NETHERLANDS GEODETIC COMMISSION
PUBLICATIONS ON GEODESY
NEW SERIES
NUMBER 38

GLOBAL GRAVITY FIELD MODELLING USING SATELLITE GRAVITY GRADIOMETRY

RADBOUD KOOP

1993

NEDERLANDSE COMMISSIE VOOR GEODESIE, THIJSSEWEG 11, 2629 JA DELFT, THE NETHERLANDS
TEL. (31)-(0)15-782819, FAX (31)-(0)15-782348

PRINTED BY W.D. MEINEMA B.V., DELFT, THE NETHERLANDS

ISBN 90 6132 246 4

Contents

Acknowledgements

Many people, friends and colleagues, contributed in many different ways to the accomplishment of this work. I am very grateful for so much help. In particular I like to thank Reiner Rummel for his continuous support in practically any field and Martin van Gelderen who helped me both as a friend and as a colleague. Furthermore, I am much indebted to Wim Corbey, who wrote the time–like version of the full gradiometric analysis program and David Stelpstra for providing both the gradient synthesis program as well as the space–like version of the analysis program. The Center for Space Research of the University of Texas at Austin is gratefully acknowledged for supplying me with the simulated one month GRM gradient data. I also like to thank Gerrit Bakker, Ernst Schrama and Prof. E. Grafarend for the discussions I had with them.

Abbreviations

Aristoteles	Application and Research Involving Space Techniques Observing The Earth's fields from a Low Earth orbiting Satellite
c.p.r.	cycles per revolution
CPU	Central Processing Unit
CSR	Center for Space Research, University of Texas at Austin
ESA	European Space Agency
FFT	Fast Fourier Transform
GPS	Global Positioning System
GRS80	Geodetic Reference System 1980
GTR	General Theory of Relativity
MIMD	Multiple Instruction Multiple Data
MSE	Mean Squared Error
NASA	National Aeronautics and Space Administration
PN	Post-Newtonian
PPN	Parametrized Post-Newtonian
r.m.s.	root mean square
SGG	Satellite Gravity Gradiometry
SIMD	Single Instruction Multiple Data
SISD	Single Instruction Single Data
SST	Satellite-to-Satellite Tracking Sea Surface Topography
STR	Special Theory of Relativity
TR	Tscherning/Rapp degree variance model

Abstract

The gravitational field of the earth can be determined globally and with high precision and resolution by means of a combined Satellite Gravity Gradiometry (SGG) and Satellite-to-Satellite Tracking (SST) mission. In such a mission a spacecraft equipped with a GPS receiver and a gradiometer will be flown around the earth in a low and nearly polar orbit. The GPS receiver is used for the determination of the long spatial wavelengths of the earth's gravitational field and the gradiometer for the short wavelengths. As such the two techniques are complementary. This work focuses on the satellite gradiometry part only.

A gradiometer delivers the second order potential derivatives relative to some local orthonormal coordinate system. In particular, the gravity gradients are determined by a technique called differential accelerometry, in which the outputs of any combination of two out of (ideally) eight accelerometers are differenced. A planar gradiometer consisting of four accelerometers and working according to this principle will be on–board the Aristoteles satellite, a mission planned by the European Space Agency (ESA). The improved knowledge of the earth's gravitational field, resulting from such a mission, can contribute to many earth related sciences, like geodesy (levelling with GPS), satellite orbit determination, solid earth physics (continental lithosphere, polar regions) and oceanography, the latter not only for topics like ocean circulation but also for study of climate changes.

The earth's gravitational potential, together with its first and second order derivatives, is usually expressed as a series expansion. The coefficients of such a series (potential coefficients) describe the gravitational potential globally and are to be determined from SGG. The equations of the gradient series expansions, either in geocentric polar coordinates or in orbital coordinates (Keplerian elements), are used as model equations in the gradiometry analysis process. However, the measured gradients are delivered relative to a local orthonormal coordinate system connected to the instrument. We need, therefore, transformation equations for the potential derivatives between several coordinate systems. By means of a compact, general algorithm, which makes use of certain concepts from tensor analysis, these equations can be derived.

Using a set of known potential coefficients, the series expansions can be used to gain some insight in the signals measured by a satellite gradiometer (spherical harmonic synthesis). Reversely, analyzing a set of gradients to obtain the harmonic coefficients is called spherical harmonic analysis. Both synthesis and analysis are time consuming processes, at least if a high degree and order series expansion is used or a large set of data points is used. Grid computation (making use of FFT routines) and the use of vector computers (for which the algorithms should be adapted) can decrease computation at burden. Numerical errors may enter the computations due to the use of recursive Legendre function computations and numerical quadrature formulas.

For the estimation of potential coefficients from a set of observed gravity gradi-

ents we used an analysis technique based on least squares adjustment, as it was, for the gradiometric case, proposed by O.L. Colombo. Under certain assumptions the normal matrix attains a block–diagonal structure, making it easy to solve the large system of linear equations. The a–posteriori error covariance matrix of the estimated potential coefficients can be computed by means of error propagation without the availability of actual measurements. We carried out such error analysis for several mission scenarios, not only for an idealized mission (polar orbit, no band limitation, full tensor gradiometer), but also more realistic situations like polar gaps, band limitation of the gradiometer and the planar Aristoteles gradiometer measuring only three tensor components. The latter, non–ideal situations have a rather large impact on the results. Apart from numerical singularities due to ill–conditioned sub–blocks of the normal matrix, some blocks may become singular too, which means that certain coefficients are not estimable any more from the observations. Of course the system can be stabilized by adding prior information, but this leads to biased estimates. Furthermore it appears that, even with stabilization, the results with a band limited gradiometer did not meet the requirements. Additional GPS tracking information, as it is planned for the Aristoteles mission, is needed to resolve this problem.

Since we had at our disposal a global set of simulated gravity gradients, we could perform a global recovery of potential coefficients from this set. The procedure showed here is an iterative process, but only a first step is really implemented so far. Two strategies were applied. The space–like method uses the series expansion in geocentric polar coordinates, and it requires the set of observations to be transformed into a global equi–angular grid of averaged values. Such a grid is not required for the so–called time–like method (which uses the series expansion in orbital coordinates), but this method is more time consuming. Both methods give promising results.

In our error analysis computations we assumed an instrumental precision of 0.01 $\mathrm{E}/\sqrt{\mathrm{Hz}}$ (as foreseen for Aristoteles). However, superconducting gradiometers are under development, aiming at a precision of 0.0001 $\mathrm{E}/\sqrt{\mathrm{Hz}}$. Also, orbit determination techniques are gradually improving, aiming at the centimeter level. Such improvements of precision may, in the future, require a relativistic formulation of the model. Furthermore, from a theoretical point of view, a relativistic description of satellite gradiometry emphasizes that a gradiometer measures the curvature of four–dimensional spacetime. For those reasons, we included a relativistic view on gradiometry. The equations are derived in the weak field approximation (in particular the so–called Post–Newtonian approximation), which is sufficient for earth orbiting satellites. The equations of motion of the satellite appear to be those of a spacetime geodesic and they show the relativistic contributions to the satellite's orbit. A relativistic description of the second order potential derivatives appears to be governed by the so–called equation of geodesic deviation. The latter shows the relativistic contribution to the observed gravity gradients. It is concluded that relativistic effects can be modelled although they are not required at present.

1

Introduction

In spite of the fact that the concept of satellite gradiometry is more than 30 years old, no actual gradiometric mission has been undertaken yet. Several instrument and mission proposals have been done, of which the Aristoteles mission concept of the European Space Agency (ESA) is the most promising one for the near future. The main purpose of a Satellite Gravity Gradiometry (SGG) mission is the global determination of the earth's gravitational field with high precision and spatial resolution. Some central aspects of the process to derive from the actual observations the required gravitational information will be treated in this work. This explains the title "Gravity Field Modelling using Satellite Gravity Gradiometry". In this title, three topics attract our attention. Obviously, the key issue here is **gravitation**, which word is reflected in the word *gravity*. The word *gradiometry* reveals that we have to do with **measurements**. Finally, the word **satellite** points out that the measurements are carried out in a spacecraft.

Gravitation

The tale of the falling apple marks one of the great moments in scientific history. According to it, *Isaac Newton* (who lived from 1642 to 1726), when sitting in his garden, watched an apple fall from a tree. The apple fell straight down, along a line which, if extended inside the earth (considered a homogeneous sphere at the moment), would go through the center of the earth. While thinking about this phenomenon, he came up with his famous law about the mutual attraction of two masses. The attractive force, called *gravitation*, is directed along the line connecting the two centers of the objects and is, up to some constant (the gravitational constant G), proportional to both masses as well as to the squared inverse of the distance between the two objects. This famous *inverse-square-law* thus tells us that on the one hand the gravitational force increases with increasing mass of one of the objects, whereas on the other hand the force decreases with increasing distance between the objects. Objects moving under the influence of a force will be accelerated in the

same direction as the force acts. Whereas the gravitational force depends on the mass of the object, the *gravitational acceleration* does not.

Although Newton lived more than two and a half centuries ago, we still use the inverse–square–law for all earth related applications. Not only does this law explain the falling of an apple. By using the same law, the motion of the planets around the sun and of moons and satellites around the planets or the trajectories of projectiles can be explained. In the beginning of the 20th century, however, measurement techniques became more accurate, and some aspects of the motion of the planets in the solar system which showed up in the measurements, could no longer be explained by Newton's law. *Albert Einstein* (1879–1955) came up with a new theory about gravitation. Einstein's gravitational theory is the famous *general theory of relativity* (GTR). The addition "general" refers to the fact that gravitation is included in the theory, where this was not the case in the so–called special theory of relativity (STR), which he conceived some years earlier. Einstein's GTR did not state that Newton was wrong. In fact, Newtonian theory represents a limiting, approximate case of the GTR. For cosmological applications, the use of the GTR is inevitable, but for most earth related applications we can still confine ourselves to Newtonian theory. However, with increasing measurement accuracy, as for instance is the case with many modern satellite related measurement techniques, relativistic effects might have to be taken into account.

If the earth would have a perfectly spherical shape and if the mass inside the earth would be distributed homogeneously (i.e. equal mass density throughout the whole sphere) or rotationally symmetric, the line along which Newton's apple fell would indeed be a straight line, directed radially and going exactly through the earth's center. This center point would be the *center of mass* of the earth. The gravitational force exerted by the earth onto a proof mass (like the apple) would be exactly the same as the force exerted by a point mass of infinite small dimension with equal mass as the earth and located at the center of the earth. At any other point at the earth's surface the force on the apple would have exactly the same magnitude and (radial) direction. Moving away from the earth's surface, the force would still be directed radially, although its magnitude would decrease with increasing altitude. The *gravitational field* obtained in this way would be perfectly spherically symmetric.

In reality, however, the situation is more complex. Although seen from the moon or from any other point in space the earth may look much like a perfect sphere, the deviations from this idealization are significant. Due to the fact that the earth is deformable and rotates about an axis going approximately through the north and south pole, it is better represented as an *ellipsoid*, flattened at the poles. Looking in more detail we also see mass irregularities at the surface of the earth, like oceans, mountains, plains, etc. Also the material inside the earth is not distributed homogeneously: the dynamics of the earth are more considered a convective system, with an enormous variation in temperature, resulting in phenomena such as plate tectonics, subduction zones, ocean ridges and a considerable radial and lateral density differentiation. As a result, the gravitational force at different places on the earth

will not be the same: both magnitude and direction of the force will deviate from one place to the other from that of a homogeneous sphere or *spheroid.*

The gravitational field of the earth is rather irregular, although the deviations from a spherical symmetric shape are not very large. In fact, for some applications a spherical earth may be a sufficient approximation. For other applications an ellipsoidal earth may be needed, but often higher–order approximations are required. Therefore, the gravitational field is represented by a *series*, of which the spherical part is the first term, the ellipsoidal part the second term, and so on. The more detailed information or the more accurate information about the gravitational field we need, the more terms are required in the series expansion for an adequate representation.

If we would exactly know the (irregular) mass distribution of the earth, we could compute the gravitational field from it (and thus obtain all terms in the series expansion) and see how much this field deflects from a perfectly spherically symmetric one. Unfortunately we still do not know the exact mass distribution. On the contrary, one of the main objectives of gravitational field determination is to learn more about the internal mass constitution. Unfortunately an exact inference is impossible. One speaks of the gravitational inverse problem.

Measurements

If Newton would have measured the time at which the falling apple would have passed through imaginary levels of known distance, he could have computed the gravitational acceleration. Modern instruments still use such *free fall* technique to measure gravitation. More precisely, when situated on the rotating earth, not gravitation is measured but *gravity.* Gravity is the sum of gravitation (attractive force between two masses) and *centrifugal acceleration.* The latter is due to the rotation of the earth and its direction is perpendicular to the axis of rotation. Today's instruments are very accurate. In the case of an *absolute* gravity apparatus, a proof mass is dropped in a vacuum chamber and its path is measured interferometrically with a laser. In this way one obtains the magnitude (or length) of the gravity vector, at the specific place where the measurement is done, to 10^{-9}. Such absolute gravity measurements are difficult and therefore expensive. They are carried out only at a few points on earth. Since gravity on earth is approximately 10^{1} m/s^2 the achieved precision corresponds to accelerations as small as 10^{-8} m/s^2.

In order to get an idea about gravity in between absolute stations, one measures gravity *differences* at arbitrary measurement points relative to the absolute stations. Such *relative* gravity measurements are easier to perform than absolute measurements and are carried out with *spring gravimeters.* Thereby, the length of a spring, mounted under a particular angle and suspending a horizontal lever with a proof mass, is measured. With modern spring gravimeters a precision of 5 to 10 $\cdot$ 10^{-8} m/s^2 can be reached.

The direction of the gravity vector is defined by the astronomical latitude and

longitude, i.e. the direction of the plumb line as determined by astronomical observations. Thus the gravity vector in any point is determined by three quantities: astronomical latitude and longitude (direction) and magnitude (usually referred to as *gravity*). In our surrounding three dimensional world any vector can be determined by three quantities: its components in all three spatial directions. Mathematically speaking these three components are the *gradient* of some scalar function, in our case the *gravity potential*. They describe how the scalar function changes when moving in all three directions. But we have seen that gravity itself also varies from place to place. These changes of gravity can also be described by gradients, and are thus called *gravity gradients*. They also can be used to describe the gravitational field of the earth and are more sensitive to small variations of the earth's mass constitution.

At the beginning of our century, the Hungarian physicist Eötvös (1848–1919) developed an instrument to measure gravity gradients. This instrument was a so–called *torsion balance*[1]. Instead of one proof mass used in a gravimeter, a torsion balance consists of *two* proof masses, (asymmetrically) suspended to the arms of a balance. The gravity gradients produce a torque on the beam of the balance, resulting in a rotation of the beam. The gravity torque is counterbalanced by a restoring torque exerted by the fiber with which the beam is suspended. The restoring torque is a measure for the gravity gradients. Eötvös achieved with his instrument a precision of 1 E (1 E = 1 Eötvös unit = 10^{-9} /s^2). With the largest gradient on earth being approximately 3000 E, this means a precision of 10^{-3}. In general instruments which measure gravity gradients are called *gradiometers*, and the measurement technique is called *gradiometry* (measuring the gradients of gravity), analogous to the word *gravimetry* (measuring gravity).

In principle, Newton could have measured gravity gradients if he would have seen *two* apples fall simultaneously from the tree and would have closely followed their adjacent trajectories. Both trajectories, if extended inside the earth, would converge as to run through the earth's center of mass, so the distance between the paths would gradually decrease. If the change in the distance between two falling proof masses would be measured, it would be a measure of the gravity gradient, i.e. of the variation of gravity in the direction of the line connecting the two masses[2]. In a next step, one could think of constraining the motion of the falling proof masses,

[1] With the same type of instrument Eötvös did a historical test to 10^{-8} of the so–called *principle of equivalence* of inertial and gravitational mass (Eötvös, 1953), (Jung, 1961). In recent years his tests were reanalyzed in relation to a search of the so–called *fifth force* (Fischbach et al., 1986).

[2] In the language of the GTR, the two proof masses, being in free fall in the gravitational field of the earth, follow nearby *geodesics* ("shortest paths"). In a flat space, like a two dimensional sheet of paper, geodesics are straight lines, which, if being parallel initially, remain parallel (constant distance between the lines) and never cross. In the GTR, the space under consideration is not flat but curved, just like the two dimensional surface of a sphere. Furthermore it is four dimensional, where three spatial directions and time merge into one so–called *spacetime* (which is hard to visualize). Geodesics in curved spacetime are therefore not straight lines but are curves. The distance between initially parallel geodesics changes and this distance change is a measure for the curvature of spacetime. So in terms of the GTR, gravity gradients describe the curvature of spacetime.

such that – during their fall – they are held in a fixed position relative to each other and their distance would remain constant. The force needed to constrain the motion could be measured and would again be a measure of the gravity gradients. This is the principle used in *spaceborne gradiometry*. There, a gradiometer, consisting of two or more proof masses, arranged in a two or three dimensional orthogonal set–up, is flown around the earth on board a spacecraft. Whereas the gravity gradient is (mathematically) defined in exactly one point of infinitesimal dimension, a real life gradiometer has some definite size. Practical limitations (i.e. material properties) prevent us from reducing the size of an apparatus ad infinitum, so the distance of two proof masses in a gradiometer, although small, is not infinitesimal, but has some definite value, say 1 meter[3]. This means we are actually measuring the *difference* in gravitational acceleration between the locations of the two proof masses. For that reason the technique is often called *differential accelerometry*. ESA has planned a satellite mission called Aristoteles, which should carry a gradiometer measuring gravitational gradients using the differential accelerometry principle.

Satellites

The Aristoteles project is a so–called dedicated gravity field mission, its main purpose being the determination of the earth's gravitational field. The idea, however, of using earth orbiting satellites for such a purpose is not very new. Already from the time of the launch of the first artificial earth orbiting satellite in 1957, people analyzed their orbits to determine the main characteristics of the earth's gravitational field. As said before, the orbits of satellites around the earth or around any other planet, as well as the orbit of planets around the sun, are determined predominantly by the gravitational field of the earth, respectively of other planets or the sun. In the 17th century, even before Newton, *Kepler* established three laws concerning the motion of planets around the sun. Actually Newton used Kepler's laws when deriving his famous inverse–square–law. According to Kepler, planets move in elliptic orbits around a spherical body. The earth is nearly spherical, its gravitational field is nearly spherically symmetric, and thus the orbits of earth orbiting satellites are nearly elliptical. The main deviation from spherical symmetry, the earth's flattening, results in a precession of the orbital plane and a precession of the orbit ellipse in the orbital plane. Other departures from spherical symmetry in the gravitational field (in general called *anomalies*) yield other, smaller, deviations in the satellite's orbit. Precise analysis of satellite orbits thus tells us something about the gravitational field.

