same direction as the line connecting the bodies. For this reason one often writes equation 1.1 as a *vector equation*, which is more expressive:

$$\mathbf{a} = -GM\frac{\mathbf{r} - \mathbf{R}}{\ell^3}, \tag{1.2}$$

---

[3]Henry Cavendish FRS (1731–1810) was a British natural scientist from a wealthy nobility background. He did also pioneering work in chemistry. He was extremely shy, and the renowned neurologist Oliver Sacks retrodiagnosed him as afflicted with the Asperger syndrome (Sacks, 2001).

[4]At least in vacuum. The Apollo astronauts showed impressively how on the Moon a feather and a hammer fall equally fast! YouTube, Hammer vs. Feather.

[5]Albert Einstein (1879–1955) was a theoretical physicist of Jewish German descent, who created both the special and general theories of relativity, applying the latter to cosmology, and did fundamental work in quantum theory.

where the three-dimensional vectors of place of both the attracting and attracted masses are defined as follows in rectangular co-ordinates[6]:

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$
$$\mathbf{R} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k},$$

where the triad of unit vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is an *orthonormal bas*is[7] in Euclidean space, $\mathbb{R}^3$ and

$$\ell = \|\mathbf{r} - \mathbf{R}\| = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2} \tag{1.3}$$

is the distance between the masses computed by the Pythagoras theorem.

Note that the vector equation 1.2 contains a minus sign! This only tells us, that the direction of the force is opposite to that of the vector $\mathbf{r} - \mathbf{R}$. This vector is the location of the attracted mass m reckoned from the location of the attracting mass M. In other words, this tells us that we are dealing with an *attraction* and not a repulsion.

1.3 The potential of a point mass

The gravitational field is a special field: if it is stationary, i.e., not time dependent, it is *conservative*. This means that a body moving inside the field along a closed path will, at the end of the journey, not have lost or gained energy. Because of this, one may attach to each point in the field a "label" onto which is written the amount of energy gained or lost by a unit or test mass, when travelling from an agreed-upon starting point to the point under discussion. The value written on the label is called the *potential*. (Note, that the choice of starting point is arbitrary! We will return to this still.)

---

[6]As vector notation, one may use either $\vec{v}$ (an arrow above) or $\mathbf{v}$ (bold). Here we use the bold notation, except for vectors designated by Greek letters, which cannot be bolded.

[7]This means that $\|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1$ and $\langle \mathbf{i} \cdot \mathbf{j} \rangle = \langle \mathbf{i} \cdot \mathbf{k} \rangle = \langle \mathbf{j} \cdot \mathbf{k} \rangle = 0$, where the norm is defined as $\|\mathbf{a}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{a} \cdot \mathbf{a} \rangle}$, and $\langle \mathbf{a} \cdot \mathbf{b} \rangle$, $\mathbf{a}, \mathbf{b} \in \mathbb{E}$ is the *inner* or *scalar product* of the space.

The *potential function* defined in this way for a pointlike body M is:

$$V = G\frac{M}{\ell} = \frac{GM}{\ell},\tag{1.4}$$

where ℓ is again, like above, the length of the vector $\mathbf{r} - \mathbf{R}$, $\ell = \|\mathbf{r} - \mathbf{R}\|$.

The constant GM has in case of the Earth (according to the GRS80 system, conventionally) the value

$$GM_\oplus = 3.986\,005 \cdot 10^{14}\,\mathrm{m^3/s^2}.$$

The currently best available physical value again is

$$GM_\oplus = 3.986\,004\,40 \cdot 10^{14}\,\mathrm{m^3/s^2}.$$

1.4 Potential of a spherical shell

We may write, based on equation 1.4, the potential of an extended body M into the following form:

$$V(\mathbf{r}) = G\int_M \frac{dm(\mathbf{R})}{\ell} = G\int_M \frac{dm(\mathbf{R})}{\|\mathbf{r} - \mathbf{R}\|}.\tag{1.5}$$

This is an integral over mass elements dm, where every such mass element is located at place $\mathbf{R}$. The potential V is evaluated at location $\mathbf{r}$, and the distance $\ell = \|\mathbf{r} - \mathbf{R}\|$.

We now derive the formula for the potential of a thin spherical shell, see figure 1.2, where we have placed the centre of the sphere at the origin O.

Because the circumference of a narrow ring, width $b \cdot d\theta$, is $2\pi b \sin\theta$, its surface area is

$$(2\pi b \sin\theta)\,(b \cdot d\theta).$$

Let the thickness of the shell be p (small) and its density ρ. We obtain for the total mass of the ring

$$2\pi p\rho b^2 \sin\theta\, d\theta.$$

Because every point of the ring is at the same distance ℓ from point P, we may write for the potential at point P:

$$V_P = \frac{2\pi G p\rho b^2 \sin\theta\, d\theta}{\ell}.$$
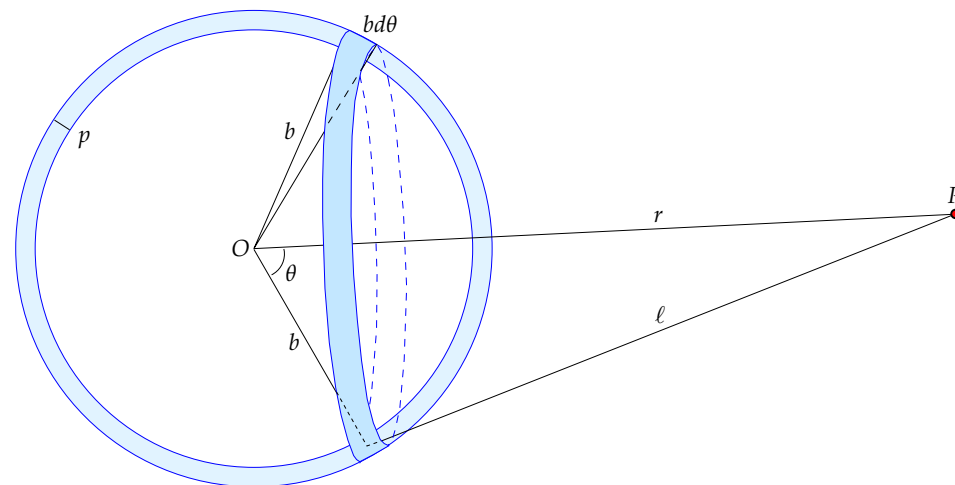
Figure 1.2. A thin spherical shell consists of rings.

With the cosine rule,

$$\ell^2 = r^2 + b^2 - 2rb\cos\theta,\tag{1.6}$$

we obtain, using equation 1.5, for the potential of the whole shell

$$V_P = 2\pi G\rho pb^2 \int \frac{\sin\theta\, d\theta}{\sqrt{r^2 + b^2 - 2rb\cos\theta}}.$$

In order to evaluate this integral, we must replace the integration variable θ by ℓ. Differentiating equation 1.6 yields

$$\ell\, d\ell = br \sin\theta\, d\theta,$$

and remembering that $\ell = \sqrt{r^2 + b^2 - 2rb\cos\theta}$ we obtain

$$V_P = 2\pi G\rho pb^2 \int_{\ell_1}^{\ell_2} \frac{d\ell}{br}.$$

In the case that point P is outside the shell, the integration bounds of ℓ are $\ell_1 = r - b$ and $\ell_2 = r + b$, and we obtain for the potential of point P

$$V_P = 2\pi G\rho pb^2 \left[\frac{\ell}{br}\right]_{\ell=r-b}^{\ell=r+b} = \frac{4\pi G\rho pb^2}{r}.$$

Because the mass of the whole shell is $M_b = 4\pi b^2 \rho p$, it follows that the potential of the shell is *the same as that of an equal sized mass in its centre O*:

$$V_P = \frac{GM_b}{r},$$

where r is now the distance of computation point P from the centre of the sphere O. We see that this is the same as equation 1.4.

In the same way, the attraction, or rather, *acceleration*, caused by the spherical shell is[8]

$$\mathbf{a}_P = \left(\vec{\nabla}V\right)_P = -4\pi G\rho pb^2 \frac{\mathbf{r}_P - \mathbf{r}_O}{r^3} = -GM_b \frac{\mathbf{r}_P - \mathbf{r}_O}{r^3},$$

in which $r = \|\mathbf{r}_P - \mathbf{r}_O\|$. This result is identical with the acceleration caused by an equal sized point mass located in point O, see equation 1.2.

In the case that point P is inside the shell, $\ell_1 = b - r$ and $\ell_2 = b + r$ and the above integral changes to the following:

$$V_P = 2\pi G\rho pb^2 \left[\frac{\ell}{br}\right]_{\ell=b-r}^{\ell=b+r} = 4\pi G\rho pb.$$

As we see, this is a *constant* and not dependent upon the location of point P. Therefore $\nabla V_P = 0$ and the attraction, being the gradient of the potential, vanishes.

The end result is, that the attraction of a spherical shell is, outside the shell,

$$a = \|\mathbf{a}\| = \frac{GM}{r^2},$$

where M is the total mass of the shell and $r = \|\mathbf{r}_P - \mathbf{r}_O\|$ the distance of the observation point from the shell's centre; and 0 inside the shell.

In figure 1.3 we have drawn the curves of potential and attraction (i.e., acceleration, attractive force per unit of mass). If a body consists of many concentric spherical shells (like rather precisely the Earth and many celestial bodies), then inside the body only those layers of mass internal to the observation point participate in causing attraction, and this attraction is the same as it

---

[8]Here, the ∇ (nabla) operator is used, to be explained in section 1.5.
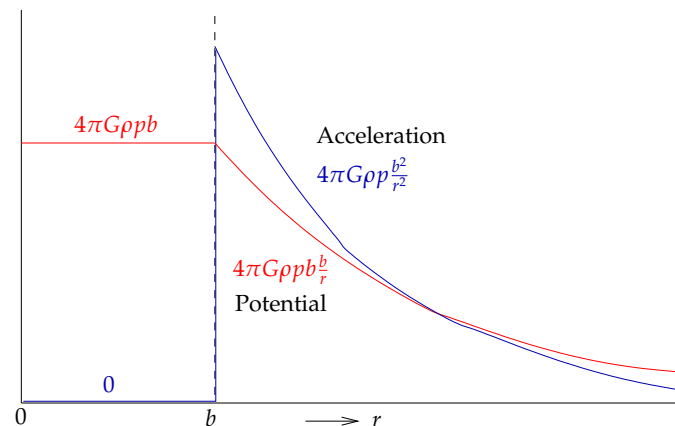
Figure 1.3. Dependence of potential and attraction on distance r from the centre of a spherical shell.

would be if all the mass of these layers was concentrated in the centre of the body. This case, where the distribution of mass density inside a body only depends on the distance from its centre, is called an *isotropic density distribution*.

1.5   Computing the attraction from the potential

As we argued earlier, the potential V is a *path integral*. Conversely we can compute, from the potential, the components of the gravitational acceleration vector by *differentiating V* with respect to *place*, i.e., by applying the *gradient operator*:

$$\mathbf{a} = \vec{\nabla}V = \overrightarrow{\text{grad}}V = \mathbf{i}\frac{\partial V}{\partial x} + \mathbf{j}\frac{\partial V}{\partial y} + \mathbf{k}\frac{\partial V}{\partial z}. \qquad (1.7)$$

Here, the symbol ∇ (nabla), is a frequently used *partial differential operator*

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}.$$

Here, $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is again a *basis* of mutually orthogonal unit vectors in Euclidean space $\mathbb{R}^3$ parallel to the (x, y, z) axes.

Let us try this differentiation in the case of the potential field of the point mass M. Substitute the above equations for V 1.4 and ℓ 1.3[9]:

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial \ell}\frac{\partial \ell}{\partial x} = GM \cdot -\frac{1}{\ell^2} \cdot \frac{x-X}{\ell} = -GM\frac{x-X}{\ell^3}.$$

Similarly we compute the y and z components:

$$\frac{\partial V}{\partial y} = -GM\frac{y-Y}{\ell^3}, \quad \frac{\partial V}{\partial z} = -GM\frac{z-Z}{\ell^3}.$$

These are the components of gravitational acceleration when the *source* of the field is one point mass M. So, in this concrete case the vector equation given above applies:

$$\mathbf{a} = \overrightarrow{\mathrm{grad}}V = \vec{\nabla}V.$$

Remark: in physical geodesy — unlike in physics — the potential is reckoned always positive if the attracting mass M is positive (as it is known to always be). However, the *potential energy* of body m inside the field of mass M is negative! More precisely, the potential energy of body m is

$$E_{\mathrm{pot}} = -Vm.$$

In practice one calls the vector of gravitational acceleration the "gravitational vector".

---

[9]From the equation

$$\ell = \sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2} = \left((x-X)^2 + (y-Y)^2 + (z-Z)^2\right)^{\frac{1}{2}}$$

it follows with the chain rule, that

$$\frac{\partial \ell}{\partial x} = \frac{\partial\left((x-X)^2 + (y-Y)^2 + (z-Z)^2\right)^{\frac{1}{2}}}{\partial\left((x-X)^2 + (y-Y)^2 + (z-Z)^2\right)} \cdot \frac{\partial (x-X)^2}{\partial x} =$$

$$= \frac{1}{2}\left((x-X)^2 + (y-Y)^2 + (z-Z)^2\right)^{-\frac{1}{2}} \cdot 2(x-X) = \frac{x-X}{\ell}.$$

1.6   Potential of a solid body

In the following we study a *solid body*, the mass of which is distributed throughout space and thus not concentrated in a single point. The Earth serves as an example of this, as its mass distribution in space may be described by a matter *density function* ρ:

$$\rho(x,y,z) = \frac{dm(x,y,z)}{d\mathcal{V}(x,y,z)},$$

in which dm is a mass element and $d\mathcal{V}$ is the corresponding element of spatial volume. The dimension of ρ is density, its unit in the SI system, $\mathrm{kg/m^3}$.

Because the gravitational acceleration 1.7 is a linear expression in the potential V, and both force and acceleration vectors may be summed linearly, it follows that also the total potential of the body can be obtained by summing together the potentials of all its parts. For example, the potential of a set of n mass points is

$$V = G\sum_{i=1}^{n}\frac{m_i}{\ell_i},$$

from which we obtain the gravitational acceleration by simply using the gradient theorem 1.7.

The potential of a solid body is obtained similarly by replacing the sum by an integral, in the following way. (*Note* that unfortunately almost the same symbols V and $\mathcal{V}$ are used here for the potential and for volume, respectively):

$$V = G\iiint_{\mathrm{body}}\frac{dm}{\ell} = G\iiint_{\mathrm{body}}\frac{\rho}{\ell}\,d\mathcal{V}. \tag{1.8}$$

The symbol ρ inside the integral designates the matter density at the location of dm. $\ell = \|\mathbf{r}-\mathbf{R}\| = \sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2}$ is the distance between point of measurement and attracting mass element. More clearly:

$$V(x,y,z) = G\iiint_{\mathrm{body}}\frac{\rho(X,Y,Z)}{\sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2}}\,dX\,dY\,dZ.$$

As we showed already above for mass points, also the first derivative with respect to place or *gradient* of the potential V of a solid body,

$$\overrightarrow{\mathrm{grad}}V = \vec{\nabla}V = \mathbf{a}, \tag{1.9}$$

is the acceleration vector caused by the attraction of the body. This applies generally.

1.6.1   Behaviour at infinity

If a body is of finite extent (i.e., it lies completely within a sphere of size ϵ around the origin) and also its density is bounded everywhere, it follows that

$$\|\mathbf{r}\| \to \infty \implies V(\mathbf{r}) \to 0,$$

because, according to the triangle inequality,

$$\ell = \|\mathbf{r} - \mathbf{R}\| \geq \|\mathbf{r}\| - \|\mathbf{R}\| \geq \|\mathbf{r}\| - \epsilon$$

and thus

$$\frac{1}{\ell} \to 0 \quad \text{when} \quad \|\mathbf{r}\| \to \infty.$$

For the acceleration of gravitation the same applies for all three components, and thus also for the length of this vectorial quantity:

$$\|\mathbf{r}\| \to \infty \implies \left\|\vec{\nabla} V\right\| \to 0.$$

This result can still be sharpened: if $\|\mathbf{r}\| \to \infty$, then, again by the triangle inequality,

$$\ell = \|\mathbf{r} - \mathbf{R}\| \leq \|\mathbf{r}\| + \|\mathbf{R}\| \leq \|\mathbf{r}\| + \epsilon,$$

and thus

$$\frac{1}{\|\mathbf{r}\| + \epsilon} \leq \frac{1}{\ell} \leq \frac{1}{\|\mathbf{r}\| - \epsilon} \implies \frac{1}{\|\mathbf{r}\|} \frac{1}{1 + \epsilon/\|\mathbf{r}\|} \leq \frac{1}{\ell} \leq \frac{1}{\|\mathbf{r}\|} \frac{1}{1 - \epsilon/\|\mathbf{r}\|}.$$

It is seen, again with the notation $r = \|\mathbf{r}\|$, that

$$r \to \infty \implies \frac{1}{\ell} \to \frac{1}{r}.$$

When we substitute this into the above integral 1.8, it follows that for large distances $r \to \infty$:

$$V = G \iiint_{\text{body}} \frac{\rho}{\ell} \, d\mathcal{V} \approx \frac{G}{r} \iiint_{\text{body}} \rho \, d\mathcal{V} = \frac{GM}{r},$$

in which M, the integral of density over the volume of the body, is precisely its *total mass*. From this we see, that at great distance, the field of a finite

sized body M is *nearly identical* with the field generated by a point mass the *total mass* of which is equal to the total mass of the body. This important observation was already made by Newton. As a result of this phenomenon we can treat, in the celestial mechanics of the Solar system, the Sun and planets[10] as mass points, although we know that they are not.

1.7   Example: The potential of a line of mass

The potential of a vertical line of mass having a linear mass density of unity is

$$V(x, y, z) = G \int_0^H \frac{1}{\sqrt{(X - x)^2 + (Y - y)^2 + (Z - z)^2}} \, dZ, \tag{1.10}$$

where (X, Y) is the location of the mass line, (x, y, z) is the location of the point in which the potential is being evaluated, and the mass line extends from sea level $Z = 0$ to height $Z = H$.

Firstly we write $\Delta x = X - x$, $\Delta y = Y - y$, $\Delta z = Z - z$, and the potential becomes

$$V(\Delta x, \Delta y, \Delta z) = G \int_{-z}^{H-z} \frac{1}{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}} \, d(\Delta z).$$

The indefinite integral is

$$\ln\left(\Delta z + \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}\right)$$

and substituting the integration bounds yields

$$V = G \ln \frac{H - z + \sqrt{\Delta x^2 + \Delta y^2 + (H - z)^2}}{-z + \sqrt{\Delta x^2 + \Delta y^2 + z^2}}.$$

Now we can expand this into a Taylor series in H around the point $H = 0$: the first derivative of equation 1.10 is

$$\frac{\partial V}{\partial H} = \frac{G}{\sqrt{(X - x)^2 + (Y - y)^2 + (H - z)^2}} = \frac{G}{\ell}$$

---

[10] The only important exception is formed by the forces between a planet and its moons, both due to the flattening of the planet and due to tidal effects.

in which $\ell(H) = \sqrt{(X-x)^2 + (Y-y)^2 + (H-z)^2}$. The second derivative is (chain rule)

$$\frac{\partial^2 V}{\partial H^2} = G\frac{\partial}{\partial H}\ell^{-1} = G \cdot -\frac{1}{2}\ell^{-3} \cdot 2(H-z) = -G\frac{H-z}{\ell^3}.$$

The third derivative, obtained in the same way:

$$\frac{\partial^3 V}{\partial H^3} = G\frac{\partial}{\partial H}\left(-\frac{H-z}{\ell^3}\right) = G\left(\frac{3(H-z)^2}{\ell^5} - \frac{1}{\ell^3}\right) = G\frac{3(H-z)^2 - \ell^2}{\ell^5},$$

and so on. The Taylor expansion is

$$V = \underbrace{\int_0^{H=0}\frac{1}{\ell}dZ}_{0} + G\frac{1}{\ell_0}H + \frac{1}{2}G\frac{z}{\ell_0^3}H^2 + \frac{1}{6}G\frac{3z^2 - \ell_0^2}{\ell_0^5}H^3 + \dots \qquad (1.11)$$

in which $\ell_0 = \sqrt{(X-x)^2 + (Y-y)^2 + z^2}$, i.e., into the derivatives we have substituted $H = 0$.

Question: how can we exploit this result for computing the gravitational potential of a complete, realistic topography?

Answer: in this expansion, the coefficients $\frac{1}{\ell_0}, \frac{1}{2}\frac{z}{\ell_0^3}, \dots$, like ℓ_0, depend only on the *differences* $\Delta x = X - x$ and $\Delta y = Y - y$ between the co-ordinates of the location of the mass line (X, Y) and those of the evaluation location (x, y) — and of the height z of the evaluation point. If the topography is given in the form of a grid, we may evaluate the above expansion 1.11 term-wise for the given z value and for all possible value pairs $(\Delta x, \Delta y)$.

Then, if the grid is, e.g., $N \times N$ in size, we need only N^2 operations for calculating every coefficient. The brute-force evaluation of the Taylor expansion itself *for the whole topography*, i.e., for every point of the terrain grid and every point of the evaluation grid, requires after that N^4 operations, but those are much simpler: the coefficients themselves have been precomputed. And brute force isn't even the best approach: as we shall see, the above *convolution* can be computed much faster using the Fast Fourier Transform.

We shall return to this subject more extensively in connection with terrain effect evaluation, sections 6.3 and 9.7.

_____

1.8   Equations of Laplace and Poisson

The *second derivative* with respect to place of the geopotential, the first derivative with respect to place of the gravitational acceleration vector, i.e., its *divergence*, is also of geophysical interest. We may write:

$$\text{div } \mathbf{a} = \left\langle \vec{\nabla} \cdot \mathbf{a} \right\rangle = \left\langle \vec{\nabla} \cdot \left(\vec{\nabla}V\right) \right\rangle = \left\langle \vec{\nabla} \cdot \vec{\nabla} \right\rangle V =$$
$$= \Delta V = \frac{\partial^2}{\partial x^2}V + \frac{\partial^2}{\partial y^2}V + \frac{\partial^2}{\partial z^2}V, \qquad (1.12)$$

in which

$$\Delta \overset{\text{def}}{=} \left\langle \vec{\nabla} \cdot \vec{\nabla} \right\rangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is a well known symbol called the *Laplace*[11] *operator*.

In equation 1.4 for the potential of a point mass we may show, by performing all partial derivations 1.12, that

$$\Delta V = 0, \qquad (1.13)$$

the well known *Laplace equation*. This equation applies outside a point mass, and more generally everywhere in empty space: all masses can in the limit be considered to consist of point-like mass elements. Or, in equation 1.8 we may directly differentiate inside the triple integral sign, exploiting the circumstance that it is allowed to interchange integration and partial differentiation, if both are defined.

A potential field for which the Laplace equation 1.13 applies, is called a *harmonic* field or function.

_____

[11]Pierre-Simon, marquis de Laplace (1749–1827) was a French universal genius who contributed to mathematics and natural sciences. He is one of the 72 French scientists, engineers and mathematicians whose names were inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

In the case where the mass density doesn't vanish everywhere, we have a different equation, with ρ the mass density:

$$\Delta V = -4\pi G\rho. \tag{1.14}$$

This equation is called the Poisson[12] equation.

The pair of equations

$$\overrightarrow{\mathrm{grad}}\,V = \mathbf{a}$$

$$\mathrm{div}\,\mathbf{a} = -4\pi G\rho$$

is known as the *field equations* of the gravitational field. They play the same role as Maxwell's[13] field equations in electromagnetism. Unlike in Maxwell's equations, however, in the above there is no time co-ordinate. Because of this, it is not possible to derive a formula for the propagation in space of gravitational waves, like the one for electromagnetic waves in Maxwell's theory.

We know today that these "Newton field equations" are only approximate, and that a more precise theory is Einstein's general theory of relativity. Nevertheless, in physical geodesy Newton's theory is generally precise enough and we shall use it exclusively.

1.9 Gauge invariance

An important property of the potential is, that, if we add a *constant C* to it, nothing related to gravitation that can actually be measured, changes. This

mittainvarianssi is called *gauge invariance*. Gravitation itself is obtained by differentiation,

---

[12]Siméon Denis Poisson (1781–1840) was a French mathematician, physicist and geodesist, one of the 72 names inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

[13]James Clerk Maxwell FRS FRSE (1831–1879) was a Scottish physicist, the discoverer of the field equations of electromagnetism. He found a wave-like solution to the equations, and, based on propagation speed, identified *light* as such.

an operation that eliminates the constant term. Therefore the definition of potential is in a sense ambiguous: all potential fields V obtained by different choices of C are equally valid.

Observations only give us potential *differences*, as spirit levellers know all too well.

An often chosen definition of potential is based on requiring that, if $\|\mathbf{r}\| \to \infty$, then also $V \to 0$, which makes physical sense and yields simple equations. However, in work on the Earth's surface, a more practical alternative may be $V = 0$ at the mean sea surface — although also that is not free of problems.

For example, for the mass of the Earth $M_\oplus$ a physically sensible form of the potential is, in spherical approximation,

$$V = \frac{GM_\oplus}{r}, r \stackrel{\text{def}}{=} \|\mathbf{r}\|,$$

which vanishes at infinity $r \to \infty$, when again a geodetically sensible definition would be

$$V = \frac{GM_\oplus}{r} - \frac{GM_\oplus}{R},$$

in which $R \stackrel{\text{def}}{=} \|\mathbf{R}\|$ is the radius of the Earth sphere. The latter potential vanishes where $r = R$, on the surface of a spherical Earth. In the limit $r \to \infty$ its value is $-\frac{GM_\oplus}{R}$, not zero.

1.10 Single mass density layer

If we apply to the surface S of a body a "coating" of mass surface density, of mass density value

$$\kappa = \frac{dm}{dS},$$

we obtain for the potential an integral equation similar to equation 1.8, but a *surface integral*:

$$V = G \iint_{\text{surface}} \frac{dm}{\ell} = G \iint_{\text{surface}} \frac{\kappa}{\ell}\,dS. \tag{1.15}$$

Here again ℓ is the distance between the point of consideration or test mass P and the moving mass element in integration dm (or surface element dS).

Note that the dimension of the mass surface density κ is kg/m², different from the dimension of ordinary (volume) mass density.

This case is theoretically interesting, though physically unrealistic. The function V is everywhere continuous, also at the surface S, however already its first derivatives with place are discontinuous. The discontinuity appears in the direction perpendicular to the surface, i.e., in the *normal derivative*.

Let us look at the simple case where a sphere, radius R, has been coated with a layer of standard surface density κ. By computing the above integral 1.15 we may prove (in a complicated way) that the exterior potential is the same as it would be if all of the mass of the body were concentrated in the sphere's centre. Earlier, in section 1.4, we proved that the potential interior to the sphere is *constant*.

Thus, the exterior attraction ($\ell > R$) is

$$a_\mathrm{e}(\ell) = G\frac{M}{\ell^2} = G\frac{\kappa \cdot 4\pi R^2}{\ell^2} = 4\pi G\kappa \left(\frac{R}{\ell}\right)^2.$$

The interior attraction ($\ell < R$) is

$$a_\mathrm{i}(\ell) = 0.$$

This means that on the surface of the sphere, $\ell = R$, the attraction is *discontinuous*:

$$a_\mathrm{e}(R) - a_\mathrm{i}(R) = 4\pi G\kappa.$$

In this symmetric case we see that

$$a = \|\mathbf{a}\| = \frac{\partial V}{\partial n} \tag{1.16}$$

in which the differentiation variable n represents the *normal direction*, i.e., the direction perpendicular to the surface S. If the surface S is an *equipotential surface* of the potential V, equation 1.16 applies generally. Then, the attraction vector — more precisely, the acceleration vector — is perpendicular to the surface S, and its magnitude is equal to that of the normal derivative of the potential.
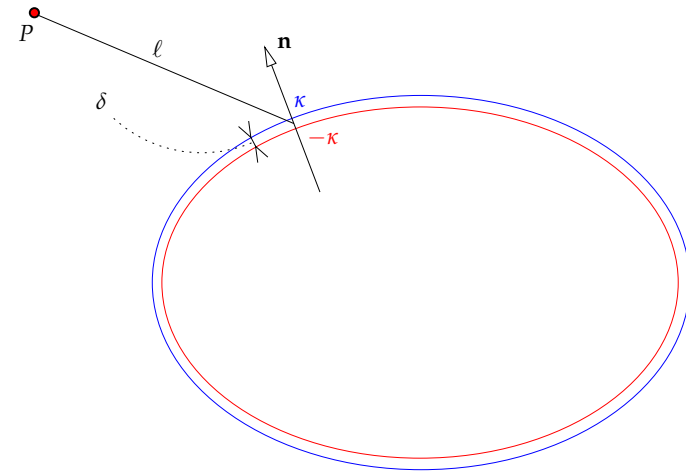
Figure 1.4. A double mass density layer.

1.11 Double mass density layer

A double mass density layer may be interpreted as a *dipole density layer*. The dipoles are oriented in the direction of the surface's normal.

If the dipole consists of two "charges" m and $-m$ in locations $\mathbf{r}_1$ and $\mathbf{r}_2$, in such a way, that the vectorial separation between them is $\Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, then the dipole *moment* is $\mathbf{d} = m\Delta\mathbf{r}$, a vectorial quantity. See figure 1.4.

Let the dipole surface density be

$$\mu = \frac{d\mathcal{M}}{dS},$$

in which $d\mathcal{M}$ is a "dipole layer element". This layer may be seen as made up of two single layers. If we have a positive layer at density κ and a negative layer at density $-\kappa$, and the distance between them is δ, we get for small values of δ an approximate correspondence:

$$\mu \approx \delta\kappa.$$

The combined potential of the two single mass density layers computed as

explained in the previous section is

$$V = G \iint_{\text{surface}} \kappa \left(\frac{1}{\ell_1} - \frac{1}{\ell_2} \right) dS.$$

Between ℓ_1, ℓ_2 and δ exists the following relationship (Taylor expansion of function $\frac{1}{\ell}$):

$$\frac{1}{\ell_1} = \frac{1}{\ell_2} + \delta \cdot \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) + \cdots,$$

in which $\frac{\partial}{\partial n}$ is again the derivative of the quantity in the normal direction of the surface.

Substitution into the equation yields

$$V = G \iint_{\text{surface}} \kappa \delta \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS = G \iint_{\text{surface}} \mu \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS.$$

If δ is small enough (and κ correspondingly large), this holds exactly.

One can easily show that the above potential isn't even continuous. The discontinuity happens at the surface S. Let us look again, for the sake of simplicity, at a sphere, radius R, coated with a double layer of constant mass density μ. The exterior potential is

$$V_{\text{e}} = -G\mu \iint_{\text{surface}} \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS = 0, \tag{1.17}$$

because the integral vanishes. To prove this, we need the Gauss integral theorem, on which more below.

The interior potential is

$$V_{\text{i}} = -G\mu \iint_{\text{surface}} \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS = -4\pi R^2 G\mu \left. \left(\frac{1}{\ell} \right)^2 \right|_{\ell=R} = -4\pi G\mu,$$

by computing the surface integral using the sphere's centre as the evaluation point, and using the earlier established circumstance that inside a sphere covered by a *single* constant-density mass density layer the potential is constant.

Now in the limit $\ell \to R$ the result is different for the exterior and interior potentials. The difference is

$$V_{\text{e}}(R) - V_{\text{i}}(R) = 4\pi G\mu.$$

≡ ↑ 🖼 ⊞ 🔍 🗐 ✛

1.12 The Gauss integral theorem

1.12.1 Presentation

The Gauss[14] integral theorem, famous from physics, looks in its vector form like this:

$$\iiint_{\mathcal{V}} \text{div } \mathbf{a} \, d\mathcal{V} = \iint_{\partial \mathcal{V}} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS, \tag{1.18}$$

in which $\mathbf{n}$ is the exterior normal to surface S, now as a vector: the length of the vector is assumed $\|\mathbf{n}\| = 1$. $\partial \mathcal{V}$ is the total surface of body $\mathcal{V}$.

This theorem applies to all differentiable vector fields $\mathbf{a}$ and all "well behaved" bodies $\mathcal{V}$ on whose surface $\partial \mathcal{V}$, everywhere a normal direction $\mathbf{n}$ exists. In other words, this is not a special property of the gravitational acceleration vector, though it applies also to this vector field.

1.12.2 Intuitive description

Let us note that

$$\text{div } \mathbf{a} = \Delta V = -4\pi G\rho$$

is a *source function*. It describes the amount, in the part of space inside surface $\partial \mathcal{V}$, of positive and negative "sources" and "sinks" of the gravitational field.   lähteet, nielut

The situation is fully analoguous with the flow pattern of a liquid: positive charges correspond to points where liquid is added to the flow, negative charges[15] correspond to "sinks" through which liquid disappears. The vector $\mathbf{a}$ is in this metaphor the velocity of flow; in the absence of "sources" and "sinks" it satisfies the condition div $\mathbf{a} = 0$, which describes the conservation and incompressibility of matter.

On the other hand, the function

$$\langle \mathbf{a} \cdot \mathbf{n} \rangle = \frac{\partial V}{\partial n}$$

is often called the *flux*, in other words, how much field stuff "flows out" —   vuo

---

[14]Johann Carl Friedrich Gauss (1777 – 1855) was a German mathematician and universal genius. "*Princeps mathematicorum*".

[15]But the charges for gravitation, i.e., masses, are always positive.

≡ ↑ 🖼 ⊞ 🔍 🗐 ✛
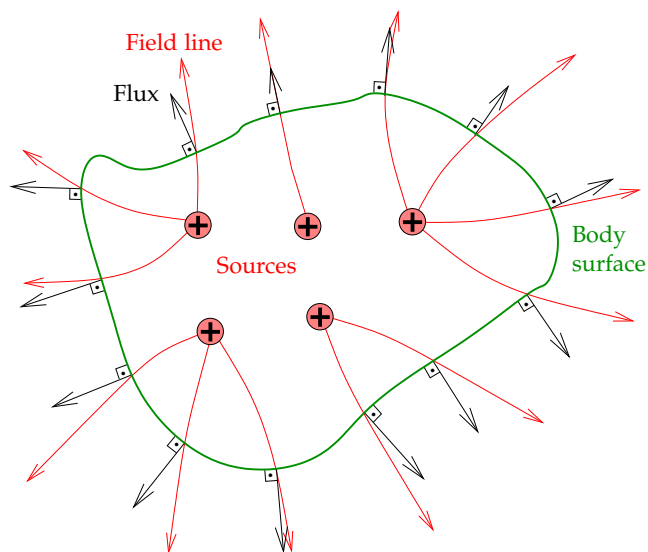
Figure 1.5. A graphical explanation of the Gauss integral theorem. The concept of *field line* was Michael Faraday's invention.

just like a liquid flow — from the space inside the surface S to the outside through S.

The Gauss integral theorem states the two amounts to be equal: it is in a way a *book-keeping statement* demanding that everything which is produced inside a surface — div $\mathbf{a}$ — has also to come out through the surface — $\langle \mathbf{a} \cdot \mathbf{n} \rangle$.

In figure 1.5 it is graphically explained, that the sum of "sources" over the inner space of the body, i.e., $\sum (+ + + \ldots)$, has to be the same as the sum of "flux" $\sum (\uparrow\uparrow\uparrow \ldots)$ over the whole boundary surface delimiting this inner space.

1.12.3   The potential version of the Gauss theorem

Let us write the Gauss integral theorem a bit differently, using the *potential* instead of the gravitational vector:

$$\iiint_{\mathcal{V}} \Delta V \, d\mathcal{V} = \iint_{\partial \mathcal{V}} \frac{\partial V}{\partial n} \, dS, \tag{1.19}$$

in which we have done the above substitutions. We also here see the popular notation $\partial \mathcal{V}$ for the surface of the body $\mathcal{V}$. The presentational forms 1.19 and 1.18 are connected by the equations 1.12 and 1.9, connecting V and $\mathbf{a}$.

1.12.4   Example 1: a little box

Let us look at a little rectangular box with sides $\Delta x, \Delta y$ and Δz, so little, that the field $\mathbf{a}(x, y, z)$ is inside it an almost linear function of place. Let us write $\mathbf{a}$ as the gradient of the potential V:

$$\mathbf{a} = \nabla V = \frac{\partial V}{\partial x}\mathbf{i} + \frac{\partial V}{\partial y}\mathbf{j} + \frac{\partial V}{\partial z}\mathbf{k} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$$

in which

$$a_x = \frac{\partial V}{\partial x}, a_y = \frac{\partial V}{\partial y}, a_z = \frac{\partial V}{\partial z}.$$

Now the volume integral

$$\iiint_{\mathcal{V}} \operatorname{div} \mathbf{a} \, d\mathcal{V} \approx \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \right) \Delta x \, \Delta y \, \Delta z \tag{1.20}$$

while the surface integral

$$\iint_{\partial V} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS \approx \left(a_x^+ - a_x^- \right) \Delta y \, \Delta z + \left(a_y^+ - a_y^- \right) \Delta x \, \Delta z + \left(a_z^+ - a_z^- \right) \Delta x \, \Delta y.$$

Here a_x^+ is the value of component a_x on the one face in the x direction and a_x^- its value on the other face, etc. For example, a_z^+ is the value of a_z on the box's upper and a_z^- on its lower face. A box has of course six faces, in each of three co-ordinate directions both "up- and downstream".
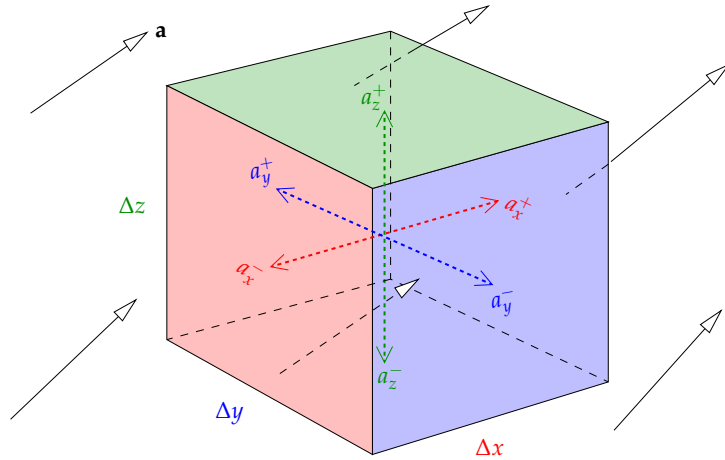
Figure 1.6. A little rectangular box.

Then

$$a_x^+ - a_x^- \approx \frac{\partial a_x}{\partial x}\Delta x,$$

$$a_y^+ - a_y^- \approx \frac{\partial a_y}{\partial y}\Delta y,$$

$$a_z^+ - a_z^- \approx \frac{\partial a_z}{\partial z}\Delta z,$$

and by substitution we see that

$$\iint_{\partial V} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS \approx \frac{\partial a_x}{\partial x}\Delta x \cdot \Delta y \, \Delta z + \frac{\partial a_y}{\partial y}\Delta y \cdot \Delta x \, \Delta z + \frac{\partial a_z}{\partial z}\Delta z \cdot \Delta x \, \Delta y =$$

$$= \left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \right)\Delta x \, \Delta y \, \Delta z,$$

the same equation as 1.20. So, in this simple case the Gauss equation applies.

Obviously the equation works also, if we build out of these "LEGO™ bricks" a larger body, because the faces of the bricks touching each other are oppositely oriented and cancel from the surface integral of the whole body. It is a bit harder to prove that the equation also applies to bodies having inclined surfaces.

1.12.5   Example 2: The Poisson equation for a sphere

According to the Poisson equation 1.14 we have

$$\Delta V = -4\pi G\rho.$$

Assume a sphere, radius R, within which the mass density ρ is constant. The volume integral thoughout the sphere gives

$$\iiint_{\mathcal{V}} \Delta V \, d\mathcal{V} = -4\pi G\rho \iiint_{\mathcal{V}} d\mathcal{V} = -4\pi G\rho \mathcal{V} = -4\pi GM, \qquad (1.21)$$

in which $M = \rho\mathcal{V}$ is the total mass of the sphere.

On the surface of the sphere, the normal derivative is

$$\frac{\partial V}{\partial n} = \frac{\partial}{\partial r}\frac{GM}{r}\bigg|_{r=R} = -\frac{GM}{R^2},$$

a constant, and its integral over the surface of the sphere is

$$\iint_{\partial \mathcal{V}} \frac{\partial V}{\partial n} \, dS = -\frac{GM}{R^2} \cdot S = -\frac{GM}{R^2} \cdot 4\pi R^2 = -4\pi GM. \qquad (1.22)$$

The results 1.22 and 1.21 are identical, as the Gauss theorem 1.19 requires.

1.12.6   Example 3: a point mass in an eight-unit cube

See figure 1.7.

Let us assume that we have a point mass of size GM in the centre of a cube bounded by the co-ordinate planes $x, y, z \in \{-1, +1\}$. In that case the volume integral is

$$\iiint_{\mathcal{V}} \Delta V \, d\mathcal{V} = -4\pi GM \iiint_{\mathcal{V}} \delta(\mathbf{r}) \, d\mathcal{V} = -4\pi GM,$$

where $\delta(\mathbf{r})$ is Dirac[16]'s delta function in space, having an infinite spike at the origin, being zero elsewhere, and producing a value of 1 upon volume

---

[16]Paul Adrien Maurice Dirac (1902–1984) was an English quantum physicist who found the relativistic wave equation for the electron, and theoretically predicted the existence of antimatter. Physics Nobel laureate 1933, shared with Erwin Schrödinger. He is also believed to have been on the autism spectrum (Farmelo, 2011).
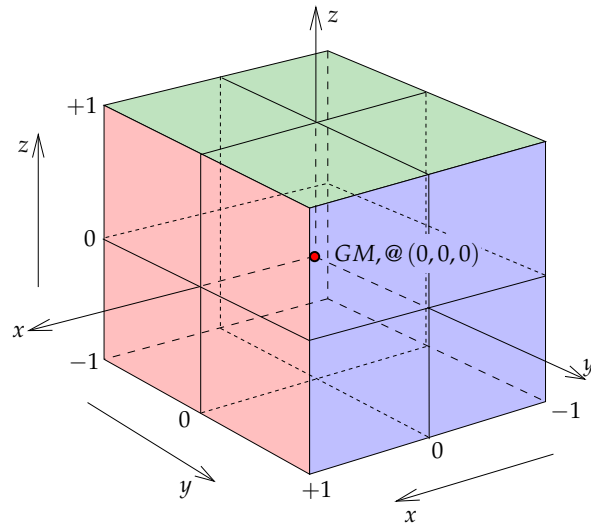
Figure 1.7. Eight-unit cube.

integration. The surface integral is six times that over the top face

$$-GM \int_{-1}^{+1} \left(\int_{-1}^{+1} \frac{1}{\left(x^2 + y^2 + 1\right)^{3/2}} \, dx \right) dy.$$

Integrating with respect to x (expression in square brackets) yields

$$\int_{-1}^{+1} \frac{1}{\left(x^2 + y^2 + 1\right)^{3/2}} \, dx = \left. \frac{x}{\left(y^2 + 1\right)\sqrt{x^2 + y^2 + 1}} \right|_{-1}^{+1} = \frac{2}{\left(y^2 + 1\right)\sqrt{y^2 + 2}}.$$

Integrating this with respect to y yields

$$\int_{-1}^{+1} \frac{2}{\left(y^2 + 1\right)\sqrt{y^2 + 2}} \, dy = \left. 2\arctan \frac{y}{\sqrt{y^2 + 2}} \right|_{-1}^{+1} =$$

$$= 4 \arctan \frac{1}{\sqrt{3}} = 4 \cdot \frac{\pi}{6} = \frac{2}{3}\pi.$$

Adding the six faces together yields

$$-6 \cdot GM \int_{-1}^{+1} \left(\int_{-1}^{+1} \frac{1}{\left(x^2 + y^2 + 1\right)^{3/2}} \, dx \right) dy = -6 \cdot GM \cdot \frac{2}{3}\pi = -4\pi GM,$$

agreeing with the volume result above.

1.13   Green's theorems

Apply the Gauss integral theorem to the vector field

$$\mathbf{F} = U \vec{\nabla} V.$$

Here U and V are two different scalar fields. We obtain:

$$\iiint_{\mathcal{V}} \operatorname{div} \mathbf{F} \, d\mathcal{V} = \iiint_{\mathcal{V}} \left\langle \vec{\nabla} \cdot \left(U \vec{\nabla} V \right) \right\rangle d\mathcal{V} =$$

$$= \iiint_{\mathcal{V}} U \Delta V \, d\mathcal{V} + \iiint_{\mathcal{V}} \left\langle \vec{\nabla} U \cdot \vec{\nabla} V \right\rangle d\mathcal{V} =$$

$$= \iiint_{\mathcal{V}} U \Delta V \, d\mathcal{V} + \iiint_{\mathcal{V}} \left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z} \right) d\mathcal{V}$$

and

$$\iint_{\partial \mathcal{V}} \left\langle \mathbf{F} \cdot \mathbf{n} \right\rangle dS = \iint_{\partial \mathcal{V}} \left\langle U \vec{\nabla} V \cdot \mathbf{n} \right\rangle dS = \iint_{\partial \mathcal{V}} U \left\langle \vec{\nabla} V \cdot \mathbf{n} \right\rangle dS = \iint_{\partial \mathcal{V}} U \frac{\partial V}{\partial n} \, dS.$$

The end result is *Green*[17]*'s first theorem*:

$$\iiint_{\mathcal{V}} U \Delta V \, d\mathcal{V} + \iiint_{\mathcal{V}} \left(\frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z} \right) d\mathcal{V} = \iint_{\partial \mathcal{V}} U \frac{\partial V}{\partial n} \, dS.$$

This may be cleaned up, because the second term on the left is *symmetric* for the interchange of U and V. Let us therefore interchange U and V, and subtract the equations obtained from each other. The result is Green's *second theorem*:

$$\iiint_{\mathcal{V}} \left(U \Delta V - V \Delta U \right) d\mathcal{V} = \iint_{\partial \mathcal{V}} \left(U \frac{\partial V}{\partial n} - V \frac{\partial U}{\partial n} \right) dS.$$

We assume in all operations, that the functions U and V are "well behaved", i.e., all necessary derivatives etc. exist everywhere in body $\mathcal{V}$.

A useful special case arises by choosing for the function U:

$$U = \frac{1}{\ell},$$

---

[17]George Green (1793 – 1841) was a British mathematical physicist, an autodidact, working as a miller near Nottingham. He also invented the word "potential". Green (1828); O'Connor and Robertson (1998); Green's Windmill.
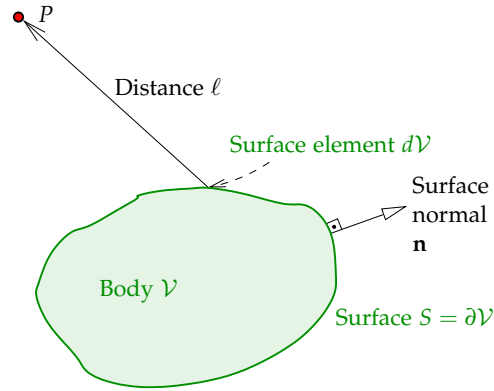
Figure 1.8. Geometry for deriving Green's third theorem if point P is outside surface $\partial\mathcal{V}$.

in which ℓ is the distance from the given point of evaluation P. This function U is well behaved everywhere *except* precisely in point P, where it is not defined.

In the case where point P is outside the surface $\partial\mathcal{V}$, the result, Green's third theorem, is now obtained by substitution:

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) \right) \, dS.$$

This case is depicted in figure 1.8.

In the case that point P is inside surface $\partial\mathcal{V}$, the computation becomes a little more complicated. One ought to learn about the clever technique that — in this case as in others — comes to the rescue. This is why we describe it shortly.

We form a small sphere of radius ϵ called $\mathcal{V}_2$ around point P; now we can formally define the body (containing a hole) $\mathcal{V} \stackrel{\text{def}}{=} \mathcal{V}_1 - \mathcal{V}_2$, and also its surface $\partial\mathcal{V}$ which consists of two parts, $\partial\mathcal{V} = \partial\mathcal{V}_1 - \partial\mathcal{V}_2$.

Now we may write the volume integral into two parts:

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = \iiint_{\mathcal{V}_1} \frac{1}{\ell}\Delta V \, d\mathcal{V} - \iiint_{\mathcal{V}_2} \frac{1}{\ell}\Delta V \, d\mathcal{V},$$
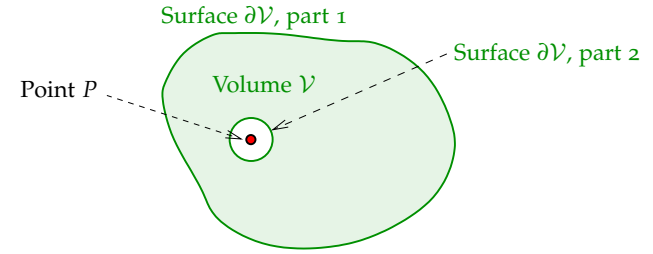
Figure 1.9. Geometry for deriving Green's third theorem if point P is inside surface $\partial\mathcal{V}$.

where the second term can be integrated in spherical co-ordinates:

$$\iiint_{\mathcal{V}_2} \frac{1}{\ell}\Delta V \, d\mathcal{V} \approx \Delta V_P \int_0^\epsilon 4\pi\ell^2\frac{1}{\ell} \, d\ell = 2\pi\Delta V_P\epsilon^2,$$

which will go to zero in the limit $\epsilon \to 0$.

For the first surface integral we obtain using the Gauss integral theorem 1.19 :

$$\iint_{\partial\mathcal{V}_2} \frac{1}{\ell}\frac{\partial V}{\partial n} \, dS = \frac{1}{\epsilon}\iint_{\partial\mathcal{V}_2} \frac{\partial V}{\partial n} \, dS = \frac{1}{\epsilon}\iiint_{\mathcal{V}_2} \Delta V \, d\mathcal{V} \approx \frac{1}{\epsilon}\Delta V_P \cdot \frac{4}{3}\pi\epsilon^3,$$

which also goes to zero for $\epsilon \to 0$.

The second surface integral (the normal is pointing away from P):

$$-\iint_{\partial\mathcal{V}_2} V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) \, dS = -\iint_{\partial\mathcal{V}_2} V \cdot -\frac{1}{\epsilon^2} \, dS \approx 4\pi\epsilon^2 \cdot \frac{1}{\epsilon^2}V_P = 4\pi V_P.$$

By combining all results with their correct algebraic signs we obtain — for the case where P is inside surface $\partial\mathcal{V}_1 \sim \partial\mathcal{V}$ — :

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = -4\pi V_P + \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) \right) \, dS. \qquad (1.23)$$

After this it must be intuitively clear, and we present without formal proof, that

$$\iiint_{\mathcal{V}} \frac{1}{\ell}\Delta V \, d\mathcal{V} = -2\pi V_P + \iint_{\partial\mathcal{V}} \left(\frac{1}{\ell}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{\ell}\right) \right) \, dS,$$

if point P is *on* the boundary surface of body $\mathcal{V}$, i.e., on $\partial\mathcal{V}$. This however presupposes that the normal derivative, and especially the *normal direction*, actually exist in precisely point P!
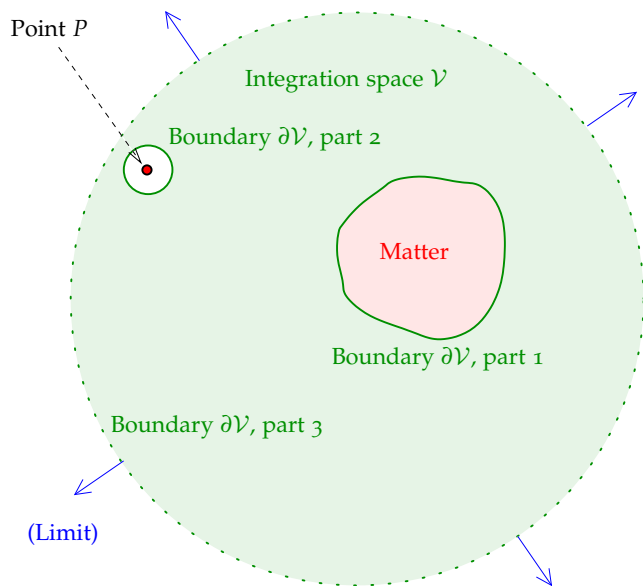
Figure 1.10. Green's third theorem for the space external to a body.

In geodesy, the typical situation is that where the body $\mathcal{V}$ over the volume of which one wants to evaluate the volume integral, is *the whole space outside the Earth*. In this case, conveniently $\Delta V = 0$ and the whole volume integral appearing above vanishes.

The result 1.23 may be generalized to this case, where $\mathcal{V}$ is the whole space outside surface $\partial \mathcal{V}$. This generalization is done by now choosing as the surface $\partial \mathcal{V}$ the three-part surface $\partial \mathcal{V} = \partial \mathcal{V}_1 + \partial \mathcal{V}_2 + \partial \mathcal{V}_3$, in which $\partial \mathcal{V}_3$ is a sphere of large radius centred on P. Its radius is then allowed to *grow in the limit to infinity*, so that all integrals over both the surface $\partial \mathcal{V}_3$ and the part of space outside it vanish. Also the normal direction to the surface $\partial \mathcal{V}_3$, like that of the little ball used earlier, is inverted, i.e., now aimed to the inside, towards the Earth.

The end result is:

$$\iiint_{\mathcal{V}} \frac{1}{\ell} \Delta V \, d\mathcal{V} = -4\pi V_P - \iint_{\partial \mathcal{V}} \left(\frac{1}{\ell} \frac{\partial V}{\partial n} - V \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) \right) dS, \qquad (1.24)$$

Because in this case, in which $\mathcal{V}$ is the part of the space external to the Earth, the left-hand side volume integral vanishes, we may express the potential at point P suitably as a two-term surface integral over surface $\partial \mathcal{V}$. See below.

1.14   The Chasles theorem

We study the above mentioned case where the "body" is the space outside the surface $\partial \mathcal{V}$ — i.e., in practice: the space outside the Earth.

From the Green theorem 1.24 derived above, we may derive for a *harmonic function V* (so, $\Delta V = 0$) in the exterior space:

$$V_P = -\frac{1}{4\pi} \iint_{\partial \mathcal{V}} \frac{1}{\ell} \frac{\partial V}{\partial n} \, dS + \frac{1}{4\pi} \iint_{\partial \mathcal{V}} V \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS. \qquad (1.25)$$

Interpretation: The exterior, harmonic potential of an arbitrarily shaped body can be represented as the sum of a single and a double mass density layer on the body's surface.

Explanation: We obtain the surface density of a single mass layer by equation 1.15,

$$\kappa = -\frac{1}{4\pi G} \frac{\partial V}{\partial n},$$

and the surface density of a double mass layer by equation 1.17,

$$\mu = \frac{V}{4\pi G}.$$

If we plug these into equation 1.25, we obtain

$$V_P = G \iint_{\partial \mathcal{V}} \left(\frac{\kappa}{\ell} + \mu \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) \right) dS.$$

———

In case that the surface ∂V is an equipotential surface of potential V, i.e., $V = V_0$, it follows that a single mass density layer suffices, because in that case

$$\iint_{\partial \mathcal{V}} V \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS = V_0 \iint_{\partial \mathcal{V}} \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS = 0.$$

The right-hand side integral vanishes based on the Gauss integral theorem. This is because the function $\frac{1}{\ell}$, with ℓ the distance from point P, is *harmonic*

inside the Earth's body of which ∂V is the surface. This is the Chasles theorem[18], also called the Green equivalent-layer theorem.

The theorem is used in Molodensky's[19] theory. Also the representation of the Earth's gravity field by underground point-mass layers, e.g., Vermeer (1984), might be justified with this theorem.

The case where ∂V is an equipotential surface is realized if the body is fluid and seeks by itself an external form equal to an equipotential surface. For planet Earth, this applies for the ocean surface. Also in electrostatic theory, for a conductor in which the electrons can move freely, the physical surface will become one equipotential surface. This is why it is often stated that the electric charges are on the surface of the conductor. This isn't necessarily so, but from a practical viewpoint the result is the same.

Equation 1.25 simplifies in this case as follows:

$$V_P = -\frac{1}{4\pi} \iint_{\partial V} \frac{1}{\ell} \frac{\partial V}{\partial n} \, dS = G \iint_{\partial V} \frac{\kappa}{\ell} \, dS. \tag{1.26}$$

The equation tells us that we can compute the whole potential exterior to the Earth, if only on the surface of the Earth — the shape of which we also assume given in order to compute $\frac{1}{\ell}$! — is given the gradient, in the normal, i.e., vertical, direction, of the potential, $\frac{\partial V}{\partial n}$. This gradient is precisely the gravitational acceleration, a quantity derivable from measurement. All of gravimetric geopotential determination ("geoid determination"), ever since G. G. Stokes, is based on this.

1.15   Boundary-value problems

The boundary-value problem (BVP) is the problem of computing the potential V throughout space (or throughout the body's exterior or interior part of

---

[18]Michel Chasles (1793 – 1880) was a French mathematician and geometrician, one of the 72 whose names are inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

[19]Mikhail Sergeevich Molodensky (1909 – 1991) was an illustrious Russian physical geodesist.

space) from given values relating to V on the boundary surface, e.g., on the surface of the Earth. The simplest boundary-value problem is Dirichlet[20]'s problem: on the boundary surface the potential V itself is given. More complicated boundary-value problems are based on *linear functionals* of the potential: on the boundary, some linear expression in V is given, e.g., a derivative or a linear combination of derivatives, generally

$$L\{V\},$$

with $L\{\cdot\}$ being a linear operator.

The Dirichlet boundary-value problem *in the form popular in geodesy* is: determine the potential field V if its values are given on a closed surface S, and furthermore is given that V is *harmonic* ($\Delta V = 0$) outside surface S. In the vacuum of space, the potential is always harmonic, as already earlier noted: the potential of a point mass m_P, $V = \frac{Gm_P}{\ell}$, is a harmonic function everywhere except at point P — and an extended body consists, in the limit, of many point masses or mass elements.

In the general case this is a theoretically challenging problem. The existence and uniqueness of the solution has been proven very generally, see Heiskanen and Moritz (1967) page 18.

1.16   What the boundary-value problem cannot compute

Based on the values of the potential function V on the surface S we may thus compute the function $V(x, y, z)$ throughout space outside the surface. The boundary-value problem is a powerful general method also applied in physical geodesy. One must however note, that from potential values given on the surface it is not possible to uniquely resolve *the mass distribution inside the Earth*, which generates this potential.

This is clear already in the simple case of a constant potential on the surface of a sphere. If additionally is given that the mass distribution is spherically

---

[20]Peter Gustav Lejeune Dirichlet (1805 – 1859) was a German mathematician also known for his contributions to number theory.

symmetric, then nevertheless the density profile along the radius remains indeterminate. All mass may be concentrated in the centre, or it may be as a thin layer just under the sphere's surface, or any alternative between these two extremes. Without additional information — e.g., from seismic studies or geophysical density models — we cannot resolve this issue.

Also the Chasles theorem mentioned above, equation 1.25, and its special case, equation 1.26, are examples of this: the theorem tells how one may describe the exterior potential as generated by a mass distribution on the surface of a body, although *we know* that the field's origin is a mass distribution extending throughout the body!

This is a fundamental limitation of all methods that try to obtain information on the situation inside the Earth based only on *gravimetric* measurements on or outside the Earth.

Self-test questions

1. Which instrument was used to determine the constant G? Why is it difficult to obtain a precise value for this constant?

2. Why do all objects, irrespective of their mass, undergo the same acceleration of free fall, although the gravitational attraction on a more massive body is obviously stronger?

3. What is a *conservative* force field?

   (a) A force field for which the force can be written as the gradient of a unique potential.

   (b) A force field in which an object carried along a closed loop will not gain and not lose energy.

   (c) An attractive force field from which no object can escape.

   (d) A force field the curl of which vanishes everywhere.

4. On the surface of a homogeneous, spherical asteroid the acceleration of free fall is $1 \, \text{cm}/\text{s}^2$. What is the acceleration of free fall on another asteroid that is otherwise similar, but that has twice the diameter?

   (a) $0.25 \, \text{cm}/\text{s}^2$

   (b) $1 \, \text{cm}/\text{s}^2$

   (c) $2 \, \text{cm}/\text{s}^2$

   (d) $4 \, \text{cm}/\text{s}^2$

5. What is a *harmonic* potential?

6. What is the *order* of the Laplace differential equation?

7. Is a linear potential, $V(x,y,z) \stackrel{\text{def}}{=} a + bx + cy + dz$, a, b, c, d constants, harmonic?

8. If the potential in the previous question is a gravitational potential, calculate its *acceleration vector*.

9. Under what condition is it possible to describe the external gravitational field emanating from a body as produced by a single mass density layer on the surface of that body?

10. The dipole layer density μ, section 1.11. What is the SI unit of this quantity?

Exercise 1−1: Core of the Earth

1. Derive the equation giving the acceleration of gravity g on the surface of a homogeneous-density sphere, if given are the density ρ and radius R_{core}.

2. The Earth's iron-nickel core has a mean density of $11 \, \text{g}/\text{cm}^3$ and its radius is $3500 \, \text{km}$. Compute the acceleration of gravitation on its surface g_{core}.

3. What is the attraction g at the centre of the core? What can you say *in general* about the geopotential in this point (*don't try to calculate it*)?

4. Derive the equation for the gravity gradient $\frac{\partial g}{\partial r}$ on the surface of a homogeneous-density sphere of density ρ.

Exercise 1−2: Atmosphere

1. The mean pressure of the atmosphere at sea level is $10^3 \, \text{hPa}$ (the unit Pascal: $\text{Pa} = \text{Nm}^{-2}$.) On the Earth surface gravity is $10 \, \text{m}/\text{s}^2$. Calculate the mean surface density as a thin layer κ in units of kg/m^2.
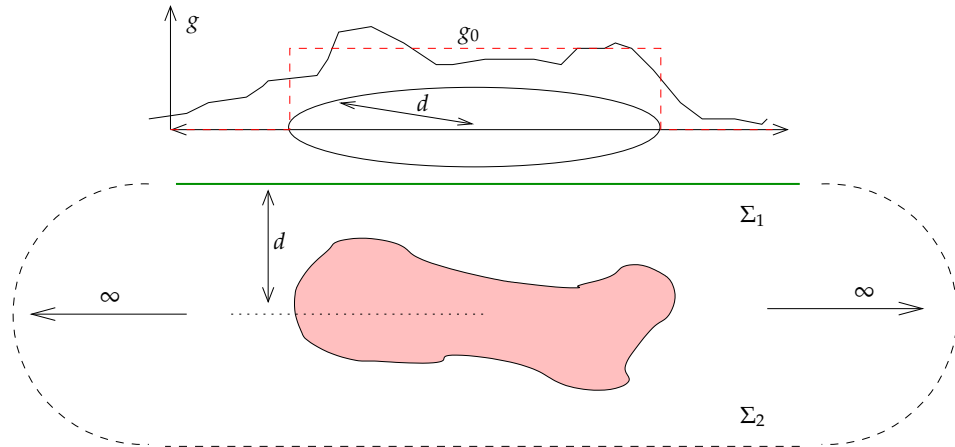
Figure 1.11. Iron-ore body.

2. Calculate the total mass of the atmosphere using the spherical shell approximation. You may take as its radius 6378 km.

3. Calculate the attraction generated by the atmosphere outside it, both as acceleration and as a fraction of the total Earth attraction.

4. How much is the attraction from the atmosphere inside the atmosphere?

Exercise 1−3:   The Gauss theorem

There is a deposit (body) of iron ore inside the Earth, which generates (in the flat Earth approximation!) an attraction on the Earth's surface, which has been drawn as the g curve. See figure 1.11.

The true attraction curve is *approximated* by a simple function

$$g = \begin{cases} g_0 & \text{if } r \leq d \\ 0 & \text{if } r > d \end{cases}$$

(red dashed line), where r is the distance from the point on the Earth surface

straight above the ore deposit. So, the area where $g \neq 0$ for a *disk* of radius d on the surface of the Earth.

1. Compute, using the above approximation for g, the surface integral

$$\iint_{\Sigma_1} g \, d\sigma,$$

where Σ_1 is the surface of the Earth, see 1.11.

2. According to the Gauss equation

$$\iint_{\Sigma_1} (-g) \, d\sigma + \iint_{\Sigma_2} (+g') \, d\sigma = \iiint_{\text{volume}} \Delta V \, d\mathcal{V} =$$
$$= \iiint_{\text{volume}} -4\pi G \rho_{\text{iron}} \, d\mathcal{V} =$$
$$= -4\pi G M_{\text{body}},$$

where $\Sigma_1 + \Sigma_2$ is the (two part) closed surface around the body. (The parts meet at infinity.) g and g' are the gravity acceleration functions on the surface of the Earth and on the surface Σ_2 — positive in the upward direction[21].

Assuming that

$$\iint_{\Sigma_1} g \, d\sigma - \iint_{\Sigma_2} g' \, d\sigma = 2 \iint_{\Sigma_1} g \, d\sigma \implies 2 \iint_{\Sigma_1} g \, d\sigma = 4\pi G M_{\text{body}},$$

calculate $G M_{\text{body}}$.

3. Assuming that the deposit is a sphere at depth d, calculate GM using Newton's law of gravitation from the value g_0 straight above the deposit at the Earth surface.

4. Compare results 2. and 3. and draw conclusions. Is the function g given above a good approximation?

---

[21]Be careful with the algebraic signs!

The Laplace equation and its solutions

2

2.1 The nature of the Laplace equation

An equation central to the study of the Earth's gravitational field is the Laplace equation,

$$\Delta V = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) V = 0.$$

We call the symbol Δ the *Laplace operator*. Often, the alternative notation ∇^2 is used.

If we study gravitation as a field, then the Laplace equation is more natural than Newton's formalism. Newton's equations are used when the mass distribution is known: it yields directly the gravitational force caused by the masses.

The Laplace equation on the other hand is a *partial differential equation*. Its solution gives the *potential* $V(x, y, z)$ of the gravitational field thoughout space or a part of space. From this potential one may then calculate the effect of the field on a body moving in space at the location where the body is. This is a two-phased process. The conceptual difference is, that a certain property, a *field*, is attributed to empty space, and we no longer talk about action at a distance directly between two bodies.

Solving the Laplace equation in the general case may be difficult. The approach generally taken is, that we choose some co-ordinate frame — a

rectangular frame (as above), spherical co-ordinates, cylindrical co-ordinates, toroidal co-ordinates, or whatever — which fits best with the geometry of the problem at hand. Then, we *transform* the Laplace equation to those co-ordinates, we find special solutions of a certain form, and finally we compose a general — or not-so-general — solution as a linear combination of those special solutions, i.e., a *series expansion*.

Fortunately the theory of linear partial differential equations is well developed. Similar theoretical problems are encountered in the theory of the electromagnetic field (Maxwell theory) and quantum mechanics (Schrödinger[1] equation), not to mention fluid and heat flow.

An important observation is, that the Laplace equation is *linear*. This means that, if given are two solutions

$$\Delta V_1 = 0,$$
$$\Delta V_2 = 0,$$

then also their linear combinations

$$V = \alpha V_1 + \beta V_2, \ \alpha, \beta \in \mathbb{R}$$

are good solutions, i.e., $\Delta V = 0$. This linearity property makes it possible to seek general solutions as linear combinations or series expansions of basic solutions.

A peculiarity that also distinguishes the Laplace equation from Newton's equation is, that it it a *local* equation. It describes the behaviour of the potential field in a small neighbourhood of one point. However, the solution is sought for a whole area. The solution approach commonly used is the so-called *boundary-value problem*. This means that the field values ("boundary values") have to be given only on the boundary of a certain part of space,

---

[1]Erwin Rudolf Josef Alexander Schrödinger (1887–1961) was a German physicist and quantum theorist, the inventor of the wave equation of matter named after him which earned him the 1933 physics Nobel (shared with Paul Dirac), and of the eponymous unobserved cat, which finds itself in a superposition state of being both alive and dead.

e.g., on the Earth's surface. From this, one calculates the values of the field in outer space — the behaviour of the field inside the Earth remains outside the scope of our interest. From the perspective of the exterior gravitational field one doesn't even need to know the precise mass distribution inside the Earth — and one cannot even determine it using only measurement values obtained on and above the Earth's surface!

2.2   The Laplace equation in rectangular co-ordinates

It is a learning experience to write and solve the Laplace equation in rectangular co-ordinates. The case is analoguous to that of spherical co-ordinates but the math is much simpler.

Assume that the Earth's surface is the level surface for z co-ordinates $z = 0$. Then

$$\Delta V = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) V = \Delta\big(X(x) \cdot Y(y) \cdot Z(z) \big),$$

in which we have "experimentally" written

$$V(x, y, z) = X(x) \cdot Y(y) \cdot Z(z).$$

In other words, we write experimentally V as the product of three factor functions, with each factor function depending only on one co-ordinate — "separation of variables". A realistic potential function V will of course usually not be of this form. We may however hope to write it as a linear combination of terms that *are* of the above form, thanks to the linearity of the Laplace equation.

muuttujien erottaminen

By taking all partial derivatives we obtain

$$YZ\frac{\partial^2}{\partial x^2}X + XZ\frac{\partial^2}{\partial y^2}Y + XY\frac{\partial^2}{\partial z^2}Z = 0.$$

Dividing by the expression XYZ yields

$$\frac{\frac{\partial^2 X(x)}{\partial x^2}}{X(x)} + \frac{\frac{\partial^2 Y(y)}{\partial y^2}}{Y(y)} + \frac{\frac{\partial^2 Z(z)}{\partial z^2}}{Z(z)} = 0.$$

Because this has to be true for *all* values x, y, z, it follows that *each term must be a constant*. If we take for the first and second constants $-k_1^2$ and $-k_2^2$, we

get in conclusion for the third constant $k_1^2 + k_2^2$. By writing this definition and result out and moving the denominator to the other side, we obtain

$$\frac{\partial^2}{\partial x^2} X(x) = -k_1^2 X(x),$$

(why the minus sign? We shall presently see...)

$$\frac{\partial^2}{\partial y^2} Y(y) = -k_2^2 Y(y),$$

and

$$\frac{\partial^2}{\partial z^2} Z(z) = \left(k_1^2 + k_2^2\right) Z(z).$$

Now, the solution is readily found at least to the first two equations: they are *harmonic oscillators*, and their basis solutions[2] are

$$X(x) = \exp(\pm i k_1 x),$$
$$Y(y) = \exp(\pm i k_2 y).$$

harmoninen
värähtelijä

The solution of the Z equation on the other hand is exponential:

$$Z(z) = \exp\left(\pm z \sqrt{k_1^2 + k_2^2}\right).$$

We can now form basis solutions in space:

$$V_{k_1 k_2}(x, y, z) = \exp\left(i\left(\pm k_1 x \pm k_2 y\right) \pm z \sqrt{k_1^2 + k_2^2}\right).$$

The general solution is obtained by summing the terms $V_{k_1 k_2}$ with varying coefficients, for different values of k_1, k_2.

We cannot choose the values k_1, k_2 entirely freely. Which values are allowed, will depend on the *boundary conditions* given.

Let us assume that both in the x and in the y direction the size of our world is L ("shoebox world[3]"). Let us make things a little simpler by assuming that, on the borders of our shoebox world, we have the *boundary conditions*

$$V(0, y) = V(L, y) = V(x, 0) = V(x, L) = 0.$$

---

[2] Alternative basis solutions are $X(x) = \sin k_1 x$, $X(x) = \cos k_1 x$ etc. They are equivalent with those presented because $\exp(i k_1 x) = \cos k_1 x + i \sin k_1 x$, $\exp(-i k_1 x) = \cos k_1 x - i \sin k_1 x$.

[3] ...though real-world shoeboxes are rarely square.

It then follows that the only pairs (k_1, k_2) yielding a solution that fits the box are

$$k_1 = \frac{\pi j}{L}, k_2 = \frac{\pi k}{L}, \; j, k \in \mathbb{Z},$$

and the only suitable functions are sine functions. Thus we obtain as a solution:

$$V_{jk}(x, y, z) = \sin\left(\pi j \frac{x}{L}\right) \sin\left(\pi k \frac{y}{L}\right) \exp\left(\pm \pi \sqrt{(j^2 + k^2)} \frac{z}{L}\right).$$

This particular solution may now be generalized by multiplying it with suitable coefficients, and summing it over different index values $j = 0, \pm 1, \pm 2, \ldots$; $k = 0, \pm 1, \pm 2, \ldots$. We may however remark, that the terms for which $j = 0$ or $k = 0$ will always vanish, and the terms that contain $j = +n$ and $j = -n$, or $k = +n$ and $k = -n$, $n \in \mathbb{N}$, are (apart from their algebraic signs) identical. Therefore in practice we sum over the values $j = 1, 2, \ldots$; $k = 1, 2, \ldots$.

Different boundary conditions will give slightly different general solutions. Their general form is however always similar.

The zero-level $z = 0$ expansion resulting from the general solution is the familiar Fourier[4] sine expansion:

$$V(x, y, 0) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} V_{jk}(x, y) =$$
$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} \sin\left(\pi\left(\frac{jx}{L}\right)\right) \sin\left(\pi\left(\frac{ky}{L}\right)\right), \quad (2.1)$$

in which the v_{jk} are Fourier coefficients, and the expressions

$$V_{jk}(x, y) \stackrel{\text{def}}{=} \sin\left(\pi\left(\frac{jx}{L}\right)\right) \sin\left(\pi\left(\frac{ky}{L}\right)\right)$$

are two-dimensional basis functions on the Earth's surface, more precisely, on level $z = 0$.

---

[4] Joseph Fourier (1768–1830) was a French mathematician and physicist — and some would say, climatologist — one of the Eiffel Tower's 72 names, Eiffel Tower, 72 names.
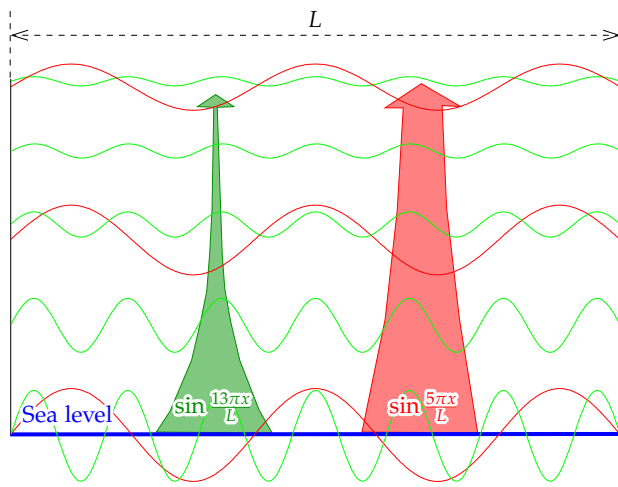
Figure 2.1. The exponential attenuation of gravitational field Fourier waviness with height. Rectangular geometry, one dimension. Long waves (small wave numbers, red) attenuate more slowly with height than short waves (green), meaning that the height acts as a low-pass filter.

We refer to section B.2.2 in appendix B for a description with illustration how a Fourier analysis and synthesis on a simple function is done, and how the Fourier expansion approximates the original function as terms are added.

A complete three-dimensional expansion again is

$$V(x,y,z) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} V_{jk}(x,y) \exp\left(\pm\pi\sqrt{j^2+k^2}\frac{z}{L}\right) =$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_{jk} \sin\left(\pi\left(\frac{jx}{L}\right)\right) \sin\left(\pi\left(\frac{ky}{L}\right)\right) \exp\left(\pm\pi\sqrt{j^2+k^2}\frac{z}{L}\right).$$

$$(2.2)$$

Note, that the z expression may have a positive as well as a negative algebraic sign! Of course the solution with a positive sign goes to $\to \infty$ when $z \to \infty$, which is not physically realistic in the exterior space.
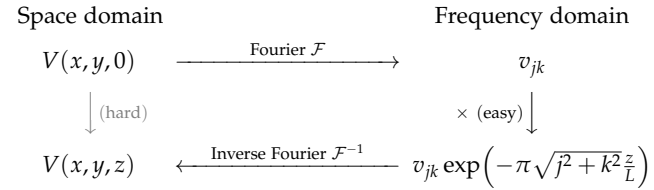
Figure 2.2. Vertically shifting the potential field V, in the space and frequency domains. Rectangular geometry.

Note also that $V(x,y,0)$ and v_{jk} describe *the same gravitational field* in two essentially different ways: in the space domain, and in the — spatial — frequency, or wave-number, domain. The information content in the two is the same. They can be transformed into each other by the forward and inverse Fourier transforms $\mathcal{F}$ and $\mathcal{F}^{-1}$.

In fact, the information content in $V(x,y,0)$ is in principle the same as that in $V(x,y,z)$ for *any* level z: knowing the potential of one surface means — with the Laplace equation — knowing the potential throughout space.

We summarize equations 2.1, 2.2 still in the commutative diagram 2.2.

The takeaway from this is, that the operation of vertically shifting the potential field V from zero level to the level z, which is not straightforward in the space domain, becomes simple — as in a straightforward multiplication — in the frequency domain[5]. The same applies in spherical co-ordinates, where the frequency domain means spherical-harmonic coefficients, as we shall see.

2.3   The Laplace equation in polar co-ordinates

In polar co-ordinates, i.e., two-dimensionally, the Laplace equation is

$$\Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \alpha^2} = 0.$$

---

[5]The reason for this is, as we shall later discuss more generally, that the vertical shift operation is a *convolution*.

We perform on this the same kind of separation of variables as in section 2.2, i.e., write first

$$V(\alpha, r) = A(\alpha)\, R(r)$$

and then split the above equation into two equations, one for the right-hand side function $R(r)$ and one for the function $A(\alpha)$.

What form does the $A(\alpha)$ function of the general solution have? Substitution yields

$$A(\alpha)\frac{\partial^2 R(r)}{\partial r^2} + \frac{A(\alpha)}{r}\frac{\partial R(r)}{\partial r} + \frac{R(r)}{r^2}\frac{\partial^2 A(\alpha)}{\partial \alpha^2} = 0.$$

Multiply by the expression $\dfrac{r^2}{A(\alpha)\, R(r)}$:

$$\left(\frac{r^2}{R(r)}\frac{\partial^2 R(r)}{\partial r^2} + \frac{r}{R(r)}\frac{\partial R(r)}{\partial r} \right) + \frac{1}{A(\alpha)}\frac{\partial^2 A(\alpha)}{\partial \alpha^2} = 0.$$

Both terms must be constant:

$$r\left(r\frac{\partial^2 R(r)}{\partial r^2} + \frac{\partial R(r)}{\partial r} \right) - k^2 R(r) = 0,$$

$$\frac{\partial^2 A(\alpha)}{\partial \alpha^2} + k^2 A(\alpha) = 0.$$

Here, the algebraic sign of k^2 has been chosen so, that $A(\alpha)$ gets a periodic solution. Such a general solution would be

$$A_k(\alpha) = a_k \cos k\alpha + b_k \sin k\alpha,$$

in which, because the angle α has a period of 2π, k has to be a non-negative integer: $k = 0, 1, 2, 3, \ldots$ Negative k do not give different solutions, because

$$a_k \cos k\alpha = a_k \cos(-k\alpha),$$
$$b_k \sin k\alpha = -b_k \sin(-k\alpha).$$

The other equation, in the function $R(r)$, is harder to solve. A test solution is a power law:

$$R(r) = r^q.$$

Substitution yields

$$r\left(rq\,(q-1)\, r^{q-2} + qr^{q-1} \right) - k^2 r^q = 0 \implies q^2 - k^2 = 0 \implies q^2 = k^2.$$

This works for positive $q = 2, 3, \ldots$ and negative $q = -1, -2, \ldots$. For $q = 1$ we find

$$r - k^2 r = 0 \implies k^2 = 1 = q^2.$$

For $q = 0$, besides the trivial constant solution, the non-trivial solution $R(r) = \ln r$ is found:

$$r\left(r \cdot -\frac{1}{r^2} + \frac{1}{r} \right) - k^2 \ln r = 0 \implies k = 0.$$

Thus we obtain the general solution

$$R_k(r) = \begin{cases} 1 \text{ or } \ln r & \text{if } k = 0, \\ r^k \text{ or } r^{-k} & \text{if } k = 1, 2, \ldots. \end{cases}$$

We see that, if we require the solution to exist at the origin $r = 0$, we need the *first* solutions, obtaining

$$V_1(\alpha, r) = a_0 + \sum_{k=1}^{\infty} r^k \left(a_k \cos k\alpha + b_k \sin k\alpha \right),$$

but if we require existence — or, at least, good behaviour — at infinity[6] $r \to \infty$, we need the *second* solutions,

$$V_2(\alpha, r) = a_0 + b_0 \ln r + \sum_{k=1}^{\infty} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha \right). \tag{2.3}$$

There is a clear similarity here to the three-dimensional (spherical co-ordinates) case.

2.4 Spherical, geodetic, ellipsoidal co-ordinates

In physical geodesy we use geometrical and physical concepts side by side. For example, co-ordinates of place can be given in the form (X, Y, Z), which

---

[6]In fact, $\lim\limits_{r\to\infty} V_2 \to \infty$ but $\lim\limits_{r\to\infty} \dfrac{\partial V_2}{\partial r} = 0$.

The Laplace equation and its solutions
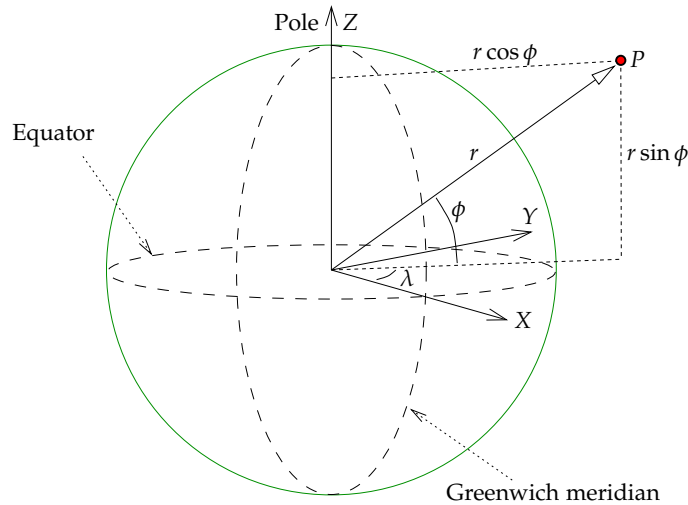
Figure 2.3. Definition of spherical co-ordinates.

are in principle geometric — except for the physical assumption that the origin of the co-ordinate system is in the centre of mass of the Earth.

As the Earth is not precisely a sphere but rather an oblate ellipsoid of revolution, one cannot use geographical co-ordinates as if they were spherical co-ordinates. Because the flattening of the Earth — some 0.3% — cannot be ignored, this difference is significant. The connection between spherical co-ordinates (r, ϕ, λ) and rectangular ones (X, Y, Z) is the following:

$$
\begin{aligned}
X &= r \cos \phi \cos \lambda, \\
Y &= r \cos \phi \sin \lambda, \\
Z &= r \sin \phi.
\end{aligned}
\tag{2.4}
$$

Here ϕ and λ are geocentric latitude and (ordinary, i.e., geocentric or geodetic or geographic) longitude. r is the distance from the Earth's centre. The X axis points in the direction of the Greenwich meridian. See figure 2.3.

On the Earth's surface, these spherical co-ordinates are not very useful because of the Earth's flattening, but in space, spherical co-ordinates are

Spherical, geodetic, ellipsoidal co-ordinates
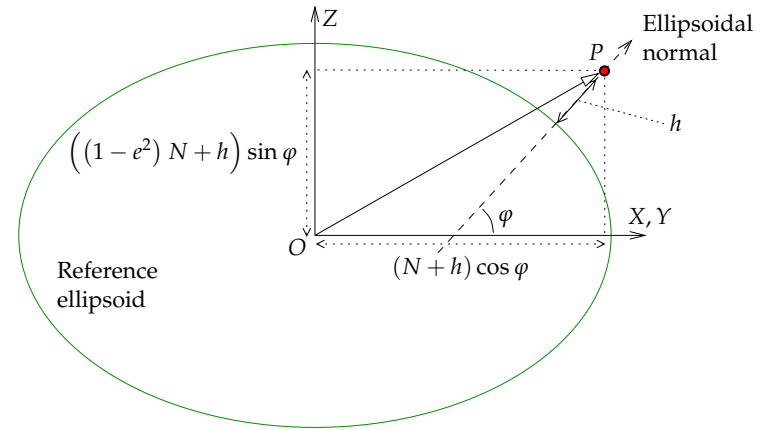
Figure 2.4. Definition of geodetic co-ordinates.

much used. On the Earth's surface, on the other hand, most often *geodetic* — or geographical — co-ordinates φ, λ, h are used:

$$
\begin{aligned}
X &= (N + h) \cos \varphi \cos \lambda, \\
Y &= (N + h) \cos \varphi \sin \lambda, \\
Z &= (N + h - e^2 N) \sin \varphi,
\end{aligned}
\tag{2.5}
$$

where

$$
N(\varphi) = \frac{a}{\sqrt{1 - e^2 \sin^2 \varphi}} = \frac{a^2}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}.
\tag{2.6}
$$

The quantity N defined in equation 2.6 is the East-West direction, or *transversal*, radius of curvature of the reference ellipsoid. In the equation, a is the equatorial radius of the reference ellipsoid used, $e^2 = \frac{a^2 - b^2}{a^2}$ is the square of the so-called *first eccentricity*[7], and in equations 2.5, h is the height of the point above the reference ellipsoid, see figure 2.4.

Converting rectangular co-ordinates into geodetic ones is easiest to do iteratively, although the literature also offers closed formulas.

---

[7]The parameter is connected to the Earth's flattening f through the equation $e^2 = 2f - f^2$.

Spherical co-ordinates and geodetic, i.e., geographical, co-ordinates are considerably different. In latitude, the difference is up to 11 minutes of arc, or almost 20 km. This maximum is attained for latitudes $\pm 45°$.

In theoretical work one also uses *ellipsoidal co-ordinates* u and β. The co-ordinate β is called the *reduced latitude*. The relationship with rectangular co-ordinates is

redukoitu leveysaste

$$
\begin{aligned}
X &= \sqrt{u^2 + E^2} \cos \beta \cos \lambda, \\
Y &= \sqrt{u^2 + E^2} \cos \beta \sin \lambda, \\
Z &= u \sin \beta.
\end{aligned}
\tag{2.7}
$$

If the semi-major axis of the Earth ellipsoid is a and its semi-minor axis b, it follows — exercise! — that $E^2 = a^2 - b^2$.

2.5 The Laplace equation in spherical co-ordinates

The Laplace equation transformed to spherical co-ordinates reads (see appendix E for a simple geometric proof):

$$
\Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{2}{r} \frac{\partial V}{\partial r} + \frac{1}{r^2} \frac{\partial^2 V}{\partial \phi^2} - \frac{\tan \phi}{r^2} \frac{\partial V}{\partial \phi} + \frac{1}{r^2 \sin^2 \phi} \frac{\partial^2 V}{\partial \lambda^2} = 0,
\tag{2.8}
$$

in which ϕ is the (geocentric) latitude, λ is the longitude, and r is the distance from the origin or centre of the Earth.

We shall here not derive the solution of this equation by separation of variables, as it is pretty complicated and can be found in ready form in the literature (Heiskanen and Moritz, 1967, section 1-9). What is significant is, that the solution looks somewhat similar to the solution in rectangular co-ordinates presented earlier, section 2.2. The basis solutions of the Laplace equation are

$$
V_{n,1}(\phi,\lambda,r) = r^n Y_n(\phi,\lambda), \qquad V_{n,2}(\phi,\lambda,r) = \frac{Y_n(\phi,\lambda)}{r^{n+1}},
\tag{2.9}
$$

where the first is again *nonphysical* in outer space, because, unlike the true geopotential, these expressions grow to infinity for $r \to \infty$.

pintapallofunktiot In the above equations, the functions Y_n are called *surface spherical harmonics*,

whereas the functions V_n are *solid spherical harmonics*. The latter are *harmonic* functions everywhere in space except at the origin (2.9, rightmost equation) or at infinity (leftmost, physically unrealistic equation).

avaruus-pallofunktiot

The functions Y_n are

$$
Y_n(\phi,\lambda) = \sum_{m=0}^{n} P_{nm}(\sin \phi) \, (a_{nm} \cos m\lambda + b_{nm} \sin m\lambda).
\tag{2.10}
$$

The functions P_{nm} are so-called Legendre functions, on which more later on. With the help of expression 2.10 we obtain, by using the second, physically realistic alternative from equations 2.9, the following solution or *series expansion* for the potential V in space:

pallofunktio-kehitelmä

$$
V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin \phi) \, (a_{nm} \cos m\lambda + b_{nm} \sin m\lambda).
\tag{2.11}
$$

The coefficients a_{nm} and b_{nm} are called the coefficients of the spherical-harmonic expansion, shortly *spectral coefficients*. Together they describe the function V, in somewhat the same way that the Fourier coefficients v_{jk} do in rectangular co-ordinates in equation 2.2. The subscripts n and m are called *degree* and *order*.

asteluku, järjestysluku

Often we will be using a somewhat freeer notation for the functions Y_n. For example, if we expand the disturbing potential T into spherical harmonics, we shall use the notation $T_n(\phi,\lambda)$ for its surface harmonics. Similarly, $\Delta g_n(\phi,\lambda)$ is the surface harmonic of the gravity anomaly Δg for degree n, and so on.

2.6 Dependence on height

From the above equation 2.9 one sees that for different values of the degree n the function V_n has a different dependence on the distance r from the Earth's centre, or equivalently, on the height $H = r - R$, if by R we denote the radius of the Earth sphere. The dependence is

$$
V_n(\phi,\lambda,r) = \frac{Y_n(\phi,\lambda)}{r^{n+1}}.
$$

On the surface of the Earth sphere

$$
V_n(\phi,\lambda,R) = \frac{Y_n(\phi,\lambda)}{R^{n+1}}.
$$

Therefore, we may write

$$V_n(\phi, \lambda, r) = \left(\frac{R}{r}\right)^{n+1} V_n(\phi, \lambda, R) = \left(\frac{R+H}{R}\right)^{-(n+1)} V_n(\phi, \lambda, R) =$$

$$= \left(1 + \frac{H}{R}\right)^{-(n+1)} V_n(\phi, \lambda, R) \approx \exp\left(-\frac{H}{R}(n+1)\right) V_n(\phi, \lambda, R).$$

We see that the attenuation of the potential with height is *again exponential*, and the harmonic degree number n appears in the exponent, as did also the wave number in rectangular geometry, see equation 2.2 and figure 2.1. The analogy works.

3 Legendre functions and spherical harmonics

3.1 Legendre's functions

In the above equations, the functions P are so-called *Legendre*[1] functions that pop up whenever we solve a Laplace-like equation in spherical co-ordinates. There exist various effective, so-called recursive algorithms, e.g., the following (this one is only for ordinary Legendre polynomials $P_n = P_{n0}$):

$$nP_n(t) = -(n-1)P_{n-2}(t) + (2n-1)tP_{n-1}(t). \tag{3.1}$$

Similar equations exist also for the functions $P_{nm}, m > 0$. There are even alternatives to choose from, though most equations are complicated. One should be careful that in their computation, the factorials don't go overboard. Already 30! (factorial of 30) is a larger number than computers can handle even as 64-bit integers... not to mention 360!. Heiskanen and Moritz (1967) equation 1-62, contrary to what is stated there, is *not* directcy suitable for computer use!

The first Legendre polynomials are listed in table 3.1. Higher polynomials than this are rarely needed in manual computation.

---

[1]Adrien-Marie Legendre (1752–1833) was a French mathematician known for his work on number theory, statistics — he invented independently from Gauss the method of least squares — and on elliptical functions. His name is inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

Table 3.1. Legendre polynomials. $t = \sin\phi$.

Function of t	Expressed in sines and cosines
$P_0(t) = 1$	$P_0(\sin\phi) = 1$
$P_1(t) = t$	$P_1(\sin\phi) = \sin\phi$
$P_2(t) = \frac{3}{2}t^2 - \frac{1}{2}$	$P_2(\sin\phi) = -\frac{3}{4}\cos 2\phi + \frac{1}{4}$
$P_3(t) = \frac{5}{2}t^3 - \frac{3}{2}t$	$P_3(\sin\phi) = -\frac{5}{8}\sin 3\phi + \frac{3}{8}\sin\phi$
$P_4(t) = \frac{1}{8}\left(35t^4 - 30t^2 + 3\right)$	
$P_5(t) = \frac{1}{8}\left(63t^5 - 70t^3 + 15t\right)$	
$P_6(t) = \frac{1}{16}\left(231t^6 - 315t^4 + 105t^2 - 5\right)$	

For comparison, also the Fourier *basis functions* (like, in a more complicated way, also sines and cosines!)

$$F_j(x) = \exp\left(2\pi i j \frac{x}{L}\right),$$
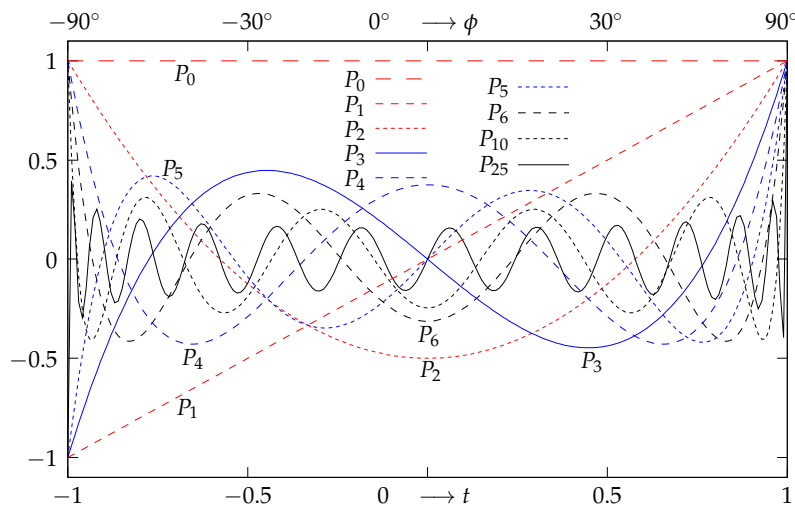
Figure 3.1. A number of Legendre polynomials $P_0(t) \ldots P_{25}(t)$ as functions of the argument $t = \sin\phi$.

Table 3.2. Associated Legendre functions.

Function of t	Trigonometric function
$P_{11}(t) = \sqrt{1 - t^2}$	$P_{11}(\sin\phi) = \cos\phi$
$P_{21}(t) = 3t\sqrt{1 - t^2}$	$P_{21}(\sin\phi) = 3\sin\phi\cos\phi$
$P_{22}(t) = 3\left(1 - t^2\right)$	$P_{22}(\sin\phi) = 3\cos^2\phi$
$P_{31}(t) = \frac{3}{2}\left(5t^2 - 1\right)\sqrt{1 - t^2}$	$P_{31}(\sin\phi) = \frac{3}{2}\left(5\sin^2\phi - 1\right)\cos\phi$
$P_{32}(t) = 15t\left(1 - t^2\right)$	$P_{32}(\sin\phi) = 15\sin\phi\cos^2\phi$
$P_{33}(t) = 15\left(1 - t^2\right)^{3/2}$	$P_{33}(\sin\phi) = 15\cos^3\phi$

in which $i^2 = -1$, can be computed recursively:

$$F_{j+1}(x) = F_j(x) \cdot F_1(x).$$

3.1.1 Properties of Legendre polynomials

1. The even polynomials are mirror symmetric through the origin, $P_n(-t) = P_n(t)$ — or equivalently $P_n\big(\sin(-\phi)\big) = P_n(\sin\phi)$ — and the odd ones are antisymmetric, i.e., $P_n(-t) = -P_n(t)$ or $P_n\big(\sin(-\phi)\big) = -P_n(\sin\phi)$.

2. Looking at figure 3.1 we see, that the polynomials $P_n(t)$ go, on the whole interval $t \in [-1, 1]$, or $\phi \in [-90°, 90°]$, precisely n times through zero.

3. As the values in the end points $t = \pm 1, \phi = \pm 90°$ are ± 1, it follows that there are precisely $n + 1$ "algebraic-sign intervals", i.e., intervals of t or ϕ on which the polynomial assumes only positive or only negative values.

3.1.2 Properties of associated Legendre functions

Of the associated Legendre functions $P_{nm}, m \neq 0$ we give several in table 3.2 for illustration.

One defining equation for these is

$$P_{nm}(t) = \left(1 - t^2\right)^{\frac{m}{2}}\frac{d^m P_n(t)}{dt^m}. \tag{3.2}$$

Legendren
liitännäisfunktiot
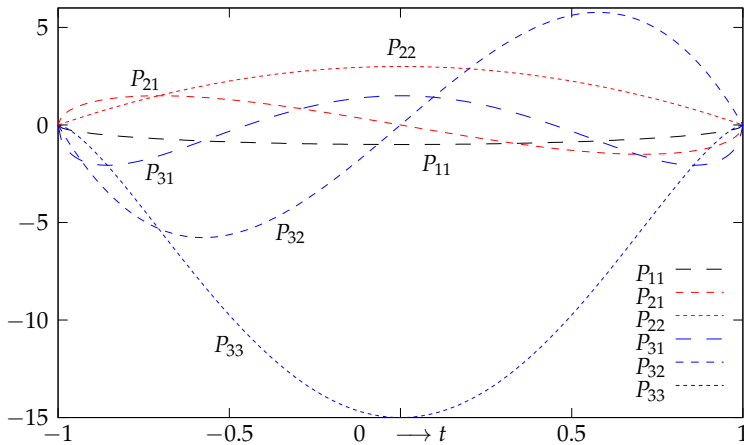
Figure 3.2. Associated Legendre functions.