The principle is not so very much different from Newton's falling apple. Also satellites "fall" in the gravitational field of the earth. The difference is that Newton's apple started, relative to the earth, from a rest position at the branch of the tree

[3]Actually, according to the principle of equivalence, an infinitesimal small gradiometer would not exist. Strictly speaking this principle holds in exactly one (infinitesimal small) point only, where one could no longer discriminate between inertial and gravitational accelerations.

and fell straight down on the earth's surface. Would it have been thrown away horizontally (i.e. parallel to the earth's surface) with high velocity, it would have hit the ground some distance away from him. The higher above the earth's surface a proof mass is ejected ("launched") and the higher its initial velocity, the further away the point of hitting the ground. Eventually, due to the spherical shape of the earth, the proof mass never hits the earth's surface, but continues to fall, so to say not *to* the earth, but *around* it. Obviously, as already pointed out by Newton, the same law which describes the fall of the apple, describes the motion of satellites around the earth.

Considerable effort has been invested into the determination of the earth's gravitational field from the analysis of satellite orbits. The same method will, additional to the gradiometry part, be used with ESA's Aristoteles mission. Especially the main deviations from a spherically shaped gravitational field can be determined from orbit analysis. Other, smaller deviations, i.e. the more detailed structure, represented by higher–order gravitational field parameters, are determined from the gradiometer measurements.

Nevertheless, the range of detail with which the gravitational field can be determined from space, is not unlimited. According to the inverse–square–law, gravitation decreases with increasing distance between the objects. This so–called *attenuation* effect is especially noticeable with satellites, since they move at very large distances from the earth's surface. Even with Aristoteles, which will move at a "very low" altitude of "only" 200 km, many details of the gravitational field will remain undetectable. Furthermore, at 200 km there is still some atmospheric density left, which disturbs the motion of the satellite and due to which measurements are extremely difficult. Atmospheric drag causes the satellite to descend, so that it needs to be kept in a 200 km orbit by maneuvers, using small rockets. These maneuvers also disturb the measurements, as does e.g. the sloshing and consumption of the fuel in the tanks needed for the rockets.

A great advantage of using satellites for gravity field determination is that with one single mission (nearly) the whole earth can be covered with measurements in short time. Of course, launching a satellite is expensive, but travelling the earth by conventional means with an instrument and carrying out measurements with corresponding density as a satellite, is much more expensive. Just think about high mountains, oceans and polar regions, which may constitute insurmountable hindrances for doing terrestrial measurements. Even after more than 50 years of terrestrial measurements large parts of our planet remain unsurveyed. Furthermore, a set of terrestrial measurements is not likely to have comparable precision all over the world, whereas the satellite mission will deliver us a global data set of homogeneous quality.

This study in brief

This study deals with the main subjects of a global determination of the earth's gravitational field from a satellite gravity gradiometry (SGG) mission, and as such includes aspects of gravitation, measurements and satellites as discussed above. Besides introductory remarks on the principles of satellite gradiometry and the applications of a precise and highly detailed gravitational field, this study focuses mainly on the *analysis* process of converting the observables (gravity gradients at satellite altitude) into the gravitational information at zero altitude. Furthermore, special attention is given to the *precision* of the derived gravitational information, being a function of measurement precision, mission parameters and chosen mathematical model. Finally, as a look into the near future, a description of Einstein's GTR is given, in as far as it may concern an SGG mission. Relativistic corrections to Newton's theory are derived, as may be necessary to account for in future gradiometer missions.

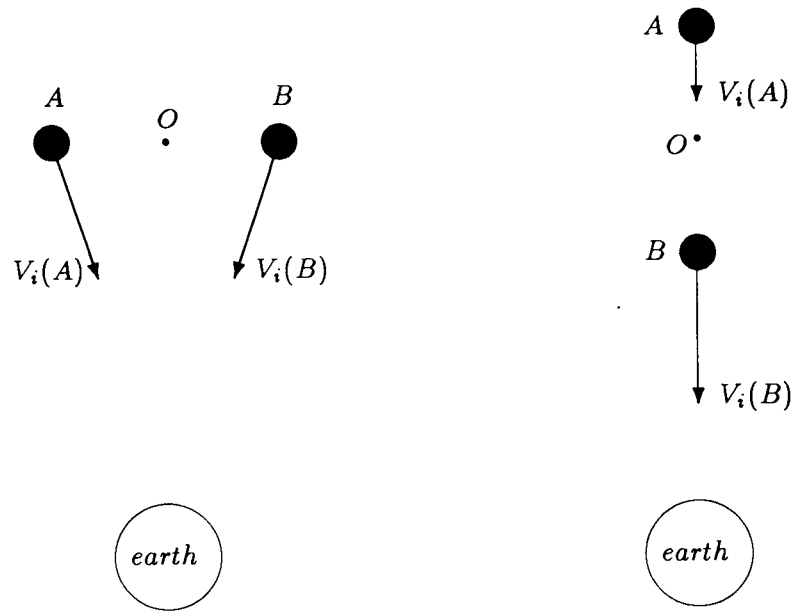
2 |

Satellite gradiometry: principles and applications

The purpose of gradiometry is the precise and detailed determination of the earth's gravitational field. This is done by measuring the second–order derivatives of the gravitational potential V of the earth. An instrument which measures these second–order derivatives is called a gradiometer. In satellite gradiometry such a gradiometer is flown in a satellite around the earth in a low, almost circular and polar orbit. In this chapter the principle of satellite gradiometry will be illustrated by looking at the motion of test masses in space. Furthermore a short overview of the possible applications of gradiometry is given, as well as a short discussion about instrumental aspects concerning the gradiometer and the satellite in which it is flown, with special attention to the planned Aristoteles mission.

2.1 Principle of satellite gradiometry

Consider first two proof masses situated in two nearby points A and B in space, see figure 2.1.a. The gravitational acceleration in point A due to the attraction of the earth is $V_i(A)$. The components are given with respect to a local cartesian coordinate system x^i. See appendix A for the definition of various coordinate systems and appendix B for some remarks on notation. The gravitational acceleration vector is directed along the line of force going through point A, perpendicular to the equipotential surface through A. The gravitational acceleration in B is $V_i(B)$. Suppose A and B are situated on the same equipotential surface. Without any support and without any other forces appearing, the two proof masses will fall towards the earth. Since the gravitational field of the earth is almost a perfect central force field, the distance between the proof masses will decrease while falling towards the earth due to the convergence of the lines of force. The change in distance between the proof masses is a measure for the difference in gravitational acceleration in A and B, cf. (Carroll and Savet, 1959) and (Savet, 1969).



a. Proof masses on the same equipotential surface

b. Proof masses on the same line of force

Figure 2.1 *Falling proof masses.*

Now consider the two points A and B situated on the same line of force but on different equipotential surfaces, see figure 2.1.b. The proof mass in B is closer to the earth as the one in A and is therefore pulled harder. If the proof masses are dropped and are falling towards the earth along the same line of force, the distance between them will increase. Again, the change in this distance is a measure for the difference in gravitational acceleration between the proof masses.

In a next step, consider a spacecraft (satellite) carrying out an orbital motion around the earth. The spacecraft, or more precisely the center of mass O of the spacecraft, is in free fall in the gravitational field of the earth. In case of a circular orbit, the gravitational acceleration at O, which is directed towards the earth, is, at any moment, compensated by the centrifugal acceleration resulting from the orbital motion, which is directed outwards. In O no resultant forces appear. Suppose that the center of mass O of the satellite is not a material point, so that we can place one of the two proof masses from above at O. The proof mass will then also be in free fall around the earth, carrying out the same orbital motion as the satellite and it will remain at O (and thus at rest relative to the spacecraft) during the motion.

Suppose now that we place the second proof mass at another point in the interior of the spacecraft, close to O. The second proof mass will be situated on another line of force and/or another equipotential surface of the gravitational field, so it will

start to move relatively to the first proof mass at O. This relative motion results in a distance change between the proof masses according to the examples above (fig. 2.1).

If both proof masses are placed inside the satellite at arbitrary points but not at O, again a relative motion between the proof masses will appear as a result of the difference in gravitational acceleration at the two points, cf. (Rummel, 1989a), but furthermore they will move relatively to O. So they will start drifting inside the satellite, eventually hitting each other or the satellite skin.

This skin consists of points, which, together with all other (material) points of the satellite, form a rigid body. All points of this body carry out the same orbital motion as O, i.e. they have the same angular velocity (disregarding any deformations). Although they are at any moment situated at other points in the gravitational field than O they remain, as a result of the material composition, in a fixed position with respect to O.

Suppose now that the two proof masses are also kept in a fixed position with respect to O by means of some mechanical or electrostatical suspension. Then also the proof masses are forced to carry out the same orbital motion as O. Proof masses constrained in this way can be seen as accelerometers. The outputs of these accelerometers are the forces needed to keep the proof masses in these fixed positions with respect to O (and thus with respect to each other). They are measures for the acceleration differences between the two points and O and can therefore be used as observations to measure the gravitational field of the earth.

These relative accelerations of the two proof masses can be expressed by expanding the gravitational acceleration V_i in a Taylor series with respect to O. For the points A and B we have:

$$\begin{aligned} V_i(A) &= V_i(O) + V_{ij}(O)dx^j(O,A) \\ V_i(B) &= V_i(O) + V_{ij}(O)dx^j(O,B) \end{aligned}$$

where we only kept the linear term. The $dx^i(O,A)$ are the coordinate differences between O and A. The overall motion of the satellite (of the center of mass, i.e. $V_i(O)$), which is common to both points, can be eliminated by taking the difference between the acceleration in A and in B. We obtain:

$$\begin{aligned} dV_i(A,B) &= V_i(B) - V_i(A) \\ &= V_{ij}(O)(dx^j(O,B) - dx^j(O,A)) \\ &= V_{ij}(O)dx^j(A,B)\ . \end{aligned} \tag{2.1}$$

If $dx^i(A,B)$, the distance between the points A and B, is known, we may compute from this equation the gravitational gradients $V_{ij}(O)$ since dV_i is measured by means of the accelerometers. An instrument consisting of two or more of such accelerometers and having as output the gravitational gradients is called a *gradiometer*.

The technique of measuring the second–order potential derivatives in the way described above is known as *differential accelerometry*, see for example (Forward, 1974)

or (Balmino et al., 1985). A gradiometer working according to this principle will consist of a symmetrically arranged array of accelerometers. The gradiometer will be built in the satellite in such a way that the center of the instrument coincides with the center of mass O of the satellite.

The neglect of the terms proportional to the third– and higher–order derivatives of the potential is of the order of 10^{-10} E (1 E = 1 Eötvös Unit = 10^{-9} s^{-2}) for a distance between A and B of about 1 m and at an altitude of the satellite of about 200 km, which is acceptable in view of the present and near future gradiometer instrument precision, cf. (Paik and Richard, 1986).

So in principle it should be possible to obtain information about the gravitational field of the earth by observing the gravitational gradients V_{ij} in the way described above. However, up to now we considered the two proof masses and the satellite moving through space under the influence of gravitational and rotational forces only. In reality also other, external and non–conservative, forces may act on the satellite. Such forces are due to, for example, solar radiation pressure, and, at lower altitudes, air drag due to the earth's atmosphere. They act on the outer skin of the satellite, the satellite housing. But if the gradiometer instrument is rigidly attached to this housing, these forces will also act on the instrument and therefore also on the proof masses due to the suspension. The measured accelerations are in this case a mixture of gravitational accelerations and external forces. Since we are only interested in gravitational forces, such external forces are viewed upon as disturbing forces.

One way to deal with these disturbing forces is to build a so–called *drag free* satellite. In such a satellite the instrument is not rigidly attached to the satellite housing. This housing, the "outer part" of the satellite, undergoes the non–gravitational disturbing forces causing a relative motion between the housing and the instrument (or the "inner part" of the satellite). In order to prevent collision of the two the motion of the outer satellite is continuously regulated. The inner satellite then carries out a perfect free motion, under the influence of gravitation (besides that of the earth also of the sun, moon and other planets) only.

On the other hand, since the external forces act on every accelerometer in the same manner (i.e. same direction and same magnitude), they are, just like the common acceleration $V_i(O)$, eliminated if we take the difference between two accelerometer outputs, as in eq. 2.1. This elimination, however, only works properly if the accelerometers are exactly identical (in dimensions and in orientation) and if they are perfectly aligned relatively to one another and relatively to the instrument. The use of a non drag–free satellite therefore puts more stringent requirements to the construction of the instrument.

A final aspect which we will consider here is the rotation of the gradiometer instrument. Suppose the gradiometer is at rest with respect to the local cartesian coordinate system x^i, of which the origin coincides with the center of mass O of the satellite, and thus with the center of the instrument. Suppose furthermore that, at a certain moment (e.g. the initial point of the mission), this local coordinate system is oriented such, that the z–axis is directed radially outwards, away from the

earth, the x–axis points in the direction of the motion of the satellite (along track) and the y–axis is perpendicular to the orbital plane (cross track) as to complete a right–handed coordinate system. The satellite (or the instrument) is said to have a space–fixed orientation if it orbits the earth such that it keeps the same orientation with respect to the fixed stars, i.e. the local coordinate system remains parallel to the initial position. This means that the z–axis after the initial point no longer points in the radial direction, the x–axis no longer along track, only the y–axis remains pointing in the cross track direction. On the other hand the coordinate system (and thus the instrument) does not rotate in this case.

If we let the z–axis keep its radial direction and the x–axis its along track direction, the satellite is said to be earth pointing. In this case the coordinate system *does* rotate. In particular it performs one complete rotation during one orbital revolution of the satellite. Its angular velocity in that case is $2\pi/T$, where T is the orbital period. In our example the rotation takes place about the y–axis. From an earth point of view, such earth pointing motion may be attractive, but it has the disadvantage that inertial accelerations occur due to the rotational motion. These accelerations are also present in the output of the accelerometers. Since the accelerometers all have different positions with respect to O, the effect of the rotation is not cancelled during the differencing of the accelerometer measurements as in eq. 2.1. But if we can discriminate between the gravitational and the rotational accelerations by either some numerical method or some a–priori rotational information (so that the rotational motion can be considered known), cf. e.g. (Rummel, 1986), we have a means of measuring the earth's gravitational field, called *satellite gradiometry*. For a comprehensive general treatment of satellite gradiometry, see for example (Rummel, 1985a, 1985b, 1986). For related topics on the principles of satellite gradiometry see e.g. (Forward, 1981, 1982), or (Moritz, 1968).

2.2 Applications

The aim of a gradiometric satellite mission is the determination of the earth's gravitational field, globally with high precision and high spectral resolution. In the case of the first planned mission, Aristoteles, one would like to obtain the following precisions (see also section 4.2)

$$\begin{aligned} \text{geoid heights} &: \sigma < 10 \text{ cm} \\ \text{gravity anomalies} &: \sigma < 5 \text{ mgal} \qquad (1 \text{ mgal} = 10^{-5} \text{ m s}^{-2}) \end{aligned}$$

both with a spatial resolution of between 50 and 100 km (half–wavelength). This corresponds to a spherical harmonic expansion (see section 3.2) complete up to degree and order 200 to 300.

So far our knowledge of the earth's gravity field relies on the one hand on the

analysis of the motion of satellites and on the other hand on terrestrial measurements. Several groups collect terrestrial point gravity anomalies and process them to equal-angular mean gravity anomalies, e.g. of size $1° \times 1°$, $6' \times 10'$ or $3' \times 5'$ (corresponding in our latitudes to 100 km, 10 km or 5 km side-length). The original point anomalies are derived from relative gravity measurements that are tied in some countries into national first-order networks, containing some absolute stations, or into the International Gravity Standardization Network. For the computation of anomalies also heights are necessary. Ideally they should be levelled heights referring to a well-defined height datum. In practice one has often to sustain with barometric heights or not well-defined local levelling networks. As a consequence it must be feared that the mean gravity anomalies in some areas of the world contain systematic errors. For large areas no gravity anomalies are available at all, either because of political reasons the data are not made available or because areas are not easily accessible, e.g. polar regions, high mountain ranges. In particular ocean areas, where gravity measurements are difficult to do and very costly, large areas are not covered. A map of the current coverage with $1° \times 1°$ mean anomalies is given in figure 2.2. Despite large white areas the impression is seemingly not too bad. However, inspection of a histogram of the precision of these values shows that only a rather small portion meets present day requirements, see figure 2.3.

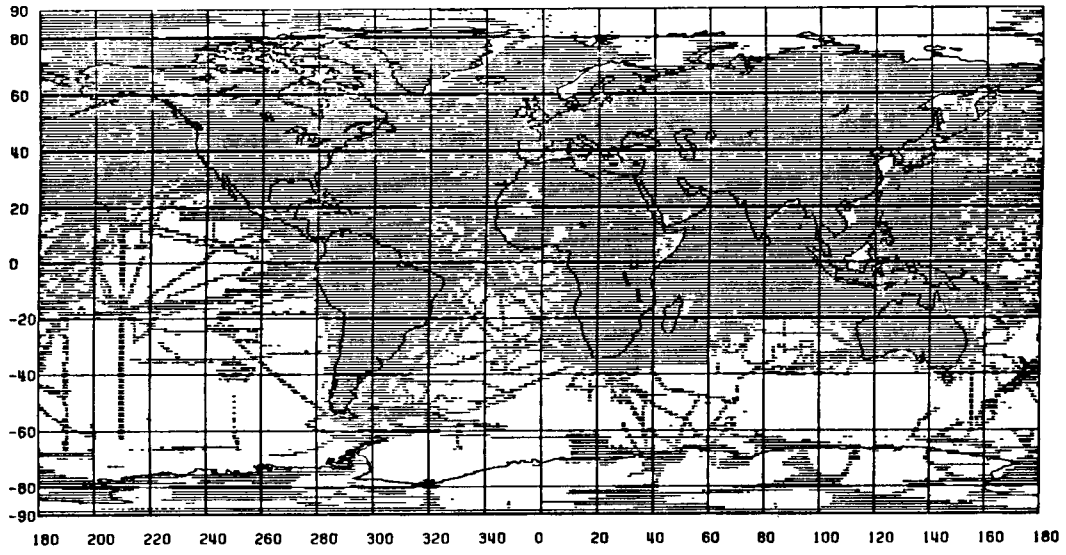


Figure 2.2 *World map of the coverage with $1° \times 1°$ mean anomalies (Rapp, 1977).*

One way to get gravity information in ocean areas is to convert heights, obtained from satellite altimetry, into mean gravity anomalies. There exist, however, two principal objections against doing this. First, the altimetric heights, even after

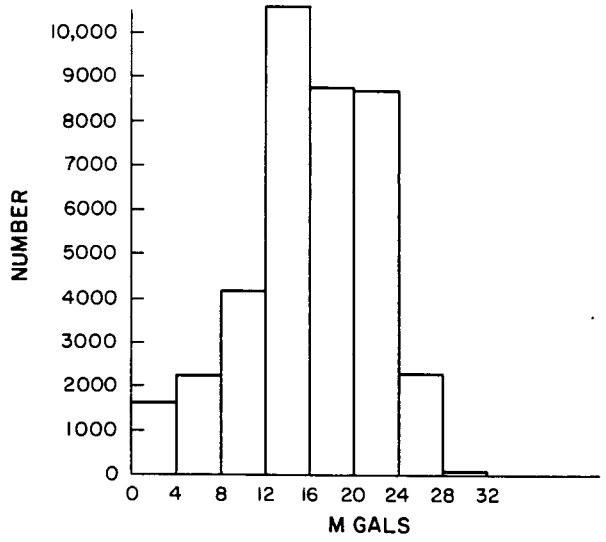


Figure 2.3 *Histogram showing the precision of the $1° \times 1°$ mean anomalies from figure 2.2 (Rapp, 1977).*

subtraction of our best models of the ocean topography, will not coincide with the geoid. Hence systematic errors will affect the derived gravity anomalies. These types of "gravity anomalies" should certainly never be used for geoid determination in the context of ocean studies. Secondly, the conversion method can theoretically not be confined to ocean areas and is intrinsically unstable. Also, because of this, the actually applied numerical methods will introduce uncontrolled biases. However, for some purposes, e.g. in geophysics, the gravity anomalies determined in this manner are useful.

Neither with terrestrial (including shipborne) nor with altimetric derived gravity anomalies a global coverage can be achieved in the foreseeable future. Global gravity information comes from the analysis of satellite orbits and is expressed in sets of spherical harmonic coefficients currently up to degree and order 30 to 70. These sets are called geopotential models. They are derived at a few computing centers in the world in a complicated estimation process consisting of several phases. In essence, the tracking data (laser, microwave, etc.) from a large number of stations to a large number of satellites is analyzed. Their orbits are determined and combined in a least squares adjustment to yield e.g. spherical harmonic coefficients of the gravity field. The separability of the individual coefficients and their precision depends largely on a good distribution of the employed satellites in terms of orbit characteristics (inclination, altitude, eccentricity). The error standard deviation per degree of some of the recent geopotential models is shown in figure 2.4. The spatial resolution is still limited. For some models a signal to noise ratio of one is reached near degree 50.

We conclude that our current knowledge of the gravity field is far from reach-

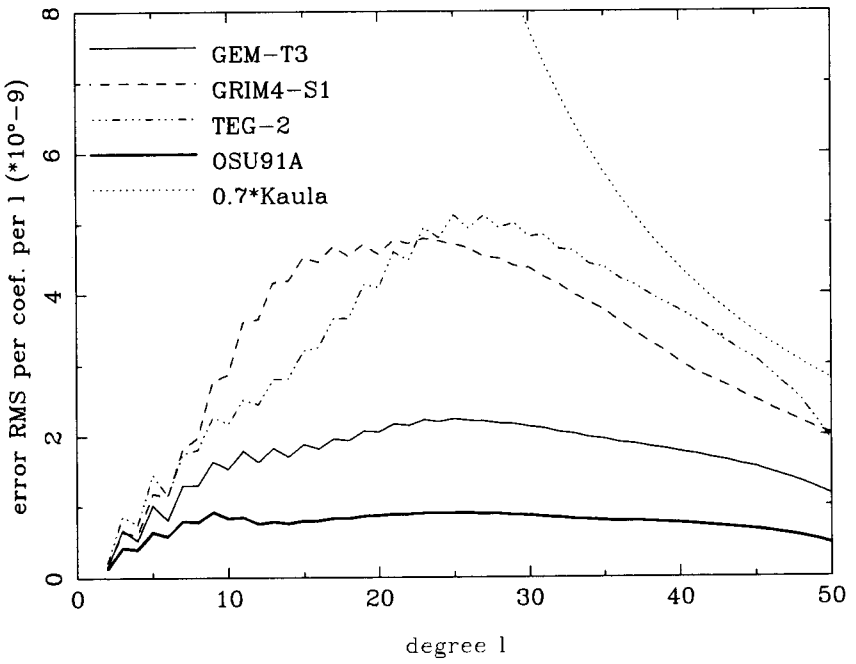


Figure 2.4 *Error standard deviation of some recent geopotential models.*

ing the level of precision, resolution and completeness aimed for by gradiometry. This leads to the question why one needs such good knowledge of the gravity field. Examples in four fields of applications serve as illustration.

Example One: Geodesy and levelled height

Levelled heights are used in geodesy for mapping, civil constructions, monitoring of land subsidence, control of tide gauges, etc. The process of levelling is very time consuming and therefore expensive. In recent years satellite positioning by GPS (see next section) became available. GPS measurements deliver meanwhile relative positions between points at a 10^{-7} to 10^{-8} level (depending on the baseline length). The cartesian coordinate differences $\Delta X, \Delta Y, \Delta Z$ between two points in the global GPS system can be converted into differences in geodetical latitude, longitude and height. Unfortunately are the height differences conventional, referring in a purely geometrical sense to an ellipsoid. A height difference, Δh, of this kind could, however, be converted into a quasi–levelled height difference ΔH were the geoid height ΔN available:

$$\Delta h = \Delta N + \Delta H \ .$$

The geometric situation is sketched in figure 2.5. The precise computation of ΔN requires precise knowledge of the earth's gravity field. The gravity field as obtained by satellite gradiometry would suffice to determine the "absolute" geoid with a precision of $\pm$ 35 cm or relative with $\pm$ 52 cm, $\pm$ 18 cm or $\pm$ 2 cm over a distance

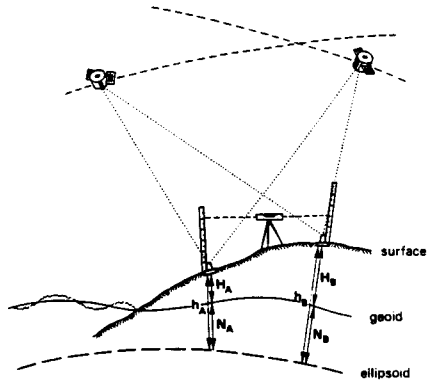


Figure 2.5 *Relation between reference surface (ellipsoid), geoid and topography (ESA, 1991).*

of respectively 100, 10 and 1 km. In combination with local gravity information, as available in North–America, Australia or Europe, cm–precision is feasible and levelling can in many instances be substituted by a combination of GPS and a precise geoid.

Example Two: Precise satellite orbits

The main obstacle on the way to orbit determination with cm–precision is the inaccuracy in our knowledge of the gravity field. Of course also non–gravitational effects form a limitation, in particular for bulky spacecraft, but could be eliminated in principle by a drag–free set–up or adequate parametrization. A precise gravity field – in combination with accurate and dense tracking – would result in orbit accuracies of a few cm. This would not only significantly increase the value of ocean, ice and land altimetry, but be useful for geo–kinematical applications too.

Example Three: Solid earth physics

The applications of an improved gravity field in solid earth physics were described in a number of reports. It is referred to (SESAME, 1986), (NASA, 1987) and (Lambeck, 1990). In geophysics it is useful to distinguish between studies concerned with core or core/mantle boundary, mantle convection processes and lithosphere. Satellite altimetry in combination with bathymetric data brought a much improved understanding of the oceanic lithosphere. Surprisingly it is the continental lithosphere, as well as the polar regions, where currently better insight is desirable.

Due to the inverse character of the problem gravity alone shall never suffice for a determination of the density structure of the earth's interior. However, precise gravity and topographic data in combination with regional and global seismic tomography would drastically reduce the uncertainty range of possible solutions, see e.g. (Hager, 1983) (Dziewonski, 1984) (Woodhouse and Dziewonski, 1984)

or (Spakman, 1988).

Example Four: Oceanography and climate change

The area where the need for an improved gravity field probably is most pressing is oceanography. Like terrain topography is defined as the deviation H of the topographic relief from an equipotential surface (geoid) the sea surface topography (SST) is the deviation of the actual ocean surface from the geoid. The ocean surface with no external forces such as tides, winds, storms, etc. would be level, coinciding with an equipotential surface. Hence any deviation from a level surface can directly be attributed to ocean dynamics. Satellite altimetry provides sea surface heights h relative to a chosen reference surface (ellipsoid). If a precise gravity field were available precise geoid heights N could be determined and from the relationship

$$h = N + H$$

the SST height H could be directly derived. Leaving aside the somewhat more complicated issue of the wind driven Ekman layer (Wunsch, 1992), the slope in the geoid can be translated into surface ocean circulation. In other words, altimetry in combination with a known geoid would let us see the ocean surface flow. For oceanography this would be a milestone.

However, there is more to it. To the present day, ocean circulation is studied on basis of hydrographic measurements (salinity, temperature, pressure, depth) along selected ocean sections. In order to derive circulation from the data – employing equations of motion – an assumption on a level of motion at some depth has to be introduced (Pond and Pickard, 1983). As one knows, such a layer is purely hypothetical and consequently systematic errors enter the calculations. Altimetry in combination with the geoid removes this uncertainty. It provides the needed boundary constraint for the equations of motion. Thus the gravity field information also permits to see the deep ocean in the proper way.

Circulation is the key to many ocean transport processes, whether it is heat transport, transport of plankton or polluted water. Heat exchange between ocean and atmosphere is probably the main uncertainty in a better understanding of climate changes. Oceans are considered a main buffer of atmosphere heat but the correct mechanism of heat transport in the oceans and exchange of heat between water and air is not well understood.

With these examples we tried to illustrate the need for an improved gravity field knowledge throughout earth sciences. Satellite gradiometry could have a substantial impact in this respect.

2.3 Aristoteles

At the time of writing the European Space Agency (ESA) is planning a solid earth mission to be launched in the late nineties. The main purpose of this project (called

Aristoteles) will be the improvement of our knowledge of the earth's gravitational field through gradiometer measurements. A second objective will be a global analysis of the near–earth magnetic field since, besides gravity gradiometry, the satellite will be equipped with a magnetometer. Furthermore, our knowledge of various geophysical phenomena such as earth rotation and pole irregularities can be enriched by means of precise point positioning measurements from space to which extent a GPS receiver will be on board the satellite (Dornier, 1989).

The Global Positioning System (GPS) consists of 24 (21 + 3 spare) satellites, orbiting the earth at an altitude of 20,240 km with an orbital period of 12 h. They are evenly distributed over 6 orbit planes with an inclination of 55°. The satellites transmit three different kinds of pseudorandom noise codes at two carrier frequencies. With a GPS receiver on board, tracking at any time four of the 24 GPS satellites, the position of the receiver antenna can be determined by pseudo–ranging, and if, in addition, differential carrier phase measurements relative to a network of ground stations are applied, orbit reconstitution at centimeter level is feasible.

A full tensor gradiometer would consist of eight ultra sensitive three–axis accelerometers placed on the corners of a cube (Balmino et al., 1985). However, due to the heavy affection of the satellite by air drag in the along–track direction (accelerations which are more than 10 times as large as the differential gravitational accelerations) the non drag–free Aristoteles configuration will consist of only four accelerometers mounted in the corners of a plate perpendicular to the satellite's velocity vector (Dornier, 1989). This instrument is called GRADIO. The four accelerometers will be very sensitive in radial and cross–track direction whereas the along–track component will be measured with lower accuracy.

The accelerometers which will be used consist of a cubic (or parallelepipedic) proof–mass which is kept in a fixed position electrostatically by means of electrodes arranged around it (GRADIO, 1989). The force necessary to maintain the proof–mass at the center of the accelerometer is measured by means of the output of the electrodes. The differential measurement of these forces between two accelerometers is a measure for the gravitational gradient (see section 2.1).

Perturbing forces resulting from air drag, solar radiation pressure, etc., are eliminated in the measurement process by differencing the accelerometer outputs (as we have seen in section 2.1, so–called common mode rejection). However, this requires a very good linearity of the accelerometers, a low coupling between the sensitivity axes and a good matching of the scale factors and of the alignments of the instrument axes (Dornier, 1989). To this extent a calibration device is provided, situated in the center of the gradiometer plate. The calibration device is furthermore needed to scale the accelerometer outputs to proper gravitational units.

The material out of which the satellite is built up also causes a gravitational acceleration which is measured by the gradiometer (so–called self gravitation). Whereas the signal coming from the rigid satellite parts (housing, electronics, solar panels) is constant and can be computed relatively simple, the part coming from the fuel in the hydrazine tanks causes a problem, because it has time varying components. Firstly,

there is the fuel consumption necessary to maintain the satellite's low orbit. It causes a gradual decrease of the mass and therefore a change in the self gravitation. This problem is accounted for by a symmetrical organization and emptying of the fuel tanks, which furthermore have a spherical shape and are equipped with elastomeric bladders to keep the fuel centered. In this case the change in self gravitation can be computed easily.

Secondly, and a more critical problem, is the sloshing of the fuel in the tanks. The tanks should be of such a size as to make sure that sloshing mode is at a frequency out of the measurement band of the GRADIO instrument (Dornier, ibid.). This measurement bandwidth, for which an 0.01 $E/\sqrt{Hz}$ white noise error spectrum is to be expected, is between 0.005 – 0.125 Hz allowing only a good recovery of the potential coefficients above degree 27. In (Schrama, 1990) it is shown that the on–board GPS receiver can be used in combination with GRADIO to obtain long wavelength (below degree 27) gravitational information from the Aristoteles mission. However, combination of gradiometric and GPS measurements will not be considered in this thesis.

In table 2.1 an overview of error sources in satellite gradiometry is given. For a detailed explanation and discussion of these errors see e.g. (Touboul et al., 1991), (Paik and Richard, 1986), (Schrama, 1990) and (Rummel, 1989b), in which further references can be found.

The complete Aristoteles mission will last for more than four years, of which the gravity phase, in which the gradiometer measurements with the GRADIO instrument will be carried out, has a duration of about 6 months. This 6 months gravity mission will be performed at a low altitude of 200 km, the orbit being a dawn–dusk orbit with inclination 95°.3. Because of the relatively high air density at 200 km, drag will cause an altitude decay of 400 m per revolution, so that regular orbit maintenance maneuvers are necessary to keep the satellite in a band of $\pm$ 3 km around the nominal altitude of 200 km. After this 6 months period the satellite is planned to fly for another two weeks in a near polar orbit (inclination 92°.3). This decreases the influence of the relatively large polar gaps of the 6 months period.

Table 2.1 *Error sources in satellite gradiometry*

instrument	geometrical	misorientation of accelerometer scale factor mismatch between sensitivity axes of accelerometer non–orthogonality of the sensitivity axes of accelerometer misalignment of accelerometers in gradiometer frame scale errors of instrument axes
	other	displacement of instrument from center of mass of the satellite non–linearity of accelerometers incorrect calibration non–mechanical instrument noise, due to temperature and electromagnetic fluctuations of the environment finite baseline of the instrument
satellite	attitude	unmodelled rotations orientation
	external forces	surface forces due to air drag, solar radiation pressure, etc. environmental disturbances such as vibrations, electromagnetic and thermal irregularities
	self gravitation	time varying components due to fuel consumption and sloshing resonating masses (e.g. solar arrays, antennas) reaction wheels noise
geodetic	anomalies	separation of gravitational and rotational parts integrated observables orbit errors orientation unknowns
	model	linearization error simplifications (truncation, symmetries)
	downward continuation	

3

The gradient tensor and its series representation in different coordinate systems

The first–order derivatives (gradient) of the gravitational potential V with respect to arbitrary coordinates x^r are $V_r = \frac{\partial V}{\partial x^r}$. If, for example, the system x^r is a cartesian coordinate system, the three components of V_r for $r = 1, 2, 3$ together form the acceleration vector or gravitational vector. The second–order derivatives of the potential are V_{rs}. With respect to a cartesian coordinate system (e.g. some instrument system with respect to which measurements are taken) they are the first–order derivatives of the acceleration vector, i.e. the gradients of the gravitation. This is the reason we call the V_{rs} the *gravity gradients* (or gravitational gradients[1]). It also is the reason for using the word *gradiometry*: "measurement of gravity gradients".

In the sense of tensor analysis the quantity V_{rs} is a tensor. It is sometimes called *gravity tensor*. The two indices r, s result from a differentiation process of the scalar quantity V, and in arbitrary coordinates (not necessarily linear coordinates) this should be the process of covariant differentiation, see appendix B. Only in cartesian coordinates the covariant derivative equals the usual partial derivative, because in those coordinates the Christoffel symbols vanish.

A few general remarks about V_{rs} can be made here. First, since the gravitational potential function V is harmonic outside all masses, it fulfils Laplace's equation, which in our arbitrary coordinates x^r is written as:

$$g^{rs} V_{rs} = 0 \tag{3.1}$$

where g^{rs} is the contravariant (or associated) metric tensor. In cartesian coordinates this equation is written out in the well known form $V_{xx} + V_{yy} + V_{zz} = 0$. We see that

[1] Gravity = gravitation + centrifugal acceleration. In many texts the terms gravity and gravitation are not well distinguished. If the difference is essential, it will be clear from the context.

the diagonal elements of the gradient tensor in cartesian coordinates are linearly dependent, leaving only two out of three independent components.

Furthermore, since the gravitational field is irrotational, it satisfies:

$$e^{rst}V_{st} = 0^r \tag{3.2}$$

which implies that V_{rs} is symmetric: $\{V_{rs}\} = \{V_{sr}\}$, leaving only three out of six independent off-diagonal components (e^{rst} is the three dimensional permutation symbol, which equals 1 if the value of the indices constitute an even permutation, -1 if the permutation is odd and 0 in other cases. 0^r is the null-tensor, but one usually writes 0, in which case, however, the index balance no longer holds.). According to both properties (eq. 3.1 and 3.2), out of the nine components of the gradient tensor only five independent components remain.