1. Also the associated Legendre functions are either mirror symmetric through the origin, $P_{nm}(-t) = P_{nm}(t)$ or $P_{nm}(\sin(-\phi)) = P_{nm}(\sin\phi)$, or antisymmetric, i.e., $P_{nm}(-t) = -P_{nm}(t)$ or $P_{nm}(\sin(-\phi)) = -P_{nm}(\sin\phi)$, depending on the values of m and n.

2. Figure 3.2 suggests that the polynomials $P_{nm}(t)$ go on $t \in [-1,1]$, or $\phi \in [-90°, 90°]$, precisely $n - m$ times through zero. This is indeed the case.

3. As also the values in the end points $t = \pm 1$, $\phi = \pm 90°$ are zero, it follows that there are precisely $n - m + 1$ "algebraic-sign intervals".

3.1.3   Surface spherical harmonics

Starting from equation 2.10 we may write

$$Y_n(\phi, \lambda) = \sum_{m=0}^{n} (a_{nm} P_{nm}(\sin\phi) \cos m\lambda + a_{n,-m} P_{nm}(\sin\phi) \sin m\lambda) =$$

$$= \sum_{m=-n}^{n} a_{nm} Y_{nm}(\phi, \lambda),$$

where now m runs from $-n$ to $+n$. Here

$$Y_{nm}(\phi, \lambda) = \begin{cases} P_{nm}(\sin\phi) \cos m\lambda & \text{if } m \geq 0, \\ P_{n|m|}(\sin\phi) \sin |m| \lambda & \text{if } m < 0. \end{cases}$$

These are the *surface spherical harmonics of degree n and order m.*

Such surface spherical harmonics come in three kinds:

- *Zonal functions*: $m = 0$. These functions depend only on latitude.

- *Sectorial functions*: $m = n$. the *algebraic signs* of these functions depend only on longitude and not on latitude. The functions themselves however *do* depend on both latitude and longitude!

- *Tesseral functions*: $0 < m < n$. These functions, the algebraic sign of which changes with both latitude and longitude, form a checkerboard pattern on the surface of the sphere, if the positive values are painted grey and the negative ones white (Lat. *tessera* = a tile, as used in a mosaic).

Every function will, on the interval $\sin\phi \in [-1, +1]$, go precisely $n - m$ times through zero. Every function is either symmetric or antisymmetric through the origin as a function of ϕ or of $t = \sin\phi$.

Spherical harmonics thus represent a wave phenomenon of sorts. The are however not wave functions (sines or cosines), the connection to those is complicated at least. It nevertheless makes sense to speak of their *wavelength*.

In figure 3.3 is depicted how the algebraic signs of the different spherical harmonics behave on the Earth's surface — and above. Note that this is a perspective sketch and not all white and grey areas are visible!   etumerkit

When looking in equation 2.10 at the expressions $\cos m\lambda$ and $\sin m\lambda$, we observe that around a full circle, the equator, $0 \leq \lambda < 2\pi$, they go precisely $2m$ times through zero. The "semi-wavelength" is thus   puoliaallonpituus

$$\frac{2\pi R}{2m} = \pi \frac{R}{m},$$

where R is again the radius of the Earth.

(a)
Zonals: $P_{5,0}(\sin\phi)$

(b)
Sectorials:
$P_{6,6}(\sin\phi)\cos 6\lambda$

(c)
Tesserals:
$P_{11,6}(\sin\phi)\cos 6\lambda$
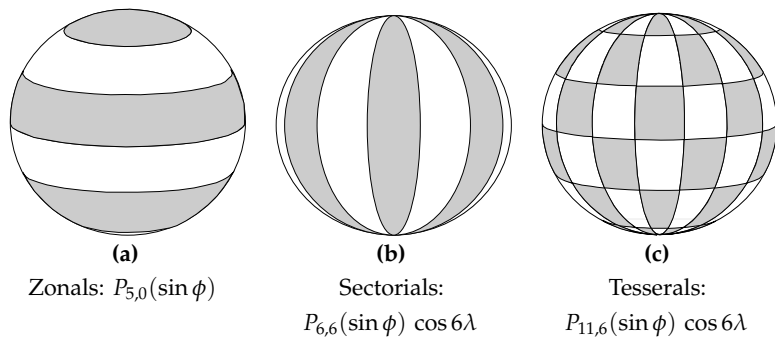
Figure 3.3. The algebraic signs of spherical harmonics on the Earth's surface. Grey means positive, white negative. The functions "wave" in a sine or cosine function like fashion.

A similar formula applies also for functions $P_{nm}(\sin\phi)$: as the function passes through zero $n-m$ times on the interval — from pole to pole — $-\frac{\pi}{2} < \phi < \frac{\pi}{2}$,
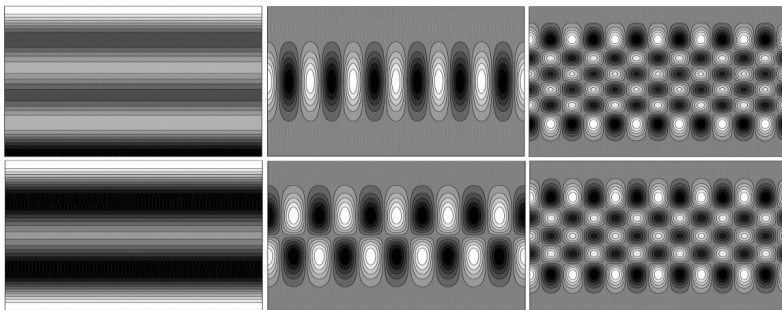
Figure 3.4. Surface spherical harmonics as maps. Horizontal axis $\lambda \in [0, 360°)$, vertical axis $\phi \in [-90°, 90°]$. Functions depicted are

$$
\begin{array}{llll}
P_{50}(\sin\phi) & P_{66}(\sin\phi)\cos 6\lambda & P_{11,6}(\sin\phi)\cos 6\lambda \\
P_{40}(\sin\phi) & P_{65}(\sin\phi)\cos 5\lambda & P_{10,6}(\sin\phi)\cos 6\lambda
\end{array}.
$$

Table 3.3. Semi-wavelengths for different degrees and orders of spherical harmonics.

m or $n-m$	Semi-wavelength (km)	In degrees
10	2000	18°
40	500	4°.5
180	111	1°
360	55	0°.5 = 30′
1800	11	0°.1 = 6′
10 800	1.85	1′

it follows that also here, the semi-wavelength is

$$
\frac{\pi R}{n-m}.
$$

If we plug various values for m and the expression $n-m$ into this, we obtain table 3.3.

This table also gives the *resolution* that can be achieved with a spherical-harmonic expansion, i.e., in how detailed a fashion the expansion can describe the Earth's gravity field. The expansions available today, like the model EGM2008, go to degree $n = 2159$; the "sharpness" of a geopotential image based on them is thus 9 km. Models based on satellite orbit perturbations often extend only to degree 40, meaning that only details the size of continents — order of magnitude 500 km — will be visible. On the other hand, experimental spherical-harmonic expansions of the topography go even up to degree 10 800 (Balmino et al., 2012).

3.2 Symmetry properties of the spherical-harmonic expansion

We recapitulate the spherical-harmonic expansion given at the beginning, equation 2.11:

$$
V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)\left(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda\right). \quad (2.11)
$$

3.2.1 Dependence on latitude ϕ

It is seen that the dependence on ϕ only works through the associated Legendre function $P_{nm}(\sin\phi)$. This function can, in terms of mirror symmetry between the Northern and Southern hemispheres, be either *symmetric* in ϕ, or *antisymmetric* in ϕ. This means that either (symmetric case)

$$P_{nm}(\sin\phi) = P_{nm}(\sin(-\phi))$$

or (antisymmetric case)

$$P_{nm}(\sin\phi) = -P_{nm}(\sin(-\phi)).$$

Equivalently it means, with $t = \sin\phi$, that either (symmetric case)

$$P_{nm}(t) = P_{nm}(-t)$$

or (antisymmetric case)

$$P_{nm}(t) = -P_{nm}(-t).$$

Which case applies, depends on the values of both n and m. To figure it out, one can look at, e.g., equation 3.2:

$$P_{nm}(t) = (1-t^2)^{\frac{m}{2}} \frac{d^m P_n(t)}{dt^m}. \tag{3.2}$$

We need to answer two questions:

1. For which values n is the polynomial $P_n(t)$ symmetric, for which is it antisymmetric in t? For this, you need to look at the recursive algorithm for computation of the polynomials, eq. 3.1. We already know that $P_0(t) = 1$ is symmetric, and that $P_1(t) = t$ is antisymmetric. The rule for other n values follows recursively (or you could cheat by looking at table 3.1).

2. What does *differentiation* $\frac{d}{dt}$ do to the symmetry or antisymmetry of the function?
   (Multiplication by $\sqrt{1-t^2} = \cos\phi$ changes nothing, as this factor is symmetric in t or ϕ.)

So, in order to make expansion 2.11 *mirror symmetric* between Northern and Southern hemispheres, one has to set those coefficients a_{nm}, b_{nm} for which the corresponding P_{nm} is antisymmetric, to zero. In other words, the corresponding terms vanish from the expansion. The coefficients, and terms, remaining are those for which the corresponding P_{nm} is symmetric.

In tableau 3.4 we give a code fragment in the octave rapid-prototyping language to plot an arbitrary surface spherical harmonic, e.g., in order to visually judge its symmetry properties. Don't believe, test.

3.2.2 Dependence on longitude λ

This dependence works though the "Fourier basis functions" $\cos m\lambda$ and $\sin m\lambda$. The interesting property here is *rotational symmetry*: does the spherical-harmonic expansion 2.11 change when we change λ?

We see immediately that, for $m \neq 0$, there will be dependence on λ if any coefficient a_{nm}, b_{nm} is non-zero. So, in order to obtain rotational symmetry, all coefficients a_{nm}, b_{nm} for values $m > 0$ must be suppressed: $a_{11} = b_{11} = a_{21} = b_{21} = a_{22} = b_{22} = \cdots = 0$.

Of the remaining coefficients, we can say that for $m = 0$, $\sin m\lambda = 0$ identically, so the coefficients $b_{00}, b_{10}, b_{20}, \ldots$ simply don't matter. They may be any value, including zero. The coefficients $a_{00}, a_{10}, a_{20}, \ldots$ however *do* matter, as for $m = 0$, $\cos m\lambda = 1$ identically. So we obtain as the *rotationally symmetric expansion*

$$V(\phi, \lambda, r) = V(\phi, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} a_n P_n(\sin\phi),$$

in which $P_n = P_{n0}$ are the familiar Legendre polynomials, and $a_n \overset{\text{def}}{=} a_{n0}$.

3.3 Orthogonality, orthonormality of Legendre functions

Legendre's *polynomials* are *orthogonal*: the integral — formally, a scalar product of vectors —

$$\int_{-1}^{+1} P_n(t) P_{n'}(t)\, dt = \begin{cases} \frac{2}{2n+1} & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases} \tag{3.3}$$

Tableau 3.4. Plotting a surface spherical-harmonic map.

```
% Plotting surface spherical harmonics
phi=linspace(-90,90,72);
lab=linspace(0,360,144);
[f,l]=meshgrid(phi,lab);
n=5; m=-3;
leg=legendre(n,sin(phi.*pi./180));
if m >= 0
cs=cos(m.*lab.*pi./180);
else
cs=sin(abs(m).*lab.*pi./180);
end
v=leg(abs(m)+1,:)'*cs;
contourf(l,f,v')
xlabel('Longitude', 'FontSize', 16)
ylabel('Latitude', 'FontSize', 16)
str=sprintf('Surface spherical harmonic n=%d, m=%d', n, m)%
title(str, 'FontSize', 20)
axis ([0 360 -90 90])
colorbar()
print('legendre2D.jpg','-djpg')
```

This orthogonality is just one example of a more general way to look at functions and integrals over functions. There exists a useful analogy with vector spaces; see appendix B.

Alternatively we may write, on the surface of a *unit sphere* σ, using a

parametrization (ψ, α):

$$\iint_\sigma P_n(\cos\psi)\, P_{n'}(\cos\psi)\ d\sigma = \int_0^{2\pi}\int_0^\pi P_n(\cos\psi)\, P_{n'}(\cos\psi)\ \sin\psi\, d\psi\, d\alpha =$$

$$= -2\pi\int_{+1}^{-1} P_n(t)\, P_{n'}(t)\ dt =$$

$$= 2\pi\int_{-1}^{+1} P_n(t)\, P_{n'}(t)\ dt,$$

in which $t = \cos\psi$ and $d\sigma = \sin\psi\, d\psi\, d\alpha$. So, we have

$$\iint_\sigma P_n(\cos\psi)\, P_{n'}(\cos\psi)\ d\sigma = \begin{cases} \frac{4\pi}{2n+1} & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases} \tag{3.4}$$

in which ψ is the angular distance from some point on the surface of the sphere. Equation 3.4 tells us that Legendre polynomials are mutually *orthogonal* if the vectorial product is defined as an integral over the surface of the unit sphere σ. Alternatively we may define also *fully normalized* Legendre polynomials

$$\overline{P}_n(\cos\psi) = \sqrt{2n+1}P_n(\cos\psi), \tag{3.5}$$

in which case the now modified scalar product — the mean square value over the unit sphere — is

$$\frac{1}{4\pi}\iint_\sigma \overline{P}_n(\cos\psi)\, \overline{P}_{n'}(\cos\psi)\ d\sigma = \begin{cases} 1 & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases}$$

showing the polynomials now to be *orthonormal*[2]. Similarly fully normalized *associated* Legendre functions also exist (without proof, see Heiskanen and Moritz 1967, page 31):

$$\overline{P}_{nm}(\cos\psi) = \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}}P_{nm}(\cos\psi). \tag{3.6}$$

---

[2]And also

$$\frac{1}{2}\int_{-1}^{+1} \overline{P}_n(t)\, \overline{P}_{n'}(t)\ dt = \begin{cases} 1 & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases}$$

again, the mean square value.

Legendre functions and spherical harmonics

In this case, the orthonormal functions look like

$$\overline{Y}_{nm}(\psi, \alpha) = \begin{cases} \overline{P}_{nm}(\cos\psi)\, \cos m\alpha & \text{if } m \geq 0, \\ \overline{P}_{n|m|}(\cos\psi)\, \sin|m|\,\alpha & \text{if } m < 0. \end{cases}$$

The scalar product that applies is

$$\frac{1}{4\pi}\iint_\sigma \overline{Y}_{nm}(\psi, \alpha)\, \overline{Y}_{n'm'}(\psi, \alpha)\, d\sigma = \begin{cases} 1 & \text{if } n = n' \text{ and } m = m', \\ 0 & \text{otherwise.} \end{cases}$$

3.4 Low-degree spherical harmonics

The potential field of a point mass is (equation 1.4):

$$V = \frac{GM}{r}.$$

The corresponding term in the potential expansion 2.11 for degree $n = 0$ is

$$V_0 = \frac{1}{r}a_{00}P_0\,(\sin\phi) = \frac{a_{00}}{r},$$

from which

$$a_{00} = GM.$$

So, a_{00} represents the force field of a point mass or spherically symmetric mass distribution centred at the origin. The higher spherical-harmonic coefficients are "perturbations" on top of this.

The expansion for the degree-one coefficients looks as follows:

$$V_1(\phi, \lambda, r) = \frac{1}{r^2}\left(a_{11}\cos\phi\cos\lambda + b_{11}\cos\phi\sin\lambda + a_{10}\sin\phi\right).$$

Write this in vector form using the expression for the location vector

$$\mathbf{r} = (r\cos\phi\cos\lambda)\,\mathbf{i} + (r\cos\phi\sin\lambda)\,\mathbf{j} + (r\sin\phi)\,\mathbf{k}$$

— in which $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is an orthonormal basis of the Euclidean space $\mathbb{R}^3$ — yielding

$$V_1(\mathbf{r}) = \frac{1}{r^3}\big\langle(a_{11}\mathbf{i} + b_{11}\mathbf{j} + a_{10}\mathbf{k})\cdot\mathbf{r}\big\rangle.$$

Remember that the potential field of a *dipole* is

$$V(\mathbf{r}) = \frac{G}{r^3}\langle\mathbf{d}\cdot\mathbf{r}\rangle,$$

in which $\mathbf{d}$ is the dipole moment. Comparison yields

$$a_{11}\mathbf{i} + b_{11}\mathbf{j} + a_{10}\mathbf{k} = G\mathbf{d},$$

i.e., the first-degree $n = 1$ spherical-harmonic coefficients represent the Earth's gravitational field's *dipole moment*.

Every mass element dm of our Earth may be taken to consist of

- a *monopole* at the origin of the co-ordinate system, size dm, and
- a *dipole*, size $\mathbf{r}\cdot dm$, in which $\mathbf{r}$ is the location vector of the mass element.

In that case we may compute the dipole moment of the whole Earth by integration:

$$\mathbf{d}_\oplus = \iiint_{\text{Earth}} \mathbf{r}\, dm = \iiint_{\text{Earth}} \rho\mathbf{r}\, d\mathcal{V} = M_\oplus \cdot \mathbf{r}_{\text{centre of mass}},$$

in which, by definition, $\mathbf{r}_{\text{centre of mass}}$ is the location of the *centre of mass* of the Earth! From this follows that, if we choose our co-ordinate system so, that the origin is in the centre of mass of the Earth, the spherical-harmonic coefficients a_{11}, b_{11}, a_{10} *vanish*. If the equations of motion of satellites are formulated in a certain co-ordinate system, like in the case of GPS satellites the WGS84 system, then the origin of the system is automatically in the centre of mass of the Earth, and the degree one spherical-harmonic coefficients are really zero.

The same logic applies to higher degrees of spherical harmonics. The degree two coefficients represent the so-called *quadrupole moment* of the Earth — corresponding to her inertial tensor — etc.

3.5 Splitting a function into degree constituents

There exists a useful integral equation for surface spherical harmonics, if the function itself f on the surface of the sphere has been given. The equation is (Heiskanen and Moritz, 1967, equation 1-71, but using our notation $Y_n \longrightarrow f_n$):
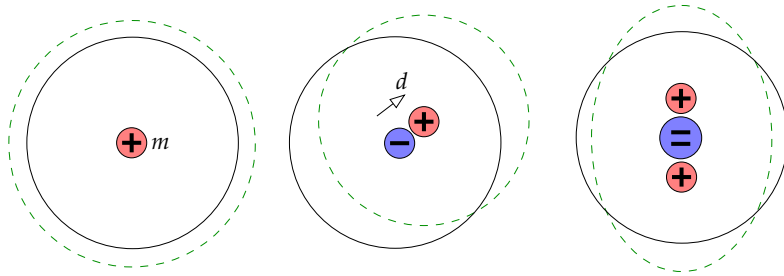
Figure 3.5. Monopole, dipole, and quadrupole, at the Earth's centre, and their effects on the geoid.

$$f_n(\phi, \lambda) = \frac{2n+1}{4\pi} \iint_\sigma f(\phi', \lambda') P_n(\cos \psi)\, d\sigma', \qquad (3.7)$$

where ψ is the geocentric angular distance between evaluation point (ϕ, λ) and moving data or integration point (ϕ', λ'), see figure 8.2. In this *degree* asteosuusyhtälö *constituent equation* 3.7 there is a certain similarity with the projection or coefficient computation formula B.10. Nevertheless, here we don't have a computation of spectral coefficients, but of "spectral constituent functions" f_n.

We bring to mind the core property of the functions f_n,

$$f(\phi, \lambda) = \sum_{n=0}^{\infty} f_n(\phi, \lambda)$$

on the surface of the sphere.

For the proof, we choose as the "north pole" of the co-ordinate system the point (ϕ, λ); then, $\phi' = 90° - \psi$. By writing (see equation 2.11):

$$f(\phi', \lambda') = \sum_{n=0}^{\infty} \sum_{m=0}^{n} P_{nm}(\sin \phi') \left(a_{nm} \cos m\lambda' + b_{nm} \sin m\lambda' \right)$$

and substituting this into the degree constituent equation 3.7, we obtain, by exploiting the orthogonality of the Legendre functions, on the right-hand

side of the degree constituent equation:

$$
\begin{aligned}
I_R &= \frac{2n+1}{4\pi} \iint_\sigma f(\phi', \lambda') P_n(\cos \psi)\, d\sigma' = \\
&= \frac{2n+1}{4\pi} a_{n0} \iint_\sigma P_n^2(\cos \psi)\, d\sigma' = \\
&= \frac{2n+1}{4\pi} a_n \int_0^{2\pi} P_n^2(\cos \psi) \cdot \left(\sin \psi \int_0^{2\pi} d\lambda' \right) d\psi \\
&= \frac{2n+1}{4\pi} a_n \int_{-1}^{+1} P_n^2(t) \cdot \left(\sin \psi \int_0^{2\pi} d\lambda' \right) \cdot \frac{1}{\sin \psi}\, dt = \\
&= \frac{2n+1}{4\pi} \cdot 2\pi a_n \cdot \frac{2}{2n+1} = a_n,
\end{aligned}
$$

where we used the notation $a_n \overset{\text{def}}{=} a_{n0}$ as well as equation 3.3.

On the left-hand side of the degree constituent equation we obtain, because, on the assumed north pole, $\phi = 90°$ and thus $\sin \phi = 1$:

$$
\begin{aligned}
I_L &= f_n(\phi, \lambda) \overset{\text{def}}{=} Y_n(90°, \lambda) = \\
&= \sum_{m=0}^{n} P_{nm}(1) \left(a_{nm} \cos m\lambda + b_{nm} \sin m\lambda \right) = P_{n0}(1)\, a_{n0} = a_n,
\end{aligned}
$$

by using

$$P_{n0}(1) = 1,$$
$$P_{nm}(1) = 0 \text{ if } m \neq 0.$$

As this applies for every point (ϕ, λ) — and note that the values a_n depend on this choice! — it follows that the degree constituent equation 3.7 *is generally true.*

3.6 Spectral representations of various quantities

3.6.1 The potential

Starting from equation 2.11 we write the following *spectral expansion* of the geopotential V in space:

$$V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \left(\frac{R}{r} \right)^{n+1} V_n(\phi, \lambda), \qquad (3.8)$$

Space domain   Frequency domain

$$V(\phi,\lambda,R) \xrightarrow{\quad V_n(\phi,\lambda) = \frac{2n+1}{4\pi} \iint_\sigma V(\phi',\lambda',R)\,P_n(\cos\psi)\,d\sigma' \quad} = \sum_{n=0}^{\infty} V_n(\phi,\lambda)$$

$\Big\downarrow$ (hard) $\times$ (easy) $\Big\downarrow$

$$V(\phi,\lambda,r) \xleftarrow{\quad\quad \Sigma \quad\quad} = \sum_{n=0}^{\infty} \left(\tfrac{R}{r}\right)^{n+1} V_n(\phi,\lambda)$$
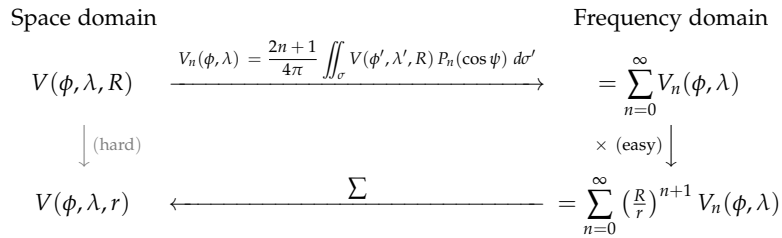
Figure 3.6. Vertically shifting the potential field V, in the space and frequency domains. Spherical geometry.

in which the spectral constituents V_n are the familiar Y_n from equation 2.10:

$$V_n(\phi,\lambda) = \sum_{m=0}^{n} P_{nm}(\sin\phi)\left(a_{nm}^V \cos m\lambda + b_{nm}^V \sin m\lambda\right) = \sum_{m=-n}^{n} v_{nm} Y_{nm}(\phi,\lambda),$$

in which the basis functions

$$Y_{nm}(\phi,\lambda) = \begin{cases} P_{nm}(\sin\phi)\,\cos m\lambda & \text{if } m \geq 0, \\ P_{n|m|}(\sin\phi)\,\sin|m|\,\lambda & \text{if } m < 0, \end{cases} \tag{3.9}$$

and coefficients

$$v_{nm} = \begin{cases} a_{nm}^V & \text{if } m \geq 0, \\ b_{n|m|}^V & \text{if } m < 0. \end{cases} \tag{3.10}$$

On the Earth's surface ($r = R$) we obtain

$$V(\phi,\lambda,R) = \sum_{n=0}^{\infty} V_n(\phi,\lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} v_{nm} Y_{nm}(\phi,\lambda). \tag{3.11}$$

We may summarize the relationships found in *commutative diagram.* 3.6. Again, like in section 2.2 for rectangular geometry, it is seen that the shift of the potential function V from the spherical level R to the level $r = R + H$ is essentially easier in the frequency domain — the degree constituents V_n — than it is in the space domain.

3.6.2 Gravitation

In the *Neumann*[3] boundary-value problem we solve a function V of which the normal derivative, $\frac{\partial V}{\partial n}$, is given on the surface of a body or a closed surface in space.

In the case of a spherical body, we may assume $\frac{\partial}{\partial n} = \frac{\partial}{\partial r}$ and work with spherical-harmonic expansions. By differentiating equation 3.8 we obtain

$$\frac{\partial V}{\partial r} = -\sum_{n=0}^{\infty} \frac{n+1}{r}\left(\frac{R}{r}\right)^{n+1} V_n(\phi,\lambda) = -\sum_{n=0}^{\infty} \frac{n+1}{R}\left(\frac{R}{r}\right)^{n+2} V_n(\phi,\lambda).$$

$$\tag{3.12}$$

On the spherical Earth's surface this means

$$\left.\frac{\partial V}{\partial r}\right|_{r=R} = -\sum_{n=0}^{\infty} \frac{n+1}{R} V_n(\phi,\lambda).$$

If we also write on the Earth's surface for the *gravitation*

$$g(\phi,\lambda,R) \stackrel{\text{def}}{=} \left.\frac{\partial V}{\partial r}\right|_{r=R} \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} g_n(\phi,\lambda),$$

it follows by analogy that

$$g_n(\phi,\lambda) = -\frac{n+1}{R} V_n(\phi,\lambda),$$

and conversely, that

$$V_n(\phi,\lambda) = -\frac{R}{n+1} g_n(\phi,\lambda).$$

As a result of this we obtain the *spectral representation of the solution* to a certain Neumann problem:

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} V_n(\phi,\lambda) = -R\sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{g_n(\phi,\lambda)}{n+1}. \tag{3.13}$$

We may write in an analogue fashion

$$g(\phi,\lambda,R) = \sum_{n=0}^{\infty} g_n(\phi,\lambda) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty}\sum_{m=-n}^{n} g_{nm} Y_{nm}(\phi,\lambda),$$

---

[3]Carl Gottfried Neumann (1832–1925) was a German mathematician.

and comparison with the equivalent expression 3.11 for the potential yields consistently

$$g_{nm} = -\frac{n+1}{R} v_{nm},\qquad(3.14)$$

see equation 3.10. This is an interesting result worth thinking about:

1. Firstly, note how simple the connection 3.14 between potential v_{nm} and gravitation g_{nm} has become in the frequency domain!

2. Secondly, if there are available, over the whole surface area of the Earth, measurement values of gravitational acceleration g, we may derive from these the spherical-harmonic coefficients g_{nm} and the degree constituent functions $g_n(\phi, \lambda)$ using the method explained earlier. In this way we can then obtain the solution by means of equation 3.13 for the whole exterior geopotential field! This is the basic idea of geopotential — or geoid — determination, from the spectral perspective.

3.7   Often used spherical-harmonic expansions

Of the existing global spherical-harmonic expansions must be mentioned the already outdated EGM96. It was developed by researchers from Ohio State University using very extensive, mostly gravimetric, data collected by the American NIMA (*National Imagery and Mapping Agency*, the former *Defense Mapping Agency* DMA, the current NGA, *National Geospatial-Intelligence Agency*). This expansion goes up to harmonic degree 360. Its standard presentation is[4]

$$V = \frac{GM_\oplus}{r}\left(1 + \sum_{n=2}^{360}\left(\frac{a}{r}\right)^n\sum_{m=0}^{n}\overline{P}_{nm}(\sin\phi)\;\left(\overline{C}_{nm}\cos m\lambda + \overline{S}_{nm}\sin m\lambda\right)\right).$$
$$(3.15)$$

This form of presentation — the algebraic sign in front of the expansion, which starts from degree number $n = 2$, the number one inside the parentheses which represents the point mass in the origin equal in magnitude to the total mass of the Earth, and the "fully normalized" coefficients $\overline{C}$ and $\overline{S}$ —

täysin normalisoidut kertoimet

----

[4]Note that here is used a, the equatorial radius of the Earth's reference ellipsoid, not R, and ϕ, the *geocentric* latitude. The co-ordinates (r, ϕ, λ) form a spherical co-ordinate system.

has been already for some time been an industry standard in the global research community in the field of computing spherical-harmonic expansions as models of the Earth's gravitational field. A pioneer has been Professor Richard H. Rapp at Ohio State University, which is why the models are often called OSU models.

Generally in these models the lower terms — $2 \le n \le 20$ — are derived primarily from analysis of satellite orbit perturbations. Because of this, the models are in a co-ordinate system with the origin in the Earth's centre of mass. This explains the absence of the degree-one coefficients, as exposited earlier.

ratahäiriöt

The higher coefficients again — $20 < n \le 360$ — were before the year 2000 mostly the result of the analysis of gravimetric data (over land) and satellite radar altimetric data (over the ocean). After the launches of the gravimetric satellite missions CHAMP, GRACE and GOCE, and as a result of their measurements, nowadays at least degree number interval $20 < n \le 200$ is the product of space geodesy. Only the still higher degree numbers — the new model EGM2008 (Pavlis et al., 2008, 2012) goes up to degree 2159 — are still due to terrestrial data.

In table 3.5 we give the first and last coefficients of the EGM96 model, the last and best model from the time just before the satellite gravity missions. The values tabulated are $n, m, \overline{C}_{nm}, \overline{S}_{nm}$ and the mean errors (standard deviations) of both coefficients from their computation. Note that all $\overline{S}_{n0}$ vanish!

Sometimes also non-normalized coefficients are used, and we write

$$V = \frac{GM_\oplus}{r}\left(1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n\sum_{m=0}^{n}P_{nm}(\sin\phi)\;\left(J_{nm}\cos m\lambda + K_{nm}\sin m\lambda\right)\right).$$
$$(3.16)$$

Then we use the notation $J_n \stackrel{\text{def}}{=} J_{n0}$, and J_2 is the most important, the parameter of the Earth's gravity field describing the flattening of the Earth. Based on

equations 3.5, 3.6, the relationship with the parameters $\overline{C}, \overline{S}$ is

$$\left\{ \begin{array}{c} J_{n0} \\ K_{n0} \end{array} \right\} = -\sqrt{2n+1} \left\{ \begin{array}{c} \overline{C}_{n0} \\ \overline{S}_{n0} \end{array} \right\},$$

$$\left\{ \begin{array}{c} J_{nm} \\ K_{nm} \end{array} \right\} = -\sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}} \left\{ \begin{array}{c} \overline{C}_{nm} \\ \overline{S}_{nm} \end{array} \right\}, m \neq 0. \tag{3.17}$$

3.8 Ellipsoidal harmonics

The Laplace differential equation 1.12 may be written, and solved, instead of in spherical co-ordinates, in ellipsoidal co-ordinates. The result is known as an *ellipsoidal-harmonic expansion*[5]. They are little used, because the math needed is more complicated. Also ellipsoidal co-ordinates are mostly only theoretically interesting and not in any broad use within geodesy.

The form of presentation is

$$V(u, \beta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)} P_{nm}(\sin\beta) \left(a_{nm}^{e} \cos m\lambda + b_{nm}^{e} \sin m\lambda\right), \tag{3.18}$$

in which $Q_{nm}(\mathbf{z})$ are the so-called *Legendre functions of the second kind*, sampled in table 3.6. Though the general argument $\mathbf{z}$ is complex, equation 3.18 gives a real result for real-valued coefficients a_{nm}^{e}, b_{nm}^{e}.

Those interested in the derivation of the above equation can find it in Heiskanen and Moritz (1967) or other textbooks on potential theory. Heiskanen and Moritz give a slightly different form to the equation, the auxiliary equations needed for the normalization used here can be found on their pages 66-67.

3.8.1 The scaling to standard form of the expansion

Assume $a_{10}^{e} = a_{11}^{e} = b_{11}^{e} = 0$, i.e., the vanishing of the dipole moment[6].

---

[5]This expansion for the ellipsoid of revolution differs from the expansion into Lamé functions found for the triaxial ellipsoid.

[6]Strictly speaking this works only in case of a spherical-harmonic expansion, or in the limit $E \to 0$.

Tableau 3.5. Coefficients and mean errors of the EGM96 spherical-harmonic expansion.

n	m	$\overline{C}_{nm}$	$\overline{S}_{nm}$	$\overline{C}_{nm}$ mean error	$\overline{S}_{nm}$ mean error
2	0	-0.484165371736E-03	0.000000000000E+00	0.35610635E-10	0.00000000E+00
2	1	-0.186987635955E-09	0.119528012031E-08	0.10000000E-29	0.10000000E-29
2	2	0.243914352398E-05	-0.140016683654E-05	0.53739154E-10	0.54353269E-10
3	0	0.957254173792E-06	0.000000000000E+00	0.18094237E-10	0.00000000E+00
3	1	0.904627768605E-06	0.248513158716E-06	0.13965165E-09	0.13645882E-09
3	2	0.904627768605E-06	-0.619025944205E-06	0.10962329E-09	0.11182866E-09
3	3	0.721072657057E-06	0.141435626958E-05	0.95156281E-10	0.93285090E-10
4	0	0.539873863789E-06	0.000000000000E+00	0.10423678E-09	0.00000000E+00
4	1	-0.536321616971E-06	-0.473440265853E-06	0.85674404E-10	0.82408489E-10
4	2	0.350694105785E-06	0.662671572540E-06	0.16000186E-09	0.16390576E-09
4	3	0.990771803829E-06	-0.200928369177E-06	0.84657802E-10	0.82662506E-10
4	4	-0.188560802735E-06	0.308853169333E-06	0.87315359E-10	0.87852819E-10
5	0	0.685323475630E-07	0.000000000000E+00	0.54383090E-10	0.00000000E+00
5	1	-0.621012128528E-07	-0.944226127525E-07	0.27996887E-09	0.28082882E-09
5	2	0.652438297612E-06	-0.323349612668E-06	0.23747375E-09	0.24356998E-09
5	3	-0.451955406071E-06	-0.214847190624E-06	0.17111636E-09	0.16810647E-09
5	4	-0.295301647654E-06	0.496658876769E-07	0.11981266E-09	0.11849793E-09
5	5	0.174971983203E-06	-0.669384278219E-06	0.11642563E-09	0.11590031E-09
6	0	-0.149957994714E-06	0.000000000000E+00	0.14497863E-09	0.00000000E+00
6	1	-0.760879384947E-07	0.262890545501E-07	0.22415138E-09	0.21957296E-09
6	2	0.481732442832E-07	-0.373728201347E-06	0.27697363E-09	0.28105811E-09
6	3	0.571730990516E-07	0.902694517163E-08	0.19432407E-09	0.18682712E-09
6	4	-0.862142660109E-07	-0.471408154267E-06	0.15229150E-09	0.15328004E-09
6	5	-0.267133325490E-06	-0.536488432483E-06	0.89838470E-10	0.87820905E-10
6	6	0.967616121092E-08	-0.237192006935E-06	0.11332010E-09	0.11518036E-09
...					
360	358	0.709604781531E-10	0.691761006753E-10	0.50033977E-10	0.50033977E-10
360	359	0.183971631467E-10	-0.310123632209E-10	0.50033977E-10	0.50033977E-10
360	360	-0.447516389678E-24	-0.830224945525E-10	0.50033977E-10	0.50033977E-10

We can also show that in the expansion 3.18 the first coefficient is

$$a^e_{00} = \frac{GM_\oplus}{E} \arctan \frac{E}{b}$$

and the expansion specialized for a rotationally symmetric field becomes

$$V(u,\beta) = \sum_{n=0}^{\infty} V_n(u,\beta) = \sum_{n=0}^{\infty} \frac{Q_n\left(i\frac{u}{E}\right)}{Q_n\left(i\frac{b}{E}\right)} a^e_{n0} P_n(\sin\beta),$$

which yields now

$$V_0(u) = \frac{Q_0\left(i\frac{u}{E}\right)}{Q_0\left(i\frac{b}{E}\right)} \frac{GM_\oplus}{E} \arctan \frac{E}{b},$$

the gravitational potential of the field constituent of ellipsoidal degree zero.

With the substitutions (Heiskanen and Moritz, 1967, page 66)

$$Q_0\left(i\frac{u}{E}\right) = -i \arctan \frac{E}{u}, \qquad Q_0\left(i\frac{b}{E}\right) = -i \arctan \frac{E}{b} \qquad (3.19)$$

we obtain

$$V_0(u) = \frac{GM_\oplus}{E} \arctan \frac{E}{u}.$$

This corresponds to the "central field" of a spherical harmonic expansion $\frac{GM}{r}$.

Using this, we may "scale" equation 3.18 as follows — by substituting the above identities 3.19 and moving the constant expression

$$\left(\frac{GM_\oplus}{E} \arctan \frac{E}{b}\right)^{-1}$$

into the new, standard scaled coefficients $\overline{C}^e_{nm}, \overline{S}^e_{nm}$:

$$V(u,\beta,\lambda) = \frac{GM_\oplus}{E} \arctan \frac{E}{u} \cdot$$
$$\cdot \left[1 + \sum_{n=2}^{\infty} \sum_{m=0}^{n} \frac{\arctan \frac{E}{b}}{\arctan \frac{E}{u}} \frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)} \overline{P}_{nm}(\sin\beta) \left(\begin{array}{c} \overline{C}^e_{nm} \cos m\lambda + \\ + \overline{S}^e_{nm} \sin m\lambda \end{array} \right) \right],$$

an ellipsoidal-harmonic expansion that agrees with the spherical-harmonic expansion 3.15, with the total mass of the Earth outside brackets and the coefficients dimensionless. This equation has however apparently not been used for any geopotential determination.

≡↑🖾▦🔍📕✧

Table 3.6. Legendre functions of the second kind.

$$Q_0(\mathbf{z}) = \frac{1}{2} \ln \frac{\mathbf{z}+1}{\mathbf{z}-1}$$

$$Q_1(\mathbf{z}) = \frac{\mathbf{z}}{2} \ln \frac{\mathbf{z}+1}{\mathbf{z}-1} - 1$$

$$Q_2(\mathbf{z}) = \frac{3\mathbf{z}^2 - 1}{4} \ln \frac{\mathbf{z}+1}{\mathbf{z}-1} - \frac{3\mathbf{z}}{2}$$

$$Q_3(\mathbf{z}) = \frac{5\mathbf{z}^3 - 3\mathbf{x}}{4} \ln \frac{\mathbf{z}+1}{\mathbf{z}-1} - \frac{5\mathbf{z}^2}{2} + \frac{2}{3}$$

$$(n+1)Q_{n+1}(\mathbf{z}) - (2n+1)\mathbf{z}Q_n(\mathbf{z}) + nQ_{n-1}(\mathbf{z}) = 0$$

$$Q_{mn}(\mathbf{z}) = \left(1 - \mathbf{z}^2\right)^{m/2} \frac{d^m}{d\mathbf{z}^m} Q_n(\mathbf{z})$$

3.8.2 Equivalence of the Rapp and ellipsoidal expansions

We can demonstrate the equivalence of spherical expansions 3.15, or 3.16, and the ellipsoidal expansion 3.18, if the flattening of the Earth $\to 0$, and thus also $a, b \to a$, $\beta \to \phi$, and $u \to r$. We assume that Heiskanen and Moritz (1967) equation 1-112,

$$\lim_{E\to 0} \frac{Q_{nm}\left(i\frac{u}{E}\right)}{Q_{nm}\left(i\frac{b}{E}\right)} = \left(\frac{a}{r}\right)^{n+1}$$

is valid. Substitution into equation 3.18 yields

$$V(u,\beta,\lambda) = V(r,\phi,\lambda) =$$
$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} \left(\frac{a}{r}\right)^{n+1} P_{nm}(\sin\phi) \left(a^e_{nm} \cos m\lambda + b^e_{nm} \sin m\lambda\right),$$

which, with the identifications $a^e_{00} = \frac{GM}{a}, a^e_{10} = a^e_{11} = b^e_{11} = 0$ and, with relations 3.17,

$$\left\{ \begin{array}{c} a^e_{n0} \\ b^e_{n0} \end{array} \right\} = -\frac{GM_\oplus}{a} \left\{ \begin{array}{c} J_{n0} \\ K_{n0} \end{array} \right\} = \frac{GM_\oplus}{a} \sqrt{(2n+1)} \left\{ \begin{array}{c} \overline{C}_{n0} \\ \overline{S}_{n0} \end{array} \right\},$$

$$\left\{ \begin{array}{c} a^e_{nm} \\ b^e_{nm} \end{array} \right\} = -\frac{GM_\oplus}{a} \left\{ \begin{array}{c} J_{nm} \\ K_{nm} \end{array} \right\} = \frac{GM_\oplus}{a} \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}} \left\{ \begin{array}{c} \overline{C}_{nm} \\ \overline{S}_{nm} \end{array} \right\}, m \neq 0$$

otherwise, corresponds to equations 3.15 and 3.16 for spherical harmonics.

3.8.3 Advantages of using an ellipsoidal harmonic expansion:

1. The expression for the normal gravitational potential is in this form of presentation simple, see Heiskanen and Moritz (1967) equation 2-56. A

≡↑🖾▦🔍📕✧

spherical-harmonic expansion of the same field would instead require theoretically an infinite number of coefficients. In practice, 3 to 4, i.e., an expansion up to J_6 or J_8, will suffice.

2. The convergence will be more rapid, as less terms are needed. This is because, due to the Earth's flattening, the equator is some $23\,\mathrm{km}$ farther from the Earth's centre than the poles. Therefore, especially high degree spherical harmonics will have difficulty converging efficiently both at the poles and in the equatorial region. This problem is worst for very high degree expansions (e.g., Wenzel, 1998). Already for a degree number of 360, the semi-wavelength of a spherical harmonic will be only $55\,\mathrm{km}$!

3.8.4   Disadvantage of using an ellipsoidal harmonic expansion:

Evaluation of an ellipsoidal-harmonic expansion is clearly more laborious, i.e., expensive, than a spherical-harmonic one, in terms of computer resources.

Self-test questions

1. How does separation of variables work?

2. Why does solving the Laplace equation require *boundary conditions*?

3. What are the harmonic degree and harmonic order in a spherical-harmonic expansion? How do they relate to the *resolution* of the expansion on the Earth's surface?

4. What types of spherical harmonics are there? Describe their dependence on latitude and longitude.

5. How many times does a surface spherical harmonic $Y_{nm}(\phi, \lambda)$ change its algebraic sign traveling along a meridian from South pole to North pole? How many times when travelling around the Earth along the equator?

6. What does it mean if it is said that two functions are mutually orthogonal? Give a possible definition of the *scalar product* of two functions.

7. How does the *attenuation* of spherical harmonics with height behave? Why does a gravimetric satellite that tries to map the gravity field of the Earth at a high resolution, fly in as low an orbit as possible?

8. What does the *degree constituent equation* describe?

9. Which spherical-harmonic coefficients are associated with the *dipole moment* of the Earth's mass distribution? Why are they missing from tableau 3.5?

Exercise 3 − 1:   Attenuation with height of a spherical-harmonic expansion

If

$$
V(\phi, \lambda, r) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=0}^{n} P_{nm}(\sin\phi) \, (a_{nm}\cos m\lambda + b_{nm}\sin m\lambda) =
$$
$$
= \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} V_n(\phi, \lambda, R), \tag{3.20}
$$

we may call

$$
\frac{V_n(\phi, \lambda, r)}{V_n(\phi, \lambda, R)} = \left(\frac{R}{r}\right)^{n+1}
$$

the *attenuation factor* of the potential with height.

Differentiation with respect to r yields

$$
\frac{\partial V_n(\phi, \lambda, r)}{\partial r} = -\frac{n+1}{R}\left(\frac{R}{r}\right)^{n+2} V_n(\phi, \lambda, R), \tag{3.21}
$$

or, because, on the Earth's surface, similarly

$$
\left.\frac{\partial V_n(\phi, \lambda, r)}{\partial r}\right|_{r=R} = -\frac{n+1}{R} V_n(\phi, \lambda, R), \tag{3.22}
$$

it follows that the attenuation factor for the *attraction* is the ratio of expressions 3.21 and 3.22:

$$
\left(\frac{R}{r}\right)^{n+2}.
$$

1. Draw a *log-linear graph* of the attenuation factors of both the potential and the attraction for values $n = 0, \ldots, 100$, by hand or by machine. Choose $R = 6378\,\text{km}$, $r = 7378\,\text{km}$ — i.e., the height is 1000 km above the Earth's surface.

2. Based on this, if the satellite is 1000 km above the Earth's surface, for what degree number n will the accelerations $\frac{\partial V_n}{\partial r}$ caused by the attraction on satellite level be less than 1% of what they are on the Earth's surface?

3. For what n value will they be less than $10^{-4} \times$ of what they are on the Earth's surface?

Exercise 3 − 2: Symmetries of spherical harmonics

See equation 3.20 above. In it, $P_{nm}(\sin\phi) = P_{nm}(t)$, $t = \sin\phi$ *is only a function of latitude* ϕ. When ϕ runs from the South pole through the equator to the North pole, values $-90° \ldots 0° \ldots +90°$, parameter t will run through values $-1 \ldots 0 \ldots +1$ on the interval $[-1, 1]$.

For these Legendre functions exists closed expression:

$$P_{nm}(t) = \left(1 - t^2\right)^{m/2} \frac{d^m}{dt^m} P_n(t),\qquad (3.2)$$

in which the $P_n(t)$ are the ordinary Legendre polynomials:

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n}\left(\left(t^2 - 1\right)^n\right).$$

1. Note first, that

   (a) Differentiating a symmetric function of t will produce an anti-symmetric function, and vice versa.

   (b) The function $(t^2 - 1)$ and its powers are symmetric.

   (c) Thus: for even n values $P_n(t) = +P_n(-t)$ — , i.e., P_n is *symmetric* between the Northern and Southern hemispheres — and for odd n values $P_n(t) = -P_n(-t)$, i.e., P_n is *anti-symmetric* between hemispheres.

   (d) Similarly, for even n, $P_n(\sin\phi) = +P_n(\sin(-\phi))$, and for odd n, $P_n(\sin\phi) = -P_n(\sin(-\phi))$.

Question: what is the corresponding rule for the functions P_{nm}, i.e., for which (n, m) values is it symmetric and for which values anti-symmetric?

(a) Fill out the diagram ($n = 0 \ldots 5$, $m = 0 \ldots n$) with symbols either 'S' or 'A' in each framed cell:

$n =$	0	1	2	3	4	5
$m = 0$						
1						
2						
3						
4						
5						

(b) What is the logic of symmetry?

2. If the field is *symmetric* between Northern and Southern hemispheres, i.e., $V(\phi, \lambda, r) = V(-\phi, \lambda, r)$, which of the spherical-harmonic coefficients a_{nm}, b_{nm} drop out of the series expansion? *Why?*
   (Hint: see the example formulas and graphs for $P_{nm}(\sin\phi)$ in the lecture notes and try to guess a general rule. Then, verify.)

3. The same question if the potential is *rotationally symmetric* about the Earth's rotation axis, i.e., $V(\phi, \lambda, r) = V(\phi, r)$.

Exercise 3 − 3: Algebraic-sign domains of spherical harmonics

We have seen in section 3.1 that the associated Legendre functions $P_{nm}(t)$ have precisely $n - m + 1$ algebraic-sign intervals on their interval of definition $\phi \in [-90°, 90°]$. We can show that the functions $\cos m\lambda$ and $\sin m\lambda$ each have $2m$ zero crossings and $2m$ algebraic-sign intervals on their domain of definition $\lambda \in [0, 360°)$. How many *algebraic-sign domains* — grey or white areas, visible or occluded, in figure 3.3 — are there for each surface spherical

harmonic

$$Y_{nm}(\phi, \lambda) = \begin{cases} P_{nm}(\sin\phi) \cos m\lambda & \text{if } m \geq 0, \\ P_{n|m|}(\sin\phi) \sin |m| \lambda & \text{if } m < 0 \end{cases}$$

?

Exercise 3 − 4: Escape velocity

1. Given a spherically symmetric planet, mass GM, radius R, from the surface of which a cannon shoots projectiles at a flight velocity v. What is the minimum value for v — the so-called escape velocity — if it is desired that the projectile can travel to arbitrarily large distances from the planet, and never falls back? The kinetic energy of the projectile is $E_{\text{kin}} = \frac{1}{2}mv^2$, in which m is the projectile's mass.

2. Given, in *two-dimensional geometry*, a circularly symmetric planet, mass GM, radius R. The gravitational field of the planet is described by a potential V as given in section 2.3. What does V look like, based on these assumptions?

3. There is again a cannon on the edge of the circle planet. What can you now say about the escape velocity v (*don't* try to compute it!)?

4

The normal gravity field

4.1 The basic idea of a normal field

Like the figure of the Earth can be approximated by an ellipsoid of revolution, also the gravity field of the Earth can be just as well approximated by a field of which one equipotential surface, or *level surface*, is precisely this ellipsoid of revolution, the *reference ellipsoid*.

This brings a logical idea to mind: why not define intercompatibly a *reference ellipsoid*, a geopotential or *normal potential* — one of the equipotential surfaces of which is the reference ellipsoid — and a *gravity formula*, computed by taking the gradient of this normal potential?

After this we may define *anomalous* quantities, like disturbing potential and gravity anomaly, which then again will be intercompatible, while being numerically much smaller.

Let the normal potential be $U(x, y, z)$. Then, normal gravity will be

$$\gamma(x, y, z) = \left\| \vec{\gamma} \right\| = \left\| \vec{\nabla} U \right\| = -\langle \vec{\gamma} \cdot \mathbf{n} \rangle = -\frac{\partial U}{\partial n},$$

in which $\frac{\partial}{\partial n}$ denotes differentiation in the direction of the exterior surface normal $\mathbf{n}$ to a level surface of the normal field, itself also an ellipsoid, see figure 4.1. This direction will differ from the direction of the normal to the level surfaces of the gravity field, or *plumb line*, by precisely the plumb-line deflection, typically a very small angle.

luotiviiva, luotiviivan poikkeama
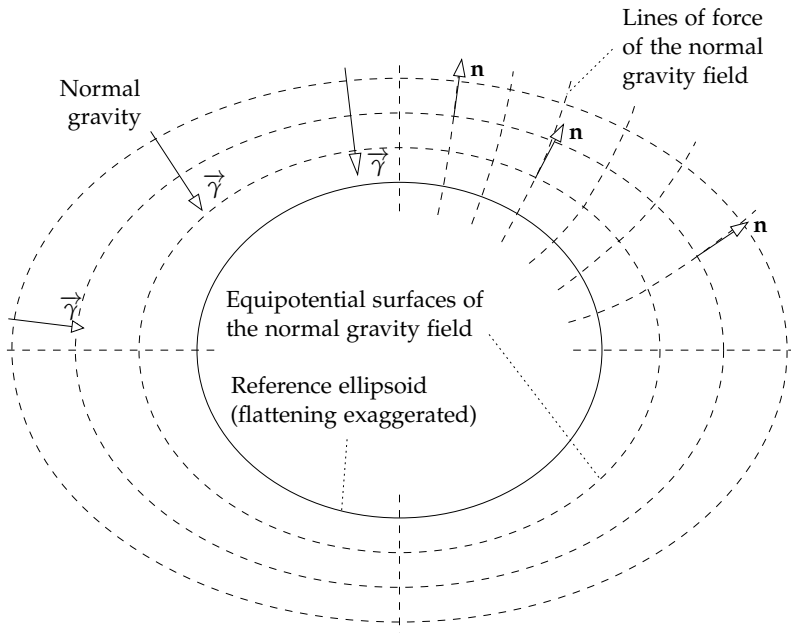
Figure 4.1. The normal gravity field of the Earth.

We shall see in the next section that the pseudo-force generated by the Earth's rotation may, in a system rotating along with the Earth, be described by a *rotational potential* Φ — also called centrifugal potential. Also the normal potential U is defined in such a way, that the rotational potential Φ is included in it: the normal potential is the reference potential of the *gravity field*, not the *gravitational field*. If we denote the normal *gravitational* potential by Ψ — a quantity rarely used in geodesy — then the normal *gravity* potential ("normal potential") U is

$$U = \Psi + \Phi,$$

in which Φ is the centrifugal potential. In other words: Ψ, like V, is defined in a non-rotating (inertial) system, whereas U, like W, is defined in a system that co-rotates with the Earth and is non-inertial. Just like also the word *gravity* refers to a force acting in a co-rotating system, when in an inertial

painovoima

Figure 4.2. Gravitation and centrifugal force.

system we use the word *gravitation*.

4.2   The centrifugal force and its potential

The rotation of the Earth affects the gravity field. In an inertial reference system one speaks of gravitation and gravitational potential V, on the Earth's surface however, in a non-inertial or *co-rotating* system, we talk of *gravity* and *gravity potential W*. They are different things, and the rotational motion and its centrifugal force are the cause of the difference. See figure 4.2.

To derive the equation for centrifugal force, write first

$$\mathbf{p} = X\mathbf{i} + Y\mathbf{j},$$

when the vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ form an orthonormal basis along the (X, Y, Z) axes. Then

$$p = \|\mathbf{p}\| = \sqrt{\langle \mathbf{p} \cdot \mathbf{p} \rangle} = \sqrt{X^2 + Y^2}.$$

Now the centrifugal force — or rather, acceleration — is, in metres per second squared,

$$\mathbf{f} = \omega_\oplus^2 \mathbf{p} = \omega_\oplus^2 \left(X\mathbf{i} + Y\mathbf{j} \right),$$

with $\omega_\oplus$ the rotation rate of the Earth in radians per second.

Here on Earth, gravity measurements are generally done with an instrument that is *at rest* with respect to the Earth's surface: it follows the rotation of the Earth. If the instrument moves, one must, in addition to the centrifugal force, take into account another pseudo-force: the Coriolis[1] force. Also fluids — water, air — on the Earth's surface, if they are at rest, sense only the centrifugal force. Currents also sense the Coriolis force, which deflects them sideways and causes the well known eddy phenomena in the oceans and atmosphere.

If we forget for the moment about the Coriolis force, we may describe the centrifugal force as *the gradient of a potential*. If we write for this *centrifugal potential*

$$\Phi = \frac{1}{2}\omega_\oplus^2\left(X^2 + Y^2\right),$$

we may directly calculate the gradient

$$\mathbf{f} = \vec{\nabla}\Phi = \frac{\partial\Phi}{\partial X}\mathbf{i} + \frac{\partial\Phi}{\partial Y}\mathbf{j} + \frac{\partial\Phi}{\partial Z}\mathbf{k} = \frac{1}{2}\omega_\oplus^2\cdot 2X\cdot\mathbf{i} + \frac{1}{2}\omega_\oplus^2\cdot 2Y\cdot\mathbf{j} + 0 = \omega_\oplus^2\left(X\mathbf{i} + Y\mathbf{j}\right),$$

which corresponds to the above centrifugal-force equation.

If we start out with the gravitational potential V and add to it the centrifugal potential Φ, we obtain the *gravity potential* W:

painovoima-potentiaali

$$W = V + \Phi.$$

We may also derive from the centrifugal potential Φ the following equation by differentiating it twice:

$$\Delta\Phi = \nabla^2\Phi = \vec{\nabla}\mathbf{f} = \frac{\partial}{\partial X}\omega_\oplus^2 X + \frac{\partial}{\partial Y}\omega_\oplus^2 Y + 0 = 2\omega_\oplus^2, \tag{4.1}$$

from which follows, with the Poisson equation 1.14,

$$\Delta W = -4\pi G\rho + 2\omega_\oplus^2, \tag{4.2}$$

[1]Gaspard-Gustave Coriolis (1792–1843) was a French mathematician, physicist and mechanical engineer. His name is inscribed on the Eiffel Tower, Eiffel Tower, 72 names.

the Poisson equation for the gravity potential.

The difference between gravitation and gravity is essential. The force, or acceleration, of gravitation $\mathbf{a} = \vec{\nabla}V$ is just an attractive force, whereas the acceleration of gravity $\mathbf{g} = \vec{\nabla}W$ is the vector sum of gravitation and centrifugal force. Attraction and centrifugal force act in the same fashion; the force is proportional to the mass of the test object, in other words, the acceleration is always the same independently of the mass of the test object. This is the famous *equivalence principle* (Galileo, Einstein), which has been proven to hold to very great precision. Especially we may mention the clever tests by the Hungarian baron Loránd Eötvös[2].

Water masses on the Earth's surface, like also the atmosphere — and on a vastly longer time scale also the "solid" Earth rock forming mountain ranges and ocean depths — react to gravity without distinguishing between attraction and centrifugal force. For this reason the sea surface coincides within a metre or so with an equipotential or *level surface* of the W function or *geopotential*. Also on dry land we measure heights from this surface, the *geoid* (Gauss: "mathematical figure of the Earth").

4.3 Level surfaces and plumb lines

Surfaces of the same gravity potential or *geopotential*, equipotential surfaces or *level surfaces* are the following surfaces:

tasopinnat

$$W(x, y, z) = \text{const.}$$

Let $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ again be an orthonormal basis along the (x, y, z) axes. Then, in the direction of the unit vector

$$\mathbf{e} = e_1\mathbf{i} + e_2\mathbf{j} + e_3\mathbf{k}$$

the potential changes as follows:

$$\frac{\partial W}{\partial e} = e_1\frac{\partial W}{\partial x} + e_2\frac{\partial W}{\partial y} + e_3\frac{\partial W}{\partial z},$$

[2]Loránd baron Eötvös de Vásárosnamény (1848–1919) was a Hungarian physicist and student of gravitation.
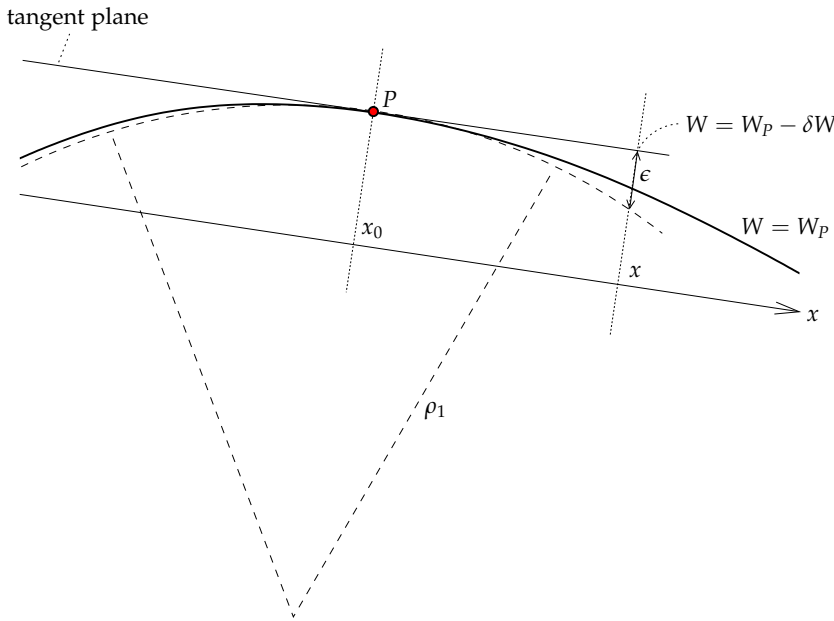
Figure 4.3. The curvature of level surfaces.

which vanishes if and only if

$$\left\langle \mathbf{e} \cdot \vec{\nabla} W \right\rangle = 0,$$

in other words, the potential is stationary only in directions that are perpendicular to the Earth's gravity vector

$$\vec{\nabla} W = \mathbf{g}.$$

> *Level surfaces and gravity vectors, or plumb lines, are always perpendicular to each other.*

4.3.1 Curvature of level surfaces

Given in point P a plane that in P has the same direction as the level surface, i.e., its *tangent plane*. If the local curvature of the level surface in the x

direction is ρ_1, and the x co-ordinate of point P is x_0, we may develop the distance between the surfaces in a Taylor series:

$$\epsilon \approx \frac{1}{2\rho_1} \left(x - x_0 \right)^2.$$

From this we obtain the difference in W values between the surfaces ($g \stackrel{\text{def}}{=} \|\mathbf{g}\|$):

$$\delta W = -\epsilon g \approx - \left(x - x_0 \right)^2 \frac{g}{2\rho_1}.$$

By differentiating (note that W here is now the geopotential *on the tangent plane*) we obtain

$$\frac{\partial^2}{\partial x^2} \delta W = \frac{\partial^2}{\partial x^2} W = W_{xx} = -\frac{g}{\rho_1}$$

from which

$$\rho_1 = -\frac{g}{W_{xx}},$$

in which the W that is being differentiated with respect to the x co-ordinate is its *restriction to the tangent or horizontal plane*.　　rajoittuma

By determining the curvature in the x direction

$$\kappa_1 \stackrel{\text{def}}{=} \frac{1}{\rho_1} = -\frac{W_{xx}}{g}, \tag{4.3}$$

and similarly in the y direction

$$\kappa_2 \stackrel{\text{def}}{=} \frac{1}{\rho_2} = -\frac{W_{yy}}{g}, \tag{4.4}$$

we obtain the mean or Germain[3] curvature, in most locations a positive number:

$$J = \frac{1}{2} \left(\kappa_1 + \kappa_2 \right) = -\frac{W_{xx} + W_{yy}}{2g},$$

and by using the Poisson equation 4.2,

$$\Delta W = W_{xx} + W_{yy} + W_{zz} = -4\pi G\rho + 2\omega_{\oplus}^2,$$

---

[3]Marie-Sophie Germain (1776−1831) was a brilliant French mathematician, number theorist and student of elasticity. She corresponded with Gauss, among others, on number theory, and did foundational work toward a proof of Fermat's last theorem. Her name is missing from the Eiffel Tower.

we obtain

$$-2gJ + W_{zz} = -4\pi G\rho + 2\omega_\oplus^2.$$

By using

$$W_{zz} = -\frac{\partial g}{\partial z} = -\frac{\partial g}{\partial H}$$

— in which H is the height co-ordinate — we obtain for the vertical gradient of gravity (Heiskanen and Moritz, 1967, equation 2-20):

$$\frac{\partial g}{\partial H} = -2gJ + 4\pi G\rho - 2\omega_\oplus^2,$$

an equation found by Ernst Heinrich Bruns.

4.4 Natural co-ordinates

Before the satellite era it was impossible to directly measure the geocentric co-ordinates X, Y and Z. Today this is possible, and we obtain at the same time the height from the reference ellipsoid h, a purely geometric quantity.

In earlier times one could measure only the direction of the plumb line as shown in figure 4.4, and the potential difference between an observation point and sea level. The direction of the plumb line **n** was measured astronomically: astronomical latitude Φ (don't confuse with the centrifugal potential) and astronomical longitude Λ. The third co-ordinate, the gravity potential difference $W(x, y, z) - W_0$ from the potential W_0 of sea level, was determined by levelling. The co-ordinates Φ, Λ and W are called *natural co-ordinates*.

Often, instead of the potential, *orthometric height* is used. Its definition is easy to understand if one writes

$$\frac{\partial W}{\partial H} = -g \implies dH = -\frac{1}{g}\,dW \implies H_P = -\int_{W_0}^{W_P} \frac{1}{g(W')}\,dW',$$

where the integral is taken along the plumb line of point P. $\frac{\partial}{\partial H}$ is the derivative in the direction of the plumb line, i.e., the local normal to the level surfaces. g is the acceleration of gravity along the plumb line as a function of place — or of geopotential level. In this case of orthometric heights, g is the *true* gravity inside the rock, which is a non-linear function of place
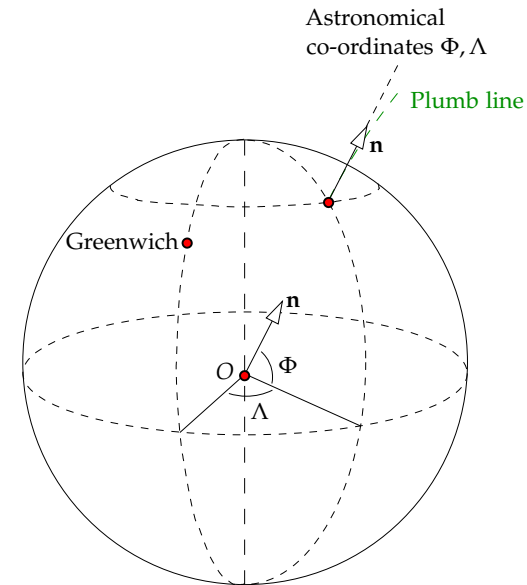
Figure 4.4. Natural co-ordinates Φ, Λ. Additionally, a natural height co-ordinate, e.g., the geopotential W, is needed.

and will also depend on rock density. This trickiness of their determination is a problem specific to orthometric heights. We will return to this later on (Heiskanen and Moritz, 1967 chapter 4).

Also the co-ordinates Φ, Λ and H form a natural co-ordinate system.

4.5 The normal potential in ellipsoidal co-ordinates

We already presented an equation 3.18 for the expansion of the geopotential into ellipsoidal harmonics. It is demanded of the normal potential U, that it is a constant on the reference ellipsoid $u = b$. We expand the centrifugal

potential Φ into ellipsoidal harmonics. We have

$$\Phi(u,\beta) = \frac{1}{2}\omega_\oplus^2\left(x^2+y^2\right) = \frac{1}{2}\omega_\oplus^2\left(u^2+E^2\right)\cos^2\beta =$$
$$= \frac{1}{2}\omega_\oplus^2\left(u^2+E^2\right)\left(1-\sin^2\beta\right) =$$
$$= \frac{1}{2}\omega_\oplus^2\left(u^2+E^2\right)\left(-\frac{2}{3}P_2(\sin\beta)+\frac{2}{3}P_0(\sin\beta)\right) =$$
$$= -\frac{1}{3}\omega_\oplus^2\left(u^2+E^2\right)\left(P_2(\sin\beta)-P_0(\sin\beta)\right).$$

Additionally we have, based on equation 3.18, for a rotationally symmetric gravitational potential Ψ:

$$\Psi(u,\beta) = \sum_{n=0}^\infty \Psi_n(u,\beta) = \sum_{n=0}^\infty \frac{Q_n\left(i\frac{u}{E}\right)}{Q_n\left(i\frac{b}{E}\right)} A_n P_n(\sin\beta). \tag{4.5}$$

Now

$$U(u,\beta) = \Psi(u,\beta) + \Phi(u,\beta).$$

On the reference ellipsoid $u=b$ we have as a requirement $U(b,\beta) = U_0$, which is possible only if

$$U_0 = A_0 + \frac{1}{3}\omega_\oplus^2\left(b^2+E^2\right),$$
$$0 = A_1,$$
$$0 = A_2 - \frac{1}{3}\omega_\oplus^2\left(b^2+E^2\right),$$
$$0 = A_n, n = 3,4,5,\dots$$

The quantity U_0 can be computed uniquely, if the Earth's mass GM and the measures of the reference ellipsoid a,b are known. The result, given in Heiskanen and Moritz (1967) as equation 2-61, is

$$U_0 = \frac{GM_\oplus}{E}\arctan\frac{E}{b} + \frac{1}{3}\omega_\oplus^2 a^2.$$

From this follows, with $a^2 = b^2 + E^2$:

$$A_0 = U_0 - \frac{1}{3}\omega_\oplus^2 a^2 = \frac{GM_\oplus}{E}\arctan\frac{E}{b}.$$

The normal *gravity* potential U is obtained as follows (remember the identities 3.19 and that $a^2 = b^2 + E^2$):

$$U(u,\beta) = \Psi(u,\beta) + \Phi(u,\beta) =$$
$$= \overbrace{\frac{GM_\oplus}{E}\arctan\frac{E}{u}}^{\Psi_0(u)} + \overbrace{\frac{1}{3}\omega_\oplus^2 a^2 \frac{Q_2\left(i\frac{u}{E}\right)}{Q_2\left(i\frac{b}{E}\right)}}^{A_2}\overbrace{\left(\frac{3}{2}\sin^2\beta - \frac{1}{2}\right)}^{P_2(\sin\beta)} + \overbrace{\frac{1}{2}\omega_\oplus^2\left(u^2+E^2\right)\cos^2\beta}^{\Phi(u,\beta)} =$$
$$= C_1(u)\sin^2\beta + C_2(u)\cos^2\beta,$$

in which C_1, C_2 are suitable functions of u. The function Ψ_0 is the term for $n = 0$ in expansion 4.5.

On the surface of the reference ellipsoid ($u = b$):

$$U(b,\beta) = \left(\frac{GM_\oplus}{E}\arctan\frac{E}{b} - \frac{1}{6}\omega_\oplus^2 a^2\right) + \frac{1}{2}\omega_\oplus^2 a^2\sin^2\beta + \frac{1}{2}\omega_\oplus^2 a^2\cos^2\beta =$$
$$= \frac{GM_\oplus}{E}\arctan\frac{E}{b} + \frac{1}{3}\omega_\oplus^2 a^2,$$

the constant U_0, as it better be!

4.6   Normal gravity on the reference ellipsoid

Without proof we mention that for *normal gravity* (the quantity $\gamma = \frac{\partial U}{\partial h}$) the following equation applies on the reference ellipsoid:

$$\gamma(\beta) = \frac{a\gamma_b\sin^2\beta + b\gamma_a\cos^2\beta}{\sqrt{a^2\sin^2\beta + b^2\cos^2\beta}}.$$

By substitution we find immediately, that γ_a is normal gravity on the equator ($\beta = 0$) and γ_b normal gravity on the poles ($\beta = \pm90°$).

Equations 2.5 and 2.7 yield

$$\tan\beta = \frac{\sin\beta}{\cos\beta} = \frac{\frac{Z}{b}}{\frac{\sqrt{X^2+Y^2}}{a}} = \frac{a}{b}\frac{Z}{\sqrt{X^2+Y^2}} = \frac{a}{b}\tan\phi$$

and

$$\tan\varphi = \frac{\sin\varphi}{\cos\varphi} = \frac{\frac{Z}{(1-e^2)N}}{\frac{\sqrt{X^2+Y^2}}{N}} = \frac{1}{1-e^2}\frac{Z}{\sqrt{X^2+Y^2}} = \frac{a^2}{b^2}\tan\phi,$$
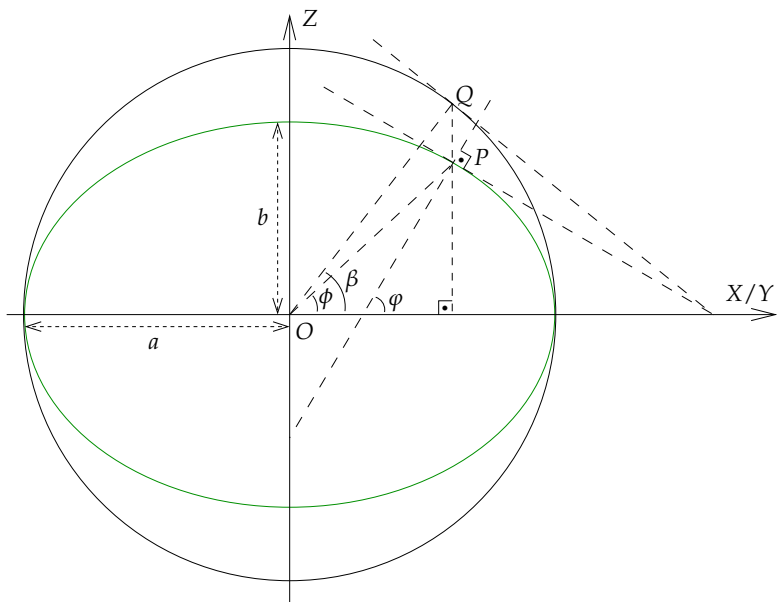
Figure 4.5. The geometry of the meridian ellipse and various types of latitude.

in which ϕ is the *geocentric* latitude, see equation 2.4. From this follows directly

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

in which the latitude angle φ is the geodetic or geographic latitude. (β is still the so-called *reduced latitude*). Now it is easy to show (exercise!) that

$$\gamma(\varphi) = \frac{a\gamma_a \cos^2 \varphi + b\gamma_b \sin^2 \varphi}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}}. \tag{4.6}$$

This is the famous *Somigliana–Pizzetti*[4] equation. These geodesists demonstrated for the first time that an "ellipsoidal" normal gravity field, which has the reference ellipsoid as one of its equipotential or level surfaces, exists

---

[4]Carlo Somigliana (1860 – 1955) was an Italian mathematician and physicist. Paolo Pizzetti (1860 – 1918) was an Italian geodesist.

exactly, and that also in geographical co-ordinates the gravity formula is a closed expression in latitude.

4.7 Numerical values and calculation formulas

When the reference ellipsoid has been chosen, we may calculate the normal potential and normal gravity corresponding to it. The fundamental quantities are

a the equatorial radius of the ellipsoid of revolution, i.e., its semi-major axis

f the flattening, $f = \frac{a-b}{a}$, in which b is the polar radius or semi-minor axis

$\omega_\oplus$ the rotation rate of the Earth

$\underline{GM}$ the total mass of the Earth (including the atmosphere).

Alternatively one may choose also γ_a, i.e., equatorial gravity.

Nowadays the most commonly used reference ellipsoid *cum* normal potential is GRS80, the Geodetic Reference System 1980:

$$a = 6\,378\,137\,\text{m},$$

$$\frac{1}{f} = 298.257\,222\,101,$$

$$\omega_\oplus = 7\,292\,115 \cdot 10^{-11}\,\text{s}^{-1},$$

$$GM = 3\,986\,005 \cdot 10^8\,\text{m}^3\,\text{s}^{-2}.$$

In reality f is not a defining constant of GRS80, but instead is used J_2, which is a defining quantity for the gravity field, see equation 3.16.

WGS84 (World Geodetic System 1984) used by the GPS system has a reference ellipsoid that is *almost* identical to that of GRS80.

The normal potential is (Heikkinen, 1981), in units, in the SI system, of metre

The normal gravity field

Numerical values and calculation formulas

[m] and second [s]:

$$U = 62\,636\,860.8500\, +$$

$$+ \left(\begin{array}{l} -9.780\,326\,77 - 0.051\,630\,75 \sin^2 \varphi\, - \\ -0.000\,227\,61 \sin^4 \varphi - 0.000\,001\,23 \sin^6 \varphi \end{array} \right) h\, +$$

$$+ \left(\begin{array}{l} 0.015\,438\,99 \cdot 10^{-4} - 0.000\,021\,95 \cdot 10^{-4} \sin^2 \varphi\, - \\ - 0.000\,000\,10 \cdot 10^{-4} \sin^4 \varphi \end{array} \right) h^2\, +$$

$$+ \left(-0.000\,024\,22 \cdot 10^{-8} + 0.000\,000\,07 \cdot 10^{-8} \sin^2 \varphi \right) h^3,$$

and normal gravity (note the minus sign, U is positive and diminishes going upward):

$$\gamma = -\frac{\partial U}{\partial h} =$$

$$= +\ 9.780\,326\,77 + 0.051\,630\,75 \sin^2 \varphi\, +$$

$$+\ 0.000\,227\,61 \sin^4 \varphi + 0.000\,001\,23 \sin^6 \varphi\, +$$

$$+ \left(\begin{array}{l} 0.030\,877\,98 \cdot 10^{-4} - 0.000\,043\,90 \cdot 10^{-4} \sin^2 \varphi\, - \\ - 0.000\,000\,200 \cdot 10^{-4} \sin^4 \varphi \end{array} \right) h\, +$$

$$+ \left(-0.000\,072\,65 \cdot 10^{-8} + 0.000\,000\,21 \cdot 10^{-8} \sin^2 \varphi \right) h^2. \qquad (4.7)$$

Here, the unit of potential is m²/s², and the unit of gravity, m/s². More precise equations can be found from Heikkinen (1981). In these equations, the coefficient $9.780\,32\ldots$ m/s² is equatorial gravity, and the value $0.030\,87\ldots$ s^{-2} is the vertical gradient of gravity on the equator. φ is geodetic latitude, h (in metres) is the height above the reference ellipsoid.

Other gravity formulas and reference ellipsoids still in legacy use (and slowly vanishing) are Helmert's 1906 ellipsoid, the Krasovsky ellipsoid or sk-42 in the countries of Eastern Europe, the International or Hayford ellipsoid (1924) and its gravity formula, and Geodetic Reference System 1967.

4.7.1 Numerical example

According to the above equation, the normal potential over the equator is

$$U = 62\,636\,860.8500 - 9.780\,326\,77h + 0.015\,438\,99 \cdot 10^{-4}h^2 -$$

$$- 0.000\,024\,22 \cdot 10^{-8}h^3.$$
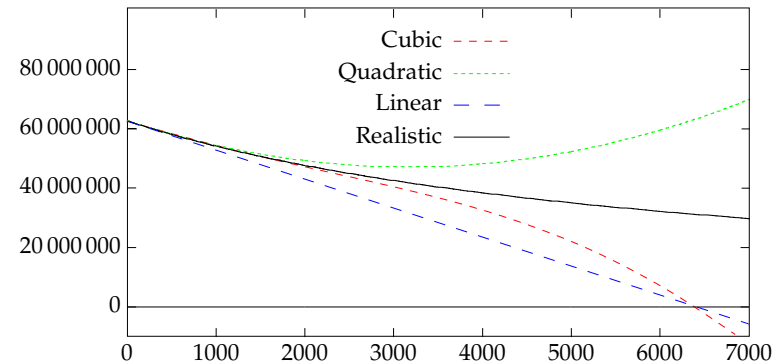
Figure 4.6. The normal field's potential curve over the equator. Heights in kilometres.

○ Draw this function for values of h in the range $0 - 7000$ km.

○ Draw for comparison the quadratic version, from which the last term is left off.

Questions:

1. What is the minimum of the quadratic function?

2. How physically realistic is this?

Answers:

1. See figure 4.6. The minimum of the quadratic function is at height 3000 km. The cubic function doesn't have a minimum.

2. Not very physical: the stationary point for potential U (the normal potential in a co-rotating reference system) should be located at approx. 36 000 km height, at the geostationary orbit.

This tells us that polynomial approximation cannot be extrapolated very far. In this case the interval of extrapolation is of the same order as the radius of the Earth, and that won't work any more.

4.8   The normal potential as a spherical-harmonic expansion

The spherical-harmonic expansion of an ellipsoidal gravitational field contains, besides the second degree harmonic, also higher degree harmonics. If we write, as is customary, the potential outside the Earth in the following form (Heiskanen and Moritz, 1967 section 2-39, also equation 3.16):

$$V = \frac{GM_{\oplus}}{r}\left(1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^{n}\sum_{m=0}^{n}P_{nm}(\sin\phi)\left(J_{nm}\cos m\lambda + K_{nm}\sin m\lambda\right)\right),$$

then we may also write the normal gravitational potential, Ψ, into the form

$$\Psi = \frac{GM_{\oplus}}{r}\left(1 - \sum_{n=1}^{\infty}J_{2n}\left(\frac{a}{r}\right)^{2n}P_{2n}(\sin\phi)\right),$$

which contains only even coefficients $J_{2n} = J_{2n,0}$ as the normal field is *symmetric* about the equatorial plane.

The coefficients for the GRS80 normal gravitational potential are found[5] in table 4.1. Higher terms are usually not needed. The relationship between fully normalized and non-normalized coefficients is $J_n = \bar{J}_n\sqrt{2n+1}$.

Note for comparison that in the expansion of the same field into *ellipsoidal* harmonics only the degree zero and degree two coefficients are non-zero! This is the main reason why these functions are used at all.

Instead of using an ellipsoidal model, we may use as a normal gravity potential formula also the first two, three terms of the spherical-harmonic expansion of the real geopotential. Then we obtain, taking the centrifugal

---

[5] They can also be calculated from equation (2-97) given in Heiskanen and Moritz (1967) :

$$J_{2n} = (-1)^{n+1}\frac{3\left(e^2\right)^n}{(2n+1)(2n+3)}\left(1 - n + 5n\frac{J_2}{e^2}\right),$$

starting from the values J_2 and e^2. The results are the same as in the table's left column.

Table 4.1. GRS80 normal potential spherical-harmonic coefficients (Heikkinen, 1981; Heiskanen and Moritz, 1967).

Non-normalized	Fully normalized
$J_2 = J_{2,0} = 1082.63 \cdot 10^{-6}$	$\bar{J}_2 = 484.166\,854\,90 \cdot 10^{-6}$
$J_4 = J_{4,0} = -2.370\,912\,22 \cdot 10^{-6}$	$\bar{J}_4 = -0.790\,304\,073 \cdot 10^{-6}$
$J_6 = J_{6,0} = +0.006\,083\,47 \cdot 10^{-6}$	$\bar{J}_6 = +0.001\,687\,251\,00 \cdot 10^{-6}$
$J_8 = J_{8,0} = -0.000\,014\,27 \cdot 10^{-6}$	$\bar{J}_8 = -0.000\,003\,46 \cdot 10^{-6}$

potential along:

$$U = \frac{Y_0}{r} + \frac{Y_2(\phi,\lambda)}{r^3} + \frac{1}{2}\omega_{\oplus}^2\left(X^2 + Y^2\right),$$

with the corresponding equipotential surface being the "Bruns spheroid", or

$$U = \frac{Y_0}{r} + \frac{Y_2(\phi,\lambda)}{r^3} + \frac{Y_4(\phi,\lambda)}{r^5} + \frac{1}{2}\omega_{\oplus}^2\left(X^2 + Y^2\right),$$

the "Helmert spheroid". Here, $Y_0 \stackrel{\text{def}}{=} GM$ and $Y_2(\phi,\lambda), Y_4(\phi,\lambda)$ are taken from the true geopotential.

These equations are easy to compute, but their equipotential or level surfaces are not ellipsoids of revolution, and in fact not even rotationally symmetric. They are, in fact, quite complicated surfaces (Heiskanen and Moritz, 1967, section 2-12)!

However, in geometric geodesy we always use a reference ellipsoid, so this is also a wise thing to do in physical geodesy.

4.9   The disturbing potential

Write the gravity potential

$$W = V + \Phi,$$

in which Φ is the centrifugal potential (see above), and the normal potential

$$U = \Psi + \Phi.$$

The difference between them is

$$T \stackrel{\text{def}}{=} W - U = V - \Psi,$$

häiriöpotentiaali   the *disturbing potential*.

Both V and Ψ can be expanded into spherical harmonics. If we write the gravity potential

$$W = V + \Phi =$$

$$= \Phi + \frac{GM_{\oplus}}{r}\left(1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(J_{nm}\cos m\lambda + K_{nm}\sin m\lambda)\right),$$

and the normal potential

$$U = \Phi + \frac{GM_{\oplus}}{r}\left(1 - \sum_{n=2,\text{even}}^{\infty}\left(\frac{a}{r}\right)^n J_n^* P_n(\sin\phi)\right),$$

we obtain by subtraction for the disturbing potential

$$T = W - U =$$

$$= -\frac{GM_{\oplus}}{r}\left(\sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(\delta J_{nm}\cos m\lambda + \delta K_{nm}\sin m\lambda)\right),$$

in which

$$\begin{cases} \delta J_{n0} = J_{n0} - J_n^* & \text{if } n \text{ even,} \\ \delta J_{nm} = J_{nm} & \text{if } n \text{ odd,} \end{cases}$$

$$\text{and}\quad \delta K_{nm} = K_{nm}.$$

The above equation for the disturbing potential T is shortened as follows (Heiskanen and Moritz, 1967, equation 2-152):

$$T(\phi,\lambda,r) = \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^{n+1} T_n(\phi,\lambda),\tag{4.8}$$

where, in every term, the degree constituent T_n has the same dimension as T, and

$$T_n(\phi,\lambda) = -\frac{GM_{\oplus}}{a}\sum_{m=0}^{n} P_{nm}(\sin\phi)\,(\delta J_{nm}\cos m\lambda + \delta K_{nm}\sin m\lambda).$$

On the surface of the reference sphere of radius a[6]:

$$T(\phi,\lambda) = \sum_{n=2}^{\infty} T_n(\phi,\lambda),\tag{4.9}$$

from which we see, that on the reference level, the terms $T_n(\phi,\lambda)$ are really the *degree constituents* for a certain degree number n.

The above expansion is missing the terms $n = 0, 1$. Of these, $T_0(\phi,\lambda) = T_0$ is a constant — the global average of the disturbing potential — and $T_1(\phi,\lambda)$ has the form of a dipole field, the value being proportional to the cosine of the angular distance from the point on the Earth's surface pointed to by the dipole vector. Both vanish because it is assumed that

1. the total mass of the Earth $GM_{\oplus}$ assumed by the normal field is realistic, and

2. the origin of the co-ordinate reference system is assumed to be at the centre of mass of the Earth.

See section 3.4 for a discussion.

The higher T_n are oscillating functions of ever shorter wavelengths.

Self-test questions

1. What is the basic idea behind using a normal gravity field?

2. What is the difference between gravity and gravitation?

3. Given the centrifugal potential

$$\Phi = \frac{1}{2}\omega^2\left(X^2 + Y^2\right),$$

derive the centrifugal acceleration as a vector. X, Y, Z are rectangular co-ordinates of a frame rotating at angular rate ω around the Z axis.

4. Explain the idea of natural co-ordinates.

---

[6]Earlier on we have used for this reference radius (in spherical approximation) also the symbol R.

5. What relationhip was there between M. Le Blanc and C. F. Gauss? Use Google.

6. What makes the Somigliana–Pizzetti equation 4.6 valuable?

7. What are the defining parameters of the Geodetic Reference System 1980?

8. Why does the spherical-harmonic expansion of the normal potential contain only a small number of terms and coefficients?

9. Why doesn't the spherical-harmonic expansion of the normal potential contain any terms with order $m \neq 0$?

10. Why does the spherical-harmonic expansion of the normal potential contain only terms with even degree numbers n?

Exercise 4 – 1: The Somigliana–Pizetti equation

1. Given gravity on the equator γ_a and on the poles γ_b. What is gravity on geodetic latitude $\varphi = 45°$? Derive an expression that may also contain a and b.

2. And what is gravity on the reduced latitude $\beta = 45°$? Compare with the previous.

3. Given the semi-major axis a and semi-minor axis b, what are the differences, for the same point, between the different latitudes (geodetic φ, geocentric ϕ, and reduced β) at most, in minutes of arc? You may assume that the maximum happens at latitudes $\pm 45°$.

4. Compute for both a geodetic and a reduced latitude of $45°$ numerical values of gravity for the case of the GRS80 reference ellipsoid.

Exercise 4 – 2: The centrifugal force

Given is the rotation rate of the Earth, radians per second: $\omega_\oplus = 7292115 \cdot 10^{-11}\,\mathrm{s}^{-1}$.

1. Compute (roughly) the *centrifugal force* caused by the Earth rotation at Southern Finland ($\varphi = 60°$, $R = 6378\,\mathrm{km}$, spherical Earth). In what direction does the force point (sketch!)?

2. How much does the centrifugal force contribute to local gravity? Both as an acceleration and as a percentage.

3. Compute from the $\omega_\oplus$ value given above, the rotation time of the Earth in hours. Why is it not precisely 24^h [difficult]?

Anomalous quantities of the gravity field

5

5.1 Disturbing potential, geoid height, and deflections of the plumb line

The first *anomalous quantity*, which we already discussed above, is the difference between the true gravity potential W and the normal gravity potential U, the *disturbing potential*:

$$T \stackrel{\text{def}}{=} W - U.$$

All other anomalous quantities are various functions of this, like the geoid height N and the plumb-line deflections ξ, η. They are generally obtained by subtracting from each other

1. a natural quantity related to the Earth's real gravity field, and

2. a corresponding quantity related to the normal gravity field of the reference ellipsoid of the Earth.

For example, *deflections of the plumb line*

$$\xi = \Phi - \varphi,$$
$$\eta = (\Lambda - \lambda) \cos \varphi,$$

in which (Φ, Λ) are astronomical latitude and longitude, i.e., the direction of the local plumb line, and the geodetic latitude and longitude (φ, λ) form correspondingly the direction of the surface normal on the reference ellipsoid. See figure 5.1.
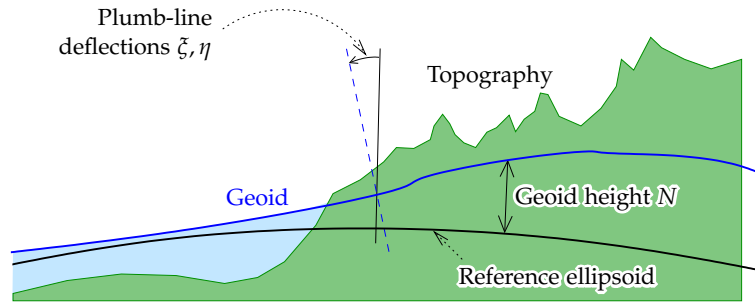
Figure 5.1. Geoid undulations and deflections of the plumb line.

The geoid height or *geoid undulation* is

$$N = H - h,$$

in which H is the orthometric height — reckoned from mean sea level — and h the height above the reference ellipsoid.

Deflections of the plumb line are in Finland a few seconds of arc ($''$) in magnitude, geoid undulations range from 15 to 32 m (for comparison, globally the range of variation is -107 m to $+85$ m), relative to the GRS80 ellipsoid as is today customary. At sea level, the plumb-line deflections — expressed in radians! — equal the horizontal gradients of the geoid undulation. See figures 5.1, 5.2.

For any reference ellipsoid, e.g., the GRS80 ellipsoid, there exists its own mathematically exact standard or *normal gravity field*, of which one equipotential or level surface is precisely that reference ellipsoid. With the aid of this field we may calculate for each gravity field quantity the corresponding normal quantity, and by subtracting the two from each other we obtain again the corresponding *anomalous* quantity.

For heights above the reference ellipsoid there exists an analoguous expression to that for orthometric heights, where U is the normal potential and γ normal gravity:

$$h_P = -\int_{U_0}^{U_P} \frac{1}{\gamma(U)}\, dU.$$
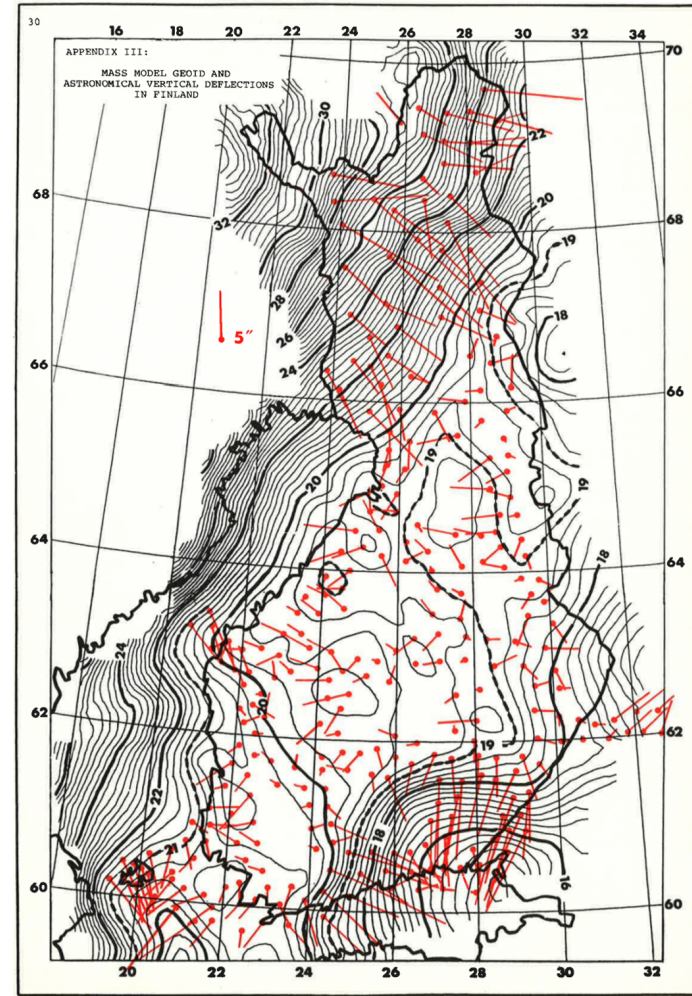
Figure 5.2. A geoid model for Finland from 1984. Deflections of the plumb line from observations in red (Vermeer, 1984).

The geoid height in point P is now

$$N_P = h_P - H_P =$$

$$= \int_{W_0}^{W_P} \frac{1}{g}\, dW - \int_{U_0}^{U_P} \frac{1}{\gamma}\, dU =$$

$$= \int_{W_0}^{W_P} \frac{1}{g}\, dW - \int_{W_0}^{W_P} \frac{1}{\gamma}\, dU - \int_{W_P}^{U_P} \frac{1}{\gamma}\, dU + \int_{W_0}^{U_0} \frac{1}{\gamma}\, dU$$

$$= \int_{W_0}^{W_P} \frac{\gamma - g}{g\gamma}\, dW' - \int_{W_P}^{U_P} \frac{1}{\gamma}\, dU + \int_{W_0}^{U_0} \frac{1}{\gamma}\, dU$$

by re-naming the integration variables $W, U \to W'$ and changing it to a metric one: $dW' = g\, dH$.

In equation 5.1 the last term vanishes if we assume[1] $U_0 = W_0$. Now in the definition of geoid heights, point P is at mean sea level — the zero point of the height system used. It follows that also the first term vanishes (it would in any case always be small, except in the mountains). So

$$N_P = -\int_{W_P}^{U_P} \frac{1}{\gamma}\, dU \approx \frac{1}{\gamma_P}\left(W_P - U_P\right) = \frac{T_P}{\gamma_P} \text{ or } N = \frac{T}{\gamma}. \tag{5.2}$$

where we have substituted $T = W - U$, the disturbing potential. All quantities are assumed to be at sea level. This is the famous *Bruns*[2] equation (Heiskanen and Moritz, 1967, equation 2-144).

Figure 5.3 depicts the situation still better. In this figure, the gradient vectors $\mathbf{g} = \overrightarrow{\text{grad}\, W}$ and $\overrightarrow{\gamma} = \overrightarrow{\text{grad}\, U}$ have lengths $\frac{\partial W}{\partial H}$ and $\frac{\partial U}{\partial H}$, from which it follows, with equation $T = W - U$, that the separation between "matching" surfaces $W = W_P$ and $U = U_Q$, when $W_P = U_Q$, is

$$N = \frac{U_Q - U_P}{\gamma} = \frac{W_P - U_P}{\gamma} = \frac{T}{\gamma}.$$

5.2 Gravity disturbances

The difference between the true and normal gravity accelerations is called the *gravity disturbance*, δg. An exact equation would be

$$\delta g = -\left(\frac{\partial W}{\partial H} - \frac{\partial U}{\partial h}\right);$$

---

[1]This is not self-evident! In a local vertical datum the potential of the zero point could well differ by even as much as a metre from the normal potential of a global reference ellipsoid.

[2]Ernst Heinrich Bruns (1848 – 1919) was an eminent German mathematician and mathematical geodesist.
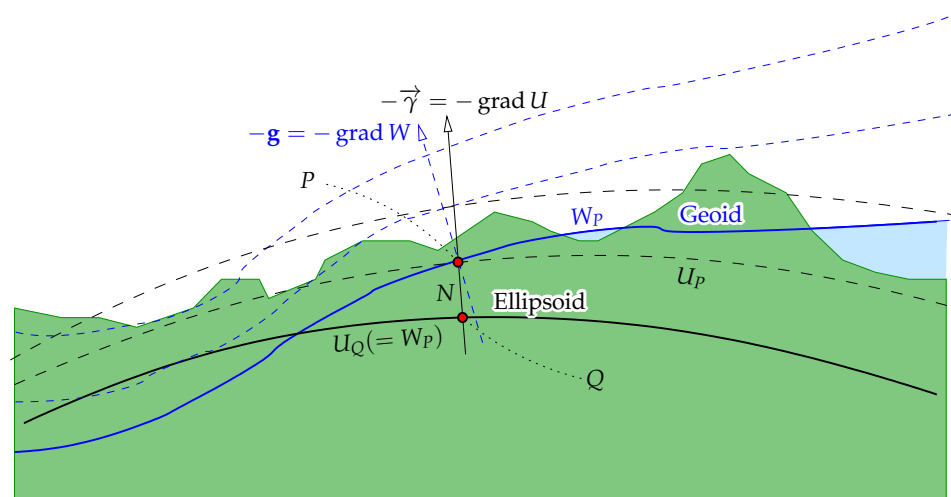
Figure 5.3. Equipotential surfaces of the gravity field (W) and the normal gravity field (U).

where differentiation takes place along the plumb line for W, and along the ellipsoidal normal for U. The directions of plumb line and ellipsoidal surface normal are actually very close to each other.

In spherical approximation we have

$$\delta g = -\left(\frac{\partial W}{\partial r} - \frac{\partial U}{\partial r}\right) = -\frac{\partial T}{\partial r}.$$

We already expanded the disturbing potential into constituents for different spherical-harmonic degree numbers — equation 4.8 — and now we obtain

by differentiating with respect to r[3]:

$$\delta g(\phi, \lambda, r) = -\frac{\partial T(\phi, \lambda, r)}{\partial r} =$$

$$= -\frac{\partial}{\partial r}\left(\sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda)\right) =$$

$$= \frac{1}{r}\sum_{n=2}^{\infty}(n+1)\left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda), \qquad (5.3)$$

or on the Earth's surface ($r = R$):

$$\delta g(\phi, \lambda, R) = \frac{1}{R}\sum_{n=2}^{\infty}(n+1)\, T_n(\phi, \lambda). \qquad (5.4)$$

This is the *spectral representation* of the gravity disturbance on the surface of the Earth — more precisely, on a sphere of radius R. As a suitable value for the reference radius R one may take the equatorial radius a of a reference ellipsoid for the Earth.

We can observe gravity disturbances only if, in addition to measuring the acceleration of gravity $g_P\left(=\left.\frac{\partial W}{\partial H}\right|_P\right)$ in point P, we have a way to measure P:s *location* in space, relative to the geocentre, so one may calculate normal gravity $\gamma_P = \left.\frac{\partial U}{\partial h}\right|_P$ in the same point. Nowadays this is even easy using GPS, but traditionally it has been impossible. For this reason gravity disturbances are little used. One rather uses gravity *anomalies*, about which more below.

5.3   Gravity anomalies

Normal gravity is calculated as a function of location expressed in *geodetic* co-ordinates (φ, λ, h). However, in traditional gravimetric field work, before the satellite positioning era, one only had access to the geodetic co-ordinates

---

[3]Note that, if the normal field that defines T has a reference ellipsoid centred on the Earth's centre of mass, and the implied total mass GM of the normal field equals that of the true Earth, then the first two constituents $T_0(\phi, \lambda) = T_1(\phi, \lambda) = 0$, and the sum may be taken as $\sum_{n=2}^{\infty}$. This was assumed here. See section 3.4.
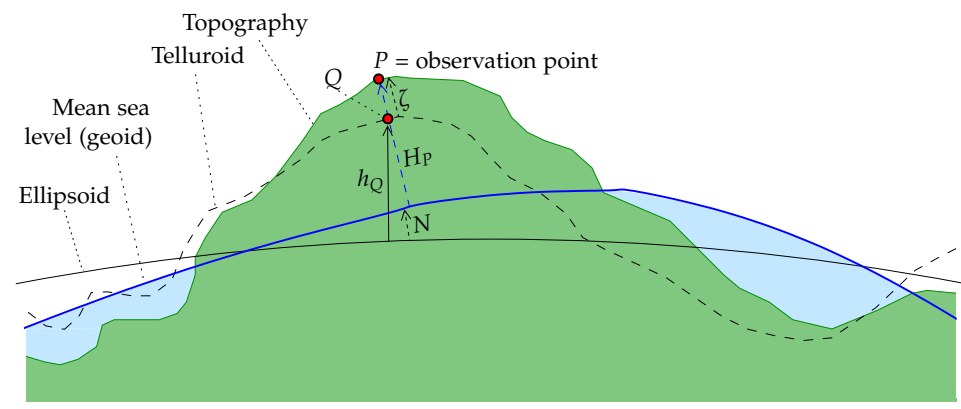
Figure 5.4. Reference ellipsoid, mean sea level (geoid), and gravity measurement.

φ and λ, not any accurate height h above the reference ellipsoid. One only had access to the height H above sea level (the geoid), obtained, e.g., through a national levelling network — or, in the worst case, barometrically.

This means that, though the true gravity g is measured in point P the height of which above sea level is H_P, normal gravity γ must of necessity be calculated *in another point Q*, the height of which *above the reference ellipsoid* is $h_Q = H_P$. See figure 5.4.

In other words, the measured height of point P *above mean sea level* is substituted, brute-force style, into the normal gravity formula, that however expects a height *above the reference ellipsoid*! This special trait of the definition of gravity anomalies may be called a "free boundary-value problem".

According to this we calculate gravity anomalies as follows:

$$
\begin{aligned}
\Delta g_P = g_P - \gamma_Q &= (g_P - \gamma_P) + (\gamma_P - \gamma_Q) = \\
&= -\left(\left.\frac{\partial W}{\partial H}\right|_P - \left.\frac{\partial U}{\partial h}\right|_P\right) - \left(\left.\frac{\partial U}{\partial h}\right|_P - \left.\frac{\partial U}{\partial h}\right|_Q\right) \\
&\approx -\left.\frac{\partial (W - U)}{\partial H}\right|_P + (h_P - h_Q)\left.\frac{\partial \gamma}{\partial H}\right|_P = \\
&= -\left.\frac{\partial T}{\partial H}\right|_P + (h_P - H_P)\left.\frac{\partial \gamma}{\partial H}\right|_P = \\
&= -\left.\frac{\partial T}{\partial H}\right|_P + N_P\left.\frac{\partial \gamma}{\partial H}\right|_P = \\
&= \left.\left(-\frac{\partial T}{\partial H} + \frac{T}{\gamma}\frac{\partial \gamma}{\partial H}\right)\right|_P,
\end{aligned}
$$

using almost all equations above.

The equation derived looks familiar: it is the boundary condition of the *third boundary-value problem* (Heiskanen and Moritz, 1967, section 1-17). It enables the solution of T in the exterior space, if Δg is given everywhere on the Earth's surface.

If we assume that the Earth's normal gravity field is spherically symmetric, we may approximate (exercise: show this!):

$$
\Delta g = -\frac{\partial T}{\partial r} - \frac{2}{r}T, \tag{5.5}
$$

in which $r = R + H$ is the distance from the Earth's centre.

By substituting into this the equation for δg, and $r = R$, we obtain on the Earth's surface:

$$
\Delta g = \delta g - \frac{2}{R}T. \tag{5.6}
$$

From this we obtain directly by using the above spectral representations 4.9, and 5.4 for T and δg:

$$
\Delta g = \frac{1}{R}\sum_{n=2}^{\infty}((n+1) - 2)T_n = \frac{1}{R}\sum_{n=2}^{\infty}(n-1)\,T_n.
$$

Choose the following notation, still on the Earth's surface:

$$
\Delta g = \sum_{n=2}^{\infty}\Delta g_n,
$$

in which the *degree constituents* of the gravity anomaly are

$$
\Delta g_n \overset{\text{def}}{=} \frac{n-1}{R}T_n. \tag{5.7}
$$

The presence of the factor $n - 1$ shows that gravity anomalies cannot contain $n = 1$ constituents, even if T would contain them. It is always wise to choose the origin of the co-ordinate system to be in the centre of mass of the Earth, but if it is not, at least gravity anomalies do not change.

Equation 5.7 applies only on a spherical Earth of radius R. In the space outside the Earth we obtain, using equations 4.8, 5.3 and 5.5, the corresponding equation

$$
\begin{aligned}
\Delta g &= \frac{1}{r}\sum_{n=2}^{\infty}((n+1) - 2])\left(\frac{R}{r}\right)^{n+1}T_n = \\
&= \frac{1}{r}\sum_{n=2}^{\infty}(n-1)\left(\frac{R}{r}\right)^{n+1}T_n = \\
&= \sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n+2}\Delta g_n. \tag{5.8}
\end{aligned}
$$

5.4   Units used for gravity anomalies

The most common unit of measurement for gravity variations is the *milligal*. The connection with the SI system is $1\,\text{mGal} = 10^{-5}\,\text{m/s}^2$. Also μGal or $10^{-8}\,\text{m/s}^2$ is used. In modern books are also used $^{-}\text{m/s}^2$ and nm/s^2, which formally belong to the SI system. Nevertheless, milligals and microgals are more familiar still, and correspond to $1\,\text{ppm}$ (part per million) and $1\,\text{ppb}$ (part per billion) of ambient gravity close to the Earth's surface.

A popular unit for measuring gravity *gradients* is the Eötvös, symbol E. In SI units it is $10^{-9}\,\text{s}^{-2}$, corresponding to $10^{-4}\,\text{mGal/m}$. In table 5.1 we give a few values in order to get an idea of the orders of magnitude of phenomena. On the Earth's surface the vertical gradient $\frac{\partial g}{\partial h}$ is on average some $-0.3\,\text{mGal/m} = -3000\,\text{E}$.

5.5   The boundary-value problem of physical geodesy

As we explained in the above section, gravimetric measurement is more complicated than just measuring the quantity $\frac{\partial W}{\partial r}$. When we measure the

Table 5.1. Orders of magnitude of gravity variations.

Phenomenon	Fraction of gravity	SI units	mGal
Ambient gravity	1	9.8	980 000
Variation with location	$\pm 10^{-4}$	$\pm 10^{-3}$	± 100
Difference equator – poles	0.5%	0.05	5000
Difference sea surface – 10 km high	0.3%	0.03	3000
Gravimeter accuracy	$\pm 10^{-8} - 10^{-7}$	$\pm 10^{-7} - 10^{-6}$	$\pm 0.01 - 0.1$

radial derivative of the geopotential, we do it *in a place we don't precisely know.* Even if we knew the height of the measurement location above sea level, that still doesn't give us the measurement point's location in space. This location depends additionally on the location in space of sea level, i.e., the *geoid*, specifically its height above or below the reference ellipsoid.

This is how we arrive at the third boundary-value problem[4]. The *boundary-value problem of physical geodesy* is to determine the potential V outside a body if given on its surface is the *linear combination*

$$aV + b\frac{\partial V}{\partial n},$$

with a, b suitable constants. The variable n represents here differentiation in the direction of the normal to the Earth's surface, in practice the same as r or h.

In physical geodesy is given, in spherical approximation, the following linear combination (gravity anomaly, equation 5.6):

$$\Delta g = -\frac{\partial T}{\partial n} - \frac{2}{R}T. \tag{5.9}$$

[4]The third or mixed boundary-value problem is associated with Victor Gustave Robin (1855 – 1897), a French mathematician. Then, the Dirichlet problem could be called the first, the Neumann problem the second boundary-value problem.

The equation, or *boundary condition,* 5.9 is called the *fundamental equation of physical geodesy.*

Above we already obtained equations 3.8 and 3.12, that apply equally well to the disturbing potential T as to the general potential V:

$$\frac{\partial T}{\partial n} = -\sum_{n=0}^{\infty} \frac{n+1}{r} \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda),$$

$$T = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda).$$

By combining these we obtain

$$\Delta g = \sum_{n=0}^{\infty} \left(\frac{n+1}{r} - \frac{2}{R}\right) \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda),$$

or on the Earth's surface ($R = r$):

$$\Delta g = \sum_{n=0}^{\infty} \frac{n-1}{R} T_n(\phi, \lambda) = \sum_{n=0}^{\infty} \Delta g_n(\phi, \lambda),$$

where the quantities $\Delta g_n(\phi, \lambda) = \frac{n-1}{R} T_n(\phi, \lambda)$ are defined in equation 5.7. Remember that the functions $\Delta g_n(\phi, \lambda)$ are computable with the help of the degree constituent equation 3.7 when $\Delta g(\phi, \lambda)$ is known all over the Earth.

Observe also that the term $n = 1$ vanishes: $\Delta g_1 = 0$. We assume also $\Delta g_0 = -\frac{T_0}{R} = 0$, i.e., the true exterior potential, and thus the total mass of the Earth GM, and her volume[5], is in global average the same as the normal potential and its assumed total mass, and the volume of the reference ellipsoid. The assumption is largely justified because GM can be, and has been, determined very precisely by satellites, and modern models for the normal potential, like GRS80, are based on these determinations[6].

[5]In fact, the atmosphere complicates this matter.

[6]Note, however, that GRS80 has an equatorial radius of 6 378 137.0 m, while the newer models like EGM2008 give a smaller value of 6 378 136.3 m as the location of global mean sea level. Uncertainty continues to be in the decimetres.

Thus we obtain the solution also of this boundary-value problem in spectral representation (which is thus valid in the *whole* exterior space) by using the degree constituent equation[7] 3.7:

$$T(\phi, \lambda, r) = R \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{\Delta g_n(\phi, \lambda)}{n-1} =$$

$$= \frac{R}{4\pi} \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \left(\frac{R}{r}\right)^{n+1} \iint_{\sigma} \Delta g(\phi', \lambda') P_n(\cos \psi) \, d\sigma'. \qquad (5.10)$$

This is precisely the boundary-value problem that is created if everywhere on the Earth, land and sea, surface gravity anomalies are given.

The integral equation corresponding to the above spectral equation 5.10 is known as the Stokes[8] equation:

$$T(\phi, \lambda, r) = \frac{R}{4\pi} \iint_{\sigma} S(\psi, r) \, \Delta g(\phi', \lambda') \, d\sigma',$$

in which the *Stokes kernel* is

$$S(\psi, r) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \left(\frac{R}{r}\right)^{n+1} P_n(\cos \psi). \qquad (5.11)$$

In section 8.1 we will give a closed expression 8.2 for this function, for the case $r = R$, and a graph.

5.6   The telluroid mapping and the "quasi-geoid"

If we measure the astronomical latitude and longitude Φ, Λ and *interpret* them as geodetic (geographical) co-ordinates φ, λ, and also *interpret* the potential difference $-(W - W_0)$ as a measure for the height above the reference ellipsoid h, we perform, as it were, a *mapping*, which adds to every point P a

---

[7]We write $\Delta g(\phi', \lambda') \stackrel{\text{def}}{=} \Delta g(\phi', \lambda', R)$, the integral being evaluated on the Earth sphere $r = R$.

[8]Sir George Gabriel Stokes PRS (1819–1903) was an Irish-born, gifted mathematician and physicist making his career in Cambridge.

corresponding point Q, the *geodetic* co-ordinates of which are the same as the *natural* co-ordinates of point P.

This approach is called the *telluroid mapping*. The telluroid is the surface that follows the shapes of Earth's topography, but is everywhere below the topography by an amount ζ if positive, or above it by an amount $-\zeta$ otherwise. The quantity ζ is called a *height anomaly*.

The telluroid mapping is an important tool in Molodensky's gravity field theory. It is however a pretty abstract concept. One may say that the telluroid is a *model* of the Earth's surface, obtained by starting from the assumptions that

- the true potential field of the Earth is the normal potential, and
- the mathematical mean sea surface or *geoid*, the reference surface for height measurement, coincides with the reference ellipsoid.

In other words, the telluroid is a model for the Earth's topographic surface that is obtained by taking *levelled* heights — more precisely, geopotential numbers obtained from levelling — as if they represented differences in *normal* potential with that of the reference ellipsoid.

In practice, a map of values ζ is often called a "quasi-geoid" model. The quasi-geoid is usually close to the geoid, except in the mountains, where the differences can exceed a metre.

One should however remember, that the height anomaly ζ is defined *on the topographic surface*, a surface that is quite rough in many places. This means also that all variations in topographic height will be reflected also as variations in the "quasi-geoid", in such a way, that the quasi-geoid *correlates* strongly with the small details in the topography. One can thus not say that the shape of the quasi-geoid only describes the Earth's potential field. In it, variations in potential and variations in topographic height are hopelessly mixed up.

This is why the quasi-geoid is an unfortunate compromise, a concession to "reference-surface thinking", which only really works within the classical geoid concept. Better stick — within Molodensky's theory — to the concept

height anomaly, which is a three-dimensional function or field

$$\zeta(X, Y, Z) = \zeta(\varphi, \lambda, h).$$

5.7 Free-air anomalies

If we measure gravity g in point P, the height of which over sea level is H and its latitude Φ, we may calculate the gravity anomaly Δg in the point as follows:

$$\Delta g \stackrel{\text{def}}{=} g - \gamma(H, \Phi),$$

in which $\gamma(H, \Phi)$ is normal gravity calculated according to its definition at height H and latitude Φ.

This is how we define free-air anomalies.

We *linearize* this as follows:

$$\Delta g = g - \gamma(H, \Phi) \approx$$
$$\approx g - \gamma(h, \varphi) - (H - h)\frac{\partial\gamma}{\partial h} - (\Phi - \varphi)\frac{\partial\gamma}{\partial\varphi} \approx$$
$$\approx g - \gamma(0, \varphi) - h\frac{\partial\gamma}{\partial h} - (H - h)\frac{\partial\gamma}{\partial h} =$$
$$= g - \gamma(0, \varphi) - H\frac{\partial\gamma}{\partial h},$$

where we make the approximation, that the vertical gradient $\frac{\partial\gamma}{\partial h}$ of normal gravity is constant[9].

Thus, free-air anomalies can be calculated in a simpler way. The gravity formula of the normal field 4.7 gives for latitude 60°:

$$\gamma = 974\,147.516 - 0.308\,449\,4\,H + \dots \text{ mGal}.$$

---

[9]For greatest precision one should consider that also the latitude Φ may not be a latitude on a geocentric reference ellipsoid, but, e.g., astronomical latitude, or latitude in some old national co-ordinate system computed on a non-geocentric ellipsoid, like in Finland KKJ, the National Grid Co-ordinate System, which was computed on the Hayford ellipsoid. The error caused by this is however of order a thousand times smaller than the effect caused by $H - h$.

So, in linear approximation (close to the Earth's surface) gravity attenuates some 0.3 mGal for every metre in height. This value is worth committing to memory.

An approximate equation for calculating free-air anomalies then is

$$\Delta g_P = g_P - \gamma_0(\varphi) + 0.3084\,[\text{mGal}/\text{m}]\,H, \tag{5.12}$$

in which $\gamma_0(\varphi) \stackrel{\text{def}}{=} \gamma(0, \varphi)$, normal gravity at sea level, is only a function of latitude. In a country like Finland, equation 5.12 is often sufficiently precise, though today also the evaluation of original equation 4.7 is easy.

Free-air anomalies are widely used. Generally, when one discusses gravity anomalies, one means just this, free-air anomalies. They describe the Earth's exterior gravity field, including mountains, valleys and everything.

Questions:

1. If gravity on the Earth's surface is $9.8\,\text{m}/\text{s}^2$, at what height will gravity disappear, as computed according to the above mentioned vertical gravity gradient $-0.3\,\text{mGal}/\text{m}$?

2. How physically realistic is this?

Answers:

1. At $-0.3\,\text{mGal}/\text{m}$, it takes $\frac{10^5 \times 9.8}{0.3}\text{m} = 3267\,\text{km}$ to go to zero.

2. Not very. The gravity gradient itself drops quickly from the value of $-0.3\,\text{mGal}/\text{m}$ going up, so this linear extrapolation is simply wrong.

Self-test questions

F

1. How do deflections of the plumb line and geoid heights relate to each other?

2. What is the fundamental equation of physical geodesy in spherical approximation?

GMT 2015 Mar 22 17:40:39
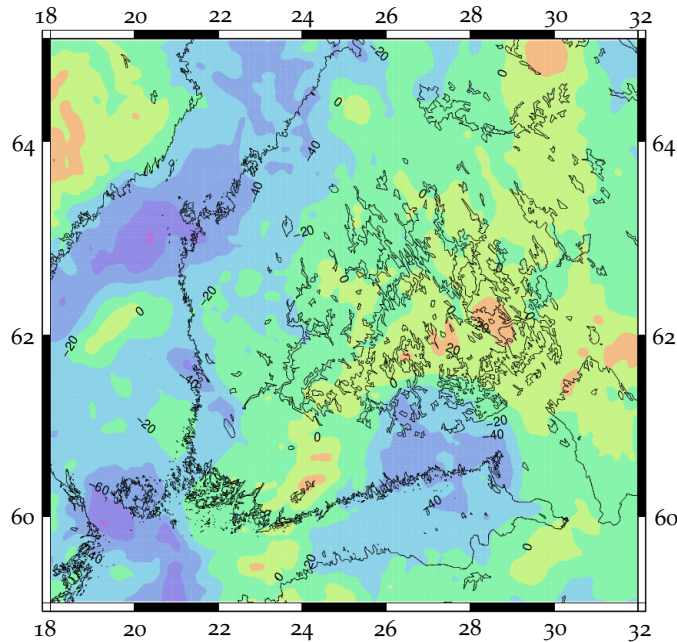
Figure 5.5. Free-air gravity anomalies for Southern Finland, computed from the EGM2008 spherical-harmonic expansion. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy. Web service BGI, EGM2008.

3. In what way is a gravity disturbance different from a gravity anomaly?
4. What units are used for measuring gravity anomalies and gravity gradients? How are they related to the SI system?
5. How does the geoid height and the disturbing potential relate to each other?
6. Explain the telluroid mapping and height anomalies.

Exercise 5 – 1: The spectrum of gravity anomalies

Use equation 5.7. If we assume that the mean magnitude of the spectral components Δg_n of gravity anomalies

$$\overline{\Delta g_n} \stackrel{\text{def}}{=} \sqrt{\frac{1}{4\pi} \iint_\sigma \Delta g_n^2 (\phi, \lambda)\, d\sigma}$$

does not depend on the chosen degree number n, how then does the similarly defined $\overline{T_n}$ depend on n?

In other words: which degree numbers of the gravity field are relatively strongest in the disturbing potential, and which in the gravity anomalies?

Exercise 5 – 2: Deflections of the plumb line and geoid tilt

If, in the North-South components of plumb-line deflections in some country, there is a systematic error of one arc second, what error does this cause in the difference $N_2 - N_1$ between the geoid heights in points 1 and 2, inter-point distance approx. 1000 km? See figure 5.1.

Exercise 5 – 3: Gravity anomaly, geoid height

In Finland there is a place where the gravity anomaly (free air) is $\Delta g = 100\,\text{mGal} = 10^{-3}\,\text{m/s}^2$. In the same place the disturbing potential T is $200\,\text{m}^2\,\text{s}^{-2}$.

1. Using the equation

$$\Delta g = \frac{\partial T}{\partial n} - \frac{2}{R}T,$$

calculate $\frac{\partial T}{\partial n}$, and compare it with the quantity $\frac{2}{R}T$. Which of the two ($\frac{\partial T}{\partial n}$ or $\frac{2}{R}T$) dominates?

2. Using the *Bruns equation*

$$N = \frac{T}{\gamma},$$

where γ is average gravity $9.8\,\text{m/s}^2$, compute the *geoid height N* of the point.

Geophysical reductions

6

6.1 General

We see that integral equations, like Green's third theorem 1.24, offer a possibility to calculate the whole exterior potential of the Earth — as well as all quantities that may be calculated from the potential, like the acceleration of gravity, etc. — from observed values V and $\frac{\partial V}{\partial n}$ on the boundary surface only. Green's third theorem is but one example out of many: every integral theorem is the solution of some *boundary-value problem*.

There are three alternatives concerning the choice of boundary surface:

1. Choose the topographic surface of the Earth.

2. Choose mean sea level, more precisely, a equipotential surface close to mean sea level called the *geoid*.

3. Choose the reference ellipsoid.

- Alternative 1 has been developed most of all by the Molodensky school (Molodensky et al., 1962) in the Soviet Union. The advantage of the method is that *we need no gravity reduction*, as all significant masses are already inside the boundary surface. Its disadvantage is, that the, often complex, shape of the topography must be taken into account when the boundary-value problem is formulated and solved.

- Alternative 2 is classical geoid or geopotential determination. In this case *geophysical reductions* are needed to the input gravity data: some

masses are outside the computation boundary and need to be computationally moved to the inside.

A further complication of the method then is, that the geopotential or geoid solution obtained is not that of the original mass distribution, but of the *reduced* one. This surface is called the *co-geoid*. We need a "restoration step" where this influence of the reduction step on the geopotential and geoid is determined and reversed[1].

epäsuora vaikutus

In the literature this method is also referred to as the *Remove-Restore* method.

○ Alternative 3 has been used rarely, because it has not been traditionally possible to do gravity measurements in a location known in the absolute sense, relative to the geocentre or the reference ellipsoid. Nowadays this is possible using GNSS, e.g., in Antarctica and Greenland's interior, where there is no sea-level bound height system.

We may expect this approach to gain more traction as heights, also of gravimetric stations, are determined more and more directly with GNSS. See, e.g., Märdla (2017).

6.2 Bouguer anomalies

Free-air anomalies depend on the topography. This is clear, because gravity itself contains the attractive effect of topographic masses. A map of free-air anomalies shows the same small details as seen in the topography. One way of removing the effect of the topography is the so-called *Bouguer*[2] *reduction*.

---

[1]This influence is called the "indirect effect".

[2]Pierre Bouguer (1698 – 1758) was a French professor in hydrography, who participated in the public discussion on the figure of the Earth, and in 1735 – 1743 led an expedition of the French Academy of Sciences doing a grade measurement in Peru, South America, at the same time when De Maupertuis carried out a similar grade measurement in Lapland. In addition to geodesy, he was also active in astronomy.
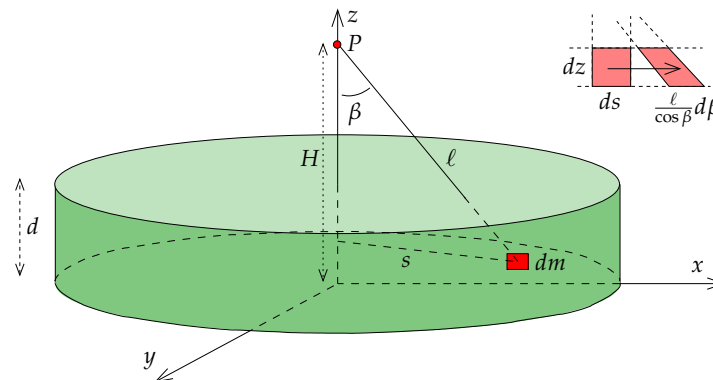
Figure 6.1. The attraction of a Bouguer plate.

6.2.1 Calculation

We calculate the effect of a homogeneous plate on gravity. Assume that the plate is infinite in size; thickness d, matter density ρ, and height of point P above the lower surface of the plate H. See figure 6.1. The attraction in point P (which is directed straight downward for symmetry reasons) is obtained by integrating. The volume integral to be computed has a volume element

$$d\mathcal{V} = ds \cdot dz \cdot s\, d\alpha$$

in the cylindrical co-ordinates (s, z, α). We *transform* this to the co-ordinates (β, z, α). We forget about α and study the surface element (figure 6.1, top right)

$$ds\, dz = \frac{\ell}{\cos\beta}\, d\beta\, dz,$$

in which the determinant of Jacobi needed is seen.

We carry out the integration:

$$a \overset{\text{def}}{=} \|\mathbf{a}\| =$$

$$= G \iiint \frac{\cos \beta}{\ell^2} \rho \, d\mathcal{V} =$$

$$= G\rho \int_0^{2\pi} \int_0^d \int_0^\infty \frac{\cos \beta}{\ell^2} \cdot ds \, dz \cdot s \, d\alpha =$$

$$= G\rho \int_0^{2\pi} \int_0^d \int_0^{\pi/2} \frac{\cos \beta}{\ell^2} \cdot \frac{\ell}{\cos \beta} \, d\beta \, dz \cdot s \, d\alpha =$$

$$= 2\pi G\rho \int_0^d \int_0^{\pi/2} \frac{s}{\ell} \, d\beta \, dz =$$

$$= 2\pi G\rho \int_0^d \left(\int_0^{\pi/2} \sin \beta \, d\beta \right) dz.$$

Here, the integral

$$\int_0^{\pi/2} \sin \beta \, d\beta = [-\cos \beta]_0^{\pi/2} = 1,$$

and the end result is

$$a = 2\pi G\rho d. \tag{6.1}$$

This is the formula for the attraction of a Bouguer plate. As a side result we obtain the attraction of a circular disk of radius r:

$$\int_0^{\beta_0(z)} \sin \beta \, d\beta = [-\cos \beta]_0^{\beta_0(z)} = 1 - \cos \left(\beta_0(z) \right),$$

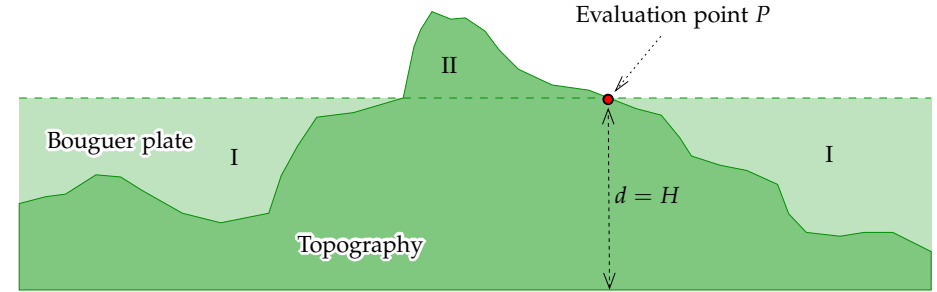and the whole integral

$$2\pi G\rho \int_0^d \left(1 - \frac{H - z}{\sqrt{(H - z)^2 + r^2}} \right) dz.$$

The indefinite integral is

$$\int \frac{H - z}{\sqrt{(H - z)^2 + r^2}} \, dz = -\sqrt{(H - z)^2 + r^2}.$$

Substituting the bounds yields

$$\int_0^d \left(1 - \frac{H - z}{\sqrt{(H - z)^2 + r^2}} \right) dz = d + \sqrt{(H - d)^2 + r^2} - \sqrt{H^2 + r^2}.$$

≡ ↑ 🖼 ▦ 🔍 🗐 ✧

Figure 6.2. The Bouguer plate as an approximation to the topography.

If we define

$$\ell(z) \overset{\text{def}}{=} \sqrt{(H - z)^2 + r^2}$$

we obtain for the whole integral

$$2\pi G\rho \left(d + \ell(d) - \ell(0) \right).$$

In the limit $r \to \infty$, and thus $\ell(d) \to \ell_0$, this is identical to equation 6.1.

Bouguer anomalies are computed in order to remove the attraction of masses of the Earth's crust above sea level, i.e., the geoid. The true topography is *approximated* by a Bouguer plate, see figure 6.2. There is no standard way to treat sea-covered areas:

- ◦ Sometimes maps are drawn on which there are Bouguer anomalies over land and free-air anomalies over the sea.

- ◦ A more correct way is to replace sea water by a rocky Bouguer plate, the thickness of which equals the local sea depth, i.e., the bathymetry.

The calculation goes as follows:

$$\Delta g_{\text{B}} = \Delta g_{\text{FA}} - 2\pi G\rho H = \Delta g_{\text{FA}} - 0.1119 \, H, \tag{6.2}$$

where we assume for the density ρ of the plate an often used value for the average density of the Earth's crust, $\rho = 2670 \, \text{kg/m}^3$. By substituting into this equation 5.12 we obtain

$$\Delta g_{\text{B}} = g_P - \gamma_0(\varphi) + (0.3084 - 0.1119) \, H = g_P - \gamma_0(\varphi) + 0.1965 \, H. \tag{6.3}$$

≡ ↑ 🖼 ▦ 🔍 🗐 ✧
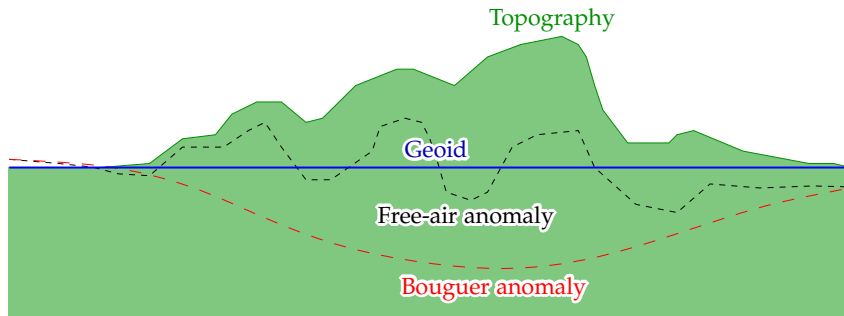
Figure 6.3. The behaviour of different anomaly types in mountainous terrain.

The quantity Δg_{B} is called a (simple) *Bouguer anomaly*.

The difference between the attraction of a Bouguer plate and that of the true topography is called the *terrain correction TC* (the volumes I and II in figure 6.2). We shall return to its computation later on.

6.2.2 Properties

Bouguer anomalies are, unlike free-air anomalies that vary on both sides of zero, especially in the mountains strongly negative. For example, if the mean elevation of a mountain range is $\overline{H} = 1000\,\mathrm{m}$, the Bouguer anomalies will, as a consequence of this, contain a *bias* of $1000 \times (-0.1119\,\mathrm{mGal}) = -112\,\mathrm{mGal}$, about $-100\,\mathrm{mGal}$ for every kilometre of elevation.

systematiikka

The *advantage* of Bouguer anomalies is their smaller variation with place. For this reason they are suited especially for *interpolation* and *prediction* of gravity anomalies, in situations where the available gravimetric material is sparse. However, one then has to have access to topographic heights.
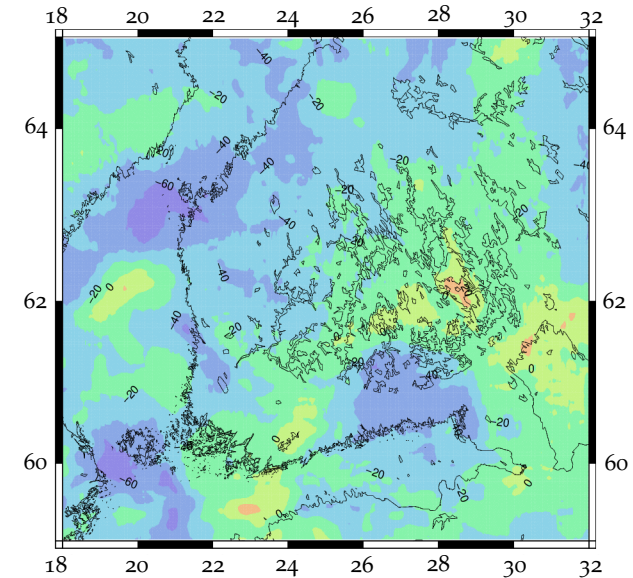
6.3 Terrain effect and terrain correction

Using the simple Bouguer reduction does not remove precisely the attractive effect of the whole topography. Figure 6.2 shows that we make *two types of error*:

Figure 6.4. Terrain corrected Bouguer anomalies for Southern Finland, computed from the spherical-harmonic expansion EGM2008. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy. Web service BGI, EGM2008. Note, in comparison to figure 5.5 on page 118, the strong negative bias of Bouguer anomalies — although part of this is due to post-glacial isostatic unbalance and also visible in the free-air map. Bouguer anomalies are also smoother, but that is harder to see here, as Southern Finland is already a smooth area.

- The attraction of volumes I is taken along, though there is nothing there.
- The attraction of volumes II, where there actually is stuff, is ignored.

Both errors work in the same direction! Because volumes I are below the point of evaluation, their attraction — which the simple Bouguer reduction considers
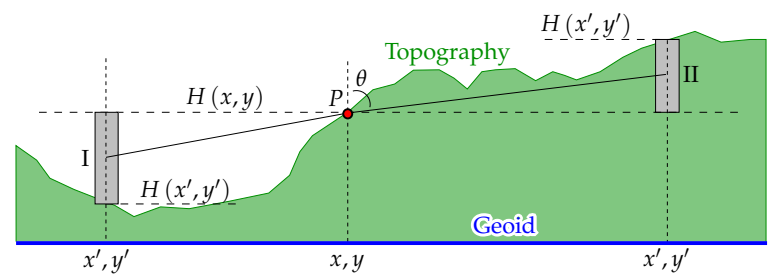
Figure 6.5. Calculating the classical terrain correction by the prism method.

present, and removes — would act downward. And because volumes II are above the point of evaluation, their attraction — which in the simple Bouguer reduction is not corrected for — acts upward. The error made is in the same direction as in the previous case.

The terrain correction is always positive!

We write

$$\Delta g_B' = \Delta g_B + TC,$$

where TC — the "terrain correction" — is positive. $\Delta g_B'$ is called the *terrain corrected Bouguer anomaly*.

The terrain correction is calculated by numerical integration. Figure 6.5 shows the *prism integration method*, and how both prisms, I and II, lead to a positive correction, because prism I is computationally added and prism II removed when applying the terrain correction. One needs a digital terrain model, DTM, which must be, especially around the evaluation point, extremely dense: according to experience, 500 m is the maximum inter-point separation in a country like Finland, in the mountains one needs even 50 m. The systematic nature of the terrain correction makes a too sparse terrain model cause, possibly serious, *biases* in the insufficiently corrected gravity anomalies.

For computing the terrain correction with the prism method we use the following equation, assuming a constant crustal density ρ and a flat Earth, in

rectangular map co-ordinates x, y:

$$TC(x,y) = \frac{1}{2}G\rho \int_{-D}^{+D} \int_{-D}^{+D} \left(H(x',y') - H(x,y) \right)^2 \ell^{-3} \, dx' \, dy',$$

in which

$$\ell = \sqrt{(x'-x)^2 + (y'-y)^2 + \left(\frac{1}{2}(H(x',y') - H(x,y)) \right)^2}$$

is the distance between the evaluation point $\begin{bmatrix} x & y & H(x,y) \end{bmatrix}^\mathsf{T}$ and the centre point of the prism $\begin{bmatrix} x & y & \frac{1}{2}\left(H(x,y) + H(x',y') \right) \end{bmatrix}^\mathsf{T}$. Of course this is only an approximation, but it works well enough in terrain where slopes generally do not exceed 45°. In the integral above, the limit D is typically tens or hundreds of kilometres. In the latter case, the curvature of the Earth already starts having an effect, which the formula does not consider.

The values of the terrain correction TC vary from fractions of a milligal (Southern Finland) to hundreds of milligals (high mountain ranges). In the "arm" of Finland — the North-Western, somewhat mountainous border area with Sweden and Norway — the terrain correction may be tens of milligals.

In figure 6.6 we depict the stages of calculating Bouguer anomalies from gravity observations through terrain correction, Bouguer plate correction and free-air reduction.

6.3.1   Example: applying the terrain correction in a special case

Given the special terrain shape rendered in quasi 3D in figure 6.7. Here, the height differences are $PQ' = 300$ m and $QQ' = 200$ m. Rock density is the standard crustal density, $2670 \, \mathrm{kg/m^3}$.

Questions:

1. Calculate the terrain correction at point P (hint: use the attraction formula for the Bouguer plate). *Algebraic sign?*

2. Calculate the terrain correction at point Q. *Algebraic sign?*

Page 131

130   Geophysical reductions   Terrain effect and terrain correction   131
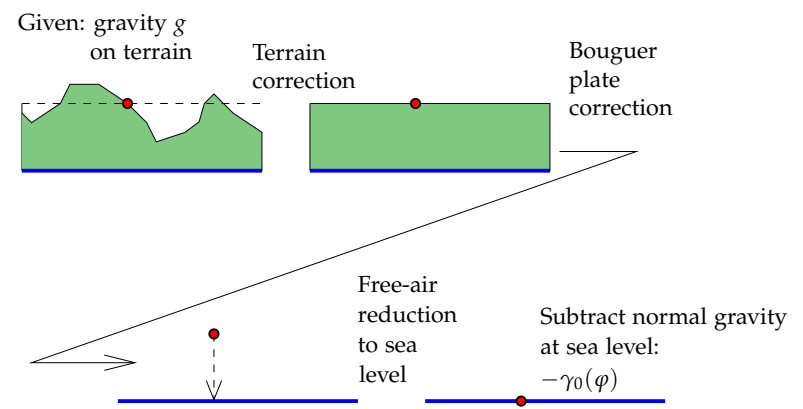
Figure 6.6. The steps in calculating the Bouguer anomaly. Note that the reduction to sea level uses the standard free-air vertical gravity gradient, -0.3084 mGal/m, i.e., the *vertical gradient of normal gravity*.

3. If in point P is given that the free-air anomaly is 50 mGal, how much is then the Bouguer anomaly in the point?

4. If in point Q is given that the Bouguer anomaly is 22 mGal, how much is then the free-air anomaly in the point?

Answers:

1. The terrain correction at point P is the change in gravity, if the

Figure 6.7. A special terrain shape. The vertical rock wall at PQ is also straight on a map and extends to infinity in both directions.
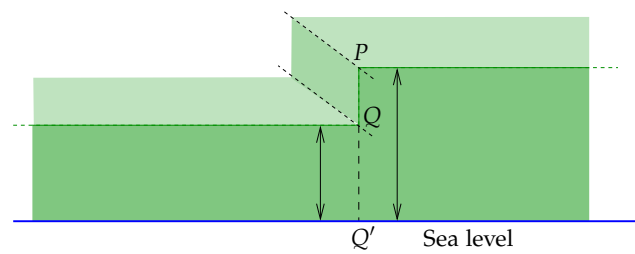
terrain is filled up on the left side up to level 300 metres. This means *the adding of half a Bouguer plate*, thickness 100 m, below the level of P. The effect (projected onto the vertical direction) is

$$TC = \frac{1}{2} \cdot 2\pi G\rho \cdot H = \frac{1}{2} \cdot 0.1119 \,\text{mGal/m} \cdot 100 \,\text{m} = 5.595 \,\text{mGal}.$$

2. The terrain correction at point Q is the change in gravity, if we remove *the half Bouguer plate* to the right of the point, which is 100 m thick. Its vertical gravity effect is, as calculated above,

$$TC = 5.595 \,\text{mGal},$$

and, because a semi-plate is *removed* that is *above* the level of point Q, the algebraic sign of TC is again positive.

3. Free air to Bouguer:

$\Delta g_{\text{FA}}(P)$	50.000 mGal
TC	+5.595 mGal
Bouguer plate removal, 300 m	−33.570 mGal
$\Delta g_{\text{B}}(P)$	22.025 mGal
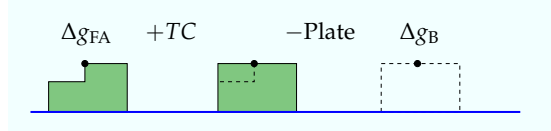
4. Bouguer to free air:

$\Delta g_{\text{B}}(Q)$	22.000 mGal
Bouguer plate addition, 200 m	+22.380 mGal
TC "uncorrection"	−5.595 mGal
$\Delta g_{\text{FA}}(Q)$	38.785 mGal
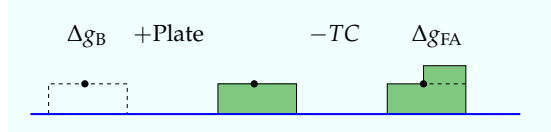
6.4 Spherical Bouguer anomalies

More recently, also *spherical Bouguer anomalies* have been calculated, e.g., Balmino et al. (2012); Kuhn et al. (2009); Hirt and Kuhn (2014). In this calculation, the topography *and bathymetry* of the whole Earth is taken into account, in spherical geometry (the error caused by neglecting the Earth's flattening is in this calculation negligible). This causes four differences with the Bouguer *plate* anomalies:

1. the attraction of a Bouguer shell of thickness H is $4\pi G\rho H$, twice as much as the corresponding Bouguer plate attraction. The remote part of the shell contributes as much attraction as the neighbourhood of the evaluation point!

2. The bathymetry of the oceans is accounted for[3] by replacing the water by standard-density crustal rock; this contribution to the anomalies is *positive.*

3. Also the topography and bathymetry of remote parts of the globe are taken into account realistically. As most of the Earth is covered by deep ocean, this causes a strong *positive general bias*, which in moderately elevated areas like Southern Finland more than cancels the negative one caused by the local topography!

4. As now also the terrain correction is calculated over the whole globe, in spherical geometry, it is no longer a small number and may be strongly negative as well as positive.

Between the planar and spherical Bouguer anomalies exists a large systematic difference, which however is very long-wavelength in nature, and even in an area the size of Australia almost a constant, -18.6 mGal within a variation interval of a few milligals. The details in the Bouguer maps look the same (Kuhn et al., 2009).

Just for fun, we compute the net mass effect of doing the complete spherical Bouguer reduction globally. The mean height of the land topography is

---

[3]One can do so, and often does, also in connection with the Bouguer plate correction.

800 m, land occupying 29% of the globe. The mean ocean depth is 3700 m, corresponding to an equivalent rock depth to be "filled in" of

$$3700 \times \frac{2.67 - 1.03}{2.67}\,\text{m} = 2272\,\text{m},$$

assuming a density for crustal rock of 2670 kg/m³ and a sea-water density of 1030 kg/m³, and ocean occupying 71% of the globe. The sum weighted by area is thus

$$(0.29 \times 800 - 0.71 \times 2272)\,\text{m} = -1381\,\text{m}.$$

Interpretation: there is not enough topography to fill all of the oceans, even if we're allowed to compress sea water into standard crustal rock. If we try this bulldozing experiment, we'll end up 1381 m short.

If, instead, we add standard crustal rock to end up at current sea level — the definition of spherical Bouguer reduction! — we'll add to the Earth's attraction as sensed from space an amount $4\pi G\rho \times 1381$ m = 309 mGal.

The global mean planar Bouguer reduction, as well as the difference between planar and spherical Bouguer reductions, on average over the globe, will be half of this, $\approx +155$ mGal. Even coastal locations, at zero height above sea level, will have similarly positive spherical Bouguer anomalies!

6.5 Helmert condensation

An often used method, proposed by Friedrich Robert Helmert[4] , for removing the effect of the masses exterior to the geoid is *condensation*. In this method, we shift mathematically all the continental masses vertically downward to mean sea level into a simple mass density layer $\kappa = H\rho$ — more precisely:

$$\kappa = H\rho \left(1 + \frac{H}{R}\right) \tag{6.4}$$

on a spherical Earth, radius R — where H is the height of the topography above sea level and ρ its mean density. The advantage of Helmert
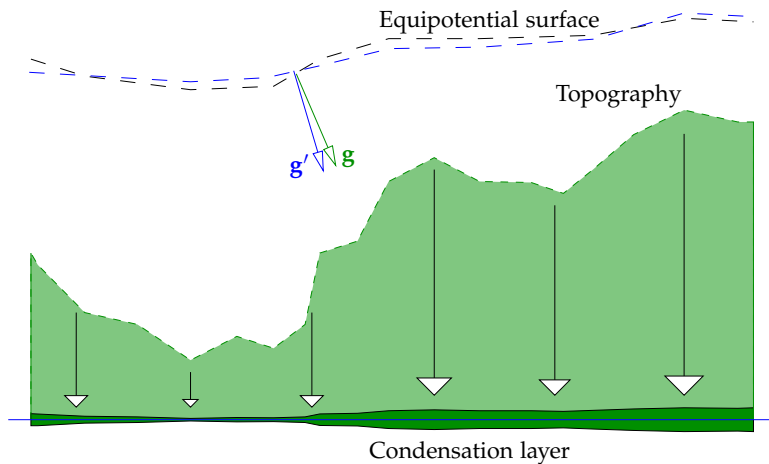
Figure 6.8. Helmert condensation and the changes it causes in the gravity field.

condensation over Bouguer reduction is, that *no mass is being removed*. The Bouguer reduction amounts to the computational removal of topographic masses on a large scale. Therefore, unlike with Bouguer reduction, in Helmert condensation gravity anomalies will not change systematically.

In appendix D we derive series expansions in spherical geometry, which describe both the external and the internal potential as functions of the "degree constituents" of the various powers of the topography $H(\phi, \lambda)$. The extensively presented derivation in the appendix is much used in gravity field theory to model the gravity effect of the topography. In this theory, issues of convergence are difficult, though we gloss over those here.

---

[4]Friedrich Robert Helmert (1843 – 1917) was an eminent German geodesist known for his work on mathematical and statistical geodesy.
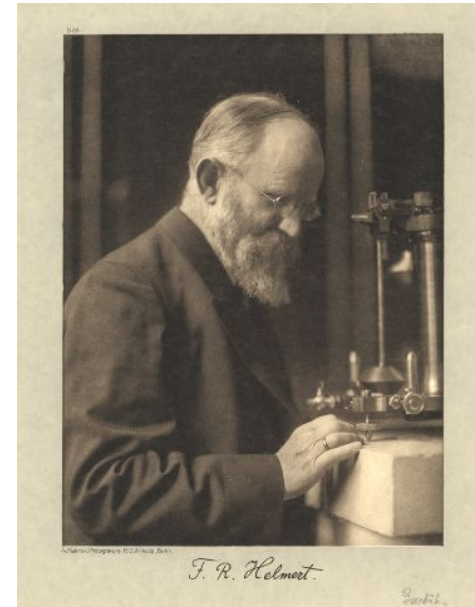
Figure 6.9. Friedrich Robert Helmert. Humboldt University Berlin (2017).

6.6 Isostasy

6.6.1 Classical hypotheses

Already in the 18th and 19th centuries, e.g., thanks to Bouguer's work in South America, as well as that of British geodesists in the Indian Himalayas, it was understood that mountain ranges weren't just piles of rock on top of the Earth's crust. The gravity field surrounding the mountains, specifically the plumb line deflections, could only be explained by assuming that under every mountain range there was also a "root" made from lighter rock species. The origin of this root was speculated to be the almost hydrostatic behaviour of the Earth's crust over geological time scales. This assumption of hydrostatic equilibrium was called the *hypothesis of isostasy*, also *isostatic compensation*.

Back then, unlike now, it was not yet possible to get a precise or even correct picture using physical methods (seismology) of how these mountain
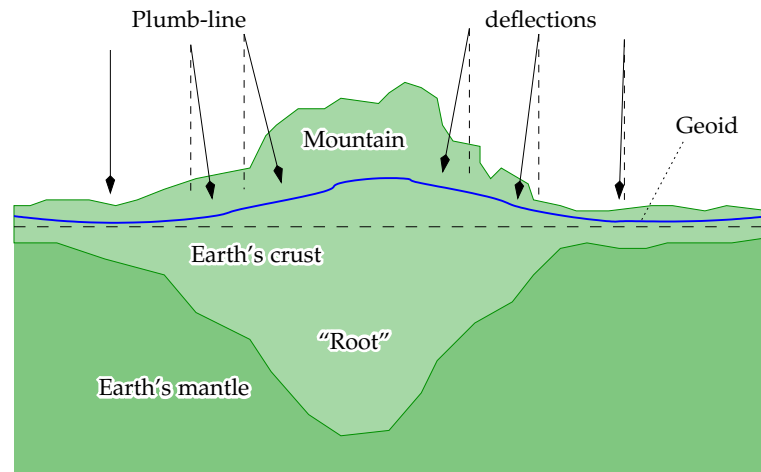
Figure 6.10. Isostasy and the bending of plumb lines towards the mountain.

Figure 6.11. Pratt–Hayford isostatic hypothesis.

roots were really shaped. That's why simplified working hypotheses were formulated.
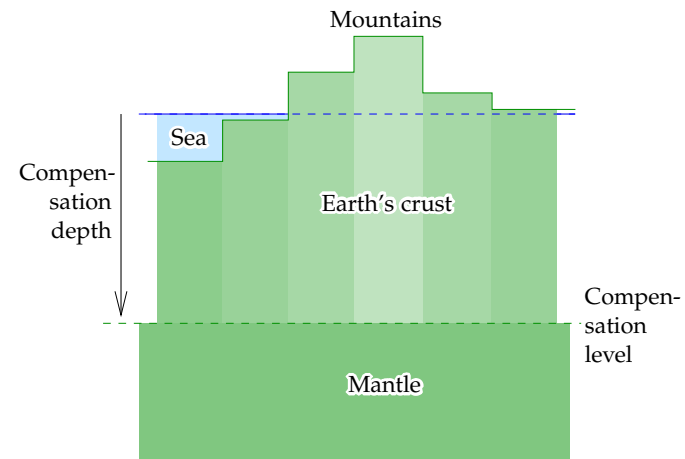
One older isostatic hypothesis is the Pratt–Hayford hypothesis. This was proposed by J. H. Pratt[5] in the middle of the 19th century (Pratt, 1855, 1859, 1864), and J.F. Hayford[6] developed the mathematical tools needed for computation. According to this hypothesis, the density of the "root" under a mountain would vary with the height of the mountain, so that under the highest mountains would be the lightest material, and the boundary between this light root material and the denser Earth mantle material would be at a fixed depth. This model, which nowadays finds little acceptance anymore, is illustrated in figure 6.11.

---

[5]John Henry Pratt (1809 – 1871) was a British clergyman and mathematician who worked as the archdeacon of Kolkata, India. Wikipedia, John Pratt.

[6]John Fillmore Hayford (1868 – 1925) was a United States geodesist who studied isostasy and the figure of the Earth.

Another classical isostatic hypothesis is due to G. B. Airy[7]. Because V. A. Heiskanen[8] used it extensively and developed its mathematical form, it is called the Airy–Heiskanen model. In this model it is assumed that the mass density of the "root" is fixed, and that the isostatic compensation is realized by varying the depth to which the root extends down into the Earth's mantle. In our current understanding this corresponds better to what is really happening inside the Earth. This hypothesis is illustrated in figure 6.12.

6.6.2   Calculation formulas

Airy's isostatic hypothesis assumes that in every place the total mass of a column of matter is the same. So, let the density of the Earth crust be ρ_c,

---

[7]George Biddell Airy PRS (1801 – 1892) was an English mathematician and astronomer, "Astronomer Royal" 1835 – 1881.

[8]Veikko Aleksanteri Heiskanen (1895 – 1971), "the great Heiskanen" (Hermans, 2007). was an eminent Finnish geodesist who also worked in Ohio, USA He is known for his work on isostasy and global geoid modelling. See Kakkuri (2008).
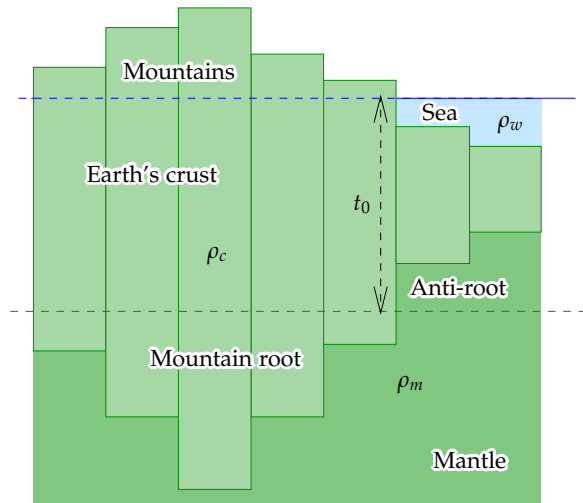
Figure 6.12. Airy–Heiskanen isostatic hypothesis.

the density of the mantle ρ_m, and the density of sea water ρ_w; sea depth d, crustal thickness t and topographic height H. We have

$$t\rho_\mathrm{c} + d\rho_\mathrm{w} - (t+d)\,\rho_\mathrm{m} = c \implies t = -\frac{d\,(\rho_\mathrm{m} - \rho_\mathrm{w}) + c}{\rho_\mathrm{m} - \rho_\mathrm{c}}$$

on the sea, and

$$t\rho_\mathrm{c} - (t - H)\,\rho_\mathrm{m} = c \implies t = \frac{H\rho_\mathrm{m} - c}{\rho_\mathrm{m} - \rho_\mathrm{c}}$$

on land. c is a suitable constant[9]. Here we have conveniently forgotten about the curvature of the Earth, i.e., we use the "flat Earth model".

Under land, the depth of a mountain root is

$$r = t - H = \frac{H\rho_\mathrm{m} - c}{\rho_\mathrm{m} - \rho_\mathrm{c}} - \frac{H\rho_\mathrm{m} - h\rho_\mathrm{c}}{\rho_\mathrm{m} - \rho_\mathrm{c}} = \frac{H\rho_\mathrm{c} - c}{\rho_\mathrm{m} - \rho_\mathrm{c}}.$$

---

[9]Its dimension, after multiplication with ambient gravity g, is *pressure*: according to Archimedes' law, the pressure of the crustal (plus sea-water) column minus the pressure of the column of displaced mantle material.
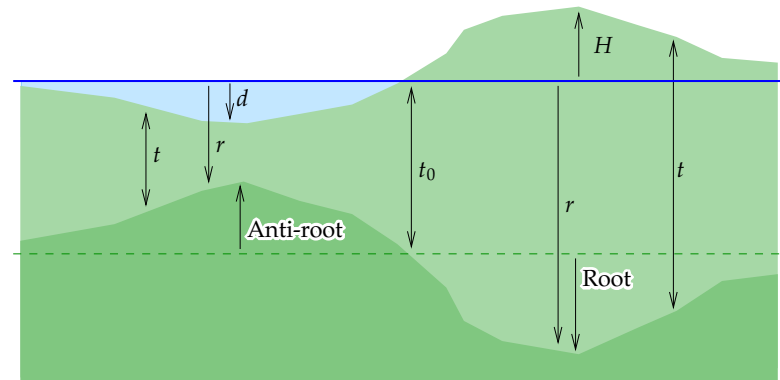

Figure 6.13. Quantities in isostatic compensation.

Similarly under the sea

$$r = t + d = -\frac{d\,(\rho_\mathrm{m} - \rho_\mathrm{w}) + c}{\rho_\mathrm{m} - \rho_\mathrm{c}} + \frac{d\rho_\mathrm{m} - d\rho_\mathrm{c}}{\rho_\mathrm{m} - \rho_\mathrm{c}} = -\frac{d\,(\rho_\mathrm{c} - \rho_\mathrm{w}) + c}{\rho_\mathrm{m} - \rho_\mathrm{c}}.$$

Note that the constant c is arbitrary and expresses the fact that the level from which one computes the depth of the root — less precisely, the "average thickness of the crust" — can be chosen arbitrarily.

Another approach: instead of c, use the "zero topography compensation level" t_0, to be computed from the above equations by setting $H = d = 0$:

$$t_0\,(\rho_\mathrm{c} - \rho_\mathrm{m}) = c.$$

This yields under the land the root depth

$$r = \frac{H\rho_\mathrm{c} - t_0\,(\rho_\mathrm{c} - \rho_\mathrm{m})}{\rho_\mathrm{m} - \rho_\mathrm{c}} = t_0 + H\frac{\rho_\mathrm{c}}{\rho_\mathrm{m} - \rho_\mathrm{c}}, \tag{6.5}$$

and under the sea

$$r = -\frac{d\,(\rho_\mathrm{c} - \rho_\mathrm{w}) + t_0\,(\rho_\mathrm{c} - \rho_\mathrm{m})}{\rho_\mathrm{m} - \rho_\mathrm{c}} = t_0 - d\frac{\rho_\mathrm{c} - \rho_\mathrm{w}}{\rho_\mathrm{m} - \rho_\mathrm{c}}, \tag{6.6}$$

somewhat simpler equations that are also more intuitive:

$$H\rho_\mathrm{c} + (-r)\,(\rho_\mathrm{m} - \rho_\mathrm{c}) = -t_0;$$
$$(-d)\,(\rho_\mathrm{c} - \rho_\mathrm{w}) + (-r)\,(\rho_\mathrm{m} - \rho_\mathrm{c}) = -t_0.$$

In other words,

$$\sum_{\text{interfaces}} (\text{deviation} \times \text{density contrast}) = \text{const.}$$

However, the effect of the different isostatic hypotheses on gravity is pretty much the same: the hypotheses can not be distinguished based on gravity measurements only. The effect of the hypothesis on the geoid is stronger.

6.6.3   Example: Norway

The Southern Norwegian *Hardangar plateau* (*Hardangarvidda*) is a highland at, on average, 1100 m above sea level. It is the largest peneplain in Europe, a national park and a popular tourist attraction, being traversed by *Bergensbanen*, the highest regular railway in Northern Europe.

puolitasanko

The *Norwegian Sea* is the part of the Atlantic Ocean adjoining Norway, and does not belong to the continental shelf. It is on average 2 km deep.

mannerjalusta

Questions:

1. What is the depth of the root under the Hardanger plateau, relative to the compensation depth t_0?

2. What is the negative depth of the anti-root under the Norwegian Sea, relative to the same compensation depth?

3. What is the *relative* depth of the root of the Hardanger plateau, compared to the nearby Norwegian Sea?

Answers:

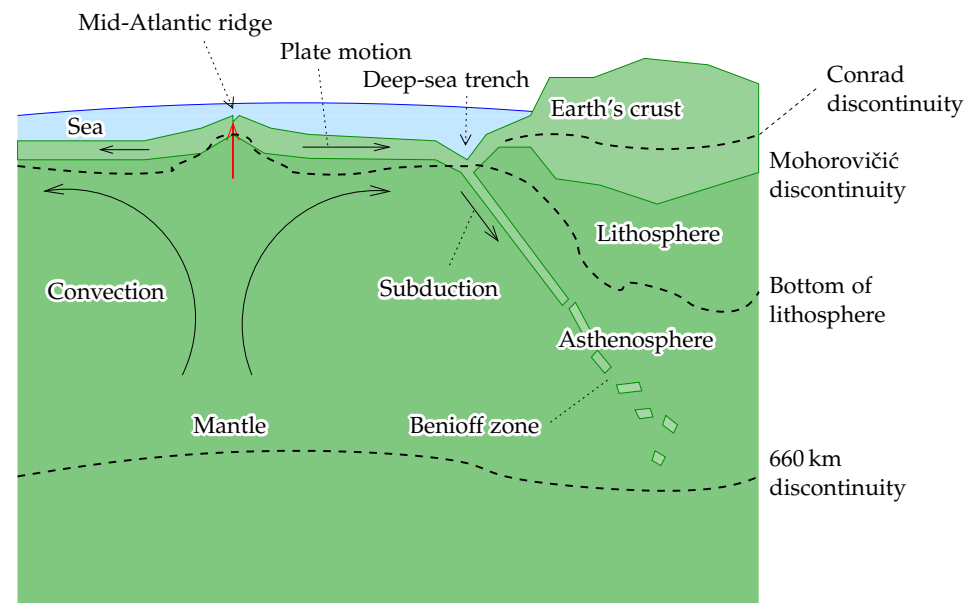1. We use the equation 6.5, finding

$$r - t_0 = H \frac{\rho_c}{\rho_m - \rho_c} = 1100\,\text{m} \times \frac{2670\,\text{kg/m}^3}{3370 - 2670\,\text{kg/m}^3} = 4196\,\text{m}.$$

   Here we have used standard densities for crustal and mantle rock, respectively.

2. We use the equation 6.6, finding

$$r - t_0 = -d \frac{\rho_c - \rho_w}{\rho_m - \rho_c} = -2000\,\text{m} \times \frac{2670 - 1027\,\text{kg/m}^3}{3370 - 2670\,\text{kg/m}^3} = -4694\,\text{m},$$

   using the standard density value for sea water.

Figure 6.14. The modern understanding of isostasy and plate tectonics. Deep-sea trenches are known to be in isostatic disequilibrium.

3. The depth contrast between root and anti-root is $4196 - (-4694)$ m $= 8890$ m (for perspective, Mount Everest is 8848 m above sea level).

_____

6.6.4   The modern understanding of isostasy

Nowadays we have a much better understanding of the internal situation in the Earth. However, isostasy continues to be a valid concept. A more realistic picture of the internal structure of the Earth is given in figure 6.14.

An important subject for current research is the effect on vertical motion of the Earth's crust of the growing and melting of the ice masses of the Earth, like the continental ice sheets. To this belongs both the direct effect of the varying ice masses, and the effect of the changes caused in the water

mannerjäätiköt

masses of the ocean. So-called paleo-research concentrates on the changes over the glacial cycle, while modern retreats of glaciers, e.g., in Alaska and on Spitsbergen, cause their own, observable local uplift of the Earth's crust. More in chapter 12.

6.6.5 Example: Fennoscandian land uplift

During the last glacial maximum, some 20 000 years ago, Fennoscandia was covered by a continental ice sheet of thickness up to 3 km.

Questions:

1. How much was the Earth's surface depressed by this load, assuming isostatic equilibrium?

2. Currently the land is rising in central Fennoscandia, where the ice thickness was maximal, at a rate of 10 mm/a. How long would it take at this rate for the depression to disappear?

Answers:

1. We assume for the ice density a value of 920 kg/m³. Then, with an upper mantle density of 3370 kg/m³ — note that it is Earth's mantle material that is being displaced by the ice, the Earth's crust just transmits the load! See figure 12.1a — we find for the depression:

$$\Delta H = 3000 \, \text{m} \times \frac{920 \, \text{kg/m}^3}{3370 \, \text{kg/m}^3} = 819 \, \text{m}.$$

2. At the rate of 10 mm/a it will take $^{819\,\text{m}}\!/_{0.01\,\text{m/a}} = 81\,900$ years total. Part of this uplift has already taken place since the last deglaciation.

   In reality, of course, the rate has decreased substantially, and will continue to decrease, over time.

———

6.7 Isostatic reductions

The computational removal of both the topography and its isostatic compensation from the measured quantities of the gravity field is called *isostatic reduction*. It serves two purposes.

1. By removing as many as possible "superficial" effects from the gravity field, we are left with a field where only the effect of the Earth's deep layers remains. This is useful for geophysical studies.

2. These "superficial" effects are also generally very local: in spectral language, very *short wavelength*. By removing those, we are left with a residual field that is much smoother, and that can be interpolated or *predicted* better. This is important especially in areas where there is a paucity of real measurement data, like the oceans, deserts, polar areas etc.

For example, *isostatic anomalies*, i.e., free-air anomalies to which isostatic reduction has been applied, are very smooth (like also Bouguer anomalies), and their *predictive properties* are good. However, unlike Bouguer anomalies, isostatic anomalies are on average zero. They lack the large bias that makes Bouguer anomalies strongly negative especially in mountainous areas (see section 6.2). This of course is because isostatic reduction is only the *shifting* of masses from one place to another — from mountains to roots beneath the same mountains, the mass deficit of which is pretty precisely the same as the mass of the mountains themselves sticking out above sea level — rather than *removal* of masses, which is what Bouguer reduction does.

The reduction methods used in isostatic calculations are the same as in other reductions, and we will discuss them later: numerical integration in the space domain — grid integration, spherical-cap integration, least-squares collocation (LSC), finite elements, etc. — or in the spectral domain (FFT, "Fast Collocation", etc.).

The question of the hypothesis assumed to apply is a more interesting one. Traditionally, the Pratt or Airy hypotheses have been used, developed into quantitative methodologies by Hayford or Heiskanen or Vening Meinesz[10]. A newer approach has been to use real measurement data from *seismic*

———

[10]Felix Andries Vening Meinesz (1887 – 1966) was a Dutch geodesist and gravimetrist. He wrote together with V. A. Heiskanen the textbook *The Earth and its Gravity Field* (1958).
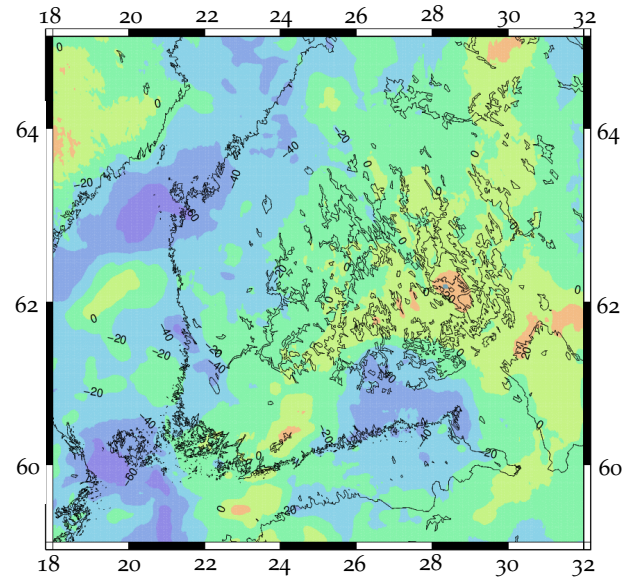
Figure 6.15. Isostatic gravity anomalies for Southern Finland. Airy–Heiskanen hypothesis, compensation depth 30 km. Data © Bureau Gravimétrique International (BGI) / International Association of Geodesy, World Gravity Map project. Web service BGI, WGM2012. Note that here, on the thick, rigid Fennoscandian Shield, the local features of the topography are not isostatically compensated and the map looks rather similar to the free-air anomaly map 5.5 on page 118.

tomography in order to model the interior structure of the Earth. With real measurement data, if reliable, one should get better results.

6.8   The "isostatic geoid"

Let us look at how the "isostatic geoid", more precisely the *co-geoid of isostatic reduction*, is computed. Isostatic reduction is one possible method for com-

putationally removing the masses outside the geoid, in order to formulate a boundary-value problem on the geoid.

We can show (Heiskanen and Moritz, 1967 page 142), that the isostatic co-geoid is under the continents as much as several metres below the geoid, i.e., the indirect effect ("Restore" step) is of this order. Under the oceans, similarly the isostatic co-geoid is somewhat above the geoid.

As one of the requirements for geoid determination methods is a small indirect effect, it follows that isostatic methods are not (contrary to what is said on Heiskanen and Moritz page 152) the best possible if the intent is to calculate a model of the geoid or quasi-geoid representing the exterior potential[11]. However, isostatic methods are very suitable for elucidating the interior structure of the Earth, because both the topography and the "impression" it makes on the Earth's mantle, the isostatic compensation, are computationally removed. Research has shown that the great topographic features of the Earth are some $85 - 90\%$ isostatically compensated. This is valuable information if no other knowledge is available.

This is the second reason why the isostatic geoid is of interest: the gravity field of an Earth from which the effect of mountains has been removed completely — mountain roots and all — can uncover physical unbalances existing in deeper layers, and processes causing these. Such processes are especially convection currents in the Earth's mantle as well as the possible effect of the liquid core of the Earth on these currents. Interesting correlations have been found between mantle convection patterns, the global map of the geoid, and the electric current patterns in the core causing the Earth's magnetic field (Wen and Anderson, 1997; Prutkin, 2008; Kogan et al., 1985).

Isostatic reduction consists of two parts:

1. computational removal of the topography
2. computational removal of the isostatic compensation of the topography.

It is possible to calculate both these parts exactly using prism integration, see section 6.3. Here however we shall gain understanding by a qualitative

---

[11]Of course Bouguer reduction is even worse! The indirect effect can be hundreds of metres.

approach. We *approximate* both parts with a single mass density layer, density, e.g., $\kappa = \rho H$ for the topography. We place the first layer at level $H = 0$, and the second, density $-\kappa$, at *compensation depth* $H = -D$. The situation is depicted in figure 6.16.

In the following we use the "generating function" equation 8.5,

$$\frac{1}{\ell} = \frac{1}{R}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi),$$

together with the single mass density layer equation 1.15:

$$V = G\iint_{\text{surface}}\frac{\kappa}{\ell}\,dS = GR^2\iint_{\text{surface}}\frac{\kappa}{\ell}\,d\sigma.$$

We obtain for the potential field mass density layer at sea level, when also the evaluation point is placed at sea level, $H = 0 \implies r = R$:

$$T_{\text{top}} = GR\iint_{\sigma}\kappa\sum_{n=0}^{\infty}P_n(\cos\psi)\,d\sigma$$

and with the density layer at compensation depth (source level $R - D$, evaluation level R):

$$T_{\text{comp}} = \frac{GR^2}{R-D}\iint_{\sigma}(-\kappa)\sum_{n=0}^{\infty}\left(\frac{R-D}{R}\right)^{n+1}P_n(\cos\psi)\,d\sigma =$$

$$= -GR\iint_{\sigma}\kappa\sum_{n=0}^{\infty}\left(\frac{R-D}{R}\right)^{n}P_n(\cos\psi)\,d\sigma,$$

from which the combined effect ($n = 0$ drops out)

$$\delta T_{\text{iso}} = -\left(T_{\text{top}} + T_{\text{comp}}\right) = -GR\iint_{\sigma}\kappa\sum_{n=1}^{\infty}\left[1 - \left(\frac{R-D}{R}\right)^{n}\right]P_n(\cos\psi)\,d\sigma.$$

$$(6.7)$$

Here, the mass density per unit of surface area κ is

$$\kappa = \begin{cases}\rho_c H & \text{if } H \geq 0, \\ (\rho_c - \rho_w)H & \text{if } H < 0,\end{cases}$$

i.e., we replace ocean depths by equivalent "dry" depths[12]. Now we use

---

[12]This works on dry land and on the ocean. Lakes, glaciers and areas like the Dead Sea are more complicated.
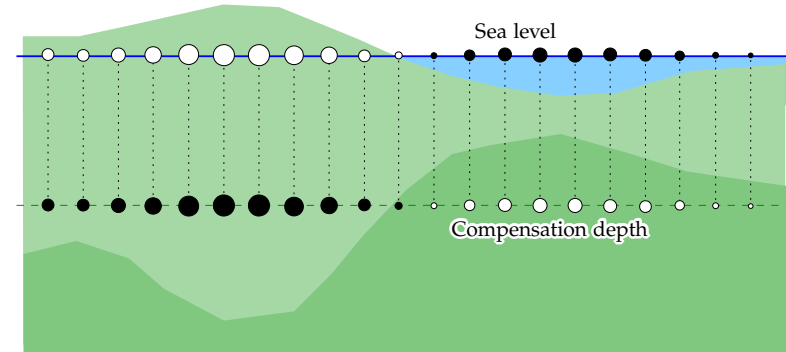

Figure 6.16. Isostatic reduction as a pair of surface density layers.

again the degree constituent equation, Heiskanen and Moritz (1967) equation 1-71, or our equation 3.7, in the following form:

$$\kappa_n(\phi, \lambda) \overset{\text{def}}{=} \frac{2n+1}{4\pi}\iint_{\sigma}\kappa(\phi', \lambda')P_n(\cos\psi)\,d\sigma.$$

Multiplying both sides of this with the factor

$$\frac{4\pi}{2n+1}\left[1 - \left(\frac{R-D}{R}\right)^{n}\right]$$

and moving it inside the integral, we obtain

$$-\frac{4\pi GR}{2n+1}\left[1 - \left(\frac{R-D}{R}\right)^{n}\right]\kappa_n(\phi, \lambda) =$$

$$= -GR\iint_{\sigma}\kappa(\phi', \lambda')\left[1 - \left(\frac{R-D}{R}\right)^{n}\right]P_n(\cos\psi)\,d\sigma.$$

Summation yields the expression 6.7 above:

$$-\sum_{n=1}^{\infty}\frac{4\pi GR}{2n+1}\left[1 - \left(\frac{R-D}{R}\right)^{n}\right]\kappa_n(\phi, \lambda) =$$

$$= -GR\iint_{\sigma}\kappa(\phi', \lambda')\sum_{n=1}^{\infty}\left[1 - \left(\frac{R-D}{R}\right)^{n}\right]P_n(\cos\psi)\,d\sigma,$$

so that

$$\delta T_{\text{iso}} = -\sum_{n=1}^{\infty} \frac{2}{2n+1} R \left[1 - \left(\frac{R-D}{R}\right)^n\right] 2\pi G \kappa_n =$$

$$= -\sum_{n=1}^{\infty} \frac{2}{2n+1} R \left[1 - \left(\frac{R-D}{R}\right)^n\right] [A_{\text{B}}]_n.$$

Here we have used the notation $A_{\text{B}} = 2\pi G \kappa$, the equivalent Bouguer plate attraction of a mass density layer κ, and its degree constituent $[A_{\text{B}}]_n = 2\pi G \kappa_n$.

Let us first look at the contribution from[13] $0 < n \le N = \frac{R}{D}$. Then the following approximation holds, as $\left(\frac{R-D}{R}\right)^n \approx 1 - \frac{nD}{R}$:

$$\delta T_{\text{iso}} \approx -\sum_{n=1}^{N} \frac{2nD}{2n+1} [A_{\text{B}}]_n \approx -\sum_{n=1}^{N} D [A_{\text{B}}]_n \approx -D A_{\text{B}},$$

and

$$\delta N_{\text{iso}} = \frac{\delta T_{\text{iso}}}{\overline{\gamma}} \approx -\frac{D A_{\text{B}}}{\overline{\gamma}}. \tag{6.8}$$

This is the *indirect effect of isostatic reduction*.

Let's substitute realistic values. Let the Mohorovičić[14] discontinuity's depth be on average $\sim 20\,\text{km}$[15].

On land $H \approx 0.8\,\text{km}$, The Earth's mean topographic height, and we obtain
$\delta N_{\text{iso}} \approx -1.8\,\text{m}$.

---

[13]The contribution from degree numbers $n > R/D$ is

$$\delta T_{\text{iso}} \approx -\sum_{n=N+1}^{\infty} \frac{2R}{2n+1} [A_{\text{B}}]_n,$$

where the terms are small and rapidly falling to zero. In this degree range also the mass density layer approximation for the topography breaks down.

[14]Andrija Mohorovičić (1857−1936) was a Croatian meteorologist and a pioneer of modern seismology.

[15]Under the continents 35 km, under the oceans 7 km below the sea floor, according to *Encyclopædia Brittanica*. Using these values, we find $\delta N_{\text{iso}} = -3.2\,\text{m}$ on land, $+2.8\,\text{m}$ on the ocean.

On the ocean $H \approx -3.7\,\text{km}$ on average. We must remember to multiply by $\frac{2.67 - 1.03}{2.67}$ because of the water, and we obtain $\delta N_{\text{iso}} \approx +5.0\,\text{m}$.

In other words, *this effect can be sizable*!

Note that equation 6.8 is *linear* in the height H. This means that, under the continents, the isostatic co-geoid will run on the order of a couple of metres below the classical geoid, when on the oceans again it must be a few metres above the geoid (mean sea level). We may also conclude that in the isostatic reduction's effect on the geoid – at least for longer wavelengths $2\pi R/n$, longer than the compensation depth D — all wavelengths are represented in the spectrum in approximately in the same proportions as in the topography itself, and the effect is in fact proportional to the topography.

Self-test questions

1. Which effects are computationally removed in
   (a) the simple Bouguer reduction?
   (b) the terrain corrected Bouguer reduction?
   (c) the isostatic reduction?

2. Why is the terrain correction always positive?

3. Why do Bouguer anomalies have good interpolation properties, and on what condition (i.e., which additional information must be available at the interpolation stage)?

4. How was it discovered that mountains have roots?

5. Describe the isostatic hypotheses of Pratt–Hayford and Airy–Heiskanen.

Exercise 6 − 1: Gravity anomaly

Given point P, height above sea level $H = 500\,\text{m}$. Local gravity is $g_P = 9.82\,\text{m/s}^2$, normal gravity at sea level for local latitude φ is $\gamma_0(\varphi) = 9.820192\,\text{m/s}^2$.

1. Compute point P's free-air anomaly Δg.

2. Compute point P's Bouguer anomaly (without terrain correction) Δg_B.

Exercise 6 – 2: Bouguer reduction

1. Point P is 500 m above sea level. Its free-air anomaly is $\Delta g_{FA} = 25$ mGal. Calculate the Bouguer anomaly Δg_B of the point. Forget about the terrain correction.

2. See section 6.2: Bouguer anomalies. Derive the equations 6.2 and 6.3 anew, assuming that the mean density of the Earth crust were $\rho = 3370$ kg/m³.

Exercise 6 – 3: Terrain correction, Bouguer reduction

Given the terrain shape:

P

Q

Q' Sea level
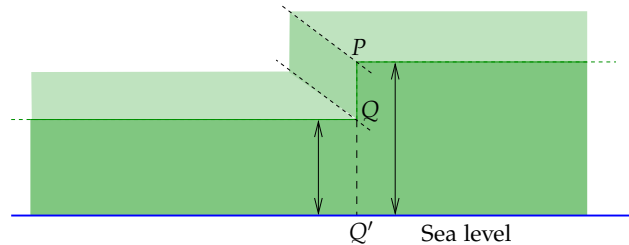
The vertical rock wall PQ is also straight on a map and extends in both directions ("into" and "out of" the paper) to infinity.

Height differences: $PQ' = 600$ m, $QQ' = 300$ m.

1. Compute in point P the *terrain correction* (hint: use the formula for the attraction of a Bouguer plate. We have here a *half* Bouguer plate, with only half the attraction of a full one.)

2. Compute in point Q the terrain correction. Algebraic sign?

3. If in point P is given that the free-air anomaly is 60 mGal, how much is then the Bouguer anomaly in the point? (Use the complete Bouguer reduction.)

4. If it is given in point Q that the Bouguer anomaly is 10 mGal, how much is the point's free-air anomaly?

Exercise 6 – 4: Isostasy

Assume Airy-Heiskanen isostatic compensation (figure 6.12). Density of Earth's crust $\rho_c = 2670$ kg/m³, density of mantle $\rho_m = 3370$ kg/m³, i.e., the density contrast at the crust-mantle interface is 700 kg/m³. Let the *reference level* for the interface corresponding to zero topography be -25 km, i.e. $t_0 = 25$ km.

1. Calculate the depth of the "root" of an 8 km high mountain below the reference level -25 km, assuming it is isostatically compensated.

2. Mauna Kea is 4 km above sea level, however the surrounding sea is 5 km deep. How deep is the "root" of Mauna Kea below the reference level?

3. How much is the "anti-root" of the surrounding sea *above* the reference level? Let the density of sea water be 1027 kg/m³.

4. So, how deep is the "root" of Mauna Kea *relative to its surroundings*?

Vertical reference systems

7

7.1 Levelling, orthometric heights and the geoid

Heights have traditionally been determined by *levelling*. Levelling is a technique for determining height differences using a level (levelling instrument) and two rods or staffs. The level comprises a telescope and a spirit level, and in the measurement situation the telescope's optical axis, the sight axis, is pointing along the local horizon. Levelling staffs are placed on two measurement points, and through the measuring telescope, measurement values are read off them. The difference between the two values gives the height difference between the two points in metres.

The distance between level and staffs is $40 - 70\,\text{m}$, as longer distances would cause too large errors due to the effect of atmospheric refraction. Longer distances are measured by repeat measurements using several instrument stations and intermediate points.

The height differences ΔH thus obtained are not, however, directly useable. The "height difference" between two points P and Q, obtained by directly summing height differences ΔH, depends namely also on the path chosen when levelling from P to Q. Also the sum of height differences $\sum_{\bigcirc} \Delta H$ around a closed path *is generally not zero*.
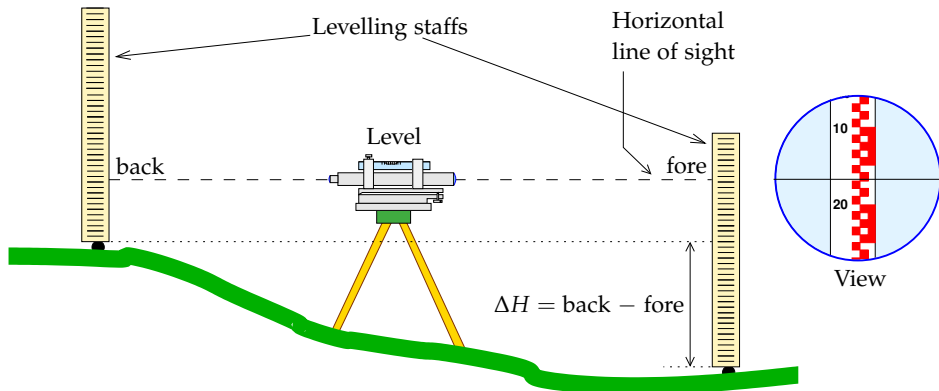
> *Geometric height is not a conservative field.*

Figure 7.1. The principle of levelling.

This is why, in precise levelling, the height differences are always converted to *potential differences*: $\Delta W = -\Delta H \cdot g$, in which g is the local gravity, which is either measured or — e.g., in Finland — interpolated from an existing gravity survey data base. The sum of potential differences around a closed loop *is always zero*: $\sum_{\bigcirc} \Delta W = 0$.

For the potential of an arbitrary terrain point P we find

$$W_P = W_0 - \sum_{\text{sea level}}^{P} (\Delta H \cdot g),$$

the summation being done directly from sea level (potential W_0) up to point P. The quantity

$$C_P = -(W_P - W_0) = \sum_{\text{sea level}}^{P} (\Delta H \cdot g),$$

geopotentiaaliluku positive above sea level, is called the *geopotential number* of point P.

W_0 is the potential of the national height reference level. In Finland, the reference level of the old N60 system is in principle mean sea level in Helsinki harbour at the beginning of 1960, which is why the system is called N60. However, the precise realization is a special pillar in the garden of Helsinki

astronomical observatory in Kaivopuisto[1]. The new Finnish height system is called N2000, and the realization of its reference level is a pillar at the Metsähovi research station. In practice N2000 heights are, at the decimetric precision level, heights over the Amsterdam NAP datum.

Other countries have their own, similar height reference or datum points: Russia has Kronstadt, Western Europe the widely used Amsterdam datum NAP, Southern Europe has Trieste, North America the North American Vertical Datum 1988 (NAVD88), datum point Father Point (Pointe-au-Père)[2] in Rimouski, Quebec, Canada, etc.

7.2 Orthometric heights

For creating a vertical reference, it would be simplest to use the original geopotential differences from sea level, i.e., the geopotential numbers defined above, $C = -(W - W_0)$, directly as height values. However, this is psychologically and practically difficult: people want their heights to be in metres. Geopotential numbers have their clear advantages: they represent the *amount of energy* that is needed (for a unit test mass) to move to the point from the reference level. Fluids (sea water, but also air, or, on geological time scales, even bedrock!) flow always downward and seek the state of minimum energy.

In Finland, as in many other countries, *orthometric heights* have been long in use. They are physically defined heights above "mean sea level" or the *geoid*. See figure 7.3.

The classical *geoid* is defined as

---

[1]However, the value engraved in the pillar is the reference height of the still older system NN, not of N60. The correct reference value for N60 for this pillar, 30.51376 m, is given in the publication Kääriäinen (1966).

[2]The district Pointe-au-Père of the city of Rimouski was named after the Jesuit priest Father Henri Nouvel (1621? – 1701?), who served forty years with the native population of New France, today's Quebec. Pointe-au-Père is also notorious as the location of the RMS *Empress of Ireland* shipwreck in 1914, in which over a thousand passengers perished.

Figure 7.2. Height reference pillar in the garden of Helsinki astronomical observatory in Kaivopuisto, Kääriäinen (1966). Text:

Suomen	*Utgångspunkt för*
tarkka	*precisionsnivellementet*
vaakituksen	*i Finland*
pääkiintopiste	*30,4652 m öfver noll*
30,4652 m yli nollan	

(Reference bench mark of precise levelling of Finland, 30.4652 m above zero).

> *"The level surface of the Earth's gravity field which fits on average best to mean sea level."*
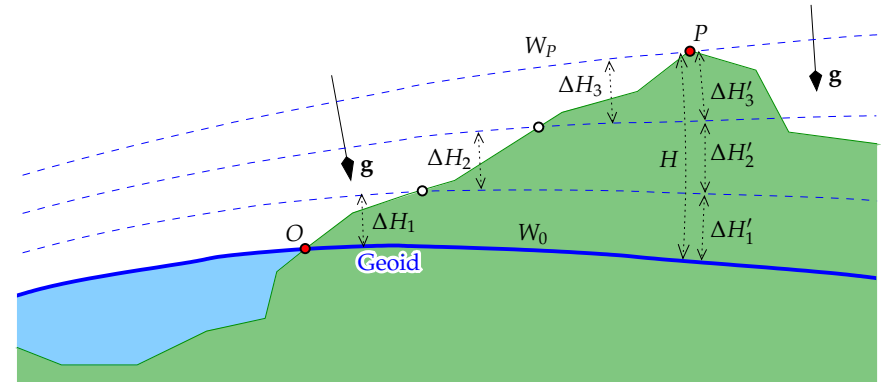
Figure 7.3. Levelled heights and geopotential numbers. The height obtained by summing levelled height differences, $\sum_{i=1}^{3} \Delta H_i$, is not the "correct" height above the geoid, i.e., $\sum_{i=1}^{3} \Delta H_i'$ computed along the plumb line.

Note how the equipotential or *level surfaces* of the geopotential are not parallel: because of this, a journey along the Earth's surface may well go "upward", to increasing heights above the geoid, although the geopotential number decreases. Thus, water may flow "upward".

The gravity vector $\mathbf{g}$ is everywhere perpendicular to the level surfaces, and its length is inversely proportional to the distance separating the surfaces.

The orthometric height H of point P is defined as the height obtained by measuring along the plumb line the *distance of P from the geoid*.

This is a very physical definition, hoewever not a very operational one, because we (generally) do not get to measure along a plumb line inside the Earth, and the geoid isn't visible there. This is why orthometric heights are calculated from geopotential numbers: if the geopotential number of point P is C_P, we calculate the orthometric height with the formula

$$H = \frac{C_P}{\overline{g}},$$
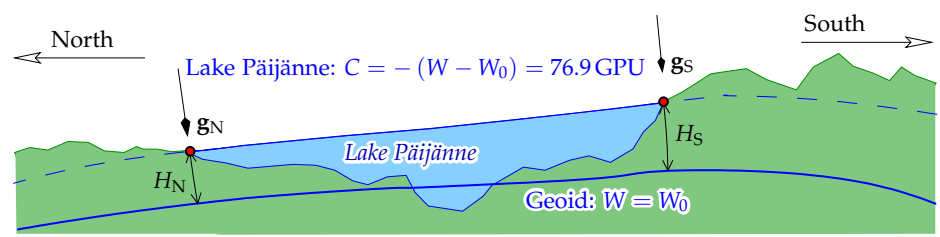
Figure 7.4. Considered in terms of orthometric heights, water may sometimes flow "upward". Although the North and South ends of Lake Päijänne are on the same geopotential level — 76.9 geopotential units below that of mean sea level — the orthometric height of the South end H_S is greater than that of the North end H_N, because local gravity g is stronger in the North than in the South. The height difference in the case of Lake Päijänne is 8 mm (Jaakko Mäkinen, personal comm.). Calculation using the normal gravity field yields 6 mm. The balance of 2 mm comes from the difference between gravity anomalies at the Northern and Southern ends.

where $\overline{g}$, *the average gravity along the plumb line*, is

$$\overline{g} = \frac{1}{H} \int_0^H g(z) \, dz,$$

and z is the measured distance from the geoid along the plumb line. Because the formula for $\overline{g}$ already itself contains H, we obtain the solution iteratively, using initially a crude estimate for H. The iteration converges fast.

We shall see that determining very precise orthometric heights is challenging, especially in the mountains.

7.3   Normal heights

In Finland, currently, with the height system N2000, *normal heights* are used. They are, like orthometric heights, heights *above mean sea level*. The mathematical representation of mean sea level in this case is the *quasi-geoid*. In sea areas, the quasi-geoid is identical to the geoid. Over land, it differs a little from the geoid, and in mountainous areas the difference may be substantial.

7.3.1   Molodensky's theory

M. S. Molodensky developed a theory in which the height of a point from "mean sea level" would be defined by the following equation:

$$H^* \stackrel{\text{def}}{=} \frac{C}{\overline{\gamma_{0H}}},$$

where $\overline{\gamma_{0H}}$ is the average normal gravity computed between the zero level (reference ellipsoid) and H^* along the *ellipsoidal normal*. So, the same way of computing as in the case of orthometric heights, but using the *normal gravity field* instead of the true gravity field.

Heights "above sea level" are for practical reasons given *in metres*. For large, continental networks we want to give heights above a computational reference ellipsoid in metres, and thus also heights above "sea level" have to be in metres.

Molodensky proposed also that instead of the geoid, *height anomalies* would be used, the definition of which is

$$\zeta \stackrel{\text{def}}{=} \frac{T}{\overline{\gamma_{Hh}}}, \tag{7.1}$$

where now $\overline{\gamma_{Hh}}$ is the average normal gravity at terrain level. More precisely: the average of normal gravity along the ellipsoidal normal over the interval $z \in [H^*, h]$, in which H^* is the normal height of the point and h its height from the reference ellipsoid. The parameter z is the distance from the reference ellipsoid reckoned along the ellipsoidal normal. T is the disturbing potential at the point.

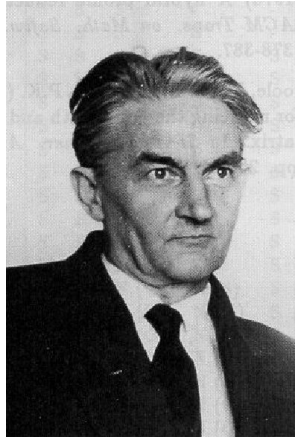Based on these assumptions, Molodensky showed that

$$H^* + \zeta = h.$$

This equation is very similar to the corresponding one for orthometric heights and geoid heights

$$H + N = h.$$

Also otherwise ζ, the *height anomaly*, also called "quasi-geoid height", is very close to N, and correspondingly H^* close to H.

we obtain

$$h = -\frac{U - U_0}{\overline{\gamma_{0h}}}.$$

By using $W = U + T$ and dividing by $\overline{\gamma_{0h}}$ we obtain

$$\frac{W - W_0}{\overline{\gamma_{0h}}} = \frac{T}{\overline{\gamma_{0h}}} - h$$

assuming $W_0 = U_0$, the normal potential on the reference ellipsoid.

Next, one could define

$$H^+ \overset{?}{=} -\frac{W - W_0}{\overline{\gamma_{0h}}}$$

as a new height type, and

$$N^+ \overset{?}{=} h - H^+ = \frac{T}{\overline{\gamma_{0h}}}$$

as the corresponding new geoid height type. It has however the esthetic flaw, that we divide here by the average normal gravity computed between the levels 0 and h. This quantity is not operational without a means of determining the ellipsoidal height h.

This suggests the following improvement based on the circumstance that $\gamma(z)$ is a nearly linear function. This means that the vertical derivative $\frac{\partial \gamma}{\partial r}$ is nearly constant in the height interval considered.

We define

$$\overline{\gamma_{0h}} \overset{\text{def}}{=} \frac{1}{h} \int_0^h \gamma(z)\ dz, \quad \overline{\gamma_{0H}} \overset{\text{def}}{=} \frac{1}{H^+} \int_0^{H^+} \gamma(z)\ dz, \quad \overline{\gamma_{Hh}} \overset{\text{def}}{=} \frac{1}{N^+} \int_{H^+}^h \gamma(z)\ dz.$$

Now

$$\overline{\gamma_{0H}} \approx \overline{\gamma_{0h}} - \frac{1}{2}N^+\frac{d\gamma}{dr} \approx \overline{\gamma_{0h}}\left(1 - \frac{N^+}{R}\right) \tag{7.2}$$

(R is the Earth's radius, and $\frac{d\gamma}{dr} \approx \frac{2\gamma}{R}$ is assumed constant), and

$$\overline{\gamma_{Hh}} \approx \overline{\gamma_{0h}} + \frac{1}{2}H^+\frac{d\gamma}{dr} \approx \overline{\gamma_{0h}}\left(1 + \frac{H^+}{R}\right). \tag{7.3}$$

Next, we also exploit that both $\frac{N^+}{R} \ll 1$ and $\frac{H^+}{R} \ll 1$, so

$$\left(1 - \frac{N^+}{R}\right)^{-1} \approx \left(1 + \frac{N^+}{R}\right), \quad \left(1 - \frac{H^+}{R}\right)^{-1} \approx \left(1 + \frac{H^+}{R}\right),$$

Figure 7.5. Mikhail Sergeevich Molodensky, source obscure. More photographs and background information Brovar et al. (2000).

7.3.2 Molodensky's proof

The realization of the Molodensky school was, that, because normal gravity is along the plumb line very close to a linear function of place, one could define a height type that can be computed directly from geopotential numbers, and that also would be compatible with similarly defined, so-called height anomalies, and with geometric heights h reckoned from the reference ellipsoid.

The geometric height h from the reference ellipsoid may be connected to the potential U of the normal gravity field indirectly, though the following integral equation:

$$U = U_0 - \int_0^h \gamma(z)\ dz.$$

Here, U is the normal potential and γ normal gravity. One level surface of U, $U = U_0$, is also the reference ellipsoid. The variable z is the distance from the ellipsoid along the local normal.

By defining

$$\overline{\gamma_{0h}} \overset{\text{def}}{=} \frac{1}{h} \int_0^h \gamma(z)\ dz$$

and with equations 7.2, 7.3 and the definitions above of H^+, N^+,

$$H^* \stackrel{\text{def}}{=} -\frac{W-W_0}{\overline{\gamma_{0H}}} = -\frac{W-W_0}{\overline{\gamma_{0h}}} \cdot \frac{\overline{\gamma_{0h}}}{\overline{\gamma_{0H}}} \approx H^+ \left(1 + \frac{N^+}{R}\right) = H^+ + \frac{N^+ H^+}{R},$$

$$\zeta \stackrel{\text{def}}{=} \frac{T}{\overline{\gamma_{Hh}}} = \frac{T}{\overline{\gamma_{0h}}} \cdot \frac{\overline{\gamma_{0h}}}{\overline{\gamma_{Hh}}} \approx N^+ \left(1 - \frac{H^+}{R}\right) = N^+ - \frac{N^+ H^+}{R}.$$

Because the, already small, correction terms $\frac{N^+ H^+}{R}$ cancel, we finally obtain

$$H^* + \zeta = H^+ + N^+ = h. \tag{7.4}$$

The quantity $\overline{\gamma_{0H}}$, and thus also normal height H^*, can be, unlike $\overline{\gamma_{0h}}$, computed using *only information obtained by* (spirit or trigonometric) *levelling*, without having to know the height h above the reference ellipsoid, which would require again knowledge of the local geoid.

This was Molodensky's realization (Molodensky et al., 1962) already in 1945, long before the global positioning system GPS, or a global, geocentric reference ellipsoid, existed. Back then, continental triangulation networks, like the one of the Soviet Union, were computed on their own, regionally defined reference ellipsoids.

The size of the correction term $\frac{N^+ H^+}{R}$ is, for heights of the global geoid up to 110 m, 17 mm for each kilometre of terrain height. The errors remaining after applying this term are microscopically small, because normal gravity is, unlike true gravity, *extremely linear along the plumb line* — as equations 7.2 and 7.3 already assumed.

Figure 7.6 attempts to visualize the derivation.

7.3.3 Normal height and height anomaly

Normal height:

$$H^* = \frac{C}{\overline{\gamma}} = -\frac{W-W_0}{\overline{\gamma}}, \tag{7.5}$$

in which (recursive definition!)

$$\overline{\gamma} = \overline{\gamma_{0H}} = \frac{1}{H^*} \int_0^{H^*} \gamma(z)\, dz.$$
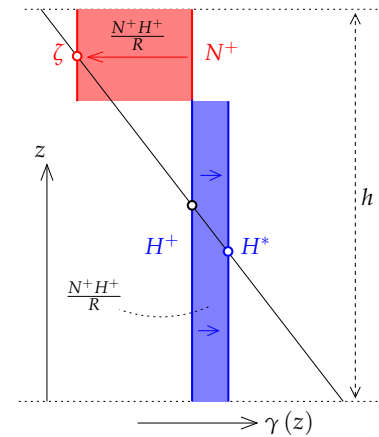
Figure 7.6. A graphic cartoon of Molodensky's proof. The blue and red areas, which are equal, represent the correction terms which convert N^+ to ζ and H^+ to H^* respectively. The red and blue arrows describe the conversion process. The balls represent midpoints of averaging intervals for the function $\gamma(z)$.