A last general remark concerns the tensor character of the gradient tensor. Being a tensor, V_{rs} transforms to some other coordinate system, e.g. x^R, as

$$V_{RS} = \frac{\partial x^r}{\partial x^R}\frac{\partial x^s}{\partial x^S}V_{rs} \ . \tag{3.3}$$

If one knows the coordinate transformation equations $x^r = x^r(x^R)$ from which the transformation matrix $\frac{\partial x^r}{\partial x^R}$ is computed these equations can easily be evaluated. This will be done for several coordinate systems in section 3.1.

We have seen in the previous chapter that, in principle, it is possible to measure the second-order derivatives of the earth's gravitational potential using a gradiometer in an earth orbiting satellite. Actually, the gradiometer will deliver the measurements in a cartesian coordinate system connected to the instrument. The orientation of this instrument system, or satellite system, will in general differ from that of a local orbital system[2] due to changes in the satellite's attitude. The local orbital system, denoted by x^i, will be orientated with the x-axis along track, the y-axis cross track and the z-axis outwards. We will at this stage assume, however, that either the differences between the instrument system and the local orbital system are negligible, or that it is possible to transform the measurements from the instrument system to the local orbital system, using an equation of the type 3.3. In this particular case, such transformation will consist of a simple rotation matrix between two cartesian coordinate systems having the same origin but a different orientation, and we will assume it to be known. Thus we consider it possible to obtain from the gradiometer the gravitational gradients V_{ij} in the local orbital system x^i.

Furthermore, the orientation of this local orbital system will, due to the inclination of the satellite's orbit, in general deviate from the orientation of a (commonly used) local, north-oriented coordinate system $x^{i'}$ with the x'-axis directed north, the y'-axis west and the z'-axis radially outwards. However, we will also assume that it is possible to transform the components of V_{ij} from the local orbital system

[2]For the definition of this system and other coordinate systems which will be used see appendix A.

to the local, north-oriented system by means of a known transformation, again using an equation of the type 3.3, in this case written as:

$$V_{i'j'} = \frac{\partial x^i}{\partial x^{i'}} \frac{\partial x^j}{\partial x^{j'}} V_{ij} \tag{3.4}$$

where the $\frac{\partial x^i}{\partial x^{i'}}$ is a known rotation matrix, cf. (Rummel and Colombo, 1985). If, in an ideal case, all nine components of the symmetric gradient tensor are measured, it is, in principle, possible to solve the attitude of the instrument from the measurements, at least if some additional information from star trackers is available (Rummel, 1985c). For the moment we will assume here that we are given the gravity gradients in either a local orbital system x^i or a local north-oriented system $x^{i'}$.

From the available gradient measurements we like to obtain information about the earth's gravitational field. This information is usually given in terms of a set of potential coefficients, cf. section 2.2. One therefore needs a relation between these spectral coefficients, often denoted $\bar{C}_{lm}$ and $\bar{S}_{lm}$, and the observations V_{ij}. This relationship is obtained by expanding the gravitational potential V into a series of spherical harmonics as function of the geocentric polar coordinates $x^A = (r, \theta, \lambda)$. This will be done in section 3.2. In this section also a series expansion of the potential as function of the orbital coordinates $x^a = (r, \omega_e, \omega_o)$ or $x^{a'} = (r, \omega_o, I)$ (see appendix A for their definition) will be given, which appears to be better suitable for problems involving satellite observations.

Given V as function of e.g. x^a, one may compute the second-order derivatives V_{ab} (using *covariant* differentiation in this case because the x^a are curvilinear coordinates). By means of an equation of the type 3.3, in this case

$$V_{ij} = \frac{\partial x^b}{\partial x^i} \frac{\partial x^b}{\partial x^j} V_{ab} \ , \tag{3.5}$$

one obtains the desired relationship between potential coefficients, contained in the V_{ab}, and measurements V_{ij}.

In order to gain some insight in the signals to be expected from a gradiometer mission, the last section of this chapter, section 3.3, deals with the synthesis problem, i.e. how to compute from some a-priori given set of potential coefficients a set of (simulated) gradiometer data. This section furthermore shows a simple analysis strategy, i.e. the inverse problem of recovering the potential coefficients again from the simulated gradient data.

3.1 Potential derivatives in different coordinate systems

In the following section 3.2 we will be given two series expansions of the gravitational potential, one as function of the geocentric polar coordinates x^A and one as function of the orbital coordinates x^a or $x^{a'}$ (respectively equations 3.14 and 3.17). Both series

are given in terms of the same potential coefficient set $\bar{C}_{lm}, \bar{S}_{lm}$. Taking two times the covariant derivative of these series yields corresponding series expansions for the gravitational gradients in curvilinear coordinates, respectively V_{AB}, V_{ab} or $V_{a'b'}$. They have to be related to the observed gradients, expressed in local coordinates x^i or $x^{i'}$, respectively V_{ij} and $V_{i'j'}$. So what we need are transformation equations for second–order potential derivatives.

In this section we derive these transformations, which are all of the general type

$$V_{RS} = \frac{\partial x^r}{\partial x^R} \frac{\partial x^s}{\partial x^S} V_{rs} \ , \tag{3.6}$$

where in our case we take for the x^r coordinates either x^A, x^a or $x^{a'}$ and for the x^R we take x^i or $x^{i'}$. For the evaluation of the V_{rs} we need the metrical tensor and the Christoffel symbols for the respective coordinate systems. They are given in appendix A. In section 3.1.1 the algorithm to derive the gradient transformation equations is explained. The subsequent section 3.1.2 gives a listing of several explicit transformations. Compare e.g. (Reed, 1973) or (Tscherning, 1976) were some of these transformation formulae are also derived.

3.1.1 Algorithm

The algorithm for computing the transformation equations for the second–order potential derivatives may serve as an example for the use of tensor analysis and index notation (see e.g. sections B.2 and B.3). For the definition of the coordinate systems used in this section, we refer to appendix A. In the same appendix the elements of the metrical tensor and the Christoffel symbols for those coordinate systems are listed.

We assume we have the gravitational potential V given as function of the coordinates x^a. We could equally well assume here that V is given as a function of $x^A, x^{a'}, x^{A'}, x^{\bar{a}}$ or $x^{\bar{A}}$, but we take x^a as example. Since the equations are all tensor equations they also hold for all other coordinate systems. Given V, we may compute by partial differentiation the first–order derivatives $V_{,a}$, which equal the covariant derivatives $V_{;a}$ and which will be simply denoted by V_a, so $V_a = V_{,a} = V_{;a}$.

The second–order derivatives have to be computed by covariant differentiation, since partial differentiation of a tensor with rank ≥ 1 in general does not lead to a tensor. This yields:

$$V_{a;b} = V_{a,b} - \Gamma^c_{ab} V_c$$

which we will simply denote by V_{ab}, so $V_{ab} = V_{;ab} = V_{a;b}$. We see that, in order to evaluate V_{ab}, we need the Christoffel symbols Γ^c_{ab}. They are computed using

$$\Gamma^c_{ab} = \frac{1}{2} g^{cd} (g_{ad,b} + g_{bd,a} - g_{ab,d}) \ .$$

For this purpose we need the metrical tensor g_{ab}. It is calculated, according to its

transformation rule, from g_{IJ} in geocentric coordinates, i.e.

$$g_{ab} = \frac{\partial x^I}{\partial x^a}\frac{\partial x^J}{\partial x^b} g_{IJ} \ .$$

Thereby g_{IJ} equals the unit matrix and the transformation matrix $\frac{\partial x^I}{\partial x^a}$ is computed from the coordinate transformation $x^I = x^I(x^a)$. Thus the coordinate transformation has to be known. In the present example this transformation equals equation A.6. The associated metric tensor g^{ab} is computed as the inverse of g_{ab}. If it is hard to invert g_{ab} one can try to compute first the inverse transformation matrix $\frac{\partial x^a}{\partial x^I}$ and calculate g^{ab} from

$$g^{ab} = \frac{\partial x^a}{\partial x^J}\frac{\partial x^b}{\partial x^K} g^{JK}$$

in which g^{IJ} is again the unit matrix.

The V_{ab} can now be computed. Since V_{ab} is a tensor, its transformation equation to another coordinate system has the form

$$V_{JK} = \frac{\partial x^a}{\partial x^J}\frac{\partial x^b}{\partial x^K} V_{ab} \ .$$

We already know $\frac{\partial x^a}{\partial x^I}$, so with this transformation equation we easily find the second–order potential derivatives with respect to geocentric cartesian coordinates as function of the first– (!) and second–order potential derivatives with respect to the I–orbital coordinates x^a.

Transformation to a local cartesian coordinate system, for example the local orbital system x^i, follows analogously:

$$V_{jk} = \frac{\partial x^a}{\partial x^j}\frac{\partial x^b}{\partial x^k} V_{ab} \ ,$$

where the transformation matrix $\frac{\partial x^a}{\partial x^i}$ is most easily computed using the chain rule

$$\frac{\partial x^a}{\partial x^j} = \frac{\partial x^a}{\partial x^J}\frac{\partial x^J}{\partial x^j}$$

in which $\frac{\partial x^a}{\partial x^I}$ is already known and $\frac{\partial x^I}{\partial x^i}$ has to be computed from the coordinate transformation $x^I = x^I(x^i)$. We therefore have to specify the latter transformation, or directly the matrix $\frac{\partial x^I}{\partial x^i}$, which in the present example equals the matrix in equation A.3.

Moreover, the transformation equations for the first–order derivatives are simply

$$V_J = \frac{\partial x^a}{\partial x^J} V_a \tag{3.7}$$

$$V_j = \frac{\partial x^a}{\partial x^J}\frac{\partial x^J}{\partial x^j} V_a \ . \tag{3.8}$$

If, during the computations above (either numerically or symbolically), use is made of pre–programmed subroutines for matrix manipulations, care has to be taken with the use of the transpose, inverse and original form of the matrices, especially of the transformation matrices $\frac{\partial x^a}{\partial x^I}$ etc. Strictly following the conventions from section B.1.3, this may not lead to any problems, but it is safer to convert the equations directly into nested DO–loops, the inner loops representing the summations over the dummy indices and the outer loops representing the repetition of the computations for all the values of the free indices. Both methods will be illustrated below.

Let us denote all the transformation matrices with the kernel letter X, i.e. $X^I{}_a = \frac{\partial x^I}{\partial x^a}, X^I{}_i = \frac{\partial x^I}{\partial x^i}$, etc. Furthermore we choose the following transition from index to matrix notation:

$$
\begin{array}{lcllcl}
V_{ab} & \rightarrow & V & , & V_a & \rightarrow & g \\
V_{IJ} & \rightarrow & W & , & V_I & \rightarrow & a \\
V_{ij} & \rightarrow & U & , & V_i & \rightarrow & \gamma \\
X^I{}_a & \rightarrow & X & , & x^a & \rightarrow & r \\
X^I{}_i & \rightarrow & R & , & x^I & \rightarrow & l \\
X^i{}_a & \rightarrow & T & , & x^i & \rightarrow & u \\
g_{ab} & \rightarrow & G & , & g_{IJ} & \rightarrow & I\,.
\end{array}
$$

The transformation equations from this section then become, in order of appearance:

$$
\begin{array}{lcll}
g_{ab} = X_a{}^I g_{IJ} X^J{}_b & \rightarrow & G & = X^T I X \\
g^{ab} = X^a{}_J g^{JK} X_K{}^b & \rightarrow & G^{-1} & = X^{-1} I X^{-T} \\
V_{JK} = X_J{}^a V_{ab} X^b{}_K & \rightarrow & W & = X^{-T} V X^{-1} \\
V_{jk} = X_j{}^a V_{ab} X^b{}_k & \rightarrow & U & = T^{-T} V T^{-1} \\
X^a{}_j = X^a{}_J X^J{}_j & \rightarrow & T^{-1} & = X^{-1} R \\
V_J = X_J{}^a V_a & \rightarrow & a & = X^{-T} g \\
V_j = X_j{}^a V_a & \rightarrow & \gamma & = T^{-T} g
\end{array}
$$

with the transformation matrices defined through the coordinate transformation equations:

$$
\begin{aligned}
x^I = x^I(x^a) \quad &\rightarrow \quad l = Xr \\
x^I = x^I(x^i) \quad &\rightarrow \quad l = Ru \\
x^i = x^i(x^a) \quad &\rightarrow \quad u = Tr \ .
\end{aligned}
$$

As an example of how a tensor equation can easily be converted into programmable code, consider the equation for the Christoffel symbols

$$\Gamma^c_{ab} = \frac{1}{2} g^{cd} (g_{ad,b} + g_{bd,a} - g_{ab,d}) \ .$$

With the arrays `gamma(a,b,c)`, `gi(a,b)` and `dg(a,b,c)` representing respectively Γ^c_{ab}, g^{ab} and $g_{ab,c}$ the loop becomes:

```
do a = 1,3
   do b = 1,3
      do c = 1,3
         gamma(a,b,c) = 0
         do d = 1,3
            gamma(a,b,c) = gamma(a,b,c) + 0.5 * gi(c,d) *
                           (dg(a,d,b) + dg(b,d,a) - dg(a,b,d))
         end do
      end do
   end do
end do
```

The algorithm sketched in this section is shown in table 3.1 as flow chart.

Table 3.1 *Computation scheme for transformation of potential derivatives in arbitrary coordinate systems.*

input:	(1)	$V = V(x^a)$	
	(2)	$x^I = x^I(x^a)$	
	(3)	$x^I = x^I(x^i)$	
	(4)	$\{g_{IJ}\} = \{g^{IJ}\} = \{\delta^I_J\}$	
output:	(5)	$X^I{}_a$	from (2)
	(6)	$X^a{}_I = (X^I{}_a)^{-1}$	from (5)
	(7)	$g_{ab} = X_a{}^I X^J{}_b g_{IJ}$	from (4), (5)
	(8)	$g^{ab} = (g_{ab})^{-1}$	from (7)
	(9)	$g_{ab,c}$	from (7)
	(10)	Γ^c_{ab}	from (8), (9)
	(11)	$V_a = V_{,a}$	from (1)
	(12)	$V_I = X_I{}^a V_a$	from (6), (11)
	(13)	$V_{a,b}$	from (11)
	(14)	$V_{ab} = V_{a;b}$	from (10), (11), (13)
	(15)	$V_{IJ} = X_I{}^a X^b{}_J V_{ab}$	from (6), (14)
	(16)	$X^I{}_i$	from (3)
	(17)	$X^a{}_i$	from (6), (16)
	(18)	$V_i = X_i{}^a V_a$	from (11), (17)
	(19)	$V_{ij} = X_i{}^a X^b{}_j V_{ab}$	from (14), (17)

3.1.2 Transformation formulae for the potential derivatives

We assume the gravitational potential V to be given as function of one of the following coordinate sets: $x^A = (r, \theta, \lambda)$, $x^a = (r, \omega_o, \omega_e)$, $x^{a'} = (r, \omega_o, I)$ or $x^{A'} = (r, \phi, \omega_o)$ (see appendix A). Then we can compute the first- and second-order potential derivatives with respect to these coordinates, respectively $V_A, V_a, V_{a'}, V_{A'}$ and $V_{AB}, V_{ab}, V_{a'b'}, V_{A'B'}$. What we need are the potential derivatives with respect to either the local north-oriented coordinate system $x^{i'}$ or the local orbital coordinate system x^i, i.e. $V_{i'j'}$ or V_{ij}. They are obtained from the derivatives with respect to curvilinear coordinates by means of transformation equations of the type 3.7 and 3.6, and are given in the boxes in the remaining of this section. Notation is abbreviated to what is commonly used: V_r means $\frac{\partial V}{\partial r} = \frac{\partial V}{\partial x^{A=1}}$, etc. For reference, also the transformations of the derivatives in ellipsoidal and geodetic coordinates are given.