Height anomaly:

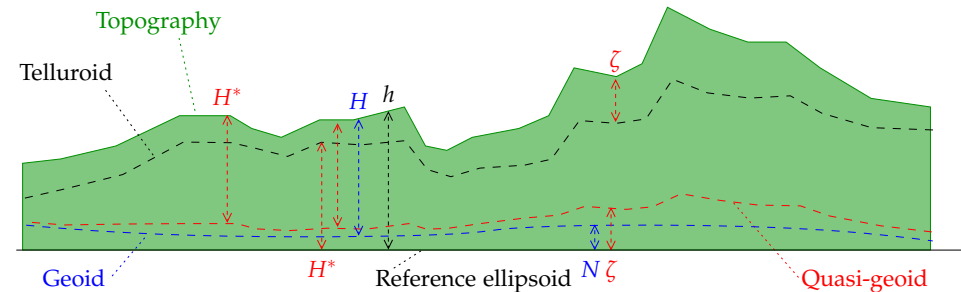$$\zeta = \frac{W-U}{\overline{\gamma_{Hh}}}$$

Figure 7.7. Geoid, quasi-geoid, telluroid and topography. Note the correlation between quasi-geoid and topography.

in which

$$\overline{\gamma_{Hh}} = \frac{1}{\zeta} \int_{H^*}^{h} \gamma(z) \, dz.$$

The height anomaly ζ, which otherwise is a quantity similar to the geoid height N, however is located at the level of the *topography*, not at sea level. The surface formed by points which are a distance H^* above the reference ellipsoid (and thus a distance ζ below the topography), is called the *telluroid*. It is a mapping of sorts of the topographic surface: the set of points Q whose *normal* potential U_Q is the same as the *true* potential W_P of the *true* topography's corresponding point P. See figure 5.4.

Often, as a concession to old habits, we construct a surface that is at a distance ζ above the reference ellipsoid. This surface is called the *quasi-geoid*. It lacks physical meaning; it is not a level surface, although out at sea it coincides with the geoid. Its short-wave features, unlike those of the geoid, *correlate* with the short-wavelength features of the topography.

Height above the ellipsoid (assumed $U_0 = W_0$):

$$h = \frac{U - U_0}{\overline{\gamma_{0h}}},$$

where

$$\overline{\gamma_{0h}} = \frac{1}{h} \int_{0}^{h} \gamma(z) \, dz.$$

_____

The *relationship* between the three quantities is

$$h = H^* + \zeta.$$

In all three cases, the quantity is defined by dividing the potential difference by some sort of "average normal gravity", suitably computed along a segment of the local plumb line. In the case of the height anomaly ζ, a piece of plumb line is used high up, *close to the topographic surface*, between level H^* (telluroid) and level h (topography).

7.4   Difference between geoid height and height anomaly

Normal heights are very *operational*. They are always used together with "quasi-geoid" heights — more correctly: height anomalies — ζ. Orthometric heights — e.g., Helmert heights — on the other hand are always used together with geoid heights N. For computing both, H and N, one needs the topographic mass density ρ, for which often a standard constant value is assumed ($2670 \, \text{kg/m}^3$), and the local vertical gradient of gravity, for which generally the vertical normal gravity gradient ($-0.3086 \, \text{mGal/m}$) is assumed.

The difference between height anomaly and geoid height is calculated as follows.

1. First, calculate the separation between the quasi-geoid and the "free-air geoid". The free-air geoid is an equipotential surface of the harmonically downward continued, exterior potential. If T is the disturbing potential of the exterior, harmonically downward continued field, then its difference between topography level and sea level is:

$$T_H - T_0 = \int_0^H \frac{\partial T}{\partial h} \, dh \approx -\Delta g_{\text{FA}} H, \tag{7.6}$$

and by using the Bruns equation twice, $\zeta = \dfrac{T_H}{\overline{\gamma}}$ (height anomaly or quasi-geoid height) and $N_{\text{FA}} = \dfrac{T_0}{\overline{\gamma}}$ ("free-air geoid height", FA = Free Air) we obtain[3]

$$\zeta - N_{\text{FA}} \approx -\frac{\Delta g_{\text{FA}} H}{\overline{\gamma}}. \tag{7.7}$$

2. Thus we have obtained the difference between height anomalies and heights of the "free-air geoid"; what is left is determining the separation between "free-air geoid" and geoid.

Let us approximate the topography by a Bouguer plate. Then

_____

[3]Here we made the approximation, that $\overline{\gamma}$ is the same on topography level as on sea level.

○ in the case of the "free-air geoid" N_{FA} the thickness of the plate is the height H of point P. This is because the free-air geoid is based on the harmonically downward continued exterior field, meaning that also the Bouguer-plate attraction acting at P must be continued downward, i.e., taken fully into account.

Because the surface mass density of the plate is $H\rho$, its *assumed* attraction is *everywhere* on point P:s plumb line:

$$2\pi G H \rho.$$

○ Now in the case of the *geoid*, we have to be physically realistic: in an arbitrary point P' on the plumb line of point P, the Bouguer plate is partly below the point, and partly above the point. The attraction is then only

$$2\pi G H' \rho - 2\pi G \left(H - H' \right) \rho = 2\pi G \left(2H' - H \right) \rho,$$

in which H' is now the height of point P'.

By integrating the difference, like we did for equation 7.6, we obtain

$$T - T_{\mathrm{FA}} = 2\pi G\rho \int_0^H \left(\left(2H' - H \right) - H \right) dH' =$$
$$= 2\pi G\rho \left[\left(H' \right)^2 - 2HH' \right]_{H'=0}^H = -2\pi G\rho H^2 \approx -A_{\mathrm{B}} H,$$

in which A_{B} is the attraction of a Bouguer plate of thickness H. We obtain again by dividing by the average normal gravity:

$$N - N_{\mathrm{FA}} = -\frac{A_{\mathrm{B}} H}{\overline{\gamma}}.$$

By subtracting this latest result from equation 7.7 we find:

$$\zeta - N = \frac{\left(-\Delta g_{\mathrm{FA}} + A_{\mathrm{B}} \right) H}{\overline{\gamma}} = -\frac{\Delta g_{\mathrm{B}} H}{\overline{\gamma}}. \tag{7.8}$$

See also Heiskanen and Moritz (1967, pages 8–13). As in the mountains the Bouguer anomaly is strongly negative, it follows that the quasi-geoid is there always *above* the geoid: approximately, using equation 6.2:

$$\zeta - N \approx \frac{0.1119 \, ^{\mathrm{mGal}/\mathrm{m}}}{9.8 \, ^{\mathrm{m/s^2}}} H^2 \approx 10^{-7} \, \mathrm{m}^{-1} H^2.$$

Or, if H is in units of [km] and $\zeta - N$ in units of [m]:

$$\zeta - N \, [\mathrm{m}] \approx 0.1 H^2 \, [\mathrm{km}].$$

7.5   Difference between orthometric and normal heights

The geoid is the level from which orthometric heights are measured. Therefore we may write

$$h = H + N,$$

where h is height above the reference ellipsoid and H orthometric height.

We may also bring back to memory equation 7.4:

$$h = H^* + \zeta,$$

in which ζ is the height anomaly, and H^* the normal height.

We obtain simply:

$$H - H^* = \zeta - N = -\frac{\Delta g_{\mathrm{B}} H}{\overline{\gamma}}, \tag{7.9}$$

using equation 7.8.

7.6   Calculating orthometric heights precisely

Orthometric heights are a traditional way of expressing height "above sea level". Orthometric heights are heights above a real geoid, i.e., a level surface inside the Earth and in the mean located at the same level as mean sea level.

We may write

$$W = W_0 - \int_0^H g(z) \, dz$$

where g is the true gravity inside the topographic masses. From this we obtain

$$H = \frac{C}{\overline{g}} = \frac{-\left(W - W_0 \right)}{\overline{g}},$$

in which the mean gravity along the plumb line is

$$\overline{g} = \frac{1}{H} \int_0^H g(z) \, dz.$$

The method is recursive: H appears both on the left and on the right side. This is not a problem: both H and $\overline{g}$ are obtained iteratively. Convergence is fast.

In practice one calculates orthometric height using an approximate formula. In Finland, *Helmert orthometric heights* have long been used, for which gravity measured on the Earth's surface, $g(H)$, is extrapolated downward by using the estimated vertical gravity gradient interior to the rock. It is assumed that its standard value outside the rock, the value, $-0.3086\,\mathrm{mGal/m}$ (the free-air gradient), changes to a value that is $+0.2238\,\mathrm{mGal/m}$ greater (double Bouguer plate effect): the end result is the total inside-rock gravity gradient, $-0.0848\,\mathrm{mGal/m}$.

This is called the *Prey*[4] *reduction*. The end result is the following equations (the coefficient is *half* the gravity gradient, i.e., the mean gravity along the plumb line is the same as gravity at the midpoint of the plumb line):

$$\overline{g} = g(H) - 0.0848\,\mathrm{mGal/m}\left(-\frac{1}{2}H\right) = g(H) + 0.0424\,\mathrm{mGal/m} \cdot H, \text{ thus}$$

$$H = \frac{C}{\overline{g}} = \frac{C}{g(H) + 0.0424\,\mathrm{mGal/m} \cdot H}, \tag{7.10}$$

in which C is the geopotential number (potential difference with mean sea level) and $g(H)$ is gravity at the Earth's surface. See also Heiskanen and Moritz (1967) pages 163–167. *Note* that the term $0.0424\,\mathrm{mGal/m} \cdot H$ is typically *much smaller* than $g(H)$, which is about $9.8\,\mathrm{m/s^2} = 980\,000\,\mathrm{mGal}$! So, iteration, in which the above denominator is first calculated using a crude H value, converges really fast.

The use of Helmert heights as an approximation to orthometric heights is imprecise for the following reasons:

- The assumption that gravity changes linearly along the plumb line. This is not the case, especially not because of the terrain correction. In the precise computation of orthometric heights, one ought to compute the terrain correction separately for every point on the plumb line.

- The assumption that the free-air vertical gravity gradient is a constant, $-0.3086\,\mathrm{mGal/m}$. This is not the case, the gradient can easily vary by $\pm 10\%$.

---

[4]Adalbert Prey (1873 – 1949) was an Austrian astronomer and geodesist and an author of textbooks.

- The assumption that rock density is $\rho = 2.67\,\mathrm{g/cm^3}$. The true density value may easily vary by $\pm 10\%$ or more around this assumed value.

The first approximation, neglecting the terrain effect, can be corrected by using Niethammer[5]'s method (see Heiskanen and Moritz (1967) page 167). It requires that, also in geoid computation, the terrain is correspondingly taken into account.

The third approximation, the density, can be removed as a problem by conventionally agreeing to use also in the corresponding geoid computation a *standard density* $\rho = 2.67\,\mathrm{g/cm^3}$. The surface thus obtained isn't any more a true geoid then, but a "fake geoid", for which no suitable name comes to mind.

The second approximation could be eliminated by using the *true* free-air gravity gradient instead of a standard value. However, the true gravity gradient depends on local density variations. One can use, e.g., the Poisson equation for computing the gradient, on which more later.

The precise calculation of orthometric heights is thus laborious. Just as laborious as the precise determination of the geoid, and for the same reasons. Fortunately in non-mountainous countries Helmert heights are good enough. In Finland they were even computed using as ρ values "true" crustal densities according to a geological map. . .

7.7   Calculating normal heights precisely

For this we use the equation 7.5:

$$H^* = \frac{C}{\overline{\gamma}} = -\frac{W - W_0}{\overline{\gamma}},$$

where the average value of normal gravity along the plumb line is

$$\overline{\gamma} = \overline{\gamma_{0H}} = \frac{1}{H^*}\int_0^{H^*} \gamma(z)\, dz.$$

---

[5]Theodor Niethammer (1876 – 1947) was a Swiss astronomer and geodesist who was the first to map the gravity field of the Swiss Alps.

Because normal gravity is in good approximation a linear function of z, we may write

$$\overline{\gamma} = \gamma_0 + \frac{1}{2}H^*\frac{\partial\gamma}{\partial z},$$

in which $\frac{\partial\gamma}{\partial z} = -0.3086\,\text{mGal/m}$ and $\gamma_0 \overset{\text{def}}{=} \gamma(0,\varphi)$ is normal gravity computed at height zero. We obtain

$$\overline{\gamma} = \gamma_0 - 0.1543\,\text{mGal/m} \cdot H^*.$$

The solution is again obtained iteratively:

$$H^* = \frac{C}{\overline{\gamma}} = \frac{C}{\gamma_0 - 0.1543\,\text{mGal/m} \cdot H^*} \tag{7.11}$$

in which $\gamma_0 = \gamma(0,\varphi)$ can be calculated exactly when local latitude φ is known. H^* appears on both sides of the equation, but it converges fast because again, the first term of the denominator γ_0, some $9.8\,\text{m/s}^2 = 980\,000\,\text{mGal}$, is a lot larger than $0.1543\,\text{mGal/m} \cdot H^*$.

Calculating normal heights is not in the same way sensitive to Earth crustal density and similar hypotheses, like calculation of orthometric heights is. It depends however on the choice of normal field, i.e., the reference ellipsoid.

7.8 Calculation example for heights

At point P the potential difference with sea level is $C = 5000\,\text{m}^2/\text{s}^2$. Local gravity is $g = 9.820\,000\,\text{m/s}^2$.

Normal gravity calculated at level zero under point P equals $\gamma_0 = 9.821\,500\,\text{m/s}^2$.

Questions:

1. Calculate the orthometric height of the point.
2. Calculate the free-air gravity anomaly Δg_{FA} of point P.
3. Calculate the Bouguer anomaly (without terrain correction) Δg_{B} of the point.
4. Calculate the normal height of point P.

5. If the geoid height at point P is $N = 25.000\,\text{m}$, how much is then the *height anomaly* ("quasi-geoid height") ζ?

Answers:

1. First attempt: $H^{(0)} = \frac{C}{g} = \frac{5000}{9.82}\,\text{m} = 519.165\,\text{m}$. Second attempt (equation 7.10):

$$H^{(1)} = \frac{5000\,\text{m}^2/\text{s}^2}{9.820\,000\,\text{m/s}^2 + 0.0424 \cdot 10^{-5}\text{s}^{-2} \cdot 519.165\,\text{m}} = 509.154\,\text{m}.$$

After that, the millimetres don't change any more.

2. The free-air anomaly is

$$\Delta g_{\text{FA}} = \left(9.820\,000 - (9.821\,500 - 0.3086 \cdot 10^{-5} \cdot 509.154)\right)\,\text{m/s}^2 = $$
$$= 7.125\,\text{mGal}.$$

3. The Bouguer anomaly is (equation 6.2):

$$\Delta g_{\text{B}} = \Delta g_{\text{FA}} - 0.1119\text{mGal/m} \cdot H = -49.849\,\text{mGal}.$$

4. The first attempt is again $H^{*(0)} = \frac{C}{\gamma_0} = 509.087\,\text{m}$. The second, equation 7.11:

$$H^{*(1)} = \frac{5000\,\text{m}^2/\text{s}^2}{9.821\,500\,\text{m/s}^2 - 0.1543 \cdot 10^{-5}\text{s}^{-2} \cdot 509.087\,\text{m}} = 509.128\,\text{m},$$

also final on the millimetre level.

5. The difference equation 7.9 yields

$$\zeta - N = -\frac{\Delta g_{\text{B}}H}{\overline{\gamma}} = -0.026\,\text{m}.$$

Also (check) $H^* - H = -0.026\,\text{m}$. So

$$\zeta = N - (-0.026\,\text{m}) = 25.026\,\text{m}.$$

7.9   Orthometric and normal corrections

In practical orthometric height calculations, one often adds together at first the height differences ΔH measured by levelling ("staff reading differences") between points A and B as a *tentative* or crude height difference

$$\sum_A^B \Delta H,$$

after which the non-exactness of this method is accounted for by applying the "orthometric correction" (*OC*):

$$H_B = H_A + \sum_A^B \Delta H + OC_{AB}.$$

The fact that the difference in orthometric heights between two points A and B is not equal to the sum of the levelled height differences is due to gravity not being the same everywhere.

With $C_A, C_B, \Delta C$ the geopotential numbers at A and B, and the geopotential differences along the levelling line, we have $C_B - C_A - \sum_A^B \Delta C = 0$ because of the conservative nature of the geopotential. Dividing by a constant γ_0 yields

$$\frac{C_B}{\gamma_0} - \frac{C_A}{\gamma_0} - \sum_A^B \frac{\Delta C}{\gamma_0} = 0.$$

On the other hand we have

$$OC_{AB} = H_B - H_A - \sum_A^B \Delta H = \frac{C_B}{\overline{g}_B} - \frac{C_A}{\overline{g}_A} - \sum_A^B \frac{\Delta C}{g},$$

with $\overline{g}_A, \overline{g}_B$ average gravities along the plumb lines of A and B, g gravity along the levelling line. In this expression, we compare $\sum_A^B \Delta H$, the naively calculated sum of levelled height differences, with the difference between the orthometric heights of the end points A and B, calculated according to the definition.

Subtraction yields

$$OC_{AB} - 0 = \left(\frac{C_B}{\overline{g}_B} - \frac{C_B}{\gamma_0}\right) - \left(\frac{C_A}{\overline{g}_A} - \frac{C_A}{\gamma_0}\right) - \sum_A^B \left(\frac{\Delta C}{g} - \frac{\Delta C}{\gamma_0}\right),$$

in which

$$\frac{C_B}{\overline{g}_B} - \frac{C_B}{\gamma_0} = \left(\frac{\gamma_0 - \overline{g}_B}{\gamma_0}\right)\frac{C_B}{\overline{g}_B} = \left(\frac{\gamma_0 - \overline{g}_B}{\gamma_0}\right) H_B,$$

$$\frac{C_A}{\overline{g}_A} - \frac{C_A}{\gamma_0} = \left(\frac{\gamma_0 - \overline{g}_A}{\gamma_0}\right) H_A,$$

$$\frac{\Delta C}{g} - \frac{\Delta C}{\gamma_0} = \left(\frac{\gamma_0 - g}{\gamma_0}\right) \Delta H,$$

yielding the *orthometric correction*

$$OC_{AB} = \sum_A^B \left(\frac{g - \gamma_0}{\gamma_0}\right)\Delta H + \left(\frac{\overline{g}_A - \gamma_0}{\gamma_0}\right) H_A - \left(\frac{\overline{g}_B - \gamma_0}{\gamma_0}\right) H_B, \qquad (7.12)$$

which is identical to Heiskanen and Moritz (1967) equation 4-33.

The choice of the constant γ_0 is arbitrary; it is wise to choose it close to the average gravity in the general area of (A, B) to keep the numerics small.

Similarly we may also compute the *normal correction* (*NC*) in calculating normal heights. Start from the equation

$$NC_{AB} = H_B^* - H_A^* - \sum_A^B \Delta H = \frac{C_B}{\overline{\gamma}_B} - \frac{C_A}{\overline{\gamma}_A} - \sum_A^B \frac{\Delta C}{g},$$

from which, like above, follows by subtraction

$$NC_{AB} = \sum_A^B \left(\frac{g - \gamma_0}{\gamma_0}\right)\Delta H + \left(\frac{\overline{\gamma}_A - \gamma_0}{\gamma_0}\right) H_A^* - \left(\frac{\overline{\gamma}_B - \gamma_0}{\gamma_0}\right) H_B^*. \qquad (7.13)$$

Note that the identical first term in both equation 7.12 and equation 7.13 derives from the term

$$\sum_A^B \frac{\Delta C}{g} = \sum_A^B \Delta H,$$

the naive summation of height differences ΔH in the case of both orthometric and normal correction, which is the generic basis of the concept of both corrections.

Applying result 7.13:

$$H_B^* = H_A^* + \sum_A^B \Delta H + NC_{AB}.$$

What changes between the orthometric and normal corrections is the definition of heights: H^* instead of H, requiring division by the average of normal gravity along the plumb line $\overline{\gamma}$, not by that of true gravity $\overline{g}$.

Note that both the orthometric correction 7.12 and the normal correction 7.13 can be calculated one staff interval at a time: one must know, in addition to the levelled height difference ΔH, local gravity g along the levelling line, and also at the end points $g(H)$ or $\gamma(H^*)$ for calculating mean gravity $\overline{g}$ or $\overline{\gamma}$ along the plumb lines of those end points. This goes well with the equations given above. Remember that gravity g along the levelling line is needed in order to reduce the individual levelled height differences ΔH to geopotential number differences ΔC. This reduction is part of the computation of both the orthometric and the normal correction.

7.10   A vision for the future: relativistic levelling

yleinen
suhteellisuusteoria
metriikka

According to general relativity, the deeper a clock is inside the potential well of masses, the slower it ticks. This is most easily seen by looking at the Schwarzschild[6] metric for a spherically symmetric field:

$$
\begin{aligned}
c^2\,d\tau^2 &= \left(1 - \frac{2GM}{c^2 r}\right) c^2\,dt^2 - \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 - r^2\left(d\phi^2 + \cos^2\phi\,d\lambda^2\right) = \\
&= \left(1 - \frac{2W}{c^2}\right) c^2\,dt^2 - \left(1 - \frac{2W}{c^2}\right)^{-1} dr^2 - r^2\left(d\phi^2 + \cos^2\phi\,d\lambda^2\right),
\end{aligned}
$$

in spherical co-ordinates plus time (r, ϕ, λ, t). Here we see, how the rate of the proper time τ is slowed down compared to stationary co-ordinate time t (time at infinity $r \to \infty$), when the geopotential W increases closer to the mass. The slowing-down ratio is

$$
\frac{\partial \tau}{\partial t} = \sqrt{1 - \frac{2W}{c^2}} \approx 1 - \frac{W}{c^2}.
$$

---

[6]Karl Schwarzschild (1873 – 1916) was a German physicist who was the first to derive, in 1915 while serving on the Russian front, a closed spherically symmetric, i.e., non-rotating, solution to the field equation of Albert Einstein's general theory of relativity, the *Schwarzschild metric*.
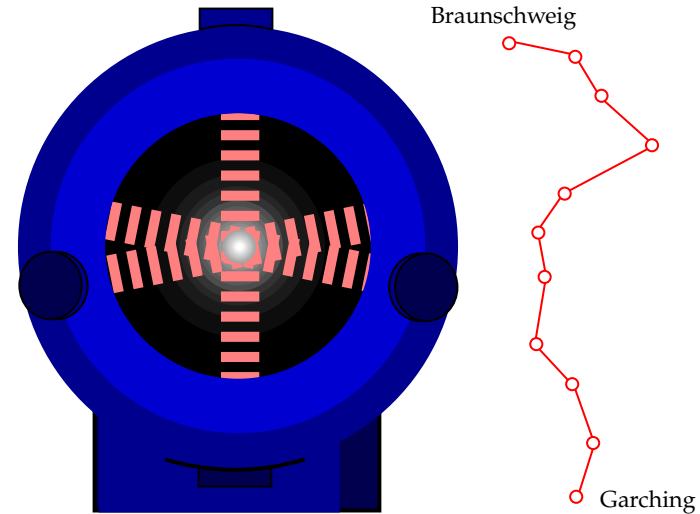
Figure 7.8. An optical lattice clock: the ultra-precise atomic clock of the future operates at optical wavelengths. To the right, the trajectory of the Predehl et al. experiment.

Now c^2 is, in the units of daily life, a huge number: $10^{17}\,\mathrm{m^2/s^2}$. This means that measuring a potential difference of $1\,\mathrm{m^2/s^2}$ — corresponding to a height difference of $10\,\mathrm{cm}$ — using this method, requires a precision of $1 : 10^{17}$. More traditional, microwave based atomic clocks can do precisions of $10^{-12} - 10^{-14}$ (Vermeer, 1983). With the new optical clocks the objective should be achievable and relativistic levelling may become a reality.

The clock works in this way, that an extremely cold, so-called Bose–Einstein condensate of atoms is trapped inside an optical lattice formed by six laser beams, an electromagnetic pattern of standing waves. The readout beam of the clock oscillation uses a different frequency. A Bose–Einstein condensate has the property that all atoms are in precisely the same quantum state — like the photons in an operating laser — i.e., their matter waves are *coherent*. In a way, all atoms together act as one virtual atom.

optinen hila

The condensate may consist of millions of atoms, and can actually be seen

through the window of the vacuum chamber as a small plasma blob.

Unfortunately it is not enough, that just one laboratory measures time to extreme precision — one also has to be able to *compare* the ticking rates of different clocks over geographical distances. Also for this, a solution has been found: existing optic-fibre cables already in global use for Internet and telephony are useable for this with small modifications. The modifications concern the amplifiers in the cables at distances of some 100 km, which must be replaced by modified ones (Predehl et al., 2012). In this way both the traditional precise levelling networks and the height systems based on GNSS technology and geoid determination may be replaced by this hi-tech (and hi-science!) solution.

Self-test questions

1. Why are heights calculated directly from levelled height differences not good enough as a height *system*?

2. What is a geopotential number?

3. What are orthometric heights?

4. What are normal heights?

5. What is the classical definition of the geoid?

6. What is a height anomaly?

7. What is the quasi-geoid?

8. Why may water sometimes flow in the "wrong" direction, i.e., to a greater height?

9. What is the telluroid?

10. What are the orthometric correction and the normal correction?

Exercise 7 – 1: Calculating orthometric heights

The potential difference with sea level at point P, $-(W - W_0)$, equals $1000 \, \mathrm{m^2/s^2}$. Gravity in the point is $g_P = 9.820\,000 \, \mathrm{m/s^2}$. Calculate the orthometric height of the point. Aim for millimetre precision.

Exercise 7 – 2: Calculating normal heights

In point P the potential difference with sea level is

$$-(W - W_0) = 5000 \, \mathrm{m^2/s^2}.$$

Below the point at sea level, normal gravity is $\gamma_0 = 9.821\,500 \, \mathrm{m/s^2}$. Calculate the normal height of the point.

Exercise 7 – 3: Difference between orthometric and normal height

In point P the Bouguer anomaly is $\Delta g_B = -120 \, \mathrm{mGal}$. The orthometric height of the point is 1150 m.

1. Calculate the normal height of point P.

2. If the geoid height in point P is $N = 21.75 \, \mathrm{m}$, calculate the height anomaly ζ of the point.

8 The Stokes equation and other integral equations

8.1 The Stokes equation and the Stokes integral kernel

By suitably combining the equations in section 5.3 one obtains on the surface of a spherical Earth

$$T = \sum_{n=2}^{\infty} T_n = R \sum_{n=2}^{\infty} \frac{\Delta g_n}{n-1},$$

with $T_n = T_n(\phi, \lambda)$ the degree constituents of the disturbing potential field $T = T(\phi, \lambda)$, and $\Delta g_n = \Delta g_n(\phi, \lambda)$ those of the gravity anomaly field $\Delta g = \Delta g(\phi, \lambda) = \sum_{n=2}^{\infty} \Delta g_n(\phi, \lambda)$ [1].

This is now the Stokes equation's *spectral form*.

Substituting into this the degree constituent equation 3.7 one obtains the

---

[1] For the degree numbers $n = 0, 1$, the Δg_n are assumed to vanish, as $\Delta g_0 \neq 0$ would mean a different total mass for the normal field than for the Earth, and $\Delta g_1 \neq 0$ an offset of the co-ordinate origin from the Earth's centre of mass, see section 3.4.
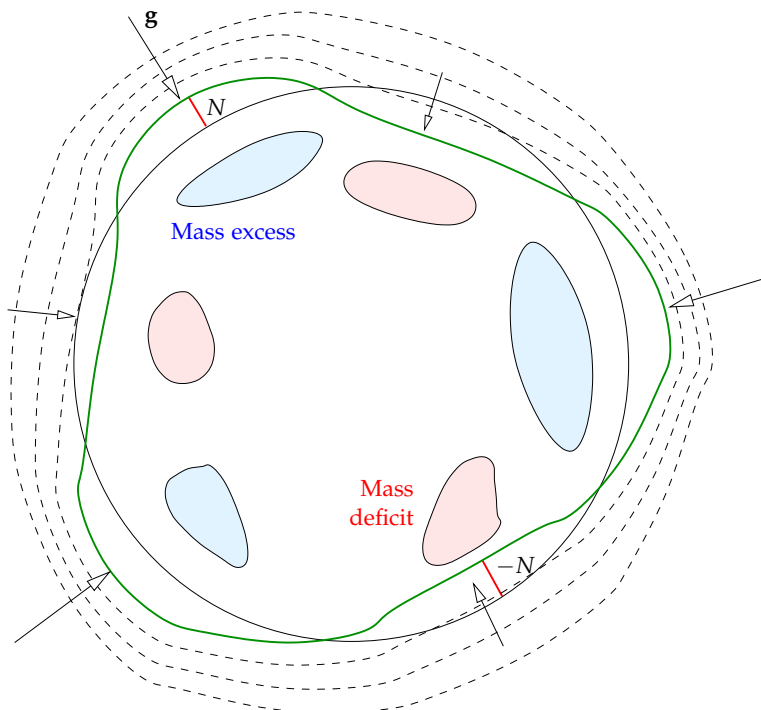
Figure 8.1. The principle of gravimetric geoid determination.

Figure 8.2. Integrating the Stokes equation geometrically.

integral equation

$$
T = \frac{R}{4\pi} \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \iint_\sigma \Delta g P_n(\cos\psi)\, d\sigma =
$$

$$
= \frac{R}{4\pi} \iint_\sigma \left(\sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi) \right) \Delta g\, d\sigma =
$$

$$
= \frac{R}{4\pi} \iint_\sigma S(\psi)\, \Delta g\, d\sigma,
$$

in which

$$
S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi),
$$

the Stokes kernel function. The angle ψ is the geocentric angular distance

between evaluation point and moving observation point, see figure 8.2. The equation above allows the calculation, from global gravimetric data and for every point on the surface of the Earth sphere, of the disturbing potential T, and from that the *geoid height* N using the Bruns equation 5.2, $N = \frac{T}{\gamma}$, with the result

$$
N(\phi,\lambda) = \frac{T(\phi,\lambda)}{\gamma} = \frac{R}{4\pi\gamma} \iint_\sigma S(\psi)\, \Delta g(\phi',\lambda')\, d\sigma', \qquad (8.1)
$$

in which (ϕ,λ) and (ϕ',λ') are the evaluation point and the moving point ("observation point"), respectively, and the distance between them is ψ. Equation 8.1 is the classical Stokes equation of gravimetric geoid determination.

The above illustrates the correspondence between integral equations and spectral equations. There are other examples of this. Earlier we presented the spectral representation of the function $\frac{1}{\ell}$, Heiskanen and Moritz (1967) equation 1-81. Of course $\frac{1}{\ell}$ is also the kernel function of an integral equation, the one yielding the potential V if given is the single layer mass density κ.

Also a version of the Stokes equation for the *exterior space* exists. We gave it earlier, equation 5.10. The spectral form of its kernel function, see equation
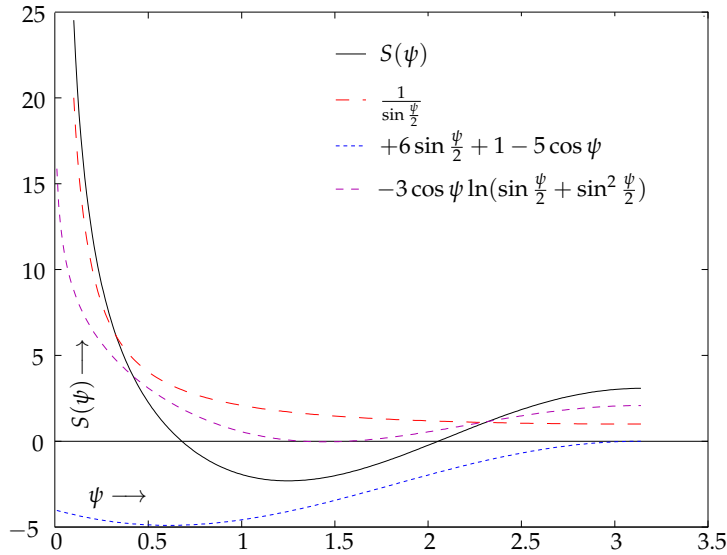
Figure 8.3. The Stokes kernel function $S(\psi)$. The argument ψ is in radians $[0, \pi)$. Also plotted are the three parts of the analytical expression 8.2 with their different asymptotic behaviours.

5.11, is

$$S(r, \psi, R) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{2n+1}{n-1} P_n(\cos \psi).$$

The Stokes kernel function on the Earth's surface is depicted in figure 8.3, in which the angle ψ is in radians (1 rad $\approx 57°.3$).

This curve was calculated using the following closed expression (Heiskanen and Moritz, 1967, section 2-16, equation 2-164):

$$S(\psi) = \frac{1}{\sin \frac{\psi}{2}} - 6 \sin \frac{\psi}{2} + 1 - 5 \cos \psi - 3 \cos \psi \ln \left(\sin \frac{\psi}{2} + \sin^2 \frac{\psi}{2}\right). \quad (8.2)$$

This closed expression helps us to understand better how the function behaves close to the origin $\psi = 0$: the first term, $\left(\sin \frac{\psi}{2}\right)^{-1}$, goes to infinity when $\psi \to 0$. The next three terms, $-6 \sin \frac{\psi}{2} + 1 - 5 \cos \psi$, are all bounded on the

whole interval $[0, \pi)$ and the limit for $\psi \to 0$ is -4. The last, complicated term $-3 \cos \psi \ln(\sin \frac{\psi}{2} + \sin^2 \frac{\psi}{2})$ goes also to infinity — *positive* infinity! — for $\psi \to 0$, but much more slowly, because of the logarithm.

8.2   Example: the Stokes equation in polar co-ordinates

In section 2.3 we derived a general solution to the Laplace equation in two dimensions in polar co-ordinates. Below we develop a "toy" computational framework for gravimetric geoid determination in two dimensions, which allows us to do simple numerical simulations, to get a feel for the behaviour of these things.

Firstly we derive the disturbing potential, gravity anomaly, and Stokes integral kernel for this solution, equation 2.3, assuming a normal potential $U(r) = a_0 + b_0 \ln r$.

○ Disturbing potential:

$$T(\alpha, r) = V_2(\alpha, r) - (a_0 + b_0 \ln r) = \sum_{k=1}^{\infty} r^{-k} \left(a_k \cos k\alpha + b_k \sin k\alpha\right).$$

○ Normal gravity:

$$\gamma(r) = \frac{\partial U}{\partial r} = \frac{b_0}{r}.$$

○ Normal gravity gradient:

$$\frac{\partial \gamma}{\partial r} = \frac{\partial^2 U}{\partial r^2} = -\frac{b_0}{r^2}.$$

The Stokes equation and other integral equations

○ Gravity anomaly, equation 5.5:

$$\Delta g(\alpha, r) = -\frac{\partial T}{\partial r} + \frac{T}{\gamma}\frac{\partial \gamma}{\partial r} =$$

$$= \sum_{k=1}^{\infty} \frac{k}{r} r^{-k} (a_k \cos k\alpha + b_k \sin k\alpha) +$$

$$+ \frac{1}{\gamma}\frac{\partial \gamma}{\partial r} \sum_{k=1}^{\infty} r^{-k} (a_k \cos k\alpha + b_k \sin k\alpha) =$$

$$= \sum_{k=1}^{\infty} \left(\frac{k}{r} + \frac{1}{\gamma}\frac{\partial \gamma}{\partial r}\right) r^{-k} (a_k \cos k\alpha + b_k \sin k\alpha) =$$

$$= \sum_{k=1}^{\infty} \frac{k-1}{r} r^{-k} (a_k \cos k\alpha + b_k \sin k\alpha).$$

We see that, if we write

$$T(\alpha, r) = \sum_{k=1}^{\infty} \left(\frac{R}{r}\right)^k T_k \quad \text{with} \quad T_k \overset{\text{def}}{=} R^{-k} (a_k \cos k\alpha + b_k \sin k\alpha),$$

it follows that

$$\Delta g(\alpha, r) = \sum_{k=1}^{\infty} \left(\frac{R}{r}\right)^{k+1} \Delta g_k(\alpha)$$

$$\text{with} \quad \Delta g_k(\alpha) \overset{\text{def}}{=} (k-1) R^{-(k+1)} (a_k \cos k\alpha + b_k \sin k\alpha),$$

and, like in the case of spherical co-ordinates,

$$\Delta g_k = \frac{k-1}{R} T_k. \tag{8.3}$$

According to Fourier theory, the base functions $\cos k\alpha, \sin k\alpha$ are *orthonormal* on the circle $r = R$ when choosing the following integral as the scalar product:

$$\frac{1}{\pi}\int_0^{2\pi} \cos k\alpha \cos m\alpha \, d\alpha = \frac{1}{\pi}\int_0^{2\pi} \sin k\alpha \sin m\alpha \, d\alpha = \begin{cases} 0 & \text{if } k \neq m, \\ 1 & \text{if } k = m, \end{cases}$$

$$\frac{1}{\pi}\int_0^{2\pi} \cos k\alpha \sin m\alpha \, d\alpha = 0 \quad \text{always.}$$

This means that, with $\Delta g(\alpha) = \sum_{k=1}^{\infty} \Delta g_k(\alpha)$, we may decompose Δg into its

Fourier terms by projection, yielding the following Fourier coefficients[2]

$$\begin{Bmatrix} A_k \\ B_k \end{Bmatrix} = \frac{1}{\pi}\int_0^{2\pi} \Delta g(\alpha) \begin{Bmatrix} \cos k\alpha \\ \sin k\alpha \end{Bmatrix} d\alpha = (k-1) R^{-(k+1)} \begin{Bmatrix} a_k \\ b_k \end{Bmatrix}.$$

It follows that

$$\begin{Bmatrix} a_k \\ b_k \end{Bmatrix} = \frac{R^{k+1}}{k-1} \begin{Bmatrix} A_k \\ B_k \end{Bmatrix},$$

and, using the cosine difference equation (Wolfram, Difference formula for cosine),

$$T(\alpha) = T(\alpha, R) = \sum_{k=1}^{\infty} T_k(\alpha) = \sum_{k=1}^{\infty} R^{-k} \left(\frac{R^{k+1}}{k-1} A_k \cos k\alpha + \frac{R^{k+1}}{k-1} B_k \sin k\alpha\right) =$$

$$= \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{R}{k-1} \left(\cos k\alpha \int_0^{2\pi} \Delta g(\alpha') \cos k\alpha' \, d\alpha' + \sin k\alpha \int_0^{2\pi} \Delta g(\alpha') \sin k\alpha' \, d\alpha'\right) =$$

$$= \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{R}{k-1} \int_0^{2\pi} \Delta g(\alpha') \cdot \cos(k[\alpha - \alpha']) \, d\alpha'.$$

We collect everything dependent on k into a single sum, the Stokes kernel for this two-dimensional situation:

$$N(\alpha) = \frac{T(\alpha)}{\gamma} = \frac{R}{\pi\gamma}\int_0^{2\pi} \Delta g(\alpha') S(\alpha - \alpha') \, d\alpha',$$

$$\text{with} \quad S(\alpha - \alpha') \overset{\text{def}}{=} \sum_{k=2}^{\infty} \frac{\cos(k(\alpha - \alpha'))}{k-1}.$$

Note that, as in the spherical case, $\Delta g_1 = 0$, and $k = 1$ is also missing from the kernel.

---

[2]So, really

$$\Delta g(\alpha) = \Delta g(\alpha, r)\big|_{r=R} = \sum_{k=1}^{\infty} (A_k \cos k\alpha + B_k \sin k\alpha).$$

For small values of $\alpha - \alpha'$ we may approximate

$$S(\alpha - \alpha') = \sum_{k=1}^{\infty} \frac{\cos\left((k+1)(\alpha - \alpha')\right)}{k} \approx \sum_{k=1}^{\infty} \frac{\cos\left(k(\alpha - \alpha')\right)}{k} =$$

$$= \frac{1}{2}\ln\left(\frac{1}{2\left(1 - \cos(\alpha - \alpha')\right)}\right) \approx -\ln(\alpha - \alpha').$$

More abstractly we may write relationship 8.3 also in terms of the discrete Fourier transform and its inverse, as

$$\mathcal{F}_k\{\Delta g\} = \frac{k-1}{R}\mathcal{F}_k\{T\} \implies T(\alpha) = \mathcal{F}^{-1}\left\{\frac{R}{k-1}\mathcal{F}_k\{\Delta g\}\right\}(\alpha).$$

Here, $\mathcal{F}_k\{f\}$ represents the Fourier transform of a function $f(\alpha)$ of spatial co-ordinate α, as a function of the spatial wave number (number of waves around the circle) k. This formulation has the merit of being able to use any standard FFT software library offering compatible versions of both the forward Fourier transform $\mathcal{F}\{\cdot\}$ and the inverse transform $\mathcal{F}^{-1}\{\cdot\}$.

More about FFT in appendix C.

In figure 8.5 we show a simulation result where a randomly generated set of gravity anomalies on the circle $r = R$ has been used to estimate geoid undulations on the same circle. Both curves display fairly realistic statistical behaviour. The code used is given in tableau 8.1.

8.3　Plumb-line deflections and Vening Meinesz equations

By differentiating the Stokes equation with respect to place we obtain integral equations for the components of the deflection of the plumb line (Heiskanen and Moritz, 1967 equation 2-210′):

$$\begin{Bmatrix} \xi \\ \eta \end{Bmatrix} = \frac{1}{4\pi\gamma}\iint_\sigma \Delta g \frac{dS(\psi)}{d\psi}\begin{Bmatrix} \cos\alpha \\ \sin\alpha \end{Bmatrix} d\sigma =$$

$$= \frac{1}{4\pi\gamma}\iint_\sigma \Delta g \frac{dS(\psi)}{d\psi}\begin{Bmatrix} \cos\alpha \\ \sin\alpha \end{Bmatrix} \cdot \sin\psi\, d\alpha\, d\psi, \qquad (8.4)$$

Tableau 8.1. Stokes equation in two dimensions, octave code.

```octave
% Stokes equation simulator in two dimensions
R = 6378137;
g = 9.8;
ak(1:180) = 0;
bk(1:180) = 0;
dg(1:360) = 0.0;
T(1:360) = 0.0;
for i=1:359
% Gauss-Markov
  dg(i+1) = 0.8*dg(i) + 50*(rand()-0.5);
end
dgsum = 0.0;
for i=1:360
  % Enforce circularity
  dg(i) = dg(i) - (dg(360) - dg(1)) * (i/359);
  dgsum = dgsum + dg(i);
end
for i = 1:360
  % Enforce zero expectation
  dg(i) = dg(i) - dgsum/360;
  for k = 2:180
    ak(k) = ak(k) + dg(i) * cos(k*i*pi/180)/180;
    bk(k) = bk(k) + dg(i) * sin(k*i*pi/180)/180;
  end
end
for i=1:360
  for k = 2:180
    T(i) = T(i) +(ak(k)*cos(k*i*pi/180)+bk(k)*sin(k*i*pi/180))*R/(k-1);
  end
end
hold on
plot(1:360, dg, 'b') plot(1:360, 0.00001*T/g, 'm')
print -dpdf stokes2D-out.pdf
```
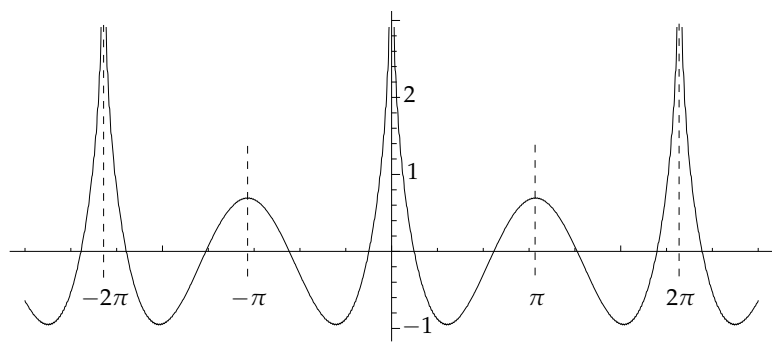
Figure 8.4. The Stokes kernel function on the circle $r = R$ in two-dimensional geometry. Note the symmetry and periodicity. Compare with the spherical Stokes kernel, figure 8.3.

where ξ, η are the North-South and East-West direction deflections of the plumb line, and the unit-sphere surface element $d\sigma = \sin \psi \, d\alpha \, d\psi$, in which
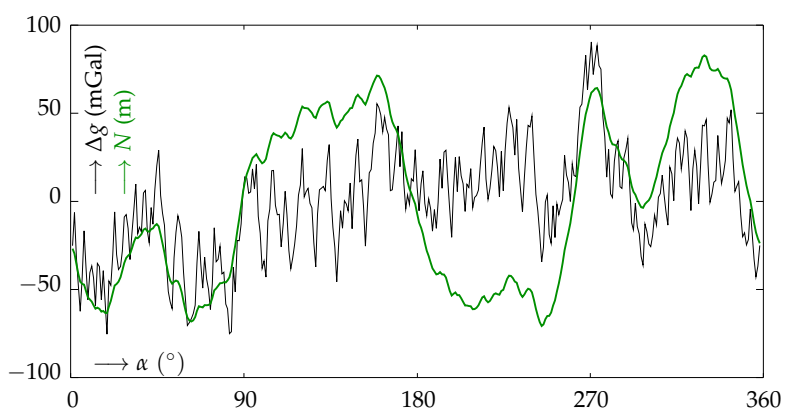
Figure 8.5. Simulation of gravity anomalies (Gauss–Markov process) and geoid undulations (green) in two-dimensional geometry on the circle. Note the spectral behaviour of both.

$\sin \psi$ is *Jacobi[3]'s determinant* of the (α, ψ) co-ordinates.

These equations were derived for the first time by the Dutch geophysicist F. A. Vening Meinesz. The angle α is the azimuth or direction angle between the calculation or evaluation point (ϕ, λ) and the moving integration or observation point (ϕ', λ'). These equations are much harder to write in spectral form, as the kernel functions are now also functions of the azimuth direction α, in other words, they are *anisotropic*.

The disturbing potential, the gravity disturbance, and the gravity anomaly, are all so-called *isotropic* quantities: they do not depend on the azimuth, and therefore, in the spectral representation, the transformations between them are only functions of harmonic degree n.

8.4   The Poisson integral equation

Look at figure 8.6. The point Q of the body is located at $\mathbf{R}$, and the observation point P at $\mathbf{r}$. The geocentric angular distance between the two location vectors, as seen from the origin, is ψ. The distance between points P and Q is ℓ.

With $R \stackrel{\text{def}}{=} \|\mathbf{R}\|$ and $r \stackrel{\text{def}}{=} \|\mathbf{r}\|$, we may write (cosine rule):

$$\ell = \sqrt{r^2 + R^2 - 2rR \cos \psi}.$$

We may also write the function $\frac{1}{\ell}$ as the following expansion (for proof, see Heiskanen and Moritz (1967) page 33):

$$\frac{1}{\ell} = \frac{1}{\sqrt{r^2 + R^2 - 2Rr \cos \psi}} = \frac{1}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} P_n(\cos \psi) \qquad (8.5)$$

in which $r = \|\mathbf{r}\|$ and $R = \|\mathbf{R}\|$ are the distances of points P and Q from the origin, the centre of the Earth. The function 8.5 is called the *generating function* of the Legendre polynomials.

Differentiating equation 8.5 with respect to r yields

$$-\frac{r - R \cos \psi}{\ell^3} = -\frac{1}{R} \sum_{n=0}^{\infty} \frac{n+1}{r} \left(\frac{R}{r}\right)^{n+1} P_n(\cos \psi).$$

---

[3]Carl Gustav Jacob Jacobi (1804–1851) was a German mathematician.
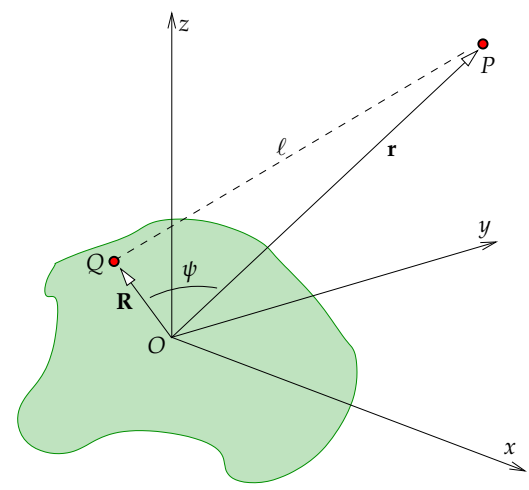
Figure 8.6. The geometry of the generating function of the Legendre polynomials.

This we multiply by $2r$:

$$-\frac{2r^2 - 2rR\cos\psi}{\ell^3} = -\frac{1}{R}\sum_{n=0}^{\infty}(2n+2)\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi).$$

Now we add together this equation and equation 8.5:

$$\frac{-2r^2 + 2rR\cos\psi + \ell^2}{\ell^3} = -\frac{1}{R}\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi).$$

The left-hand side is simplified by using $-r^2 + 2rR\cos\psi + \ell^2 = R^2$:

$$\frac{-2r^2 + 2rR\cos\psi + \ell^2}{\ell^3} = \frac{R^2 - r^2}{\ell^3},$$

and the end result is, by multiplying with $-R$,

$$\frac{R\left(r^2 - R^2\right)}{\ell^3} = \sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi). \tag{8.6}$$

Applying now the degree constituent equation 3.7 to the harmonic potential field V on the spherical Earth's surface, radius R:

$$V_n(\phi,\lambda) = \frac{2n+1}{4\pi}\iint_\sigma V(\phi',\lambda',R)P_n(\cos\psi)\,d\sigma',$$

The Poisson integral equation   **191**

as well as the spectral expansion of the field in space 3.8:

$$V(\phi,\lambda,r) = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}V_n(\phi,\lambda),$$

we obtain

$$\begin{aligned}
V(\phi,\lambda,r) &= \frac{1}{4\pi}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}(2n+1)\iint_\sigma V(\phi',\lambda',R)P_n(\cos\psi)\,d\sigma' = \\
&= \frac{1}{4\pi}\iint_\sigma V(\phi',\lambda',R)\left(\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{r}\right)^{n+1}P_n(\cos\psi)\right)d\sigma' = \\
&= \frac{R}{4\pi}\iint_\sigma \frac{\left(r^2 - R^2\right)V(\phi',\lambda',R)}{\ell^3}d\sigma'
\end{aligned}$$

by substituting the expression in square brackets directly into equation 8.6.

Thus we have obtained the Poisson equation for computing a harmonic field V from values given on the Earth's surface:

$$V_P = \frac{R}{4\pi}\iint_\sigma \frac{r^2 - R^2}{\ell^3}V_Q\,d\sigma_Q, \tag{8.7}$$

in which ℓ is again the straight distance between evaluation point P (where V_P is being computed) and moving data point Q (on the surface of the sphere, V_Q under the integral sign). In this equation we have given the points symbolic names: the co-ordinates of evaluation point P are (ϕ,λ,r), the co-ordinates of the data point Q are (ϕ',λ',R).

Still a third way to write the same equation, useful when the function or field V isn't actually defined between the topographic Earth's surface and sea level, is

$$V = \frac{R}{4\pi}\iint_\sigma \frac{r^2 - R^2}{\ell^3}V^*\,d\sigma,$$

in which V^* denotes the value of a *harmonically downward continued* function V — downward continued into the topography, all the way down to sea level, or, in spherical approximation, to the surface of the sphere $r = R$. This is a function that above the topography is identical to V, that is harmonic, and that also exists between topography and sea level. The existence of such a function has been a classical theoretical nut to crack…

Equation 8.7 solves the so-called *Dirichlet boundary-value problem*, finding a function in an area of space when the value of the function on its boundary has been given.

🔖 8.5 Gravity anomalies in the exterior space

The equation derived in the previous section 8.4, equation 8.7, applies for an arbitrary *harmonic* field V, i.e., a field for which $\Delta V = 0$. The equation may be conveniently applied to the expression $r\Delta g$, i.e., the gravity anomaly multiplied by the radius, which is also a harmonic field. This is how we can express the gravity anomaly in the external space $\Delta g(\phi, \lambda, r)$ as a function of gravity anomalies $\Delta g(\phi', \lambda', R)$ on a sphere of radius R. The function $r\Delta g$ is harmonic, because according to equation 5.8

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} (n-1) \left(\frac{R}{r}\right)^{n+1} T_n,$$

i.e.,

$$r\Delta g = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} (n-1) T_n = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \widetilde{T}_n,$$

in which $\widetilde{T}_n(\phi, \lambda) \stackrel{\text{def}}{=} (n-1) T_n(\phi, \lambda)$ is a perfectly legal surface spherical harmonic just like $T_n(\phi, \lambda)$ itself. Also, the dependence on the radius r, the factor $\left(\frac{R}{r}\right)^{n+1}$, is the same as for the (harmonic) potential. So, Poisson's integral equation 8.7 applies to function $r\Delta g$:

$$[r\Delta g(\phi, \lambda, r)] = \frac{R}{4\pi} \iint_\sigma \frac{(r^2 - R^2) \left[R\Delta g(\phi', \lambda', R)\right]}{\ell^3} \, d\sigma',$$

or

$$\Delta g(\phi, \lambda, r) = \frac{R^2}{4\pi r} \iint_\sigma \frac{r^2 - R^2}{\ell^3} \Delta g(\phi', \lambda', R) \, d\sigma'. \tag{8.8}$$

An alternative notation:

$$\Delta g = \frac{R^2}{4\pi r} \iint_\sigma \frac{r^2 - R^2}{\ell^3} \Delta g^* \, d\sigma,$$

in which Δg^* denotes the gravity anomaly at sea level, again calculated *by harmonic downward continuation* of the exterior field, in this case the expression $r\Delta g$.

See also Heiskanen and Moritz (1967) equation 2-160. Using the approximation $r + R \approx 2r$ yields still

$$\Delta g(\phi, \lambda, r) \approx \frac{R^2}{2\pi} \iint_\sigma \frac{r - R}{\ell^3} \Delta g(\phi', \lambda', R) \, d\sigma'.$$

Alternatively we derive the *spectral form*:

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} (n-1) T_n = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} \Delta g_n.$$

The degree constituent equation 3.7 gives the functions Δg_n:

$$\Delta g_n = \frac{2n+1}{4\pi} \iint_\sigma \Delta g P_n(\cos\psi) \, d\sigma,$$

with the aid of which

$$
\begin{aligned}
\Delta g &= \frac{1}{4\pi} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} (2n+1) \iint_\sigma \Delta g P_n(\cos\psi) \, d\sigma = \\
&= \frac{1}{4\pi} \iint_\sigma \left(\sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} (2n+1) P_n(\cos\psi) \right) \Delta g \, d\sigma = \\
&= \frac{1}{4\pi} \iint_\sigma K \Delta g \, d\sigma,
\end{aligned}
$$

in which

$$K(r, \psi, R) \stackrel{\text{def}}{=} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} (2n+1) P_n(\cos\psi)$$

is a (modified) Poisson kernel for gravity anomalies. Its closed form can be lifted from equation 8.8:

$$K(r, \psi, R) = \frac{R^2}{r} \frac{r^2 - R^2}{\ell^3}.$$

Compared to the Stokes kernel, the Poisson kernel drops off fast to zero for growing ℓ values. In other words, the evaluation of the integral equation may be restricted to a very local area, e.g., a cap of radius $1°$. See figure kalotti
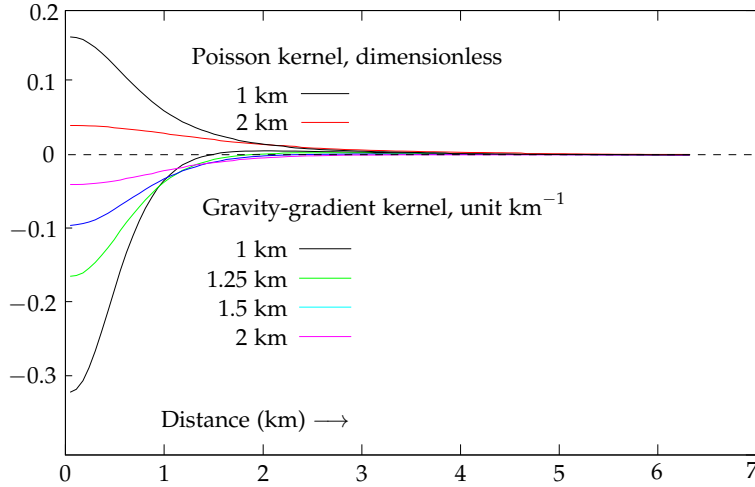
Figure 8.7. The Poisson kernel function for gravity anomalies as well as the kernel for the anomalous vertical gravity gradient, both at various height levels. These are kernel values to be used when evaluating the surface integral in map co-ordinates in km.

8.7. The main use of Poisson's kernel is the *harmonic continuation*, upward or downward, i.e., the shifting of gravity anomalies measured and computed at various levels to the same reference level.

In the limit $r \to R$ (sea level becomes the level of evaluation) this kernel function goes asymptotically to the Dirac δ function.

8.6 The vertical gradient of the gravity anomaly

Differentiate a formula obtained from formulas 5.7, 5.8:

$$\Delta g = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+2} \Delta g_n \implies \frac{\partial \Delta g}{\partial r} = -\frac{1}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+3} (n+2) \Delta g_n.$$

This equation is exact in spherical approximation. Its kernel function is well *localized*, in other words, it drops off to zero very fast, i.e., a small "cap" also here suffices for calculation.

Δg_n is, as calculated according to the degree constituent equation 3.7 from the anomaly field at sea level:

$$\Delta g_n = \frac{2n+1}{4\pi} \iint_\sigma \Delta g(\phi', \lambda', R) P_n(\cos \psi)\, d\sigma',$$

i.e.,

$$\frac{\partial \Delta g(\phi, \lambda, r)}{\partial r} = -\frac{1}{4\pi R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+3} (2n+1)(n+2) \iint_\sigma \Delta g(\phi', \lambda', R) P_n(\cos \psi)\, d\sigma' =$$

$$= \frac{1}{4\pi R} \iint_\sigma K'(r, \psi, R)\, \Delta g(\phi', \lambda', R)\, d\sigma', \tag{8.9}$$

where now the kernel function is

$$K'(r, \psi, R) = -\sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+3} (2n+1)(n+2) P_n(\cos \psi).$$

Alternatively we derive a closed expression. We start from the Poisson equation 8.8 for gravity anomalies, and differentiate[4] with respect to r:

$$\frac{\partial \Delta g(\phi, \lambda, r)}{\partial r} = \frac{\partial}{\partial r} \left(\frac{R^2}{4\pi r} \iint_\sigma \frac{r^2 - R^2}{(r^2 + R^2 - 2rR\cos\psi)^{3/2}} \Delta g(\phi', \lambda', R)\, d\sigma' \right) =$$

$$= \frac{R^2}{4\pi} \iint_\sigma \frac{1}{\ell^3} \left(2 - \frac{r^2 - R^2}{r^2} - \frac{3(2r - 2R\cos\psi)(r^2 - R^2)}{2r\ell^2} \right) \Delta g(\phi', \lambda', R)\, d\sigma' =$$

$$= \frac{R^2}{4\pi} \iint_\sigma \frac{1}{\ell^3} \left(2 - \frac{3(\ell^2 + r^2 - R^2)(r^2 - R^2)}{2r^2\ell^2} \right) \Delta g(\phi', \lambda', R)\, d\sigma' -$$

$$- \frac{1}{r}\frac{R^2}{4\pi r} \iint_\sigma \frac{(r^2 - R^2)\Delta g(\phi', \lambda', R)}{\ell^3}\, d\sigma' =$$

$$= \frac{R^2}{4\pi} \iint_\sigma \frac{1}{\ell^3} \left(2 - \frac{3(r^2 - R^2)}{2r^2} - \frac{3(r^2 - R^2)(r^2 - R^2)}{2r^2\ell^2} \right) \Delta g(\phi', \lambda', R)\, d\sigma' -$$

$$- \frac{1}{r}\Delta g(\phi, \lambda, r) =$$

$$= \frac{R^2}{4\pi} \iint_\sigma \frac{1}{\ell^3} \left(2 - \frac{3(r^2 - R^2)^2}{2r^2\ell^2} \right) \Delta g(\phi', \lambda', R)\, d\sigma' - \left(\frac{1}{r} + \frac{3}{2r} \right) \Delta g(\phi, \lambda, r) =$$

$$= \frac{R^2}{4\pi} \iint_\sigma \frac{1}{\ell^3} \left(2 - \frac{3(r^2 - R^2)^2}{2r^2\ell^2} \right) \Delta g(\phi', \lambda', R)\, d\sigma' - \frac{5}{2r}\Delta g(\phi, \lambda, r). \tag{8.10}$$

---

[4]Hint: use symbolic algebra software.

In the final right-hand side, the last term is very small, of order less than one part in a thousand, compared to the leftmost term. Both terms inside the square brackets are of the same order of magnitude.

Finally, note that, if we are integrating over the surface of the Earth sphere, radius R, rather than the unit sphere σ, radius 1, the coefficient R^2 drops out from both equation 8.8 and equation 8.10.

In Molodensky's method this or similar equations can be rapidly evaluated from very local gravimetric data.

The closed expression given in Heiskanen and Moritz (1967), expression 2-217[5], is the anomalous vertical gravity gradient evaluated *at sea level* (on the reference sphere). That is why it differs from expression 8.10 given above. Also in the expression given here, like in expression 8.9, we need gravity anomalies at sea level. Available however are anomalies at the *topographic surface* level. In practice we can proceed iteratively, by first assuming that the anomaly values observed at topography level *are* at sea level:

$$\Delta g^{(0)}(\phi', \lambda', R) \approx \Delta g(\phi', \lambda', r) = \Delta g(\phi', \lambda', R + H),$$

where $H = H(\phi', \lambda')$ is the topographic height at point (ϕ', λ'). When the first, crude anomalous gradient has been calculated, e.g., using equation 8.10, we may perform a real *reduction to sea level*, initially linearly:

$$\Delta g^{(1)}(\phi', \lambda', R) \approx \Delta g(\phi', \lambda', R + H) - \left. \frac{\partial \Delta g}{\partial r} \right|^{(0)} H,$$

and so forth.

8.7   Gravity reductions in geoid determination

8.7.1   Classical methods

Use of the Stokes equation for gravimetric geoid determination presupposes that all masses are *inside the geoid* — and the exterior field thus harmonic. For

---

[5]In the derivation there it is assumed that Δg is harmonic. Not so: $r\Delta g$ is harmonic. The error made is small.

this reason we move the topographic masses computationally to inside the geoid, in a way that needs to be specified. The classical methods for this are

- Helmert's (second) condensation method, section 6.5: the masses are shifted vertically down to the geoid into a surface density layer. After this, shifting gravity down from the topographic surface to sea level is easy. The indirect effect (the effect of the mass shifts on the geoid, the "Restore" step) is small.

- Isostatic reduction, in which the effects of both the topography and its compensation, i.e., the "roots" of mountains below sea level, are computationally removed. The indirect effect of this method is larger. See section 6.7 and equation 6.8.

- Bouguer reduction, section 6.2: the effect of the topographic masses is brutally removed from the observed gravity data, and, after geoid calculation, it is equally brutally restored to the result.

    Bouguer anomalies contain large negative biases in the mountains and therefore, the indirect effect of Bouguer reduction is excessive and extends over a large area. This is why the reduction is used more rarely.

8.7.2   The Residual Terrain Modelling (rtm) method

Imagine that, conceptually, the topographic masses are shifted into the geoid, to below sea level, in a way that *does not change the exterior field.* This is materially the same as determining the geoid associated with the *harmonically downward continued exterior field*.

The problem here is, that such a mass distribution below sea level, which produces the harmonically downward continued external potential in the space between topographic surface and geoid, doesn't always precisely exist. Or, that a suitable mass distribution will contain extremely large positive and negative masses close to each other, which are physically unrealistic.

One expresses this by saying that the problem is "*ill posed*". In such cases   huonosti asetettu
one uses *regularization*: one changes a little — as little as possible — the exterior field, so that it corresponds precisely to some sensible field that *can*

be harmonically continued below the topographic surface, and some sensible mass distribution interior to the geoid that produces it.

One can start, e.g., by filtering out the short-wave parts caused by the topography using a high resolution digital terrain model. This is called the RTM (residual terrain modelling) method.

In this method, we do not actually move all topographic masses to below the geoid. Instead, we use a bulldozer technique, figure 8.8: only masses close to the topographic surface are either removed or filled in, in a way which creates a smooth replacement topography that is long-wavelength only. The exterior field of this smoothed topography, unlike that of the original topography, lacks the shortest wavelengths. It may thus be downward continued to the geoid with sufficient precision.

puskutraktori

First we computationally remove from the topography *only* the short wavelengths (under 30 km) by moving the masses of the peaks into the valleys, i.e., we do a low-pass filtering. The effect of this on the free-air gravity anomalies Δg calculated from measurements is evaluated and taken into account: the "Remove" step.

In detail:

1. In each point P we apply the *terrain correction* as described in section 6.3 to the gravity anomalies.

2. Next, we remove the attraction of a Bouguer plate, thickness $H - H_{\text{RTM}}$, where H stands for the terrain height of point P, and H_{RTM} for the height of the smoothed, or low-pass filtered, terrain at the location of P. This effect is, according to equation 6.2 on page 125, equal to

$$2\pi G\rho \left(H - H_{\text{RTM}} \right),$$

in which ρ is the rock density assumed in the calculation.

3. After this, the *location* of the gravity anomaly is moved down[6] — "downward continuation" — from the original terrain level H to the surface of the new, smoothed terrain, H_{RTM}. For this, the Poisson equation may be used as described in section 8.8.

---

[6]Or up!

If the vertical gravity gradient of the terrain-reduced field equals the normal vertical gradient of gravity — according to section 5.4, $0.3\,{}^{\text{mGal}}/_{\text{m}}$ — this will leave the anomaly unchanged. Typically, there will be a small change: one may show — exercise 1–1 item 4 — that on the surface of a buried sphere of anomalous density $\Delta\rho$, there will be a radial anomalous gravity gradient of $\frac{8}{3}\pi G\Delta\rho$. For $\Delta\rho \ll \rho$, this will be negligible compared to the Bouguer-plate coefficient in item 2.

4. Rigorously speaking, an inverse terrain correction for the shapes of the smoothed terrain should be applied, to arrive at gravity anomalies realistic for this new replacement topography. Often this step is left out as the effect is small.

5. After that, harmonic downward continuation of the exterior field succeeds: in the exterior field remain nearly only the long wavelengths.

Because the mass shifts in the RTM method are so small, take place over such small distances, and are so short wavelength in nature, the *indirect effect* or "Restore" step — the change in geopotential due to the mass shifts that has to be applied in reverse to arrive at the final geopotential or geoid solution — is so small as to be often negligible. For the same reason the effect of unknown topographic density will remain small.

Finally we note that, because the RTM method removes the effect of the short-wavelength topography, it is also a suitable method for *interpolating* gravity anomalies. See Märdla (2017).

8.7.3 Downward continuation in linear approximation

The approach described above can, following Molodensky, be *linearized*:

$$T(\phi, \lambda, H) = \frac{R}{4\pi} \iint_\sigma \left(\Delta g(\phi', \lambda', H') - \frac{\partial \Delta g}{\partial H'} H' \right) S(\psi)\, d\sigma' + \frac{\partial T}{\partial H} H. \quad (8.11)$$

So, first we reduce the Δg measured and calculated at the topographic surface to sea level using the gradient of the anomalies and the terrain height H' of the measurement point, with the result

$$\Delta g^* = \Delta g - \frac{\partial \Delta g}{\partial H'} H'.$$
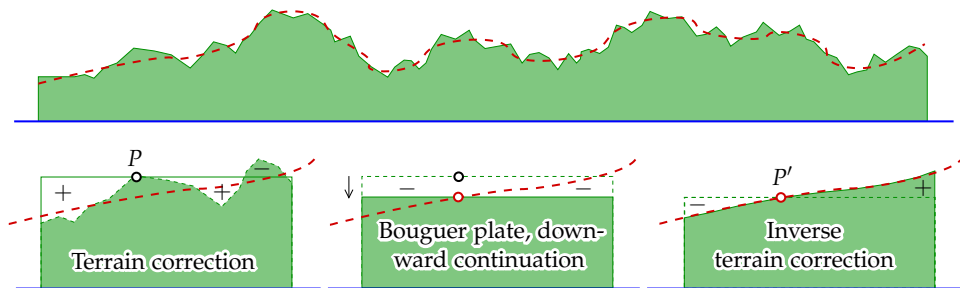
Figure 8.8. Residual terrain modelling (RTM). One removes from the terrain computationally the short wavelengths, i.e., the differences from the red dashed line: the masses rising above it are removed, the valleys below it are filled. After reduction, the red line, smoother than the original terrain, is the new terrain surface. The exterior potential of the new mass distribution will differ only little from the original one, but may be harmonically downward continued to sea level.

Left, terrain correction for point P; middle, Bouguer plate and gradient reduction to the level of the smoothed terrain point P'; and right, the inverse terrain correction for point P'.

After this, we apply, at sea level, the Stokes equation, and obtain the disturbing potential at sea level T^*. After this, the disturbing potential is "unreduced" back to terrain level, to the evaluation point, with the equation

$$T = T^* + \frac{\partial T}{\partial H} H.$$

In these equations T, its vertical derivative, and Δg and its vertical derivative always belong to the *exterior* harmonic gravity field, and the connection between them is the fundamental equation of physical geodesy, equation 5.5, in spherical geometry:

$$\Delta g = -\frac{\partial T}{\partial H} - \frac{2}{r} T,$$

in which $r = R + H$. Here, we need firstly the vertical derivative of the disturbing potential. This is easy: we have

$$\frac{\partial T}{\partial H} = -\Delta g - \frac{2}{r} T,$$

where the first term on the right is directly measured, and the second term's T is obtained iteratively from the main product of the solution process.

Calculating the vertical gradient of gravity anomalies, i.e., the *anomalous vertical gradient of gravity*, is harder. For this, we have the integral equation 8.9:

$$\frac{\partial \Delta g(\phi, \lambda, r)}{\partial r} = \frac{1}{4\pi R} \iint_\sigma K'(r, \psi, R) \, \Delta g(\phi', \lambda', R) \, d\sigma',$$

$$K'(r, \psi, R) = -\sum_{n=2}^\infty \left(\frac{R}{r}\right)^{n+3} (2n+1)(n+2) P_n(\cos\psi).$$

Luckily for practical calculations, the kernel K' of this integral is very localized and one does not need gravimetric data Δg from a very large area.

8.7.4   The evaluation point as the reference level

In the above equation 8.11 we used as the *reference level* the sea surface. This is arbitrary: we may use whatever reference level, e.g., H_0, in which case

$$T(\phi, \lambda, H) = \frac{R + H_0}{4\pi} \iint_\sigma \left(\Delta g(\phi', \lambda', H') - \frac{\partial \Delta g}{\partial H}(H' - H_0)\right) S(\psi) \, d\sigma' +$$
$$+ \frac{\partial T}{\partial H}(H - H_0).$$

If we now choose $H_0 = H$, the last term drops off, and we obtain

$$T(\phi, \lambda, H) = \frac{R + H}{4\pi} \iint_\sigma \left(\Delta g(\phi', \lambda', H') - \frac{\partial \Delta g}{\partial h}(H' - H)\right) S(\psi) \, d\sigma'.$$

In this case the reduction takes place from the height of the Δg measurement point to the height of the T evaluation point, probably a shorter distance than from sea level to evaluation height, especially in the immediate surrounding of the evaluation point. This means that the *linearization error will remain smaller*. What is bad, on the other hand, is that the expression in parentheses is now different for each evaluation point. This complicates the use of FFT based computation techniques, on which more later.

Here, we were all the time discussing the determination of the *disturbing potential* $T(\phi, \lambda, H)$; this is in practice the same as determining the *height anomaly*

$$\zeta(\phi, \lambda, H) = \frac{T(\phi, \lambda, H)}{\gamma_{Hh}} \approx \frac{T(\phi, \lambda, H)}{\gamma(\phi, H)},$$

equation 7.1. Here, γ is normal gravity calculated for point latitude ϕ ($\approx \varphi$) and height H.

8.8　The Remove-Restore method

All geoid determination methods currently in use are in one way or another "Remove-Restore" methods, even in several different ways.

1. From the observed gravity values, first the effect of the a global gravity field model is removed. This model is generally given in the form of a *spherical-harmonic expansion*. Thus, a *residual gravity field* is obtained

   ○ that has numerically smaller values — easier to work with — and

   ○ that is *more local*: the long "wavelengths", the patterns extending over large areas, have been removed from the residual field, only the local details remain.

2. From the observed gravity, the effects are removed of all masses that are *outside the geoid* — in practice, the topography. The purpose of this is to obtain a residual gravity field

   ○ on which the Stokes equation may be applied, because no masses are left outside the boundary surface, and

   ○ from which especially the very small "wavelengths" — details the size of which is of order a few kilometres — caused by the topography, are gone. After this, *prediction* of gravity values from sparse measurement values will work better.

Some gravity reduction methods — i.e., methods which computationally remove the gravity effect of the exterior masses — with good prediction properties, were already presented in subsection 8.7.1: Helmert condensation, Bouguer reduction, and isostatic reduction.

kommutoiva kaavio We may illustrate the Remove-Restore method by *commutative diagram* 8.9.

In this diagram the arrows with black text denote calculations that are recommended, because they are easy and accurate. Thearrow with grey text refers to direct computation, which again is troublesome and compute intensive.
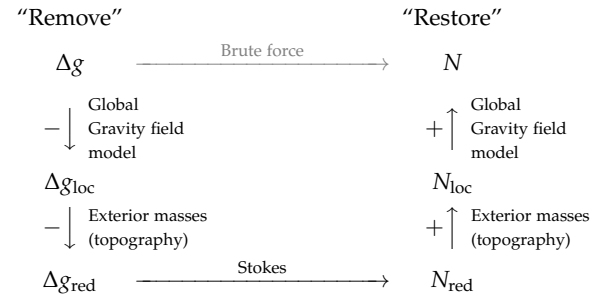
Figure 8.9. The Remove-Restore method as a commutative diagram.

8.8.1　Kernel modification in the Remove-Restore method

In the Remove-Restore method described above, the handling of reduced gravity anomalies Δg_{red} and geoid heights N_{red} happens typically within a relatively small area. For example, when using the FFT method, the area of computation is often a rectangular area in the map projection plane, drawn generously around the country or area the geoid of which is being computed.

Also if we compute the geoid directly by integrating the Stokes equation, we will evaluate this integral, after removing the effect of the global model from the given gravity data, only over a limited area or *cap*: evaluate the equation **kalotti**

$$N_{\text{red}} = \frac{R}{4\pi\gamma} \iint_{\sigma_0} S(\psi)\, \Delta g_{\text{red}}\, d\sigma, \tag{8.12}$$

where σ_0 is a cap on the unit sphere, the radius of which is, say, ψ_0.

The assumption behind this is, that, outside the cap, Δg_{red} is both small and rapidly varying, because the longer wavelengths have been removed from it with the global-model reduction. This may however be a dangerous assumption.

Write, in the above equation 8.12,

$$S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi)$$

and

$$\Delta g_{\text{red}}(\phi, \lambda) = \sum_{n=L+1}^{\infty} \Delta g_n(\phi, \lambda),$$

assuming that L is the largest degree number that is still along in the global spherical-harmonic expansion, or gravity model, that was subtracted from the data[7].

Now, because Δg_n is a certain linear combination of the surface spherical harmonics

$$Y_{nm}(\psi, \alpha) = \begin{cases} P_{n|m|}(\cos\psi)\sin|m|\alpha & \text{if } m = -n, \dots, -1, \\ P_{nm}(\cos\psi)\cos m\alpha & \text{if } m = 0, \dots, n, \end{cases}$$

i.e.,

$$\Delta g_n(\psi, \alpha) = \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha),$$

and also

$$Y_{n0}(\psi, \alpha) = P_n(\cos\psi),$$

it follows from the orthogonality of the Y functions, that

$$\iint_\sigma P_n(\cos\psi) Y_{n'm} d\sigma = \iint_\sigma Y_{n0} Y_{n'm} d\sigma = 0$$

if $n \neq n'$ or $m \neq 0$. Now we may write — note that the terms $n \leq L$ drop away:

$$S(\psi)\,\Delta g_{\text{red}}(\phi, \lambda) = S(\psi)\,\Delta g_{\text{red}}(\psi, \alpha) =$$

$$= \left(\sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi) \right) \times \left(\sum_{n=L+1}^{\infty} \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha) \right) =$$

$$= \left(\sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi) \right) \times \left(\sum_{n=L+1}^{\infty} \sum_{m=-n}^{n} \Delta g_{nm} Y_{nm}(\psi, \alpha) \right) =$$

$$= S^L(\psi)\,\Delta g_{\text{red}}(\phi, \lambda),$$

in which

$$S^L(\psi) = \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi)$$

is a so-called *modified Stokes kernel function*. The degree number L is called the *modification degree*. The size of the evaluation area σ_0 is chosen to be compatible with this.

---

[7]. . . and that the model is *accurate*!
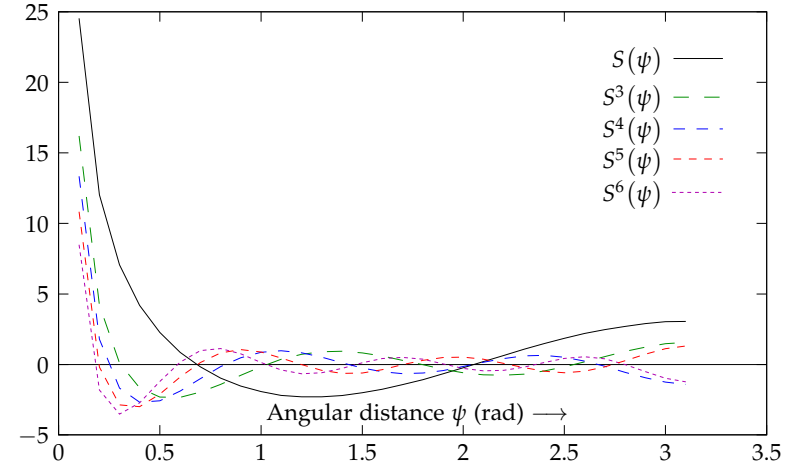
Figure 8.10. Modified Stokes kernel functions. Note how the kernel value outside the local area goes to zero for higher L values.

The modification method described here, restricting the Legendre expansion of the S function to higher degree numbers, is called the *Wong-Gore*[8] modification (Wong and Gore, 1969). A desirable property of the new kernel function S^L is, that it would be — at least compared to the original function S — *small* outside the cap area σ_0. In that case restricting the integral to the cap instead of the whole unit sphere (equation 8.12) does not do much damage. It is clear that S^L is much narrower than S, as in it, only the higher harmonic degrees are represented. This can be verified by plotting a graph of both curves. It doesn't go totally to zero outside the cap, however, but oscillates somewhat.

The reason for this oscillation is, that in the frequency (i.e., degree number) domain, the cut-off of the modified kernel is quite sharp. Transforming such a sharp edge between space and frequency domains will invariably produce

---

[8]L. Wong and R. C. Gore worked at the Aerospace Corporation, a space technology research institution in California. Wikipedia, The Aerospace Corporation.

The Stokes equation and other integral equations

Gibbsin ilmiö an oscillation, which is related to the so-called *Gibbs*[9] *phenomenon*.

8.8.2 Advanced kernel modifications

In the literature, other kernel modification methods are found. Their general form is

$$S^L(\psi) = \sum_{n=L+1}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi) + \sum_{n=2}^{L} (1-s_n) \frac{2n+1}{n-1} P_n(\cos\psi) =$$

$$= S(\psi) - \sum_{n=2}^{L} s_n \frac{2n+1}{n-1} P_n(\cos\psi), \qquad (8.13)$$

where the modification coefficients $s_n, n = 2, \ldots, L$ can be chosen[10]. They are chosen so as to minimize the values of S^L in the area outside the cap, $\sigma - \sigma_0$. In this way one may eliminate the truncation error of equation 8.12, and the oscillation of the Wong-Gore modification, almost entirely. Molodensky et al. (1962) developed already earlier such a method.

In the above equation 8.13 we want to minimize the function

$$S^L(\psi) = S(\psi) - \sum_{n=2}^{L} s_n \frac{2n+1}{n-1} P_n(\cos\psi)$$

over the area outside a local cap, $\sigma - \sigma_0$. Let us multiply this expression with each of the Legendre polynomials $P_n(\cos\psi), n = 2, \ldots, L$ in turn, and integrate over the area $\sigma - \sigma_0$ outside the local cap:

$$\int_{\sigma-\sigma_0} S(\psi) P_n(\cos\psi) \, d\sigma - \sum_{n'=2}^{L} s_{n'} \frac{2n'+1}{n'-1} \int_{\sigma-\sigma_0} P_{n'}(\cos\psi) P_n(\cos\psi) \, d\sigma = 0,$$

$$n = 2, \ldots, L,$$

a system of $L-1$ equations in the $L-1$ unknowns $s_{n'}$:

$$\sum_{n'=2}^{L} A_{nn'} s_{n'} = b_n,$$

[9]Josiah Willard Gibbs (1839–1903) was an American physicist, chemist, mathematician and engineer.

[10]The choice $s_n = 1$ again gives the simply modified Stokes kernel from which the low degrees have been completely removed.

with

$$b_n = \int_{\sigma-\sigma_0} S(\psi) P_n(\cos\psi) \, d\sigma$$

and

$$A_{nn'} = \frac{2n'+1}{n'-1} \int_{\sigma-\sigma_0} P_{n'}(\cos\psi) P_n(\cos\psi) \, d\sigma.$$

From this we can solve the s_n for every degree number n from 2 to L.

This solution sets to zero the expressions

$$\int_{\sigma-\sigma_0} S^L(\psi) P_n(\cos\psi) \, d\sigma, \qquad (8.14)$$

also for all values n from 2 to L.

The expressions 8.14 can be understood as *inner or scalar products*, between the functions S^L and P_n. Similarly the elements of $A_{nn'}$ contain the scalar products between the functions P_n and $P_{n'}$. Note that these scalar products do not vanish: when integrating over $\sigma - \sigma_0$, unlike over σ, the Legendre polynomials are *not* mutually orthogonal. Therefore, the A matrix is a full, symmetric, positive-definite matrix, not a diagonal matrix like when integrating over the full unit sphere σ.

The Legendre polynomials *are* however independent of each other also on this domain, and together span an $L-1$-dimensional linear vector space.

Now, outside the cap σ_0 of radius ψ_0, the Stokes kernel $S(\psi)$, by visual inspection, is "*smooth*". Depending of course on the values of ψ_0 and L, it may be so smooth that it does not contain any significant contribution from degree numbers higher than L. If this applies for S, it will also apply for S^L. This means that S^L will be a linear combination of the Legendre polynomials, i.e., an element of the vector space spanned by the polynomials $P_n, n = 2, \ldots, L$. But if this is so, and the scalar products 8.14 with each of the basis vectors vanish, then S^L *must be the zero function* on $\sigma - \sigma_0$.

See also Featherstone (2003).

Appendix A.1 explains more about linear vector spaces and the scalar product of vectors.

8.9 The effect of the local zone

In numerical gravimetric geoid computation one uses *averages* of anomalies computed over standard-sized cells or *blocks*, generally $5' \times 5'$, $10' \times 10'$, $30' \times 30'$ etc. At European latitudes, often sizes $3' \times 5'$, $5' \times 10'$, $6' \times 10'$ etc., are used, which are approximately square.

The following equation applies when evaluating an integral using block averages:

$$N = \sum_i c_i \overline{\Delta g_i},$$

in which $\overline{\Delta g_i}$ is the mean of block i, and the weight

$$c_i(\lambda, \phi) = \frac{R}{4\pi G} \iint_{\sigma_i} S(\psi)\, d\sigma,$$

where σ_i is the surface area of block i.

Numerical evaluation of such an integral, or *quadrature*, is done conveniently using Simpson's rule[11]:

$$c_i(\lambda, \phi) = \frac{R}{4\pi\gamma} \int_{\lambda_1}^{\lambda_2} \int_{\varphi_1}^{\varphi_2} S(\lambda, \phi, \lambda', \phi') \cos\phi\, d\phi'\, d\lambda' \approx$$

$$\approx \frac{R\Delta\lambda\Delta\phi}{4\pi\gamma} \sum_{j=-1}^{1} w_j \sum_{k=-1}^{1} w_k S_{jk}(\lambda, \phi),$$

where $\Delta\lambda$ and $\Delta\varphi$ are the block sizes in the latitude and longitude directions, and $w_{-1} = w_1 = \frac{1}{6}, w_0 = \frac{4}{6}$ are the weights.

$$S_{jk} = S(\lambda, \phi, \lambda'_0 + j\Delta\lambda, \phi'_0 + k\Delta\phi) \cos\phi, \ j,k = -1, 0, 1$$

are the values of expression $\left[S(\lambda, \phi, \lambda', \phi') \cos\phi \right]$ at the nodal points used in the evaluation, 3×3 of them. See figure 8.11. Also more complicated formulas (repeated Simpson or Romberg) can be employed.

---

[11] Thomas Simpson FRS (1710–1761) was an English mathematician and textbook writer. Actually Simpson's rule was used already a century earlier by Johannes Kepler.
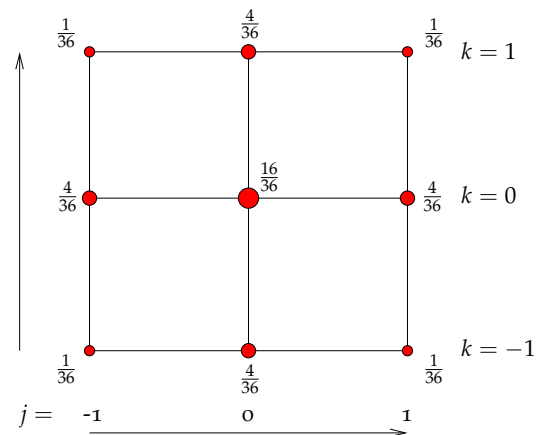
Figure 8.11. Simpson integration nodal weights in two dimensions.

One can show that the effect of the local (inner) zone on the geoid in the evaluation point (ϕ, λ) is proportional to the gravity anomaly in the point itself, Δg. Starting from the Stokes equation 8.1 with $S(\psi) \approx \left(\sin\frac{\psi}{2} \right)^{-1} \approx \frac{2}{\psi}$, we find, for a circular inner zone of radius ψ_0:

$$N_{\text{int}} = \frac{R}{4\pi\gamma} \int_0^{2\pi} \int_0^{\psi_0} \frac{2}{\psi} \Delta g \sin\psi\, d\psi\, d\alpha \approx \frac{R}{4\pi\gamma} \cdot 2\pi \cdot \Delta g_0 \cdot 2\psi_0 = \frac{s_0}{\gamma} \Delta g_0.$$

Here $s_0 = R\psi_0$ is the radius of the local block or cap *in metres*. Δg_0 is a special average of the gravity anomaly, the average of "ring averages" for radii between zero and s_0. If s_0 is small, one may take for this the anomaly value at the centre without incurring much error.

The local contributions to the deflections of the plumb line again are proportional to the *horizontal gradients* of gravity anomalies. We start from the Vening-Meinesz equations 8.4, with the above approximations for a local cap:

$$\left\{ \begin{array}{c} \xi_{\text{int}} \\ \eta_{\text{int}} \end{array} \right\} \approx \frac{1}{4\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \left(-\frac{2}{\psi^2} \right) \Delta g \left\{ \begin{array}{c} \cos\alpha \\ \sin\alpha \end{array} \right\} \cdot \sin\psi\, d\alpha\, d\psi.$$

We expand Δg into local rectangular metric co-ordinates x, y:

$$\Delta g \approx \Delta g_0 + x\frac{\partial\Delta g}{\partial x} + y\frac{\partial\Delta g}{\partial y} = \Delta g_0 + R\sin\psi \left(\cos\alpha\frac{\partial\Delta g}{\partial x} + \sin\alpha\frac{\partial\Delta g}{\partial y} \right),$$

and substitute:

$$\zeta_{\text{int}} \approx \frac{1}{4\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \left(-\frac{2}{\psi^2}\right) \left[\Delta g_0 + R\sin\psi \left(\cos\alpha \frac{\partial\Delta g}{\partial x} + \sin\alpha \frac{\partial\Delta g}{\partial y}\right)\right] \cdot$$
$$\cdot \cos\alpha \sin\psi \, d\alpha \, d\psi,$$

$$\eta_{\text{int}} \approx \frac{1}{4\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \left(-\frac{2}{\psi^2}\right) \left[\Delta g_0 + R\sin\psi \left(\cos\alpha \frac{\partial\Delta g}{\partial x} + \sin\alpha \frac{\partial\Delta g}{\partial y}\right)\right] \cdot$$
$$\cdot \sin\alpha \sin\psi \, d\alpha \, d\psi.$$

Here, the terms in Δg_0 drop out in α integration (because $\int_0^{2\pi} \sin\alpha = \int_0^{2\pi} \cos\alpha = 0$), as do the mixed terms in $\sin\alpha \cos\alpha$. What remains is

$$\zeta_{\text{int}} \approx -\frac{1}{4\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \frac{2}{\psi^2} R\sin\psi \cos\alpha \frac{\partial\Delta g}{\partial x} \cos\alpha \cdot \sin\psi \, d\alpha \, d\psi \approx$$

$$\approx -\frac{R}{2\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \frac{\partial\Delta g}{\partial x} \cos^2\alpha \cdot d\alpha \, d\psi = -\frac{R}{2\gamma} \int_0^{\psi_0} \frac{\partial\Delta g}{\partial x} \, d\psi,$$

$$\eta_{\text{int}} \approx -\frac{1}{4\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \frac{2}{\psi^2} R\sin\psi \sin\alpha \frac{\partial\Delta g}{\partial y} \sin\alpha \cdot \sin\psi \, d\alpha \, d\psi \approx$$

$$\approx -\frac{R}{2\pi\gamma} \int_0^{\psi_0} \int_0^{2\pi} \frac{\partial\Delta g}{\partial y} \sin^2\alpha \cdot d\alpha \, d\psi = -\frac{R}{2\gamma} \int_0^{\psi_0} \frac{\partial\Delta g}{\partial y} \, d\psi.$$

Carrying out the final ψ integration yields now, with $R\psi_0 = s_0$:

$$\zeta_{\text{int}} = -\frac{s_0}{2\gamma} \frac{\partial\Delta g}{\partial x}, \qquad \eta_{\text{int}} = -\frac{s_0}{2\gamma} \frac{\partial\Delta g}{\partial y}.$$

Sometimes these equations are useful, e.g., when estimating the errors of grid based methods. Let the mesh size of a grid be Δx, then we may set in the above equations $s_0 \approx \frac{\Delta x}{2}$, and for Δg we take

$$\Delta g^{\text{obs}} - \Delta g^{\text{grid}},$$

in which Δg^{grid} is the gravity anomaly value interpolated from the grid file at the evaluation point. In this way one obtains a rough estimate of how much error is due to the mesh size of the grid.

Self-test questions

1. What do the Stokes equation and its spectral form look like?

2. What does the Stokes kernel function $S(\psi)$ look like when expanded in Legendre polynomials?

3. What is a suitable aaproximation of the Stokes kernel when ψ is small?

4. What is an isotropic, what an anisotropic quantity on the Earth's surface? Give an example of the latter.

5. What does the Poisson integral equation describe?

6. Why are gravity reductions necessary when using the Stokes equation for computing a geoid model?

7. Which various gravity reduction methods are available?

8. Describe the Residual Terrain Modelling (RTM) method.

9. Describe the Remove-Restore method.

10. Why, in geoid determination, is the Stokes kernel function often midified? What does such a modification look like?

11. What is the Gibbs phenomenon?

Exercise 8–1: The Stokes equation

1. Derive a simpler form of the Stokes function $S(\psi)$, which is valid when the angular distance ψ is *small*. Really only one term!

2. Using this term, write the integral equation

$$N = \frac{R}{4\pi\gamma} \iint_\sigma S(\psi) \Delta g \, d\sigma$$

into polar co-ordinates, i.e., an integral of the form

$$\int_0^{2\pi} \int_0^\infty \cdots \, ds \, d\alpha,$$

in which $s = \psi R$ is the metric distance from the evaluation point, and α the azimuth angle (direction angle) from the evaluation point for N to the moving integration point for Δg.

[Hint: you need to consider *Jacobi's determinant* for the polar co-ordinates (s, α)]

3. Compute N (as a formula) if $\Delta g = \Delta g_0$ only within a circular disk $s \leq s_0$, and outside it $\Delta g = 0$. Assume that s_0 is *small*.

Spectral techniques, FFT

9

9.1 The Stokes theorem as a convolution

We start from the Stokes equation

$$T(\phi,\lambda) = \frac{R}{4\pi} \iint_\sigma S(\psi)\,\Delta g(\phi',\lambda')\,d\sigma',$$

in which (ϕ',λ') is the location of the moving point — i.e., the integration or observation point — on the Earth's surface, and (ϕ,λ) the location of the evaluation point, it too on the Earth's surface. Generally the locations of both points are given in spherical co-ordinates (ϕ,λ), and correspondingly the integration is executed over the surface of the unit sphere σ: a surface element is $d\sigma = \cos\phi\,d\phi\,d\lambda$, in which the factor $\cos\phi$ represents the *determinant of Jacobi*, for these spherical co-ordinates (ϕ,λ).

However *locally*, in a sufficiently small area, one may write the point co-ordinates also in rectangular form, and then express also the integral in rectangular co-ordinates. Suitable rectangular co-ordinates are, e.g., map projection co-ordinates, see figure 9.1.

A simple example of rectangular co-ordinates in the tangent plane would be

$$\begin{aligned}
x &= \psi R \sin\alpha,\\
y &= \psi R \cos\alpha,
\end{aligned}$$

(9.1)

in which α is the azimuth of the connection between evaluation point and moving data point. The centre of this projection is the point where the tangent
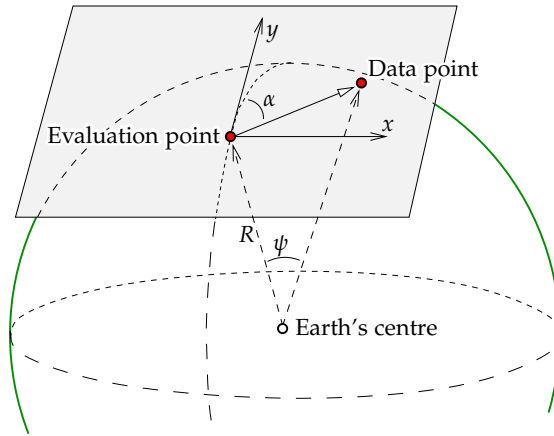
Figure 9.1. Map projection co-ordinates x, y in the local tangent plane.

plane touches the sphere. The locations of other points are measured by the angle at the Earth's centre, ψ, i.e., the *geocentric angular distance*, and by the direction angle in the tangent plane or *azimuth* α.

A more realistic example uses a popular conformal map projection, the *stereographic projection*:

$$x = 2\sin\left(\frac{\psi}{2}\right) R \sin \alpha,$$

$$y = 2\sin\left(\frac{\psi}{2}\right) R \cos \alpha.$$

In the limit for small values of ψ this corresponds to equations 9.1.

Taking the squares of equations 9.1, summing them, and dividing them by R^2 yields

$$\psi^2 \approx \frac{x^2 + y^2}{R^2}.$$

More generally ψ is the angular distance between the two points (x, y) (evaluation point) and (x', y') (data, integration or moving point) seen from the Earth's centre, approximately

$$\psi^2 \approx \left(\frac{x - x'}{R}\right)^2 + \left(\frac{y - y'}{R}\right)^2.$$

Furthermore we must account for Jacobi's determinant of the projection:

$$d\sigma = R^{-2}\, dx\, dy$$

and the Stokes equation becomes now

$$T(x, y) \approx \frac{1}{4\pi R} \iint_{-\infty}^{\infty} S(x - x', y - y')\Delta g(x', y')\, dx'\, dy', \quad (9.2)$$

a two-dimensional *convolution*[1].

Convolutions have nice properties in Fourier theory. If we designate the Fourier transform with the symbol $\mathcal{F}$, and convolution with the symbol $\circledast$, we may abbreviate the above equation as follows:

$$T = \frac{1}{4\pi R} S \circledast \Delta g,$$

and according to the *convolution theorem* ("Fourier transforms a convolution into a multiplication"):

$$\mathcal{F}\{T\} = \frac{1}{4\pi R}\mathcal{F}\{S\}\mathcal{F}\{\Delta g\}.$$

This (x, y) plane approximation works only, *if integration can be restricted to a local area*, where the curvature of the Earth's surface may be neglected. This is possible thanks to the use of global spherical-harmonic expansions, because these describe the long-wavelength part of the spatial variability of the Earth's gravity field. After we have removed from the observed gravity anomalies Δg the effect of the global spherical-harmonic model (the "Remove" step) we may safely forget the effect of areas far removed from the evaluation point: after this removal, the anomaly field will contain only the remaining short-wavelength parts, the effect of which cancels out at greater distances. Of course, once the integral has been computed and the local disturbing potential T_{loc} has been obtained, we must remember to add to it again the T_{glob} effect of the global spherical-harmonic expansion to be calculated separately (the "Restore" step).

---

[1] The integration extends from minus to plus infinity in both co-ordinates x and y. This can only be kept physically realistic on a curved Earth if it is assumed that the kernel S is of *bounded support*, i.e., it differs from zero only in a bounded area. This is the case for the modified kernels discussed in section 8.8.1.

9.2 Integration by FFT

The Fourier transform needed for applying the convolution theorem is calculated as a *discrete Fourier transform*. For this purpose exists the highly efficient Fast Fourier Transform, FFT (e.g., Vermeer, 1992b). There are several slightly differing formulations of the discrete Fourier transform to be found in the literature. It doesn't really matter which one is chosen, as long as it is a compatible pair of a forward Fourier transform $\mathcal{F}$ and a reverse Fourier transform $\mathcal{F}^{-1}$.

hilaesitys In preparation for this we first compute a discrete *grid representation* of the function $\Delta g(x, y)$, a rectangular table of Δg values on an equidistant (x, y) grid of points. The values may be, e.g., the function values themselves at the grid points[2]:

$$\Delta g_{ij} = \Delta g(x_i, y_j),$$

in which the co-ordinates of the grid points are

$$x_i = i\,\delta x, \quad i = -\frac{N}{2}, \ldots, \frac{N}{2} - 1,$$
$$y_j = j\,\delta y, \quad j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1,$$

for suitably chosen grid spacings $(\delta x, \delta y)$. The sequence of values of the subscripts i and j has been chosen so, that the centre of the area $(x = y = 0)$ is also in the centre of the constructed data table Δg_{ij}, $(i = j = 0)$. The integer N, assumed even, is here the grid size, assumed for simplicity the same in both directions.

Next, we do the same for the kernel function

$$S(\psi) = S(x - x', y - y') = S(\Delta x, \Delta y),$$

i.e., we write

$$S_{ij} = S(\Delta x_i, \Delta y_j),$$

---

[2]There exist alternatives to this. For example, one could calculate for every grid point the average over a square cell surrounding the point.

where again (N being the grid size):

$$\Delta x_i = i\,\delta x, \quad i = -\frac{N}{2}, \ldots, \frac{N}{2} - 1,$$
$$\Delta y_j = j\,\delta y, \quad j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1.$$

Also here, the choice of the value sequences for the i and j subscripts is based in the wish to get the central peak at the origin of the S function — $S(\Delta x, \Delta y) \to \infty$ when $(\Delta x, \Delta y) \to (0, 0)$ — placed at the centre of the grid of values S_{ij}[3].

Next:

1. The grid representations S_{ij} and Δg_{ij} thus obtained of the functions S and Δg are transformed to the *frequency domain* — they become functions $\mathcal{S}_{uv}$ and $\mathcal{G}_{uv}$ of the two "frequencies", the wave numbers u and v in the x and y directions. The spatial frequencies are $\omega_u = \frac{u}{L}, \omega_v = \frac{v}{L}$, in which L is the size of the area, assumed square.

2. They are multiplied with each other "one frequency pair at a time", i.e., we calculate

$$\mathcal{T}_{uv} = \frac{1}{2\pi R}\mathcal{S}_{uv}\mathcal{G}_{uv}, \quad u, v = 0 \ldots N - 1.$$

3. We transform the result, $\mathcal{T}_{uv} = \mathcal{F}\{T_{ij}\}$, back to the space domain: $T_{ij} = \mathcal{F}^{-1}\{\mathcal{T}_{uv}\}$, a grid $T_{ij} = T(x_i, y_j)$ describing the disturbing potential T. The disturbing potential of an arbitrary point can be obtained from this grid by interpolation. The co-ordinates x_i, y_j run as functions of i, j in the same way as described above for Δg[4].

This method is good for computing the disturbing potential T — and similarly the geoid height $N = \frac{T}{\gamma}$ — from gravity anomalies using the Stokes equation. It is just as good for evaluating other quantities, like, e.g., the vertical gradient

---

[3]Without this measure the result of the calculation would be correct, but in the wrong place...

[4]In fact, for both Δg and T we could choose the simpler grid geometry in which the subscript sequences are $i, j = 0, \cdots, N - 1$; however, for S it is mandatory to have the origin in the middle of the grid.

Free observation point selection $\xrightarrow{\text{Interpolation}}$ Regular point grid

Free observation point selection $\xrightarrow{\text{Direct solution}}$ (down)

Solution points in their own places $\xleftarrow{\text{Interpolation}}$ Regular point grid

Regular point grid $\xrightarrow{\text{FFT}}$ (down)
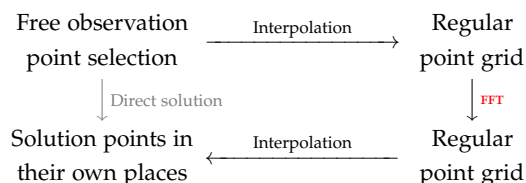
Figure 9.2. Commutative diagram for FFT.

of the gravity anomaly by the Poisson equation. The only requirement is, that the equation can be expressed as a *convolution*.

Also the *inversion calculation* is easy, as we shall see: in the frequency domain it is just a simple division.

Using the discrete Fourier transform requires that the input data in a field to be integrated — in the example, gravity anomalies — is given on a regular grid covering the area of computation, or can be converted to one. The result — e.g., the disturbing potential — is obtained on a regular grid in the same geometry. Values can then be *interpolated* to chosen locations.

kommutoiva kaavio

The FFT method may again be depicted as a *commutative diagram*, figure 9.2.

Appendix C offers a short explanation of why FFT works and what makes it as efficient as it is.

9.3   Solution in rectangular co-ordinates

In the above equation 9.2, the grid co-ordinates x and y are rectangular. For practical reasons, we would rather use latitude and longitude (ϕ, λ) as grid co-ordinates. In that way, the need to generate a new (x, y) point grid by interpolating from the given (φ, λ) one through a map projection calculation is avoided. However, working in geographic co-ordinates causes errors due to *meridian convergence* — as a latitude and longitude co-ordinate system isn't actually rectangular. Slightly more suitable would be the pair $(\phi, \lambda \cos \phi)$.

The problem has also been addressed on a more conceptual level.

9.3.1   The Strang van Hees method

The Stokes kernel function $S(\psi)$ depends only on the geocentric angular distance ψ between evaluation point (ϕ, λ) and observation point (ϕ', λ'). The angular distance may be written as follows (cosine rule on the sphere):

$$\cos \psi = \sin \phi' \sin \phi + \cos \phi' \cos \phi \cos(\lambda' - \lambda).$$

Substitute

$$\cos(\lambda' - \lambda) = 1 - 2\sin^2 \frac{\lambda' - \lambda}{2},$$

$$\cos \psi = 1 - 2\sin^2 \frac{\psi}{2},$$

$$\cos(\phi' - \phi) = 1 - 2\sin^2 \frac{\phi' - \phi}{2},$$

and obtain the *half-angle cosine rule*:

$$\cos \psi = \cos(\phi' - \phi) - 2\cos \phi' \cos \phi \sin^2 \frac{\lambda' - \lambda}{2}$$

$$\implies \sin^2 \frac{\psi}{2} = \sin^2 \frac{\phi' - \phi}{2} + \cos \phi' \cos \phi \sin^2 \frac{\lambda' - \lambda}{2}.$$

Here we may use the following approximation:

$$\cos \phi' \approx \cos \phi \overset{\text{def}}{=} \cos \phi_0,$$

in which ϕ_0 is a reference latitude in the middle of the calculation area. Now the above equation becomes

$$\sin^2 \frac{\psi}{2} \approx \sin^2 \frac{\phi' - \phi}{2} + \cos^2 \phi_0 \sin^2 \frac{\lambda' - \lambda}{2}, \tag{9.3}$$

which depends *only* on the differences $\Delta\phi \overset{\text{def}}{=} \phi' - \phi$ and $\Delta\lambda \overset{\text{def}}{=} \lambda' - \lambda$, a requirement for convolution.

After this, the FFT method may be applied by using co-ordinates (ϕ, λ)[5] and the re-written Stokes kernel

$$S^*(\Delta\phi, \Delta\lambda) \overset{\text{def}}{=} S\left(2 \arcsin \sqrt{\sin^2 \frac{\Delta\phi}{2} + \cos^2 \phi_0 \sin^2 \frac{\Delta\lambda}{2}}\right).$$

---

[5]In practice one uses the geodetic or geographic latitude φ instead of ϕ without significant error.

This clever way of using FFT in geographic co-ordinates was invented by the Dutchman G. Strang van Hees[6] in 1990.

9.3.2 "Spherical FFT", multi-band model

We divide the area in several narrow latitude bands. In each band we apply the Strang van Hees method using its own optimal central latitude.

Write the Stokes equation as follows:

$$N(\phi,\lambda) = \frac{R}{4\pi\gamma} \iint S(\phi - \phi', \lambda - \lambda'; \phi) \left[\Delta g(\phi',\lambda') \cos \phi' \right] d\phi' d\lambda', \quad (9.4)$$

where we have expressed $S(\cdot)$ as a function of latitude difference, longitude difference and *evaluation latitude*. Now, choose two support latitudes: ϕ_i and ϕ_{i+1}. Assume furthermore that S is between these a linear function of ϕ. In that case we may write

$$S(\Delta\phi, \Delta\lambda; \phi) = \frac{(\phi - \phi_i)\, S_{i+1}(\Delta\phi, \Delta\lambda) + (\phi_{i+1} - \phi)\, S_i(\Delta\phi, \Delta\lambda)}{\phi_{i+1} - \phi_i},$$

where $\Delta\phi = \phi - \phi', \Delta\lambda = \lambda - \lambda'$ and

$$S_i(\Delta\phi, \Delta\lambda) = S(\phi - \phi', \lambda - \lambda'; \phi_i),$$
$$S_{i+1}(\Delta\phi, \Delta\lambda) = S(\phi - \phi', \lambda - \lambda'; \phi_{i+1}).$$

We obtain by substitution into integral equation 9.4

$$N(\phi,\lambda) = \frac{R}{4\pi\gamma} \left(\frac{\phi_{i+1} - \phi}{\phi_{i+1} - \phi_i} I_i + \frac{\phi - \phi_i}{\phi_{i+1} - \phi_i} I_{i+1} \right) \quad (9.5)$$

with

$$I_i = \iint S_i(\Delta\phi, \Delta\lambda) \left[\Delta g(\phi', \lambda') \cos \phi' \right] d\phi' d\lambda',$$
$$I_{i+1} = \iint S_{i+1}(\Delta\phi, \Delta\lambda) \left[\Delta g(\phi', \lambda') \cos \phi' \right] d\phi' d\lambda'.$$

This equation is the *linear combination of two convolutions*. Both are evaluated by FFT and from the solutions obtained one forms the weighted mean according to equation 9.5.

---

[6]Govert L. Strang van Hees (1932–2012) was a Dutch gravimetric geodesist.

In this method we use, instead of the approximative equation 9.3, an exact equation, in which ϕ' is expressed into ϕ and $\Delta\phi$:

$$\sin^2 \frac{\psi}{2} = \sin^2 \frac{\phi' - \phi}{2} + \cos\phi' \cos\phi \sin^2 \frac{\lambda' - \lambda}{2} =$$
$$= \sin^2 \frac{\Delta\phi}{2} + \cos(\phi - \Delta\phi)\cos\phi \sin^2 \frac{\Delta\lambda}{2}.$$

Here again, we calculate S_i and S_{i+1} for the values $\phi = \phi_i$ and $\phi = \phi_{i+1}$, we evaluate the integrals with the aid of the convolution theorem, and interpolate $N(\phi,\lambda)$ according to equation 9.5 when $\phi_i \leq \phi < \phi_{i+1}$. After this, the solution isn't entirely exact, because inside every band we still use linear interpolation. However by making the bands narrower, we can keep the error arbitrarily small.

9.3.3 "Spherical FFT", Taylor expansion model

This somewhat more complicated but also more versatile approach expands the Stokes kernel into a Taylor expansion with respect to latitude about a *reference latitude* located in the middle of the computation area[7]. Each term in the expansion depends only on the *difference* in latitude. The integral to be calculated similarly expands into terms, of which each contains a pure convolution.

Let us write the *general* problem as follows:

$$\ell(\varphi, \lambda) = \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} C(\varphi, \varphi', \Delta\lambda) \left[m(\varphi', \lambda') \cos \varphi' \right] d\varphi' d\lambda',$$

in which ℓ contains values to be computed, m values given, and C is the coefficient or kernel function. Here is assumed only *rotational symmetry* around the Earth's axis for the geometry, i.e., the kernel function depends only on the difference between longitudes $\Delta\lambda$ rather than the absolute longitudes λ, λ'.

---

[7]In the literature the method has been generalized by expanding the kernel also with respect to height.

In a concrete case m contains for example gravity anomaly values Δg in various points (φ', λ'), ℓ contains geoid heights N in various points (φ, λ), and C contains coefficients calculated using the Stokes kernel function.

We first change the dependence upon φ, φ' into a dependence upon $\varphi, \Delta\varphi$:

$$C = C(\varphi, \varphi', \Delta\lambda) = C(\Delta\varphi, \Delta\lambda, \varphi).$$

Linearize:

$$C = C_0(\Delta\varphi, \Delta\lambda) + (\varphi - \varphi_0)\, C_\varphi(\Delta\varphi, \Delta\lambda) + \dots$$

where we define for a suitable *reference latitude* φ_0,

$$C_0(\Delta\varphi, \Delta\lambda) \overset{\text{def}}{=} C(\Delta\varphi, \Delta\lambda, \varphi_0),$$
$$C_\varphi(\Delta\varphi, \Delta\lambda) \overset{\text{def}}{=} \left.\frac{\partial}{\partial\varphi}C(\Delta\varphi, \Delta\lambda, \varphi)\right|_{\varphi=\varphi_0}.$$

Note that this expansion into two terms will work only for a limited range in $\Delta\varphi$, and that the kernel function C is assumed to be of bounded support. In this case, the integrals may be calculated within a limited area instead of over the whole Earth.

Substitution yields

$$\ell(\varphi, \lambda) = \iint C(\Delta\varphi, \Delta\lambda, \varphi) \cdot m(\varphi', \lambda') \cos\varphi'\, d\varphi'\, d\lambda' =$$
$$= \iint \big(C_0 + (\varphi - \varphi_0)\, C_\varphi\big) \cdot m \cos\varphi'\, d\varphi'\, d\lambda' =$$
$$= \iint C_0 \cdot m \cos\varphi'\, d\varphi'\, d\lambda' +$$
$$+ (\varphi - \varphi_0) \iint C_\varphi \cdot m \cos\varphi'\, d\varphi'\, d\lambda'. \tag{9.6}$$

Important here is now, that the integrals in the first and second terms,

$$\iint C_0(\Delta\varphi, \Delta\lambda)\, \big[m(\varphi', \lambda') \cos\varphi'\big]\, d\varphi'\, d\lambda' = C_0 \circledast [m \cos\varphi],$$
$$\iint C_\varphi(\Delta\varphi, \Delta\lambda)\, \big[m(\varphi', \lambda') \cos\varphi'\big]\, d\varphi'\, d\lambda' = C_\varphi \circledast [m \cos\varphi],$$

are both convolutions: both C functions depend only on $\Delta\varphi$ and $\Delta\lambda$. Both integrals can be calculated only if the corresponding $\Delta\varphi = \varphi - \varphi'$ and

$\Delta\lambda = \lambda - \lambda'$, and the corresponding coefficient grids C_0, C_φ, are calculated first in preparation. After this — in principle expensive, but, thanks to FFT and the convolution theorem, a lot cheaper — integration, computing the compound 9.6 is cheap: one multiplication and one addition for each evaluation point (φ, λ).

Example: let the evaluation area at latitude $60°$ be $10° \times 20°$ in size. If the grid mesh size is $5' \times 10'$, the number of cells is 120×120. Let us choose, e.g., a 256×256 grid (i.e., $N = 256$) and fill the missing values with extrapolated values.

Also the values of the kernel functions C_0 and C_φ are calculated on a 256×256 size $(\Delta\varphi, \Delta\lambda)$ grid. The number of these is thus also $65\,536$. Calculating the convolutions $C_0 \circledast [m \cos\varphi]$ and $C_\varphi \circledast [m \cos\varphi]$ by means of FFT — i.e.,

$$\iint C_0(\Delta\varphi, \Delta\lambda)\, m(\varphi', \lambda') \cos\varphi'\, d\varphi'\, d\lambda' = C_0 \circledast [m \cos\varphi] =$$
$$= \mathcal{F}^{-1}\big\{\mathcal{F}\{C_0\}\mathcal{F}\{m \cos\varphi\}\big\},$$
$$\iint C_\varphi(\Delta\varphi, \Delta\lambda)\, m(\varphi', \lambda') \cos\varphi'\, d\varphi'\, d\lambda' = C_\varphi \circledast [m \cos\varphi] =$$
$$= \mathcal{F}^{-1}\big\{\mathcal{F}\{C_\varphi\}\mathcal{F}\{m \cos\varphi\}\big\},$$

requires $(N^2) \times 2 \log(N^2) = 65\,536 \times 16 = $ more than a million operations, multiplication with $(\varphi - \varphi_0)$ and adding together, each again $65\,536$ operations.

The grid matrices corresponding to functions C_0 and C_φ are obtained as follows: for three reference latitudes $\varphi_{-1}, \varphi_0, \varphi_{+1}$ we compute numerically the grids

$$C_{-1} = C(\Delta\varphi, \Delta\lambda, \varphi_{-1}),$$
$$C_0 = C(\Delta\varphi, \Delta\lambda, \varphi_0),$$
$$C_{+1} = C(\Delta\varphi, \Delta\lambda, \varphi_{+1}),$$

in which C_0 is directly available, and

$$C_\varphi \approx \frac{C_{+1} - C_{-1}}{\varphi_{+1} - \varphi_{-1}}.$$

Also *inversion calculation* is thus directly feasible. Let ℓ be given in suitable point grid form. We compute the first approximation to m as follows:

$$\mathcal{F}\{C_0\}\mathcal{F}\{m\cos\varphi\} = \mathcal{F}\{\ell\} \implies [m\cos\varphi]^{(0)} = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\ell\}}{\mathcal{F}\{C_0\}}\right\}.$$

The second approximation is obtained by first calculating

$$\ell^{(0)} = C_0 \circledast [m\cos\varphi]^{(0)} + (\varphi - \varphi_0) \cdot C_\varphi \circledast [m\cos\varphi]^{(0)},$$

after which we make the improvement:

$$[m\cos\varphi]^{(1)} = [m\cos\varphi]^{(0)} + \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{\ell - \ell^{(0)}\}}{\mathcal{F}\{C_0\}}\right\},$$

and so on, iteratively. Two, three interations are usually enough. This method has been used to compute underground mass points from gravity anomalies to represent the exterior gravity field of the Earth. More is explained in Forsberg and Vermeer (1992).

9.3.4   "1D-fft"

This is a limiting case of the previous ones, in which FFT is used only in the longitude direction. In other words, a zones method where the zones have a width of only a single grid row. This method is exact if all longitudes $(0° - 360°)$ are along in the calculation. It requires a bit more computing time compared to the previous methods. In fact, it is identical to a Fourier transform in variable λ, longitude. Details are found in Haagmans et al. (1993).

9.4   Bordering and tapering of the data area

The discrete Fourier transform presupposes the data to be *periodically continuous*. In other words, it is assumed that when connecting the Eastern edge of the data area to the Western edge, and the Northern edge to the Southern edge, the data has to be continuous across these edges[8]. In practice this is not the case. We are faced with two different problems:

---

[8]Topologically the area with the edges thus connected is equivalent to a *torus*, and the data is presupposed to be continuous on the surface of the torus.

1. the data on the opposite side of an edge must be so far away to have no noticeable influence on the result of the calculation, and

2. the data must be continuous across the edges.

Therefore, always when using FFT with the convolution theorem 9.2, two measures need to be taken.

1. we continue the data by adding a border area to the data area, so-called *bordering*. Often the border area is 25% of the size of the data area; then, the size of the whole calculation area will become four times that of the data area itself. The border is often filled with zeroes, although predicted values — or even measured values, if those exist — are a better choice.

   Also the calculation area for the *kernel function* is made similarly four times larger. In this case, as the function is *symmetric*, the border area is filled with real (computable) values, making it automatically periodically continuous.

2. Because the discrete Fourier transform assumes periodicity, one must make sure that the data really is periodic. If the values at the borders are not zero, they may be forced to zero by multiplying the whole data area by a so-called *tapering* function, which goes smoothly to zero towards the edges. Such a function can easily be built, e.g., a cubic spline polynomial or a Tukey or cosine tapering. See figure 9.3, showing a 25% tapering function, as well as example images 9.4, where one sees how non-periodicity — differing left and right, and upper and lower, edges — causes horizontal and vertical artefacts in the Fourier transform. These artefacts are called the *Gibbs phenomenon*, already mentioned in section 8.8.1: a sharp cut-off or edge in the space domain will produce signal in all frequencies up to the highest ones.

Many journal articles have appeared on these technicalities. Groups that have been involved in early development of FFT geoid determination already in the 1980s are Forsberg's group in Copenhagen, the group of Schwarz and Sideris in Calgary, Canada, the Delft group (Strang van Hees, Haagmans, De Min, Van Gelderen), the Milanese (Sansò, Barzaghi, Brovelli), Heiner Denker at the Hannover "Institut für Erdmessung", and many others.

25% ⟷ 50% ⟷ 25%

Data area
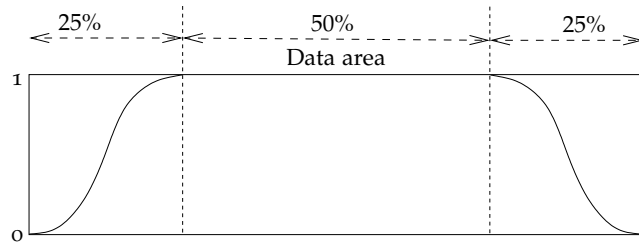
Figure 9.3. "Tapering" 25%.

Figure 9.4. Example images for FFT transform without (above) and with (below) tapering. Used on-line FFT service Watts (2004).

9.5   Computing a geoid model with fft

Nowadays computing a geoid or quasi-geoid model is easy thanks to increased computing power, especially using FFT. On the other hand the spread of precise geodetic satellite positioning has made the availability of precise geoid models an important issue, so that one can use GNSS technology for rapid and inexpensive height determination.

9.5.1   The GRAVSOFT software

The GRAVSOFT geoid computation software has been mainly produced in Denmark. Authors include Carl Christian Tscherning[9], René Forsberg, Per Knudsen, the Norwegian Dag Solheim, and the Greek Dimitris Arabelos. The manual is Forsberg and Tscherning (2008).

This package is in widespread use and offers, in addition to variants of FFT geoid determination, also, e.g., least-squares collocation, routines for evaluation of various terrain effects, etc. Its spread can be partly explained by it being free for scientific use, and being distributed as source code. It is also well documented. Therefore it has also found commercial use, e.g., in the petroleum extraction industry.

GRAVSOFT has been used a lot also for teaching, e.g., at many research schools organized by the IAG (International Association of Geodesy) in various countries. ISG, Geoid Schools.

9.5.2   The Finnish FIN2000 geoid

Currently two geoid models are in use in Finland: FIN2000 (figure 9.5) and FIN2005N00 (Bilker-Koivula and Ollikainen, 2009). The first model is a reference surface for the N60 height system: using it together with GNSS positioning allows determination of the N60 heights of points. The model

---

[9]Carl Christian Tscherning (1942–2014) was a Danish physical geodesist well known for his research into the gravity field of the Earth. He did groundbreaking work on statistical computation methods for modelling the Earth's gravity field from many different measurement types.
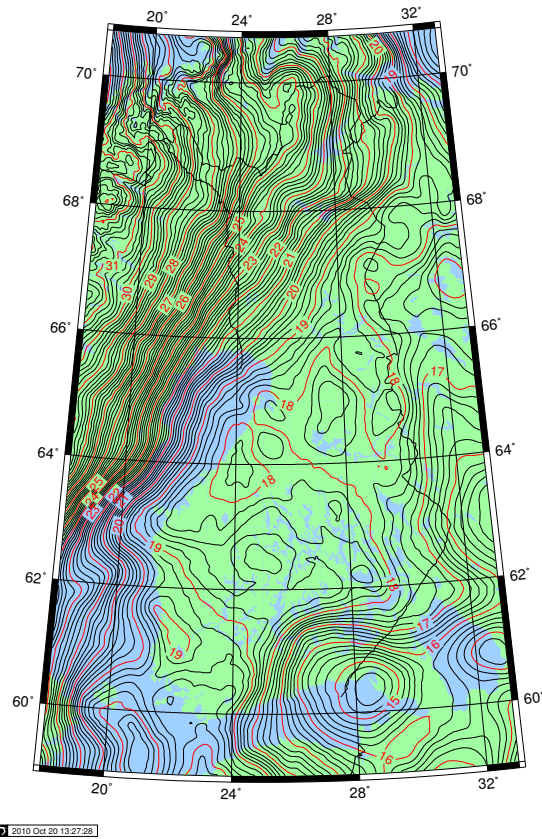
Figure 9.5. The Finnish FIN2000 geoid. Data source: Finnish Geodetic Institute.

gives geoid heights above the GRS80 reference ellipsoid. The second model is similarly a reference surface for the new N2000 height system. It too gives heights from the GRS80 reference ellipsoid.

The precisions (mean errors) of FIN2000 and FIN2005N00 are on the level of $\pm 2 - 3$ cm.

9.6 Use of FFT computation in other contexts

9.6.1 Satellite altimetry

The Danish researchers Per Knudsen and Ole Balthasar Andersen have computed a gravity map of the world ocean by starting from satellite altimetry derived "geoid heights" and inverting them to gravity anomalies (Andersen et al., 2010). A pioneer of the method has been David Sandwell from the Scripps Institute of Oceanography in California (e.g., Garcia et al., 2014). The short-wavelength features in the map can tell us about the sea-floor topography.

9.6.2 Satellite gravity missions and airborne gravimetry

Also the data from satellite gravity missions (like CHAMP, GRACE and GOCE) can be regionally processed using the FFT method: in the case of GOCE, the inversion of gradiometric measurements, i.e., calculating geoid heights on the Earth's surface from measurements made at satellite level. Also airborne gravity measurements are processed in this way. The problem is called "harmonic downward continuation" and is in principle unstable.

Airborne gravimetry is a practical method for gravimetric mapping of large areas. In the pioneering days, the gravity field over Greenland was mapped, as well as many areas around the Arctic and Antarctic. Later, areas were measured like the Brazilian Amazonas, Mongolia and Ethiopia (Bedada, 2010), where no full-coverage terrestrial gravimetric data existed. The advantage of this method is that one measures rapidly large areas in a homogeneous way. Also for the processing of airborne gravimetry data, FFT is suitable.

9.7 Computing terrain corrections with FFT

The *terrain correction* is a very localized phenomenon, the calculation of which requires high-resolution terrain data from a relatively small area surrounding the computation point. Thus, calculating the terrain correction is ideally suited for the FFT method.

We show how, with FFT, we can simply and efficiently evaluate the terrain correction. We make the following simplifying *assumptions*:

1. Terrain slopes are relatively small.

2. The density ρ of the Earth's crust is constant.

3. The Earth is flat – the "shoebox world".

These assumptions are not mandatory. The general case however leads us into a jungle of equations without aiding the conceptual picture.

The *terrain correction*, the joint effect of all the topographic masses, or lacking topographic masses, above and below the height level H of the evaluation point, can be calculated under these assumptions using the following rectangular equation, which describes the attraction of rock columns projected vertically (figure 6.5):

$$
\begin{aligned}
TC(x,y) &= \iint_{-\infty}^{+\infty} \frac{G\rho\,(H'-H)}{\ell^2} \cos\theta\, dx'\, dy' \\
&= \iint_{-\infty}^{+\infty} \frac{G\rho\,(H'-H)}{\ell^2} \cdot \frac{1}{2}\frac{H'-H}{\ell}\, dx'\, dy' \\
&= \frac{1}{2}G\rho \iint_{-\infty}^{+\infty} \frac{(H'-H)^2}{\ell^3}\, dx'\, dy'.
\end{aligned}
\tag{9.7}
$$

Here $G\rho\,(H'-H)\,\ell^{-2}$ is the attraction of the column and $\frac{1}{2}(H'-H)\,\ell^{-1}$ is the cosine of the angle θ between the force vector — assumed coming from the midpoint of the rock column — and the vertical direction. This is the so-called *prism method*.

We will make a linear approximation, wherein ℓ, the slant distance between the evaluation point (x,y) and the moving data point (x',y'), is also the *horizontal distance*:

$$
\ell^2 \approx (x'-x)^2 + (y'-y)^2.
$$

Equation 9.7 is easy to check straight from Newton's law of gravitation. When it is assumed that the terrain is relatively free of steep slopes, then ℓ is large compared to $H'-H$.

From equation 9.7 we obtain by development into terms:

$$
\begin{aligned}
TC(x,y) = {}& \frac{1}{2}G\rho h H^2 \iint_{-\infty}^{+\infty} \frac{1}{\ell^3}\, dx'\, dy' \\
& - G\rho H \iint_{-\infty}^{+\infty} \frac{H'}{\ell^3}\, dx'\, dy' + \\
& + \frac{1}{2}G\rho \iint_{-\infty}^{+\infty} \frac{(H')^2}{\ell^3}\, dx'\, dy',
\end{aligned}
\tag{9.8}
$$

in which every integral is a *convolution* with kernel ℓ^{-3}, and functions to be integrated are 1, H' and $(H')^2$.

Unfortunately the function ℓ^{-3} as implicitly defined above has no Fourier transform, wherefore we change the above definition a tiny bit by adding a small term:

$$
\ell^2 = (x'-x)^2 + (y'-y)^2 + \delta^2.
\tag{9.9}
$$

Then, the terms in the above sum are large numbers that almost cancel each other, giving a nearly correct result. Numerically this is however an unpleasant situation.

If ℓ is defined according to equation 9.9, then the Fourier transform of kernel ℓ^{-3} is (Harrison and Dickinson, 1989):

$$
\mathcal{F}\{\ell^{-3}\} = \frac{2\pi}{\delta}\exp(-2\pi\delta q) = \frac{2\pi}{\delta}\left(1 - 2\pi\delta q + \frac{4\pi^2 q^2 \delta^2}{1\cdot 2} - \cdots\right),
$$

in which $q \stackrel{\text{def}}{=} \sqrt{u^2 + v^2}$, and u, v are wave numbers (i.e., "frequencies") in the x and y directions in the (x,y) plane. If we substitute this into equation 9.8, we notice that the terms containing δ^{-1} sum to zero, and of course also the terms containing positive powers of δ vanish when $\delta \to 0$. As follows (Harrison and Dickinson, 1989):

$$
\begin{aligned}
\mathcal{F}\{TC\} \approx {}& \frac{1}{2}G\rho H^2 \mathcal{F}\{1\}\left(\frac{2\pi}{\delta}(1-2\pi\delta q)\right) - \\
& - G\rho H \mathcal{F}\{H'\}\left(\frac{2\pi}{\delta}(1-2\pi\delta q)\right) \\
& + \frac{1}{2}G\rho \mathcal{F}\{(H')^2\}\left(\frac{2\pi}{\delta}(1-2\pi\delta q)\right)
\end{aligned}
$$

where we leave off all terms in higher powers of δ.

Re-order the terms:

$$\mathcal{F}\{TC\} = \frac{\pi}{\delta}G\rho\left(H^2\mathcal{F}\{1\} - 2H\mathcal{F}\{H'\} + \mathcal{F}\left\{\left(H'\right)^2\right\}\right) + $$
$$ + 4\pi^2q\left(-\frac{1}{2}G\rho H^2\mathcal{F}\{1\} + G\rho H\mathcal{F}\{H'\} - \frac{1}{2}G\rho\mathcal{F}\left\{\left(H'\right)^2\right\}\right).$$

Because $\mathcal{F}\{1\} = 0$ if $q \neq 0$, the first term inside the second term will always vanish. We obtain (remember that H is a constant, the height of the evaluation point):

$$\mathcal{F}\{TC\} = \frac{\pi}{\delta}G\rho\left(\mathcal{F}\left\{H^2 - 2HH' + \left(H'\right)^2\right\}\right) + $$
$$ + 4\pi^2q\left(G\rho H\mathcal{F}\{H'\} - \frac{1}{2}G\rho\mathcal{F}\left\{\left(H'\right)^2\right\}\right)$$

and the reverse Fourier transform yields

$$TC = \frac{2\pi G\rho}{\delta}\left(\frac{1}{2}H^2 - H'H + \frac{1}{2}\left(H'\right)^2\right) + $$
$$ + G\rho H\mathcal{F}^{-1}\left\{\mathcal{F}\{H'\} \cdot 4\pi^2q\right\}$$
$$ - \frac{1}{2}G\rho\mathcal{F}^{-1}\left\{\mathcal{F}\left\{\left(H'\right)^2\right\} \cdot 4\pi^2q\right\}.$$

In the first term

$$\frac{1}{2}H^2 - H'H + \frac{1}{2}\left(H'\right)^2 = \frac{1}{2}\left(H' - H\right)^2 = 0$$

in point (x,y) in which $H' = H$, and we obtain

$$TC = 4\pi^2G\rho\mathcal{F}^{-1}\left\{q \cdot \left(H\mathcal{F}\{H'\} - \frac{1}{2}\mathcal{F}\left\{\left(H'\right)^2\right\}\right)\right\},$$

from which now the troublesome δ^{-1} has vanished.

A *condition* for this "regularization" or "renormalization" is, that at point (x,y) $H' = H$, i.e., the evaluation happens at the Earth's surface. The convolutions above are evaluated by the ꜰꜰᴛ technique. A more detailed account is found, e.g., in the article Vermeer (1992a).

For calculating the terrain correction TC in the *exterior space* — airborne gravimetry, but also the effect of the sea floor at the sea surface, or the effect of the Mohorovičić discontinuity at the Earth's surface — there are techniques that express TC as a sum of convolutions, as a Taylor series expansion. An early paper on this is Parker (1972).

Self-test questions

1. What is the definition of a convolution?

2. Explain the convolution theorem.

3. Explain the basic idea of the Strang van Hees method.

4. What other approaches are there to apply the ꜰꜰᴛ method on a curved (spherical or ellipsoidal) surface?

5. Why are bordering of the data area and tapering of the data necessary?

6. In addition to geoid determination, where in physical geodesy is the ꜰꜰᴛ method also used?

7. When computing the terrain correction on the Earth's surface, describe the "δ trick" used in the derivation. Why is it necessary, and how does one make the δ vanish again?

Statistical methods

10

10.1 The role of uncertainty in geophysics

In geophysics, we often obtain results based on uncertain, incomplete, or otherwise deficient observational data. This applies also in the study of the Earth's gravity field: e.g., the density of gravity observations on the Earth surface varies a lot, and large areas on the oceans and polar regions are covered only by a very sparse network of measurements. We speak of *spatial undersampling*.

Measurement technologies that work from space typically provide coverage of the whole globe, oceans, poles and all. They however don't measure at a very high *resolution*. Either the resolution of the method is limited — e.g., the gravity field parameters calculated from satellite orbit perturbations — or the instruments measure only straight underneath the satellite's path — like satellite altimetry.

Another often relevant uncertainty factor is, that one can do precise measurements on the Earth surface, but inside the Earth the uncertainty is much larger and the data is obtained much more indirectly.

In previous chapters we described techniques by which we could calculate desired values or parameters for the Earth's gravity field, assuming that, e.g., gravity anomalies are available everywhere on the Earth surface, and with arbitrarily high resolution. In this chapter we look at mathematical means to handle real-world situations where this is not the case.

10.2   Linear functionals

In mathematics, an operator or mapping that associates with every function in a given function space a certain numerical value is called a *functional*. One such is, e.g., a (partial) derivative in a certain point:

$$f \mapsto \frac{\partial}{\partial x} f(x) \Big|_{x=x_0}.$$

A trivial functional is also the *evaluation functional*, the function value itself for a certain argument value,

$$f \mapsto f(x_0).$$

Other functionals are, e.g., the integral over a given area:

$$f \mapsto \int_\sigma f(x)\, dx,$$

and so on.

We may write symbolically

$$L = \frac{\partial}{\partial x} \Big|_{x=x_0}, \text{ i.e., } L\{f\} = \frac{\partial f}{\partial x} \Big|_{x=x_0}.$$

lineaariset funktionaalit   A functional or operator is *linear* if

$$L\{\alpha f + \beta g\} = \alpha L\{f\} + \beta L\{g\}, \forall \alpha, \beta \in \mathbb{R}.$$

Note that all partial derivatives, as also the *Laplace operator* Δ, are linear.

In physical geodesy, all interesting functionals are functionals of the function $T(\phi, \lambda, R) = T(\phi, \lambda, r)|_{r=R}$, i.e., the disturbing potential on the surface of the Earth sphere. The theory thus uses the spherical approximation[1], and the surface of the sphere, radius R, corresponds to mean sea level. For example, the disturbing potential $T_P \stackrel{\text{def}}{=} T(\phi_P, \lambda_P, R)$ in a point P at sea-level location (ϕ_P, λ_P) is such a functional:

$$T(\cdot, \cdot, R) \mapsto T(\phi, \lambda, R).$$

---

[1]This is not mandatory, but the error of approximation is small.

Even if point P is not at sea level, a suitable functional exists:

$$T(\cdot, \cdot, R) \mapsto T(\phi, \lambda, r).$$

Even if the quantity is not the disturbing potential, but, e.g., the gravity anomaly or the deflection of the plumb line:

$$T(\cdot, \cdot, R) \mapsto \Delta g(\phi, \lambda, r),$$
$$T(\cdot, \cdot, R) \mapsto \xi(\phi, \lambda, r),$$
$$T(\cdot, \cdot, R) \mapsto \eta(\phi, \lambda, r).$$

All these are also *linear* functionals. In fact, if we write

$$T = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=0}^{n} P_{nm}(\sin\phi)\ (a_{nm}\cos m\lambda + b_{nm}\sin m\lambda),$$

then even the spherical-harmonic coefficients a_{nm}, b_{nm} are all linear functionals of the disturbing potential T:

$$T(\cdot, \cdot, R) \longmapsto a_{nm},$$
$$T(\cdot, \cdot, R) \longmapsto b_{nm}.$$

Here, $T(\cdot, \cdot, R)$ is shorthand for the whole function

$$T(\phi, \lambda, R), \quad \phi \in \left[-\frac{\pi}{2}, +\frac{\pi}{2}\right], \lambda \in [0, 2\pi).$$

10.3   Statistics on the Earth's surface

In statistics we define a *stochastic process* as a stochastic quantity, or random variable, the *domain* of which is a function space. In other words, a random määrittelyjoukko variable the realization values of which are *functions*. A stochastic process may be a quantity developing over time, the precise behaviour of which is uncertain, e.g., a satellite orbit. In the same way as for a (real-valued) stochastic quantity $\underline{x}$, we may calculate an expected value or *expectancy* $E\{\underline{x}\}$ odotusarvo and a *variance*

$$C_{xx} = \text{Var}\{\underline{x}\} = E\left\{(\underline{x} - E\{\underline{x}\})^2\right\},$$

we may also do so for a stochastic process. The only difference is, that by doing so we obtain *functions*.

Let, e.g., the stochastic process $\underline{x}(t)$ be a function of time. Then we may compute its variance function as follows:

$$C_{xx}(t) = \mathrm{Var}\{\underline{x}(t)\}.$$

For a stochastic process however, much more can be computed: e.g., the covariance of values of the same function taken at different points in time, the so-called *autocovariance*:

$$A_{xx}(t_1, t_2) = \mathrm{Cov}\{\underline{x}(t_1), \underline{x}(t_2)\} = E\left\{\left(\underline{x}(t_1) - E\{\underline{x}(t_1)\}\right)\left(\underline{x}(t_2) - E\{\underline{x}(t_2)\}\right)\right\}.$$

Similarly if we have two different functions, we may compute between them the so-called *cross covariance*, etc.

The argument of a stochastic process is commonly *time, t*. However in geophysics we study stochastic processes the arguments of which are locations on the Earth's surface, i.e., we talk of processes of the form $\underline{x}(\phi, \lambda)$. The definitions of auto- and cross-covariances work otherwise in the same way, but in case of the Earth we have a special problem. A stochastic quantity is generally defined as a quantity $\underline{x}$, from which *realizations* $x_1, x_2, x_3, \dots$ are obtained, which have certain statistical properties. The classical example is the *dice* throw. A die can be thrown again and again, and one can practice the art of statistics on the results of the throws. Another classic example is *measurement*. Measurement of the same quantity can be repeated, and is repeated, in order to improve precision.

For a stochastic process defined on the Earth's surface, the situation is different.

> We have only one Earth.

For this reason, statistics must be done in a somewhat different fashion.

Given a stochastic process on the surface of the Earth, $\underline{x}(\phi, \lambda)$, we define a quantity similar to the statistical expectancy $E\{\cdot\}$, the *geographic mean*

$$M\{x\} \stackrel{\text{def}}{=} \frac{1}{4\pi}\iint_\sigma x(\phi, \lambda)\,d\sigma = \frac{1}{4\pi}\int_0^{2\pi}\int_{-\pi/2}^{+\pi/2} x(\phi, \lambda)\,\cos\phi\,d\phi\,d\lambda. \qquad (10.1)$$

Here $x(\phi, \lambda)$ is the one and only realization of process $\underline{x}$ that we have available on this Earth.

Clearly this definition makes sense only in the case where the statistical behaviour of the process $\underline{x}(\phi, \lambda)$ is the same everywhere on Earth, independently of the value of (ϕ, λ). This is called the *assumption of homogeneity*. It is   homogeenisuus in fact the assumption that the spherical symmetry of the Earth extends to the statistical behaviour of her gravity field.

Similarly to the statistical variance based on expectancy, we may define the *geographic variance*:

$$C_{xx}(\phi, \lambda) \stackrel{\text{def}}{=} \mathrm{Var}\{\underline{x}(\phi, \lambda)\} \stackrel{\text{def}}{=} M\left\{(x - M\{x\})^2\right\}. \qquad (10.2)$$

The global average of gravity anomalies $\Delta g(\phi, \lambda)$ *vanishes* based on their definition:

$$M\{\Delta g\} = 0.$$

In that case, equation 10.2 is simplified as follows:

$$C_{\Delta g \Delta g}(\phi, \lambda) = \mathrm{Var}\{\underline{\Delta g}(\phi, \lambda)\} = M\{\Delta g^2\} = \frac{1}{4\pi}\iint_\sigma \left(\Delta g(\phi, \lambda)\right)^2 d\sigma.$$

The definition given here of the geographic mean $M\{\cdot\}$ is based on *integration* of the one and only realization over the surface of the Earth. As has been seen, in statistics the mean is defined slightly differently, as the *expectancy* of a stochastic process. For gravity anomalies this means $E\{\underline{\Delta g}\}$, where $\underline{\Delta g}$ is the anomaly *considered as a stochastic process*, i.e., the series of values of Δg that results if we look at an infinitely long series of randomly formed Earths. Not very practical!

If the expectancy of a stochastic process is the same as the mean of one realization computed by integration, we speak of an *ergodic process*. Establishing   ergodisuus empirically that a process is ergodic is in geophysics typically difficult to impossible.

10.4   The covariance function of the gravity field

Defining a *covariance function* between points P and Q is more complicated. Something like equations 10.1, 10.2 cannot be used directly, because *both* Δg_P
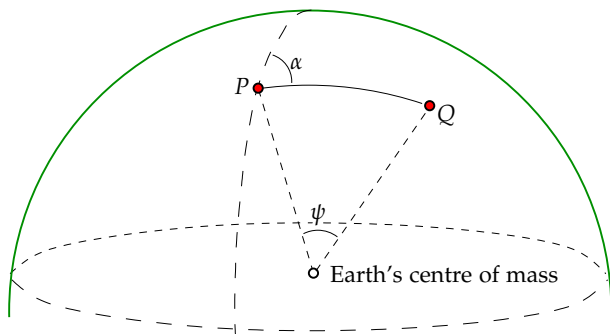
Figure 10.1. Definition of geocentric angular distance and azimuth.

and Δg_Q can move over the whole Earth's surface. We have

$$\Delta g_P = \Delta g(\phi_P, \lambda_P),$$
$$\Delta g_Q = \Delta g(\phi_Q, \lambda_Q).$$

In the following we assume that the covariance to be calculated will only depend on the *relative* location of points P and Q. In a homogeneous gravity field, the covariance function will not depend on the absolute location of the points, but only on the *difference* in location between points P and Q.

Write

$$\phi_Q = \phi_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}),$$
$$\lambda_Q = \lambda_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}).$$

(ϕ_Q, λ_Q) can be computed[2] if we know (ϕ_P, λ_P) and both the *geocentric angular distance* ψ_{PQ} and the *azimuth angle* α_{PQ}. See figure 10.1.

Now we may write

$$\Delta g_Q = \Delta g_Q\Big(\phi_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}), \lambda_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ})\Big) =$$
$$= \Delta g_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}),$$

---

[2]This is called the *geodetic forward problem* on the sphere.

and we may define as the *covariance function*

$$C_{\Delta g \Delta g}(\psi_{PQ}, \alpha_{PQ}) \overset{\text{def}}{=} M\{\Delta g_P(\phi_P, \lambda_P)\Delta g_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ})\} =$$
$$= \frac{1}{4\pi} \iint_\sigma \Delta g_P(\phi_P, \lambda_P)\Delta g_Q(\phi_P, \lambda_P, \psi_{PQ}, \alpha_{PQ}) \, d\sigma_P.$$

Also here, M is a geographic-mean operator. First we fix point Q in relation to point P: both azimuth α_{PQ} and distance ψ_{PQ} are held fixed. The point P — and with it, point Q — is moved over the whole Earth's surface. We compute the corresponding integral over the unit sphere σ_P, and divide by 4π:

$$C_{\Delta g \Delta g}(\psi_{PQ}, \alpha_{PQ}) = M\{\Delta g_P \Delta g_{Q(P)}\} = \frac{1}{4\pi} \iint_\sigma \Delta g_P \Delta g_{Q(P)} \, d\sigma_P =$$
$$= \frac{1}{4\pi} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} \Delta g_P \Delta g_{Q(P)} \, d\lambda_P \cos\phi_P \, d\phi_P,$$

in which was used $d\sigma = \cos\phi \, d\lambda \, d\phi$, with $\cos\phi$ *Jacobi's determinant* for the co-ordinates (ϕ, λ) on the unit sphere.

In addition to the assumption of homogeneity, we may make still the *assumption of isotropy*: the covariance function — or more generally, the statistical behaviour of the gravity field — does *not* depend on the relative direction or *azimuth* α_{PQ} of point pair (P, Q), but only on the *angular distance* ψ_{PQ} between them. (This too is, like homogeneity, one of the forms in which the Earth's spherical symmetry is expressed.) In this case we may compute the geographic mean in a slightly different way, by also averaging over all azimuth angles $\alpha_{PQ} \in [0, 2\pi)$:

$$C_{\Delta g \Delta g}(\psi_{PQ}) = M'\{\Delta g_P \Delta g_{Q(P)}\} = \frac{1}{2\pi} \int_0^{2\pi} M\{\Delta g_P \Delta g_{Q(P)}\} \, d\alpha_{PQ} =$$
$$= \frac{1}{8\pi^2} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} \Delta g_P \Delta g_{Q(P)} \, d\lambda_P \cos\phi_P \, d\phi_P \, d\alpha_{PQ}. \quad (10.3)$$

Remark: The true gravity field of the Earth isn't terribly homogeneous or isotropic, but in spite of this, both hypotheses are widely used.

10.5 Least-squares collocation

10.5.1 Stochastic processes in one dimension

Collocation is a statistical estimation technique used to estimate the values of pienimmän neliösumman kollokaatio

Statistical methods

Least-squares collocation

a *stochastic process*, and calculate the uncertainties (e.g., mean errors) of the estimates.

Let $\underline{s}(t)$ be a stochastic process, the autocovariance function of which is $C(t_i, t_j)$. Let the process furthermore be *stationary*, i.e., for any two moments in time t_i, t_j we have $C(t_i, t_j) = C(t_j - t_i) = C(\Delta t)$. The argument t is generally time, but could be any parameter, e.g., distance of a journey.

Of this process, we have *observations* made at times $t_1, t_2, \ldots, t_N$, when the corresponding process values for those times are $\underline{s}(t_1), \underline{s}(t_2), \ldots, \underline{s}(t_N)$. Let us assume, for the moment, that these values are *error free* observations. Then the observations are function values of process $\underline{s}$, stochastic quantities, the variance matrix of which we may write as follows:

$$\text{Var}\{\underline{s}_i\} = \begin{bmatrix} C(t_1, t_1) & C(t_2, t_1) & \cdots & C(t_1, t_N) \\ C(t_1, t_2) & C(t_2, t_2) & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ C(t_1, t_N) & C(t_2, t_N) & \cdots & C(t_N, t_N) \end{bmatrix}.$$

We call this autocovariance matrix also the *signal variance matrix* of $\underline{s}$. We use for this the symbol C_{ij}, both for one element $C_{ij} = C(t_i, t_j)$ of the matrix, and for the whole matrix, $C_{ij} = [C(t_i, t_j), i, j = 1, \ldots, N]$. The symbol $\underline{s}_i$ again denotes a vector $[\underline{s}(t_i), i = 1, \ldots, N]$ consisting of process values — or one of its elements $\underline{s}(t_i)$.

Note that, if the function $C(t_i, t_j)$, or $C(\Delta t)$, is known, then the whole matrix and all of its elements can be calculated provided all argument values (observation times) t_i are known.

Let the shape of the problem now be, that one should *estimate,* i.e., *predict,* the value of process $\underline{s}$ at the moment in time T, i.e., $\underline{s}(T)$, based on our knowledge of the above described *observations* $\underline{s}(t_i), i = 1, \ldots, N$.

In the same way as we calculated above the covariances between $\underline{s}(t_i)$ and $\underline{s}(t_j)$ (elements of the signal variance matrix C_{ij}), we also compute the covariances between $\underline{s}(T)$ and all $\underline{s}(t_i), i = 1, \ldots, N$. We obtain

$$\text{Cov}\{\underline{s}(T), \underline{s}(t_i)\} = \begin{bmatrix} C(T, t_1) & C(T, t_2) & \cdots & C(T, t_N) \end{bmatrix}.$$

For this we may again use the notation C_{Tj}. It is assumed here, that there is only one point in time T for which estimation is done. Generalization to the case where there are several $T_p, p = 1, \ldots, M$, is straightforward. In that case, the signal covariance matrix will be of size $M \times N$:

$$\text{Cov}\{\underline{s}(T_p), \underline{s}(t_i)\} = \begin{bmatrix} C(T_1, t_1) & C(T_1, t_2) & \cdots & C(T_1, t_N) \\ C(T_2, t_1) & C(T_2, t_2) & \cdots & C(T_2, t_N) \\ \vdots & \vdots & \vdots & \vdots \\ C(T_M, t_1) & C(T_M, t_2) & \cdots & C(T_M, t_N) \end{bmatrix}.$$

For this we may use the more general notation C_{pj}.

10.5.2 Signal and noise

The process $\underline{s}(t)$ is called the *signal*. It is a *physical phenomenon* that we are *interested* in. There exist also physical phenomena that are otherwise similar, but that we are *not* interested in: to the contrary, we wish to *remove* their influence. Such stochastic processes are called *noise*.

When we make an observation, the purpose of which is to obtain a value for the quantity $\underline{s}(t_i)$, we obtain *in realit*y a value that is not absolutely precise. The real observation thus is

$$\underline{\ell}_i = \underline{s}(t_i) + \underline{n}_i. \tag{10.4}$$

Here, $\underline{n}_i$ is a stochastic quantity: *observational error* or *noise*. Let its variance — or more precisely, the joint noise variance matrix of multiple observations — be D_{ij}. This is a very similar matrix to the above C_{ij}, and also symmetric and positive definite. The only difference is that D_{ij} designates *noise*, which we are *not* interested in. Often it may be assumed, that the errors $\underline{n}_i, \underline{n}_j$ of two different observations $\underline{\ell}_i, \underline{\ell}_j$ do not correlate, in which case D_{ij} is a diagonal matrix.

10.5.3 Estimator and variance of prediction

Now we construct an *estimator*

$$\widehat{s}(T_p) \stackrel{\text{def}}{=} \sum_i \Lambda_{pi} \underline{\ell}_i,$$

a linear combination of the observations at our disposal $\underline{\ell}_i$. The purpose in life of this estimator is to get as closely as possible to $\underline{s}(T_p)$. So, the quantity to be minimized is the difference

$$\widehat{s}(T_p) - \underline{s}(T_p) = \Lambda_{pi}\underline{\ell}_i - \underline{s}(T_p) = \Lambda_{pi}\big(\underline{s}(t_i) + \underline{n}_i\big) - \underline{s}(T_p).$$

Here, for the sake of writing convenience, we left the summation sign $\sum$ off (Einstein summation convention): We always sum over adjacent, identical indices, in this case i.

Study the *variance* of this difference, the so-called variance of prediction:

$$\Sigma_{pp} \overset{\text{def}}{=} \text{Var}\big\{\widehat{s}(T_p) - \underline{s}(T_p)\big\}.$$

varianssien kasautuminen We exploit *propagation of variances*, the notations introduced above, and our knowledge that surely there is no physical relationship, or *correlation*, between observation process noise $\underline{n}$ and signal $\underline{s}$:

$$\text{Cov}\Big\{\big(\underline{s}(t_i) + \underline{n}_i\big), \big(\underline{s}(t_j) + \underline{n}_j\big)\Big\} = \text{Cov}\big\{\underline{s}(t_i), \underline{s}(t_j)\big\} + \text{Cov}\big\{\underline{n}_i, \underline{n}_j\big\} = C_{ij} + D_{ij},$$

and

$$
\begin{aligned}
\Sigma_{pq} &= \text{Cov}\Big\{\big(\widehat{s}(T_p) - \underline{s}(T_p)\big), \big(\widehat{s}(T_q) - \underline{s}(T_q)\big)\Big\} = \\
&= \Lambda_{pi}\,\text{Cov}\Big\{\big(\underline{s}(t_i) + \underline{n}_i\big), \big(\underline{s}(t_j) + \underline{n}_j\big)\Big\}\Lambda_{jq}^{\mathsf{T}} + \text{Cov}\big\{\underline{s}(T_p), \underline{s}(T_q)\big\} - \\
&\quad - \Lambda_{pi}\,\text{Cov}\big\{\underline{s}(t_i), \underline{s}(T_q)\big\} - \text{Cov}\big\{\underline{s}(T_p), \underline{s}(t_j)\big\}\Lambda_{jq}^{\mathsf{T}} = \\
&= \Lambda_{pi}\big(C_{ij} + D_{ij}\big)\Lambda_{jq}^{\mathsf{T}} + C_{pq} - \Lambda_{pi}C_{iq} - C_{pj}\Lambda_{jq}^{\mathsf{T}}.
\end{aligned}
\tag{10.5}
$$

The variances, or diagonal elements, Σ_{pp} of the matrix are now obtained by setting $q = p$.

10.5.4 Showing optimality

Here we show that the optimal estimator is indeed the one producing the minimum possible variances.

Choose

$$\Lambda_{pj} \overset{\text{def}}{=} C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}.$$

Then, from equation 10.5, and exploiting the symmetry of the C and D matrices, we obtain

$$
\begin{aligned}
\Sigma_{pp} &= C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp} + C_{pp} - \\
&\quad - C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp} - C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp} = \\
&= C_{pp} - C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp}.
\end{aligned}
\tag{10.6}
$$

Let us study next the *alternative choice*

$$\Lambda_{pj} = C_{pi}\big(C_{ij} + D_{ij}\big)^{-1} + \delta\Lambda_{pj}.$$

In this case we obtain

$$
\begin{aligned}
\Sigma'_{pp} &= C_{pp} - C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp} + \\
&\quad + \delta\Lambda_{pi}\Big[\big(C_{ij} + D_{ij}\big)\Lambda_{jp}^{\mathsf{T}}\Big] + \big[\Lambda_{pi}\big(C_{ij} + D_{ij}\big)\big]\delta\Lambda_{jp}^{\mathsf{T}} + \\
&\quad + \delta\Lambda_{pi}\big(C_{ij} + D_{ij}\big)\delta\Lambda_{jp}^{\mathsf{T}} - \delta\Lambda_{pi}C_{ip} - C_{pj}\delta\Lambda_{jp}^{\mathsf{T}} = \\
&= C_{pp} - C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp} + \delta\Lambda_{pi}C_{ip} + C_{pj}\delta\Lambda_{jp}^{\mathsf{T}} - \\
&\quad - \delta\Lambda_{pi}C_{ip} - C_{pj}\delta\Lambda_{jp}^{\mathsf{T}} + \delta\Lambda_{pi}\big(C_{ij} + D_{ij}\big)\delta\Lambda_{jp}^{\mathsf{T}} = \\
&= C_{pp} - C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}C_{jp} + \delta\Lambda_{pi}\big(C_{ij} + D_{ij}\big)\delta\Lambda_{jp}^{\mathsf{T}}.
\end{aligned}
$$

Here, the last term — the only difference with result 10.6 — is positive, because the matrices C_{ij} and D_{ij} are positive definite: $\Sigma'_{pp} > \Sigma_{pp}$, except when $\delta\Lambda_{pi} = 0$. In other words, the solution given above,

$$\Lambda_{pj} = C_{pi}\big(C_{ij} + D_{ij}\big)^{-1} \implies \widehat{s}(T_p) = C_{pi}\big(C_{ij} + D_{ij}\big)^{-1}\underline{\ell}_j,$$

is *optimal* in the sense of least squares — more precisely, in the sense of minimizing the variance of prediction Σ_{pp}.

10.5.5 The covariance function of gravity anomalies

Least-squares collocation is used much to optimally estimate gravity values and other functionals of the gravity field on the Earth's surface.

If we have two points, P and Q, with measured gravity anomalies $\underline{\Delta g}_P = \underline{\Delta g}(\phi_P, \lambda_P)$ and $\underline{\Delta g}_Q = \underline{\Delta g}(\phi_Q, \lambda_Q)$, we would like to have the covariance between these two anomalies,

$$\text{Cov}\big\{\underline{\Delta g}_P, \underline{\Delta g}_Q\big\}.$$

As argued in section 10.4, we can only empirically derive such a covariance by looking at all point pairs (P, Q) that are in the same relative position around the globe, and averaging over them using the M or M' operator.

Normally the covariance is assumed to depend *only* on the geocentric angular distance ψ between points P, Q. Then, we speak of an *isotropic process* $\underline{\Delta g}(\phi, \lambda)$. Then also, the covariance will be

$$\mathrm{Cov}\{\underline{\Delta g}_P, \underline{\Delta g}_Q\} = M'\{\Delta g_P \Delta g_{Q(P)}\} = C(\psi_{PQ}).$$

A popular covariance function for gravity anomalies is Hirvonen[3]'s equation:

$$C(\psi) = \frac{C_0}{1 + \left(\frac{\psi}{\psi_0}\right)^2},\tag{10.7}$$

where $C_0 = C(0)$ and ψ_0 are parameters describing the behaviour of the gravity field. C_0 is called the *signal variance*, ψ_0 the *correlation length*. ψ_0 gives the distance at which the correlation between the gravity anomalies in two points is still 50%.

In local applications, instead of the angular distance ψ one uses the metric distance

$$s = \psi R,$$

where R is the mean Earth radius. Then

$$C(s) = \frac{C_0}{1 + \left(\frac{s}{d}\right)^2}.$$

This equation was derived from gravimetric data for Ohio state, USA, but it has broader validity. $C(0) = C_0$, the signal variance when $s = 0$. Also the variable d is called the correlation length. It is the distance d for which $C(d) = \frac{1}{2}C_0$, as seen from the equation.

The quantity C_0 varies considerably between areas, from hundreds to thousands of mGal^2, and is largest in mountainous areas. The quantity d is generally order of magnitude tens of km.

[3] Reino Antero Hirvonen (1908–1989) was a Finnish physical and mathematical geodesist.
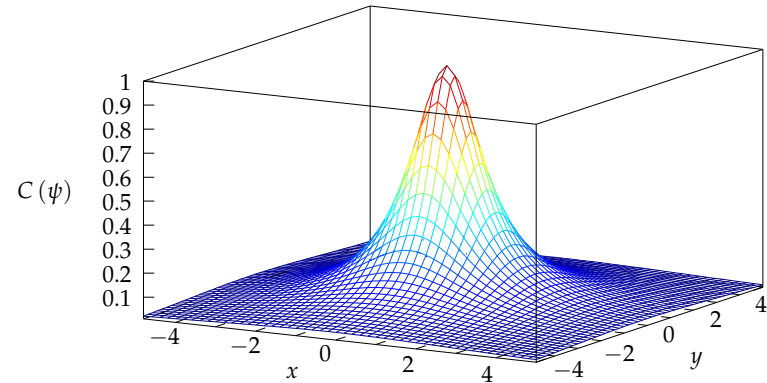
Figure 10.2. Hirvonen's covariance function in two dimensions. Assumed is $C_0 = d = 1$.

Warning: The Hirvonen covariance function is meant for use with (free-air) gravity anomalies, i.e., quantities obtained by subtracting normal gravity from the measured gravity. Nowadays anomalies are often obtained by subtracting from the observations a high-degree "normal field", i.e., a spherical-harmonic expansion. Then one uses Hirvonen's formula at one's own risk!

Alternative functions that are also often used in local applications are the covariance functions of first and second order autoregressive processes, or AR(1) or AR(2) processes:

$$C(\psi) = C_0 e^{-\frac{\psi}{\psi_0}} \quad \text{or} \quad C(\psi) = C_0 e^{-\left(\frac{\psi}{\psi_0}\right)^2}.$$

An AR(1) process is also called a Gauss–Markov process.

10.5.6   Least-squares collocation for gravity anomalies

If given are N points $P_i, i = 1, \ldots, N$, where were measured gravity values — more precisely, anomalies — $\underline{\Delta g}_i = \underline{\Delta g}(\phi_i, \lambda_i)$, we may, like above, construct
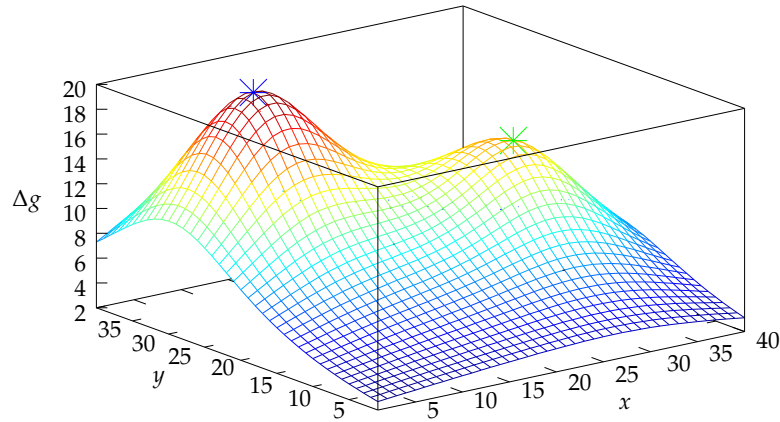
Figure 10.3. An example of least-squares collocation. Here are given two data points (stars); the surface plotted gives the estimated value $\widehat{\Delta g}_P$ for each point P in the area. Here we use least-squares collocation for inter- and extrapolating gravimetric data.

a *signal variance matrix*

$$
\mathrm{Var}\{\underline{\Delta g}_i\} =
\begin{bmatrix}
C_0 & C(\psi_{21}) & \cdots & C(\psi_{N1}) \\
C(\psi_{12}) & C_0 & \cdots & C(\psi_{N2}) \\
\vdots & \vdots & & \vdots \\
C(\psi_{1N}) & C(\psi_{2N}) & \cdots & C_0
\end{bmatrix} =
$$

$$
=
\begin{bmatrix}
C_0 & C_{21} & \cdots & C_{N1} \\
C_{12} & C_0 & \cdots & C_{N2} \\
\vdots & \vdots & d & \vdots \\
C_{1N} & C_{2N} & \cdots & C_0
\end{bmatrix} \overset{\text{def}}{=} C_{ij},
$$

where all elements $C(\psi_{ij})$ are calculated using the covariance function 10.7 given above.

If we also compute for the point P in which gravity is unknown:

$$
\mathrm{Cov}\{\underline{\Delta g}_P, \underline{\Delta g}_i\} = \begin{bmatrix} C(\psi_{P1}) & C(\psi_{P2}) & \cdots & C(\psi_{PN}) \end{bmatrix} \overset{\text{def}}{=} C_{Pi},
$$
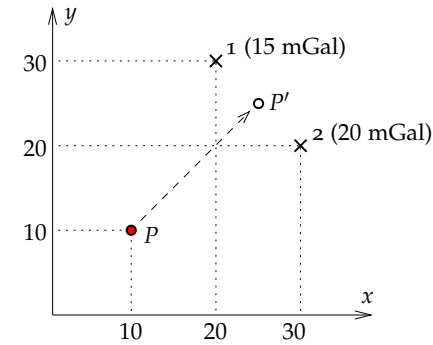
Figure 10.4. Collocation example.

we obtain, in the same way as before, for the *least-squares collocation* solution

$$
\widehat{\Delta g}_P = C_{Pi}\left(C_{ij} + D_{ij}\right)^{-1}\ell_j \approx C_{Pi}C_{ij}^{-1}\ell_j,
$$

where the $\underline{\ell}_j = \underline{\Delta g}_j + \underline{n}_j$ are gravity anomaly observations made in points $j = 1, \ldots, N$. The matrix D_{ij} (which we leave out of consideration here) again describes the random observation error, observation uncertainty, or *noise* $\underline{n}_i$ associated with making those observations. Usually D_{ij} is a diagonal matrix, i.e., the observations are statistically independent and don't correlate with each other.

We may also compute a precision assessment of this solution, i.e., the *variance of prediction*, equation 10.10:

$$
\Sigma_{PQ} \approx C_{PQ} - C_{Pi}C_{ij}^{-1}C_{jQ}.
$$

In the case of one unknown prediction point P, $Q = P$ and

$$
\Sigma_{PP} = C_0 - C_{Pi}C_{ij}^{-1}C_{jP}.
$$

Its square root

$$
\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}}
$$

is the *mean error* of estimator $\widehat{\Delta g}_P$.

10.5.7 Calculation example

See figure 10.4. *Given* two points where gravity has been measured and gravity anomalies calculated: $\underline{\Delta g}_1 = 15\,\text{mGal}$, $\underline{\Delta g}_2 = 20\,\text{mGal}$. The co-ordinates in the x and y directions are in kilometres. Assumed is that between the gravity anomalies of different points, Hirvonen's covariance function,

$$C(s) = \frac{C_0}{1 + \left(\frac{s}{d}\right)^2},$$

applies, in which $d = 20\,\text{km}$ and $C_0 = \pm 1000\,\text{mGal}^2$. Additionally it is assumed that the gravity measurements done (including height determination of the gravity points!) were *errorless*. So, $D_{ij} = 0, i, j = 1, 2$.

Calculate an estimate of the gravity anomaly $\widehat{\Delta g}_P$ at point P and its mean error $\sigma_{PP} = \sqrt{\Sigma_{PP}}$.

Calculate first the distances s and the corresponding covariances C.

$$s_{12}^2 = \left((30 - 20)^2 + (20 - 30)^2\right) \text{km}^2 = 200\,\text{km}^2,$$

$$C_{12} = C_{21} = \frac{1000\,\text{mGal}^2}{1 + \frac{200}{400}} = 666.66\ldots\,\text{mGal}^2,$$

$$s_{1P}^2 = \left((30 - 10)^2 + (20 - 10)^2\right) \text{km}^2 = 500\,\text{km}^2,$$

$$C_{1P} = \frac{1000\,\text{mGal}^2}{1 + \frac{500}{400}} = 444.44\ldots\,\text{mGal}^2,$$

$$s_{2P}^2 = \left((20 - 10)^2 + (30 - 10)^2\right) \text{km}^2 = 500\,\text{km}^2,$$

$$C_{2P} = \frac{1000\,\text{mGal}^2}{1 + \frac{500}{400}} = 444.44\ldots\,\text{mGal}^2.$$

From this follows

$$C_{ij} + D_{ij} = C_{ij} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} 1000 & 666.66 \\ 666.66 & 1000 \end{bmatrix} \text{mGal}^2,$$

and its inverse matrix

$$\left(C_{ij} + D_{ij}\right)^{-1} = \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \text{mGal}^{-2}.$$

We also have

$$C_{Pi} = \begin{bmatrix} C_{P1} & C_{P2} \end{bmatrix} = \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \text{mGal}^2.$$

As the vector of observations is

$$\underline{\Delta g}_j = \begin{bmatrix} \underline{\Delta g}_1 \\ \underline{\Delta g}_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} \text{mGal},$$

we get the result

$$\widehat{\Delta g}_P = \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \begin{bmatrix} 15 \\ 20 \end{bmatrix} \text{mGal} = 9.333\,\text{mGal}.$$

Precision, i.e., the *variance of prediction*, equation 10.10:

$$\Sigma_{PP} = C_{PP} - C_{Pi}\left(C_{ij} + D_{ij}\right)^{-1} C_{jP} =$$

$$= C_0 - \begin{bmatrix} 444.44 & 444.44 \end{bmatrix} \begin{bmatrix} 0.0018 & -0.0012 \\ -0.0012 & 0.0018 \end{bmatrix} \begin{bmatrix} 444.44 \\ 444.44 \end{bmatrix} \text{mGal}^2 =$$

$$= 762.96\,\text{mGal}^2,$$

i.e.,

$$\sigma_{\Delta g_P} = \sqrt{\Sigma_{PP}} = \pm 27.622\,\text{mGal}.$$

Summarizing the result:

$$\widehat{\Delta g}_P = 9.333 \pm 27.622\,\text{mGal}.$$

Observe that the gravity anomaly estimate found is much smaller than its own uncertainty, and thus *does not differ significantly from zero*. In fact, not using the observational data at all would leave us with the *a priori* estimate

$$\widehat{\Delta g}_P = 0 \pm \sqrt{1000}\,\text{mGal} = 0 \pm 31.623\,\text{mGal},$$

almost as good.

If, instead, we would choose to locate point P' in between points 1 and 2, at location $(25\,\text{km}, 25\,\text{km})$, then $C_{P'1} = C_{P'2} = \frac{1000\,\text{mGal}^2}{1 + \frac{50}{400}} = 888.89\,\text{mGal}^2$ and

$\widehat{\Delta g}_{P'} = 18.667 \pm 7.201\,\text{mGal}$, which *is* clearly better than the *a priori* estimate of zero.

And *if* we had chosen instead the Gauss–Markov covariance function

$$C = C_0 e^{s/d}$$

we would have obtained the results $\widehat{\Delta g}_P = 7.663 \pm 29.272 \, \text{mGal}$ for the original point location, and $\widehat{\Delta g}_{P'} = 16.460 \pm 18.426 \, \text{mGal}$ for the shifted point location.

10.5.8   Theory of least-squares collocation

Above we presented one popular application of least-squares collocation. Here we look at the method more generally. The basic equation is

$$\widehat{f} = C_{fg} \left[C_{gg} + D_{gg} \right]^{-1} \left[g + n \right]. \tag{10.8}$$

The vector $\underline{g}$ contains *observed quantities* g_i, the vector $\underline{n}$ contains the observational *noise*, and $\widehat{f}$ is a vector of quantities $\widehat{f}_i$ to be *predicted*. The hat is a commonly used symbol for an estimator.

Both vectors $\underline{g}$ and $\widehat{f}$ can, e.g., be gravity anomalies, in which case we have *homogeneous prediction*, a type of interpolation or extrapolation. More generally $\widehat{f}$ and $\underline{g}$ are of different type, e.g., $\widehat{f}$ consists of geoid heights N_i and $\underline{g}$ of gravity anomalies Δg_i. In the latter case, the Stokes equation is "covertly" along in the structure of the C matrices.

These matrices are built from covariance functions. Their elements can be expressed as follows:[4]

$$\left[C_{fg} \right]_{ij} = M\{f_i g_j\},$$
$$\left[C_{gg} \right]_{jk} = M\{g_j g_k\},$$
$$\left[D_{gg} \right]_{jk} = E\{\underline{n}_j \underline{n}_k\},$$

---

[4]Note that here we use the *geographic mean* $M\{\cdot\}$ for evaluating the signal covariances. In doing so, f and g are no longer considered stochastic. It is however assumed that their global geographic mean vanishes: $M\{f\} = M\{g\} = 0$.

where $\underline{n}_i$, an element of vector $\underline{n}$, represents the *uncertainty* of the observation process appearing in the observation equation 10.4:

$$\ell_i = g_i + \underline{n}_i \longleftrightarrow \underline{\ell} = g + \underline{n}.$$

$\underline{\ell}$ is the vector of the observation values themselves, including observation uncertainty $\underline{n}$.

The D is the variance matrix of observational uncertainty, the *noise variance matrix* describing a property of the observational process, not of the gravity field. While the values of $M\{\Delta g_i \Delta g_j\}$ can be as large as $1200 \, \text{mGal}^2$, the values of $E\{\underline{n}_i \underline{n}_j\}$ can be much smaller, depending on the measurement technique used, e.g., as small as $0.01 \, \text{mGal}^2$.

Not however in the case of block averages — e.g., averages over blocks of size $1° \times 1°$, computed from scattered measurements — which often are very imprecise.

The great advantage of least-squares collocation is its *flexibility*. Different observation types may be handled with a single unified theory and method, the locations of observation points (or blocks) are totally free, and the result is obtained directly as freely choosable quantities in locations where one wants them.

10.6   Prediction of gravity anomalies

If the quantity to be calculated or estimated, $\widehat{f}$, is of the same type as the observed quantity, $\underline{g}$, we often speak of (homogeneous) *prediction*. For example, the prediction equation for gravity anomalies already presented in subsection 10.5.6 is obtained from equation 10.8 by substitution:

$$\widehat{\Delta g}_P = C_{Pi} \left(C_{ij} + D_{ij} \right)^{-1} \ell_j. \tag{10.9}$$

Here are several points j where gravity is given: let us say, N observations $\ell_j = \Delta g_j + \underline{n}_j$, $j = 1, \ldots, N$. The number of points to be predicted may be one, P, or also many. The matrices C_{ij} and D_{ij} are square, and the inverse of their sum exists. C_{Pi} is a rectangular matrix. If there is only one point P, it is a size $1 \times N$ row matrix.

The *prediction error* is now the difference quantity[5] $\widehat{\Delta g}_P - \underline{\Delta g}_P$, and its variance ("variance of prediction") is

ennustusvarianssi

$$\Sigma_{PP} \overset{\text{def}}{=} \text{Var}\{\widehat{\Delta g}_P - \underline{\Delta g}_P\} =$$
$$= \text{Var}\{\widehat{\Delta g}_P\} + \text{Var}\{\underline{\Delta g}_P\} - \text{Cov}\{\widehat{\Delta g}_P, \underline{\Delta g}_P\} - \text{Cov}\{\underline{\Delta g}_P, \widehat{\Delta g}_P\}.$$

Here (propagation of variances applied to equation 10.9):

$$\text{Var}\{\widehat{\Delta g}_P\} = C_{Pi}\left(C_{ij} + D_{ij}\right)^{-1}\left(C_{jk} + D_{jk}\right)\left(C_{k\ell} + D_{k\ell}\right)^{-1}C_{\ell P} =$$
$$= C_{Pi}\left(C_{ij} + D_{ij}\right)^{-1}C_{\ell P}$$

and

$$\text{Cov}\{\widehat{\Delta g}_P, \underline{\Delta g}_P\} = \text{Cov}\{\underline{\Delta g}_P, \widehat{\Delta g}_P\} = C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP}.$$

Here, C_{iP}^{T} (or C_{jP}^{T}, or $C_{\ell P}^{\mathrm{T}}$) is the transpose of C_{Pi}. The matrix $\left(C_{ij} + D_{ij}\right)^{-1}$ is symmetric and its own transpose.

The end result is (remember that the signal variance $\text{Var}\{\underline{\Delta g}_P\} = C_{PP}$):

$$\Sigma_{PP} = C_{PP} + C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP}-$$
$$- C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP} - C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP} =$$
$$= C_{PP} - C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP}.$$

In case $D_{ij} \ll C_{ij}$, we obtain a simpler, often used result:

$$\Sigma_{PP} \approx C_{PP} - C_{Pi}C_{ij}^{-1}C_{jP}. \tag{10.10}$$

Borderline cases:

1. Point P is far from all points i. Then $C_{Pi} \approx 0$ and $\Sigma_{PP} \approx C_{PP}$, i.e., prediction is impossible in practice and the prediction equation 10.9 will yield the value zero. The *mean error* of prediction $\sqrt{\Sigma_{PP}}$ is the same as the variability $\sqrt{C_{PP}}$ of the gravity anomaly signal.

----

[5]Note that here, $\underline{\Delta g}_P$ is the *true value* of the gravity anomaly at point P, which we don't know empirically. The *measured value* is $\underline{\ell}_P = \underline{\Delta g}_P + \underline{n}_P$, in which $\underline{n}_P$ is the random error or "noise" of the gravimetric observation.

2. Point P is identical with one of the points i. Then, if we use only that point i, we obtain

$$\Sigma_{PP} = C_{PP} - C_{PP}C_{PP}^{-1}C_{PP}^{\mathrm{T}} = 0,$$

no prediction error whatsoever — as the value at the prediction point was already known!.

However, if $D_{PP} \neq 0$ (but small), the result is $\Sigma_{PP} \approx D_{PP}$. Show.

——

10.7   Covariance function and degree variances

10.7.1   The covariance function of the disturbing potential

In theoretical work we use, instead of gravity anomalies, rather the covariance function of the *disturbing potential T* on the Earth's surface:

$$K(P, Q) = M\{T_P T_{Q(P)}\}.$$

We write this in the following form using the definition of $M'\{\cdot\}$, equation 10.3:

$$K(\psi_{PQ}) = M'\{T_P T_{Q(P)}\} =$$
$$= \frac{1}{8\pi^2}\int_0^{2\pi}\int_{-\pi/2}^{+\pi/2}\int_0^{2\pi} T_P T_{Q(P)}\, d\lambda_P \cos\phi_P\, d\phi_P\, d\alpha_{PQ}. \tag{10.11}$$

Here it is assumed that the disturbing potential is *isotropic*: K does not depend on α but only on ψ.

We choose on the unit sphere a co-ordinate system where point P is a "pole". In this system, the parameters α_{PQ} and ψ_{PQ} are the spherical co-ordinates of point Q. The covariance function is expanded into the following sum:

$$K(\psi) = \sum_{n=2}^{\infty}\sum_{m=-n}^{n} k_{nm}Y_{nm}(\alpha, \psi)$$

with Y_{nm} defined as in equation 3.9.

Based on isotropy, all coefficients vanish[6] for which $m \neq 0$:

$$K(\psi) = \sum_{n=2}^{\infty} k_{n0} Y_{n0}(\psi) \stackrel{\text{def}}{=} \sum_{n=2}^{\infty} k_n P_n(\cos \psi).$$

astevarianssit   The coefficients k_n are called the *degree variances* (of the disturbing potential). For isotropic covariance functions $K(\psi)$ the information content of the degree variances $k_n, n = 2, 3, \ldots$ is the same as that of the function itself, and is in fact its *spectral representation*.

10.7.2   Degree variances and spherical-harmonic coefficients

We can in a simple way specialize the degree constituent equation 3.7:

$$f_n = \frac{2n+1}{4\pi} \iint_\sigma f(\psi) \, P_n(\cos \psi) \, d\sigma = \frac{2n+1}{2} \int_0^\pi f(\psi) \, P_n(\cos \psi) \, \sin \psi \, d\psi$$

if the expansion of function f is

$$f(\psi) = \sum_{n=2}^{\infty} f_n P_n(\cos \psi).$$

Comparison with the previous yields

$$k_n = \frac{2n+1}{2} \int_0^\pi K(\psi) \, P_n(\cos \psi) \, \sin \psi \, d\psi,$$

i.e., if $K(\psi)$ is given, we can calculate all k_n.

Substituting $K(\psi_{PQ})$ from equation 10.11 yields

$$k_n = \frac{2n+1}{16\pi^2} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} T_P \left(\int_0^\pi \int_0^{2\pi} T_{Q(P)} \, d\alpha_{PQ} P_n(\cos \psi) \, \sin \psi_{PQ} \, d\psi_{PQ} \right) d\lambda_P \cos \phi_P \, d\phi_P.$$

Here we have already interchanged the order of the integrals, as is allowed, and moved T_P to another place.

---

[6]because

$$Y_{nm}(\alpha, \psi) = \begin{cases} P_{nm}(\cos \psi) \cos m\alpha & \text{if } m \geq 0, \\ P_{n|m|}(\cos \psi) \sin |m| \, \alpha & \text{if } m < 0, \end{cases}$$

an expression that can only be independent of α if $m = 0$.

The expression inside the large parentheses is a surface integral over the unit sphere

$$\int_0^\pi \int_0^{2\pi} T_{Q(P)} P_n(\cos \psi_{PQ}) \, d\alpha_{PQ} \sin \psi_{PQ} \, d\psi_{PQ} =$$
$$= \iint_\sigma T_Q P_n(\cos \psi_{PQ}) \, d\sigma_Q \stackrel{\text{def}}{=} \frac{4\pi}{2n+1} T_n(P),$$

where T_n is the constituent of T for the harmonic degree number n, compare the degree constituent equation 3.7. Substitution yields

$$k_n = \frac{1}{4\pi} \int_{-\pi/2}^{+\pi/2} \int_0^{2\pi} T T_n \cos \phi \, d\lambda \, d\phi =$$
$$= \frac{1}{4\pi} \iint_\sigma T T_n \, d\sigma = M\{T T_n\} = \frac{1}{4\pi} \iint_\sigma T_n^2 d\sigma = M\{T_n^2\},$$

according to the definition of operator M, and considering the orthogonality of the functions T_n.

If we now write, with the familiar definitions,

$$T(\phi, \lambda) = \sum_{n=2}^{\infty} T_n(\phi, \lambda) =$$
$$= \sum_{n=2}^{\infty} \sum_{m=0}^{n} \left(\bar{a}_{nm} \overline{P}_{nm}(\sin \phi) \cos m\lambda + \bar{b}_{nm} \overline{P}_{nm}(\sin \phi) \sin m\lambda \right) =$$
$$= \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \bar{a}_{nm} \overline{Y}_{nm}(\phi, \lambda),$$

we obtain

$$K(\psi) = \sum_{n=2}^{\infty} k_n P_n(\cos \psi) = \sum_{n=2}^{\infty} \left(\frac{1}{4\pi} \iint_\sigma T_n^2 d\sigma \right) \cdot P_n(\cos \psi) =$$
$$= \sum_{n=2}^{\infty} \left(\sum_{m=0}^{n} \left(\bar{a}_{nm}^2 + \bar{b}_{nm}^2 \right) \right) P_n(\cos \psi) = \sum_{n=2}^{\infty} \left(\sum_{m=-n}^{n} \bar{a}_{nm}^2 \right) P_n(\cos \psi).$$

Here, we have exploited the *orthonormality* of the fully normalized basis functions

$$\overline{Y}_{nm} = \overline{P}_{n|m|}(\sin \phi) \begin{cases} \cos m\lambda & \text{if } m \geq 0 \\ \sin |m| \, \lambda & \text{if } m < 0 \end{cases}$$

on the surface of unit sphere σ. One sees from the equation, that

$$k_n = \sum_{m=0}^{n} \overline{a}_{nm}^2 + \overline{b}_{nm}^2 = \sum_{m=-n}^{n} \overline{a}_{nm}^2,$$

i.e.,

> *The degree variances k_n of the disturbing potential can be calculated directly from the spherical-harmonic coefficients.*

The literature offers many alternative notations for the degree variances, like

$$k_n \overset{\text{def}}{=} \sigma_n^2 \overset{\text{def}}{=} \sigma_i^{TT}.$$

10.8   Propagation of covariances

The covariance function K derived above can be used to also derive the covariance functions of other quantities. This works in principle for quantities that can be expressed as linear functionals of the disturbing potential $T(\cdot, \cdot, R)$ on the surface of the Earth sphere, as explained in section 10.2.

10.8.1   Example: *upward continuation of the potential*

ylöspäin
jatkaminen

Let us write the disturbing potential in space $T(\phi, \lambda, r)$ as a functional of the surface disturbing potential. $T(\phi, \lambda, R) = T(\cdot, \cdot, R)$ We know, with the definition of T_n, with equation 3.7, being

$$T(\phi, \lambda, R) \overset{\text{def}}{=} \sum_{n=2}^{\infty} T_n(\phi, \lambda),$$

that

$$T(\phi, \lambda, r) = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda).$$

Symbolically

$$T(\phi, \lambda, r) = L\{T(\phi, \lambda, R)\},$$

Here, L is the linear operator

$$L\{f\} = \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} f_n,$$

where the f_n are defined according to the degree constituent equation 3.7, so that on the surface of the sphere

$$f = \sum_{n=2}^{\infty} f_n.$$

Symbolically

$$L\{f\} = \sum_{n=2}^{\infty} L^n f_n,$$

where

$$L^n = \left(\frac{R}{r}\right)^{n+1}$$

is the *spectral representation* of the operator L.

We may still write *in a certain point* P (ϕ_P, λ_P, r_P) in space:

$$L_P\{f\} = \sum_{n=2}^{\infty} L_P^n f_n,$$

in which

$$L_P^n = \left(\frac{R}{r_P}\right)^{n+1}.$$

Concretely, for the disturbing potential $T(\phi_P, \lambda_P, r_P)$ in point P, this means

$$T(\phi_P, \lambda_P, r_P) = L_P T\{(\phi, \lambda, R)\} =$$
$$= \sum_{n=2}^{\infty} L_P^n T_n =$$
$$= \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} T_n.$$

Now the covariance function *in space* of T is obtained:

$$K(r_P, r_Q, \psi_{PQ}) = M\{T_P, T_Q\} = M\{T(\phi_P, \lambda_P, r_P) T(\phi_Q, \lambda_Q, r_Q)\} =$$
$$= M\{L_P\{T(\phi, \lambda, R)\} L_Q\{T(\phi, \lambda, R)\}\} =$$
$$= M\left\{\sum_{n=2}^{\infty} (L_P^n T_n) \sum_{n'=2}^{\infty} \left(L_Q^{n'} T_{n'}\right)\right\} =$$
$$= \sum_{n=2}^{\infty} \sum_{n'=2}^{\infty} L_P^n L_Q^{n'} M\{T_n T_{n'}\}.$$

Based on the orthogonality of the functions T_n it holds that

$$M\{T_n T_{n'}\} = \begin{cases} k_n P_n(\cos \psi_{PQ}) & \text{if } n = n', \\ 0 & \text{if } n \neq n', \end{cases}$$

i.e., the harmonic components of the surface covariance function

$$K(\psi_{PQ}) = \sum_{n=2}^{\infty} M\{T_n T_n\} = \sum_{n=2}^{\infty} k_n P_n(\cos \psi_{PQ}). \tag{10.12}$$

Thus we obtain[7]

$$K(r_P, r_Q, \psi_{PQ}) = \sum_{n=2}^{\infty} L_P^n L_Q^n k_n P_n(\cos \psi_{PQ}) =$$

$$= \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \left(\frac{R}{r_Q}\right)^{n+1} k_n P_n(\cos \psi_{PQ}) =$$

$$= \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+1} k_n P_n(\cos \psi_{PQ}). \tag{10.13}$$

Here we have expressed the covariance function of the disturbing potential in space $T(\phi, \lambda, r)$ into an expansion into the degree variances k_n of the corresponding Earth's surface disturbing potential $T(\phi, \lambda, R)$, by applying *propagation of covariances* on the expansion 10.12 of the function K. Thus we have obtained the *three-dimensional covariance function* for the disturbing potential, needed, e.g., in mountainous countries and in air and space applications.

🔲 **10.8.2   Example: *the covariance function of gravity anomalies***

We know (equation 5.8) that there exists the following relationship between gravity anomalies and the disturbing potential:

$$\Delta g = \frac{1}{r} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} (n-1) T_n,$$

---

[7]This works only this cleanly because in this case the operator L^n is of multiplier type, $\left(\frac{R}{r}\right)^{n+1}$.

symbolically: $\Delta g = L_g\{T\}$ for a suitable operator L_g:

$$L_g\{f\} = \sum_{n=2}^{\infty} L_g^n f_n,$$

where now

$$L_g^n = \frac{n-1}{r} \left(\frac{R}{r}\right)^{n+1}.$$

Again, in a concrete point P,

$$\Delta g(\phi_P, \lambda_P, r_P) = L_{g,P}(T(\phi, \lambda, R)) =$$

$$= \sum_{n=2}^{\infty} L_{g,P}^n T_n =$$

$$= \sum_{n=2}^{\infty} \frac{n-1}{r_P} \left(\frac{R}{r_P}\right)^{n+1} T_n.$$

Now we can show in the same way as above, that

$$\text{Cov}\{\underline{\Delta g}_P, \underline{\Delta g}_Q\} = M\{\Delta g_P \Delta g_Q\} = \sum_{n=2}^{\infty} L_{g,P}^n L_{g,Q}^n M\{T_n T_n\} =$$

$$= \sum_{n=2}^{\infty} \frac{n-1}{r_P} \left(\frac{R}{r_P}\right)^{n+1} \frac{n-1}{r_Q} \left(\frac{R}{r_Q}\right)^{n+1} k_n P_n(\cos \psi_{PQ}) =$$

$$= \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+2} \left(\frac{n-1}{R}\right)^2 k_n P_n(\cos \psi_{PQ}).$$

Often we write

$$C(\psi_{PQ}, r_P, r_Q) \overset{\text{def}}{=} M\{\Delta g_P \Delta g_Q\} = \sum_{n=2}^{\infty} \left(\frac{R^2}{r_P r_Q}\right)^{n+2} c_n P_n(\cos \psi_{PQ}),$$

where the *degree variances of gravity anomalies* are

$$c_n = \left(\frac{n-1}{R}\right)^2 k_n.$$

Similarly we calculate also the "mixed covariances" between disturbing potential and gravity anomaly:

$$\mathrm{Cov}\{\underline{T}_P, \underline{\Delta g}_Q\} = M\{T_P \Delta g_Q\} = \sum_{n=2}^{\infty} L_P^n L_{g,Q}^n M\{T_n T_n\} =$$

$$= \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \frac{n-1}{r_Q} \left(\frac{R}{r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}) =$$

$$= \sum_{n=2}^{\infty} \frac{n-1}{r_Q} \left(\frac{R^2}{r_P r_Q}\right)^{n+1} k_n P_n(\cos\psi_{PQ}).$$

kovarianssien kulkeutuminen All these are examples of *propagation of covariances*, when applied to a series expansion:

$$\mathrm{Cov}\left\{L_1\{\underline{T}_P\}, L_2\{\underline{T}_Q\}\right\} = \sum_n L_{1,P}^n L_{2,Q}^n M\{T_n T_n\} =$$

$$= \sum_n L_{1,P}^n L_{2,Q}^n k_n P_n(\cos\psi_{PQ}),$$

for *arbitrary* linear functionals

$$L_1\{T_P\} = \sum_{n=2}^{\infty} L_{1,P}^n T_n, \qquad L_2\{T_Q\} = \sum_{n=2}^{\infty} L_{2,Q}^n T_n,$$

where the $T_n = T_n(\phi, \lambda)$ are the degree constituents of the disturbing potential on the Earth's surface. The problem, in each case, is identifying the spectral form of this linear functional. This is done by expanding the quantity concerned into T_n, and lifting the coefficient found from the equation. These coefficients are indicated above by red and blue colourings.

🔲A 10.9   Global covariance functions

Empirical covariance functions have been calculated a lot. Empirical covariance functions for the whole Earth there have been only a few. Typically they are given in the form of a *degree variance formula*. The best known is the rule observed by William Kaula[8]:

$$k_n = \alpha n^{-4}.$$

---

[8]William M. Kaula (1926–2000) was an American geophysicist and space geodesist who studied the determination of the Earth's gravity field by means of satellite geodesy.

By writing

$$c_n = \left(\frac{n-1}{R}\right)^2 k_n,$$

where c_n are the degree variances of *gravity anomalies*, we obtain

$$c_n = \frac{\alpha}{R^2} \frac{(n-1)^2}{n^4} \approx \frac{\alpha}{R^2} n^{-2}.$$

Here, $\frac{\alpha}{R^2}$ is a planet specific constant, value about $1200\,\mathrm{mGal}^2$ for the Earth.

The Kaula rule does not hold very precisely for very high degree numbers. It applies, by the way, fairly well for the gravity field of Mars, of course with a different constant (Yuan et al., 2001).

Another well known rule is the Tscherning–Rapp equation (Tscherning and Rapp, 1974):

$$c_n = \frac{A(n-1)}{(n-2)(n+B)} = \left(\frac{n-1}{R}\right)^2 k_n.$$

The constants are, according to the authors, $A = 425.28\,\mathrm{mGal}^2$ and $B = 24$ (exactly). As a technical detail, one usually chooses $R = R_B = 0.999\overline{R}$, the radius of a *Bjerhammar*[9] *sphere* inside the Earth ($\overline{R}$ is the Earth mean radius). The form of the above equation is chosen so the covariance functions of various quantities will be closed expressions.

🔲A 10.10   Collocation and the spectral viewpoint

Also the calculations in least-squares collocation can be executed efficiently by way of FFT. For this one should study the symmetries present in the geometry, especially the *rotational symmetry*, which exists, e.g., in the direction of longitude on the whole Earth: nothing changes when we turn the whole Earth by a certain angle θ around its axis of rotation: for all longitudes, what happens is $\lambda \mapsto \lambda + \theta$.

In the following we discuss a simplified example in two dimensions. Let observations $\underline{\ell}_i = \underline{g}_i + \underline{n}_i$ of a field $g(\psi), \psi \in [0, 2\pi)$ be given on the edge of a *circle*, in points $\psi_i \stackrel{\text{def}}{=} 2\pi\frac{i}{N}, i = 0, 1, 2, \ldots, N-1$. Let us assume that also

---

[9]Arne Bjerhammar (1917–2011) was an eminent Swedish geodesist.
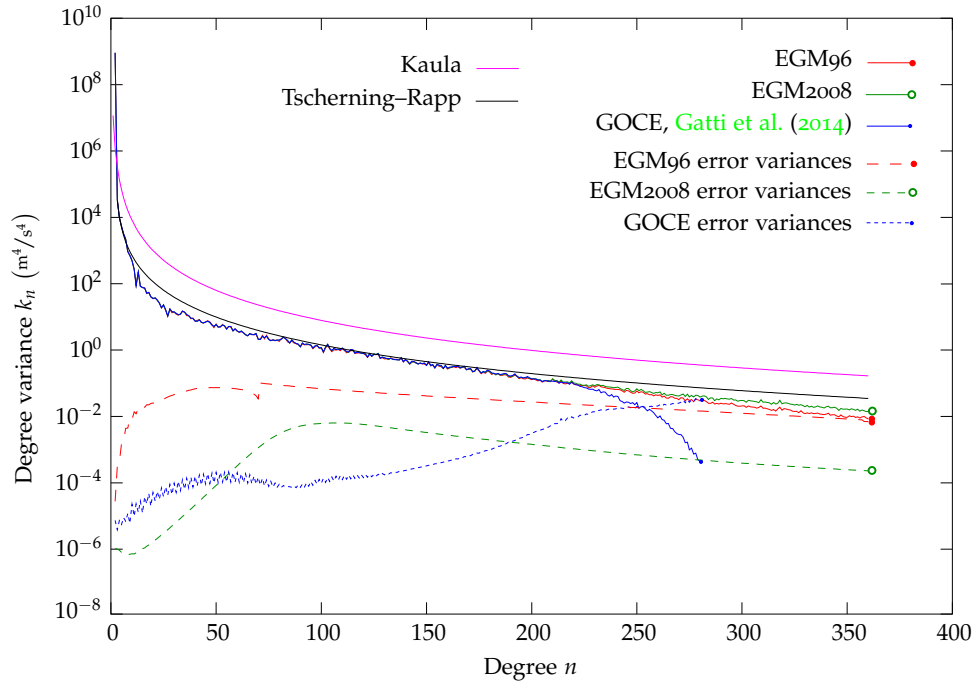
Figure 10.5. Global covariance functions as degree variances. The GOCE model cuts off at degree 280.

the results of the calculation, i.e., estimates $\widehat{f}_i$ of the result function $f(\psi)$ are desired in the same points. Then equation 10.8 yields

$$\widehat{f} = C_{fg} \left[C_{gg} + D_{gg} \right]^{-1} \left(\underline{g} + \underline{n} \right) \tag{10.14}$$

with

$$\left[C_{fg} \right]_{ij} = C_{fg} \left(f(\psi_i), g(\psi_i) \right) = C_{fg}(\psi_i, \psi_j),$$
$$\left[C_{gg} \right]_{ij} = C_{gg} \left(g(\psi_i), g(\psi_i) \right) = C_{gg}(\psi_i, \psi_j),$$
$$\left[D_{gg} \right]_{ij} = D_{gg} \left(g(\psi_i), g(\psi_i) \right) = D_{gg}(\psi_i, \psi_j).$$

If the *physics* of the whole situation, including the physics of the measurement

process, is rotationally symmetric, we must have

$$\left[C_{fg} \right]_{ij} = M_{\bigcirc} \{ f(\psi_i) \, g(\psi_{j(i)}) \} = \frac{1}{N} \sum_{i=0}^{N-1} f(\psi_i) \, g(\psi_{j(i)}),$$

in which $j(i) = (i + k) \bmod N$. Here, the operator $M_{\bigcirc}$ is again the "circle average" of a function,

$$M_{\bigcirc} \{ h \} \overset{\text{def}}{=} \frac{1}{N} \sum_{i=0}^{N-1} h(\psi_i),$$

which, like the geographic average in section 10.4, replaces the statistical average.

In the same way we obtain

$$\left[C_{gg} \right]_{ij} = M_{\bigcirc} \{ g(\psi_i) \, g(\psi_{j(i)}) \} = \frac{1}{N} \sum_{i=0}^{N-1} g(\psi_i) \, g(\psi_{j(i)}).$$

Now C_{fg}, C_{gg} are only functions of k, and we may write them

$$\left[C_{fg} \right]_{ij} = C_{fg}(\psi_i, \psi_j) = C_{fg}(\Delta\psi_k) = \left[C_{fg} \right]_k,$$
$$\left[C_{gg} \right]_{ij} = C_{gg}(\psi_i, \psi_j) = C_{gg}(\Delta\psi_k) = \left[C_{gg} \right]_k,$$

in which $\Delta\psi_k \overset{\text{def}}{=} (\psi_j - \psi_i) \bmod 2\pi$ and $k = (j - i) \bmod N$.