$\boxed{V_i = \frac{\partial x^{A'}}{\partial x^i} V_{A'}}$

$$
\begin{aligned}
V_x &= \frac{1}{r\cos\phi} V_{\omega_o} \\
V_y &= \frac{1}{r} V_\phi \\
V_z &= V_r
\end{aligned}
$$

$\boxed{V_{ij} = \frac{\partial x^{A'}}{\partial x^i} \frac{\partial x^{B'}}{\partial x^j} V_{A'B'}}$

$$
\begin{aligned}
V_{xx} &= \frac{1}{r} V_r - \frac{\tan\phi}{r^2} V_\phi + \frac{1}{r^2\cos^2\phi} V_{\omega_o\omega_o} \\
V_{xy} &= \frac{1}{r^2\cos\phi} V_{\phi\omega_o} + \frac{\sin\phi}{r^2\cos^2\phi} V_{\omega_o} \\
V_{xz} &= -\frac{1}{r^2\cos\phi} V_{\omega_o} + \frac{1}{r\cos\phi} V_{r\omega_o} \\
V_{yy} &= \frac{1}{r} V_r + \frac{1}{r^2} V_{\phi\phi} \\
V_{yz} &= -\frac{1}{r^2} V_\phi + \frac{1}{r} V_{r\phi} \\
V_{zz} &= V_{rr}
\end{aligned}
\tag{3.9}
$$

If we put $\phi = 0$ in equations 3.9 we obtain:

$$
\begin{aligned}
V_{xx} &= \frac{1}{r} V_r + \frac{1}{r^2} V_{\omega_o\omega_o} & V_{xy} &= \frac{1}{r^2} V_{\phi\omega_o} \\
V_{xz} &= -\frac{1}{r^2} V_{\omega_o} + \frac{1}{r} V_{r\omega_o} & V_{yy} &= \frac{1}{r} V_r + \frac{1}{r^2} V_{\phi\phi} \\
V_{yz} &= -\frac{1}{r^2} V_\phi + \frac{1}{r} V_{r\phi} & V_{zz} &= V_{rr}
\end{aligned}
\tag{3.10}
$$

Equations 3.10 will be used in chapter 4. They are obviously *only valid* on the equator of the coordinate system which, in our case, is the *orbit of the satellite*.

$$\boxed{V_i = \frac{\partial x^{a'}}{\partial x^i} V_{a'}}$$

$$V_x = \tfrac{1}{r} V_{\omega_o}$$
$$V_y = \frac{1}{r \sin \omega_o} V_I$$
$$V_z = V_r$$

$$\boxed{V_{ij} = \frac{\partial x^{a'}}{\partial x^i} \frac{\partial x^{b'}}{\partial x^j} V_{a'b'}}$$

$$\begin{aligned}
V_{xx} &= \frac{1}{r} V_r + \frac{1}{r^2} V_{\omega_o \omega_o} \\
V_{xy} &= \frac{1}{r^2 \sin \omega_o} V_{I\omega_o} - \frac{\cos \omega_o}{r^2 \sin^2 \omega_o} V_I \\
V_{xz} &= \frac{1}{r} V_{r\omega_o} - \frac{1}{r^2} V_{\omega_o} \\
V_{yy} &= \frac{1}{r} V_r + \frac{1}{r^2 \tan \omega_o} V_{\omega_o} + \frac{1}{r^2 \sin^2 \omega_o} V_{II} \\
V_{yz} &= \frac{1}{r \sin \omega_o} V_{rI} - \frac{1}{r^2 \sin \omega_o} V_I \\
V_{zz} &= V_{rr}
\end{aligned} \tag{3.11}$$

$$\boxed{V_i = \frac{\partial x^a}{\partial x^i} V_a}$$

$$V_x = \tfrac{1}{r} V_{\omega_o}$$
$$V_y = \frac{\cot I}{r \cos \omega_o} V_{\omega_o} - \frac{1}{r \sin I \cos \omega_o} V_{\omega_e}$$
$$V_z = V_r$$

$$\boxed{V_{ij} = \frac{\partial x^a}{\partial x^i} \frac{\partial x^b}{\partial x^j} V_{ab}}$$

$$\begin{aligned}
V_{xx} &= \frac{1}{r} V_r + \frac{1}{r^2} V_{\omega_o \omega_o} \\
V_{xy} &= \frac{1}{r^2 \sin I \cos^2 \omega_o} (\cos I \sin \omega_o V_{\omega_o} - \cos I \cos \omega_o V_{\omega_o \omega_o} \\
&\quad - \cos \omega_o V_{\omega_o \omega_e} - \sin \omega_o V_{\omega_e}) \\
V_{xz} &= \frac{1}{r} V_{r\omega_o} - \frac{1}{r^2} V_{\omega_o} \\
V_{yy} &= \frac{1}{r} V_r + \frac{1}{r^2 \sin^2 I \cos^3 \omega_o} \Big(\sin \omega_o (\sin^2 I \sin^2 \omega_o - 2 \sin^2 I + 1) V_{\omega_o} \\
&\quad + \cos^2 I \cos \omega_o V_{\omega_o \omega_o} - 2 \cos I \cos \omega_o V_{\omega_o \omega_e} + \cos \omega_o V_{\omega_e \omega_e} - \cos I \sin \omega_o V_{\omega_e} \Big) \\
V_{yz} &= \frac{1}{r^2 \sin I \cos \omega_o} (r \cos I V_{r\omega_o} - \cos I V_{\omega_o} - r V_{r\omega_e} + V_{\omega_e}) \\
V_{zz} &= V_{rr}
\end{aligned} \tag{3.12}$$

$V_{\bar{i}} = \frac{\partial x^A}{\partial x^{\bar{i}}} V_{\bar{A}}$

$$V_{\bar{x}} = \frac{1}{h+M} V_\varphi$$
$$V_{\bar{y}} = -\frac{1}{(h+N)\cos\varphi} V_\lambda$$
$$V_{\bar{z}} = V_h$$

$V_{\bar{i}\bar{j}} = \frac{\partial x^A}{\partial x^{\bar{i}}} \frac{\partial x^B}{\partial x^{\bar{j}}} V_{\bar{A}\bar{B}}$

$$V_{\bar{x}\bar{x}} = \frac{1}{h+M} V_h + \frac{1}{(h+M)^2} V_{\varphi\varphi} - \frac{3e^2 M^2 \sin\varphi\cos\varphi}{N(1-e^2)(h+M)^3} V_\varphi$$
$$V_{\bar{x}\bar{y}} = \frac{-1}{(h+M)(h+N)\cos\varphi} V_{\varphi\lambda} - \frac{\tan\varphi}{\cos\varphi(h+N)^2} V_\lambda$$
$$V_{\bar{x}\bar{z}} = \frac{1}{h+M} V_{h\varphi} - \frac{1}{(h+M)^2} V_\varphi$$
$$V_{\bar{y}\bar{y}} = \frac{1}{h+N} V_h + \frac{1}{(h+N)^2\cos^2\varphi} V_{\lambda\lambda} - \frac{\tan\varphi}{(h+M)(h+N)} V_\varphi$$
$$V_{\bar{y}\bar{z}} = \frac{1}{(h+N)^2\cos\varphi} V_\lambda - \frac{1}{(h+N)\cos\varphi} V_{h\lambda}$$
$$V_{\bar{z}\bar{z}} = V_{hh}$$

$V_{\bar{i}} = \frac{\partial x^a}{\partial x^{\bar{i}}} V_{\bar{a}}$

$$V_{\bar{x}} = \frac{1}{L} V_\beta$$
$$V_{\bar{y}} = -\frac{1}{v\cos\beta} V_\lambda$$
$$V_{\bar{z}} = \frac{v}{L} V_u$$

$V_{\bar{i}\bar{j}} = \frac{\partial x^a}{\partial x^{\bar{i}}} \frac{\partial x^b}{\partial x^{\bar{j}}} V_{\bar{a}\bar{b}}$

$$V_{\bar{x}\bar{x}} = \frac{uv^2}{L^4} V_u + \frac{1}{L^2} V_{\beta\beta} - \frac{E^2 \sin\beta\cos\beta}{L^4} V_\beta$$
$$V_{\bar{x}\bar{y}} = \frac{-1}{vL\cos\beta} V_{\beta\lambda} - \frac{\tan\beta}{vL\cos\beta} V_\lambda$$
$$V_{\bar{x}\bar{z}} = \frac{v}{L^2} V_{u\beta} - \frac{uv}{L^4} V_\beta - \frac{vE^2 \sin\beta\cos\beta}{L^4} V_u$$
$$V_{\bar{y}\bar{y}} = \frac{u}{L^2} V_u + \frac{1}{v^2\cos^2\beta} V_{\lambda\lambda} - \frac{\tan\beta}{L^2} V_\beta$$
$$V_{\bar{y}\bar{z}} = \frac{u}{Lv^2\cos\beta} V_\lambda - \frac{1}{L\cos\beta} V_{u\lambda}$$
$$V_{\bar{z}\bar{z}} = \frac{v^2}{L^2} V_{uu} - \frac{uE^2\cos^2\beta}{L^4} V_u + \frac{E^2\sin\beta\cos\beta}{L^4} V_\beta$$

$$\boxed{V_{i'} = \frac{\partial x^A}{\partial x^{i'}} V_A}$$

$$\begin{aligned} V_{x'} &= -\tfrac{1}{r} V_\theta \\ V_{y'} &= -\frac{1}{r \sin\theta} V_\lambda \\ V_{z'} &= V_r \end{aligned}$$

$$\boxed{V_{i'j'} = \frac{\partial x^A}{\partial x^{i'}} \frac{\partial x^B}{\partial x^{j'}} V_{AB}}$$

$$\begin{aligned} V_{x'x'} &= \frac{1}{r} V_r + \frac{1}{r^2} V_{\theta\theta} \\ V_{x'y'} &= \frac{1}{r^2 \sin\theta} V_{\theta\lambda} - \frac{\cos\theta}{r^2 \sin^2\theta} V_\lambda \\ V_{x'z'} &= \frac{1}{r^2} V_\theta - \frac{1}{r} V_{r\theta} \\ V_{y'y'} &= \frac{1}{r} V_r + \frac{1}{r^2 \tan\theta} V_\theta + \frac{1}{r^2 \sin^2\theta} V_{\lambda\lambda} \\ V_{y'z'} &= \frac{1}{r^2 \sin\theta} V_\lambda - \frac{1}{r \sin\theta} V_{r\lambda} \\ V_{z'z'} &= V_{rr} \end{aligned} \tag{3.13}$$

3.2 Series expansion of the potential and its derivatives

3.2.1 Spherical harmonics

The gravitational potential V, being a solution of Laplace's equation, is expanded into a series of spherical harmonics as function of the geocentric polar coordinates $x^A = (r, \theta, \lambda)$ as:

$$V(r,\theta,\lambda) = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^{l} \left[\bar{C}_{lm} \cos m\lambda + \bar{S}_{lm} \sin m\lambda\right] \bar{P}_{lm}(\cos\theta) \tag{3.14}$$

where

GM	gravitational constant times mass of the earth
R	reference radius
$\bar{C}_{lm}, \bar{S}_{lm}$	normalized potential coefficients of degree l and order m
l, m	degree and order
$\bar{P}_{lm}(\cos\theta)$	normalized Legendre functions
r, θ, λ	geocentric polar coordinates (θ is co-latitude).

For computational purposes it is convenient to interchange the summation over l

and m leading to the Fourier series, cf. (Colombo, 1981)

$$V(r,\theta,\lambda) = \sum_{m=0}^{\infty} [A_m(r,\theta)\cos m\lambda + B_m(r,\theta)\sin m\lambda] \tag{3.15}$$

with coefficients

$$\left.\begin{matrix} A_m(r,\theta) \\ B_m(r,\theta) \end{matrix}\right\} = \sum_{l=m}^{\infty} \left\{\begin{matrix} K_{lm}^A(r,\theta) \\ K_{lm}^B(r,\theta) \end{matrix}\right. \tag{3.16}$$

$$\left.\begin{matrix} K_{lm}^A(r,\theta) \\ K_{lm}^B(r,\theta) \end{matrix}\right\} = H_{lm}(r,\theta) \left\{\begin{matrix} \bar{C}_{lm} \\ \bar{S}_{lm} \end{matrix}\right.$$

$$H_{lm}(r,\theta) = u_l(r)\bar{P}_{lm}(\cos\theta)$$

$$u_l(r) = \frac{GM}{R}\left(\frac{R}{r}\right)^{l+1} .$$

We saw that, in order to compute the potential derivatives in the local north–oriented system $x^{i'}$, we need the derivatives V_A and V_{AB} of the potential with respect to the polar coordinates x^A. These can be derived from equation 3.14 or 3.15 by means of simple partial differentiation. The derivatives, computed in this way, can also be expressed in series analogous to 3.15, with appropriate expressions for the factors K_{lm}^A, K_{lm}^B and H_{lm}. In table 3.2 one finds these factors for all the first– and second–order derivatives of V with respect to r,θ,λ. In this table we abbreviated $\bar{P}'_{lm} = \partial\bar{P}_{lm}/\partial\theta$ and $\bar{P}''_{lm} = \partial^2\bar{P}_{lm}/\partial\theta^2$. The local derivatives $V_{i'j'}$ are expressed in the same way, the expressions for K_{lm}^A, K_{lm}^B and H_{lm} now to be taken from table 3.3. They are found by using equations 3.13.

Already at this point one might find out, by taking a glance at this table, why satellite gradiometry is especially suitable for the determination of higher degrees and orders of the gravitational potential. On the one hand, one notices the factor $(R/r)^{l+1}$, which is present in all six gradients (since $\Gamma_l = u_l/r^2 = (GM/R/r^2)\cdot(R/r)^{l+1}$). This factor causes an attenuation effect with height for the gravitational potential and related quantities. It means that, since R/r is smaller than 1, the power contents of the potential signal decreases with increasing altitude above R. Due to the power of $l+1$ this decrease is stronger for higher degrees l, making them harder to detect if one goes further away from the earth. The attenuation effect is inherent to all spaceborne gravity determination techniques.

Table 3.2 *Potential derivatives w.r.t.* (r, θ, λ)

differentiation w.r.t.	$H_{lm}^{(\cdot\cdot)}$	K_{lm}^{A}	K_{lm}^{B}
r	$\frac{-(l+1)}{r} u_l \bar{P}_{lm}$	$H_{lm}^{(r)} \bar{C}_{lm}$	$H_{lm}^{(r)} \bar{S}_{lm}$
θ	$u_l \bar{P}'_{lm}$	$H_{lm}^{(\theta)} \bar{C}_{lm}$	$H_{lm}^{(\theta)} \bar{S}_{lm}$
λ	$m u_l \bar{P}_{lm}$	$H_{lm}^{(\lambda)} \bar{S}_{lm}$	$-H_{lm}^{(\lambda)} \bar{C}_{lm}$
rr	$\frac{(l+1)(l+2)}{r^2} u_l \bar{P}_{lm}$	$H_{lm}^{(rr)} \bar{C}_{lm}$	$H_{lm}^{(rr)} \bar{S}_{lm}$
$r\theta$	$\frac{-(l+1)}{r} u_l \bar{P}'_{lm}$	$H_{lm}^{(r\theta)} \bar{C}_{lm}$	$H_{lm}^{(r\theta)} \bar{S}_{lm}$
$r\lambda$	$\frac{-m(l+1)}{r} u_l \bar{P}_{lm}$	$H_{lm}^{(r\lambda)} \bar{S}_{lm}$	$-H_{lm}^{(r\lambda)} \bar{C}_{lm}$
$\theta\theta$	$u_l \bar{P}''_{lm}$	$H_{lm}^{(\theta\theta)} \bar{C}_{lm}$	$H_{lm}^{(\theta\theta)} \bar{S}_{lm}$
$\theta\lambda$	$m u_l \bar{P}'_{lm}$	$H_{lm}^{(\theta\lambda)} \bar{S}_{lm}$	$-H_{lm}^{(\theta\lambda)} \bar{C}_{lm}$
$\lambda\lambda$	$-m^2 u_l \bar{P}_{lm}$	$H_{lm}^{(\lambda\lambda)} \bar{C}_{lm}$	$H_{lm}^{(\lambda\lambda)} \bar{S}_{lm}$

On the other hand, in the case of the second-order derivatives, this attenuation effect is to some extent compensated by multiplication factors like $(l+1)(l+2), (l+2), (l+1)$ and m, which tend to increase the power contents for higher degrees. This makes satellite gradiometry (the technique of observing gravity gradients at satellite altitude) an attractive candidate for gravitational recovery with high precision and resolution compared to other techniques like orbit determination.

Table 3.3 *Potential derivatives with respect to local north-oriented coordinates* (x', y', z'). $\Gamma_l = \frac{u_l}{r^2}$

	$H_{lm}^{(\cdot\cdot)}$	K_{lm}^{A}	K_{lm}^{B}
$x'x'$	$\Gamma_l(\bar{P}''_{lm} - (l+1)\bar{P}_{lm})$	$H_{lm}^{(x'x')} \bar{C}_{lm}$	$H_{lm}^{(x'x')} \bar{S}_{lm}$
$x'y'$	$m \sin^{-1}\theta \Gamma_l(\bar{P}'_{lm} - \cot\theta \bar{P}_{lm})$	$H_{lm}^{(x'y')} \bar{S}_{lm}$	$-H_{lm}^{(x'y')} \bar{C}_{lm}$
$x'z'$	$(l+2)\Gamma_l \bar{P}'_{lm}$	$H_{lm}^{(x'z')} \bar{C}_{lm}$	$H_{lm}^{(x'z')} \bar{S}_{lm}$
$y'y'$	$\Gamma_l(\cot\theta \bar{P}'_{lm} - (l+1+m^2 \sin^{-2}\theta)\bar{P}_{lm})$	$H_{lm}^{(y'y')} \bar{C}_{lm}$	$H_{lm}^{(y'y')} \bar{S}_{lm}$
$y'z'$	$m(l+2)\sin^{-1}\theta \Gamma_l \bar{P}_{lm}$	$H_{lm}^{(y'z')} \bar{S}_{lm}$	$-H_{lm}^{(y'z')} \bar{C}_{lm}$
$z'z'$	$(l+1)(l+2)\Gamma_l \bar{P}_{lm}$	$H_{lm}^{(z'z')} \bar{C}_{lm}$	$H_{lm}^{(z'z')} \bar{S}_{lm}$

3.2.2 Orbital coordinates

For the description of the motion of planets around the sun or satellites around the earth the orbital elements $a, e, I, \Omega, \omega, M$ can be employed, which are respectively the semi–major axis of the elliptical orbit, eccentricity, inclination, right ascension of the ascending node, argument of perigee and mean anomaly (see figure A.1). It is convenient to have also an expression of the gravitational potential V in terms of these orbital elements. Upon keeping the harmonic coefficients $\bar{C}_{lm}$ and $\bar{S}_{lm}$ from eq. 3.14 as coefficients in the potential expansion, but changing from r, θ, λ to r, I, ω, Ω, M (we assume the orbit is (nearly) circular, so $e \approx 0$) we obtain:

$$V = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^{l} \sum_{p=0}^{l} \bar{F}_{lmp}(I) \cdot$$

$$\cdot \left\{ \begin{bmatrix} \bar{C}_{lm} \\ -\bar{S}_{lm} \end{bmatrix}_{l-m:\text{odd}}^{l-m:\text{even}} \cos\psi_{lmp} + \begin{bmatrix} \bar{S}_{lm} \\ \bar{C}_{lm} \end{bmatrix}_{l-m:\text{odd}}^{l-m:\text{even}} \sin\psi_{lmp} \right\} \qquad (3.17)$$

where

$\bar{F}_{lmp}(I)$ normalized inclination functions

$\psi_{lmp} = (l - 2p)\,\omega_o + m\,\omega_e$

$\omega_o = \omega + M$ subscript "o" referring to "orbit"

$\omega_e = \Omega - \theta_G$ subscript "e" referring to "earth"

} see (Schrama, 1989)

θ_G earth's argument of longitude .