Furthermore

$$\left[D_{gg} \right]_{ij} = D_{gg}(\psi_i, \psi_j) = D_{gg}(\Delta\psi_k) = \left[D_{gg} \right]_k = E \{ \underline{n}_i \underline{n}_{j(i)} \},$$

the traditional statistical variance of the observation noise. Also because generally the observations do not correlate with each other, we have[10]

$$D_{gg} = \sigma^2 I_N,$$

---

[10] In fact, the unit or identity matrix is also known as the Kronecker delta, and as a Toeplitz matrix may be interpreted as a discrete version of the Dirac delta function. Its discrete Fourier transform is

$$\mathcal{F}\{I\} = 1$$

(and yes, this extends to infinite frequencies, and contains an infinite amount of power...).
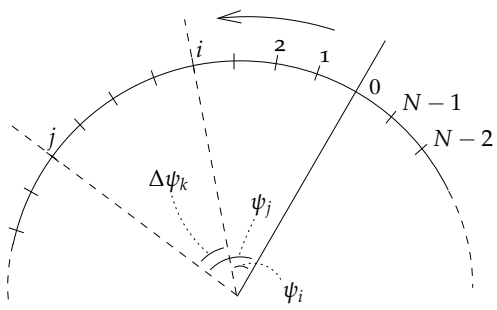
Figure 10.6. Circular geometry.

σ^2 (the variance of observations, assumed equal for all) times the $N \times N$ sized unit matrix.

Toeplitz-sirkulantti matriisi Matrices of this form are called *Toeplitz circulant*[11]. Thanks to this property, equation 10.14 is a string of *convolutions*.

Without proof we present that the spectral version of equation 10.14 looks like this:

$$\mathcal{F}\{\widehat{f}\} = \frac{\mathcal{F}\{C_{fg}\}}{\mathcal{F}\{C_{gg}\} + \mathcal{F}\{D_{gg}\}} \mathcal{F}\{\underline{g} + \underline{n}\} =$$
$$= \frac{\mathcal{F}\{C_{fg}\}}{\mathcal{F}\{C_{gg}\} + \sigma^2} \mathcal{F}\{\underline{g} + \underline{n}\}. \qquad (10.15)$$

This is an easy and rapid way to calculate the solution using FFT. In the limit in which the observations are exact, i.e., $\sigma^2 = 0$, by equation 10.15 $\widehat{f}$ follows straight from $\underline{g} + \underline{n} = \underline{g}$. If for a suitable operator L we have $f = L\{g\}$, the equation simplifies as follows:

$$\mathcal{F}\{\widehat{f}\} = \frac{\mathcal{F}\{L\}\mathcal{F}\{C_{gg}\}}{\mathcal{F}\{C_{gg}\} + \sigma^2} \mathcal{F}\{\underline{g} + \underline{n}\},$$

and if also $\sigma^2 = 0$, i.e., $\underline{n} = 0$, then

$$\mathcal{F}\{\widehat{f}\} = \mathcal{F}\{L\}\mathcal{F}\{\underline{g}\} \iff \widehat{f} = L\{\underline{g}\}.$$

[11]Otto Toeplitz (1881 – 1940) was a German Jewish mathematician who contributed to functional analysis.

For example, if the g are gravity anomalies and the f are values of the disturbing potential, then[12]

$$\mathcal{F}\{L\} = \frac{R}{n-1}.$$

The approach is called *Fast Collocation*, e.g., Bottoni and Barzaghi (1993). Of course it is used in two dimensions on the Earth's surface, though our example is one-dimensional. As always, it requires that the observations are given on a *grid*, and in this case also, that the precision of the material is *homogeneous* — the same everywhere — over the area. This requirement is hardly ever precisely fulfilled.

Self-test questions

1. What is the difference between *signal* and *noise*?

    (a) Signal is not a random stochastic process, whereas noise is.

    (b) Signal is a stochastic process that we are interested in and wish to estimate, while noise is a stochastic process that we are not interested in and that we would like to filter out.

    (c) Signal is a stochastic process with a greater variance and is therefore more easily detectable than noise.

    (d) Signal is a property of a real-world system, while noise is a property of an observation instrument or method.

2. What is a functional?

    (a) A mapping from a function space to a set of numbers, e.g., the real numbers.

    (b) A random-valued function.

    (c) A functional associates with every (well behaved) function defined on some domain, a number.

---

[12]In real computation it is not so simple... the degree number n, which refers to global spherical geometry, must first be converted to the Fourier wave number expressed on the computational grid used.

(d) A function of a vectorial argument.

3. What is a linear functional?

(a) A linear functional associates a number $L\{f\}$ with any linear function $f(x) = a + bx$ defined on some domain

(b) If, for functions f and g, it holds for a functional L that

$$L\{af + bg\} = aL\{f\} + bL\{g\}$$

for any real values a, b, then L is a linear functional.

(c) A linear functional associates with any (well behaved) function defined on some domain, a linear expression $L\{f\} = a + bx$.

4. The statistical behaviour of a stochastic process defined on the Earth's surface is the same independently of where on Earth you are. This property is called isotropy | ergodicity | homogeneity | stationarity.

5. The statistical behaviour of a stochastic process of time is the same independently of where on the time axis you are. This property is called isotropy | ergodicity | homogeneity | stationarity.

6. Why, in the study of the Earth's gravity field, one uses as the average of quantities the geographical average rather than the statistical average?

7. Which two different kinds of covariance functions are used for gravity anomalies on the Earth's surface? Give the formulas and name the free parameters.

8. Explain *degree variances*. What is the difference between degree variances k_n and c_n?

9. Describe Kaula's rule.

10. What is a Toeplitz circulant matrix?

Exercise 10 − 1: Variance of prediction

The equation for the variance of prediction in a point P is

$$\Sigma_{PP} = C_{PP} - C_{Pi}(C_{ij} + D_{ij})^{-1}C_{jP},$$

≡ ↑ 🖼 ▦ 🔍 🗐 ✛

in which the observation points are $i = 1, \ldots, N$. Assume there is only one observation point, point P. Then

$$\Sigma_{PP} = C_{PP} - C_{PP}(C_{PP} + D_{PP})^{-1}C_{PP}.$$

Show that, if $D_{ij} \neq 0$ but however $D_{ij} \ll C_{ij}$,

$$\Sigma_{PP} \approx D_{PP}.$$

Exercise 10 − 2: Hirvonen's covariance equation and prediction

Hirvonen's covariance equation is

$$C(s) = \frac{C_0}{1 + \left(\frac{s}{d}\right)^2},$$

with the Ohio parameters $C_0 = 337\,\text{mGal}^2$ and $d = 40\,\text{km}$. The equation gives the covariance between the gravity anomalies in two points P and Q

$$C(s_{PQ}) = \text{Cov}\{\underline{\Delta g}_P, \underline{\Delta g}_Q\}.$$

s_{PQ} is the metric distance between the points.

1. Calculate $\text{Var}\{\underline{\Delta g}_P\}$ and $\text{Var}\{\underline{\Delta g}_Q\}$. Remember that according to the definition $\text{Var}\{\underline{x}\} = \text{Cov}\{\underline{x}, \underline{x}\}$!

2. Calculate $\text{Cov}\{\underline{\Delta g}_P, \underline{\Delta g}_Q\}$ if $s_{PQ} = 20\,\text{km}$.

3. Calculate the *correlation*

$$\text{Corr}\{\underline{\Delta g}_P, \underline{\Delta g}_Q\} \overset{\text{def}}{=} \frac{\text{Cov}\{\underline{\Delta g}_P, \underline{\Delta g}_Q\}}{\sqrt{\text{Var}\{\underline{\Delta g}_P\}\,\text{Var}\{\underline{\Delta g}_Q\}}}.$$

4. Assume now, that *we only have a measurement in point P*. What is the "variance of prediction" of the gravity anomaly in point Q which is at a distance $s_{PQ} = 10\,\text{km}$ from the (precisely!) given anomaly in point P? Apply equation 10.10 as follows:

$$\sigma_{QQ}^2 = C_{QQ} - C_{QP}C_{PP}^{-1}C_{PQ}.$$

5. And item 4 if the distance is $s_{PQ} = 80\,\text{km}$?

≡ ↑ 🖼 ▦ 🔍 🗐 ✛

Exercise 10 – 3:   Predicting gravity anomalies

Let in two points 1 and 2 be given the measured gravity anomalies $\ell_1 = \Delta g_1 + n_1$ and $\ell_2 = \Delta g_2 + n_2$. The distance between the points is 80 km and between them, at the same distance of 40 km from both, is located point P. Compute the gravity anomaly of point P, Δg_P by means of prediction. The prediction equation is

$$\widehat{\Delta g}_P = C_{Pi}\left(C_{ij} + D_{ij}\right)^{-1}\ell_j,$$

where $\ell_j = \Delta g_j + n_j$ is the (abstract) vector of gravity anomaly observations,

$$C_{ij} = \begin{bmatrix} \mathrm{Var}\{\Delta g_i\} & \mathrm{Cov}\{\Delta g_i, \Delta g_j\} \\ \mathrm{Cov}\{\Delta g_i, \Delta g_j\} & \mathrm{Var}\{\Delta g_j\} \end{bmatrix}$$

is the signal variance matrix of the vector Δg_i, and

$$C_{Pi} = \begin{bmatrix} \mathrm{Cov}\{\Delta g_P, \Delta g_1\} & \mathrm{Cov}\{\Delta g_P, \Delta g_2\} \end{bmatrix}$$

is the signal covariance matrix between Δg_P and Δg_i. D_{ij} is the variance matrix of the observation random uncertainty or *noise* $n_i, i = 1,2$:

$$D_{ij} = \begin{bmatrix} \mathrm{Var}\{n_i\} & \mathrm{Cov}\{n_i, n_j\} \\ \mathrm{Cov}\{n_i, n_j\} & \mathrm{Var}\{n_j\} \end{bmatrix}.$$

1. Compute the matrix C_{ij}, assuming again Hirvonen's covariance formula (previous exercise) and a parameter value of $d = 40$ km.
2. Compute C_{Pi}.
3. Compute $\widehat{\Delta g}_P$ expressed in the observed values ℓ_1 and ℓ_2. Assume $D_{ij} = 0$ (and thus $n_i = 0$). (Inverting the C_{ij} matrix is possible by hand, but just use Matlab.)
4. Compute the *variance of prediction* (note $C_{jP} = C_{Pi}^{\mathsf{T}}$) using

$$\sigma_{PP}^2 = C_{PP} - C_{Pi}C_{ij}^{-1}C_{jP}.$$

Exercise 10 – 4:   Predicting gravity anomalies (2)

Let us again have points 1 and 2 with measured gravity anomalies $\ell_1 = \Delta g_1$ and $\ell_2 = \Delta g_2$. Now however the points 1, 2 and P are in a *triangular*

configuration, with a right angle at point P, and the distances from P to points 1 and 2 still 40 km. The distance between points 1 and 2 is now only $40\sqrt{2}$ km.

1. Compute C_{ij}, C_{Pi}, $\widehat{\Delta g}_P$ and σ_{PP}^2.
2. Compare the result with the previous one. Conclusion?

Exercise 10 – 5:   Propagation of covariances

Given the covariance function 10.13 of the disturbing potential

$$\mathrm{Cov}\{T_P, T_Q\} = \sum_{n=2}^{\infty}\left(\frac{R^2}{r_P r_Q}\right)^{n+1} k_n P_n\left(\cos\psi_{PQ}\right),$$

1. calculate the covariance function of the gravity disturbance δg (equation 5.3). Hint: write first an expansion of form

$$\delta g = \sum_{n=2}^{\infty} L_{\delta g}^n T_n$$

in order to find the expression for the coefficient $L_{\delta g}^n$. After this

$$\mathrm{Cov}\{\delta g_P, \delta g_Q\} = \sum_{n=2}^{\infty} L_{\delta g,P}^n L_{\delta g,Q}^n k_n P_n\left(\cos\psi_{PQ}\right).$$

2. Compute the covariance function of the *gravity gradient* $\dfrac{\partial^2 T}{\partial r^2}$ (i.e., the vertical gradient of the gravity *disturbance*!).

Exercise 10 – 6:   Kaula's rule for gravity gradients

For the disturbing potential

$$T(\phi, \lambda, r) = \sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n+1} T_n(\phi, \lambda) \tag{10.16}$$

or on the Earth's surface $(r = R)$

$$T(\phi, \lambda, R) = \sum_{n=2}^{\infty} T_n(\phi, \lambda)$$

Kaula's rule applies, with the *degree variances*

$$k_n = \alpha n^{-4}.$$

From these one can derive, using propagation of variances, the degree variances of *gravity anomalies*

$$\Delta g = \sum_{n=2}^{\infty} L_g^n T_n = \sum_{n=2}^{\infty} \left(\frac{n-1}{R} \right) T_n$$

as follows:

$$c_n = \left(L_g^n \right)^2 k_n = \left(\frac{n-1}{R} \right)^2 k_n \approx \frac{\alpha}{R^2} n^{-2}.$$

By differentiating the above expansion 10.16 for the disturbing potential

$$T(r, \varphi, \lambda) = \sum_{n=2}^{\infty} \left(\frac{R}{r} \right)^{n+1} T_n(\varphi, \lambda)$$

we obtain the second derivative

$$\frac{\partial^2 T}{\partial r^2} = \sum_{n=2}^{\infty} \frac{(n+1)(n+2)}{r^2} \left(\frac{R}{r} \right)^{n+1} T_n,$$

the connection between the disturbing potential and the *gravity gradient* in the spectral domain.

On the Earth's surface $r = R$, or

$$\left. \frac{\partial^2 T}{\partial r^2} \right|_{r=R} = \sum_{n=2}^{\infty} \frac{(n+1)(n+2)}{R^2} T_n \stackrel{\text{def}}{=} \sum_{n=2}^{\infty} L_{gg}^n T_n$$

with

$$L_{gg}^n = \frac{(n+1)(n+2)}{R^2}.$$

1. Derive an (approximate) equation for the degree variances for the *gravity gradient*, which we may call g_n, in an analogue fashion as above for the gravity anomaly degree variances c_n:

$$g_n = ? \cdot k_n \approx ? \cdot n^?.$$

2. Conclusion?

Exercise 10 − 7: Underground mass points

1. If a mass point is placed inside the Earth at a depth D beneath an observation point P, what then is the *correlation length s* of the gravitational field it causes on the Earth's surface, for which $C(s) = \frac{1}{2} C_0$?

2. Thus, if we wish to construct a model made of mass points, where under each observation point Δg_P there is one mass point, how deep should we place them if the correlation length d is given?

📖 Gravimetric measurement devices

11

📖 11.1 History

The first measurement device ever built based on a *pendulum* was a clock. The pendulum equation,

$$P = 2\pi\sqrt{\frac{\ell}{g}},$$

tells that the swinging time or period P of a pendulum of a given length is a constant that depends only on the length ℓ and local gravity g, on condition that the swings are *small*. The Dutch Christiaan Huygens[1] built in 1657 the first useable pendulum clock based on this Wikipedia, Pendulum clock.

When the young French researcher Jean Richer[2] visited French Guyana in 1671 with a pendulum clock, he noticed that the clock ran clearly slower. The matter was corrected simply by shortening the pendulum. The cause of the effect could not be the climatic conditions in the tropics, i.e., the thermal expansion of the pendulum. The right explanation was that in the tropics, gravity g is weaker than in Europe. After return to France, Richer had

---

[1]Christiaan Huygens (1629 – 1695) was a leading Dutch natural scientist and mathematician. Besides inventing the pendulum clock, he also was the first to realize (in 1655) that the planet Saturn has a ring.

[2]Jean Richer (1630 – 1696) was a French astronomer. He is really only remembered for his pendulum finding.
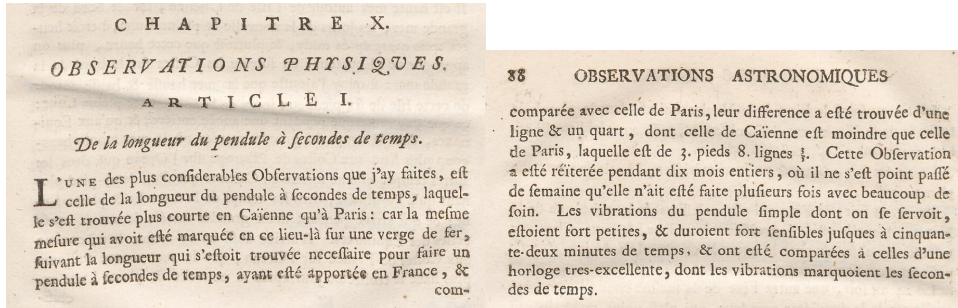
Figure 11.1. Jean Richer's report.

again to make his pendulum longer. The observation is described in just one paragraph on pages 87–88 in his report "*Observations astronomiques et physiques faites en l'isle de Caïenne*", Richer (1731).

This is how the *pendulum gravimeter* was invented. Later, much more precise special devices were built, e.g., Kater[3] 's reversion pendulum, and the four-pendulum Von Sterneck[4] device, which was also used in Finland in the 1920's and 1930's. We must mention also the submarine measurements, e.g., in the Java Sea by the Dutch F. A. Vening Meinesz in which it was observed that above the trenches in the ocean floor there is a notable shortage of gravity, and that they thus are in a state of strong isostatic disequilibrium (Vening Meinesz, 1928).

syvänmeren hauta

For production gravimetric observations, pendulum gravimeters are however too hard to operate and too slow. For that purpose the *spring gravimeter* has been developed, see section 11.2.

Pendulum gravimeters are in principle absolute measurement devices, i.e., gravity is obtained directly as an acceleration. There are, however, systematic effects associated with the suspension of the pendulum that make that one cannot trust in the absoluteness of measurement after all. One tried out solution is the *very long wire pendulum*, e.g., Hytönen (1972). However, nowadays

---

[3]Henry Kater (1777–1835) was an English physicist.

[4]Robert von Sterneck (1839–1910) was an Austrian-Hungarian scientist.

Figure 11.2. Autograv CG5 spring gravimeter. Image Autograv CG5.

absolute measurements are made with ballistic gravimeters, cf. section 11.3. It has been observed that the older measurements made with pendulum apparatus in the so-called Potsdam system are systematically 14 mGal too large...

11.2   The relative or spring gravimeter

A spring gravimeter is at its simplest the same as a *spring balance*.   jousivaaka

In a linear spring balance the equation of motion of the test mass is

$$m\left(\frac{d^2\ell}{dt^2} - g\right) = -k\left(\ell - \ell_0\right),$$

where m is the test mass, g the local (to be measured) gravity, k the spring constant. The quantity ℓ_0 is the "rest length" of the spring, its length if there were no external forces acting on it. ℓ is the true, instantaneous length of the string.

The equilibrium between the spring force and gravity is

$$\frac{d^2\ell}{dt^2} = 0 \implies mg = k\left(\ell - \ell_0\right) = k\left(\overline{\ell} - \ell_0\right), \tag{11.1}$$

in which $\overline{\ell}$ is the mean length of the spring during the oscillation, and also the *equilibrium length* in the absence of oscillations.

When the test mass is disturbed, it starts oscillating about its equilibrium position. The oscillation equation, obtained by summing the above two equations, is

värähtely-yhtälö

$$\frac{d^2}{dt^2}\left(\ell - \overline{\ell}\right) = -\frac{k}{m}\left(\ell - \overline{\ell}\right).$$

The period is

$$P = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{\overline{\ell} - \ell_0}{g}} = 2\pi\sqrt{\frac{\delta\ell}{g}}, \tag{11.2}$$

in which $\delta\ell = \overline{\ell} - \ell_0$ denotes the difference between the equilibrium length and the length in the state of rest, i.e., the *lengthening of the spring by gravity*.

The *sensitivity* of the instrument is obtained by differentiating equation 11.1 in the form

$$mg = k\left(\overline{\ell} - \ell_0\right) = k\,\delta\ell$$

with the result

$$\frac{d\overline{\ell}}{dg} = \frac{d\left(\delta\ell\right)}{dg} = \frac{m}{k} = \frac{P^2}{4\pi^2}. \tag{11.3}$$

Substitution, e.g., of $\delta\ell = 5\,\text{cm}$ and $g = 10\,\text{m/s}^2$ into equation 11.2 yields $P = 0.44\,\text{s}$. One milligal of change in gravity g produces according to equation 11.3 a lengthening of only $5 \cdot 10^{-8}\,\text{m}$ (check)! Clearly then, the sensor observing or compensating this displacement must be extremely sensitive!

11.2.1 Astatization

An *astatized gravimeter* offers a different measurement geometry. We use as our example the LaCoste-Romberg gravimeter which long enjoyed great popularity. In it, the test mass is at the end of a lever beam, see figure 11.3. Two torques are operating on the beam, which are in equilibrium. The torque by the spring is

$$\tau_{\text{s}} = k\left(\overline{\ell} - \ell_0\right) b\sin\beta,$$

in which $\overline{\ell}$ is the spring's true, stretched equilibrium length, and ℓ_0 the theoretical or state-of-rest length without loading.

According to the sine rule

$$\overline{\ell}\sin\beta = c\sin(90° + \epsilon) = c\cos\epsilon,$$

from which upon substitution in the previous:

$$\tau_{\text{s}} = k\left(\overline{\ell} - \ell_0\right)\frac{bc}{\overline{\ell}}\cos\epsilon.$$

Gravity pulling at the mass again is mg, and the corresponding torque

$$\tau_{\text{g}} = mgp\cos\epsilon.$$

Between these there has to be *equilibrium*:

$$\tau_{\text{g}} - \tau_{\text{s}} = mgp\cos\epsilon - k\left(\overline{\ell} - \ell_0\right)\frac{bc}{\overline{\ell}}\cos\epsilon = 0,$$

or

$$mgp\overline{\ell} - kbc\left(\overline{\ell} - \ell_0\right) = 0. \tag{11.4}$$

By differentiation

$$mp\overline{\ell}\,dg + mgp\,d\overline{\ell} - kbc\,d\overline{\ell} = 0$$

from which we obtain, by substituting equation 11.4, a *sensitivity equation*:

$$\frac{d\overline{\ell}}{dg} = -\frac{mp\overline{\ell}}{mgp - kbc} = -\frac{mp\overline{\ell}}{mgp - mgp\frac{\overline{\ell}}{\overline{\ell} - \ell_0}} = \frac{\overline{\ell}}{g}\frac{\overline{\ell} - \ell_0}{\ell_0}.$$

From this we see that the sensitivity can be driven up arbitrarily by choosing ℓ_0 as short as possible, almost zero — a so-called *zero-length spring* solution (Wikipedia, Zero-length springs).

Of course, *levelling* the instrument, using its bull's eye level and three foot-screws, is critical.

tasaus
rasiatasain

For example, assuming $\overline{\ell} = 5\,\text{cm}$, $\ell_0 = 0.1\,\text{cm}$, $g = 10\,\text{m/s}^2$ gives

$$\frac{d\overline{\ell}}{dg} = 2.5 \cdot 10^{-6}\,\text{m/mGal},$$
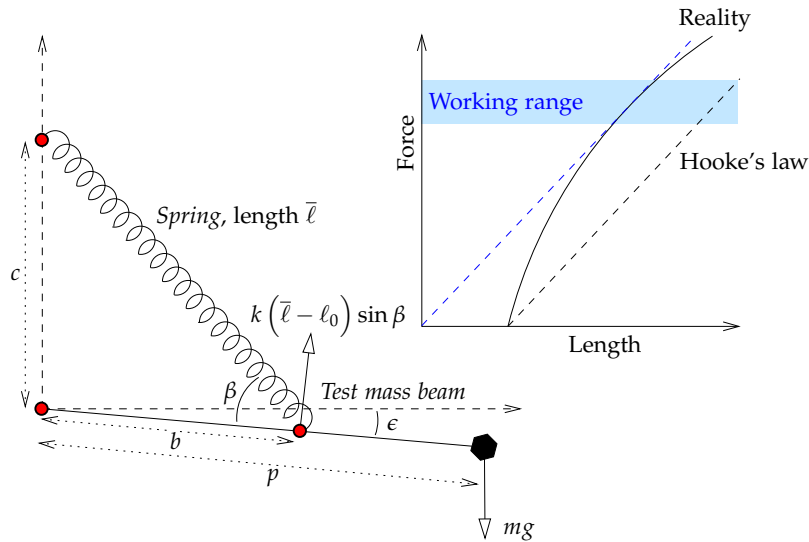
Figure 11.3. Operating principle of spring gravimeter. On the right, how to build a "zero-length spring".

a 50 times[5] better result than earlier! The improvement or *astatization ratio* is precisely $\frac{\bar{\ell}-\ell_0}{\ell_0}$.

This is the operating principle of an *astatized* gravimeter, like the LaCoste-Romberg[6].

11.2.2 Period of oscillation

There is another way to look at this: if the instrument is not in equilibrium, the lever beam will slowly oscillate about the equilibrium position. We start

---

[5]For comparability we should still multiply by $\frac{p}{b\sin\beta}$, if we measure the position of the test mass.

[6]Lucien LaCoste (1908–1995) was an American physicist and metrologist, who, as an undergraduate, together with his physics professor Arnold Romberg (1882–1974) discovered the principle of the astatized gravimeter and zero-length spring.

---

from equation 11.4:

$$mgp\bar{\ell} - kbc\left(\bar{\ell}-\ell_0\right) = 0, \tag{11.5}$$

but for a state of disequilibrium. Then, the test mass will be undergoing an acceleration a, and we have

$$m\left(g-a\right)p\ell - kbc\left(\ell-\ell_0\right) = 0,$$

where, instead of the equilibrium spring length $\bar{\ell}$, we have the instantaneous length ℓ. Subtracting the above two equations yields

$$mgp\left(\bar{\ell}-\ell\right) - map\,\ell - kbc\left(\bar{\ell}-\ell\right) = 0.$$

We use equation 11.5 again to eliminate kbc, yielding

$$mgp\left(\bar{\ell}-\ell\right) - map\,\ell - mgp\frac{\bar{\ell}}{\bar{\ell}-\ell_0}\left(\bar{\ell}-\ell\right) = 0.$$

Rearranging terms gives

$$map\,\ell = mgp\frac{-\ell_0}{\bar{\ell}-\ell_0}\left(\bar{\ell}-\ell\right)$$

or

$$a = \frac{g}{\ell}\frac{\ell_0}{\bar{\ell}-\ell_0}\left(\ell-\bar{\ell}\right).$$

Here we see again the "astatization ratio" $\frac{\bar{\ell}-\ell_0}{\ell_0}$ appear, which for a zero-length spring ($\ell_0 \approx 0$) is very large.

Now the string length disequilibrium $\ell - \bar{\ell}$ is connected with the vertical displacement z (reckoned upward) of the test mass, as follows:

$$z = -\left(\ell-\bar{\ell}\right)\frac{p}{b\sin\beta}.$$

With this we obtain

$$a = \frac{d^2}{dt^2}z = -\frac{g}{\ell}\frac{\ell_0}{\bar{\ell}-\ell_0}\frac{b\sin\beta}{p}z.$$

This is again an oscillation equation in z, with a period of

$$P = 2\pi\sqrt{\frac{\ell}{g}\frac{p}{b\sin\beta}\frac{\bar{\ell}-\ell_0}{\ell_0}}.$$
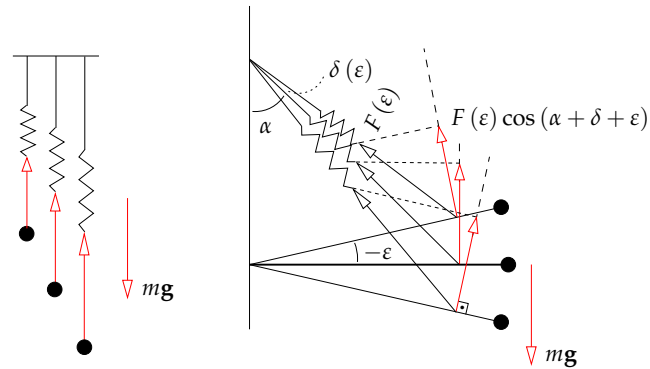
Figure 11.4. The idea of astatization. The elastic force of an ordinary spring grows steeply with extension (left), whereas the weight of the test mass is constant. The lever beam and diagonal arrangement (right) causes the part of the force of the spring in the direction of motion of the lever (red) to *diminish* with extension, while the spring force itself *grows* similarly with extension. This near-cancellation boosts sensitivity. The spring used is a *zero-length spring*.

For the same values as above, $\ell_0 = 0.1\,\text{cm}$, $\overline{\ell} = 5\,\text{cm} \approx \ell$, $g = 10\,\text{m/s}^2$, and $\frac{p}{b \sin \beta} = 2$, we find

$$P = 4.4\,\text{s}.$$

What this long oscillation period also means is, that the instrument is less sensitive to high-frequency vibrations by passing traffic, microseismicity, etc. This is a significant operational advantage.

11.2.3   Practicalities of measurement

An ordinary spring gravimeter is based on *elasticity*. Because there is no material that is perfectly elastic, but always also plastic[7] (viscous), the gravimeter

käynti   itself *changes* during the measurement process. This change is called *drift*.

---

[7]Plastic deformation in a metal crystal is mediated by crystal-lattice defects called *dislocations*. As dislocations travel through the crystal lattice under

The drift is managed in practical measurements by the following measures:

- we measure along lines starting from a known point and ending on a known point, producing a *closing error*. The line is traversed as rapidly as possible. The closing error is eliminated by adjusting the values obtained from the measurement in proportion to their times of measurement.
- The gravimeter is transported carefully without bumping it, and
- we remember always to *arrest* (clamp down the lever beam) during   arretointi transport!
- Because the elastic properties of the spring and the instrument geometry both depend on temperature, precision gravimeters are always *thermostated*.

A sea gravimeter differs from an ordinary (land) gravimeter in having a powerful *damping*. This applies also for an airborne gravimeter. Both types   vaimennus are mounted on a stabilized platform, keeping the axis of measurement along the local vertical in spite of vehicle motions.

11.3   The absolute or ballistic gravimeter

The ballistic or absolute gravimeter is a return to roots, the definition of gravity: it measures directly the *acceleration of free fall*. The instrument comprises a vacuum tube, inside of which an object, a prism reflecting light, falls freely. See figure 11.5.

Here we describe shortly the JILA gravimeter, built at the University of Colorado at Boulder by Jim Faller[8] of which the Finnish Geodetic Institute has acquired two. Figure 11.6 shows the newer model, FG5, built by the

---

load, the properties of the metal change, which may eventually result in *metal fatigue*, a known problem, i.a., in aviation. Wikipedia, Dislocation. The art of making metals stronger by inhibiting the motion of dislocations, e.g., by adding carbon to iron to form steel, forms a large part of metallurgy. Wikipedia, Strengthening mechanisms of materials.

[8]James E. Faller (1934 –) is an American physicist, metrologist, geodesist and

Gravimetric measurement devices
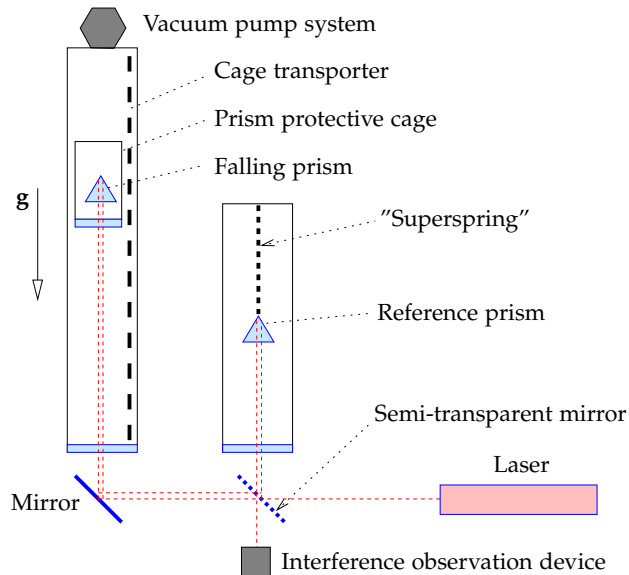
Figure 11.5. Operating principle of a ballistic absolute gravimeter.

Figure 11.6. FG5 absolute gravimeter. Figure © National Oceanic and Atmospheric Administration.

same group. In Finland this instrument, serial number 221, has served as the national standard for the acceleration of free fall. It was upgraded to a model FG5X in 2012.

During the fall of the prism, a "cage" with a window in the bottom moves along with the prism inside it without touching it. The purpose of the cage is to prevent the last remaining traces of air from affecting the motion of the prism. Approaching the bottom, the cage, which moves along a rail under computer control, decelerates, and the prism lands relatively softly on its bottom. After that, the cage moves back to the top of the tube and a new measurement cycle starts.

A laser interferometer measures the locations of the prism during its fall;

student of gravitation. He proposed the installation of laser retroreflectors on the lunar surface in the context of the Apollo project, in order to measure the distance to the Moon — LLR, lunar laser ranging.

the measurements are repeated thousands of times to get a good precision through averaging. Another prism, the reference prism, is suspended in another tube from a very soft spring (actually an electronically simulated "superspring") to protect it from microseismicity. The instrument is designed to achieve the greatest precision possible, e.g., the vibration caused by the drop is controlled by a well-designed mount. Precisions are of order several μGal, similar to what ordinary LaCoste-Romberg relative gravimeters are capable of. The instrument is however large and, though transportable, it cannot be called a field instrument. Of late, development has gone in the direction of smaller devices, which are essentially better portable.

The motion of a freely falling mass is described by the equation

$$\frac{d^2}{dt^2}z = g(z),$$

where it is assumed — realistically — that gravity g depends on the location

z within the drop tube. If we nevertheless take g to be constant, we obtain by integration

$$\frac{d}{dt}z = v_0 + gt,$$

$$z = z_0 + v_0 t + \frac{1}{2}gt^2,$$

from which we obtain the *observation equations* of the measurement process

$$\underline{z}_i = \left[\begin{array}{ccc} 1 & t_i & \frac{1}{2}t_i^2 \end{array} \right] \cdot \left[\begin{array}{c} z_0 \\ v_0 \\ g \end{array} \right] + \underline{n}_i.$$

Here, the unknowns[9] are $\widehat{z}_0$, $\widehat{v}_0$ and $\widehat{g}$. The quantities $\underline{z}_i$ are the interferometrically measured vertical locations of the falling prism, and $\underline{n}_i$ are the residuals of the measurements. Determining precisely the corresponding measurement time or *epoch* t_i is of course essential. The volume of measurements obtained from each drop is large.

We write the observation equations in matric form:

$$\underline{\ell} = A\mathbf{x} + \underline{n},$$

in which

$$\underline{\ell} = \left[\begin{array}{c} \underline{z}_1 \\ \underline{z}_2 \\ \vdots \\ \underline{z}_i \\ \vdots \\ \underline{z}_n \end{array} \right], \quad \underline{n} = \left[\begin{array}{c} \underline{n}_1 \\ \underline{n}_2 \\ \vdots \\ \underline{n}_i \\ \vdots \\ \underline{n}_n \end{array} \right], \quad A = \left[\begin{array}{ccc} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_i & t_i^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{array} \right] \quad \text{and} \quad \mathbf{x} = \left[\begin{array}{c} z_0 \\ v_0 \\ g \end{array} \right].$$

From this, the solution follows according to the method of least-squares adjustment, from the *normal equations*

$$A^{\mathsf{T}}A\widehat{\mathbf{x}} = A^{\mathsf{T}}\underline{\ell}$$

---

[9]It would be easy (exercise!) to add an unknown representing the vertical gradient of gravity to this.

giving the solution (estimate)

$$\widehat{\mathbf{x}} = \left(A^{\mathsf{T}}A \right)^{-1} A^{\mathsf{T}}\underline{\ell}.$$

The uncertainty of the estimates is given by the variance matrix

$$\mathrm{Var}\{\widehat{\mathbf{x}}\} = \sigma^2 \left(A^{\mathsf{T}}A \right)^{-1},$$

σ being the uncertainty (mean error) of a single observation, also known as the mean error of unit weight. z_i

An alternative type of absolute gravimeter *throws* the prism *up* (inside the tube), after which it moves along a symmetric parabolic path. Such a "rise-and-fall" instrument is, e.g., the Italian IMGC-02 (d'Agostino et al., 2008). Theoretically this method would give more precise results, however, the technical challenges are larger than in case of the dropping method. Inter-comparisons between instruments of these two types have helped to identify error sources.

Recently also so-called atomic or quantum gravimeters have been built, in which interferometrically the falling of individual atoms is measured (de Angelis et al., 2009). The idea of the device is, that it measures the effect of gravity on the phase angle of the matter wave of falling atoms. Firstly an extremely cold, so-called Bose–Einstein condensate is prepared; perhaps a million atoms in identical quantum states, with the same phase angle like marching soldiers. The condensate is dropped, and the first laser pulse splits it into two. Half of the atoms[10] fall first slowly, then faster; the other half fast at first and then slower. In order to achieve this, a second laser pulse pair is used that acts like a mirror, or perhaps a tennis racket. The third and last laser pulse is for reading out interferometrically the phase difference between the two merging atomic beams. The interaction between light and atoms is based on the Raman effect.

As the atoms travel two different paths through space-time where the gravity

---

[10]This is a quantum theoretically erroneous statement. The matter wave of *each individual atom* splits into two!
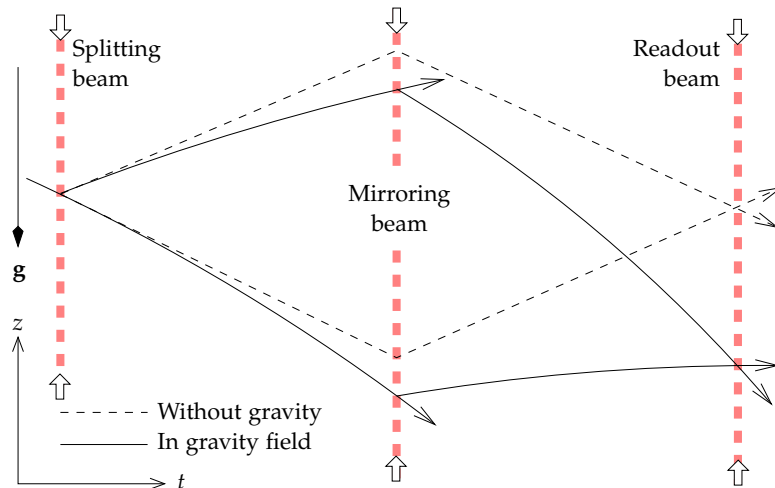
Figure 11.7. Principle of operation of an atomic gravimeter.

potential is different[11], a phase difference is formed between these which can be measured. Without gravity (dashed lines) this phase difference would be zero. See figure 11.7, where the horizontal axis is time.

11.4 Network hierarchy in gravimetry

In gravimetry, network hierarchy is just as important as in measurements of location or height. The procedure has typically been, that the highest measurement order consisted of points measured by absolute gravimeters — in the old days this meant pendulum measurements. Stepwise densification of this network, i.e., measurement of the base network, was then done with relative or spring gravimeters, like also the lowest-order measurements, gravity mapping surveys. In base network measurement, fast transportation

---

[11]In fact, the spinning of the atom's phase angle acts like a clock, and the speed at which time elapses depends on the local geopotential (Vermeer, 1983).

Figure 11.8. International intercomparison of absolute gravimeters. Image © 2003 EGCS, Luxembourg.

was used, such as aircraft: national or regional reference points often were located at airports.

Because pendulum instruments were not genuinely absolute, the old, so-called Potsdam system collected a 14 mGal systematic error: all values were that much too high. Nowadays we use instead ballistic free-fall gravimeters, the possible systematics of which are much smaller — but not nonexistent, order of magnitude microgals. As there are no better, i.e., more absolute, instruments than these, the issue cannot really be resolved. Nevertheless, international instrument intercomparisons, like the International Intercomparison of Absolute Gravimeters, are organized regularly and are valuable.

In Finland, regular absolute gravimetric measurements have been made besides in Metsähovi, also in Vaasa (two points), Joensuu (two points), Kuusamo, Sodankylä, Kevo and Eurajoki.

11.5 The superconducting gravimeter

This gravimeter type is based on a superconducting metal sphere levitating on a magnetic field, the precise place of which is measured electronically.
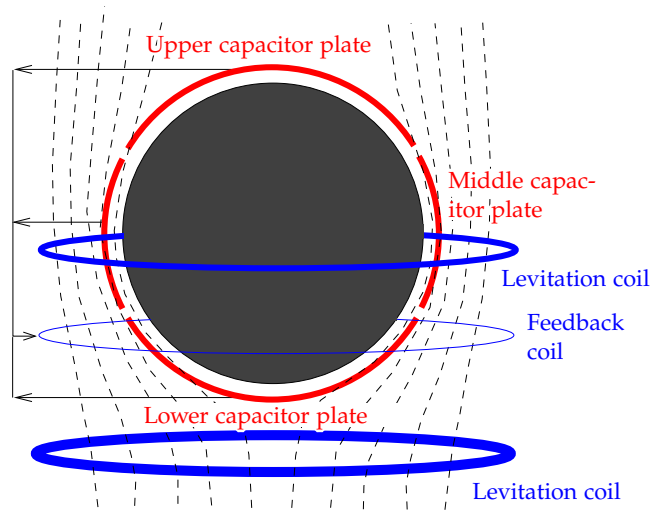
Figure 11.9. Principle of operation of a superconducting gravimeter. Reading out the sphere position is done capacitively.

Because a superconducting material is impenetrable by a magnetic field, the sphere will remain forever in the same spot inside the field: the Meissner effect. Of course the field itself must be constant; it is generated by superconducting solenoids inside a vessel made of mu-metal,Wikipedia, Mu-metal, which keeps out the Earth's magnetic field.

käämet

Superconductivity in these applications still demands working at the temperature of liquid helium (He). For this reason the device is not only expensive, but requires also an expensive laboratory in an environment where societal infrastructure works.

The number of these superconducting gravimeters in the world is over twenty. One GWR 20 type instrument has worked from 1994 in Kirkkonummi at the Metsähovi research station of the then Finnish Geodetic Institute, now the National Land Survey. See Virtanen and Kääriäinen (1995), Virtanen (1998). The instrument was upgraded in 2014.

The most important property of a superconducting gravimeter is, in addition

to its precision[12], its *stability*, i.e., its small drift. For this reason it is extremely suited for monitoring long period phenomena, like the free oscillations of the solid Earth after large earthquakes[13]. Thus it is suitable for measurements that are unsuitable for an ordinary gravimeter because of its larger drift and poorer sensitivity, and measurements for which a seismometer is unsuited because the frequencies are too low.

A recent trend is the development of lightweight, "portable" and remotely controllable superconducting gravimeters, e.g., the GWR *i*Grav, weighing 30 kg and not consuming any liquid helium at all. On the other hand it needs over a kilowatt in grid power for its refrigeration system (GWR Instruments, Inc., *i*GRAV® Gravity Sensors). Perhaps this will lead to improvement over the current situation where the bulk of instruments is located in Europe and North America.

11.6   Atmospheric influence on gravity measurement

The atmosphere has the following two effects on gravity:

1. *Instrumental effects.* These are due to the way the gravimeter is constructed. By putting the instrument in a pressure chamber, one makes these effects go away. In practice it is easier to *calibrate* the instrument (in the laboratory) and calculate a correction term according to the calibration certificate to be applied to the field measurements.

2. *Attraction of the atmosphere.* This is real gravitation. It contains irregular variations with place and time that we need to remove from the observed gravity values.

   The effect of the atmosphere can be evaluated with the aid of the Bouguer plate approximation: if air pressure is p, then the surface mass

---

[12]Virtanen (2006) reports how the instrument at Metsähovi detected the change in gravity as workmen cleared snow from its laboratory roof, including a tea break! "Weighing" visitors to the lab by their gravitational attraction is also standard fare.

[13]Their periods range from under an hour to over twenty hours, and they are of considerable geophysical interest.

density of the atmosphere is

$$\kappa = \frac{p}{g},$$

where g is a representative gravity value inside the atmosphere. We don't make a very large error by assuming

$$g \approx 9.8 \, \text{m/s}^2,$$

giving us on sea level[14] $\kappa \approx 10\,000 \, \text{kg/m}^2$. The effect of the Bouguer plate is

$$-2\pi G\kappa = -0.43 \, \text{mGal}.$$

Variations in air pressure affect proportionally. If the air pressure disturbance is $\Delta p = p - p_0$, in which p_0 is mean air pressure, 1015 hP, its effect on gravity measurement will be

$$\delta g_A = -0.43 \frac{\Delta p}{p_0} \text{mGal}.$$

During the passage of a storm or weather front, this beautiful theory collapses, and simple equations give misleading results. Then it is best to just not do any gravity measurements!

3. *Including the atmosphere into the mass of the Earth.* This *is not* a correction to be applied to gravity measurements. It is a *reduction* which is applied in the calculation of gravity *anomalies*, if we want anomalies from which the effect of the atmosphere has been removed.

   Remember that the reference or normal gravity field of GRS80 is defined in such a way, that the parameter GM contains the whole mass of the Earth including atmosphere, i.e., the Earth's gravitational field as satellites are observing it (Heikkinen, 1981). Therefore, also when calculating gravity anomalies Δg, one should reduce gravity by computationally moving *the whole atmosphere above the point of measurement* to below the measurement point, e.g., to sea level.

---

[14]So yes, the force acting on a standard 14-inch laptop screen is 540 kg... fortunately it's not an oldfashioned vacuum image tube.

The total mass of the atmosphere is

$$M_A = 4\pi\kappa R^2 = 4\pi \frac{p}{g} R^2.$$

According to Newton its attraction is

$$\frac{GM_A}{R^2} = \frac{4\pi Gp}{g},$$

twice the atmospheric reduction given above. At sea level, the effect is 0.86 mGal. At height, the effect is

$$0.86 \frac{p(H)}{p_0} \text{mGal},$$

in which $p(H)$ and p_0 are the air pressures at height H and at sea level, respectively.

11.7 Airborne gravimetry and gnss

In the early years of the 1990s GPS, the Global Positioning System, and more generally, satellite positioning, has changed *airborne gravimetry* from a difficult technology to something completely operational. To understand this, one must know the principle of operation of airborne gravimetry.

An aircraft carries an airborne gravimeter, an instrument that, in the same way as a sea gravimeter, is strongly *damped*. The measurement is done   vaimennus automatically, generally using electrostatic compensation. The instrument is mounted on a stabilized platform that follows the local vertical.

During flight, the gravimeter measures *total gravity* on board the aircraft, consisting of two parts:

1. gravity proper — i.e., gravity as felt in a reference frame connected to the Earth's surface — and

2. the pseudo-forces caused by the inevitable accelerations of the aircraft even in cruise flight.

Attached to the aircraft are a number of GNSS antennas. With these, and a geodetic GNSS instrument, the motions of the aircraft can be monitored with

centrimetre accuracy. From these can then be calculated the pseudo-forces mentioned above under item 2.

If we measure the position of the plane (or instrument) $\mathbf{x}_i$ at moment t_i, $\Delta t = t_{i+1} - t_i$, we obtain estimated acceleration values as follows:

$$\mathbf{a}_i \approx \frac{\mathbf{x}_{i+1} + \mathbf{x}_{i-1} - 2\mathbf{x}_i}{(\Delta t)^2}. \tag{11.6}$$

When the acceleration measured by the gravimeter is Γ_i and the direction of the local plumb line $\mathbf{n}_i$, local gravity g_i follows:

$$g_i = \Gamma_i - \langle \mathbf{a}_i \cdot \mathbf{n}_i \rangle.$$

Critical in the whole method is the choice of the time constant Δt. It is best to choose it as long as possible; then, the precision of the calculated GNSS accelerations $\mathbf{a}_i$ is as good as possible. Also the damping of the gravimeter is chosen in accordance with Δt, and the observations are filtered digitally: all frequencies above the bound Δt^{-1} are removed, because they are largely caused by the motions of the aircraft.

In practice, often the high-frequency part removed from the signal is 10 000 times larger than the gravity signal we are after!

If the uncertainty (mean error) of one GNSS position *co-ordinate* measurement is σ_x (and the different co-ordinates don't correlate with each other!), then according to equation 11.6 the uncertainty of the vertical acceleration is

$$\sigma_a = \frac{\sigma_x \sqrt{6}}{(\Delta t)^2}.$$

Making the time interval Δt as long as possible without *resolution* suffering, requires a low flight speed. Generally a propeller aircraft or even a helicopter is used. Of course the price of the measurement grows with the duration of the flight — a helicopter rotor hour is expensive!

For the flight height H we choose in accordance with resolution Δx:

$$H \sim \Delta x = v \, \Delta t,$$

where v is the flight speed. The separation between adjacent flight lines is chosen similarly.

The first major airborne gravimetry project was probably the *Greenland Aerogeophysics Project* (Brozena, 1992). In this ambitious American-Danish project in the summers of 1991 and 1992, over 200 000 km was flown, all the time measuring gravity and the magnetic field, and the height of the ice surface using a radar altimeter (Ekholm et al., 1995).

After that, also other large uninhabited areas in the Arctic and Antarctic regions have been mapped, see Brozena et al. (1996), Brozena and Peters (1994). Already in subsection 9.6.2 on page 229 we made mention of other large surveys. Activity continues, see Coakley et al. (2013), Kenyon et al. (2012). The method is well suitable for large, uninhabited areas, but also, e.g., for sea areas close to the coast where ship gravimeters would have difficulty navigating long straight tracks. In 1999 an airborne gravimetry campaign was undertaken over the Baltic, including the Gulf of Finland (Jussi Kääriäinen, personal comm.).

In addition to the economic viewpoint, an important advantage of airborne gravimetry is, that a *homogeneous coverage* by gravimetric data is obtained from a large area. The homogeneity of surface gravimetric data collected over many decades is difficult to guarantee in the same way. Also the effect of the very local terrain, which for surface measurements is a hard to remove systematic error source especially in mountainous terrain, (see section 6.3 on page 126), does not come into play for airborne gravimetry.

The operating principle of *satellite gravimetry*, e.g., GOCE (Geopotential and Steady-state Ocean Circulation Explorer) is similar. An essential difference is however, that the instrumentation on the satellite is in a state of weightlessness: $\Gamma = 0$ (in a high orbit, or when using an air drag compensation mechanism), or Γ is small and is measured using a sensitive accelerometer (in a low orbit, where air drag is notable).   ilmanvastus

The greatest challenge in planning a satellite mission is choosing the flight height. The lowest possible height is some 150 km. At that height, already a tankload of propellant is needed, or the flight will not last long. However,   ajoaine the resolution of the measurements on the Earth's surface is limited, e.g., the   erotuskyky smallest details in the Earth's gravity field "seen" by the GOCE satellite are $50 - 100$ km in diameter.

11.8　Measuring the gravity gradient

The acceleration of gravity **g** is the gradient of the geopotential W. It varies with place, especially close to masses. We speak of the gravity-gradient tensor or Eötvös tensor:

$$M \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial x \partial z} \\ \frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y^2} & \frac{\partial^2}{\partial y \partial z} \\ \frac{\partial^2}{\partial z \partial x} & \frac{\partial^2}{\partial z \partial y} & \frac{\partial^2}{\partial z^2} \end{bmatrix} W.$$

We know that gravity increases going down, at least in free air. Going up, gravity diminishes, about $0.3\,\mathrm{mGal}$ for every metre of height.

In topocentric co-ordinates (x, y, z), where z points to the zenith, this matrix is approximately

$$M \approx \begin{bmatrix} -0.15 & 0 & 0 \\ 0 & -0.15 & 0 \\ 0 & 0 & 0.3 \end{bmatrix} \mathrm{mGal}/\mathrm{m},$$

where $\frac{\partial^2 W}{\partial z^2} = \frac{\partial}{\partial z} g_z \approx 0.3\,\mathrm{mGal}/\mathrm{m}$ is the standard value for the free-air vertical gravity gradient: Newton's law gives for a spherical Earth (the minus sign is because **g** points downward while the z co-ordinate increases going up):

$$g_z = -\frac{GM}{(R+z)^2}.$$

Derivation gives

$$\frac{\partial}{\partial z} g_z = 2 \frac{GM}{(R+z)^3} \frac{\partial(R+z)}{\partial z} = -\frac{2 g_z}{(R+z)} \approx 3 \cdot 10^{-6\,\mathrm{m/s^2}}/\mathrm{m} = 0.3\,\mathrm{mGal}/\mathrm{m}.$$

The quantities $\frac{\partial^2 W}{\partial x^2}$ and $\frac{\partial^2 W}{\partial y^2}$ again describe the *curvatures* of the equipotential or level surfaces in the x and y directions, equations 4.3, 4.4:

$$\frac{\partial^2 W}{\partial x^2} = -\frac{g}{\rho_1}, \qquad \frac{\partial^2 W}{\partial y^2} = -\frac{g}{\rho_2},$$

where ρ_1 and ρ_2 are the *radii of curvature* in the x and y directions. Substitution $\rho_1 = \rho_2 = R \approx 6378\,\mathrm{km}$ yields

$$\frac{\partial^2 W}{\partial x^2} = \frac{\partial^2 W}{\partial y^2} \approx -1.5 \cdot 10^{-6\mathrm{m/s^2}}/\mathrm{m} = -0.15\,\mathrm{mGal}/\mathrm{m}.$$

The Hungarian researcher Loránd Eötvös did a number of clever experiments (Eötvös, 1998) in order to measure components of the gravity-gradient tensor with *torsion balances* built by him. The method continues to be in use in geophysical research, as the gravity gradient as a measured quantity is very sensitive to local variations in matter density in the Earth's crust.

torsiovaaka

In honour of Eötvös we use as the unit of gravity gradient the Eötvös, symbol E:

$$1\,\mathrm{E} = 10^{-9\,m/s^2}/\mathrm{m} = 10^{-4}\,\mathrm{mGal}/\mathrm{m}.$$

The above tensor is now

$$M \approx \begin{bmatrix} -1500 & 0 & 0 \\ 0 & -1500 & 0 \\ 0 & 0 & 3000 \end{bmatrix} \mathrm{E}.$$

Note that

$$\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} \approx 0,$$

the familiar Laplace differential equation. However, the equation is not exact here: in a co-ordinate system co-rotating with the Earth, the term for the centrifugal force, $2\omega_\oplus^2$, must be added, equation 4.1.

The gravity-gradient field of Sun and Moon is known on the Earth's surface as the *tidal field*, see section 14.1.

Self-test questions

1. For the spring gravimeter described in section 11.2 , one milligal of change in gravity g produces according to equation 11.3 a lengthening of $5 \cdot 10^{-8}\,\mathrm{m}$. Do a calculational check.

2. Why is a pendulum gravimeter, although theoretically absolute, not very accurate as an absolute gravimeter?

3. By which method choices do we, in practical measurements, take the drift of a relative gravimeter into account?

4. Why were, before the advent of absolute gravimeters, the reference points of international fundamental gravimetric networks often on airports?

5. What is, in an absolute or ballistic gravimeter, the role of:

   (a) the "cage" surrounding the falling prism, and

   (b) the "superspring"?

6. According to Google

   ○ the Gulf War from 1990 to 1991 was the first conflict in which the military widely used GPS,

   ○ by December 1993, GPS achieved initial operational capability (IOC), indicating a full constellation (24 satellites) was available, and

   ○ the Greenland Aerogeophysics Project, the first ever large-scale airborne gravimetric mission, mapped the gravity field of Greenland during the summers of 1991 and 1992.

Why are these three dates so close together?

Exercise 11 – 1: Absolute gravimeter

The observation process of absolute gravimetry is described by

$$z = z_0 + v_0 t + \frac{1}{2} g t^2.$$

Let us assume that the distance of falling is 30 cm.

1. How much is the time of falling?

2. If we aim at an accuracy of $\pm 10 \mu$Gal, how precisely should the laser interferometer then measure the falling distance of the prism? (A very crude order-of-magnitude guesstimate is enough!)

3. Same question for the time registration of the falling time.

Exercise 11 – 2: Spring gravimeter

When we use a spring gravimeter in the field, we place the device at every measurement station on a solid base, e.g., bedrock, for measurement, and *level* it.

Furthermore we always take care that

   ○ the device is *arrested* during transport, i.e., the beam is clamped to be motionless;

   ○ the internal temperature of the device is kept constant by a *thermostat system*.

The reason for this is that the functioning of a spring gravimeter depends on the properties of the spring material, which may change as a result of careless handling or temperature variations.

Furthermore a gravimeter always has a *drift*, i.e., the relationship between measured value and true value changes slowly over time. In a non- factory fresh gravimeter this drift is however very regular and almost linear.

As a result of the drift, a spring gravimeter cannot be used for absolute gravity measurement and is therefore called a *relative gravimeter*.

Question: how is the relative nature of a spring gravimeter and its drift taken into account

   1. in planning the topology of the measurement network?

   2. In planning the time order of the different measurements in a network?

   3. In the choice of vehicles and point locations?

———

Exercise 11 – 3: Air pressure and gravity

1. How much does a low-pressure zone of 100 hPa (i.e., the pressure is 100 hPa *lower* than average air pressure 1015 hPa) affect gravity measured on the Earth surface? (You may assume the low-pressure zone to be very extended in area.)

2. How much does sea water rise due to the "upside-down barometer effect" under a low-pressure zone?

3. How much does the effect from point 2 amount to in local gravity measured *on a ship*? On the open sea, with a free-air gravity gradient of $-0.3\,\mathrm{mGal/m}$, density of sea water $1000\,\mathrm{kg/m^3}$. Analyze the situation carefully[15].

The geoid, mean sea level, sea-surface topography

12

12.1   Basic concepts

On the ocean, the geoid is on average at the same level as *mean sea level*, the surface obtained by removing from the instantaneous sea surface all periodic and quasi-periodic variations. These variations are, for example:

- tidal phenomena, caused by Sun and Moon, order of magnitude $\pm 1\,\mathrm{m}$, locally even more

- variations caused by air pressure variations ("inverted barometer effect"). Typically of order decimetres, but up to metres under tropical cyclones

- "wind pile-up", water being pushed by winds

- littoral seas: variation in the volume of sweet water flowing out from rivers into the sea

- eddies that are formed in the oceans in connection with, e.g., the Gulf Stream and the Agulhas Stream ("mesoscale eddies") that may live for months, and inside of which the sea surface may be even decimetres above or below that of the surroundings

- the continual shifting of ocean currents from one place to another

- ENSO, El Niño Southern Oscillation, is a very long time scale, quasi-periodic weather phenomenon happening in the waters of the Pacific Ocean and the air above it, but affecting weather phenomena world-

---

[15] And I mean *really* carefully.

wide. The time scale of variability ranges from two to seven years. See figure 13.1.

If we remove all these periodic and quasi-periodic variations, we are left with *mean sea level*. If the water of the seas was in a state of equilibrium, then this mean sea surface would be an equipotential or level surface of the Earth's gravity field, the *geoid.*

This is however not how things really are. Mean sea level differs from a level surface due to the following phenomena:

○ Permanent ocean currents cause, though the Coriolis force, permanent differences in mean water level.

○ Also permanent differences in temperature and salinity cause permanent differences in mean water level, the latter, e.g., in front of the mouths of rivers.

These physical phenomena, among others, cause the so-called *sea-surface topography*, a permanent separation between sea surface and geoid.

The classical definition of the geoid is

> *"the level surface of the Earth's gravity field that agrees most closely with mean sea level."*

The practical problem with this definition is, that determining the correct level of the geoid requires knowledge of mean sea level everywhere on the world ocean. This is why many "geoid" models in practice don't coincide with global mean sea level, but with some locally defined mean sea level — and often only approximately.

Mean sea level in its turn is also a problematic concept. It is sea level from which has been computationally removed all periodic effects — but who can know if a so-called secular effect in reality is perhaps long period? A sensible compromise is the average sea level over 18 years — an important periodicity, *saros*, Wikipedia, Saros, in the orbital motion of the Moon.

meritopografia   The *sea-surface topography* again is defined as that part of the difference

between mean sea level and the geoid, which is permanent. Also here, the measure of permanency is the time series that are available, as tide gauges   mareografi have been widely operating already for about a century, when again many satellite time series — TOPEX/Poseidon and its successors — are just about a quarter of a century long. See figure 13.1.

12.2   Geoids and national height datums

A locally determined geoid model is generally *relative*. Locally, at the current state of the art, one has no access to global mean sea level at an acceptable precision. This may change with technology development.

In general, a local geoid model is tied to a *national height system*, and the difference from the classical definition is thus the same as the difference of the national height system from global mean sea level. In the case of Finland, the difference is about 30 cm, almost entirely caused by the sea-surface topography in the Baltic Sea, see figure 12.4.

In *Finland*, heights were determined for a long time in the N60 height system, which is tied to mean sea level in Helsinki harbour at the start of 1960. The reference benchmark however is located in nearby Kaivopuisto. Precise levelling disseminated heights from here all over Finland. The modern Finnish height system is N2000, which is in principle tied to sea level in Amsterdam, but the reference benchmark in Finland is similarly located at the Metsähovi research station in the Kirkkonummi municipality, West of Helsinki.

At the beginning of 1960, the reference surface of the Finnish height system N60 was an equipotential or level surface of the Earth's gravity field. However, due to post-glacial land uplift, that is no longer the case: the *post-glacial land uplift* varies from some four millimetres per year in the Helsinki area to some ten millimetres per year in the area of maximum land uplift near Ostrobothnia. This is the main reason why in Fennoscandia height systems have a "best before" date and must be modernized a couple of times per century.

Generally, geoid maps for practical use, like FIN2000, the Finnish geoid model (figure 9.5), are constructed so, that they transform heights in the national

height system, e.g., N60 heights (Helmert heights) above "mean sea level" to heights above the GRS80 reference ellipsoid. As, however, land uplift is an ongoing process, it must be tied to a certain epoch, a point in time at which the GNSS measurements were done to which the original gravimetric geoid solution has been fitted. In the case of FIN2000 this was 1997.0 (Matti Ollikainen, several sources).

Strictly speaking then, FIN2000 is not a model of the geoid. A better name might be "transformation surface". This holds true, in fact, for *all* national or regional geoid models that are built primarily for the purpose of enabling the use of GNSS in height determination ("GNSS levelling"). These "geoid-like surfaces" are constructed generally in the following way:

1. We calculate a gravimetric geoid model by using the Stokes method and Remove-Restore, e.g., by the FFT method.

2. We fit this geoid surface solution to a number of comparison points, in which both the height from levelling — "above sea level" — and from the GNSS method — above the reference ellipsoid — are known. The fit takes place, e.g., by describing the differerences by a polynomial function:

$$\delta N = a + b\left(\lambda - \lambda_0\right) + c\left(\varphi - \varphi_0\right) + ...$$

or something more complicated, and solving the coefficients a, b, c from the geoid differences in the *known comparison points* by using the least-squares method.

12.3   The geoid and post-glacial land uplift

Global mean sea level is not constant. It rises slowly by an amount that, over the past century, has slowly grown. Over the whole 20th century, the rate has been $1.5 - 2.0\,\mathrm{mm/a}$, e.g., $1.6\,\mathrm{mm/a}$ (Wöppelmann et al., 2009). Over the last couple of decades, the rate has accelerated and is now over $3\,\mathrm{mm/a}$, see figure 13.1.

This value is called the *eustatic rise of mean sea level*. It is caused partly by the melting of glaciers, ice caps and continental ice sheets, partly by thermal expansion of sea water. A precise value for the eustatic rise is hard to determine: almost all tide gauges used for monitoring sea level have their own vertical motions, and distinguishing these from the rise of sea level requires a representative geographic distribution of measurement locations. Especially the ongoing response of the solid Earth to the end of the last ice age, the latest deglaciation: so-called GIA (glacial isostatic adjustment) is a global phenomenon that it only in the latest decades has been possible to observe by satellite positioning.

Because of eustatic sea-level rise, a distinction must be made between absolute and relative land uplift:

Absolute land uplift is the motion of the Earth's crust relative to the centre of mass of the Earth. This land uplift is measured when using satellites the orbits of which are determined in a co-ordinate reference system tied to the Earth's centre of mass. For example, GNSS positioning of tide gauges.

Relative land uplift is the motion of the Earth's crust relative to mean sea level. This motion is measured by tide gauges, also called *mareographs*.

Geoid rise: as the post-glacial land uplift is the shifting of masses internal to the Earth from one place to another, it is clear that also the geoid must change. The geoid rise is however small compared to the land uplift, only a few percent of it.

Equation (the point above a quantity denotes the time derivative $\frac{d}{dt}$):

$$\dot{h} = \dot{H}_r + \dot{H}_e + \dot{H}_t + \dot{N},$$

in which

$\dot{h}$ is the absolute land uplift,

$\dot{H}_r$ is the relative land uplift,

$\dot{H}_e$ is the eustatic (mean sea level) rise,

$\dot{H}_t$ is the change over time of the sea-surface topography (likely small),

$\dot{N}$ is the geoid rise.

The rise in the geoid as a result of land uplift can be simply calculated with the Stokes equation:

$$\frac{dN}{dt} = \frac{R}{4\pi\gamma} \iint_\sigma S(\psi) \left(\frac{d}{dt}\Delta g\right) d\sigma.$$

Here, $\frac{d}{dt}\Delta g$ is the change of gravity anomalies over time due to land uplift. Unfortunately we do not precisely know the *mechanism* by which mass flows into the land uplift area in the Earth's mantle. We may write

$$\frac{d}{dt}\Delta g = c\frac{dh}{dt},$$

in which the constant c may range from -0.16 to -0.31 mGal/m.

- The value -0.16 mGal/m is called the "Bouguer hypothesis": it corresponds to the situation where upper mantle material flows into the space freed up underneath the rising Earth's crust, in order to fill it.
- The value -0.31 mGal/m is the opposite extreme, the "free-air hypothesis". By this hypothesis, the ice load during the last ice age has only compressed the Earth's mantle, and now it is slowly expanding again into its former volume ("rising dough model").

Up until fairly recently, the most likely value was about -0.2 mGal/m, with substantial uncertainty. The latest results (Mäkinen et al., 2010) give -0.16 ± 0.02 mGal/m (one standard deviation), which would seem to settle the issue. It looks like the Bouguer hypothesis is closer to physical reality. The flow of mass happens probably within the *asthenosphere*.

This problem has been studied much in the Nordic countries. The method has been gravimetric measurement along the 63°N parallel ("Blue Road Geotraverse" project). The measurement stations extend from the Norwegian coast to the Russian border, and have been chosen so, that gravity along them varies within a narrow range. In this way, the effect of the scale error of the gravimeters is avoided. Clearly, absolute gravity is of no interest here, only the *change* in gravity *differences* over time between the stations.

These measurements have been made over many years using high precision spring or relative gravimeters. In recent years, there has been a shift to using absolute gravimeters, obviating the need for measurement *lines*.

(a)
Bouguer hypothesis...

(b)
...and free-air hypothesis.
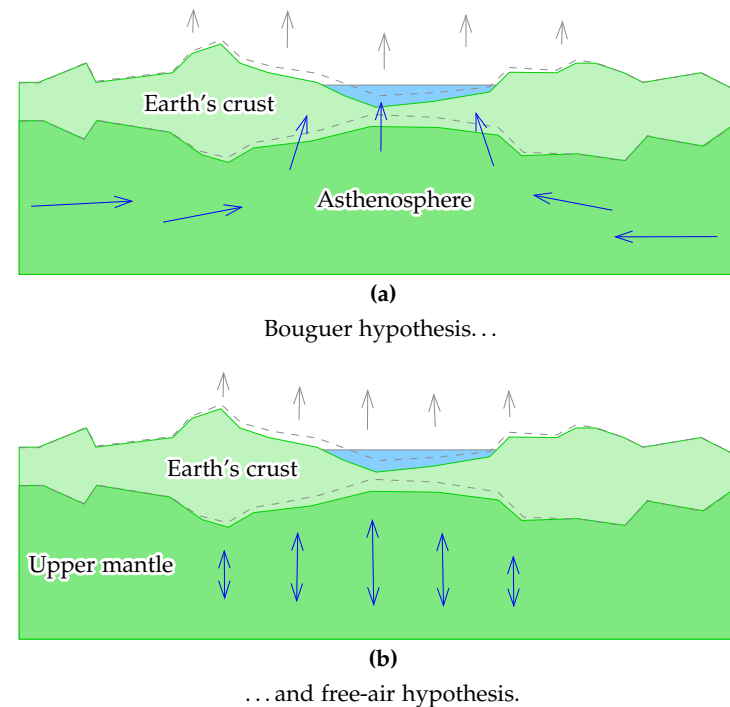
Figure 12.1. The two different hypotheses on the mechanism of post-glacial land uplift.

12.4   Methods for determining the sea-surface topography

In principle three geodetic methods exist:

1. satellite radar altimetry and gravimetric geoid determination
2. GNSS positioning along the coast (tide gauges) and gravimetric geoid determination
3. precise levelling along the coast.

In addition to this, we still have the oceanographic method, i.e., physical
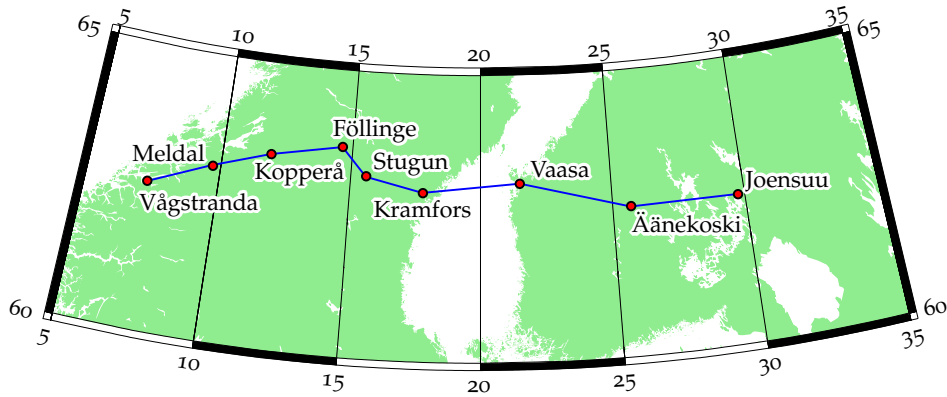
Figure 12.2. The Fennoscandian 63°N parallel gravity line.

modelling. The method is termed *steric levelling* if temperature and salinity measurements along vertical profiles are used on the open ocean, and *geostrophic levelling* if ocean current measurements are used to determine the Coriolis effect, generally close to the coast.

All methods should give the same results. The Baltic Sea is a textbook example, where all three methods have been used. A result was that the whole Baltic Sea surface is tilted: relative to a level surface, the sea surface goes up from the Danish straits to the bottoms of the Gulf of Finland and the Bothnian Bay by some $25 - 30$ cm.

Oceanographic model calculations show, that this tilt is mainly due to a *salinity gradient*: in the Atlantic Ocean, salinity is $30 - 35$ o/oo, when in the Baltic it drops to $5 - 10$ o/oo due to the massive production of sweet water (Ekman, 1992). Of course on top of this come temporal variations, like oscillations like in a bathtub, the amplitude of which can be over a metre.

In Ekman (1992) more is said about the sea-surface topography of the Baltic and its determination.

12.5   Global sea-surface topography and heat transport

One important reason why researchers are interested in the global sea-surface topography, is that it offers an opportunity to study more precisely the currents in the oceans and thus the transport of the Sun's energy from the equator to higher latitudes. There are many things the study of which is helped by precise knowledge of ocean currents: carbon dioxide dissolved into the water, chlorophyll (phytoplankton), salinity, etc.

The Coriolis force, or *acceleration*, caused by the Earth's rotation is:

$$\mathbf{a} = 2\left\langle \mathbf{v} \times \overrightarrow{\omega_{\oplus}} \right\rangle, \tag{12.1}$$

in which $\mathbf{v}$ is the velocity vector in a system attached to the rotating Earth, and $\overrightarrow{\omega_{\oplus}}$ is the rotation vector of the Earth.

If a fluid flows on the Earth's surface, then, in the above equation 12.1, only the part of $\overrightarrow{\omega_{\oplus}}$ normal to the surface will have an effect: this part has a length of $\left\langle \overrightarrow{\omega_{\oplus}} \cdot \mathbf{n} \right\rangle = \omega_{\oplus} \sin \varphi$, and the vector equation 12.1 may be replaced by a simpler scalar equation:

$$a = 2v\omega_{\oplus} \sin \varphi,$$

where $a \overset{\text{def}}{=} \|\mathbf{a} - \langle \mathbf{a} \cdot \mathbf{n} \rangle\|$, i.e., the length of the projection of $\mathbf{a}$ onto the tangent plane to the Earth, and $v \overset{\text{def}}{=} \|\mathbf{v}\|$, $\omega_{\oplus} \overset{\text{def}}{=} \|\overrightarrow{\omega_{\oplus}}\|$ etc. in the familiar way. The direction of the Coriolis acceleration is always perpendicular to the flow velocity: when watching along the flow direction, to the right on the Northern hemisphere, to the left on the Southern hemisphere.

As a result of the Coriolis force, the sea surface in the area of an ocean current is *tilted sideways* with respect to the current, at an angle

$$\frac{a}{\gamma} = \frac{2v\omega_{\oplus}}{\gamma} \sin \varphi.$$

This equilibrium between Coriolis force and the horizontal gradient of pressure is called the *geostrophic equilibrium*. On the equator it can be seen from the equation that the tilt is zero, but everywhere else, ocean currents are tilted. For example, in case of the Gulf Stream, the height variation caused by this effect is several decimetres. If we define a local (x, y) co-ordinate system
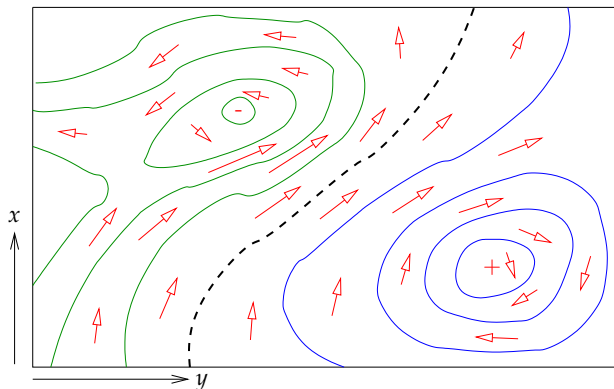
Figure 12.3. Connection between sea-surface topography and ocean currents. Red arrows depict the ocean currents; curves, the sea-surface topography.

in which $x(\varphi, \lambda)$ is pointing North and $y(\varphi, \lambda)$ East, we may write for the sea-surface topography H the geostrophic equations

$$\frac{\partial H}{\partial x} = -2v_y \frac{\omega_\oplus}{\gamma} \sin \varphi, \qquad \frac{\partial H}{\partial y} = +2v_x \frac{\omega_\oplus}{\gamma} \sin \varphi, \qquad (12.2)$$

As we will see in chapter 13, we can measure the location in space of the sea surface at this precision using satellite radar altimetry. If we furthermore have a precise geoid map, we may calculate the sea-surface topography, and with the aid of equations 12.2 solve for the flow velocity field[1]

$$\begin{bmatrix} v_x(x,y) & v_y(x,y) \end{bmatrix}^T = \begin{bmatrix} v_x(\varphi, \lambda) & v_y(\varphi, \lambda) \end{bmatrix}^T.$$

An elegant property of these equations is, that we don't even have to know the absolute level of the field $H(x,y) = H(\varphi, \lambda)$ because that vanishes in differentiation.

---

[1] A popular, though unofficial, unit for ocean current is the sverdrup (Wikipedia, Sverdrup), a million cubic metres per second. All the rivers of the world together make about one sverdrup, while the Gulf Stream is $30 - 150\,\mathrm{Sv}$.
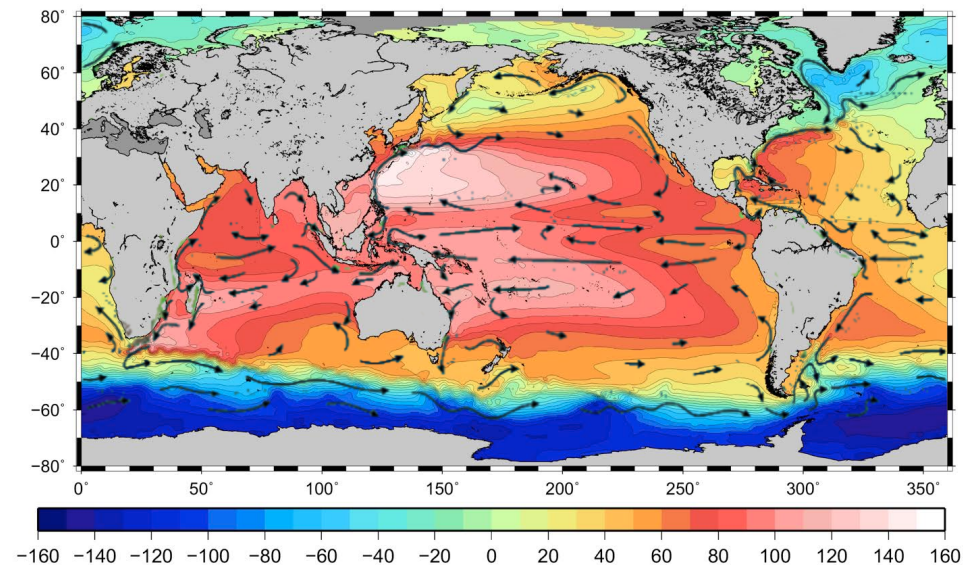
Figure 12.4. Sea-surface topography map produced by GOCE. © European Space Agency. Unit: cm. Ocean currents superimposed: NOAA / Rick Lumpkin (Lumpkin and Garraffo, 2005).

The method described, figure 12.3, requires a sufficiently precise geoid map of the oceans of the world. To this need, the GOCE satellite fits like a glove, see section 13.7. One objective of the mission was, as the name indicates, to get a full picture of ocean currents and especially their capacity for heat transport. This knowledge helps understand how the Earth's climate functions and how it is changing, also as a result of human activity. This is for Europe and Fennoscandia, and also Finland, an important issue, as the heat energy brought by the Gulf Stream helps to keep these areas habitable.

Even without a geoid model we can study, using satellite altimetry, the *variations* of ocean currents. It has been known for long that in the North Atlantic Ocean, *mesoscale eddies* have been moving alongside the Gulf Stream, eddies of size $10 - 100\,\mathrm{km}$ which show up in altimetric imagery. Interesting is that the eddies also show up in maps of the ocean surface temperature, and

biologists have observed that life inside the eddies differs from that outside (Godø et al., 2012). The life span of the eddies can be weeks, even months.

A good, though somewhat dated, introduction into "geodetic oceanography" is given by Rummel and Sansó (1992).

12.6   The global behaviour of sea level

Water exists on the Earth in three phases: liquid, ice, and vapour. During geological history, especially the ratio between liquid water and ice has varied substantially. Also today, a large amount of ice is tied up in continental ice

mannerjäätikkö   sheets, specifically Antarctica and Greenland. Of these, Eastern Antarctica is the overwhelmingly largest.

When the amount of water tied up in continental ice sheets varies, so does sea level. The end of the last ice age has raised sea level by as much as 120 m, a process that ran to completion some 6000 years ago[2] (Wikipedia, Sea level rise). Not until the last century or two has sea level again started rising, and the rise accelerating, as a consequence of global warming.

We still live in the aftermath of the last glaciation; there were large continental

mannerjäätikkö   ice sheets which have since molten away, like in Fennoscandia and Canada (the so-called Laurentide ice sheet), the land is still rising at an even pace, in places even 10 millimetres per year. Around the land uplift areas, in central

maan vajoaminen   Europe and the United States, again takes place a *subsidence* of the land at an annual pace of $1 - 1.5$ mm, as directly underneath the hard crust of the Earth or lithosphere, in the upper mantle layer called *asthenosphere*, material is flowing slowly inward under the rising Earth's crust.

In order to complicate the picture, the sea-level rise caused by the melting of continental ice sheets also presses the ocean floor down — by as much as 0.3 mm per year, the so-called Peltier effect (Peltier, 2009). Therefore, the measured sea-level rise — whether on the coast by tide gauges, or from space using satellite altimetry — *does not represent the whole change in total ocean*

---

[2]6000 years "before present", 6 ka BP. *BP* conventionally means: before 1950. Nowadays is also used *b2k*, before the year 2000.
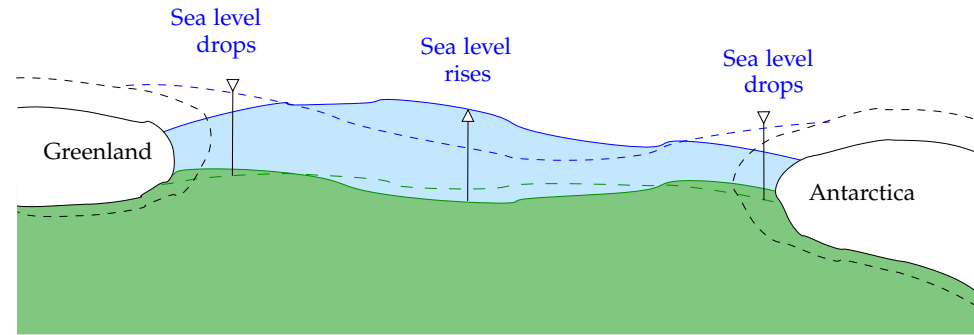
Figure 12.5. The sea-level equation. Sea level reacts in a complicated way when continental ice sheets melt.

water volume. If that is what interests us, as it always does in climate research, this Peltier correction must still be added to the observation values.

The subsidence of the sea floor hasn't even been globally uniform: at the edges of the continents happens a "lever motion" when the sea floor subsides   *vipuliike* but dry land doesn't. And in the tropics in the Indian and Pacific Oceans, sea level reached 6000 years ago its maximum level, the so-called *mid-Holocene highstand*, relative to the Earth's crust. After this, local sea level has subsided and the coral formations from that age have remained, dead, some $2 - 3$ m above modern sea level. This is how, e.g., Tuvalu and the Maldives were formed, which are now again being threatened by modern sea-level rise.

12.7   The sea-level equation

Scientifically the variations in sea level are studied using the *sea-level equation*. A pioneer in this field has been the Canadian W. Richard Peltier (W. R. Peltier, FRSC, Home Page), who has constructed physics-based models of how both the solid Earth and sea level respond when the total mass of the continental ice sheets changes.

The sea-level equation (Selen User Manual) is

$$S = S_\mathrm{E} + \frac{\rho_\mathrm{i}}{\gamma_0} \left[G_\mathrm{s} \circledast_\mathrm{i} I - \overline{G_\mathrm{s} \circledast_\mathrm{i} I} \right] + \frac{\rho_\mathrm{o}}{\gamma_0} \left[G_\mathrm{s} \circledast_\mathrm{o} S - \overline{G_\mathrm{s} \circledast_\mathrm{o} S} \right], \qquad (12.3)$$

in which

- $S = S(\omega, t) = S(\varphi, \lambda, t)$ describes the variations of sea level as a function of place $\omega = (\varphi, \lambda)$ and time t,

- $I = I(\omega, t)$ is similarly a function of place and time describing the geometry of ice sheets and glaciers,

- S_E is the *eustatic term*, i.e., the variation in ice volume converted to "equivalent global sea-level variation", in an equation

$$S_E(t) = \frac{m_i(t)}{\rho_o A_o},$$

in which $m_i(t)$ is the variation in total ice mass as a function of time, ρ_o the density of sea water, and A_o the total surface area of the oceans,

- ρ is the density of matter: ρ_i that of ice, and ρ_o that of sea water,

- $\circledast$ is the symbol of a convolution on the surface of the Earth and the time axis, $\circledast_i$ over land ice, $\circledast_o$ over the oceans — i.e., *Green's function* is multiplied with the ice and sea functions and integrated over the domain in question. These integrals are by the way very similar to the ones discussed in section 8.1, e.g.:

$$\{G_s \circledast_o S\}(\omega, t) = \int_{-\infty}^{t} \iint_{ocean} G_s\{\psi(\omega, \omega'), \{t - t'\}\} S(\omega', t')\, d\omega'\, dt',$$

in which $\psi(\omega, \omega')$ is the geocentric angular distance between evaluation point $\omega = (\varphi, \lambda)$ and data point $\omega' = (\varphi', \lambda')$. The measure of the surface integral is $d\omega = N(\varphi)\, M(\varphi)\, \cos\varphi\, d\varphi\, d\lambda$, in which $N, M \approx R$ are the principal radii of curvature of the Earth ellipsoid. As can be seen, we have here a convolution applied *both* over the Earth's surface ω *and* over the time axis t.

- The *overbar* designates *averaging* over the whole ocean surface,

- γ_0 is an average acceleration of gravity,

- G_s is the so-called *Green's function of sea level*:

$$G_s = G_V - \gamma_0 G_u,$$

where the *Green's function of the geopotential* is

$$G_V(\psi, t) = G_V^r(\psi, t) + G_V^e(\psi, t) + G_V^v(\psi, t)$$
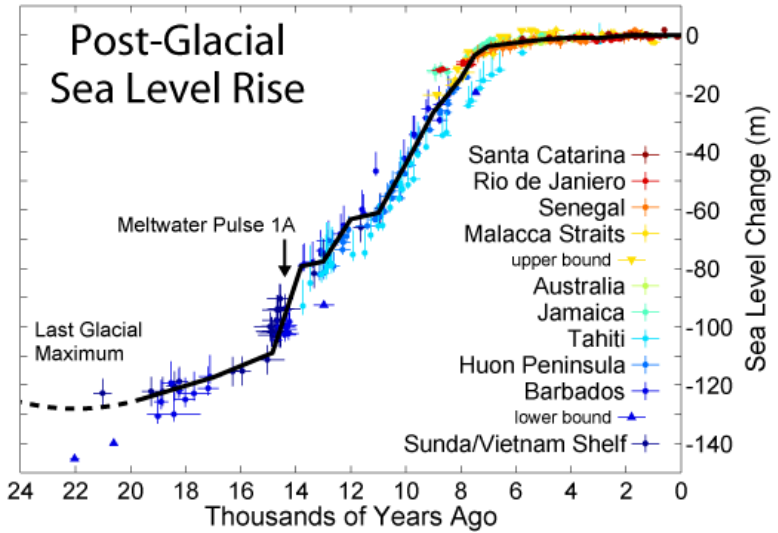
Figure 12.6. Sea-level rise after the last ice age (Wikimedia Commons, © Robert A. Rohde, GNU Free Documentation License).

where again ψ is the distance of evaluation point from integration point, and G_V^r, G_V^e ja G_V^v are the rigid, elastic and plastic ("viscous") partial Green's functions of deformation. These thus describe the *rheological behaviour* of the Earth, and their theoretical calculation requires the internal viscosity distribution $\eta(r)$ of the Earth, assuming it is *isotropic*, i.e., only dependent upon r.

$$G_u(\psi, t) = G_u^e(\psi, t) + G_u^v(\psi, t)$$

is again similarly *Green's function function of vertical displacement*, in the same way split into elastic and plastic parts.

The behaviour of sea level can now be computed in this way, that one first tries to construct an "ice-load history", i.e., $I(\omega, t)$. Then, from this one tries to calulate iteratively, using the sea-level equation 12.3, $S(\omega, t)$. Note that S describes *relative* sea-level variation, i.e., changes in the relative positions of sea level and the Earth's solid body or Earth's crust. It is a *function of*

place: one may not assume that it would be the same everywhere. In the article Mitrovica et al. (2001) it is shown how, e.g., the melting water from Greenland flees to the Southern hemisphere, when the melting water from Antarctica again comes similarly to the North. This is a consequence from the change in the *Earth's gravity field* and the geoid, when large volumes of ice melt. Another factor is that also the physical shape of the Earth changes, when the ice load changes: so-called *glacial isostatic adjustment*, GIA.

This also complicates the monitoring of global mean sea level from local measurements: the problem is familiar in Fennoscandia, as the Earth's crust, for now, moves up faster than global sea-level rise...

Green's functions in the sea-level equation are functions of both ψ and time difference Δt; this tells us that GIA is a function of both place and time. On a spherically symmetric Earth, the functions may be written as *expansions*, e.g.[3]

$$G_V^{\mathrm{v}}(\psi, \Delta t) = H(\Delta t) \frac{R \gamma_0}{M} \sum_{\ell=1}^{\infty} \left(\sum_{i=1}^{I} k_{\ell i} e^{-s_{\ell i} \Delta t} \right) P_\ell(\cos \psi),$$

in which H is a step function, the "Heaviside function", which is zero for negative and $+1$ for positive arguments. The index i counts the so-called *viscous relaxation modes* for every degree number ℓ; $k_{\ell i}$ are "viscous loading deformation coefficients" and $\tau_{\ell i} = \frac{1}{s_{\ell i}}$ correspond to the *relaxation times* in which the mode in question will decay over time. Generally the modes that are of large spatial extent — i.e., low ℓ values — decay slower, when again the local modes — high ℓ values — tend to decay faster, and the local modes of the last deglaciation have today already vanished: the geographic pattern of the Fennoscandian land uplift is already very smooth, and the seismicity accompanying the deglaciation is pretty much over. Back then, immediately after the retreat of the ice sheet at its edge, there were strong earthquakes, the traces of which are visible in the landscape (Kuivamäki et al., 1998). The now dominant viscoelastic modes are many hundreds of kilometres in geographic extent, and correspondingly of time scales thousands of years.

---

[3]We consider here only the plastic or *viscous* deformation.

Self-test questions

1. List all the causes you are aware of of sea-level variations.
2. What is the sea-surface topography?
3. What is eustatic sea-level rise?
4. What is absolute, what relative land uplift? What does the difference between the two consist of?
5. What two main models are on offer for the mechanism of land uplift?
6. What three geodetic techniques are available for determining the sea-surface topography?
7. What is the shape of the sea-surface topography of the Baltic Sea, and what is its cause?
8. What is the Coriolis force, and how does it affect ocean currents?
9. What is the geostrophic balance?
10. How can one invert a map of the sea-surface topography into a map of ocean currents? Where on Earth does this *not* work?
11. What is the Peltier effect? What is the mid-Holocene highstand?
12. What does the sea-level equation describe?
13. Why does mean sea level in the Baltic Sea not rise when the Greenland continental ice sheet melts? What will happen in the Baltic Sea when the West Antarctic ice sheet melts?

Exercise 12 − 1: Coriolis force, ocean current

If the velocity of flow of an ocean current is 0.1 m/s and its width 100 km, compute:

1. How much at latitude 45° N is the height difference between its left and right edges?
2. If the same current was 200 km broad and the velocity of flow 0.05 m/s (i.e., assuming the same depth, also the transport of water would be the same), compute for that case the height difference between the left and the right edges.

318   The geoid, mean sea level, sea-surface topography

3. [For fun] if the depth of the current is 1 km, what is the water transport in sverdrup?

Exercise 12 − 2:   Land subsidence and the mechanism of land uplift

How does the post-glacial land subsidence observed in the United States and central Europe support a Bouguer type of land-uplift mechanism (figure 12.1a), rather than a free-air mechanism?

Satellite altimetry and satellite gravity missions

13

13.1   Satellite altimetry

Satellite altimetry measures, using microwave radar, the distance from a satellite straight downward to the sea surface. Historically there have been many satellites carrying an altimetry radar, see table 13.1, which may not be complete.

- The GEOS-3 (1975-027A) and Seasat satellites were American testing satellites aimed at developing the altimetric technique. The measurement precision of GEOS-3 was still rather poor. Before that, satellite altimetry was also tested with a device on board the orbital laboratory Skylab (1973-027A).
- Seasat (1978-064A) broke down only three months after launch, probably due to a short-circuit[1]. However, the data from Seasat was the first large satellite altimetry data set used for determining the mean sea surface, also of the Baltic Sea.
- Geosat (1985-021A) was a satellite launched by the U.S. Navy, intended to map the gravity field on the world's oceans, more precisely the deflections of the plumb line, which are needed to impart the correct departure direction to ballistic missiles launched from submarines. The 17-day repeat data from the geodetic mission was initially classified.

[1]But read this: Wikipedia, Seasat conspiracy theory.

Table 13.1. Altimetric satellites through the ages.

Satellite	Launch year	Orbital inclination (°)	Orbital height (km)	Repeat periods (days)	Measurement precision (m)	Positioning technique
GEOS-3	1975	115.0	843	—	0.20	
Seasat	1978	108.0	780	3 (17)	0.08	
Geosat	1985	108.0	780	3, 17	0.04	
ERS-1	1991	98.5	780	3, 35, 2 × 168	0.03	
TOPEX/Poseidon	1992	66.0	1337	10	0.033	GPS, DORIS
ERS-2	1995	98.5	780	35	0.03	PRARE
Geosat follow-on	1998	108.0	800	17	0.035	
Envisat	2001	98.5	784	35	0.045	GPS, DORIS
Jason-1	2001	66.1	1336	9.9156	0.025	GPS, DORIS
Jason-2	2008	66.04	1336	9.9156	0.025	GPS, DORIS
Cryosat-2	2010	92.0	725	369, 30		DORIS
HY-2A	2011	99.3	970	14, 168	0.085	DORIS, GPS
SARAL/AltiKa	2013	98.5	781	35		DORIS
Jason-3	2016	66.04	1338	9.9927	0.025	GPS, DORIS
Sentinel-3A	2016	98.62	804	27	0.03	DORIS, SLR, GNSS

Later however, the data from the Southern hemisphere was published for scientists to use, and currently the whole data set is public.

○ The satellites ERS-1/2 (1991-050A, 1995-021A) and Envisat (2002-009A) were launched by ESA, the European Space Agency. The altimeter was just one among many packages. On the ERS satellites a German positioning device called PRARE was along, but only on ERS-2 it functioned after launch.

○ TOPEX/Poseidon (1992-052A) was an American-French collaboration, one goal of which was to precisely map the *sea-surface topography*. A special feature was the on-board precise GPS positioning device, which allowed the determination of the location of the sea surface

geocentrically. Together with her successors Jason-1, 2 and 3 (2001-055A, 2008-032A, 2016-002A), this satellite mission has also produced, and continues to produce, valuable information on the global rise of sea level over the last 20 years, about 3 mm per year. See figure 13.1.

The famous oceanographer Walter Munk (1917–2019) described TOPEX/Poseidon in 2002 as "the most successful ocean experiment of all time" (Wikipedia, TOPEX/Poseidon).

○ HY-2A (2011-043A) is a Chinese satellite also launched by China.

○ SARAL/AltiKa (2013-009A) is a satellite launched by India. The altimeter and DORIS are French contributions.

○ Cryosat-2 (2010-013A) is a satellite launched by the ESA, the European Space Agency, to study polar sea ice. Of special interest is the so-called *freeboard*, the amount by which the ice sticks out of the water. From this, the thickness, and with surface area, the total volume may be calculated. The launch of Cryosat-1 failed. In-orbit positioning is done with the French DORIS system.

○ Sentinel 3A (2016-011A) is a versatile ESA remote-sensing satellite, the first of a planned constellation. She carries several instruments, among them SRAL, Synthetic Aperture Radar Altimeter.

The measurement method of satellite radar altimetry is depicted in figure 13.2. Here we see all the quantities that are along in altimetry: the measured range ℓ is the height h of the satellite above the reference ellipsoid, corrected for the geoid height N, sea-surface topography H, and variations of the sea surface, like tides, eddies, annual variation etc.

Furthermore, if the satellite does not contain a precise positioning device, the true orbit of the satellite will differ from the calculated orbit — even from the orbit calculated afterward. Therefore

$$h_{\text{sat}} = h_{0,\text{sat}} + \Delta h,$$

where $h_{0,\text{sat}}$ is the calculated orbit, and Δh the orbit-error correction.

The measurements are performed by sending thousands of pulses per second down, measuring and averaging the travel times of the reflected return
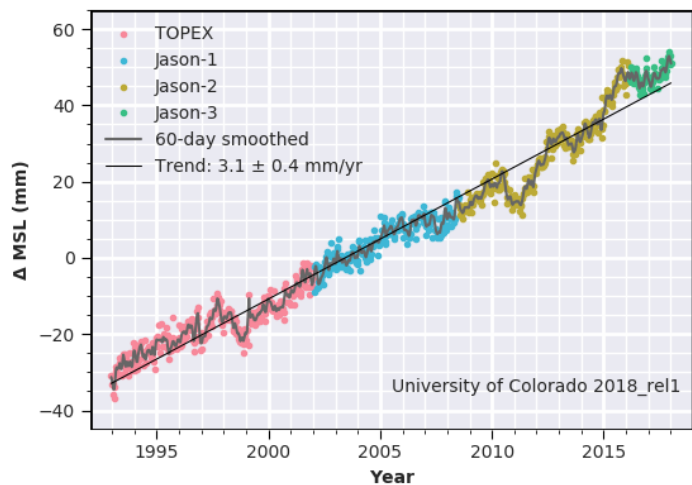
(a)
Global mean sea-level rise
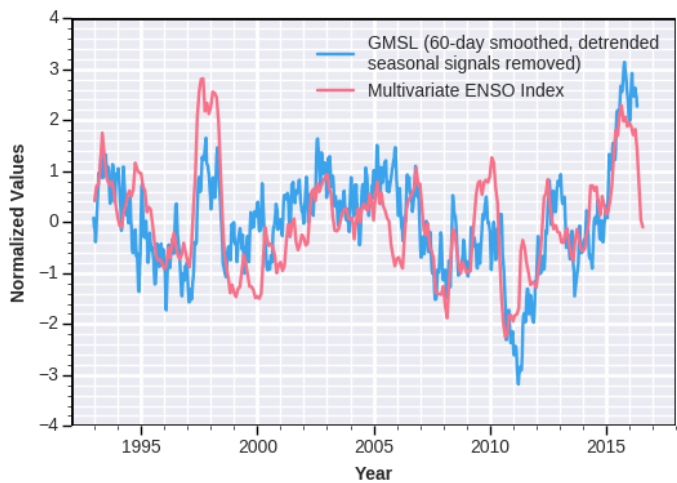
(b)
Correlation between global mean sea level and ENSO ("El Niño")

Figure 13.1. Results from the TOPEX/Poseidon and Jason satellites. © Colorado University at Boulder's Sea Level Research Group; Nerem et al. (2010).
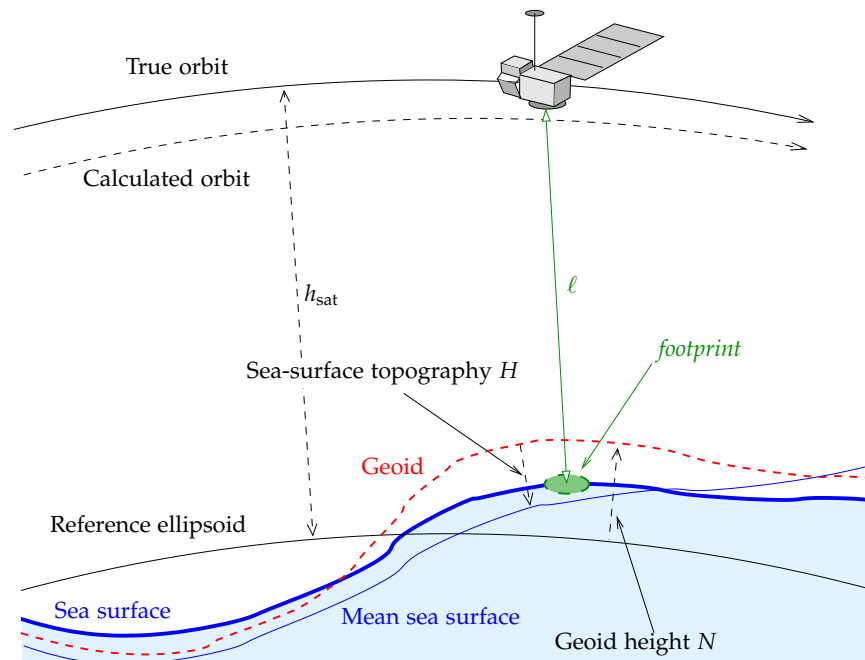
Figure 13.2. Satellite altimetry as a measurement method; concepts.

pulses on-board the satellite down to a rate of 10 – 20 values per second, and transmitting these to Earth. Of these values, the largest and smallest are thrown away as possibly erroneous, and from the remainder, a mean value is calculated for the central epoch of the pulse train using linear regression. The value thus obtained from the regression line is the actual "measurement": one every second, making the effective measurement frequency 1 Hz.

The details will vary from satellite to satellite. The pulse shape is never quite crisp; the place of the reflection on the ocean surface, or *footprint*, has a diameter of several kilometres. Especially if the ocean has wave motion (*significant wave height*, SWH), then in the processing phase one should make a careful correction so no bias is created: if the SWH is large, also the altimeter footprint — the area on the sea from which radio energy returns to the

receiver — will be larger, and the distance travelled by the radio waves will on average be a little longer.

Newer satellites use an interferometric technique that differs somewhat from the description above.

Of all the corrections related to instrumentation, atmosphere, ocean, and solid Earth, we mention

- the height of sea waves (SWH)
- solid-Earth tides
- ocean tides
- the "wet" tropospheric propagation delay, best computed from water vapour content measured with a water vapour radiometer on the satellite, otherwise from an atmospheric model
- the "dry" tropospheric propagation delay
- the ionospheric delay, only for the part of the ionosphere below the satellite, depending on flight height
- the altimeter's own calibration correction. Nowadays "in-flight" calibration is always strived for, using an ensemble of GNSS-positioned tide gauges.

The measurements and all corrections to be made to them are collected into a "geophysical data record" (GDR), one per observation epoch. The files built this way are distributed to researchers. This allows all kind of experimentation, e.g., the replacement of a correction by one calculated from improved models, etc.

13.2 Crossover adjustment

When a satellite orbits the Earth over months or years, thousands of points are formed where the tracks cross each other. If we assume that sea level is the same for both satellite overflights, then this forms a *condition* that can be used to adjust orbit errors.

The observation equations are

$$\underline{h}_a = N + H + \Delta h + \epsilon + \underline{n},$$

in which $\underline{h}_a$ is the altimetric measurement of the height of the sea surface, N is the geoid height, H is the *sea-surface topography*, or the permanent deviation of the sea surface from an equipotential surface, Δh is the orbit-error correction, ϵ is the variability of the sea surface due to, e.g., the tides among other causes, and $\underline{n}$ is the noise in the radar altimetry observations.

From this we obtain in the crossing point of tracks i and j:

$$\underline{\ell}_k \stackrel{\text{def}}{=} \underline{h}_a^i - \underline{h}_a^j = \left(\Delta h_i - \Delta h_j\right) + \left(\epsilon_i - \epsilon_j\right) + \left(\underline{n}_i - \underline{n}_j\right).$$

This is the observation equation of crossover adjustment. Here we see the complication that in both, sea-surface variability and orbit errors are along in the equation in the same way. They cannot be separately determined by crossover adjustment.

If we forget for now the sea-surface variability — or assume that it behaves randomly, in other words it is part of the noise $\underline{n}$ — we may write more simply

$$\underline{\ell}_k = \Delta h_i - \Delta h_j + \underline{n}_k.$$

The index k counts crossover points, the indices i, j count tracks.

Next, we choose a suitable *model* for the satellite orbit error. The simplest choice, sufficient for a small area, is the assumption that the orbit error is a constant for each track. See a simple example, figure 13.3.

13.2.1 A simple example

In the figure we have three tracks and two crossing points. The *observation equations*, which describe the discrepancies in the known crossover points as functions of the orbit errors, are

$$\underline{\ell}_1 = \Delta h_2 - \Delta h_3 + \underline{n}_1$$
$$\underline{\ell}_2 = \Delta h_1 - \Delta h_3 + \underline{n}_2$$
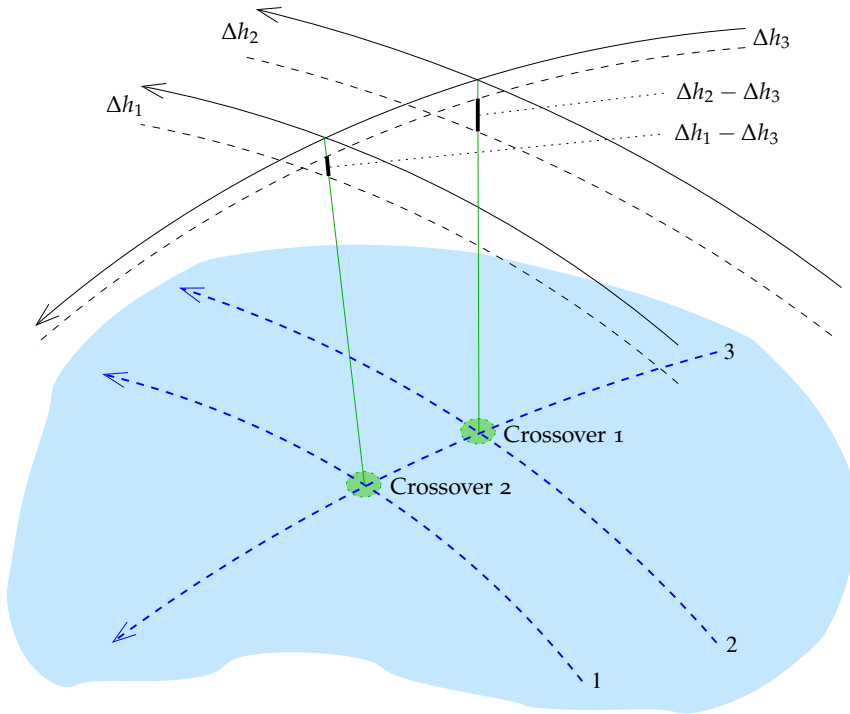
Figure 13.3. A simple crossover geometry.

or in matric form[2]

$$
\overbrace{\begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}}^{\ell} = \overbrace{\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \Delta h_3 \end{bmatrix}}^{\mathsf{x}} + \overbrace{\begin{bmatrix} n_1 \\ n_2 \end{bmatrix}}^{\mathsf{n}}, \tag{13.1}
$$

symbolically

$$
\ell = A\mathsf{x} + \underline{n}.
$$

_____

[2]Note the similarity with the observation equations for levelling! Instead of benchmarks, we have tracks, instead of levelling lines, crossover points.

If one now tries to calculate the solution with ordinary least squares,

$$
\widehat{\mathsf{x}} = \left(A^\mathsf{T} A \right)^{-1} A^\mathsf{T} \underline{\ell},
$$

one will notice that this doesn't work. The *normal matrix* $A^\mathsf{T} A$ is *singular* (check!). This makes sense, as one can move the whole track network up or down without the observations $\underline{\ell}_k$ changing. No unique solution can be found for such a system.

Finding a solution requires that *something* must be fixed. For example, one track, or, more democratically, the mean level of all tracks. This fixing is achieved by *adding* the following "observation equation":

$$
\ell_3 \overset{\text{def}}{=} 0 = \begin{bmatrix} c & c & c \end{bmatrix} \cdot \mathsf{x}, \tag{13.2}
$$

where c is some suitable constant. Then the matrix A becomes

$$
A = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix},
$$

and the least-squares solution

$$
\widehat{\mathsf{x}} = \begin{bmatrix} \widehat{\Delta h_1} \\ \widehat{\Delta h_2} \\ \widehat{\Delta h_3} \end{bmatrix} = \left(A^\mathsf{T} A \right)^{-1} A^\mathsf{T} \underline{\ell} = \left(A^\mathsf{T} A \right)^{-1} A^\mathsf{T} \begin{bmatrix} \ell_1 \\ \ell_2 \\ 0 \end{bmatrix},
$$

where the matrix inversion is now possible. Note that $\widehat{\mathsf{x}} = A^{-1}\underline{\ell}$ will in this case give the same solution, as A is square and invertible:

$$
\left(A^\mathsf{T} A \right)^{-1} A^\mathsf{T} \underline{\ell} = A^{-1} \left(A^\mathsf{T} \right)^{-1} A^\mathsf{T} \underline{\ell} = A^{-1} \left(\left(A^\mathsf{T} \right)^{-1} A^\mathsf{T} \right) \underline{\ell} = A^{-1}\underline{\ell}.
$$

Now the symbolic algebra system maxima (SourceForge, Maxima) — or brute-

force calculation — gives the readily verified inverse

$$
A^{-1} = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix}^{-1} =
$$

$$
= \left[\begin{bmatrix} 1 & & \\ & 1 & \\ & & c \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \right]^{-1} =
$$

$$
= \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & & \\ & 1 & \\ & & c \end{bmatrix}^{-1} =
$$

$$
= \frac{1}{3} \begin{bmatrix} -1 & 2 & 1 \\ 2 & -1 & 1 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & & \frac{1}{c} \end{bmatrix} =
$$

$$
= \frac{1}{3} \begin{bmatrix} -1 & 2 & \frac{1}{c} \\ 2 & -1 & \frac{1}{c} \\ -1 & -1 & \frac{1}{c} \end{bmatrix},
$$

and the solution is

$$
\begin{bmatrix} \widehat{\Delta h_1} \\ \widehat{\Delta h_2} \\ \widehat{\Delta h_3} \end{bmatrix} = A^{-1} \underline{\ell} = \frac{1}{3} \begin{bmatrix} -1 & 2 & \frac{1}{c} \\ 2 & -1 & \frac{1}{c} \\ -1 & -1 & \frac{1}{c} \end{bmatrix} \begin{bmatrix} \ell_1 \\ \ell_2 \\ 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} -1 & 2 \\ 2 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix},
$$

from which c has vanished.

Another way to look at this is to first write the observation equations 13.1 and 13.2 together as

$$
\overbrace{\begin{bmatrix} \ell_1 \\ \ell_2 \\ 0 \end{bmatrix}}^{\underline{\ell}} = \overbrace{\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ c & c & c \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \Delta h_3 \end{bmatrix}}^{\times} + \overbrace{\begin{bmatrix} n_1 \\ n_2 \\ 0 \end{bmatrix}}^{\underline{n}},
$$

and then multiply both sides, and both terms on the right, with the diagonal matrix

$$
D \overset{\text{def}}{=} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{c} \end{bmatrix}.
$$

The result is

$$
\overbrace{\begin{bmatrix} \ell_1 \\ \ell_2 \\ 0 \end{bmatrix}}^{D\underline{\ell}} = \overbrace{\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}}^{DA} \begin{bmatrix} \Delta h_1 \\ \Delta h_2 \\ \Delta h_3 \end{bmatrix} + \overbrace{\begin{bmatrix} n_1 \\ n_2 \\ 0 \end{bmatrix}}^{D\underline{n}},
$$

from which also c has vanished.

The principle applies generally:

> *Minimal constraints added to observation equations with a datum defect do not essentially change the solution.*

13.2.2   A more advanced orbit correction model

A more advanced representation of orbit errors more suitable for use in a larger area, is a *linear function*:

$$
\Delta h = a + b\tau,
$$

where the parameter τ is the location along the track reckoned from its starting point. The dimension of this location can be time (seconds) or angular distance (degrees). Now the set of observation equations for the situation described above is

$$
\overbrace{\begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}}^{\underline{\ell}} = \overbrace{\begin{bmatrix} 0 & 0 & 1 & \tau_1^2 & -1 & -\tau_1^3 \\ 1 & \tau_2^1 & 0 & 0 & -1 & -\tau_2^3 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \\ a_3 \\ b_3 \end{bmatrix}}^{\times} + \overbrace{\begin{bmatrix} n_1 \\ n_2 \end{bmatrix}}^{\underline{n}}.
$$

Here the design matrix contains, besides the values 1 and -1, also the expressions τ_k^i, in which i is the number of the track, k that of the crossover point. These are computable when the geometry of the tracks is known.

Now there are *two* unknowns for every track, *a* and *b*, a constant and a trend. Of course also this system will prove to be singular. Removing the singularity can be done by fixing all three parameters *b* and one parameter a[3].

The phenomenon that no solution can be found unless something is fixed, is called a *datum defect*. Fixing something suitable will define a certain *datum*. Between two different datums exists a *transformation formula*: in the case of one orbit error parameter per track, this transformation is a simple parallel shift or *translation* of all tracks up or down.

The situation is somewhat similar as when defining a height or vertical reference system, for a country. One needs to fix one point, e.g., Helsinki harbour. If alternatively one fixes another point, e.g., Turku harbour, the result is another *datum*, in which all height values differ from the corresponding ones in the first datum by a certain fixed amount.

The argument continues to hold if there is a large number of tracks: say, ten Northgoing and ten Southgoing tracks, crossing in 10×10 crossover points. Here, for two parameters per track, we would have 40 unknowns and no less than 100 observations. Still, we must constrain the absolute level and the various trends and possible other deformations of the whole network of tracks. It gets complicated, but a simple approach is to attach *a priori* uncertainties to the unknowns a_i, b_i to be estimated, e.g., from the known uncertainties of the orbit prediction available. The least-squares adjustment equation then becomes

$$\widehat{x} = \left(A^{\mathsf{T}} A + \Sigma^{-1} \right)^{-1} A^{\mathsf{T}} \underline{\ell},$$

where Σ is the diagonal matrix containing the *a priori* variances[4] $\sigma_{a,i}^2, \sigma_{b,i}^2$ of the parameters of each track *i*. This is referred to as Tikhonov[5] regularization.

---

[3]In order to understand this, build, e.g., a three track "wire-frame model" from pieces of iron wire, tied together by pieces of string at the crossover points. Crossover conditions don't in any way fix the values of the trends *b*, and the whole absolute level of the frame continues to be unconstrained.

[4]We assume the mean error of unit weight to be 1.

[5]Andrey Nikolayevich Tikhonov (1906 – 1993) was a Russian Soviet mathe-
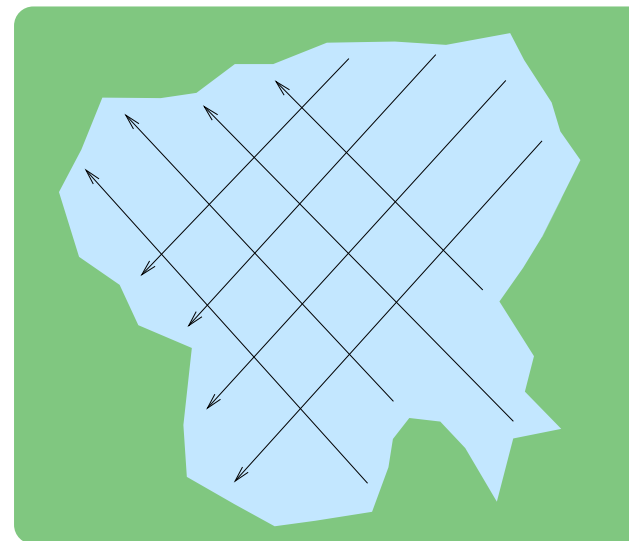
Figure 13.4. Example of track geometry of satellite altimetry

13.2.3 Another example

In diagram 13.4 describing a satellite altimetry geometry, there are 16 crossover points. We attempt a crossover adjustment.

Questions:

1. If the orbital error Δh of each satellite track is described by a model with a single bias term, how many unknowns are there?

2. If we have available 16 "observations", i.e., crossover differences, how many of them are redundant?

3. Is it geometrically possible to calculate this network?

4. If we fix one track in advance (so-called *a priori* information), how many redundant observations are there? Can this network be calculated?

---

matician and geophysicist.

5. If every track has *two* unknowns, a bias as well as an error growing linearly with time, i.e., a trend, what then needs to be fixed in order to make the network calculable? How many redundancies are there then?

6. If, in case 3, we fix one track, which one would you choose? Propose alternatively a solution where you do not have to make a choice.

Answers:

1. As many as there are tracks: 8.

2. $16 - 8 = 8$.

3. No, because the absolute level of the whole network is indeterminate.

4. $16 - (8 - 1) = 9$. Now the network can be calculated.

5. If we assume that the tracks are "straight" in (x, y) co-ordinates, then the set of allowable transformations on the whole network is

$$\Delta h = a_{00} + a_{10}x + a_{01}y + a_{11}xy$$

with four degrees of freedom. So, fix one bias and three trends, not all North- or all Southgoing.

6. Any such choice would be arbitrary. Rather use the method described above instead, Tikhonov regularization.

_____

13.2.4   Global crossover adjustment

In a global crossover adjustment, often a still more sophisticated model is used,

$$\Delta h = a + b \sin \tau + c \cos \tau, \tag{13.3}$$

where now τ is an angular measure, e.g., the place along the track measured from the last South-North equator crossing. See Schrama (1989), where this problem is treated more extensively. In this model, a represents the size of the orbit, while (b, c) describe the offset of the centre of the orbit from

the geocentre. This model is three-dimensional: the orbital arcs with their crossovers form a spherical network surrounding the Earth. The degrees of freedom left by the crossover conditions are now the size of this sphere and the offset of its centre from the geocentre: with (X, Y, Z) geocentric co-ordinates, we have

$$\Delta h = a_0 + a_1 X + a_2 Y + a_3 Z \tag{13.4}$$

with four degrees of freedom[6].

13.3   Choice of satellite orbit

In choosing a satellite orbit, Kepler's orbital laws are central. Kepler's third law says:

$$GM P^2 = 4\pi^2 a^3, \tag{13.5}$$

where $a = a_E + h$ is the satellite orbit's semi-major axis (i.e., the mean distance from the geocentre), while h is called the satellite's mean height. P is the orbital period, a_E the equatorial radius of the Earth ellipsoid.

From equation 13.5 one can already infer that using satellite observations one can precisely determine the quantity GM, the mass of the Earth multiplied by Newton's universal gravitational constant[7]. The period P can be precisely determined from long observation series, and also the size of the orbit a can be obtained very precisely, e.g., from satellite laser ranging (SLR) observations. For this purpose have been used, e.g., the well known Lageos (Laser Geodynamic Satellite) satellites, which orbit the Earth at a height of 6000 km. Ranges are nowadays obtained with better than centimetre precision.

_____

[6]One could argue that, in eq. 13.3, the parameter a should be zero, as Kepler's third law allows a very precise determination of the orbital size, see section 13.3. Then, also $a_0 = 0$ in equation 13.4.

[7]This is why it is said that Henry Cavendish was the first to "weigh the Earth"... determining GM was already straightforward back then using the orbital motion of the Moon. The challenge was separating G and the mass of the Earth M, obtaining the latter in ordinary units of mass.
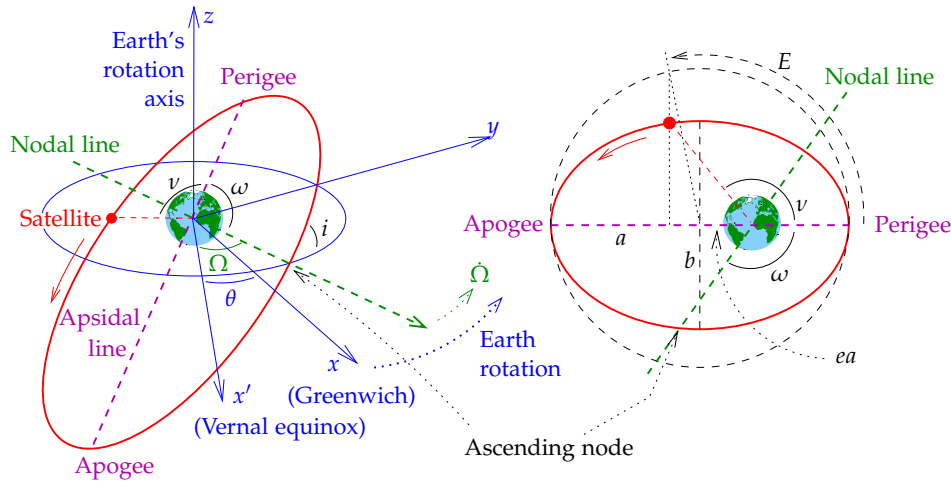
Figure 13.5. Kepler's orbital elements: a — semi-major axis, e — eccentricity, i — inclination, Ω — right ascension (celestial longitude) of the ascending node, ω — argument of perigee, and ν — true anomaly.

The orbits of *altimetric satellites* are chosen much lower, as is seen from table 13.1 at the start of the chapter. The height is fine tuned using on-board thrusters, so that the satellite passes over the same place, e.g., once a day, after 14 orbital periods. Alternatively one chooses an orbit that flies over the same place every third, seventeenth, 168th day... this is called the *repeat period*.

rakettimoottorit

The choice of the repeat period depends on the mission objective:

- If one wishes to study the precise shape of the *mean sea surface*, one chooses a *long* repeat period, in order to get the tracks as close together as possible on the Earth's surface.
- If one wishes to study the *variability* of the sea surface, one chooses an orbit that returns to the same location after a *short* time interval. Then, the grid of tracks on the Earth's surface will be sparser.

Also parameters describing the figure of the Earth affect satellite motion, e.g.,

the quantity J_2, the *dynamic flattening*, having a value of $J_2 = 1082.6267 \cdot 10^{-6}$. It is just one of many so-called spherical-harmonic coefficients that describe the figure of the Earth and affect satellite orbits. In the case of J_2 the effect is, that the plane of the satellite orbit rotates at a certain rate (orbital precession), which make the satellite, if she flies over the same location the next day, do so several minutes earlier. The equation is, for a circular orbit of radius a:

$$\frac{d\Omega}{dt} = -\frac{3}{2}\sqrt{\frac{GM}{a^3}}\frac{a_E^2}{a^2}J_2 \cos i,$$

in which again a_E is the equatorial radius of the Earth reference ellipsoid, and i the *inclination* of the orbital plane relative to the equator. If we substitute numerical values into this, we obtain

$$\frac{d\Omega}{dt} = -1.31895 \cdot 10^{18}\frac{\cos i}{(a_E + h)^{3.5}}\left[\mathrm{m}^{3.5}\,\mathrm{s}^{-1}\right],$$

where h is the *mean height* of the satellite orbit, conventionally above a sphere of size equatorial radius a_E. If we substitute into this, e.g., the satellite height $h = 800\,\mathrm{km}$ (and use $a_E = 6\,378\,137\,\mathrm{m}$) we obtain

$$\frac{d\Omega}{dt} = -1.33102 \cdot 10^{-6}\cos i\left[\mathrm{rad\,s}^{-1}\right] = \frac{-6°.589}{\mathrm{day}}\cdot\cos i.$$

For practical reasons — solar panels! — we often choose the satellite orbit such, that the orbital plane turns along with the annual apparent motion of the Sun, $\frac{360°}{365.25\,\mathrm{days}} = \frac{0°.9856}{\mathrm{day}}$. See figure 13.6.

If the inclination i is chosen in the range $96° - 102°$, depending on the orbital height, then the Earth's dynamic flattening J_2 will cause just the suitable rotational motion of the orbital plane ("no-shadow / Sun-synchronous / Sun-stationary orbit"), see figure 13.7.

An orbit with an inclination $i > 90°$ is called a *retrograde orbit*: the satellite is moving Westward in longitude, opposite to the direction of the Earth's rotation, which is Eastward. The orbital inclination i, or for a retrograde orbit, its supplement $180° - i$, is also the greatest Northern or Southern latitude a satellite can fly over. This means that, unless the inclination is precisely $90°$, there will be areas around both poles that the satellite will never overfly: the "polar holes".
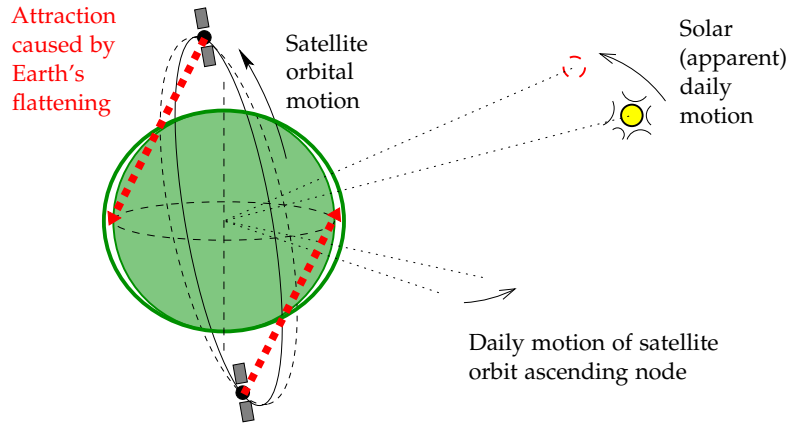
Figure 13.6. The mechanism of a Sun-stationary orbit.

A drawback of a Sun-stationary orbit is, that the altimetric observations are made at always the same local time of day. For example, the diurnal and semi-diurnal tides caused by the Sun will always have the same phase angle,
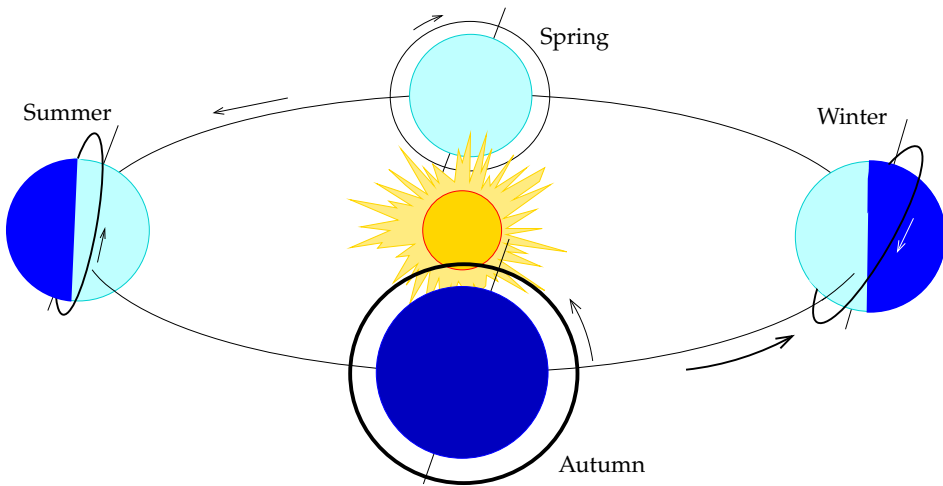
Figure 13.7. Geometry of a "no-shadow" orbit. Season names are boreal.
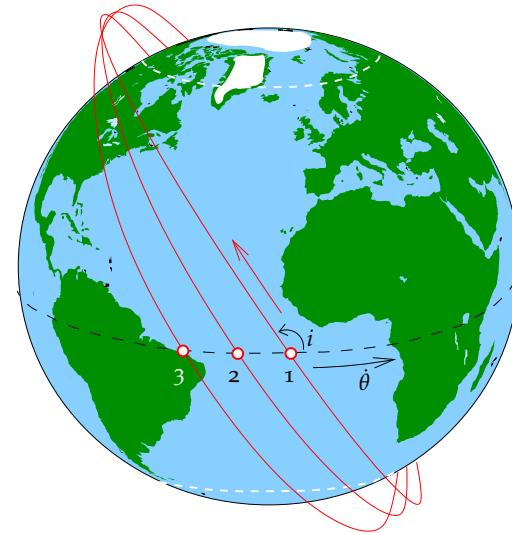
Figure 13.8. A satellite in a retrograde orbit around the rotating Earth, crossing the equator South to North three successive times. The angle between the orbit and the equator, the inclination i, or for a retrograde orbit, its supplement $180° - i$, is also the highest Northern or Southern latitude that the satellite can fly over. The unreachable "polar holes" are indicated by white dashed lines.

and thus they cannot be observed with this kind of satellite ("resonance"). Therefore, the oceanographic satellite TOPEX/Poseidon, and her follow-up Jason satellites, were placed in non-Sun-stationary orbits.

13.3.1   Example

A satellite moves in an Sun-stationary orbit, i.e., she always, day after day, flies over the same latitude at the same local (mean) Solar time.

Questions:

    1. What is the period of the satellite if she always after 14 revolutions flies again over the same spot?

2. Same question if she always flies over the same spot after 43 revolutions (3 days)?

3. And after 502 revolutions (35 days)?

4. What is the height of the satellite in a "three-day orbit"? Use Kepler's third law, equation 13.5. $GM = 3\,986\,005 \cdot 10^8\,\text{m}^3\text{s}^{-2}$, and the height of the satellite is $h = a_S - a$, with $a = 6\,378\,137\,\text{m}$.

5. What is the satellite height in a "35-day orbit"? And the height *difference* with the previous question?

6. What is, for the three-day orbit, the mean separation between North-going orbital tracks (i.e., at what level of detail is the altimeter able to image the sea surface!)?

7. Same question for a 35-day orbit.

8. Questions for reflection:

   (a) for what purpose would you use a 35-day orbit, for what purpose a three-day orbit?

   (b) Would it be possible, or easy, to fly both orbits with the same satellite (see question 5)?

Answers:

1. The satellite completes 14 orbits per day, i.e., per 1440 minutes: $P = \frac{1440}{14} = 102.857\,\text{min}$.

2. The satellite completes 43 orbits in three days, i.e., per 3×1440 minutes: $P = \frac{3 \times 1440}{43}\,\text{min} = 100.465\,\text{min}$.

3. The satellite completes 502 orbits in 35 days, i.e., per 35×1440 minutes: $P = \frac{35 \times 1440}{502}\,\text{min} = 100.398\,\text{min}$.

4. Execute the `octave` code in tableau 13.2. The result is 780.604 km.

5. The same code, with `P=100.398*60`, yields 777.421 km. The difference with the previous is 3.183 km.

6. There are 43 orbits with different ground tracks. That means a separation of $\frac{360}{43} = 8.372$ degrees, or at the equator, $\frac{40\,000}{43} = 930\,\text{km}$; less at higher latitudes.

7. $\frac{360}{502} = 0.717$ degrees, or $\frac{40\,000}{502} = 80\,\text{km}$.

Tableau 13.2. Calculating the height of a satellite from her period.

```
format long
GM=3986005e8;
ae=6378137;
P=100.465*60; % seconds
fac=4*pi*pi; % four pi square
a=(GM*P*P/fac)^0.33333333;
h = a - ae;
printf('\n\nOrbital height: %8.3f km.\n', h/1000);
```

8. (a) The 35-day orbit would be excellent for detailed mapping. The three-day orbit would be able to see, e.g., tides or weather-related phenomena, but at poor resolution.

   (b) The difference in height being only 3 km, and in period, 4 s, the change in orbit between the two repeat periods should be easily within reach of even small on-board thrusters. So, yes.

———

13.4   In-flight calibration

The highly precise, GNSS positioned satellite radar altimeters in use today require proper *calibration*. The technique of choice for this is *in-flight calibration*, using an ocean area – or sometimes a lake area – the geocentric location of the water surface of which is known thanks to surrounding GNSS-positioned tide gauges combined with a precise geoid model of the area. An example of such measurements is Vu et al. (2018).

One reason for in-flight calibration is the circumstance that radar altimeters not only have an unknown zero offset – due to the not precisely known signal paths through the electronic circuitry – but this offset may slowly change or drift over time.
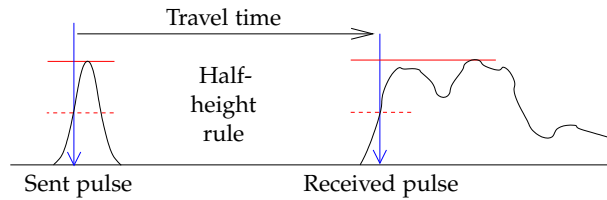
Figure 13.9. Analyzing the altimeter return pulse. The classical return pulse time measurement uses the "half-height point".

13.5   Retracking

The results of a satellite altimetry mission are published already during flight in the form of a so-called *geophysical data record* (GDR) file, containing everything related to the measurement and, e.g., atmospheric correction terms, tidal corrections, wave parameters, etc.

It is common practice today to process again altimetry measurements already collected earlier, in order to extract further useful information. The complete return pulse is analyzed again, in a method called *retracking* (Altimetry, Retracking).

The standard method of analysis is based on that point on leading edge of the return pulse, which is at half height from the maximum value of the pulse. This is according to experience a good way to get the travel time associated with the point in the centre of the *footprint*, directly underneath the satellite. In the back part of the pulse are reflections from the further-away peripheral areas of the footprint.

There are however two situations where this method doesn't work properly during flight, and a more careful *a posteriori* analysis of the pulse is worthwhile:

1. Archipelagos like Indonesia, Åland, … Here it may happen, e.g., that the centre point of the footprint is on land. Then, the first strong bounces will come under an angle from the nearest coast. A precise coastline mask is then essential for processing.

2. Sea ice areas in the Arctic and Antarctic Oceans. Bounces may come

from the surface of the sea ice, in which case one should consider *freeboard* in the processing, i.e., how much the ice sticks out of the water.

In both cases the traditional real-time processing on-board produces erroneous measurements, as the travel time of the return pulse varies too rapidly as the satellite flies on. With retracking, such measurements have been saved, and the area covered by altimetric measurements has been extended into the Arctic and Antarctic Oceans.

Freeboard is an important quantity in determining the thickness of the ice. As the density of ice is about $920 \, \text{kg/m}^3$ and the density of sea water about $1027 \, \text{kg/m}^3$, the ice thickness is about $8\times$ freeboard[8]. If additionally there is remote-sensing data on the area of ice cover, one can calculate the total volume and mass of sea ice.

The Arctic ice cover has diminished radically over the last decades. However, the most radical reduction has been that of ice volume, see figure 13.10: in addition to surface area, also thickness goes down, and especially of the multi-year, thicker ice, a large part has already vanished.

13.6   Oceanographic research using satellite altimetry

The interest of geodesy into satellite altimetry has traditionally been into its use for determining the geoid. Altimetric geoid determination works only is we assume that the sea surface

1. is constant
2. coincides with a level surface, i.e., is the same as the geoid.

In practice however the ocean surface is variable in time and is also not a level surface. Therefore, other approaches have appeared.

1. The variability of the sea surface can be studied by satellite altimetry using *three methods*:

---

[8]Assuming that there is no snow on the ice. Also, ice density varies, and differs between one-year and multi-year ice.
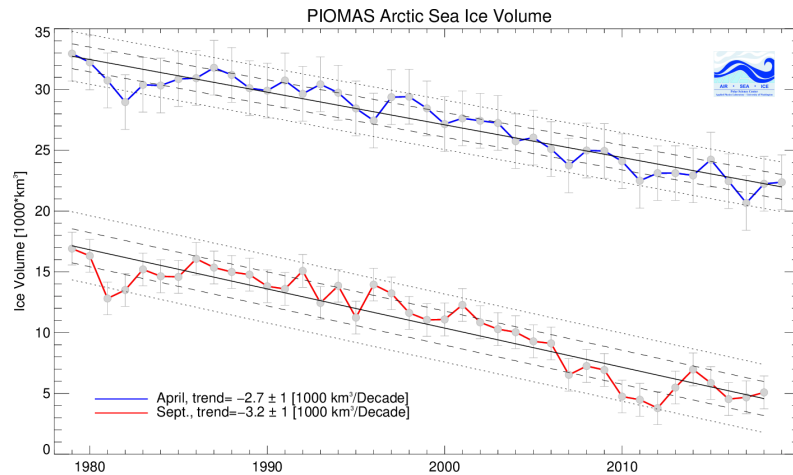
Figure 13.10. Ice volume on the Arctic Ocean. PIOMAS; Schweiger et al. (2011).

(a) Repeat tracks from the same satellite. The tracks can be stacked and adjusted using a simple orbit error model, and the remaining per-track residuals tell something (but, not everything!) about the variability of the sea surface.

(b) Also the crossovers may provide information on sea-surface variability. When the sea surface varies, the results from the crossover adjustment will get poorer: the root-mean-square *a posteriori* (after calculation) crossover difference will become larger. Using this method to actually *study* sea-surface variability is more difficult: it is mostly just able to establish that it exists, and estimate its magnitude.

(c) Nowadays altimetric satellites always carry a GNSS positioning instrument, providing the absolute, geocentric location of the microwave radar device at the moment of measurement. With it, the variations of sea level can be monitored by direct measurement, assuming that both temporal and spatial measurement densities

are sufficient.

2. The deviations of sea level from a level surface — the geoid — can be studied only, if we have access to independent information on the true geoid surface. If dense, high-quality gravity measurements are available for an area, this is the case, and we may estimate the *sea-surface topography*.

Collecting sufficiently precise and dense gravimetric data is possible with a sea gravimeter or with *airborne gravimetry*. Also, measurement with a special satellite (gravitational gradiometry, GOCE satellite) has long been planned and was finally realized, see section 13.7.

13.7   Satellite gravity missions

During the early years of the 21st century three satellites were launched for investigating the fine structure of the Earth's gravity field or geopotential, i.e., for determining a global, high resolution model of the geoid.

CHAMP (Challenging Minisatellite Payload for Geophysical Research and Applications, 2000-099A) was a German satellite project under the auspices of the German Research Centre for Geosciences GFZ. She was launched into orbit from Plesetsk, Russia, in 2000. The orbit height of CHAMP was initially 454 km, diminishing over the mission time to $\sim$ 300 km due to atmospheric drag. The orbital inclination was 87°. On September 19, 2010 the satellite returned into the atmosphere. Project description: CHAMP Mission.

CHAMP contained a GPS receiver. This allowed the determination of her location in space $\mathbf{x}(t)$ for any moment in time t. From a succession of such locations one may calculate the geometric acceleration $\mathbf{a}(t)$ by differentiation:

$$\mathbf{a}(t) = \frac{d^2}{dt^2}\mathbf{x}(t).$$

The differentiation is done numerically in the way that was described in the part on airborne gravimetry, equation 11.6.

The satellite also contained an accelerometer, which eliminated the satellite accelerations caused by the atmosphere's aerodynamic forces,
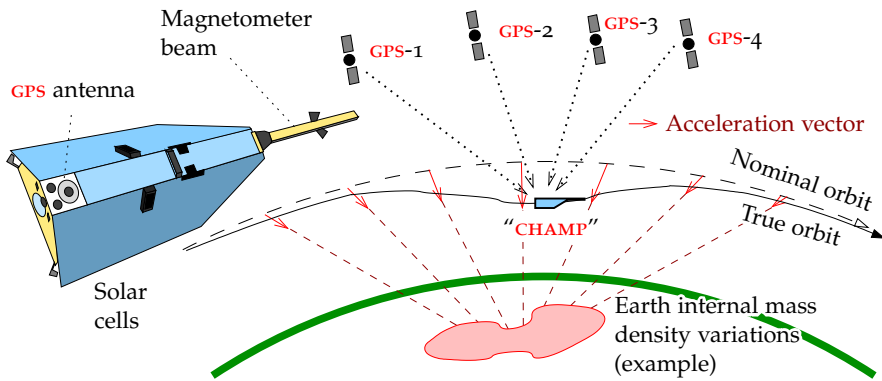
Figure 13.11. Determining the Earth's gravity field from GPS orbital tracking of a low flying satellite.

i.e., the deviations from free-fall motion. Then, only the accelerations caused by the Earth's gravitational field remain, from which a precise geopotential or *geoid* model may be calculated by the techniques described earlier.

A number of global geopotential models based on CHAMP data have been calculated and published.

GRACE (Gravity Recovery And Climate Experiment Mission, 2002-012 A and B) measured *temporal changes* in the Earth's gravity field at intervals of about a month extremely precisely, but at a rather crude geographic resolution. These temporal changes are caused by motions in the Earth's "blue film", i.e., atmosphere and hydrosphere. The quantity measured is also called the "sea-floor pressure", a somewhat surprising expression, until one sees that it really represents the total mass of a column of air and water. Project description: GRACE Mission. It was a collaborative American–German undertaking under the leadership of the Center for Space Research, University of Texas at Austin.

GRACE was a *satellite pair* ("Tom and Jerry"): the satellites flew in the same orbit in a tandem configuration at initially about 500 km height, at an inter-satellite separation of 220 km. The orbital inclination was
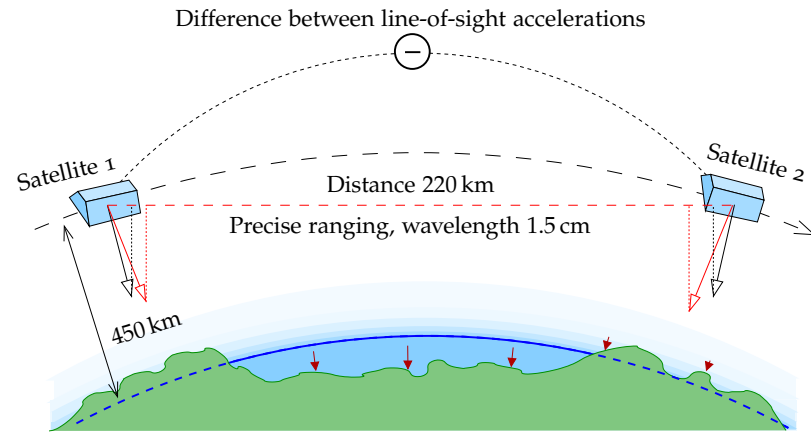
Figure 13.12. The principle of the GRACE satellites: measuring the minute variations in time of the gravity field using SST (satellite-to-satellite tracking). The changes are due to mass shifts in the "blue film" – the atmosphere and hydrosphere – and expressed as variations in "total sea-floor pressure" ($\downarrow$).

89°, i.e., the orbit was almost polar, providing complete global coverage. The changes in distance between the satellites were measured by a microwave link at a precision of $1\,\mu\mathrm{m/s}$. Both satellites also carried sensitive accelerometers for measuring and eliminating the effect of atmospheric drag.

The measurement system was so sensitive, that even the movement of a water layer of one millimetre thickness could be noticed, as long as it extended over an area the size of a continent (some 500 km).

Published results show impressively, e.g., the wet and dry monsoons, seasonal variations in opposite phase in the Northern and Southern hemispheres, in the great tropical river basins: Amazonas, Congo, the Mekong, India, Indonesia... GRACE Mission, hydrology.

The mission ended in 2017 after 15 years, three times the planned mission duration. A GRACE follow-on mission was launched in 2018, GRACE Follow-On Mission.

Figure 13.13. GRACE mission results: surface mass layer, in centimetres of water equivalent. Click for animation.

GOCE (2009-013A, Geopotential and Steady-state Ocean Circulation Explorer) was the most ambitious of all the satellites. Built by the European Space Agency ESA she was launched successfully from Plesetsk in March 2009. The orbital height was only $270 - 235$ km during the mission and the satellite contained an ionic rocket engine with a stock of propellant in order to maintain the orbit against atmospheric drag. The orbital inclination was $96°.7$, i.e., the orbit was Sun-stationary[9].

GOCE carried a very sensitive *gravitational gradiometer*, a device that measured precisely components of the *gradient* of the Earth's attraction, i.e., the dependence of components of the attraction vector on the various co-ordinates of place. The gradiometer consisted of six extremely sensitive, three-axes accelerometers mounted pairwise on a frame. The mission ended in 2013 and the satellite was seen to burn up in the atmosphere November 11 over the Falkland islands (Scuka, 2013).

Theoretical analysis has shown that a gravitational gradiometer is the best way to measure the very local features of the Earth's gravity field,

Figure 13.14. Determining the Earth's gravity field with the gravitational gradiometer on the GOCE satellite.

better than orbital tracking by GNSS. The smallest details in the geoid map seen by GOCE are only 100 km in diameter, and their precision is as good as ± 2 cm.

With a global geoid model this precise, we may calculate the deviations of the sea surface from the geoid, or equipotential or level surface, at the same precision. We saw that the true location in space of the sea surface is obtained from satellite radar altimetry, also at a

---

[9]Because of this inclination angle, there was a cap of radius $3°.3$ at each pole within which no measurements were obtained.

few centimetres precision. This separation between sea surface and equipotential surface can again be *inverted* to *ocean currents*, see section 12.5. This is the background for the name of the GOCE satellite.

Self-test questions

1. What is the *footprint* of a radar altimeter? How does it depend on wave height?

2. What is the freeboard of ocean ice? How can it be used to determine the volume of the ice?

3. What three alternative models for the satellite orbit error correction exist?

4. What is, in the case of satellite altimetry, a *datum defect*, and how can it be fixed?

5. How can Kepler's third law be used to determine the mean height of a satellite if its period is given?

6. What is the repeat period of a satellite orbit?

7. What is J_2, and how does it affect the motion of a satellite?

8. What is a Sun-synchronous orbit, and why is it useful?

9. What is a retrograde orbit?

10. Why are the orbits of the TOPEX/Poseidon and Jason satellites not Sun-synchronous?

11. With which three satellite altimetric methods can one study sea-surface variability?

12. For the study of the fine structure of the Earth's gravity field and its temporal variability, so far three satellite missions have been launched. Describe them and the methods used by them.

Exercise 13 – 1: Altimetry, crossover adjustment

Given are two Northgoing satellite tracks and three Southgoing ones. There are six crossover sites, see figure 13.15.
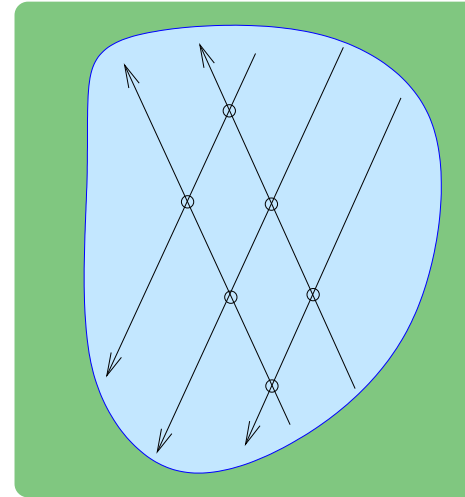
Figure 13.15. Example of satellite altimetric track geometry.

1. If the orbit errors of every track are described as a linear function of place:
$$\Delta h = a + b\tau,$$
how many coefficients a and b are then needed?

2. Write out the *observation equations*. The observations are the crossover differences, the unknowns are the coefficients a and b for the different tracks.

3. Can these observation equations give a unique solution? Why / why not?

Exercise 13 – 2: Satellite orbit

A satellite moves in a Sun-synchronous orbit, where after 419 orbits and 30 days, she again moves over exactly the same spot.

1. How much is the period of the satellite?

350   Satellite altimetry and satellite gravity missions

2. How long is the distance (West to East), in kilometres, between the Northgoing tracks at the equator?

3. What is the highest Northern latitude that the satellite can fly over? And...

4. ...*in what compass direction* does the satellite fly at that point?

Exercise 13 − 3: Kepler's third law

What is the satellite height if the period is 98 minutes? Use Kepler's third law

$$GMP^2 = 4\pi^2 a_S^3,$$

$GM = 3\,986\,005 \cdot 10^8 \text{ m}^3/\text{s}^2$, and the height of a satellite is $h_S = a_S - a_E$, where $a_E = 6\,378\,137$ m.

Tides, atmosphere and Earth crustal movements

14

14.1   Theoretical tide

The tidal potential W can be written as follows:

$$W = \frac{GMR^2}{d^3} P_2(\cos\zeta) + \ldots = \frac{GMR^2}{2d^3}\left(3\cos^2\zeta - 1\right) + \ldots,$$

in which d is the distance to either the Moon or the Sun, R the radius of the Earth, and ζ the local zenith angle of the Sun or Moon. $P_2(\cos\zeta)$ is the Legendre polynomial of degree two. GM is the mass of the Sun or Moon multiplied by Newton's gravitational constant. In the case of Sun and Moon, the extra terms $(\ldots)$ can be neglected, because these are such remote bodies: $d \gg R$.

The cosine rule on the sphere tells us that

$$\cos\zeta = \sin\phi\sin\delta + \cos\phi\cos\delta\cos h,$$

where ϕ is the latitude, δ is the declination[1] of the Moon, and h is the hour angle[2] of the Moon.

---

[1] The declination of a celestial body is its geocentric latitude.

[2] The hour angle is the difference in longitude between the Moon and the local meridian. It vanishes when the Moon is in *upper culmination*, due South when seen from Northern non-tropical latitudes.

Tides, atmosphere and Earth crustal movements
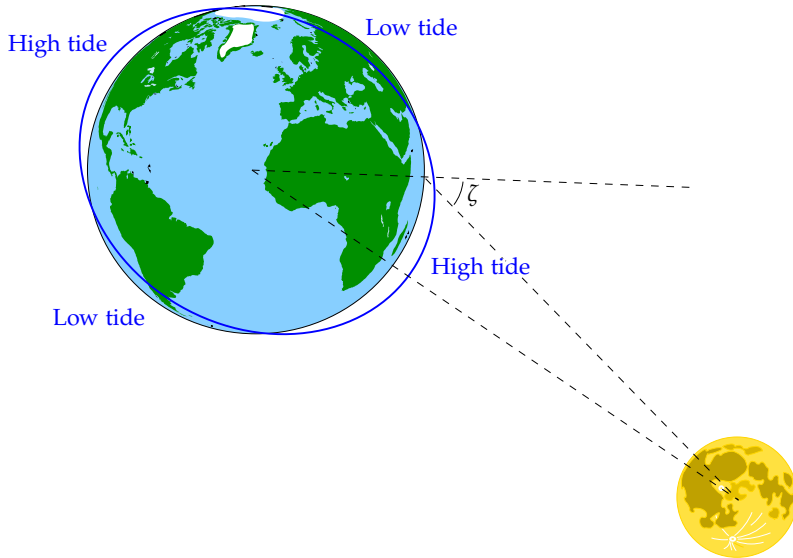
Figure 14.1. Theoretical tide. ζ is the local zenith angle of the Moon (or Sun).

According to the *spherical-harmonic addition theorem* (Wolfram Mathworld, Spherical Harmonic Addition Theorem) we have now

$$P_n(\cos \zeta) = P_n(\sin \phi) P_n(\sin \delta) +$$
$$+ 2 \sum_{m=1}^{n} \frac{(n-m)!}{(n+m)!} P_{nm}(\sin \phi) P_{nm}(\sin \delta) \cos mh,$$

or for $n = 2$,

$$P_2(\cos \zeta) = P_2(\sin \phi) P_2(\sin \delta) +$$
$$+ \frac{1}{3} P_{21}(\sin \phi) P_{21}(\sin \delta) \cos h +$$
$$+ \frac{1}{12} P_{22}(\sin \phi) P_{22}(\sin \delta) \cos 2h.$$

In this, according to table 3.2,

$$P_{21}(\sin \phi) = 3 \sin \phi \cos \phi,$$
$$P_{22}(\sin \phi) = 3 \cos^2 \phi,$$

and we obtain

$$P_2(\cos \zeta) = \frac{1}{2} \left(3 \cos^2 \zeta - 1 \right) = P_2(\sin \phi) P_2(\sin \delta) +$$
$$+ 3 \sin \phi \cos \phi \sin \delta \cos \delta \cos h +$$
$$+ \frac{3}{4} \cos^2 \phi \cos^2 \delta \cos 2h =$$
$$= \frac{1}{2} \left(3 \sin^2 \phi - 1 \right) \frac{1}{2} \left(3 \sin^2 \delta - 1 \right) +$$
$$+ \frac{3}{4} \sin 2\phi \sin 2\delta \cos h +$$
$$+ \frac{3}{4} \cos^2 \phi \cos^2 \delta \cos 2h.$$

From this

$$W = \frac{GMR^2}{4d^3} \left(\begin{array}{c} \left(3 \sin^2 \phi - 1 \right) \left(3 \sin^2 \delta - 1 \right) + \\ + 3 \sin 2\phi \sin 2\delta \cos h + \\ + 3 \cos^2 \phi \cos^2 \delta \cos 2h \end{array} \right).$$

This is the so-called Laplace tidal decomposition equation.

It has three parts:

1. A slowly varying part,

$$W_1 = \frac{GMR^2}{4d^3} \left(\left(3 \sin^2 \phi - 1 \right) \left(3 \sin^2 \delta - 1 \right) \right),$$

that still depends on δ and is therefore periodic with a 14-day (half-month) period. Again by using spherical trigonometry:

$$\sin \delta = \sin \epsilon \sin \ell \implies \sin^2 \delta = \sin^2 \epsilon \sin^2 \ell = \sin^2 \epsilon \left(\frac{1}{2} - \frac{1}{2} \cos 2\ell \right),$$

$$(14.1)$$

where ℓ is the longitude of the Moon in its orbit, reckoned from the ascending node (equator crossing), and ϵ is the inclination of the Moon's orbit with respect to the equator, on average $23°$ but rather variable, between $18°.3$ and $28°.6$. Thus we obtain

$$W_1 = \frac{GMR^2}{4d^3} \left[\left(3 \sin^2 \phi - 1 \right) \left(3 \sin^2 \epsilon \left(\frac{1}{2} - \frac{1}{2} \cos 2\ell \right) - 1 \right) \right],$$

where we have used result 14.1. We split $W_1 = W_{1a} + W_{1b}$ into two parts, a constant[3] and a periodic or semi-monthly ("fortnightly") part:

$$W_{1a} = \frac{GMR^2}{4d^3}\left[(3\sin^2\phi - 1)\left(\frac{3}{2}\sin^2\epsilon - 1\right)\right], \qquad (14.2)$$

$$W_{1b} = -\frac{GMR^2}{4d^3}\left[(3\sin^2\phi - 1)\left(\frac{3}{2}\sin^2\epsilon \cos 2\ell\right)\right].$$

2. Additionally we have a couple of terms in which the hour angle h appears (periods roughly a day and roughly half a day):

$$W_2 = \frac{GMR^2}{4d^3}\left(3\sin 2\phi \sin 2\delta \cos h\right),$$

$$W_3 = \frac{GMR^2}{4d^3}\left(3\cos^2\phi \cos^2\delta \cos 2h\right).$$

In both, we have in addition to h, still δ as a "slow" variable. These equations could be written out as sums of various functions of the longitude of the Moon ℓ.

Use again basic trigonometry, equation 14.1:

$$\cos^2\delta = 1 - \sin^2\delta = 1 - \sin^2\epsilon \sin^2\ell = 1 - \sin^2\epsilon\left(\frac{1}{2} - \frac{1}{2}\cos 2\ell\right),$$

$$\cos 2\ell \cos 2h = \frac{1}{2}\left(\cos(2\ell + 2h) + \cos(2\ell - 2h)\right),$$

$$\sin 2\delta = 2\sin\delta \cos\delta = 2\sin\delta\sqrt{\cos^2\delta} =$$

$$= 2\sin\epsilon \sin\ell\sqrt{1 - \sin^2\epsilon\left(\frac{1}{2} - \frac{1}{2}\cos 2\ell\right)},$$

leading to a trigonometric expansion in ℓ, and so on. See, e.g., Melchior's[4] famous book Melchior (1978).

---

[3]Not precisely, because ϵ is (slowly) time dependent.

[4]Paul Melchior (1925 – 2004) was an eminent Belgian geophysicist and Earth tides researcher.

Table 14.1. The various periods in the theoretical tide. The widely used symbols were standardized by George Darwin.

	Changing function	Period		Darwin symbol		Name
		Moon	Sun	Moon	Sun	
W_{1a}	-	-	-	M_0	S_0	Permanent tide
W_{1b}	$\cos 2\ell$	14^d	182^d	Mf^a	Ssa^b	Declination tide
W_2	$\cos h$	$24^h 50^m$	24^h	K_1, O_1	S_1, P_1	Diurnal
W_3	$\cos 2h$	$12^h 25^m$	12^h	M_2	S_2	Semi-diurnal

[a]Lunar fortnightly

[b]Solar semi-annual

From the above equations, often the coefficient

$$D \stackrel{\text{def}}{=} \frac{3GMR^2}{4d^3}, \qquad (14.3)$$

"Doodson[5]'s constant" is taken separately. The value for the Moon equals $D = 26.75\,\text{cm} \times g$ and for the Sun $12.3\,\text{cm} \times g$. See figure 14.2.

The *periods* are tabulated in table 14.1 with their Darwin[6] symbols.

In practice, the diurnal and semi-diurnal tides can be divided further into many "spectral lines" close to each other, also because the lunar orbit (like the Earth's orbit) is an ellipse, not a circle.

14.2   Deformation caused by the tidal force

The tidal force, or theoretical tide, of which we spoke above, is not the same as the deformation it causes in the solid Earth. This deformation will depend

---

[5]Arthur Thomas Doodson (1890 – 1968) was a British oceanographer, a pioneer of tidal theory, also involved in designing machines for computing the tides. He was stone deaf.

[6]Sir George Howard Darwin (1845 – 1912) was an English astronomer and mathematician, son of Charles Darwin of *Origin of Species* fame.
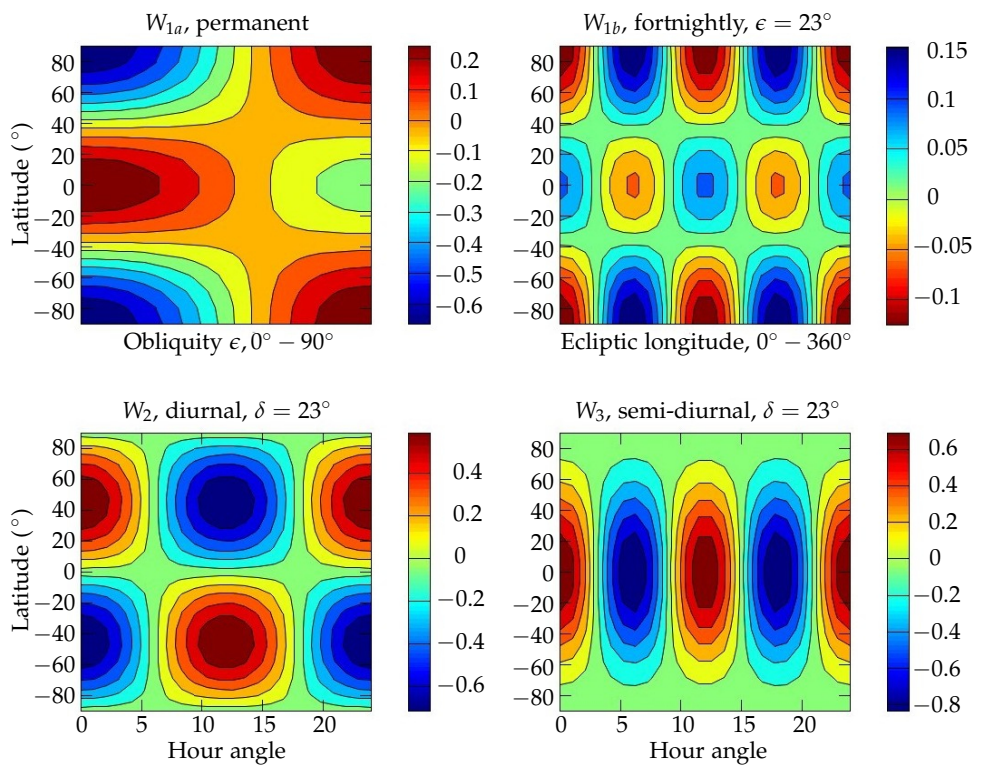
Figure 14.2. The main components of the theoretical tide. These values must still be multiplied by Doodson's constant D.

upon the elastic properties inside the Earth. These elastic properties are often described by so-called (elastic) Love[7] numbers (Melchior, 1978).

Let us first write the external (tidal, or generally, disturbing) potential in the following way:

$$W = \sum_{n=2}^{\infty} \left(\frac{r}{R}\right)^n W_n,$$

---

[7] Augustus Edward Hough Love (1863 – 1940) was a British mathematician and student of Earth elasticity.

in which the index n denotes the degree number of spherical harmonics.

Call the displacement of an element of matter of the solid Earth in the radial direction, u_r, in the North direction, u_ϕ, and in the East direction, u_λ. The following equations apply:

$$u_r = \sum_{n=2}^{\infty} H_n(r) \frac{W_n}{g},$$

$$u_\phi = \frac{1}{g} \sum_{n=2}^{\infty} L_n(r) \frac{\partial W_n}{\partial \phi},$$

$$u_\lambda = \frac{1}{g} \sum_{n=2}^{\infty} L_n(r) \frac{\partial W_n}{\cos\phi \, \partial \lambda}.$$

Here r is the distance from the geocentre. It is assumed here that the Love numbers H_n, L_n depend only on r, i.e., the elastic properties of the Earth are spherically symmetric.

The deformation of the Earth causes also a change (the "indirect effect", in addition to the Moon's original potential W) in the gravity potential. We write

$$\delta W = \sum_{n=2}^{\infty} K_n(r) W_n,$$

where we use already a third type of Love numbers.

On the Earth's surface $r = R$ we make the following specialization:

$$h_n \stackrel{\text{def}}{=} H_n(R),$$
$$\ell_n \stackrel{\text{def}}{=} L_n(R),$$
$$k_n \stackrel{\text{def}}{=} K_n(R).$$

In practice, because of the large distances to Sun and Moon, the only important part of the tidal potential W is the part for the degree number $n = 2$, i.e., W_2.

The Love numbers will depend still on the *frequency*, i.e., the tidal *period P*:

$$h_n = h_n(P),$$
$$\ell_n = \ell_n(P),$$
$$k_n = k_n(P).$$

The tides offer an excellent means of determining all these Love numbers $h_2(P), \ell_2(P)$ and $k_2(P)$ empirically, because, being periodic variations, they cause in the Earth deformations at the same periods, but different amplitudes and phase angles. In this way we may determine at least those Love numbers that correspond to periods occurring in the theoretical tide.

The h and ℓ numbers are nowadays obtained, e.g., by GNSS positioning. The GNSS processing software contains a built-in reduction for this phenomenon. From gravity measurements one obtains information on a certain linear combination of h and k (vertical displacement changes gravity though its gradient, and deformation of the Earth, the shifting of masses, also changes gravity). A useful research instrument is also the long water-tube clinometer, like the tube of the Finnish Geodetic Institute that has long been in use in the Tytyri limestone mine in Lohja (Kääriäinen and Ruotsalainen, 1989). A modern, improved version of this instrument is presented in Ruotsalainen (2017). The same applies for sensitive clinometers in general, like the Verbaandert–Melchior pendulum etc. A clinometer measures the change in orientation between the Earth's crust and the local plumb line.

Measuring the absolute direction of the plumb line, e.g., with a zenith tube, can again give information on a certain linear combination of ℓ and k, but only after various reductions (Earth polar motion).

14.3   The permanent part of the tide

As shown above, the theoretical tide equation contains a constant part that doesn't even vary in a long-period way. Of course the Earth responds also to this part of the tidal force. However, because the deformation isn't periodic, it is not possible to measure it. And the mechanical theory of the solid Earth, and our knowledge of the state of matter inside of the Earth, just aren't good enough for prediction of the response.

For this reason the understanding is generally accepted that the effect of the permanent part of the tide on the Earth's state of deformation should not be included in any tidal reduction (Ekman, 1992). However, often, e.g., in the processing of GNSS observations or in defining spherical-harmonic expansions

of the Earth's gravity field, the tidal reduction does include this term which it is theoretically and practically impossible to know. See Poutanen et al. (1996).

More generally we can say, that a geodetic quantity, e.g., the height of the geoid, can be reduced for the permanent part of the tide *in three different ways*:

○ No reduction whatever is made for the permanent part; the quantity thus obtained is called the "mean geoid". The surface obtained is in the hydrodynamic sense an equilibrium surface, and is therefore the best surface to use in oceanography.

○ The effect of the gravitational field emanating from celestial objects is removed in its entirety from the quantity, but the Earth's deformation it causes is left completely uncorrected; the quantity thus obtained is called the "zero geoid".

○ Both the gravitational effect of a celestial body, and the indirect effect of the deformation it causes, can be calculated according to a certain deformation model (Love number), and corrected for. The result obtained is called the "tide-free geoid". Its problem is precisely the empirical indeterminacy of the elasticity model used.

See figure 14.3. It is good to be critical and precisely analyze in which way the data reduction has been done!

14.4   Tidal corrections between height systems

We see from equation 14.2 that, with $\epsilon = 23°.5$, the permanent part of the tidal potential is equal to

$$W_{\text{perm}} = \frac{GMR^2}{4d^3}\left((3\sin^2\phi - 1)\left(\frac{3}{2}\sin^2\epsilon - 1\right)\right) \approx$$

$$\approx -0.7615\frac{3GMR^2}{4d^3}\left(\sin^2\phi - \frac{1}{3}\right).$$

With the combined Doodson's constant 14.3 for Sun and Moon equal to

$$D = \frac{3GM_{\text{Sun}}R^2}{4d_{\text{Sun}}^3} + \frac{3GM_{\text{Moon}}R^2}{4d_{\text{Moon}}^3} = (12.3\,\text{cm} + 26.75\,\text{cm}) \times g = 39.05\,\text{cm} \times g$$

Earth permanent tidal deformation	Tide-free Earth crust
Direct effect on geoid of Moon and Sun	Mean Earth crust
Effect of Earth permanent tidal deformation (mass displacement) on geoid	Tide-free geoid
	Zero geoid
	Mean geoid
	Reference ellipsoid
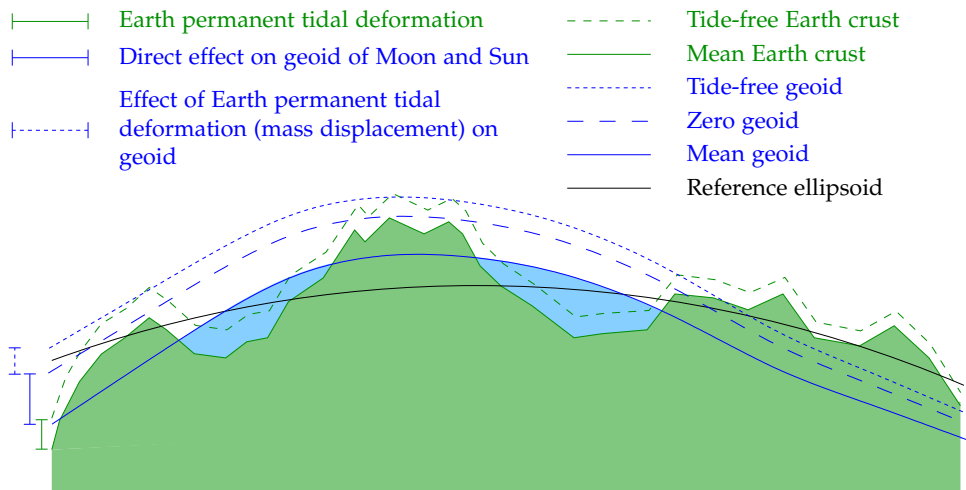
Figure 14.3. Conceptual diagram showing the constituents of the permanent tide.

we obtain

$$W_{\text{perm}} = 29.74 \, \text{cm} \times \left(\frac{1}{3} - \sin^2 \phi \right) \times g.$$

We can express this, with the Bruns equation 5.2, into a permanent tidal geoid effect:

$$N_{\text{perm}} = 29.74 \, \text{cm} \times \left(\frac{1}{3} - \sin^2 \phi \right).$$

From this, $N_{\text{perm}}(0°) = 9.91 \, \text{cm}$ on the equator, and $N_{\text{perm}}(90°) = -29.74 \, \text{cm}$ on the poles.

This, the geoid effect of the permanent part of the external potential of Sun and Moon, is also equal to the difference between the mean geoid and the zero geoid as defined above:

$$\Delta^{\text{mean}}_{\text{zero}} N \overset{\text{def}}{=} N_{\text{mean}} - N_{\text{zero}} = 29.74 \, \text{cm} \times \left(\frac{1}{3} - \sin^2 \phi \right).$$

For heights H above sea level, with $H = h - N$, we have

$$\Delta^{\text{mean}}_{\text{zero}} H \overset{\text{def}}{=} H_{\text{mean}} - H_{\text{zero}} = -29.74 \, \text{cm} \times \left(\frac{1}{3} - \sin^2 \phi \right),$$

and for two different latitudes ϕ_1 and ϕ_2 we have for the effect on the height difference

$$\Delta^{\text{mean}}_{\text{zero}} H(\phi_2) - \Delta^{\text{mean}}_{\text{zero}} H(\phi_1) = 29.74 \, \text{cm} \times \left(\sin^2 \phi_2 - \sin^2 \phi_1 \right).$$

This is the value to be *added* when going from a zero-geoid to a mean-geoid height system, and *subtracted* when going from a mean-geoid to a zero-geoid height system.

When the *tide-free* geoid enters into the picture, we need a value for the Love number k_2 for the permanent tidal deformation, describing the potential of this deformation as a fraction of the original external tidal potential. As we have seen, this number cannot really be empirically determined; a value often used is $k_2 \approx 0.3$. With this, the above equations apply with the coefficient $29.68 \, \text{cm}$ multiplied by $1 + k_2$, yielding

$$\Delta^{\text{mean}}_{\text{tidefree}} H \overset{\text{def}}{=} H_{\text{mean}} - H_{\text{tidefree}} = -38.66 \, \text{cm} \times \left(\frac{1}{3} - \sin^2 \phi \right),$$

$$\Delta^{\text{mean}}_{\text{tidefree}} H(\phi_2) - \Delta^{\text{mean}}_{\text{tidefree}} H(\phi_1) = 38.66 \, \text{cm} \times \left(\sin^2 \phi_2 - \sin^2 \phi_1 \right).$$

Any other correction equation can be obtained from these. Like

$$\Delta^{\text{zero}}_{\text{tidefree}} H(\phi_2) - \Delta^{\text{zero}}_{\text{tidefree}} H(\phi_1) = 8.92 \, \text{cm} \times \left(\sin^2 \phi_2 - \sin^2 \phi_1 \right).$$

14.5　Loading of the Earth's crust by sea and atmosphere

In addition to the deformation caused by the tidal force, the Earth's crust also deforms due to the loading by sea and atmosphere. Especially close to the coast, the tidal motion of the sea causes a multi-period deformation that moves the Earth's crust up and down by as much as centimetres.

This phenomenon can be computationally modelled if the elastic properties of the solid Earth, the tidal motion of the sea, and the precise shape of the coastline are known. One known programme for this purpose is the package Eterna written by the German H.-G. Wenzel[8], which has been used also in Finland.

---

[8]Hans-Georg Wenzel (1945–1999) was a German physical geodesist and geophysicist.

Tides, atmosphere and Earth crustal movements

On the other hand, when such a tool exists, then tidal loading offers also an excellent opportunity for *studying* precisely the very local elastic properties of the Earth's crust.

For measuring the deformation, generally a registering gravimeter is used. The Earth's crust moves up and down, which changes gravity in proportion to the free-air gradient value $-0.3\,\mathrm{mGal}/\mathrm{m}$ (Torge, 1992).

The use of GNSS for measuring the ocean tidal loading has not yet become common.

Like the ocean, also the atmosphere causes, through changes in air pressure, varying deformations of the Earth's crust. The phenomenon is very small, at most a couple of cm. Gravity measurement is not a very good way to study this phenomenon, because many more local, often poorly known, factors affect local gravity. Measurement by GNSS is promising but also challenging.

Self-test questions

1. Describe *in words* the three components of the theoretical tide produced by the Laplace decomposition method.

2. How may the slowly varying part of the theoretical tide be further decomposed into two parts? Describe the parts *in words*.

3. What are the declination and hour angle of a celestial body, e.g., the Moon?

4. What is Doodson's constant?

5. What do Love numbers describe?

6. Why is it not possible to empirically determine the deformation caused by the permanent part of the tide?

7. Describe three different ways to take the permanent part of the tide into account when defining the geoid.

Exercise 14 − 1: Tide

The formula for the permanent tide is

$$W_{1a} = \frac{GMR^2}{4d^3} \left((3\sin^2\varphi - 1)\left(\frac{3}{2}\sin^2\epsilon - 1\right)\right),$$

where φ is latitude and ϵ is the *obliquity* of the Earth's axis of rotation, currently about 23°.

1. For what value φ the permanent tide vanishes? Interpretation?

2. For what value ϵ the permanent tide vanishes? Interpretation?

⬚ Earth gravity field research

15

⬚ 15.1 Internationally

In the framework of the IAG, the International Association of Geodesy, research into the Earth's gravity field is currently the responsibility of the International Gravity Field Service. The IGFS was created in 2003 at the IUGG General Assembly in Sapporo, Japan, and it operates under the IAG's new Commission 2 "Gravity Field". The United States National Geospatial-Intelligence Agency (NGA) serves as its technical centre.

An important IAG service of great reputation is the International Gravity Bureau, BGI, *Bureau Gravimétrique International*, located in Toulouse, France (http://bgi.omp.obs-mip.fr/). The bureau works as an international broker to which countries can submit their gravimetric materials. If some researcher needs gravimetric material from another country, e.g., for geoid computation, he can request it from the BGI, who will provide it with the permission of the country of origin, provided the country of the researcher has in its turn submitted its own gravimetric materials for BGI use.

The French state has invested significant funds into this vital international activity.

Another important IAG service in this field is the ISG, the International Service for the Geoid. It has in fact already operated since 1992 under the name International Geoid Service (IGES), the executive arm of the International Geoid Commission (IGEC). The ISG office is located in Milano (http://www.

isgeoid.polimi.it/) also with substantial support by the Italian state. The task of this service is to support geoid determination in different countries, for which purpose existing geoid solutions are collected in a common data base, and international research schools are organized to develop knowledgability and skills in the art of geoid computation, especially in developing countries.

Both services, BGI and ISG, are under the auspices of the International Gravity Field Service IGFS, as two of the many official services of the IAG. Other IGFS services are the International Center for Earth Tides (ICET), the International Center for Global Earth Models (ICGEM), and the International Digital Elevation Model Service (IDEMS).

15.2   Europe

In Europe operates the EGU, the European Geosciences Union, in the frame of which much publication and meeting activities relating to gravity field and geoid are being co-ordinated. The EGU organizes annually symposia, where always also sessions are included on subjects related to gravity field and geoid. Also American scientists participate. Conversely the American Geophysical Union's (AGU) fall and spring meetings[1] are also favoured by European researchers.

To be mentioned is the Geodetic Institute ("Institut für Erdmessung") of Leibniz University in Hannover, Germany, which since 1990 has acted as the European computing centre of the International Geoid Commission (IGEC, and produced high quality geoid models for use in Europe (Denker, 1998; European geoid calculations).

15.3   The Nordic countries

In the Nordic countries, important work is being co-ordinated by the NKG, *Nordiska Kommissionen för Geodesi* and its Working Group for Geoid and Height

---

[1]Fall (autumn) meetings in San Francisco, spring meetings somewhere in the world. The AGU, though American, is a very cosmopolitan player.

Systems. To its activities belongs geoid determination, studying the preconditions for still more precise geoid models, new levelling technologies, and the study of post-glacial land uplift.

The group has for a long time computed, at its computing centre in Copenhagen, high-quality geoid models, the next to last such being NKG2004 (Forsberg and Kaminskis, 1996; Forsberg and Strykowski, 2010). The newest model, NKG2015, is the result of calculations by the computing centres of several countries including Sweden and Estonia. It was published in October 2016.

15.4   Finland

In Finland the study of the Earth's gravity field has mainly been in the hands of the Finnish Geodetic Institute, founded in 1918, one year after Finnish independence. The institute has been responsible for the national fundamental levelling and gravimetric networks and their international connections. In 2001 the Finnish Geodetic Institute's gravity and geodesy departments were joined into a new department of geodesy and geodynamics, to which also gravity research belongs. Among matters studied are, e.g., solid-Earth tides, the free oscillations of the solid Earth, post-glacial land uplift, and vertical reference or height systems.

Geoid models have been computed all the time, starting with Hirvonen's global model (Hirvonen, 1934) and ending, for now, with the Finnish model FIN2005N00 (Bilker-Koivula, 2010). These geoid models are actually based on the Nordic NKG2004 gravimetric geoid, and are only fitted to a Finnish set of GNSS levelling control points through a transformation surface.

In 2015, the Finnish Geodetic Institute was merged into the National Land Survey as its geospatial data centre and research facility. The English-language acronym continues as FGI, the Finnish Geospatial Research Institute (http://www.fgi.fi/fgi).

Also Helsinki University of Technology (today part of Aalto University) has been active in research on the Earth's gravity field. V. A. Heiskanen, who was a professor at HUT in 1928–1949, acted 1936–1949 as the director of the International Isostatic Institute. After moving to Ohio State University

he worked with many other, including Finnish and Finnish-born, geodesists on calculating the first major global geoid model, the "Columbus geoid" (Kakkuri, 2008).

15.5   Textbooks

There are many good textbooks on the study of the Earth's gravity field. In addition to the already mentioned classic Heiskanen and Moritz (1967), which is in large part obsolete, we may mention Wolfgang Torge's book Torge (1989). Difficult but good is Moritz (1980). Similarly difficult is Molodensky et al. (1962). Worth reading also from the perspective of physical geodesy is Vaníček and Krakiwsky (1987). A new book in this field is Hofmann-Wellenhof and Moritz (2006).

Field theory and vector calculus — core knowledge

A.1   Vector calculus

In physics, we decribe many quantities as *vector quantities*. For example, force, velocity, electrostatic field, and many more. A vector behaves in the same way as the location difference between two neighbouring points. Velocity $\mathbf{v}$, force $\mathbf{F}$, location difference $\Delta \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the location vectors of points 1 and 2. In a co-ordinate transformation, the vector considered as an object does not change, but the numerical values of its *components* are co-ordinate system dependent and will change, see subsection A.2.2.

About notation: in printed text, vectors are most often written in **bold**. In handwritten text one may use an arrow above the symbol: $\overrightarrow{v}$.

A.1.1   Scalar product

Between two vectors, a *scalar product* or dot product can be defined, which is itself a scalar value. A scalar is in physics a single numeric value, e.g., pressure or temperature. In the case of a scalar product of two vector *fields*, the value is tied to a location, but, even if a co-ordinate transformation changes the co-ordinate values of the location, the scalar itself remains unchanged.

An example of a scalar product: *work* ΔE is

$$\Delta E = \langle \mathbf{F} \cdot \Delta \mathbf{r} \rangle ,$$

the scalar product of force $\mathbf{F}$ and path $\Delta\mathbf{r}$. Often, also in the sequel, we leave the angle brackets $\langle\cdot\rangle$ off.

Later we shall see that if the points 1 and 2, $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, are very close to each other, we may write

$$dE = \langle \mathbf{F} \cdot d\mathbf{r} \rangle,$$

in which $d\mathbf{r}$ and dE are infinitesimal elements of path and energy. If now there is between the points A and B a curved path, we may get from this an integral equation:

$$\Delta E_{AB} = \int_A^B \langle \mathbf{F} \cdot d\mathbf{r} \rangle.$$

This is the *work integral*.

A.1.2   The scalar product, formally

Let

$$s \stackrel{\text{def}}{=} \langle \mathbf{a} \cdot \mathbf{b} \rangle$$

be the scalar product of the vectors $\mathbf{a}$ and $\mathbf{b}$. Then

$$\langle \mu\mathbf{a} \cdot \mathbf{b} \rangle = \langle \mathbf{a} \cdot \mu\mathbf{b} \rangle = \mu \langle \mathbf{a} \cdot \mathbf{b} \rangle,$$
$$\langle \mathbf{a} \cdot \mathbf{b} \rangle = \langle \mathbf{b} \cdot \mathbf{a} \rangle,$$

and often we call

$$\|\mathbf{a}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{a} \cdot \mathbf{a} \rangle}$$

the *norm* or length of vector $\mathbf{a}$.

The following also applies:

$$\langle \mathbf{a} \cdot \mathbf{b} \rangle = \|\mathbf{a}\| \|\mathbf{b}\| \cos\alpha,$$

where α is the angle between the directions of the vectors $\mathbf{a}$ and $\mathbf{b}$.

A.1.3   Exterior or vectorial product

The exterior product, or cross product, of two vectors is itself a vector called the *vectorial product* (at least in three-dimensional Euclidean space $\mathbb{R}^3$). For example, the *angular momentum* $\mathbf{q}$:

$$\mathbf{q} = \langle \mathbf{r} \times \mathbf{p} \rangle,$$
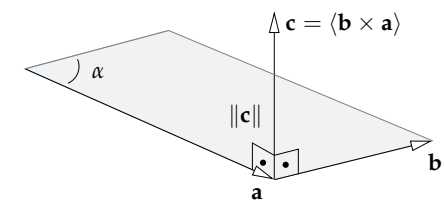
Figure A.1. Exterior or vectorial product.

where $\mathbf{p} = m\mathbf{v}$ is linear momentum, $\mathbf{r}$ the location vector of the body relative to some origin, m the mass of the body, and $\mathbf{v} = \dfrac{d\mathbf{r}}{dt}$ is the time derivative of the location, or *velocity*. We write

$$\mathbf{q} = m \left\langle \mathbf{r} \times \frac{d\mathbf{r}}{dt} \right\rangle.$$

A.1.4   The vectorial product, formally

Let

$$\mathbf{c} \stackrel{\text{def}}{=} \langle \mathbf{a} \times \mathbf{b} \rangle$$

be the vectorial product of the two vectors $\mathbf{a}$ and $\mathbf{b}$. Then ($\mu \in \mathbb{R}$):

$$\langle \mu\mathbf{a} \times \mathbf{b} \rangle = \langle \mathbf{a} \times \mu\mathbf{b} \rangle = \mu \langle \mathbf{a} \times \mathbf{b} \rangle,$$
$$\langle \mathbf{a} \times \mathbf{b} \rangle = - \langle \mathbf{b} \times \mathbf{a} \rangle,$$

and thus $\langle \mathbf{a} \times \mathbf{a} \rangle = 0$.

The resulting vector $\mathbf{c}$ is always *orthogonal* to the vectors $\mathbf{a}$ and $\mathbf{b}$; the length of vector $\mathbf{c}$ corresponds to the surface area of the parallellogram spanned by   suunnikas the vectors $\mathbf{a}$ and $\mathbf{b}$. In a formula:

$$\|\mathbf{c}\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin\alpha,$$

where again α is the angle between the directions of vectors $\mathbf{a}$ and $\mathbf{b}$. If the angle is zero, then also the vectorial product is zero (because then, $\mathbf{a} = \mu\mathbf{b}$ for some suitable value of μ).

A.1.5   Kepler's second law

If $\mathbf{r}$ is the location vector of the body (planet) relative to the centre of motion (the Sun), and $\dfrac{d\mathbf{r}}{dt}$ is the velocity vector, then the product

$$\left\langle \mathbf{r} \times \frac{d\mathbf{r}}{dt} \right\rangle \tag{A.1}$$

is precisely twice the surface area of the triangle or "area" swept over.

Let us take the time derivative of this product, i.e., the expression A.1:

$$\frac{d}{dt}\left\langle \mathbf{r} \times \frac{d\mathbf{r}}{dt} \right\rangle = \left\langle \frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} \right\rangle + \left\langle \mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} \right\rangle .$$

Here, the first term vanishes, because $\langle \mathbf{a} \times \mathbf{a} \rangle = 0$. In the second term, we can exploit our knowledge that the attractive force $\mathbf{F}$ emanating from the Sun that causes planetary orbital motion — and also the acceleration it causes, $\mathbf{a} \stackrel{\text{def}}{=} \dfrac{d^2\mathbf{r}}{dt^2}$ — is *central*:

$$\mathbf{F} = m\mathbf{a} = -\frac{GMm}{\|\mathbf{r}\|^3}\mathbf{r} .$$

Substitute into the above:

$$\frac{d}{dt}\left\langle \mathbf{r} \times \frac{d\mathbf{r}}{dt} \right\rangle = 0 - \frac{GM}{\|\mathbf{r}\|^3}\langle \mathbf{r} \times \mathbf{r} \rangle = 0 .$$

So: *the quantity* $\left\langle \mathbf{r} \times \dfrac{d\mathbf{r}}{dt} \right\rangle$ — *angular momentum per unit of mass* $\frac{q}{m}$ — *is conserved.* Like, e.g., the total amount of energy, electric charge and many other quantitites, the amount of angular momentum in a closed system is constant.

G is the universal gravitational constant, M is the mass of the Sun, m the mass of the planet.

A.2   Scalar and vector fields

A.2.1   Definitions

In the space $\mathbb{R}^3$ we may define functions, or *fields*.
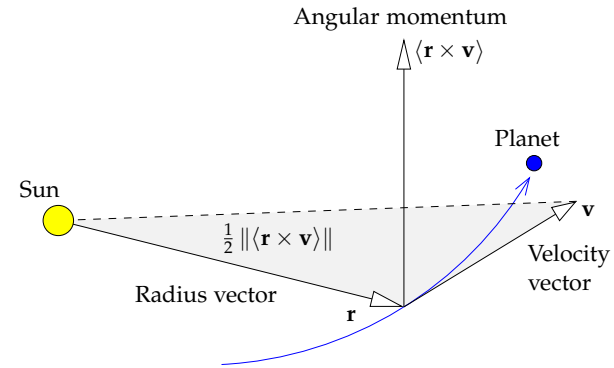
Figure A.2. Kepler's second law. In the same amount of time, the radius vector of a planet will "sweep over" a same-sized area — conservation of angular momentum.

A scalar field is a scalar-valued function, which is defined throughout the space (or a part of it), e.g., temperature T:

$$T(\mathbf{r}) .$$

I.e., for every value of the location vector $\mathbf{r}$ there is a temperature value T.

A vector field is a vector-valued function that again is defined throughout space, e.g., the electrostatic field $\mathbf{E}$:

$$\mathbf{E}(\mathbf{r}) .$$

A.2.2   A basis in space

In the space $\mathbb{R}^3$ we may choose a *basis* made up of three vectors which *span* the space in question. Generally we choose three basis vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, that are orthogonal to each other, and the norms, or lengths, of which are 1, a so-called *orthonormal basis*:

$$\mathbf{i} \perp \mathbf{j}, \ \mathbf{i} \perp \mathbf{k}, \ \mathbf{j} \perp \mathbf{k}, \ \|\mathbf{i}\| = \|\mathbf{j}\| = \|\mathbf{k}\| = 1 .$$

Now we may write vectors out into their *components*:

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$$

and also scalar and vectorial products can now be calculated with the aid of their components:

$$s = \langle \mathbf{a} \cdot \mathbf{b} \rangle = \langle (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \cdot (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) \rangle = a_1b_1 + a_2b_2 + a_3b_3 = \sum_{i=1}^{3} a_ib_i,$$

using the above identities for the basis vectors.

For the vectorial product, the calculation is more involved; we get as the final outcome the determinant

$$\mathbf{c} = \langle \mathbf{a} \times \mathbf{b} \rangle =$$

$$= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} =$$

$$= (a_2b_3 - a_3b_2)\,\mathbf{i} + (a_3b_1 - a_1b_3)\,\mathbf{j} + (a_1b_2 - a_2b_1)\,\mathbf{k}.$$

I.e.,

$$c_1 = a_2b_3 - a_3b_2,$$
$$c_2 = a_3b_1 - a_1b_3,$$
$$c_3 = a_1b_2 - a_2b_1.$$

Also these expressions are determinants:

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \left[\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \quad \begin{vmatrix} a_3 & a_1 \\ b_3 & b_1 \end{vmatrix} \quad \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right]^{\mathsf{T}}.$$

A.2.3   The nabla operator

The *location vector* $\mathbf{r}$ can be written on the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ basis as follows:

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

which defines (x, y, z) co-ordinates in space.

Let us now define a vector operator called *nabla* (∇) as follows:

$$\nabla \overset{\text{def}}{=} \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}.$$

The operator, or function, is on its own without meaning. It acquires meaning only when it *operates* on something, in which case the three partial derivatives on the right-hand side can be calculated.

A.2.4   The gradient

Let $F(\mathbf{r}) = F(x, y, z)$ be a scalar field in space. The nabla operator will give its *gradient* $\mathbf{g}$, a vector field in the same space:

$$\mathbf{g} = \text{grad}\,F = \nabla F = \mathbf{i}\frac{\partial F}{\partial x} + \mathbf{j}\frac{\partial F}{\partial y} + \mathbf{k}\frac{\partial F}{\partial z}.$$

So, the field $\mathbf{g}(\mathbf{r}) = \mathbf{g}(x, y, z)$ is a vector field in the same space, the gradient field of F.

Interpretation: the gradient describes the *slope* of the scalar field. The direction of the vector is the direction in which the value of the scalar field changes fastest, and its length describes the rate of change with location. Imaging a hilly landscape: the height if the ground above sea level is the scalar field, and its gradient is pointing everywhere *uphill*, away from the valleys toward the hilltops. The $\mathbf{g}$ arrows are the longer, the steeper is the slope of the ground surface.

The gradient operator (like also the divergence and the curl, see later) is *linear*:

$$\text{grad}\,(F + G) = \text{grad}\,F + \text{grad}\,G.$$

——

A.2.5   The divergence

Given a vector field $\mathbf{a}(x, y, z)$. We form the *scalar product* s of this and the nabla operator:

$$s = \text{div}\,\mathbf{a} = \langle \nabla \cdot \mathbf{a} \rangle = \frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_3}{\partial z}.$$

Interpretation: The divergence describes the "sources" of a vector field, both the positive and the negative ones. Imagine the velocity of flow of water as a vector field. At the locations of the "sources" the divergence
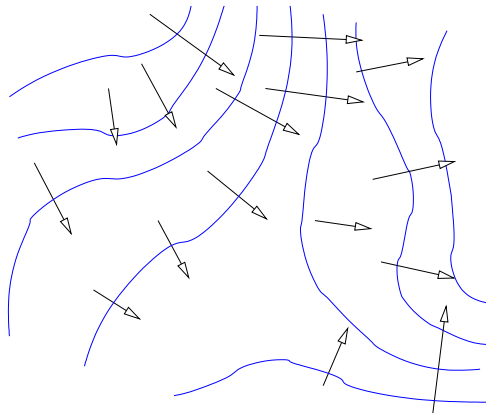
Figure A.3. The gradient. The level curves of the scalar field in blue.

is positive, at the locations of the "sewer holes" or *sinks*, negative, everywhere else zero (because liquid cannot appear out of nothing or disappear into nothing).
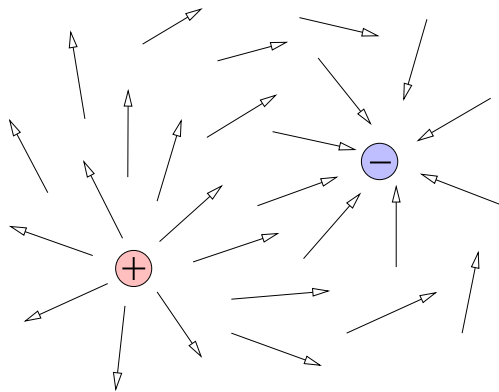
Figure A.4. The divergence. Positive divergences ("sources") and negative ones ("sinks").

A.2.6   The curl

Given a vector field $\mathbf{a}(x, y, z)$. We form the *vectorial product* $\mathbf{c}$ of this and the nabla operator, producing again a vector field:

$$
\mathbf{c} = \mathrm{rot}\,\mathbf{a} = \nabla \times \mathbf{a} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_1 & a_2 & a_3 \end{vmatrix} =
$$

$$
= \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ a_1 & a_2 \end{vmatrix} \mathbf{i} - \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ a_1 & a_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ a_2 & a_3 \end{vmatrix} \mathbf{k} =
$$

$$
= \left(\frac{\partial a_3}{\partial y} - \frac{\partial a_2}{\partial z} \right) \mathbf{i} + \left(\frac{\partial a_1}{\partial z} - \frac{\partial a_3}{\partial x} \right) \mathbf{j} + \left(\frac{\partial a_2}{\partial x} - \frac{\partial a_1}{\partial y} \right) \mathbf{k},
$$

using the evaluation rules for determinants.

Interpretation: the curl describes the *eddiness* or *turbulence* present in a vector field.

Imagine a weather map, where low- and high-pressure zones are drawn. Our vector field is the wind field. The wind circulates (on the Northern hemisphere) clockwise around the high-pressure zones, and counterclockwise around the low-pressure zones. We may say that the curl of the wind field is positive at the high pressures and negative at the low pressures.

(This is a poor metaphor as it is two-dimensional. In $\mathbb{R}^2$, the curl is a scalar, not a vector. Just like we need only one angle to describe a rotational motion, when in $\mathbb{R}^3$ we need the three Euler angles.)

Field theory and vector calculus — core knowledge
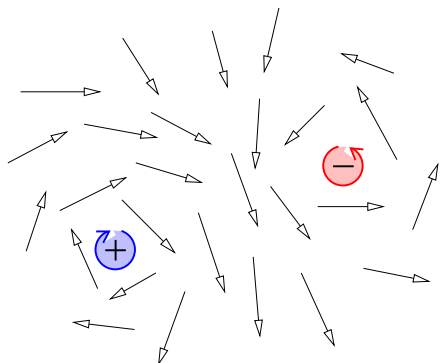
Figure A.5. The curl. Positive (clockwise) and negative (counterclockwise) eddies.

A.2.7 Conservative fields

What happens if a vector field **a** is the gradient of a scalar field F, and we try to calculate its curl? As follows:

$$\operatorname{rot} \mathbf{a} = \operatorname{rot} \operatorname{grad} F =$$

$$= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z} \end{vmatrix} =$$

$$= \left(\frac{\partial}{\partial y} \frac{\partial}{\partial z} F - \frac{\partial}{\partial z} \frac{\partial}{\partial y} F \right) \mathbf{i} + \left(\frac{\partial}{\partial z} \frac{\partial}{\partial x} F - \frac{\partial}{\partial x} \frac{\partial}{\partial z} F \right) \mathbf{j} + \left(\frac{\partial}{\partial x} \frac{\partial}{\partial y} F - \frac{\partial}{\partial y} \frac{\partial}{\partial x} F \right) \mathbf{k} =$$

$$= 0!$$

In other words, if the vector field $\mathbf{a}(x,y,z)$ is the gradient of the scalar field $F(x,y,z)$, *its curl will vanish.*

Definition: this kind of vector field **a** is called *conservative*, and the corresponding scalar field F, $\mathbf{a} = \operatorname{grad} F$, is called the *potential* of field **a**.

_____

Note that if

$$\mathbf{a}(x,y,z) = \operatorname{grad} F(x,y,z),$$

then also

$$\mathbf{a}(x,y,z) = \operatorname{grad}\left(F(x,y,z) + F_0\right),$$

where F_0 is a constant, because

$$\operatorname{grad} F_0 = \mathbf{i}\frac{\partial F_0}{\partial x} + \mathbf{j}\frac{\partial F_0}{\partial y} + \mathbf{k}\frac{\partial F_0}{\partial z} = 0.$$

So the potential is *not uniquely* defined.

A.2.8 The Laplace operator

Assume a conservative field **a**, i.e., rot $\mathbf{a} = 0$. Then we may write

$$\mathbf{a} = \operatorname{grad} F = \nabla F,$$

where F is the potential.

Let us now express the *divergence* of field **a** into the potential:

$$\operatorname{div} \mathbf{a} = \nabla \mathbf{a} = \nabla \nabla F = \frac{\partial}{\partial x}\frac{\partial}{\partial x}F + \frac{\partial}{\partial y}\frac{\partial}{\partial y}F + \frac{\partial}{\partial z}\frac{\partial}{\partial z}F = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) F \stackrel{\text{def}}{=} \Delta F,$$

where we have introduced a new *differential operator*: the Delta operator invented by the French Pierre-Simon de Laplace,

$$\Delta \stackrel{\text{def}}{=} \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

For the potential of a "source free" field — e.g., for the gravitational potential in vacuum, the electrostatic potential in an area of space free of electric charges — this Delta, or Laplace, operator vanishes.

A.3 Integrals

A.3.1 The curve integral

We saw earlier, that work ΔE can be written as the scalar product of force **F** and path $\Delta \mathbf{r}$:

$$\Delta E = \langle \mathbf{F} \cdot \Delta \mathbf{r} \rangle.$$
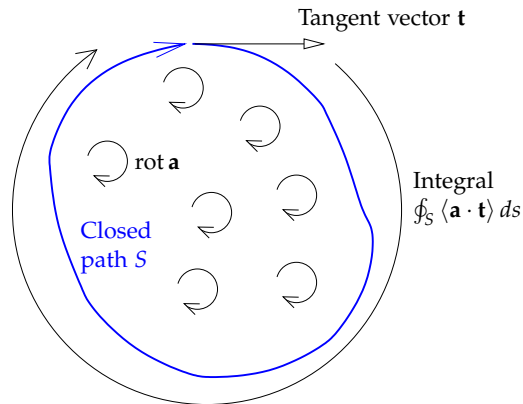
Figure A.6. The Stokes curl theorem.

The differential form of this is

$$dE = \langle \mathbf{F} \cdot d\mathbf{r} \rangle,$$

from which one obtains the integral form, the *work integral*

$$\Delta E_{AB} = \int_A^B \langle \mathbf{F} \cdot d\mathbf{r} \rangle.$$

Here is computed the amount of work needed to move a body from point A to point B by integrating $\langle \mathbf{F} \cdot d\mathbf{r} \rangle$ along the path AB.

If we parametrize the path according to arc length s, and the tangent vector to the path is called

$$\mathbf{t} = \frac{dx}{ds}\mathbf{i} + \frac{dy}{ds}\mathbf{j} + \frac{dz}{ds}\mathbf{k},$$

we may also write

$$\Delta E_{AB} = \int_A^B \langle \mathbf{F} \cdot \mathbf{t} \rangle \, ds,$$

the parametrized version of the integral.