A derivation of eq. 3.17 can be found in (Kaula, 1966). In (Sneeuw, 1991a) the same expression is derived by means of a group–theoretical approach which makes use of the so–called representation coefficients (also known as Wigner coefficients). In such an approach it becomes obvious that it is, in fact, more fundamental, and for satellite purposes also more convenient, to use in eq. 3.17 the index k instead of p, with $k = l - 2p$. The index k has a stepsize 2 and ranges from $-l$ to l. So switching from p to k and changing the notation accordingly, we make the following replacements in eq. 3.17:

$$\begin{aligned} \textstyle\sum_{p=0}^{l} &\rightarrow \textstyle\sum_{k=-l[2]}^{l} \\ \bar{F}_{lmp}(I) &\rightarrow \bar{F}_{lm}^{k}(I) \\ \psi_{lmp} &\rightarrow \psi_{km} = k\,\omega_o + m\,\omega_e \ . \end{aligned}$$

From $k = l - 2p$ we see that k and l always have the same parity, a property we will use later on. It appears convenient to abbreviate eq. 3.17 (including the replacements

above) as

$$V = \sum_{l=0}^{\infty} \sum_{m=0}^{l} \sum_{k=-l[2]}^{l} [A_{lmk}(r, I) \cos \psi_{km} + B_{lmk}(r, I) \sin \psi_{km}] \tag{3.18}$$

with

$$\left.\begin{matrix} A_{lmk}(r, I) \\ B_{lmk}(r, I) \end{matrix}\right\} = H_{lmk}(r, I) \left\{\begin{matrix} \alpha_{lm} \\ \beta_{lm} \end{matrix}\right.$$

$$H_{lmk}(r, I) = u_l(r) \bar{F}^k_{lm}(I)$$

$$\alpha_{lm} = \left[\begin{matrix} \bar{C}_{lm} \\ -\bar{S}_{lm} \end{matrix}\right]^{l-m:\text{even}}_{l-m:\text{odd}}$$

$$\beta_{lm} = \left[\begin{matrix} \bar{S}_{lm} \\ \bar{C}_{lm} \end{matrix}\right]^{l-m:\text{even}}_{l-m:\text{odd}}$$

$$u_l(r) = \frac{GM}{R} \left(\frac{R}{r}\right)^{l+1} .$$

Furthermore, we mention the possibility to rearrange the summation order l, m, k to k, m, l, as a result of which we may rewrite eq. 3.18 (truncated at some maximum degree L) in the following manner, cf. (Schrama, 1989):

$$\begin{aligned} V &= \sum_{m=0}^{L} \sum_{l=m}^{L} \sum_{k=-l[2]}^{l} [A_{lmk} \cos \psi_{km} + B_{lmk} \sin \psi_{km}] \\ &= \sum_{m=0}^{L} \sum_{k=-L}^{L} \sum_{l=lmin[2]}^{L} [A_{lmk} \cos \psi_{km} + B_{lmk} \sin \psi_{km}] \\ &= \sum_{k=-L}^{L} \sum_{m=0}^{L} [A_{km} \cos \psi_{km} + B_{km} \sin \psi_{km}] \end{aligned} \tag{3.19}$$

where

$$\begin{aligned} \left.\begin{matrix} A_{km} \\ B_{km} \end{matrix}\right\} &= \sum_{l=lmin[2]}^{L} \left\{\begin{matrix} A_{lmk} \\ B_{lmk} \end{matrix}\right. \\ &= \sum_{l=lmin[2]}^{L} H_{lmk} \left\{\begin{matrix} \alpha_{lm} \\ \beta_{lm} \end{matrix}\right. . \end{aligned} \tag{3.20}$$

and $H_{lmk} = u_l \bar{F}^k_{lm}$. In the expression above $lmin = max(|k|, m) + \delta$ where $\delta = 0$ if k and $max(|k|, m)$ have the same parity and $\delta = 1$ if they have opposite parity, cf. (Schrama, ibid.). The A_{km} and B_{km} coefficients are the so–called lumped coefficients, see (Schrama, ibid.) or (Wagner, 1983).

From eq. 3.18 we may compute the partial derivatives of V with respect to the coordinates r, ω_o, ω_e, I, resulting in different expressions for the quantities H_{lmk}, A_{lmk} and B_{lmk}. Those quantities are listed in table 3.4 for all the first– and second–order partial derivatives of V with respect to r, ω_o, ω_e, I. Following appendix A.1 we either consider the set (r, ω_o, I) or the set (r, ω_o, ω_e) so the mixed derivative with respect to I and ω_e is not listed in table 3.4. In this table we abbreviated $\bar{F}^{k\,\prime}_{lm} = \partial \bar{F}^k_{lm}/\partial I$ and $\bar{F}^{k\,\prime\prime}_{lm} = \partial^2 \bar{F}^k_{lm}/\partial I^2$. Transformation to local derivatives is, for this situation, discussed in section 4.2.

Table 3.4 *Potential derivatives w.r.t.* r, ω_o, ω_e, I

differentiation w.r.t.	$H^{(\cdot\cdot)}_{lmk}$	A_{lmk}	B_{lmk}
r	$-\frac{l+1}{r} u_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(r)}_{lmk}$	$\beta_{lm} H^{(r)}_{lmk}$
I	$u_l \bar{F}^{k\,\prime}_{lm}$	$\alpha_{lm} H^{(I)}_{lmk}$	$\beta_{lm} H^{(I)}_{lmk}$
ω_o	$k u_l \bar{F}^k_{lm}$	$\beta_{lm} H^{(\omega_o)}_{lmk}$	$-\alpha_{lm} H^{(\omega_o)}_{lmk}$
ω_e	$m u_l \bar{F}^k_{lm}$	$\beta_{lm} H^{(\omega_e)}_{lmk}$	$-\alpha_{lm} H^{(\omega_e)}_{lmk}$
rr	$\frac{(l+1)(l+2)}{r^2} u_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(rr)}_{lmk}$	$\beta_{lm} H^{(rr)}_{lmk}$
II	$u_l \bar{F}^{k\,\prime\prime}_{lm}$	$\alpha_{lm} H^{(II)}_{lmk}$	$\beta_{lm} H^{(II)}_{lmk}$
$\omega_o\omega_o$	$-k^2 u_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(\omega_o\omega_o)}_{lmk}$	$\beta_{lm} H^{(\omega_o\omega_o)}_{lmk}$
$\omega_e\omega_e$	$-m^2 u_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(\omega_e\omega_e)}_{lmk}$	$\beta_{lm} H^{(\omega_e\omega_e)}_{lmk}$
rI	$-\frac{l+1}{r} u_l \bar{F}^{k\,\prime}_{lm}$	$\alpha_{lm} H^{(rI)}_{lmk}$	$\beta_{lm} H^{(rI)}_{lmk}$
$r\omega_o$	$-\frac{l+1}{r} k u_l \bar{F}^k_{lm}$	$\beta_{lm} H^{(r\omega_o)}_{lmk}$	$-\alpha_{lm} H^{(r\omega_o)}_{lmk}$
$r\omega_e$	$-\frac{m(l+1)}{r} u_l \bar{F}^k_{lm}$	$\beta_{lm} H^{(r\omega_e)}_{lmk}$	$-\alpha_{lm} H^{(r\omega_e)}_{lmk}$
$\omega_o\omega_e$	$-mk u_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(\omega_o\omega_e)}_{lmk}$	$\beta_{lm} H^{(\omega_o\omega_e)}_{lmk}$
$I\omega_o$	$k u_l \bar{F}^{k\,\prime}_{lm}$	$\beta_{lm} H^{(I\omega_o)}_{lmk}$	$-\alpha_{lm} H^{(I\omega_o)}_{lmk}$

3.3 Synthesis and analysis

In the previous sections we have seen that gravity gradients are measurable, in principle, in an earth orbiting satellite by means of differential accelerometry (section 2.1). We have seen how this technique will be realized in practice and what kinds of problems thereby will occur (section 2.3). We have seen how to express the observed gradients in several coordinate systems (section 3.1), how they are related to the desired potential coefficients (section 3.2) and for which applications we need these potential coefficients, as they are derived from satellite gradiometry (section 2.2). But what should we expect from gradiometric measurements in terms of magnitude and spatial or spectral characteristics of the signal?

Probably the easiest way to get at least a first, rough estimate of the size of the gravity gradients is to take for the gravitational potential V the central term $\frac{GM}{r}$ only, representing the potential of a homogeneous spherical mass distribution. In terms of a spherical harmonic expansion (eq. 3.14) this means that the only non–zero potential coefficient is $\bar{C}_{00} = 1$. Then we can easily compute the gravity gradients in a local north–oriented coordinate system $x^{i'}$. Since in this case V is only a function of r all the first– and second–order partial derivatives with respect to θ and λ are zero. Using eq. 3.13 this leaves for the elements of the gradient tensor:

$$V_{i'j'} = \begin{bmatrix} \frac{1}{r}V_r & 0 & 0 \\ 0 & \frac{1}{r}V_r & 0 \\ 0 & 0 & V_{rr} \end{bmatrix} = \frac{GM}{r^3} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} .$$

At a satellite altitude of 200 km we find in this case for the diagonal elements $V_{x'x'} = V_{y'y'} = -\frac{1}{2}V_{z'z'} \approx$ -1400 E.

In reality the earth is not a perfectly homogeneous spherical mass. In terms of the series expansion 3.14 an infinite number of non–zero potential coefficients exists. The potential V and the gravity gradients are not only a function of r but also of θ and λ. A global computation of the gravity gradient signal should therefore include a full evaluation of the series expansion. Whereas in this equation (and also in eq. 3.17) the summation over l ranges from zero to infinity, one has to choose some finite maximum degree $L \geq 0$ when doing practical computations. Such evaluation can be done in arbitrary points of which the geocentric polar coordinates (r, θ, λ) are specified. If we carry out such a computation in individual points, it is called single point computation.

Very often, however, we like to compute the gravity gradients in the nodes of an equi–angular world wide grid located on a sphere with radius r. Such a computation, based on the evaluation of eq. 3.14, is called *spherical harmonic synthesis* (Colombo, 1981). Besides the polar or orbital coordinates of the (grid) points, gradient synthesis (using eq. 3.14 or 3.17) requires the availability of some known set of potential coefficients. Nowadays several potential coefficient sets exist, for example GEM–T2, complete up to degree and order $L = 36$ (Marsh et al., 1989), OSU180

with $L = 180$ (Rapp, 1981) and OSU86F with $L = 360$ (Rapp and Cruz, 1986). The first one is derived purely from orbit analysis of a variety of satellites, the second included altimeter data from Seasat and terrestrial data (gravity anomalies), whereas the third model was computed from altimeter data and land measurements including geophysically predicted gravity anomalies. A maximum degree L of 180 corresponds to a smallest recoverable wavelength of 2° thus giving the possibility of showing gravitational features in a grid of 1° × 1° blocks (half of the smallest wavelength or the *resolution*). In the same way data in 0.5° × 0.5° blocks corresponds to $L = 360$.

So if some *known* set of potential coefficients is available, the synthesis problem may give us some insight into the signals to be expected from a gradiometer mission by *computing* a set of gravity gradients. But the reverse reasoning is also of interest to us: given the values of the gradients in a set of points, regular or irregular distributed over the whole earth, compute from them the potential coefficients. If one uses an equation of the type 3.14 this problem is called *spherical harmonic analysis* (Colombo, 1981). The set of gradients could be thought of as *observed* during a real gradiometer mission, from which we like to recover *unknown* potential coefficients.

In general any functional of the gravitational potential may be generated and/or analyzed by means of, respectively, the synthesis and analysis problems discussed above. Such potential functionals could include for instance the gravitational potential itself, its first– and second–order derivatives in any coordinate system, geoid heights, gravity anomalies, deflections of the vertical, etc. Let us write $f(\theta, \lambda)$ for such arbitrary function, defined on the surface of a sphere, and let us expand this function in a spherical harmonic series as

$$f(\theta,\lambda) = \sum_{l=0}^{L} \sum_{m=0}^{l} \bar{P}_{lm}(\cos\theta)\,[a_{lm}\cos m\lambda + b_{lm}\sin m\lambda] \tag{3.21}$$

where the a_{lm} and b_{lm} are the normalized spherical harmonic coefficients. Comparison with eq. 3.14 reveals the same structure, so that, if we identify f with the gravitational potential V and a_{lm} and b_{lm} with respectively $(GM/R)(R/r)^{l+1}\bar{C}_{lm}$ and $(GM/R)(R/r)^{l+1}\bar{S}_{lm}$, the two expressions coincide. In analogy with eq. 3.15 we could write 3.21 as

$$f(\theta,\lambda) = \sum_{m=0}^{L} [a_m \cos m\lambda + b_m \sin m\lambda] \tag{3.22}$$

with

$$\left.\begin{matrix} a_m \\ b_m \end{matrix}\right\} = \sum_{l=m}^{L} \bar{P}_{lm}(\cos\theta) \left\{\begin{matrix} a_{lm} \\ b_{lm} \end{matrix}\right. . \tag{3.23}$$

The harmonic coefficients a_{lm}, b_{lm} constitute the *spectrum* of f and are therefore sometimes called spectral coefficients, thereby referring to the theory of Fourier analysis, where one also uses the term *synthesis* for the Fourier series or inverse Fourier

transform of respectively periodic or aperiodic signals and the term *analysis* for the Fourier series coefficients resp. the Fourier transform, e.g. (Oppenheim et al., 1983). Note, however, that eq. 3.21 does not correspond exactly to a two–dimensional Fourier transform in the plane, cf. (Schwarz, 1985). From the spectral coefficients one computes the (quadratic) *power spectrum* c_l, $(l = 0, \ldots, L)$ as

$$c_l = \sum_{m=0}^{l} \left(a_{lm}^2 + b_{lm}^2 \right) . \tag{3.24}$$

The c_l are also called *degree variances*. The *degree–order variances* c_{lm} are computed from the power spectrum as

$$c_{lm} = \frac{c_l}{2l+1} , \tag{3.25}$$

the square root of which is referred to as r.m.s. value per coefficient per degree. Given the $f(\theta, \lambda)$ in a grid the spectral coefficients a_{lm} and b_{lm} themselves can be computed using the following numerical quadrature formula:

$$\left.\begin{matrix} \hat{a}_{lm} \\ \hat{b}_{lm} \end{matrix}\right\} = \frac{1}{4\pi} \sum_{i=0}^{L-1} \Delta_i \bar{P}_{lm}(\cos\theta_i) \sum_{j=0}^{2L-1} f(\theta_i, \lambda_j) \left\{\begin{matrix} \cos m\lambda_j \\ \sin m\lambda_j \end{matrix}\right. \tag{3.26}$$

where $\Delta_i = \Delta\lambda\Delta\theta \sin\theta_i$ and where "$\hat{}$" indicates the fact that these coefficients are estimates of the a_{lm} and b_{lm}. The $f(\theta_i, \lambda_j)$ are the function values at the grid points. Eq. 3.26 is a discrete approximation of

$$\left.\begin{matrix} a_{lm} \\ b_{lm} \end{matrix}\right\} = \frac{1}{4\pi} \int_{\theta=0}^{\pi} \bar{P}_{lm}(\cos\theta) \int_{\lambda=0}^{2\pi} f(\theta, \lambda) \left\{\begin{matrix} \cos m\lambda \\ \sin m\lambda \end{matrix}\right\} d\lambda \, d\theta$$

where σ represents the unit sphere. This formula is based on the orthogonality property of spherical harmonics, see for example (Heiskanen and Moritz, 1967). For the $\cos m\lambda$ and $\sin m\lambda$ terms the orthogonality also applies in the case of a discrete summation, like the summation over j in eq. 3.26. For the i–summation over the $\bar{P}_{lm}$ terms the orthogonality property holds only approximately so the $\hat{a}_{lm}$ and $\hat{b}_{lm}$ will contain an error due to this approximation.

The computation of potential functionals in a grid, and the estimation of spectral coefficients from such grid, involve some equivalent numerical techniques (duality between synthesis and analysis (Colombo, 1981)). This can be seen by looking at the transform pair 3.22 and 3.26. Firstly, both problems require the computation of all Legendre functions up to degree and order L. Secondly, half of the problem can be solved by applying pre–programmed FFT routines. For spherical harmonic synthesis one first computes the coefficients a_m and b_m using eq. 3.23. Afterwards the grid values are obtained by applying an inverse FFT with those coefficients (eq. 3.22). When evaluating the spherical harmonic analysis equation 3.26, the j–summation can also be done by an FFT, after which the i–summation is carried out separately.

The value one chooses for L depends mainly on the specific application, but it may have a rather large impact on numerical computations. Typical numerical problems like underflow, numerical stability, but also problems with computation time and data storage may occur when simulating gravity gradients (or in general any potential functional) up to high maximum degree, for example $L = 360$, and at a large number of points. Therefore, we will pay here some attention to a few numerical aspects of gradient synthesis, in particular to the recursive computation of Legendre functions, single point versus grid computation and vectorization.

Legendre functions

In the process of evaluating a series expansion of the potential and its derivatives up to high degree and order the computation of the Legendre functions plays an important role. Since the computation of the Legendre functions by means of explicit formulas (for example (Heiskanen and Moritz, 1967), eq. 1–60 or 1–62) is too time consuming, one uses recurrent relations. For the normalized Legendre functions we use the relations:

$$\bar{P}_{l,l} = f_1 \sin\theta \bar{P}_{l-1,l-1} \tag{3.27}$$

$$\bar{P}_{l,l-1} = f_2 \cos\theta \bar{P}_{l-1,l-1} \tag{3.28}$$

$$\bar{P}_{l,m} = f_3(f_4 \cos\theta \bar{P}_{l-1,m} - f_5 \bar{P}_{l-2,m}) \tag{3.29}$$

with the starting values

$$\bar{P}_{0,0} = 1$$

$$\bar{P}_{1,1} = \sqrt{3}\sin\theta$$

where

$$f_1 = \sqrt{\frac{2l+1}{2l}}$$

$$f_2 = \sqrt{2l+1}$$

$$f_3 = \sqrt{\frac{2l+1}{(l-m)(l+m)}}$$

$$f_4 = \sqrt{2l-1}$$

$$f_5 = \sqrt{\frac{(l-m-1)(l+m-1)}{2l-3}}\,.$$

We may visualize these recurrent relations by means of arrows in an l, m–scheme with degree l on the horizontal axis and order m on the vertical axis (see figure 3.1).