A.3.2   The surface integral

Assume given again some vector field $\mathbf{a}$ and a surface in space S. Often, one needs to integrate over the surface S the *normal component* of a vector field, the projection of $\mathbf{a}$ onto the normal vector of the surface.

Let the normal vector of the surface be $\mathbf{n}$. Then we must integrate

$$\iint_S \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS,$$

symbolically written

$$\iint_S \langle \mathbf{a} \cdot d\mathbf{S} \rangle,$$

where the notation $d\mathbf{S}$ is called an *oriented surface element*. It is a vector pointing in the same direction as the normal vector $\mathbf{n}$.

Like a curve, also a surface can be *parametrized*. For example, the Earth's surface (assumed a sphere) can be parametrized by latitude ϕ and longitude λ: $\mathbf{r} = \mathbf{r}(\phi, \lambda)$. In this case we write as the surface element

$$dS = R^2 \cos \phi \, d\phi \, d\lambda,$$

in which $R^2 \cos \phi$ is *Jacobi's determinant*, of the parameter pair (ϕ, λ). In this parametrization, the integral is calculated as follows:

$$\iint_S \langle \mathbf{a} \cdot d\mathbf{S} \rangle = \iint_S \langle \mathbf{a} \cdot \mathbf{n} \rangle \, R^2 \cos \phi \, d\phi \, d\lambda.$$

Other surfaces and parametrizations have other Jacobi's determinants. The determinant always describes the *true* area of a "parameter surface element" $d\phi \, d\lambda$ "in nature". For example on the Earth's surface, a degree times degree patch is the largest near the equator. In polar co-ordinates (ρ, θ) in the plane $(x = \rho \cos \theta, y = \rho \sin \theta)$ the determinant of Jacobi is ρ. In the ordinary (x, y) parametrization in the plane, it is 1 and thus can be left out altogether.

A.3.3   The Stokes curl theorem

Let S be a surface in space (not necessarily flat) and ∂S its edge curve. Assume that the surface and its edge are well-behaved enough for all necessary integations and differentiations to be possible. Then (Stokes):

$$\iint_S \langle \text{rot } \mathbf{a} \cdot d\mathbf{S} \rangle = \oint_{\partial S} \langle \mathbf{a} \cdot d\mathbf{r} \rangle.$$

In words: The surface integral of the curl of a vector field over a surface is the same as the closed path integral of the field around the edge of the surface.

382

Field theory and vector calculus — core knowledge

Special case: If rot $\mathbf{a} = 0$ everywhere (a conservative vector field) then

$$\oint_{\partial S} \langle \mathbf{a} \cdot d\mathbf{r} \rangle = 0,$$

i.e., also

$$\int_{A,\,\text{path1}}^{B} \langle \mathbf{a} \cdot d\mathbf{r} \rangle = \int_{A,\,\text{path2}}^{B} \langle \mathbf{a} \cdot d\mathbf{r} \rangle .$$

In other words, *the work integral from point A to point B does not depend on the path chosen.* And *the work done by a body transported around a closed path is zero.*

This explains perhaps better the essence of a conservative force field. A conservative field can be represented as *the gradient of a potential*: $\mathbf{a} = \operatorname{grad} F$, where F is the potential of the field. The Earth's gravity vector field $\mathbf{g}(x,y,z)$ is the gradient of the Earth's gravity potential $W(x,y,z)$. At mean sea level — more precisely, at the geoid — the gravity *potential* is constant; the gravity *vector* $\mathbf{g}$ stands everywhere orthogonally on the geoid.

_____

A.3.4 The Gauss integral theorem

Let V be a certain volume of space, and ∂V its closed boundary, a union of surfaces. Assume again that both are mathematically well behaved. Then the following theorem applies (Gauss):

$$\iiint_V \operatorname{div} \mathbf{a} \, dV = \iint_{\partial V} \langle \mathbf{a} \cdot d\mathbf{S} \rangle = \iint_{\partial V} \langle \mathbf{a} \cdot \mathbf{n} \rangle \, dS.$$

In words: what is created inside a body (i.e., "sources", divergence) must come out through its surfaces.

Remark: generally the orientation of surface ∂V is taken as positive on the outside, i.e., the normal vector $\mathbf{n}$ of the surface points outward.

_____

A.4 The continuity of matter

An often used equation in, e.g., hydro- or aerodynamics is the *continuity equation*. This describes that matter cannot just disappear or increase in
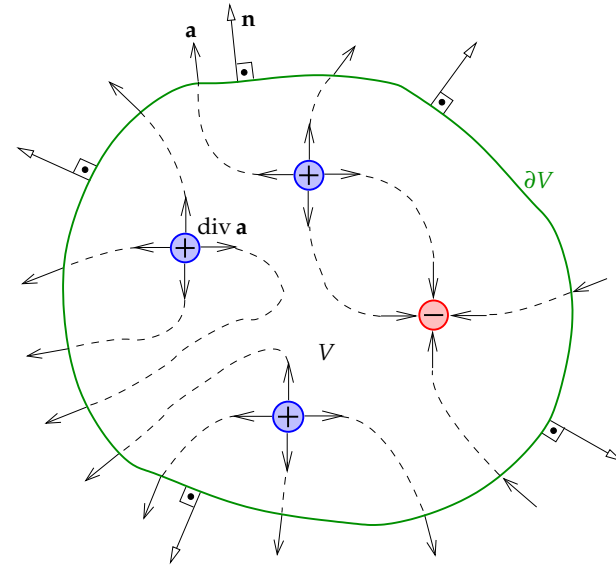
Figure A.7. The Gauss integral theorem. $\mathbf{n}$ is the normal vector to the exterior surface. Note that the Gauss intregral theorem can also be formulated with the aid of (Michael Faraday's) *field lines*: a field line starts or terminates on an electric charge (i.e., a place where $\operatorname{div} \mathbf{a} \neq 0$) or runs to infinity (i.e., through the surface ∂V).

amount. In the general case, this equation looks like this:

$$\operatorname{div}(\rho \mathbf{v}) + \frac{d}{dt}\rho = 0.$$

Here, the expression $\rho \mathbf{v}$ describes *mass currents*; ρ is the matter density, $\mathbf{v}$ is the current flow velocity. The term $\operatorname{div}(\rho \mathbf{v})$ describes how much more matter, in a unit of time, exits the volume element than enters it, per unit of volume. The second term again, $\frac{d}{dt}\rho$, describes the change in the amount of matter inside the volume element over time. The two terms must balance for the "matter accounting" to close.

If the moving fluid is incompressible, then ρ is constant and

$$\frac{d}{dt}\rho = 0 \quad \text{and} \quad \operatorname{div}(\rho \mathbf{v}) = \rho \operatorname{div} \mathbf{v},$$

Field theory and vector calculus — core knowledge

and we obtain

$$\rho \operatorname{div} \mathbf{v} = 0 \implies \operatorname{div} \mathbf{v} = 0.$$

Remember, however, that *not* necessarily $\operatorname{rot} \mathbf{v} = 0$ — i.e., the flow isn't necessarily eddy free — i.e., a potential F for which $\mathbf{v} = \operatorname{grad} F$ does not necessarily exist.

Function spaces

B

B.1 An abstract vector space

In an *abstract vector space* we may create a *basis*, with the help of which each vector can be written as a linear combination of the basis vectors: e.g., if the basis, in a three-dimensional space, is $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, we may write an arbitrary vector $\mathbf{r}$ in the form

$$\mathbf{r} = r_1 \mathbf{e}_1 + r_2 \mathbf{e}_2 + r_3 \mathbf{e}_3 = \sum_{i=1}^{3} r_i \mathbf{e}_i.$$

Precisely because three basis vectors are always enough, we call the ordinary (Euclidean) space $\mathbb{E}$ three-dimensional, also $\mathbb{R}^3$.

In a vector space one can define a scalar product, which is a linear mapping from two vectors to one number ("bilinear form"):

$$\langle \mathbf{r} \cdot \mathbf{s} \rangle .$$

Linearity means that

$$\langle \alpha \mathbf{r}_1 + \beta \mathbf{r}_2 \cdot \mathbf{s} \rangle = \alpha \langle \mathbf{r}_1 \cdot \mathbf{s} \rangle + \beta \langle \mathbf{r}_2 \cdot \mathbf{s} \rangle ,$$

and commutativity, that

$$\langle \mathbf{r} \cdot \mathbf{s} \rangle = \langle \mathbf{s} \cdot \mathbf{r} \rangle$$

If the basis vectors are orthogonal to each other, in other words, $\langle \mathbf{e}_i \cdot \mathbf{e}_j \rangle = 0$ if $i \neq j$, we may calculate the coefficients r_i simply as

$$\mathbf{r} = \sum_{i=1}^{3} \frac{\langle \mathbf{r} \cdot \mathbf{e}_i \rangle}{\langle \mathbf{e}_i \cdot \mathbf{e}_i \rangle} \mathbf{e}_i \tag{B.1}$$

If, additionally, also $\langle \mathbf{e}_i \cdot \mathbf{e}_i \rangle = \|\mathbf{e}_i\|^2 = 1 \ \forall i \in \{1, 2, 3\}$, i.e., the basis vectors are *orthonormal*, equation B.1 becomes simpler still:

$$\mathbf{r} = \sum_{i=1}^{3} \langle \mathbf{r} \cdot \mathbf{e}_i \rangle \mathbf{e}_i = \sum_{i=1}^{3} r_i \mathbf{e}_i. \tag{B.2}$$

Here, the coefficients are, according to definition, $r_i = \langle \mathbf{r} \cdot \mathbf{e}_i \rangle$. The quantity

$$\|\mathbf{e}_i\| = \sqrt{\langle \mathbf{e}_i \cdot \mathbf{e}_i \rangle}$$

is called the *norm* of the vector $\mathbf{e}_i$.

B.2 Fourier function space

B.2.1 Description

Also *functions* can be considered elements in a vector space. If we define the scalar product of two functions f, g as the following integral:

$$\left\langle \overrightarrow{f} \cdot \overrightarrow{g} \right\rangle \overset{\text{def}}{=} \frac{1}{\pi} \int_{0}^{2\pi} f(x) \, g(x) \, dx, \tag{B.3}$$

it is easily verified that the above requirements for a scalar product are met.

One basis in this vector space (a *function space*) is formed by the so-called Fourier basis functions,

$$\begin{aligned}
\overrightarrow{e_0} &= \frac{1}{2}\sqrt{2} \quad (k = 0) \\
\overrightarrow{e_k} &= \cos kx, \quad k = 1, 2, 3, \dots \\
\overrightarrow{e_{-k}} &= \sin kx, \quad k = 1, 2, 3, \dots
\end{aligned} \tag{B.4}$$

This basis is *orthonormal* (proof: exercise). It is also a *complete* basis, which we shall not prove. As the number of basis vectors is countably infinite, we say that this function space is infinitely dimensional.

Now every function $f(x)$ meeting certain conditions can be expanded in the way of equation (B.2), i.e.,

$$f(x) = \frac{1}{2}a_0\sqrt{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

— the familiar Fourier expansion — in which the coefficients are

$$a_0 = \left\langle \overrightarrow{f} \cdot \overrightarrow{e_0} \right\rangle = \frac{1}{2\pi}\sqrt{2} \int_{0}^{2\pi} f(x) \, dx = \sqrt{2} \cdot \overline{f(x)},$$

$$a_k = \left\langle \overrightarrow{f} \cdot \overrightarrow{e_k} \right\rangle = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \cos kx \, dx, \quad k = 1, 2, \dots$$

$$b_k = \left\langle \overrightarrow{f} \cdot \overrightarrow{e_{-k}} \right\rangle = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \sin kx \, dx, \quad k = 1, 2, \dots$$

This is the familiar way in which the coefficients of a Fourier series are calculated.

B.2.2 Example

As an example of Fourier analysis we may take a step function on the interval $[0, 2\pi)$:

$$f(x) = \begin{cases} 0 & x \in [0, \pi) \\ 1 & x \in [\pi, 2\pi) \end{cases}.$$

We can derive the Fourier coefficients of this function as follows:

$$a_0 = \frac{1}{2\pi}\sqrt{2} \cdot \int_{0}^{2\pi} f(x) \, dx = \frac{1}{2\pi}\sqrt{2} \cdot \pi = \frac{1}{2}\sqrt{2},$$

$$a_k = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \cos kx \, dx = \frac{1}{\pi} \int_{\pi}^{2\pi} \cos kx \, dx =$$

$$= \frac{1}{\pi} \left[\frac{1}{k} \sin kx \right]_{\pi}^{2\pi} = \frac{1}{k\pi} \left\{ \sin 2k\pi - \sin k\pi \right\} = 0,$$

$$b_k = \frac{1}{\pi} \int_{0}^{2\pi} f(x) \sin kx \, dx = \frac{1}{\pi} \int_{\pi}^{2\pi} \sin kx \, dx =$$

$$= \frac{1}{\pi} \left[-\frac{1}{k} \cos kx \right]_{\pi}^{2\pi} = \frac{1}{k\pi} \left\{ \cos k\pi - \cos 2k\pi \right\}$$

$$= \frac{1}{k\pi} \left\{ (-1)^k - 1 \right\} = \begin{cases} 0 & \text{if } k \text{ even,} \\ -\frac{2}{k\pi} & \text{if } k \text{ odd.} \end{cases}$$
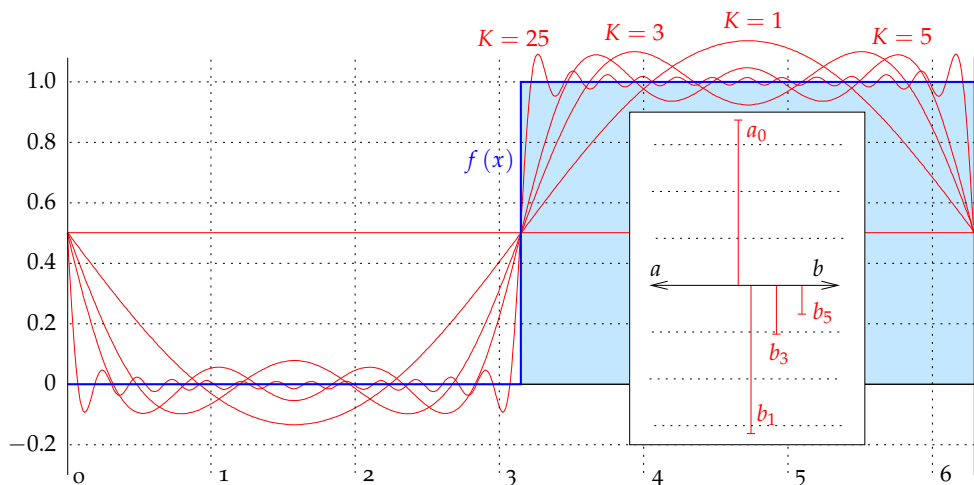
Figure B.1. Fourier analysis on a step function. Plotted are the Fourier expansions $f^{(K)}(x) \stackrel{\text{def}}{=} \frac{1}{2}a_0\sqrt{2} + \sum_{k=1}^{K} b_k \sin kx = \frac{1}{2} - \frac{2}{\pi}\sum_{k=1,3,5,\dots}^{K} \frac{1}{k}\sin kx$, for values of K of 1, 3, 5, and 25. The inset gives the spectrum of the function.

Or, in numbers, $a_0 = \frac{1}{2}\sqrt{2} = 0.70710, b_1 = -\frac{2}{\pi} = -0.63662\dots, b_3 = -\frac{2}{3\pi} = -0.21220, b_5 = -0.12732$, and so forth. The expansion becomes now

$$f(x) = \frac{1}{2}\sqrt{2}a_0 + \sum_{k=1}^{\infty} b_k \sin kx = \frac{1}{2} - \frac{2}{\pi}\sum_{k=1,3,5,\dots}^{\infty} \frac{1}{k}\sin kx =$$

$$= \frac{1}{2} - \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{2n-1}\sin(2n-1)\,x.$$

Note that it only contains sines, no cosines. This is a consequence of the function's symmetry properties.

In figure B.1 we show truncated expansions for this function.

B.2.3   Convergence

The Fourier expansion converges in the square integral sense, i.e., if we define the truncated expansion

$$f^{(K)}(x) \stackrel{\text{def}}{=} \frac{1}{2}a_0\sqrt{2} + \sum_{k=1}^{K} (a_k \cos kx + b_k \sin kx),$$

then

$$\lim_{K\to\infty} \frac{1}{\pi}\int_0^{2\pi} \left\{ f^{(K)}(x) - f(x) \right\}^2 dx = 0.$$

Note that this does not mean that, for every $x \in [0, 2\pi)$, $f^{(K)}(x) \to f(x)$ when $K \to \infty$. Looking at figure B.1, there will always remain a small neighbourhood of $x = \pi$ where the absolute difference $\left| f^{(K)}(x) - f(x) \right|$ will be almost $\frac{1}{2}$, even for arbitrarily large values of K. We say that the Fourier expansion is *convergent*, but not *uniformly* convergent.

The Fourier series converges pointwise "almost everywhere" in $x \in [0, 2\pi)$: in all points except in the two special points $x = 0$ and $x = \pi$.

Also, note the "shoulder" of the expansion even for $K = 25$. This shoulder will get narrower for higher K, but not any lower, remaining at approximately 0.09. It is known as the Gibbs phenomenon.

B.3   Sturm–Liouville differential equations

B.3.1   The eigenvalue problem

In an abstract vector space we may formulate an eigenvalue problem: if there exists a linear operator (mapping) L, we may write

$$L\mathbf{x} - \lambda\mathbf{x} = 0,$$

where the problem consists of determining the values λ_i for which solutions $\mathbf{x}_i$ exist.

In a concrete n-dimensional vector space we may write the vector

$$\mathbf{x} = \sum_{i=1}^{n} x_i\mathbf{e}_i,$$

Function spaces

and, thanks to linearity,

$$Lx = L\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} x_i \cdot L e_i;$$

on the other hand we may write n different vectors $L e_i$ on the basis $\{e_j\}$ in the following way:

$$L e_i = \sum_{j=1}^{n} a_{ij} e_j.$$

This defines the coefficients a_{ij}, which may be collected into an $n \times n$ matrix A.

Now by substitution

$$Lx = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ij} x_i\right) e_j, \tag{B.5}$$

when also

$$\lambda x = \lambda \sum_{i=1}^{n} x_i e_i = \sum_{j=1}^{n} \left(\lambda x_j\right) e_j. \tag{B.6}$$

By combining equations B.5, B.6, of which all coefficients must be identical, we obtain

$$\sum_{i=1}^{n} a_{ij} x_i - \lambda x_j = 0, j = 1, \ldots, n,$$

or, as a matric equation,

$$A\bar{x} - \lambda\bar{x} = 0, \tag{B.7}$$

in which A is a matrix consisting of the coefficients a_{ij}, and $\bar{x}$ a column vector consisting of the coefficients x_i: $\bar{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathsf{T}}$.

Of course also equation B.7 represents an eigenvalue problem, but now in the linear vector space consisting of all *coefficient vectors* $\bar{x}$. Every $\bar{x}$ is the numerical representation of a vector x on the chosen basis $\{e_i\}$. The matrix A again is the numerical representation of operator L on the same basis[1].

---

[1] An advantage of the numerical representations is of course that one can really calculate with them.

Sturm–Liouville differential equations

B.3.2 A self-adjoint operator

Let L be a linear operator in a vector space where there exists a *scalar product*, i.e., a bilinear form $\langle x \cdot y \rangle$ which is symmetric or commutative.

Then L is *self-adjoint*, if for each pair of vectors x, y it holds that

$$\langle x \cdot Ly \rangle = \langle Lx \cdot y \rangle.$$

If the *matrix A* is self-adjoint, that means that

$$\langle \bar{x} \cdot A\bar{y} \rangle = \langle A\bar{x} \cdot \bar{y} \rangle$$

i.e.,

$$\sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} a_{ij} y_j\right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} x_j\right) y_i,$$

which is trivially true if, for all i, j values $1, \ldots, n$,

$$a_{ij} = a_{ji}, \text{ i.e., } A = A^{\mathsf{T}}.$$

In other words:

> *A symmetric matrix is a self-adjoint operator.*

From linear algebra it is undoubtedly familiar, that *the eigenvectors $\bar{x}_p, \bar{x}_q$ belonging to different eigenvalues $\lambda_p \neq \lambda_q$ of a symmetric $n \times n$ matrix are mutually orthogonal: $\bar{x}_p \perp \bar{x}_q$.* If all eigenvalues $\lambda_p, p = 1, \ldots, n$ are different, then the eigenvectors $\bar{x}_p, p = 1, \ldots, n$ will constitute a *complete orthogonal basis*[2] in the vector space $\mathbb{R}^n$.

The proof is not hard. We start from the equation for the eigenvalue problem for eigenvectors and -values x_p, λ_p:

$$Lx_p = \lambda_p x_p,$$

---

[2] Actually the eigenvectors may be arbitrarily re-scaled: if x is an eigenvector, then also $e \stackrel{\text{def}}{=} \frac{x}{\|x\|}$ is. Thus we obtain an *orthonormal basis*.

and multiply from the left by vector $\mathbf{x}_q$:

$$\langle \mathbf{x}_q \cdot L\mathbf{x}_p \rangle = \lambda_p \langle \mathbf{x}_q \cdot \mathbf{x}_p \rangle .$$

Similarly for eigenvectors and -values $\mathbf{x}_q, \lambda_q$ multiplied from the left by vector $\mathbf{x}_p$:

$$\langle \mathbf{x}_p \cdot L\mathbf{x}_q \rangle = \lambda_q \langle \mathbf{x}_p \cdot \mathbf{x}_q \rangle .$$

If now L is self-adjoint, then

$$\langle \mathbf{x}_q \cdot L\mathbf{x}_p \rangle = \langle L\mathbf{x}_q \cdot \mathbf{x}_p \rangle = \langle \mathbf{x}_p \cdot L\mathbf{x}_q \rangle$$

(remember that the scalar product is symmetric) and thus, that

$$\lambda_p \langle \mathbf{x}_q \cdot \mathbf{x}_p \rangle = \lambda_q \langle \mathbf{x}_p \cdot \mathbf{x}_q \rangle$$

i.e.,

$$(\lambda_p - \lambda_q) \langle \mathbf{x}_p \cdot \mathbf{x}_q \rangle = 0.$$

If $\lambda_p \neq \lambda_q$, we thus must have $\langle \mathbf{x}_p \cdot \mathbf{x}_q \rangle = 0$, or $\mathbf{x}_p \perp \mathbf{x}_q$. What was to be proven.

Example: *the variance matrix of location in the plane.* The variance matrix of the co-ordinates of point P in the plane is

$$\mathrm{Var}\{\mathbf{x}_P\} = \Sigma_{PP} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix},$$

a *symmetric* matrix. Here, σ_x^2 and σ_y^2 are the variances, or squares of the mean errors, of the x and y co-ordinates, whereas σ_{xy} is the covariance between the co-ordinates.

The eigenvalues of this matrix Σ_{PP} are the solutions of the *characteristic equation*

$$\det \begin{bmatrix} \sigma_x^2 - \lambda & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 - \lambda \end{bmatrix} = 0,$$

i.e.,

$$(\sigma_x^2 - \lambda)(\sigma_y^2 - \lambda) - \sigma_{xy}^2 = 0.$$

This yields

$$\lambda_{1,2} = \frac{1}{2}\left(\sigma_x^2 + \sigma_y^2\right) \pm \frac{1}{2}\sqrt{\left[\sigma_x^2 + \sigma_y^2\right]^2 - 4\left[\sigma_x^2\sigma_y^2 - \sigma_{xy}^2\right]} =$$
$$= \frac{1}{2}\left(\sigma_x^2 + \sigma_y^2\right) \pm \frac{1}{2}\sqrt{\left[\sigma_x^2 - \sigma_y^2\right]^2 + 4\sigma_{xy}^2}.$$

The variance matrix has a *variance or error ellipse.* The semi-lengths of its principal axes are $\sqrt{\lambda_1}, \sqrt{\lambda_2}$ and the directions of the principal axes are the eigenvectors of Σ_{PP}, $\mathbf{x}_1, \mathbf{x}_2$, mutually orthogonal. If the co-ordinate axes are turned into the directions of $\mathbf{x}_{1,2}$, then the matrix Σ_{PP} will assume the form

$$\Sigma'_{PP} = \begin{bmatrix} \sigma_{x'}^2 & 0 \\ 0 & \sigma_{y'}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

The sum of the eigenvalues (and the *trace* of this matrix), $\lambda_1 + \lambda_2 = \sigma_x^2 + \sigma_y^2$, is an *invariant* called the *point variance.*

B.3.3   Self-adjoint differential equations

In a function space there are also self-adjoint or "symmetric" differential equations. In fact, the most famous equations of physics are of this type.

Take a good look at, e.g., the oscillation equation

$$\frac{d^2}{dt^2}x(t) - \omega^2 x(t) = 0. \tag{B.8}$$

The solution has the general form (α amplitude, ϕ phase constant)

$$x(t) = \alpha \sin(t\omega - \phi).$$

On the interval $t \in [0, T]$ we require $x_0 = x(T)$, $\left.\dfrac{d}{dt}x\right|_{x=0} = \left.\dfrac{d}{dt}x\right|_{x=T}$, i.e., *periodicity.* These *boundary conditions* are an essential part of being self-adjoint. Then, a solution is found only for certain values of ω — *quantization.*

Note that equation B.8 is an *eigenvalue problem*, form-wise:

$$Lx - \omega^2 x = 0,$$

where the operator is

$$L = \frac{d^2}{dt^2}.$$

We first show that this operator is on the interval $[0, T]$ *self-adjoint*. If the scalar product is defined as follows:

$$\langle \vec{x} \cdot \vec{y} \rangle \stackrel{\text{def}}{=} \int_0^T x(t)\, y(t)\, dt,$$

it holds that (integration by parts):

$$\langle \vec{x} \cdot L\vec{y} \rangle = \int_0^T x(t)\, \frac{d^2}{dt^2} y(t)\, dt =$$
$$= \left[x(t)\, \frac{d}{dt} y(t) \right]_0^T - \int_0^T \frac{d}{dt} x(t)\, \frac{d}{dt} y(t)\, dt,$$
$$\langle L\vec{x} \cdot \vec{y} \rangle = \int_0^T \left(\frac{d^2}{dt^2} x(t) \right) y(t)\, dt =$$
$$= \left[\left(\frac{d}{dt} x(t) \right) y(t) \right]_0^T - \int_0^T \frac{d}{dt} x(t)\, \frac{d}{dt} y(t)\, dt.$$

As on the right-hand side the first terms vanish and the second terms are identical, we have

$$\langle \vec{x} \cdot L\vec{y} \rangle = \langle L\vec{x} \cdot \vec{y} \rangle,$$

which was to be proven.

Self-adjoint operators have eigenvalues and eigenvectors, in this case *functions*, that are mutually orthogonal for different values of ω[3]. For the oscillation equation with the above periodicity conditions they are just the solution functions

$$\sin(\omega_k t - \phi) = \sin\left(\frac{2\pi k}{T} t - \phi \right), \tag{B.9}$$

---

[3]In fact, for the same value ω there exist two mutually orthogonal periodic solutions,

$$\sin \omega_k t = \sin \frac{2\pi k t}{T}, \qquad \cos \omega_k t = \cos \frac{2\pi k t}{T}.$$

Any linear combination of these is also a valid solution, and is of the general form B.9.

in which the frequency

$$\omega_k = \frac{2\pi k}{T}$$

is quantized by a "quantum number" $k \in \mathbb{N}$.

If we let $T \to \infty$, the frequencies ω_k get closer and closer to each other, and in the end morph into a continuum.

In physics there is a broad class of differential equations that are self-adjoint in some function space. The class is known as "Sturm[4]–Liouville[5] type problems". To it belongs, e.g., the oscillation equation, Legendre's equation, Bessel's equation, and many more. Every one of them generates, in a natural way, its own set of mutually orthogonal functions that serve as the basis functions for the general solution of many partial differential equations.

B.4 Legendre polynomials

Also the ordinary Legendre polynomials $P_n(t)$ constitute a basis in a function space, with the scalar product definition

$$\langle \vec{f} \cdot \vec{g} \rangle \stackrel{\text{def}}{=} \int_{-1}^{+1} f(t)\, g(t)\, dt.$$

They don't however constitute an ortho*normal* basis, but only an ortho*gonal* one:

$$\left\| \vec{P_n} \right\|^2 = \langle \vec{P_n} \cdot \vec{P_n} \rangle = \int_{-1}^{+1} P_n^2(t)\, dt = \frac{1}{2n+1}.$$

Unlike ordinary space, which is three-dimensional, a function space is an ∞-dimensional, *abstract* vector space, that nevertheless helps us make more concrete certain abstract, but very useful fundamentals of function theory!

---

[4]Jacques Charles François Sturm FRS FAS (1803 – 1855) was an eminent French mathematician, one of the 72 names engraved on the Eiffel Tower.

[5]Joseph Liouville FRS FRSE FAS (1809 – 1882) was an eminent French mathematician.

B.5 Spherical harmonics

On the surface of the *sphere* also all functions can be considered elements of a function space. Every function meeting certain "well-behavedness properties" — like integrability — is an element. The functions

$$R_{nm}(\phi,\lambda) = P_{nm}(\sin\phi)\,\cos m\lambda,$$
$$S_{nm}(\phi,\lambda) = P_{nm}(\sin\phi)\,\sin m\lambda,$$

together, for the values $n = 0\ldots\infty, m = 0\ldots n$, form a *complete basis* for this vector space in such a way, that every function can be written as an — if necessary infinite — linear combination of these basis functions. The situation is analogous to three-dimensional space, where a complete basis consists of three vectors not in the same plane.

An alternative, more compact way of writing this is

$$Y_{nm}(\phi,\lambda) = \begin{cases} P_{nm}(\sin\phi)\,\cos m\lambda & \text{if } m \geq 0, \\ P_{n|m|}(\sin\phi)\,(\sin\phi)\sin|m|\lambda & \text{if } m < 0, \end{cases}$$

for values $n = 0\ldots\infty, m = -n\ldots n$.

In this function space, a *scalar product* is defined:

$$\left\langle \vec{V}\cdot\vec{W} \right\rangle = \frac{1}{4\pi} \iint_\sigma V(\phi,\lambda)\,W(\phi,\lambda)\,d\sigma,$$

in which σ is the surface of the unit sphere ("directional sphere", or even "celestial sphere"). According to this definition we can show, that two different functions, namely Y_{nm}, Y_{sr}, are orthogonal with respect to each other:

$$\left\langle \overrightarrow{Y_{nm}}\cdot\overrightarrow{Y_{sr}} \right\rangle = \frac{1}{4\pi}\iint_\sigma Y_{nm}(\phi,\lambda)\,Y_{sr}(\phi,\lambda)\,d\sigma = 0$$

if $n \neq s$ or $m \neq r$, etc.

The basis $\left\{\overrightarrow{Y_{nm}}\right\}$ is *orthogonal* but not *orthonormal*: the "length" of every vector differs from unity.

$$\left\|\overrightarrow{Y_{nm}}\right\|^2 = \left\langle \overrightarrow{Y_{nm}}\cdot\overrightarrow{Y_{nm}} \right\rangle = \frac{1}{4\pi}\iint_\sigma Y_{nm}^2(\phi,\lambda)\,d\sigma = \begin{cases} \frac{1}{2n+1} & \text{if } m = 0, \\ \frac{1}{2(2n+1)}\frac{(n+|m|)!}{(n-|m|)!} & \text{if } m \neq 0, \end{cases}$$

see also Heiskanen and Moritz (1967). Proving this with the help of equation 3.3 is a long process.

If we now divide the functions Y_{nm} (or, equivalently, R_{nm}, S_{nm}) by the square roots of the above factors, we obtain the *fully normalized* surface spherical harmonics $\overline{Y}_{nm}$.

With those it is again easy to calculate the coefficients $\overline{f}_{nm}$ of a given general function $f(\phi,\lambda)$ (the overline means that these are fully normalized coefficients):

$$\overline{f}_{nm} = \left\langle f\cdot\overrightarrow{\overline{Y}_{nm}} \right\rangle = \frac{1}{4\pi}\iint_\sigma f(\phi,\lambda)\,\overline{Y}_{nm}(\phi,\lambda)\,d\sigma. \tag{B.10}$$

This is a straightforward *projection* on the unit vectors of the basis (geometric analogue).

In the above integral, $f(\phi,\lambda)$ is the function f *on the Earth's surface*, i.e., if the mean radius of the Earth is R, $f(\phi,\lambda) = f(\phi,\lambda,R)$.

The equation corresponding to expansion 2.11 is

$$V(\phi,\lambda,r) = \sum_{n=0}^\infty \frac{1}{r^{n+1}} \sum_{m=0}^n \overline{P}_{nm}(\sin\phi)\left(\overline{a}_{nm}\cos m\lambda + \overline{b}_{nm}\sin m\lambda\right).$$

We may write also

$$\overline{Y}_{nm}(\phi,\lambda) = \begin{cases} \overline{P}_{nm}(\sin\phi)\,\cos m\lambda & \text{if } m \geq 0, \\ \overline{P}_{n|m|}(\sin\phi)\,\sin|m|\lambda & \text{if } m < 0, \end{cases}$$

which corresponds to the definition of the fully normalized Legendre functions:

$$\overline{P}_{n0}(\sin\phi) = \sqrt{2n+1}\,P_{n0}(\sin\phi),$$
$$\overline{P}_{nm}(\sin\phi) = \sqrt{2(2n+1)\frac{(n-m)!}{(n+m)!}}\,P_{nm}(\sin\phi)\,(\sin\phi),\ m > 0.$$

Then the above equation for the potential becomes

$$V(\phi,\lambda,r) = \sum_{n=0}^\infty \frac{1}{r^{n+1}} \sum_{m=-n}^n \overline{v}_{nm}\overline{Y}_{nm}(\phi,\lambda),$$

in which

$$\bar{v}_{nm} = \begin{cases} \bar{a}_{nm} & \text{if } m \geq 0, \\ \bar{b}_{n|m|} & \text{if } m < 0. \end{cases}$$
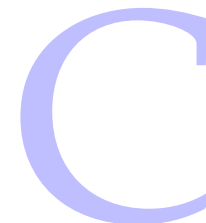
Self-test questions

1. The identity $\langle \mathbf{r} \cdot \mathbf{s} \rangle = \langle \mathbf{s} \cdot \mathbf{r} \rangle$, for two elements $\mathbf{r}$ and $\mathbf{s}$ of a vector space, expresses the property of: linearity | commutativity | associativity.

Exercise B – 1:   Orthonormality of the Fourier basis functions

Show the orthonormality of the Fourier basis functions, equation B.4 by deriving their scalar products by equation B.3.

Why does FFT work?

C

FFT is a factorization method for computing the discrete Fourier transform that spectacularly reduces the number of calculations needed and speeds up the calculation. It requires the number of grid points to be a factorizable number.

There are alternatives in choosing precisely which FFT method to use. The fastest FFT requires a grid the number of points of which is a power of 2. The size of the grid is then $2^n \times 2^m$. Alternative, "mixed-radix" methods, may also be considered and perform well if the grid size is something like 360×480, e.g., $N = 360 = 2 \times 2 \times 2 \times 3 \times 3 \times 5$. If the grid size is a prime number, FFT does not give any advantage over the ordinary discrete Fourier transform.

If the function $f(x)$ is defined on the interval[1] $x \in [0, 1)$ on an equi-spaced grid as values $f(x_k)$, $k = 0, \ldots, N - 1$, then the discrete Fourier transform in one dimension is

$$\mathcal{F}\{f(x)\} = F(\omega),$$

in which

$$F(\omega_j) = \frac{1}{N} \sum_{k=0}^{N-1} f(x_k) \exp\left(2\pi i \frac{jk}{N}\right), \quad j = 0, \ldots, N - 1. \tag{C.1}$$

Here, also the frequency argument $\omega_j, j = 0, \ldots, N - 1$ is defined on the interval $\omega \in [0, 1]$. Here, i is the imaginary unit $i^2 = -1$. We use $\exp(x)$ to denote e^x.

---

[1] For simplicity's sake. An arbitrary interval will work.

Correspondingly, the inverse discrete Fourier transform

$$\mathcal{F}^{-1}\{F(\omega)\}$$

is

$$f(x_k) = \sum_{j=0}^{N-1} F(\omega_j) \exp\left(-2\pi i \frac{jk}{N}\right), \quad k = 0, \ldots, N-1. \tag{C.2}$$

FFT is just a very efficient method for computing both these formulas C.1, C.2. A brute-force calculation of these formulas requires of order N^2 "standard operations", each of them a single multiplication plus a single addition or subtraction. If N is even, we may write

$$F(\omega_j) = \frac{1}{N}\left[\sum_{k=0}^{\frac{N}{2}-1} f(x_k) \exp\left(-2\pi i \frac{jk}{N}\right) + \sum_{k=\frac{N}{2}}^{N-1} f(x_k) \exp\left(-2\pi i \frac{jk}{N}\right)\right] =$$

$$= \frac{1}{N}\left[\sum_{k=0}^{\frac{N}{2}-1} f(x_k) \exp\left(-2\pi i \frac{jk}{N}\right) + \exp\left(-2\pi i \frac{j\frac{N}{2}}{N}\right) \sum_{k'=0}^{\frac{N}{2}-1} f\left(x_{k'+\frac{N}{2}}\right) \exp\left(-2\pi i \frac{jk'}{N}\right)\right] =$$

$$= \frac{1}{N}\left[\sum_{k=0}^{\frac{N}{2}-1} f(x_k) \exp\left(-2\pi i \frac{jk}{N}\right) + \exp(-\pi i j) \sum_{k=0}^{\frac{N}{2}-1} f\left(x_{k+\frac{N}{2}}\right) \exp\left(-2\pi i \frac{jk}{N}\right)\right] =$$

$$= \frac{1}{N}\sum_{k=0}^{\frac{N}{2}-1}\left[f(x_k) \overset{j}{\pm} f\left(x_{k+\frac{N}{2}}\right)\right] \exp\left(-2\pi i \frac{jk}{N}\right), \qquad \left(\begin{array}{l} + \text{ if } j \text{ even} \\ - \text{ if } j \text{ odd} \end{array}\right) \tag{C.3}$$

the computation of which sum requires only $N \cdot \frac{N}{2}$ multiplications and $\frac{N}{2} + \frac{N}{2}$ additions and subtractions. Here we used Euler's identity $\exp(-\pi i) = -1$, i.e., $e^{-\pi i j} = (e^{-\pi i})^j = (-1)^j$, either $+1$ or -1.

Note that the expression in square brackets, for each k value $k = 0, 1, \ldots, \frac{1}{N} - 1$, is either a summation, for even values of j, or a subtraction, for odd values of j. In total, $2 \cdot \frac{N}{2} = N$ different values are computed, $\frac{N}{2}$ sums and $\frac{N}{2}$ differences. After that, there are $N \cdot \frac{N}{2} = \frac{N^2}{2}$ multiplications. We assume that the exp expressions are pre-calculated into a lookup table. Altogether some $\frac{N^2}{2}$ standard operations are needed, half the original number.

Equation C.3 is itself recognised as a Fourier series, but the number of support points is instead of N only $\frac{N}{2}$. If also $\frac{N}{2}$ is even, we may repeat the above

trick, resulting in an expression requiring only of order $\frac{N^2}{4}$ operations. Lather, rinse, repeat, and the number of operations becomes $\frac{N^2}{8}$, $\frac{N^2}{16}$, $\frac{N^2}{32}$, etc... A more precise analysis shows that, if N is a power of 2, then the whole discrete Fourier transform may be computed in order $N \times 2 \log N$ operations!

In the literature, smart algorithms are found implementing the above method, e.g., fftw ("Fastest Fourier Transform in the West", (FFTW Home Page)).

Helmert condensation

D

D.1   The interior potential of the topography

In order to derive the equation for Helmert condensation, we first derive the equation for the interior potential of the topography, i.e., the masses between sea level and the terrain surface:

$$T_t(r,\phi,\lambda) = G \iiint_{\text{top}} \frac{\rho(r',\phi',\lambda')}{\ell(r,r',\psi)} \, d\mathcal{V},$$

where ψ is the angular distance between the evaluation point (r,ϕ,λ) and the observation point (r',ϕ',λ'). The spatial distance ℓ between those points again is written using the *interior expansion* (equation 8.5):

$$\frac{1}{\ell} = \frac{1}{r} \sum_{n=0}^{\infty} \left(\frac{r}{r'}\right)^{n+1} P_n(\cos\psi).$$

This expansion converges uniformly[1] with respect to ψ if $r < r'$.

---

[1]Uniform convergence means that, given r, r', for every $\epsilon > 0$ there is an $N_{\min}$ for which

$$\left| \frac{1}{\ell} - \frac{1}{r} \sum_{n=0}^{N} \left(\frac{r}{r'}\right)^{n+1} P_n(\cos\psi) \right| < \epsilon$$

for all $N > N_{\min}$, and *for all values of* ψ. This is a stronger property than mere convergence.

Substitute:

$$
T_{\mathrm{t}}^{\mathrm{int}}(r,\phi,\lambda) = G\rho \iiint_{\mathrm{top}} \frac{1}{r} \sum_{n=0}^{\infty} \left(\frac{r}{r'}\right)^{n+1} P_n(\cos\psi)\, d\mathcal{V} =
$$

$$
= G\rho \iint_\sigma \left(\int_R^{R+H(\phi',\lambda')} \frac{1}{r} \sum_{n=0}^{\infty} \left(\frac{r}{r'}\right)^{n+1} (r')^2\, dr' \right) P_n(\cos\psi)\, d\sigma =
$$

$$
= G\rho \iint_\sigma \left(\sum_{\substack{n=0 \\ n\neq 2}}^{\infty} r^n \left(-\frac{(r')^{-(n-2)}}{n-2} \right) + r^2 \ln r' \right) \Bigg|_{r'=R}^{R+H} P_n(\cos\psi)\, d\sigma =
$$

$$
= G\rho \iint_\sigma \sum_{\substack{n=0 \\ n\neq 2}}^{\infty} \frac{r^n}{n-2} \left(\begin{array}{c} R^{-(n-2)} - (R+H)^{-(n-2)} + \\ + r^2 \ln\frac{R+H}{R} \end{array} \right) P_n(\cos\psi)\, d\sigma.
$$

Here, we shall expand the following expression into a Taylor series:

$$
(R+H)^{-(n-2)} = R^{-(n-2)} \left(\begin{array}{c} 1 - (n-2)\frac{H}{R} + \frac{(n-2)(n-1)}{2}\frac{H^2}{R^2} - \\ -\frac{(n-2)(n-1)n}{2\cdot3}\frac{H^3}{R^3} + \ldots \end{array} \right).
$$

Also the special case $n=2$,

$$
r^2 \ln\frac{R+H}{R} = r^2 \left(\frac{H}{R} - \frac{1}{2}\frac{H^2}{R^2} + \frac{1}{3}\frac{H^3}{R^3} - \frac{1}{4}\frac{H^4}{R^4} + \ldots \right) =
$$

$$
= \frac{r^n}{R^{n-2}} \left(\frac{H}{R} - \frac{n-1}{2}\frac{H^2}{R^2} + \frac{(n-1)n}{2\cdot3}\frac{H^3}{R^4} - \frac{(n-1)n(n+1)}{2\cdot3\cdot4}\frac{H^4}{R^4} + \ldots \right),
$$

is cleanly included into the following expression obtained by substitution:

$$
T_{\mathrm{t}}^{\mathrm{int}}(r,\phi,\lambda) = G\rho \iint_\sigma \sum_{n=0}^{\infty} \frac{r^n}{R^{n-2}} \left(\frac{H}{R} - \frac{n-1}{2}\frac{H^2}{R^2} + \frac{(n-1)n}{6}\frac{H^3}{R^3} - \ldots \right) P_n(\cos\psi)\, d\sigma.
$$

D.2 The exterior potential of the topography

In the same way we may derive the exterior potential of the topography. Now we use for the expansion of the inverse distance (equation 8.5):

$$
\frac{1}{\ell} = \sum_{n=0}^{\infty} \frac{1}{r'} \left(\frac{r'}{r}\right)^{n+1} P_n(\cos\psi) = \sum_{n=0}^{\infty} \frac{1}{r} \left(\frac{r'}{r}\right)^n P_n(\cos\psi),
$$

which converges uniformly if $r > r'$.

Substitution yields

$$
T_{\mathrm{t}}^{\mathrm{ext}} = G\rho \iiint_{\mathrm{top}} \sum_{n=0}^{\infty} \frac{1}{r} \left(\frac{r'}{r}\right)^n P_n(\cos\psi)\, d\mathcal{V} =
$$

$$
= G\rho \iint_\sigma \left(\int_R^{R+H(\phi',\lambda')} \sum_{n=0}^{\infty} \frac{1}{r} \left(\frac{r'}{r}\right)^n (r')^2\, dr' \right) P_n(\cos\psi)\, d\sigma =
$$

$$
= G\rho \iint_\sigma \left(\sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \frac{1}{n+3} (r')^{n+3} \right) \Bigg|_{r'=R}^{R+H} P_n(\cos\psi)\, d\sigma =
$$

$$
= G\rho \iint_\sigma \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \frac{1}{n+3} \left((R+H)^{n+3} - R^{n+3} \right) P_n(\cos\psi)\, d\sigma.
$$

Here again we use the Taylor expansion:

$$
(R+H)^{n+3} = R^{n+3} \left(\begin{array}{c} 1 + (n+3)\frac{H}{R} + \frac{(n+3)(n+2)}{2}\frac{H^2}{R^2} + \\ + \frac{(n+3)(n+2)(n+1)}{2\cdot3}\frac{H^3}{R^3} + \ldots \end{array} \right).
$$

Substitution yields

$$
T_{\mathrm{t}}^{\mathrm{ext}} = G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\begin{array}{c} \frac{H}{R} + \frac{n+2}{2}\frac{H^2}{R^2} + \\ + \frac{(n+2)(n+1)}{6}\frac{H^3}{R^3} + \ldots \end{array} \right) P_n(\cos\psi)\, d\sigma.
$$

(D.1)

This is thus the *exterior potential* of the topography, or, inside the topographic masses, the *harmonic downward continuation* of the exterior potential, assuming that this is mathematically possible (in the case of mountainous topography, generally not) and doesn't diverge.

D.3 The exterior potential of the condensation layer

This is derived by specializing equation D.1 to the case $H \to 0$, but nevertheless $\rho \to \infty$, so that $\kappa = \rho H$ is finite. in this limit, all terms containing H^2, H^3

etc. are going to zero. The result is then

$$T_c^{ext} = G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \frac{H}{R} P_n(\cos\psi) \, d\sigma =$$

$$= GR \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \kappa P_n(\cos\psi) \, d\sigma.$$

Earlier on we had a more precise formula 6.4 for κ on the surface of a spherical Earth:

$$\kappa = \rho H \left(1 + \frac{H}{R}\right);$$

by substituting this in the previous, we obtain

$$T_c^{ext} = G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\frac{H}{R} + \frac{H^2}{R^2}\right) P_n(\cos\psi) \, d\sigma. \tag{D.2}$$

D.4   The total potential of Helmert condensation

This is obtained by subtracting equations D.1 and D.2 from each other. The result (which thus applies in the *exterior* space) is

$$T_{Helmert}^{ext} = T_t^{ext} - T_c^{ext} =$$

$$= G\rho R^2 \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\begin{array}{c} \left(\frac{n+2}{2} - 1\right)\frac{H^2}{R^2} + \\ + \frac{(n+2)(n+1)}{6}\frac{H^3}{R^3} + \dots \end{array} \right) P_n(\cos\psi) \, d\sigma =$$

$$= G\rho \iint_\sigma \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\frac{n}{2}H^2 + \frac{(n+2)(n+1)}{6}\frac{H^3}{R} + \dots\right) P_n(\cos\psi) \, d\sigma.$$

Often we define the *degree constituents* of powers of height H (compare the degree constituent equation 3.7), as follows:

$$H_n^v = \frac{2n+1}{4\pi} \iint_\sigma H^v P_n(\cos\psi) \, d\sigma, \tag{D.3}$$

after which we may expand

$$H^v = \sum_{n=0}^{\infty} H_n^v.$$

Then

$$T_{Helmert}^{ext} = 4\pi G\rho \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \left(\frac{n}{2(2n+1)}H_n^2 + \frac{(n+2)(n+1)}{6(2n+1)}\frac{H_n^3}{R} + \dots\right). \tag{D.4}$$

If the topography is *constant*, then all terms for which $n \neq 0$ vanish; in the above expression, also the first term vanishes, and the second and further terms are very small. So, in practice[2]

$$T_{Helmert}^{ext} = \frac{4}{3}\pi\overline{H}^3 \cdot \frac{G\rho}{r} + \dots \approx 0$$

as was to be expected. The exterior field remains in this special case almost unchanged as a result of Helmert condensation.

D.4.1   The gravity effect of Helmert condensation

Let us calculate *gravity anomalies* from the Helmert potential, but only using the first term of equation D.4:

$$\Delta g_{Helmert} = \frac{\partial T}{\partial r} + \frac{2}{r}T =$$

$$= 2\pi G\rho \sum_{n=0}^{\infty} \frac{n}{2n+1} \left(\frac{-(n+1)}{r} + \frac{2}{r}\right) \left(\frac{R}{r}\right)^{n+1} H_n^2 =$$

$$= -2\pi G\rho \frac{1}{r} \sum_{n=0}^{\infty} \frac{n(n-1)}{2n+1} \left(\frac{R}{r}\right)^{n+1} H_n^2.$$

Now, also $n = 1$ gives a zero result, expected as gravity anomalies do not contain any components of degree number 1.

Note in this equation also a dependence upon n: The gravity effect of Helmert condensation is dominated by short wavelengths or the very local features of the topography. The appearance of the square of height in this equation is again related to the *terrain correction*, in which also the square of the terrain

---

[2]As a curiosity, this result can be interpreted as the potential of a sphere of crustal matter with radius $\overline{H}$ (the average height of the topography) located at the geocentre. Even for a topographic mean height of 10 km the effect on the geoid would only be 12 mm (exercise!).

height figures. When we are summing squares, leaving out terms will always cause a *systematic error*: also very short wavelengths, i.e., high values of n, must be taken along in the summation.

D.4.2 The interior potential of Helmert condensation

This quantity is evaluated *on the level of the geoid*. It represents the *indirect effect* of Helmert condensation, i.e., the shift of the geoid surface caused by the mass shift in space.

$$T_{\text{Helmert}}^{\text{int}} = T_t^{\text{int}} - T_c^{\text{ext}} =$$
$$= G\rho R^2 \iint_\sigma \sum_{n=0}^\infty \left(\frac{H}{R} - \frac{n-1}{2} \frac{H^2}{R^2} + \frac{(n-1)\,n}{6} \frac{H^3}{R^3} - \dots \right) P_n(\cos\psi)\,d\sigma -$$
$$- G\rho R^2 \iint_\sigma \sum_{n=0}^\infty \left(\frac{H}{R} + \frac{H^2}{R^2} \right) P_n(\cos\psi)\,d\sigma =$$
$$= G\rho \iint_\sigma \sum_{n=0}^\infty \left(-\frac{n+1}{2} H^2 + \frac{(n-1)\,n}{6} \frac{H^3}{R} - \dots \right) P_n(\cos\psi)\,d\sigma.$$

Using again the definition of the degree constituents of the powers of height H, equation D.3:

$$H_n^\nu = \frac{2n+1}{4\pi} \iint_\sigma H^\nu P_n(\cos\psi)\,d\sigma,$$

we obtain

$$T_{\text{Helmert}}^{\text{int}} = 4\pi G\rho \sum_{n=0}^\infty \left(-\frac{n+1}{2\,(2n+1)} H_n^2 + \frac{(n-1)\,n}{6\,(2n+1)} \frac{H_n^3}{R} - \dots \right),$$

from which one obtains the *indirect effect* of Helmert condensation:

$$\delta N_{\text{HC}} = -\frac{T_{\text{Helmert}}^{\text{int}}}{\overline{\gamma}} = -\frac{4\pi G\rho}{\overline{\gamma}} \sum_{n=0}^\infty \left(-\frac{n+1}{2\,(2n+1)} H_n^2 + \frac{(n-1)\,n}{6\,(2n+1)} \frac{H_n^3}{R} - \dots \right) =$$
$$= \frac{4\pi G\rho}{\overline{\gamma}} \left(\frac{1}{2} \sum_{n=0}^\infty \frac{n+1}{2n+1} H_n^2 - \frac{1}{6R} \sum_{n=0}^\infty \frac{(n-1)\,n}{2n+1} H_n^3 - \dots \right). \tag{D.5}$$

The term $n = 0$ yields the indirect effect of a constant terrain $H = \overline{H} = H_0$: using only the first term inside the brackets yields

$$\delta N_{\text{HC,Const}} = \frac{2\pi G\rho}{\overline{\gamma}} \overline{H}^2,$$

which can *not* be neglected.

D.5 The dipole method

As a sanity test, we may describe the effect of Helmert condensation in first approximation as a *dipole density layer field* μ. The topographic mass, density $\kappa = H\rho$, moves downward by on average $\frac{1}{2}H$. The effect would be the same if mean sea level[3] was covered by a double mass density layer

$$\mu = \frac{1}{2} H^2 \rho. \tag{D.6}$$

The potential of this layer is, in spherical approximation,

$$T = G \iint_S \mu \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) dS \approx GR^2 \iint_\sigma \mu \frac{\partial}{\partial n} \left(\frac{1}{\ell} \right) d\sigma.$$

Written more explicitly:

$$T_P = GR^2 \iint_\sigma \mu_Q \frac{\partial}{\partial r_Q} \left(\frac{1}{\ell_{PQ}} \right) d\sigma_Q.$$

We use the expansion into Legendre polynomials, equation 8.5:

$$\frac{1}{\ell_{PQ}} = \frac{1}{r_Q} \sum_{n=0}^\infty \left(\frac{r_Q}{r_P} \right)^{n+1} P_n(\cos\psi_{PQ}),$$

differentiate with respect to r_Q, and substitute:

$$T_P = GR^2 \iint_\sigma \frac{1}{r_Q^2} \mu_Q \sum_{n=0}^\infty n \left(\frac{r_Q}{r_P} \right)^{n+1} P_n(\cos\psi_{PQ})\,d\sigma_Q.$$

By substituting into this the equation D.6 for the μ_Q double mass density layer, we obtain, by taking the limit $r_P, r_Q \downarrow R$:

$$T = \frac{1}{4\pi} \sum_{n=0}^\infty n \iint_\sigma (2\pi G H \rho)\, H P_n(\cos\psi)\,d\sigma =$$
$$= \frac{1}{4\pi} \sum_{n=0}^\infty n \iint A_{\text{B}} H P_n(\cos\psi)\,d\sigma.$$

---

[3]In fact, a better place for this layer would be level $\frac{1}{4}H$. This is one of the approximations made here.

Helmert condensation

Here, we have left the designations P, Q again off as no longer needed for clarity.

The symbol A_B denotes the attraction of a Bouguer plate of thickness H and density ρ.

Let us develop the quantity $[A_B H]$ into a spherical-harmonic expansion (Heiskanen and Moritz, 1967 equation 1-71). then, each degree constituent is (equation 3.7):

$$[A_B H]_n = \frac{2n+1}{4\pi} \iint_\sigma [A_B H]\, P_n(\cos\psi)\, d\sigma,$$

in which case (note that the term $n = 0$ vanishes):

$$T = \sum_{n=1}^{\infty} \frac{n}{2n+1} [A_B H]_n \approx \frac{1}{2} [A_B H],$$

at least for the higher n values, i.e., regionally if not globally.

Thus is obtained again the indirect effect of Helmert condensation, in geoid computation by means of this method the shift in geoid surface caused by the condensation, which must be accounted for with opposite algebraic sign. In other words, looked upon as a *Remove-Restore* method, its "Restore" step:

$$\delta N_{HC} = \frac{T}{\overline{\gamma}} \approx \frac{1}{2} \frac{A_B H}{\overline{\gamma}} = \frac{\pi G \rho H^2}{\overline{\gamma}}.$$

For comparison, the more precise expansion D.5 yields in approximation for larger n values

$$\delta N_{HC} \approx \frac{4\pi G\rho}{\overline{\gamma}} \cdot \frac{1}{2} \sum_{n=0}^{\infty} \frac{n+1}{2n+1} H_n^2 \approx \frac{\pi G \rho H^2}{\overline{\gamma}},$$

essentially the same result.

The Laplace equation in spherical co-ordinates

E

E.1 Derivation

Consider a small volume element with sizes in co-ordinate directions of Δr, $\Delta\phi$, $\Delta\lambda$. Look at the difference in flux of vector field $\mathbf{a} \stackrel{\text{def}}{=} \nabla V$ between what comes in and what goes out through opposite faces.

We do the analogue of what was shown in subsection 1.12.4, using a body with surfaces aligned along co-ordinate lines, allowing the size of the body to go to zero in the limit, and exploiting the Gauss integral theorem 1.18. The quantity $\operatorname{div}\mathbf{a} = \Delta V$ is a *volume density*, and its average value multiplied by the volume of a body must equal the total flux through the surface of the body.

Part of the difference in flux f between opposing faces is due to a change in the normal component of $\mathbf{a}$ between the faces, part is due to a difference in face surface area ω:

$$f^+ - f^- \approx \omega\left(a^+ - a^-\right) + a\left(\omega^+ - \omega^-\right).$$

See figure E.1.

1. In the radial direction, the surface areas of opposing faces — "inner – outer" — are

$$\omega_r^+ = (r + \Delta r)^2 \cos\phi\,\Delta\phi\,\Delta\lambda,$$
$$\omega_r^- = r^2 \cos\phi\,\Delta\phi\,\Delta\lambda,$$

The Laplace equation in spherical co-ordinates
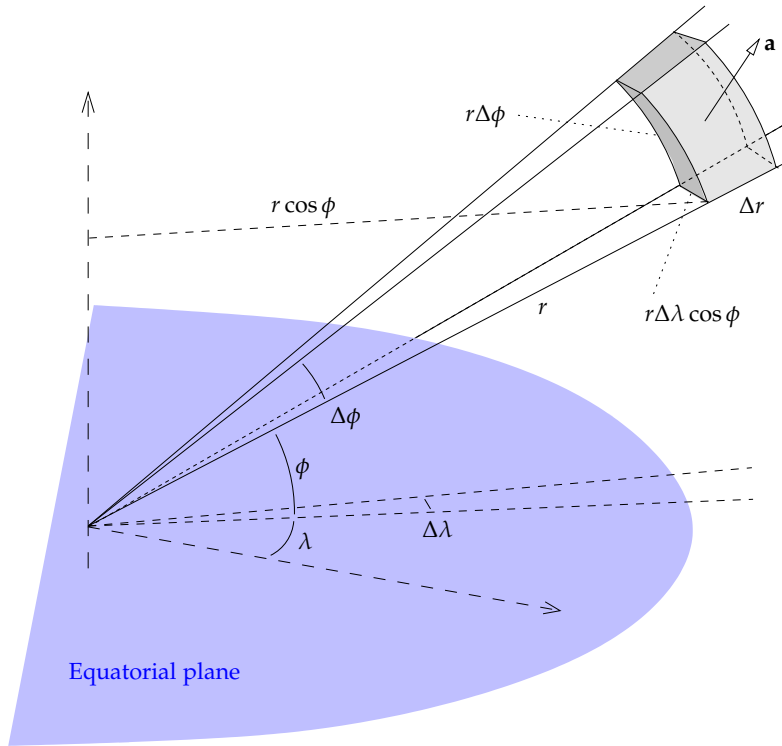
Figure E.1. Gauss integral theorem applied to a co-ordinate conformal volume element.

difference

$$\omega_r^+ - \omega_r^- \approx 2r\Delta r \cos\phi \Delta\phi \Delta\lambda.$$

This is be multiplied by $a_r = \frac{\partial V}{\partial r}$, and divided by the volume of the body $r^2 \cos\phi \Delta r \Delta\phi \Delta\lambda$, yielding for the contribution to the Laplacian

$$\Delta_{r'} V = \frac{2}{r}\frac{\partial V}{\partial r}.$$

This in addition to the "regular" contribution

$$\Delta_r V = \frac{a_r^+ - a_r^-}{\Delta r} \approx \frac{\partial^2 V}{\partial r^2}.$$

2. Longitudinal direction, λ, "West–East": no change in surface area $\omega_\lambda = r\Delta r\Delta\phi$ because of rotational symmetry. We only have

$$\Delta_\lambda V = \frac{a_\lambda^+ - a_\lambda^-}{r\cos\phi\Delta\lambda} = \frac{1}{r\cos\phi\Delta\lambda}\left(\frac{\partial V^+}{\partial r\lambda\cos\phi} - \frac{\partial V^-}{\partial r\lambda\cos\phi}\right) \approx$$

$$\approx \frac{1}{r^2\cos^2\phi}\frac{\partial^2 V}{\partial\lambda^2}.$$

3. Latitudinal direction, ϕ, "South–North":

$$\omega_\phi^+ = r\cos(\phi+\Delta\phi)\,\Delta r\Delta\lambda,$$
$$\omega_\phi^- = r\cos\phi\Delta r\Delta\lambda,$$

difference

$$\omega_\phi^+ - \omega_\phi^- \approx -r\sin\phi\Delta\phi\Delta r\Delta\lambda.$$

Multiply by $a_\phi = \frac{\partial V}{\partial r\phi}$ and divide by body volume $r^2\cos\phi\Delta r\Delta\phi\Delta\lambda$, yielding

$$\Delta_{\phi'} V = -\frac{\tan\phi}{r^2}\frac{\partial V}{\partial\phi}.$$

This of course in addition to the regular contribution

$$\Delta_\phi V = \frac{a_\phi^+ - a_\phi^-}{r\Delta\phi} = \frac{1}{r\Delta\phi}\left(\frac{\partial V^+}{\partial r\phi} - \frac{\partial V^-}{\partial r\phi}\right) \approx \frac{1}{r^2}\frac{\partial^2 V}{\partial\phi^2}.$$

All of this gives us the end result

$$\Delta V = \Delta_r V + \Delta_\lambda V + \Delta_\phi V + \Delta_{r'} V + \Delta_{\phi'} V =$$
$$= \frac{\partial^2 V}{\partial r^2} + \frac{1}{r^2\cos^2\phi}\frac{\partial^2 V}{\partial\lambda^2} + \frac{1}{r^2}\frac{\partial^2 V}{\partial\phi^2} + \frac{2}{r}\frac{\partial V}{\partial r} - \frac{\tan\phi}{r^2}\frac{\partial V}{\partial\phi},$$

equivalent to equation 2.8.

E.2   Solution

E.2.1   Separating the radial dependency

Let us attempt separation of variables as follows:

$$V(r,\phi,\lambda) = R(r)\,Y(\phi,\lambda).$$

The Laplace equation in spherical co-ordinates

Substitution into equation 2.8 and multiplication by $\frac{r^2}{RY}$ yields

$$\frac{1}{R}\left(r^2\frac{\partial^2 R}{\partial r^2} + 2r\frac{\partial R}{\partial r}\right) = -\frac{1}{Y}\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2} + \frac{\partial^2 Y}{\partial\phi^2} - \tan\phi\frac{\partial Y}{\partial\phi}\right).$$

This must again apply for all values r, ϕ, λ and thus can only be a constant, p. This yields two equations

$$\left(r^2\frac{\partial^2 R}{\partial r^2} + 2r\frac{\partial R}{\partial r}\right) - pR = 0,$$

$$\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2} + \frac{\partial^2 Y}{\partial\phi^2} - \tan\phi\frac{\partial Y}{\partial\phi}\right) + pY = 0.$$

For the first equation we try a power law,

$$R = r^q,$$

yielding

$$q(q-1))r^q + 2qr^q - pr^q = 0$$

with the solution

$$p = q(q+1).$$

Solving the second equation for $Y(\phi, \lambda)$,

$$\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2} + \frac{\partial^2 Y}{\partial\phi^2} - \tan\phi\frac{\partial Y}{\partial\phi}\right) + q(q+1)Y = 0, \qquad \text{(E.1)}$$

is trickier. It turns out that q must be an integer. One finds, for $n \in \mathbb{N}_0$, that there are positive solutions[1] $q = n$ and negative solutions $q = -(n+1)$, with $n = 0, 1, 2, \ldots$ With this, the full set of special solutions is

$$V_{n,1} = r^n Y_n(\phi, \lambda), \quad V_{n,2} = \frac{Y_n(\phi, \lambda)}{r^{n+1}}, \quad n \in \mathbb{N}_0.$$

i.e., equations 2.9.

---

[1]The above derivation logic doesn't work for the exponent values $q = 1$ or $q = 0$, but the result holds also for these values.

E.2.2   Solving for surface harmonics

Note that both solutions q, the positive and the negative one, produce on substitution into equation E.1 the same equation for n:

$$\left(\frac{1}{\cos^2\phi}\frac{\partial^2 Y}{\partial\lambda^2} + \frac{\partial^2 Y}{\partial\phi^2} - \tan\phi\frac{\partial Y}{\partial\phi}\right) + n(n+1)Y = 0.$$

We attempt again separation of variables:

$$Y(\phi, \lambda) = F(\phi)L(\lambda).$$

Substitution and multiplication by $\frac{\cos^2\phi}{FL}$ yields

$$\frac{\cos^2\phi}{F}\left(\frac{\partial^2 F}{\partial\phi^2} - \tan\phi\frac{\partial F}{\partial\phi} + n(n+1)F\right) = -\frac{1}{L}\frac{\partial^2 L}{\partial\lambda^2}.$$

Both sides must be again equal to the same constant, which we shall assume positive and call m^2:

$$\frac{\partial^2 F}{\partial\phi^2} - \tan\phi\frac{\partial F}{\partial\phi} + \left(n(n+1) - \frac{m^2}{\cos^2\phi}\right)F = 0,$$

$$\frac{\partial^2 L}{\partial\lambda^2} + m^2 L = 0.$$

The first equation is known as Legendre's equation. Its solutions are the Legendre functions $P_{nm}(\sin\phi)$, with the integer $m = 0, 1, 2, \ldots, n$. The second is the classical *harmonic oscillator*, with solutions[2]

$$L_{m,1}(\lambda) = \cos m\lambda, \quad L_{m,2} = \sin m\lambda .$$

With this, we find for the surface spherical harmonics the linear combinations

$$Y_n(\phi, \lambda) = \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda).$$

The general solution is now formed as follows:

$$V_1(r, \phi, \lambda) = \sum_{n=0}^{\infty} r^n \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda),$$

$$V_2(r, \phi, \lambda) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=0}^{n} P_{nm}(\sin\phi)\,(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda).$$

---

[2]This also explains why m must be an integer.

Here, a_{nm} and b_{nm} are the spherical-harmonic coefficients describing the linear combination of special solutions. For describing the Earth's gravitational field, only the second solution is physically realistic, going to zero at infinity $r \to \infty$.

Bibliography

Altimetry, Retracking. URL http://www.altimetry.info/radar-altimetry-tutorial/data-flow/data-processing/retracking/. Accessed May 11, 2019. 340

Ole Balthasar Andersen, Per Knudsen, and Philippa Berry. The DNSC08GRA global marine gravity field from double retracked satellite altimetry. *Journal of Geodesy*, 84(3), 2010. URL https://doi.org/10.1007/s00190-009-0355-9. Accessed May 13, 2019. 229

Autograv CG5. URL https://commons.wikimedia.org/wiki/File:Autograv_CG5_P1150838.JPG. © 2011 David Monniaux (GFDL). Accessed May 13, 2019. 277

Georges Balmino, N. Vales, S. Bonvalot, and A. Briais. Spherical harmonic modeling to ultra-high degree of Bouguer and isostatic anomalies. *Journal of Geodesy*, 2012. URL https://www.doi.org/10.1007/s00190-011-0533-4. Accessed May 11, 2019. 59, 132

Tulu Besha Bedada. *Absolute geopotential height system for Ethiopia*. PhD thesis, University of Edinburgh, 2010. URL https://www.era.lib.ed.ac.uk/bitstream/1842/4726/3/Bedada2010_small.pdf. Accessed May 11, 2019. 229

N. Benitez, T. Broadhurst, H. Ford, M. Clampin(STScI), G. Hartig, G. Illingworth, et al. Hubble Looks Through Cosmic Zoom Lens, 2003. URL https://www.spacetelescope.org/images/opo0301a/. © 2003 ESA/Hubble (CC BY 4.0). Accessed May 15, 2019. 3

BGI, EGM2008. EGM2008 anomaly maps visualization. URL
http://bgi.omp.obs-mip.fr/data-products/Toolbox/EGM2008-anomaly-
maps-visualization. Accessed May 10, 2019. 118, 127

BGI, WGM2012. WGM2012 maps visualization/extraction. URL
http://bgi.omp.obs-mip.fr/data-products/Toolbox/WGM2012-maps-
vizualisation-extraction. Accessed May 10, 2019. 144

Mirjam Bilker-Koivula. Development of the Finnish Height Conversion
Surface FIN2005N00. *Nordic Journal of Surveying and Real Estate Research*, 7
(1):76–88, 2010. URL
http://ojs.tsv.fi/index.php/njs/article/download/3663/3432. Accessed
May 11, 2019. 367

Mirjam Bilker-Koivula and Matti Ollikainen. Suomen geoidimallit ja niiden
käyttäminen korkeuden muunnoksissa. Tiedote 29, Geodeettinen laitos,
2009. URL https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/
files/fgi/GLtiedote29.pdf. Accessed May 11, 2019. 227

Gian Paolo Bottoni and Riccardo Barzaghi. Fast collocation. *Bulletin
géodésique*, 67:119–126, 1993. URL
https://www.doi.org/10.1007/BF01371375. Accessed May 13, 2019. 267

V. V. Brovar, M. I. Yurkina, M. Heifets, M. S. Molodensky, and H. Moritz. M. S.
Molodensky In Memoriam. On-line PDF, Mitteilungen der geodätischen
Institute der Technischen Universität Graz Folge 88, 2000. URL
http://www.helmut-moritz.at/SciencePage/Molodensky.pdf. H. Moritz
and M. I. Yurkina, eds. Accessed May 19, 2019. 160

John M. Brozena. The Greenland Aerogeophysics Project: Airborne gravity,
topographic and magnetic mapping of an entire continent. volume 110 of
International Association of Geodesy Symposia, pages 203–214, Vienna,
Austria, August 20, 1992. Springer, New York, NY. URL
https://doi.org/10.1007/978-1-4613-9255-2_19. Accessed May 13, 2019.
295

John M. Brozena and Mary F. Peters. State-of-the-art airborne gravimetry.
volume 113 of *International Association of Geodesy Symposia*, pages 187–197,
Graz, Austria, 1994. Springer Verlag. URL

https://doi.org/10.1007/978-3-642-79721-7_20. Accessed May 13, 2019.
295

John M. Brozena, Mary F. Peters, and R. Salman. Arctic airborne gravity
measurements program. In Segawa et al. (1996), pages 131–146. URL
https://doi.org/10.1007%2F978-3-662-03482-8_20. Accessed May 13, 2019.
295

Henry Cavendish. Experiments to determine the Density of the Earth.
Philosophical Transactions of the Royal Society, 88, 1798. URL
https://doi.org/10.1098/rstl.1798.0022. Accessed May 13, 2019. 4

CHAMP Mission. URL https://www.gfz-potsdam.de/champ/. Accessed
April 12, 2019. 343

Bernard J. Coakley, Steve C. Kenyon, and René Forsberg. Updating the
Arctic Gravity Project grid with new airborne and Extended Continental
Shelf data. *AGU Fall Meeting Abstracts*, page C3, December 2013. URL
https://ui.adsabs.harvard.edu/abs/2013AGUFM.G13C..03C/abstract. 295

Giancarlo d'Agostino, Sergio Desogus, Alessandro Germak, Claudio Origlia,
Danilo Quagliotti, Giovanna Berrino, Gennaro Corrado, Vincenzo d'Errico,
and Giuseppe Ricciardi. The new IMGC-02 transportable absolute
gravimeter: measurement apparatus and applications in geophysics and
volcanology. *Annals of Geophysics*, 51(1):39–49, 2008. URL
https://doi.org/10.4401/ag-3038. Accessed May 13, 2019. 287

M. de Angelis, A. Bertoldi, L. Cacciapuoti, A. Giorgini, G. Lamporesi,
M. Prevedelli, G. Saccorotti, F. Sorrentino, and G. M. Tino. Precision
gravimetry with atomic sensors. *Measurement Science and Technology*, 20(2):
022001, 2009. URL http://dx.doi.org/10.1088/0957-0233/20/2/022001.
Accessed May 11, 2019. 287

Heiner Denker. Evaluation and Improvement of the EGG97 Quasigeoid
Model for Europe by GPS and Leveling Data. In Vermeer and Ádám
(1998), pages 53–61. 366

Eiffel Tower, 72 names. List of the 72 names on the Eiffel Tower. URL https:
//en.wikipedia.org/wiki/List_of_the_72_names_on_the_Eiffel_Tower.
Accessed April 7, 2019. 15, 16, 32, 43, 53, 84

Bibliography

Simon Ekholm, René Forsberg, and John M. Brozena. Accuracy of satellite altimeter observations over the Greenland ice sheet. *Journal of Geophysical Research*, 100(C2):2687–2696, 1995. URL https://doi.org/10.1029/94JC03042. Accessed May 13, 2019. 295

Martin Ekman. Postglacial Rebound and Sea Level Phenomena with Special Reference to Fennoscandia and the Baltic Sea. In Kakkuri (1993), pages 7–70. 308, 358

European geoid calculations. Leibniz Universität Hannover, Institute of Geodesy. URL https://www.ife.uni-hannover.de/en/research/main-research-focus/regional-gravity-field-and-geoid-modelling/european-geoid-calculations/. Accessed May 11, 2019. 366

Loránd Eötvös. *Three Fundamental Papers of Loránd Eötvös*. Loránd Eötvös Geophysical Institute of Hungary, 1998. ISBN 963-7135-02-2. Editor Zoltán Szabó. 299 pages. 297

Graham Farmelo. *The strangest man*. Basic Books, reprint edition, 2011. ISBN 978-0-4650-2210-6. 25

Will E. Featherstone. Software for computing five existing types of deterministically modified integration kernel for gravimetric geoid determination. *Computers and Geosciences*, 29:183–193, 2003. URL https://doi.org/10.1016/S0098-3004(02)00074-2. Accessed May 11, 2019. 207

FFTW Home Page. URL https://www.fftw.org. Accessed May 15, 2019. 401

René Forsberg and Jānis Kaminskis. Geoid of the Nordic and Baltic region from gravimetry and satellite altimetry. In Segawa et al. (1996), pages 540–547. URL https://doi.org/10.1007/978-3-662-03482-8_72. Accessed May 13, 2019. 367

René Forsberg and Gabriel Strykowski. NKG Gravity Data Base and NKG Geoid. Technical University of Danmark, DTU Space, National Space Institute, 2010. URL http://www.nordicgeodeticcommission.com/wp-content/uploads/2014/10/8-WG_geoid_2010_March_presentation_ForsbergStrykowski.pdf. Accessed May 11, 2019. 367

Bibliography

René Forsberg and Carl Christian Tscherning. An overview manual for the GRAVSOFT Geodetic Gravity Field Modelling Programs, 2008. URL https://www.academia.edu/9206363/An_overview_manual_for_the_GRAVSOFT_Geodetic_Gravity_Field_Modelling_Programs. Accessed May 11, 2019. 227

René Forsberg and Martin Vermeer. A Generalized Strang van Hees Approach to Fast Geopotential Inversion. *Manuscripta geodaetica*, 17:302–314, 1992. 224

Emmanuel S. Garcia, David T. Sandwell, and Walter H. F. Smith. Retracking CryoSat-2, Envisat and Jason-1 radar altimetry waveforms for improved gravity field recovery. *Geophysical Journal International*, 2014. doi: 10.1093/gji/ggt469. URL http://topex.ucsd.edu/sandwell/publications/142.pdf. Accessed May 11, 2019. 229

A. Gatti, M. Reguzzoni, F. Migliaccio, and F. Sansò. Space-wise grids of gravity gradients from GOCE data at nominal satellite altitude. Presented at the 5th International GOCE User Workshop, November 25-28, 2014, UNESCO, Paris, France, November 25–28, 2014. URL https://www.researchgate.net/publication/275029640_SPACE-WISE_GRIDS_OF_GRAVITY_GRADIENTS_FROM_GOCE_DATA_AT_NOMINAL_SATELLITE_ALTITUDE. Accessed May 11, 2019. 264

Olav R. Godø, Annette Samuelsen, Gavin J. Macaulay, Ruben Patel, Solfrid Sætre Hjøllo, John Horne, Stein Kaartvedt, and Johnny A. Johannessen. Mesoscale Eddies Are Oases for Higher Trophic Marine Life. *PLoS One*, 7(1):e30161, 2012. URL https://doi.org/10.1371/journal.pone.0030161. Accessed May 11, 2019. 312

GRACE Follow-On Mission. URL https://gracefo.jpl.nasa.gov/mission/overview/. Accessed April 12, 2019. 345

GRACE Mission. URL https://grace.jpl.nasa.gov/. Accessed April 12, 2019. 344

GRACE Mission, hydrology. URL
https://commons.wikimedia.org/wiki/File:
Global_Gravity_Anomaly_Animation_over_LAND.gif. Accessed April 12,
2019. 345

George Green. An Essay on the Application of Mathematical Analysis to the
Theories of Electricity and Magnetism. 1828. URL
https://play.google.com/books/reader?id=GwYXAAAAYAAJ. Accessed
April 23, 2019. 27

Green's Windmill. Green's windmill and science centre. URL
https://www.greensmill.org.uk/. Accessed May 10, 2019. 27

GWR Instruments, Inc., *iGRAV® Gravity Sensors*. URL
http://www.gwrinstruments.com/igrav-gravity-sensors.html. Accessed
May 27, 2019. 291

Roger Haagmans, Erik De Min, and Martin van Gelderen. Fast Evaluation of
Convolution Integrals on the Sphere Using 1D FFT, and a Comparison
with Existing Methods for Stokes' Integral. *Manuscripta geodaetica*, 18:
227 – 241, 1993. 224

J. C. Harrison and M. Dickinson. Fourier transform methods in local gravity
modelling. *Bulletin Géodésique*, 63:149 – 166, 1989. URL
https://doi.org/10.1007/BF02519148. 231

Markku Heikkinen. Solving the Shape of the Earth by Using Digital Density
Models. Report 81:2, Finnish Geodetic Institute, Helsinki, 1981. 93, 94, 97,
292

Weikko A. Heiskanen and Helmut Moritz. *Physical Geodesy*. W. H. Freeman
and Company, San Francisco, London, 1967. 33, 50, 53, 63, 65, 72, 74, 75,
88, 89, 90, 96, 97, 98, 106, 110, 145, 147, 166, 168, 169, 173, 181, 182, 186, 189,
193, 196, 368, 397, 410

Willem Frederik Hermans. *Beyond Sleep*. Harry N. Abrams, reprint edition,
2007. ISBN 978-1-5856-7583-8. 137

C. Hirt and M. Kuhn. Band-limited topographic mass distribution generates
full-spectrum gravity field: Gravity forward modeling in the spectral and
spatial domains revisited. *Journal of Geophysical Research Solid Earth*, 119,

2014. URL https://doi.org/10.1002/2013JB010900. Accessed May 13, 2019.
132

Reino A. Hirvonen. *The continental undulations of the geoid*. PhD thesis,
Helsinki University of Technology, 1934. Finnish Geodetic Institute
Publication 19. 367

Bernhard Hofmann-Wellenhof and Helmut Moritz. *Physical geodesy*.
Springer-Verlag Wien GmbH, 2006. Second, revised edition. 368

Humboldt University Berlin. Humboldt university berlin, friedrich robert
helmert with a relative pendulum, 2017. URL
https://www.researchgate.net/publication/318994932_Friedrich_
Robert_Helmert_founder_of_modern_geodesy_on_the_occasion_of_the_
centenary_of_his_death/figures. © 2017 Humboldt University Berlin (CC
BY 3.0). Accessed May 19, 2019. 135

Erkki Hytönen. *Absolute gravity measurement with long wire pendulum*. PhD
thesis, University of Helsinki, 1972. Finnish Geodetic Institute Publication
75. 276

ISG, Geoid Schools. URL
http://www.isgeoid.polimi.it/Schools/schools.html. Accessed May 10,
2019. 227

Juhani Kakkuri, editor. *Geodesy and Geophysics, lecture notes, NKG Autumn
School 1992*, Finnish Geodetic Institute Publication 115, 1993. 420, 428, 429

Juhani Kakkuri. *Surveyor of the Globe – Story of the Life of V. A. Heiskanen*.
National Land Survey of Finland, 2008. URL
https://readymag.com/u95015526/508134. Accessed May 13, 2019. 137,
368

Steve C. Kenyon, René Forsberg, A. V. Olesen, and Simon A. Holmes. NGA's
use of aerogravity to advance the next generation of Earth Gravitational
Models. *AGU Fall Meeting Abstracts*, page A4, December 2012. URL
https://ui.adsabs.harvard.edu/abs/2012AGUFM.G12A..04K/abstract.
Accessed May 11, 2019. 295

M. G. Kogan, M. Diament, A. Bulot, and G. Balmino. Thermal isostasy in the
South Atlantic Ocean from geoid anomalies. *Earth and Planetary Science*

Letters, 74:280–290, 1985. URL
https://doi.org/10.1016/0012-821X(85)90028-7. Accessed May 13, 2019.
145

Michael Kuhn, Will Featherstone, and Jonathan Kirby. Complete spherical
Bouguer gravity anomalies over Australia. *Australian Journal of Earth
Sciences*, 56(2):213–223, 2009. URL
https://espace.curtin.edu.au/handle/20.500.11937/34751. Accessed May
11, 2019. 132

Aimo Kuivamäki, Paavo Vuorela, and Markku Paananen. Indications of
postglacial and recent bedrock movements in Finland and Russian Karelia.
Report YST-99, Geological Survey of Finland, 1998. URL
http://tupa.gtk.fi/julkaisu/ydinjate/yst_099.pdf. 316

Erkki Kääriäinen. The Second Levelling of Finland in 1935–1955.
Publication 61, Finnish Geodetic Institute, Helsinki, 1966. 155, 156

Jussi Kääriäinen and Hannu Ruotsalainen. Tilt measurements in the
underground laboratory Lohja 2, Finland, in 1977–1987. Publication 110,
Finnish Geodetic Institute, Helsinki, 1989. 358

Rick Lumpkin and Zulema Garraffo. Evaluating the decomposition of
tropical atlantic drifter observations. *Journal of Atmospheric and Oceanic
Technology*, pages 1403–1415, 2005. URL
https://doi.org/10.1175/JTECH1793.1. Accessed May 11, 2019. 311

Paul J. Melchior. *The Tides of the Planet Earth*. Pergamon Press, Oxford, 1978.
ISBN 978-0-0802-6248-2. 354, 356

Jerry X. Mitrovica, Mark E. Tamisiea, James L. Davis, and Glenn A. Milne.
Recent mass balance of polar ice sheets inferred from patterns of global
sea level change. *Nature*, 409:1026–1029, February 2001. URL
https://doi.org/10.1038/35059054. 316

Mikhail S. Molodensky, V. F. Eremeev, and M. I. Yurkina. *Methods for the
Study of the External Gravitational Field and Figure of the Earth*. Israel
Program of Scientific Translations, Jerusalem, 1962. (Transl. from Russian).
121, 162, 206, 368

Helmut Moritz. *Advanced Physical Geodesy*. H. Wichmann Verlag, Karlsruhe,
1980. ISBN 9783879071067. 368

J. Mäkinen, A. Engfeldt, L. Engman, B. G. Harsson, T. Oja, S. Rekkedal,
K. Røthing, P. Rouhiainen, H. Ruotsalainen, H. Skatt, G. Strykowski,
H. Virtanen, K. Wieczerkowski, and D. Wolf. The Fennoscandian Land
Uplift Gravity Lines: comparison of observed gravity change with
observed vertical motion and with GIA models. Technical report,
Nordiska Kommissionen för Geodesi, 2010. URL
http://www.nordicgeodeticcommission.com/wp-content/uploads/
2014/10/1-Makinen_et_al_land_uplift_gravity_lines.pdf. Accessed May
11, 2019. 306

Silja Märdla. *Regional Geoid Modelling by the Least Squares Modified Hotine
Formula Using Gridded Gravity Disturbances*. PhD thesis, Tallinn University
of Technology, 2017. URL
https://digi.lib.ttu.ee/i/file.php?DLID=9130&t=1. Accessed May 11, 2019.
122, 199

R. S. Nerem, D. Chambers, C. Choe, and G. T. Mitchum. Estimating Mean
Sea Level Change from the TOPEX and Jason Altimeter Missions. *Marine
Geodesy*, 33(1):435, 2010. URL
https://doi.org/10.1080/01490419.2010.491031. 322

John J. O'Connor and Edmund F. Robertson. George Green, 1998. URL
http://www-history.mcs.st-and.ac.uk/Biographies/Green.html. Accessed
May 10, 2019. 27

R. L. Parker. The rapid calculation of potential anomalies. *Geophysical Journal
of the Royal Astronomical Society*, 31:447–455, 1972. URL
https://doi.org/10.1111/j.1365-246X.1973.tb06513.x. Accessed May 11,
2019. 232

Nikolaos K. Pavlis, Simon A. Holmes, Steve C. Kenyon, and John K. Factor.
An Earth Gravitational Model to Degree 2160: EGM2008. In *EGU General
Assembly 2008*, Vienna, Austria, April 13–18, 2008. URL
http://earth-info.nga.mil/GandG/wgs84/gravitymod/egm2008.
Accessed May 11, 2019. 71

Nikolaos K. Pavlis, Simon A. Holmes, Steve C. Kenyon, and John K. Factor. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research*, 117(B4), 2012. URL https://doi.org/10.1029/2011JB008916. Accessed May 11, 2019. 71

W. Richard Peltier. Closure of the budget of global sea level rise over the GRACE era: the importance and magnitudes of the required corrections for global glacial isostatic adjustment. *Quaternary Science Reviews*, 28 (17−18):1658−1674, 2009. URL https://doi.org/10.1016/j.quascirev.2009.04.004. Special issue: Quaternary Ice Sheet-Ocean Interactions and Landscape Responses. Accessed May 13, 2019. 312

PIOMAS. Polar Science Center, PIOMAS Actic Sea Ice Volume Reanalysis. URL http://psc.apl.washington.edu/research/projects/arctic-sea-ice-volume-anomaly/. Accessed May 11, 2019. 342

M. Poutanen, M. Vermeer, and J. Mäkinen. The Permanent Tide in GPS Positioning. *Journal of Geodesy*, 70:499−504, 1996. URL https://doi.org/10.1007/BF00863622. Accessed May 13, 2019. 359

John Henry Pratt. II. On the attraction of the Himalaya mountains, and of the elevated regions beyond them, upon the plumb-line in India. *Philosophical Transactions of the Royal Society of London*, 145:53−100, 1855. URL https://doi.org/10.1098/rstl.1855.0002. 136

John Henry Pratt. II. On the deflection of the plumb-line in India caused by the attraction of the Himalaya mountains and the elevated regions beyond, and its modification by the compensating effect of a deficiency of matter below the mountain mass. *Proceedings of the Royal Society of London*, 9: 493−496, 1859. URL https://doi.org/10.1098/rspl.1857.0096. 136

John Henry Pratt. On the degree of uncertainty which local attraction, if not allowed for, occasions in the map of a country, and in the mean figure of the earth as determined by geodesy ; a method of obtaining the mean figure free from ambiguity by a comparison of the Anglo-Gallic, Russian, and Indian Arcs ; and speculations on the constitution of the earth's crust. *Proceedings of the Royal Society of London*, 13:253−276, 1864. URL https://doi.org/10.1098/rspl.1863.0061. 136

K. Predehl, G. Grosche, S. M. F. Raupach, S. Droste, O. Terra, J. Alnis, Th. Legero, T. W. Hänsch, Th. Udem, R. Holzwarth, and H. Schnatz. A 920 km Optical Fiber Link for Frequency Metrology at the 19th Decimal Place. *Science*, April 27, 2012. URL https://doi.org/10.1126/science.1218442. Accessed May 13, 2019. 175, 176

Ilya Prutkin. Gravitational and magnetic models of the core-mantle boundary and their correlation. *Journal of Geodynamics*, 45:146−153, 2008. URL https://www.researchgate.net/publication/257097255_Gravitational_and_magnetic_models_of_the_core-mantle_boundary_and_their_correlation. Accessed May 13, 2019. 145

Jean Richer. Observations astronomiques et physiques faites en l'isle de Caïenne. In *Ouvrages de Mathematique de M. Picard*. P. Gosse and J. Neaulme, 1731. URL http://www.e-rara.ch/zut/content/pageview/815403. Accessed May 11, 2019. 276

Reiner Rummel and Fernando Sansó, editors. *Satellite Altimetry in Geodesy and Oceanography. Proceedings, International Summer School of Theoretical Geodesy*, volume 50 of *Lecture Notes in Earth Sciences*, Trieste, Italy, May 25 − June 6, 1992. Springer-Verlag. URL https://doi.org/10.1007/BFb0117924. Accessed May 13, 2019. 312

Hannu Ruotsalainen. Interferometric Water Level Tilt Meter Development in Finland and Comparison with Combined Earth Tide and Ocean Loading Models. *Pure and Applied Geophysics*, 2017. doi: 10.1007/s00024-017-1562-6. URL https://www.doi.org/10.1007/s00024-017-1562-6. Accessed May 11, 2019. 358

Oliver Sacks. Henry Cavendish: An early case of Asperger's syndrome? *Neurology*, 57(7):1347−1347, 2001. ISSN 0028-3878. doi: 10.1212/WNL.57.7.1347. URL https://doi.org/10.1212/WNL.57.7.1347. Accessed May 11, 2019. 4

Ernst J.O. Schrama. *The Role of Orbit Errors in Processing of Satellite Altimeter Data*. PhD thesis, Delft University of Technology, 1989. URL https://www.ncgeo.nl/downloads/33Schrama.pdf. Accessed May 11, 2019. 332

Axel R. Schweiger, Ron Lindsay, Linjun Zhang, Mike Steele, Harry Stern, and Ron Kwok. Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research*, 116(C00D06), 2011. URL http://dx.doi.org/10.1029/2011JC007084. Accessed May 17, 2019. 342

Daniel Scuka. GOCE burning: last orbital view. ESA blog, 2013. URL http://blogs.esa.int/rocketscience/2013/11/11/goce-burning-last-orbital-view/. Accessed May 11, 2019. 346

Sea Level Research Group. URL http://sealevel.colorado.edu/. Accessed April 12, 2019. 322

J. Segawa, H. Fujimoto, and S. Okubo, editors. *Proceedings, IAG International Symposium on Gravity, Geoid and Marine Geodesy (GraGeoMar96)*, volume 117 of *International Association of Geodesy Symposia*, Tokyo, Japan, September 30 – October 5, 1996. Springer Verlag. 419, 420

Selen User Manual. URL http://geodynamics.org/cig/software/selen/selen-manual.pdf. Accessed May 15, 2019. 313

SourceForge, Maxima. Maxima, a computer algebra system. URL http://maxima.sourceforge.net/. Accessed May 11, 2019. 327

Wolfgang Torge. *Gravimetry*. de Gruyter, Berlin, New York, 1989. ISBN 978-3-11-010702-9. 368

Wolfgang Torge. Gravity and Tectonics. In Kakkuri (1993), pages 131 – 172. 362

Carl Christian Tscherning and Richard H. Rapp. Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variances. Report 208, Dept. of Geodetic Science and Surveying, The Ohio State University, Columbus, OH, USA, 1974. URL https://earthsciences.osu.edu/sites/earthsciences.osu.edu/files/report-208.pdf. 263

Peter Vaníček and Edward Krakiwsky. *Geodesy — The Concepts*. Elsevier Science Publishers, second edition, 1987. ISBN 978-0-4448-7777-2. 368

Felix Andries Vening Meinesz. Gravity Survey by Submarine Via Panama to Java. *The Geographical Journal*, 71(2):144–156, 1928. URL https://doi.org/10.2307/1782700. Accessed May 13, 2019. 276

Martin Vermeer. Chronometric levelling. Report 83:2, Finnish Geodetic Institute, 1983. ISSN 0355-1962. 175, 288

Martin Vermeer. *Geoid studies on Finland and the Baltic*. PhD thesis, University of Helsinki, 1984. 32, 105

Martin Vermeer. Terrain Reduction and Gridding Techniques for Geoid Determination. In Kakkuri (1993), pages 173 – 181. 232

Martin Vermeer. Geoid determination using frequency domain techniques. In Kakkuri (1993), pages 183 – 200. 216

Martin Vermeer and József Ádám, editors. *Proceedings, Second Continental Workshop on the Geoid in Europe*, Report 98:4, Finnish Geodetic Institute, Masala, March 10–14, 1998. 419, 430

Heikki Virtanen. On superconducting gravimeter observations above 8 mHz at the Metsähovi station. Report 98:5, Finnish Geodetic Institute, Masala, 1998. 19 p. 290

Heikki Virtanen. *Studies of Earth dynamics with superconducting gravimeter*. PhD thesis, University of Helsinki, 2006. URL http://ethesis.helsinki.fi/julkaisut/mat/fysik/vk/virtanen/studieso.pdf. Accessed May 11, 2019. 291

Heikki Virtanen and Jussi Kääriäinen. The installation and first results from the superconducting gravimeter GWR20 at the Metsähovi station, Finland. Report 95:1, Finnish Geodetic Institute, Helsinki, 1995. 15 p. 290

Phuong Lan Vu, Frédéric Frappart, José Darrozes, Vincent Marieu, Fabien Blarel, Guillaume Ramillien, Pascal Bonnefond, and Florence Birol. Multi-Satellite Altimeter Validation along the French Atlantic Coast in the Southern Bay of Biscay from ERS-2 to SARAL. *Remote Sensing*, 93(10), 2018. URL http://dx.doi.org/10.3390/rs10010093. Accessed May 11, 2019. 339

W. R. Peltier, FRSC, Home Page. URL http://www.atmosp.physics.utoronto.ca/~peltier/data.php. Accessed May 15, 2019. 313

Dave Watts. Fourifier, 2004. URL http://www.ejectamenta.com/Imaging-Experiments/fourierimagefiltering.html. Accessed May 10, 2019. 226

Lianxing Wen and Don L. Anderson. Layered mantle convection: A model for geoid and topography. *Earth and Planetary Science Letters*, 146(3–4): 367–377, 1997. ISSN 0012-821X. URL http://222.195.83.195/wen/Reprints/WenAnderson97EPSL.pdf. Accessed May 11, 2019. 145

Hans-Georg Wenzel. Ultra High Degree Geopotential Model GPM3E97A to Degree and Order 1800 Tailored to Europe. In Vermeer and Ádám (1998), pages 71–80. 76

Wikipedia, Dislocation. URL https://en.wikipedia.org/wiki/Dislocation. Accessed April 23, 2019. 283

Wikipedia, John Pratt. URL https://en.wikipedia.org/wiki/John_Pratt_(Archdeacon_of_Calcutta). Accessed April 23, 2019. 136

Wikipedia, Mu-metal. URL https://en.wikipedia.org/wiki/Mu-metal. Accessed April 23, 2019. 290

Wikipedia, Pendulum clock. URL https://en.wikipedia.org/wiki/Pendulum_clock. Accessed April 23, 2019. 275

Wikipedia, Saros. URL https://en.wikipedia.org/wiki/Saros_(astronomy). Accessed April 23, 2019. 302

Wikipedia, Sea level rise. URL https://en.wikipedia.org/wiki/Sea_level_rise. Accessed April 23, 2019. 312

Wikipedia, Seasat conspiracy theory. URL https://en.wikipedia.org/wiki/Seasat#Conspiracy_theory. Accessed April 23, 2019. 319

Wikipedia, Strengthening mechanisms of materials. URL https://en.wikipedia.org/wiki/Strengthening_mechanisms_of_materials. Accessed May 10, 2019. 283

Wikipedia, Sverdrup. URL https://en.wikipedia.org/wiki/Sverdrup. Accessed April 23, 2019. 310

Wikipedia, The Aerospace Corporation. URL https://en.wikipedia.org/wiki/The_Aerospace_Corporation. Accessed April 23, 2019. 205

Wikipedia, TOPEX/Poseidon. URL https://en.wikipedia.org/wiki/TOPEX/Poseidon. Accessed April 23, 2019. 321

Wikipedia, Zero-length springs. URL https://en.wikipedia.org/wiki/Spring_(device)#Zero-length_springs. Accessed April 23, 2019. 279

Wolfram, Difference formula for cosine. URL http://demonstrations.wolfram.com/DifferenceFormulaForCosine/. Accessed April 7, 2019. 185

Wolfram Mathworld, Spherical Harmonic Addition Theorem. URL http://mathworld.wolfram.com/SphericalHarmonicAdditionTheorem.html. Accessed May 11, 2019. 352

L. Wong and R. Gore. Accuracy of geoid heights from modified Stokes kernels. *Geophysical Journal of the Royal Astronomical Society*, 18(1):81–91, 1969. URL https://doi.org/10.1111/j.1365-246X.1969.tb00264.x. 205

G. Wöppelmann, C. Letetrel, A. Santamaria, M.-N. Bouin, X. Collilieux, Z. Altamimi, S. D. P. Williams, and B. Martin Miguez. Rates of sea-level change over the past century in a geocentric reference frame. *Geophysical Research Letters*, 36(12), 2009. doi: 10.1029/2009GL038720. URL https://doi.org/10.1029/2009GL038720. Accessed May 13, 2019. 304

YouTube, Hammer vs. Feather. URL https://www.youtube.com/watch?feature=player_embedded&v=KDp1tiUsZw8#! Accessed April 7, 2019. 4

Dah-Ning Yuan, William L. Sjogren, Alex S. Konopliv, and Algis B. Kucinskas. Gravity field of Mars: A 75th Degree and Order Model. *Journal of Geophysical Research*, 106(E10):23377–23401, 2001. URL https://doi.org/10.1029/2000JE001302. Accessed May 11, 2019. 263

Index