Equation 3.27 is the diagonal recurrence upwards on the main diagonal, eq. 3.28 the first step in horizontal direction for fixed order m, arriving at the first sub–diagonal and eq. 3.29 expresses all the further steps in horizontal direction for increasing degree l. There also exists a vertical downward recurrent relation (for fixed

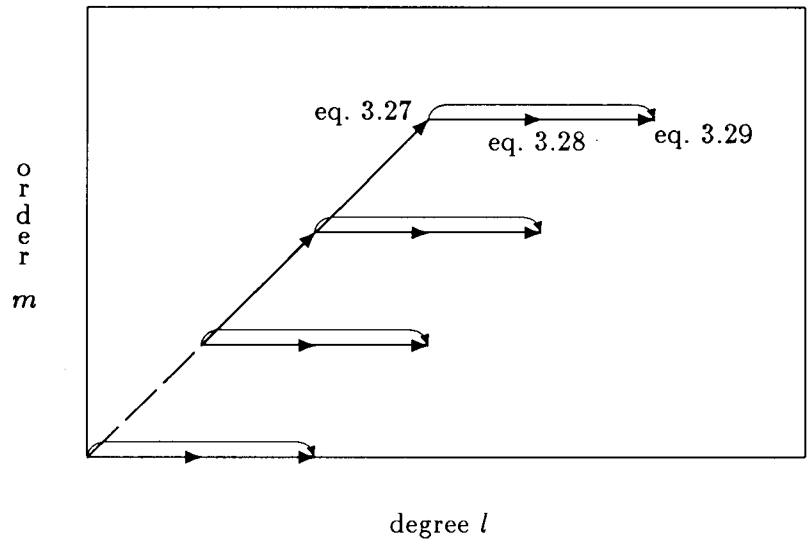


Figure 3.1 *Recurrence relations for Legendre functions*

degree l), namely

$$\bar{P}_{l,m-2} = \frac{1}{f_6}\left[-2(m-1)\frac{\cos\theta}{\sin\theta}\bar{P}_{l,m-1} + f_7\bar{P}_{l,m}\right]$$

with

$$f_6 = \begin{cases} -\sqrt{2l(l+1)} & \text{if} \quad m = 2 \\ -\sqrt{(l-m+2)(l+m-1)} & \text{if} \quad m \neq 2 \end{cases}$$

$$f_7 = \sqrt{(l+m)(l-m+1)}\,.$$

This recurrence, however, is not used here since it becomes numerically singular for small θ due to the factor $\frac{1}{\sin\theta}$. Furthermore, a downward recurrence is not recommendable when, during the computational process, an underflow occurs on the main diagonal. An overview of all possible recurrence relations can be found in (Ilk, 1983). Compare also (Gerstl, 1980).

Starting the diagonal recurrence with $\bar{P}_{00} = 1$ the value of the $\bar{P}_{ll}$ for increasing l will decrease rapidly. The degree at which a numerical underflow will occur depends on the value of the maximum exponent of the used computer and on the latitude θ (see fig. 3.2). The underflow results in a zero value for the $\bar{P}_{ll}$. In that case all subsequent $\bar{P}_{lm}$ will also be zero. If a downward recurrence would have been used a much larger part of all the $\bar{P}_{lm}$ would be zero due to the underflow on the main diagonal as in the case of horizontal recurrences, cf. (Koop and Stelpstra, 1989).

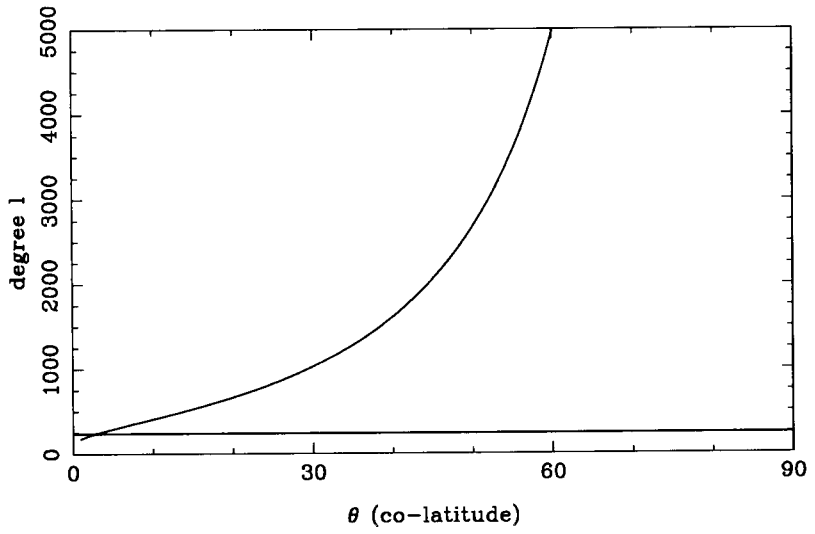


Figure 3.2 *Degree l at which the $\bar{P}_{ll}$ (during the diagonal recurrence) become smaller than 10^{-308} (and is therefore set to zero on the CONVEX) as function of θ.*

Whereas the value of the $\bar{P}_{ll}$ for increasing l on the main diagonal decreases, the value of the $\bar{P}_{lm}$ for fixed m and increasing l during a horizontal recurrence slightly increases. An underflow may therefore cause information to be lost for high degrees. This is even more the case for the second–order derivatives of the Legendre functions $\bar{P}''_{lm}$ which can reach values up to 10^6. This problem of underflows can be dealt with by scaling techniques, as described in (Koop and Stelpstra, ibid.) for recursive Legendre function computations or in (Sneeuw, 1991a) for recursive inclination function computations.

The recurrent relations for the first– and second–order derivatives of the Legendre functions may be derived from equations 3.27 – 3.29 by means of differentiation with respect to θ, obtaining:

$$\begin{aligned}
\bar{P}'_{l,l} &= f_1(\cos\theta\bar{P}_{l-1,l-1} + \sin\theta\bar{P}'_{l-1,l-1}) \\
\bar{P}'_{l,l-1} &= f_2(-\sin\theta\bar{P}_{l-1,l-1} + \cos\theta\bar{P}'_{l-1,l-1}) \\
\bar{P}'_{l,m} &= f_3(-f_4\sin\theta\bar{P}_{l-1,m} + f_4\cos\theta\bar{P}'_{l-1,m} - f_5\bar{P}'_{l-2,m})
\end{aligned}$$

with the starting values

$$\begin{aligned}
\bar{P}'_{0,0} &= 0 \\
\bar{P}'_{1,1} &= \sqrt{3}\cos\theta
\end{aligned}$$

and

$$\bar{P}''_{l,l} = f_1(-\sin\theta\bar{P}_{l-1,l-1} + 2\cos\theta\bar{P}'_{l-1,l-1} + \sin\theta\bar{P}''_{l-1,l-1})$$

$$\bar{P}''_{l,l-1} = f_2(-\cos\theta\bar{P}_{l-1,l-1} - 2\sin\theta\bar{P}'_{l-1,l-1} + \cos\theta\bar{P}''_{l-1,l-1})$$
$$\bar{P}''_{l,m} = f_3(-f_4\cos\theta\bar{P}_{l-1,m} - 2f_4\sin\theta\bar{P}'_{l-1,m} + f_4\cos\theta\bar{P}''_{l-1,m} - f_5\bar{P}''_{l-2,m})$$

with

$$\bar{P}''_{0,0} = 0$$
$$\bar{P}''_{1,1} = -\sqrt{3}\sin\theta\,.$$

Single point versus grid computation

Gradient computation, using high degree potential models, is a time consuming process. Especially if one wishes to compute the gradients in a large number of points. Depending on further applications, one may proceed in different ways. Either one evaluates the full spherical harmonic expansion 3.14 in each point separately (single point computation) or one applies spherical harmonic synthesis using FFT techniques as described above, resulting in a grid of gradient values, having constant step size in θ and λ direction (grid computation).

Grid computation is especially useful if one, for example, has to compute world wide representations of gravity gradients using different coefficient sets as input, aiming at, for instance, comparison of potential models or comparison with gradients obtained from other techniques. The main advantage of this method is of course the use of the FFT, decreasing computation time drastically. The step size in λ direction depends on the degree of the FFT, which in turn depends on the maximum degree L of the potential model which is used as input. For example, using a potential coefficient set with $L = 180$ allows to compute independent gradient values in points with $\Delta\lambda = 1°$ (Nyquist frequency). The degree of the (one–dimensional) FFT in that case will be 360. As for the θ direction, one may arbitrarily choose the value of θ (Colombo, 1981) with the limitation that the smallest possible wavelength to be present in the grid is determined by L (the maximum degree of the potential model), even if one chooses more θ's than L.

For e.g. satellite applications the grid computation may not be well suited, since the points along a satellite orbit where one wishes to compute the gradients usually do not coincide with the nodes of a grid. Using, however, a series expansion of the type 3.19, we may still apply an FFT, in this case not along a parallel but along the satellite orbit. Another possibility is to compute a grid at satellite altitude and interpolate between the grid points. Both latter methods, however, fail if the satellite orbit is elliptical so that r is no longer constant. The interpolation method could be extended to three dimensions (Schrama, 1984) or (Wichiencharoen, 1985), for which grids at different altitudes have to be computed first. A final possibility is to compute gradients at each point of the orbit individually, bringing us back to single point computation.

The choice between single point and grid computation depends, however, not only on the specific application, but also on the available computer hardware and software. As for the machine characteristics, we mention the maximum value of

the exponent, optimization properties (Koop and Stelpstra, 1989), CPU speed and available memory. Concerning the software, accurate and fast interpolation and FFT routines are nowadays widely available. Other aspects are handling of underflows, speed, portability, memory usage, handling of various types of coordinates and potential models, accuracy of the input (coordinates and coefficients), use of single or double precision numbers, vectorization and parallelization capabilities, etc. In (Balmino et al., 1990) several software packages for computation of potential functionals are compared on (some of) these aspects. The influence of machine characteristics and software capabilities on computation time and accuracy of the results increases with increasing maximum degree L. In general, single point computation will be the most time consuming method, but it does not require much memory. Interpolation methods introduce some loss of accuracy and they require more memory capacity. Methods using FFT are very fast, but have the restriction of computing the gradient values at regularly distributed points.

Due to the wide availability of ever larger and faster computers the machine influence will become less and less important in the future. Nevertheless, it is to be expected at this moment that for future gradiometer missions the use of vector computers or even parallel computers will be indispensable. Therefore, some remarks about vectorization will be made in the following.

Vectorization

Vectorization is, in fact, only one aspect of what is called more generally *optimization*. When talking about optimization we have to discriminate between time and space optimization. In particular time optimization plays an important role when dealing with very large computational problems like gradient synthesis up to high degree. First of all it must be stated that optimization is not only a matter of hardware and system software. If one likes to fully benefit the optimization capabilities of the hardware configuration, also the users software should be optimally adapted to it. In spite of the fact that there are some general techniques of optimization, the specific computer which is used eventually determines the way in which the software should be set up.

Already in the case of "conventional" computers (scalar processor or so called "Single Instruction Single Data" (SISD) machines) the impact of the algorithm set-up can be significant. Just look at the difference in execution speed between linear and binary search. Compilers running on machines equipped with a vector processor (so called "Single Instruction Multiple Data" (SIMD) machines) modify the code in order to fully benefit the use of the vectorization capabilities. But the programmer still has to design the algorithm in such a way as to fit optimally the compiler. Whereas vectorization intends to decrease CPU time, parallelization tries to reduce "time to solution" by spreading work across multiple CPU's. The development of parallel computers ("Multiple Instruction Multiple Data" (MIMD) machines), equipped with a large number of parallel connected processors, is still going on, and

they are not yet very common in use.

In (Koop and Stelpstra, 1989) it is described how the algorithms for the recursive computation of the Legendre functions as well as the final evaluation of the potential functionals have to be organized as to perform optimally on a CONVEX C240 vector computer. It appeared that, in the case of single-point potential synthesis, the best way to go through the l, m-scheme was diagonally, i.e. first the main diagonal ($m = l$, $\forall\, m \leq L$), then the first sub-diagonal ($m = l - 1$, $\forall\, m \leq L - 1$), then the second sub-diagonal ($m = l - 2$, $\forall\, m \leq L - 2$), etc. For grid computation, the same order is used, evaluating in each call the Legendre functions or potential values for all θ's.

Numerical tests

We did some numerical tests concerning spherical harmonic synthesis and analysis. The tests were carried out on the CONVEX C240 vector computer. We used the OSU86F (Rapp and Cruz, 1986) set of potential coefficients, however GRS80 reference values for the first four zonal coefficients were subtracted and $\bar{C}_{21}$ and $\bar{S}_{21}$ were assigned special values, see table 3.5. Furthermore, a maximum degree $L = 240$ was chosen for the computations.

Table 3.5 *Specifications of the numerical tests*

OSU86F with J_2, J_4, J_6, J_8 from GRS80 subtracted
$\bar{C}_{00} = 0$
$\bar{C}_{21} = -0.10 \cdot 10^{-9}$ $\qquad$ $\bar{S}_{21} = 0.102 \cdot 10^{-8}$
$GM = 3.986004404 \cdot 10^{14}$ m^{3}s^{-2}
$R = 6378137$ m

First we examined the approximation error introduced by the numerical quadrature formula 3.26. We have seen that this error is caused by the i-summation only, so we inserted eq. 3.21 into 3.26. In the resulting equation, from which the j-summation is eliminated due to the orthogonality of Fourier series, we inserted on the right hand side the OSU86F coefficients. After numerical evaluation of this equation, we obtain the estimates $\hat{a}_{lm}$ and $\hat{b}_{lm}$, which can then be compared with the original OSU86F set. Figure 3.3 shows the logarithm of the degree variances of the relative differences between the estimated and the original coefficients for two cases: a step size in θ direction of $\Delta\theta = 0°.5$ (i.e. 360 parallels) and a step size $\Delta\theta = 0°.25$ (720 parallels). Since the i-summation is in fact nothing more than a numerical integration, the error should decrease with decreasing step size, as indeed can be seen in the figure. Furthermore we see that the error slowly increases with increasing

degree l, reaching about 10^{-6} at $l = 240$ in the case of 360 parallels. We have to bear in mind, however, that part of this error may be caused by underflows in the Legendre recursions, compare figure 3.2 and the paragraph on Legendre functions.

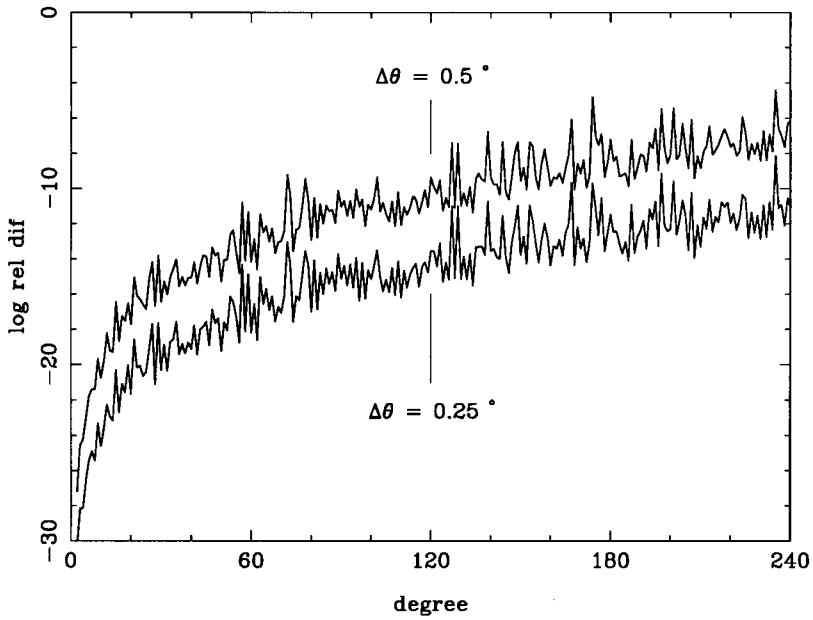


Figure 3.3 *Relative error made by using the numerical quadrature formula*

Secondly, a program was made for grid computation of potential functionals. The program is an adapted version of the single point program EVLPOT described in (Koop and Stelpstra, 1989), which is especially tailored for use on the CONVEX C240 vector computer. It computes a grid of nine potential functionals in one run (potential and first– and second–order derivatives), both in geocentric cartesian and local north–oriented coordinates. Due to the vectorized set–up is does not make use of an FFT routine. Comparison with a spherical harmonic synthesis version (which uses FFT) did not show any significant differences in computation time (!), at least up to $L = 240$ on the CONVEX. With the program we computed a world wide grid of gravity gradients in a local north–oriented coordinate system at an altitude of 200 km. This grid was then analyzed for each of the six gradients using Colombo's HARMIN program for spherical harmonic analysis (Colombo, 1981). We slightly adapted the program in order to perform better on the vector computer. HARMIN makes use of the numerical quadrature formula 3.26 so it yields estimates $\hat{a}_{lm}, \hat{b}_{lm}$ of the spectral coefficients a_{lm}, b_{lm} (eq. 3.21) for some function f, in our case gravity gradients. From these estimates we computed power spectra using 3.24, which are

shown in figure 3.4.

From this figure (and confirmed by degree–wise comparison of the power spectra) we see that the power of the various gradients approximately obey the relations (cf. (Rummel and v. Gelderen))

$$\begin{aligned} & & \sigma_l^2(x'y') &\approx \tfrac{1}{8}\sigma_l^2(z'z') \\ \sigma_l^2(x'x') &\approx & \sigma_l^2(y'y') &\approx \tfrac{3}{8}\sigma_l^2(z'z') \\ \sigma_l^2(x'z') &\approx & \sigma_l^2(y'z') &\approx \tfrac{1}{2}\sigma_l^2(z'z') \quad . \end{aligned}$$

Note that the spectral coefficients a_{lm}, b_{lm} are in general *not* in a simple way related to the potential coefficients $\bar{C}_{lm}, \bar{S}_{lm}$. Only for $V_{z'z'}$ such relation is easy:

$$\left.\begin{matrix} a_{lm}^{(z'z')} \\ b_{lm}^{(z'z')} \end{matrix}\right\} = \frac{GM}{R^3}\left(\frac{R}{r}\right)^{l+3}(l+1)(l+2)\left\{\begin{matrix} \bar{C}_{lm} \\ \bar{S}_{lm} \end{matrix}\right. \quad .$$

This relation gives us an easy way to compare the results, obtained via synthesis and analysis, with the original coefficients. The absolute relative differences between the left–hand side and the right–hand side of the above equation are shown in figure 3.5. It can be seen the error remains below the level of 10^{-2} over the whole spectrum.

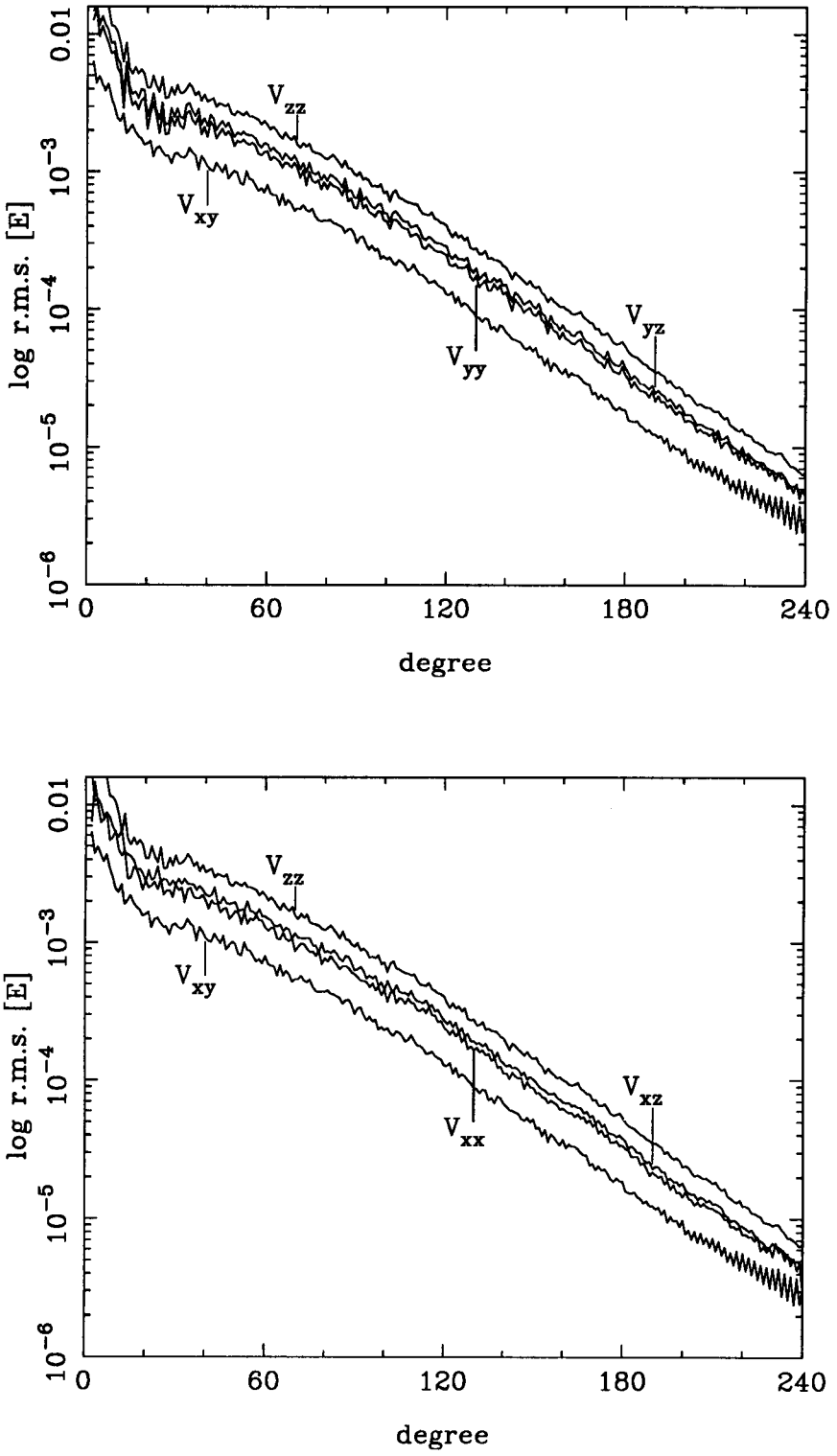


Figure 3.4 *R.m.s. spectra for six gradients at 200 km*

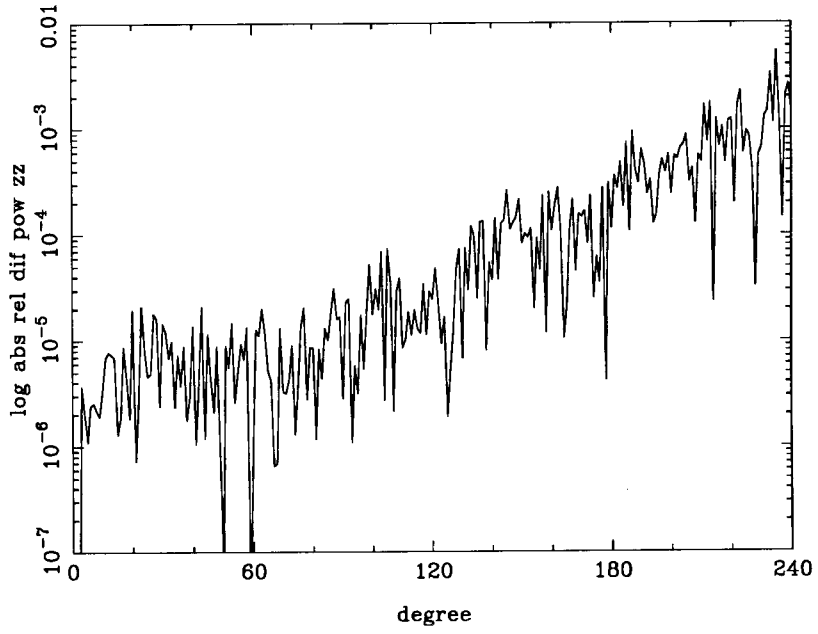


Figure 3.5 *Relative difference between original potential coefficients and coefficients computed via synthesis and analysis of a $V_{z'z'}$-grid.*

4

Global gradiometric analysis

In chapter 2 we have seen that our main interest in satellite gradiometry is the precise and detailed determination of the earth's gravitational field. The gravitational information derived from the observed gravity gradients is presented by means of a set of potential coefficients $\bar{C}_{lm}, \bar{S}_{lm}$ up to some maximum degree and order, together with an indication of their precision. The relations between the measurements and the unknown parameters (potential coefficients or potential coefficient corrections) are derived in chapter 3. In the present chapter attention shall be paid to the method of solving the potential coefficients from the measurements (analysis). We restrict ourselves to a *global analysis*, which means each spherical harmonic coefficient reflects a feature of the global field, corresponding to some spatial wavelength. Investigations into regional (local) approaches can be found in e.g. (Tscherning et al., 1990), (Robbins, 1985), (Arabelos and Tscherning, 1990), (Ilk, 1987) or (Ilk et al., 1990). Furthermore, emphasis will be laid on *error analysis*, which means that we study the error propagation during the adjustment process. This will be done for several kinds of possible mission scenarios.

The method we use here is based on *least squares adjustment* and was proposed by O.L. Colombo (1987, 1989). One of the main characteristics of this method is that, under certain assumptions (see section 4.2), the normal matrix in the least squares estimation process attains a block–diagonal structure which can be inverted without much (computational) effort. Without the block–diagonal structure it will be a tough job to invert the normal matrix, especially for high degree solutions (Balmino and Barriot, 1990). After a description of the least squares adjustment method in the first section of this chapter, the second section deals with the error analysis. Finally the third section deals with methods of global recovery (i.e. solving the potential coefficients), again based on a least squares adjustment.

4.1 Least squares analysis

We will show here how the problem of solving the potential coefficients from gradiometer measurements is set up in terms of a least squares adjustment model. For the measurements we take a set of observed second–order radial derivatives. This is just an example. Equivalent derivations hold for other gradients, as well as for combinations of gradients. For the moment, orbit and orientation uncertainties are neglected.

Observations are given along a satellite orbit and it is therefore appropriate to use an expansion of the gravitational potential in inclination functions, as was given in section 3.2. We assume the measurements are given with respect to a local orbital coordinate system x^i. In appendix A this local orbital system is defined. From section 3.1.2 we find that $V_{zz} = V_{rr}$. Using equation 3.18 and table 3.4 from section 3.2.2 we have for the second–order radial derivative:

$$V_{zz} = \sum_{l=0}^{L} \sum_{m=0}^{l} \sum_{k=-l[2]}^{l} [A_{lmk}^{(zz)} \cos\psi_{km} + B_{lmk}^{(zz)} \sin\psi_{km}] \tag{4.1}$$

with

$$\left.\begin{matrix} A_{lmk}^{(zz)} \\ B_{lmk}^{(zz)} \end{matrix}\right\} = H_{lmk}^{(zz)} \left\{\begin{matrix} \alpha_{lm} \\ \beta_{lm} \end{matrix}\right.$$

$$H_{lmk}^{(zz)} = \Gamma_l (l+1)(l+2) \bar{F}_{lm}^{k}$$

$$\Gamma_l = \frac{GM}{R^3} \left(\frac{R}{r}\right)^{l+3} = \frac{u_l}{r^2}$$

$$\alpha_{lm} = \left[\begin{matrix} \bar{C}_{lm} \\ -\bar{S}_{lm} \end{matrix}\right]_{l-m:\text{odd}}^{l-m:\text{even}}$$

$$\beta_{lm} = \left[\begin{matrix} \bar{S}_{lm} \\ \bar{C}_{lm} \end{matrix}\right]_{l-m:\text{odd}}^{l-m:\text{even}} .$$

$$\psi_{km} = k\,\omega_o + m\,\omega_e$$

For practical computations we have to truncate the infinite summation over l at some maximum degree L. In equation 4.1 the potential coefficients $\bar{C}_{lm}$ and $\bar{S}_{lm}$ are

the unknowns. Now consider a large set of measurements V_{zz} given for all kinds of values of the coordinates ω_o, ω_e (we assume the inclination I of the orbit remains constant during the mission). For each measurement we have an equation of the type above, each as a function of the same potential coefficients $\bar{C}_{lm}, \bar{S}_{lm}$. Now let us arrange all these measurements V_{zz} in one vector, denoted $\tilde{\boldsymbol{\ell}}$ (the "~" denotes stochastic quantities). We may also order the unknown coefficients, for each value of degree l and order m successively, in one vector, denoted $\boldsymbol{c}$, merging the double summation over l and m into one. In this way we obtain the following system of equations:

$$\tilde{\boldsymbol{\ell}} = A\,\boldsymbol{c} + \tilde{\varepsilon} \tag{4.2}$$

in which $\tilde{\varepsilon}$ are the residuals between the observations and the model. This model is represented by the design matrix A of which each element is of the type

$$\sum_{k=-l[2]}^{l} H_{lmk}^{(zz)} \cos\psi_{km} \tag{4.3}$$

or

$$\sum_{k=-l[2]}^{l} H_{lmk}^{(zz)} \sin\psi_{km}\ , \tag{4.4}$$

depending on whether we are dealing with a $\bar{C}_{lm}$ or $\bar{S}_{lm}$ coefficient and whether $l-m$ is even or odd. The a–priori variances and covariances of the observations are collected in the variance–covariance matrix Q_{zz}. Solving the model 4.2 in a least squares sense (i.e. minimizing $\tilde{\varepsilon}^T Q_{zz}^{-1} \tilde{\varepsilon}$), we find for the estimates $\hat{\boldsymbol{c}}$ of the potential coefficients:

$$\hat{\boldsymbol{c}} = N^{-1} A^T Q_{zz}^{-1}\, \boldsymbol{\ell} \tag{4.5}$$

with the normal matrix N

$$N = A^T Q_{zz}^{-1} A\ .$$

The inverse of the normal matrix, N^{-1}, represents the a–posteriori variances and covariances of the estimated potential coefficients. It can be calculated without having to perform a complete adjustment. This offers the possibility of carrying out an error analysis without the availability of a real set of observations. This fact will be used in the next section.

Since a gradiometer not only measures the second–order radial derivative but also, as in the case of Aristoteles, the gradients V_{yy} and V_{yz}, and in the future hopefully all six gradients, we will need observation equations for these gradients too.

The relation between the local gradients V_{ij} and the potential derivatives with respect to r, I, ω_o, ω_e are given in section 3.1.2. There we see that we have two possibilities. Either we choose the set $\{x^a | a = 1,2,3\} = (r, \omega_o, \omega_e)$ (equations 3.12) or the set $\{x^{a'} | a' = 1,2,3\} = (r, \omega_o, I)$ (equations 3.11). It can be shown that both sets of expressions give the same result if used for gradient synthesis. Just for the

sake of convenience we choose equations 3.11. For all six gradients an expression of the type 4.1 can be derived, using equations 3.11 and table 3.4, resulting in again different expressions for the quantities A_{lmk}, B_{lmk} and H_{lmk}. For all six gradients, those quantities are listed in table 4.1.

Table 4.1 *Expressions for local gradients in terms of derivatives of the potential with respect to* (r, ω_o, I).

	$H^{(\cdot\cdot)}_{lmk}$	A_{lmk}	B_{lmk}
xx	$-(l+1+k^2)\Gamma_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(xx)}_{lmk}$	$\beta_{lm} H^{(xx)}_{lmk}$
xy	$-\sin^{-1}\omega_o \Gamma_l \bar{F}^{k\,\prime}_{lm}$	$(\cot\omega_o \alpha_{lm} - k\beta_{lm}) H^{(xy)}_{lmk}$	$(k\alpha_{lm} + \cot\omega_o \beta_{lm}) H^{(xy)}_{lmk}$
xz	$-(l+2)\Gamma_l k \bar{F}^k_{lm}$	$\beta_{lm} H^{(xz)}_{lmk}$	$-\alpha_{lm} H^{(xz)}_{lmk}$
yy	$-((l+1)^2 - k^2)\Gamma_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(yy)}_{lmk}$	$\beta_{lm} H^{(yy)}_{lmk}$
yz	$-(l+2)\sin^{-1}\omega_o \Gamma_l \bar{F}^{k\,\prime}_{lm}$	$\alpha_{lm} H^{(yz)}_{lmk}$	$\beta_{lm} H^{(yz)}_{lmk}$
zz	$(l+1)(l+2)\Gamma_l \bar{F}^k_{lm}$	$\alpha_{lm} H^{(zz)}_{lmk}$	$\beta_{lm} H^{(zz)}_{lmk}$

In this table we computed the expression for V_{yy} using the Laplace equation: $V_{yy} = -(V_{xx} + V_{zz})$, thereby avoiding the second-order derivatives of the inclination functions and obtaining a simpler equation. As an example we derive here the expression for V_{xz}. From equation 3.11 we have

$$V_{xz} = \frac{1}{r} V_{r\omega_o} - \frac{1}{r^2} V_{\omega_o} .$$

From table 3.4 we have for the quantities A_{lmk}, B_{lmk} and H_{lmk} for the $r\omega_o$-derivative

$$\begin{aligned} A^{(r\omega_o)}_{lmk} &= \beta_{lm} H^{(r\omega_o)}_{lmk} \\ B^{(r\omega_o)}_{lmk} &= -\alpha_{lm} H^{(r\omega_o)}_{lmk} \\ H^{(r\omega_o)}_{lmk} &= -\frac{(l+1)}{r} u_l k \bar{F}^k_{lm} \end{aligned}$$

and for the ω_o-derivative

$$\begin{aligned} A^{(\omega_o)}_{lmk} &= \beta_{lm} H^{(\omega_o)}_{lmk} \\ B^{(\omega_o)}_{lmk} &= -\alpha_{lm} H^{(\omega_o)}_{lmk} \\ H^{(\omega_o)}_{lmk} &= u_l k \bar{F}^k_{lm} . \end{aligned}$$

This yields for V_{xz}:

$$
\begin{aligned}
V_{xz} &= \frac{1}{r}\sum_{l,m,k}\left[A_{lmk}^{(r\omega_o)}\cos\psi_{km} + B_{lmk}^{(r\omega_o)}\sin\psi_{km}\right] \\
&\quad -\frac{1}{r^2}\sum_{l,m,k}\left[A_{lmk}^{(\omega_o)}\cos\psi_{km} + B_{lmk}^{(\omega_o)}\sin\psi_{km}\right] \\
&= \sum_{l,m,k}\left[\left(-\frac{l+1}{r}\frac{1}{r} - \frac{1}{r^2}\right)\beta_{lm}ku_l\bar{F}_{lm}^k\cos\psi_{km} + \right. \\
&\quad \left. + \left(-\frac{l+1}{r}\frac{1}{r} - \frac{1}{r^2}\right)\alpha_{lm}ku_l\bar{F}_{lm}^k\sin\psi_{km}\right] \\
&= \sum_{l,m,k}\left[(-(l+2))\,\beta_{lm}k\Gamma_l\bar{F}_{lm}^k\cos\psi_{km} + \right. \\
&\quad \left. + (-(l+2))\,\alpha_{lm}k\Gamma_l\bar{F}_{lm}^k\sin\psi_{km}\right] \\
&= \sum_{l,m,k}\left[A_{lmk}^{(xz)}\cos\psi_{km} + B_{lmk}^{(xz)}\sin\psi_{km}\right],
\end{aligned}
$$

with $A_{lmk}^{(xz)}, B_{lmk}^{(xz)}$ and $H_{lmk}^{(xz)}$ as in table 4.1.

If we take a closer look at the expressions of this table we see that the V_{xy} and V_{yz} gradients contain a $\sin\omega_o$ term in the denominator. Apparently those expressions are singular for $\omega_o = 0 + n\pi, n = 0,1,2,\ldots$. Although this is not a very strange phenomenon (just take a look at all other transformation equations of the potential derivatives in section 3.1.2), it will appear to be inconvenient (see section 4.2.1). In order to avoid these kinds of singularities, we proceed following (Betti and Sansò, 1989).

In section 3.1.2 also the transformation equations for the second–order potential derivatives from the $\{x^{A'}|A' = 1,2,3\} = (r,\phi,\omega_o)$ coordinates to the local orbital coordinates x^i are given (equation 3.9). Confining ourselves to points along the satellite orbit, where $\phi = 0$, we obtain the simpler equations 3.10. These equations obviously do not contain a singularity (except of course for $r = 0$), but the restriction to points along the satellite's orbit limits their use. Furthermore, we face the problem that the potential is not given as function of ϕ. In appendix C it is explained, however, how to find the potential derivative with respect to ϕ, V_ϕ, as function of the derivative with respect to the inclination I, V_I. This results in a new kind of inclination function, $\bar{F}_{lm}^{k\,*}(I)$. In this appendix it is also shown how to compute these cross–track inclination functions.

Comparing equations 3.10 and 3.11 we see that the expressions for V_{xx}, V_{xz} and V_{zz} are the same in both coordinate systems whereas the expressions for the other gradients differ. If we derive the expression for V_{yy} (using the Laplace equation) as $-(V_{xx} + V_{zz})$, then only the expressions for V_{xy} and V_{yz} remain different. These are exactly the expressions with the singularities. The $\sin\omega_o$ terms in the denominators are accounted for in the cross-track inclination functions $\bar{F}_{lm}^{k\,*}$. The relation

between 3.10 and 3.11 for these two gradients is easy if we use equation C.1 from appendix C:

$$\begin{aligned} V_{xy} &= \frac{-\cos\omega_o}{r^2\sin^2\omega_o}V_I + \frac{1}{r^2\sin\omega_o}V_{I\omega_o} \\ &= \frac{1}{r^2}\frac{\partial}{\partial\omega_o}\left(\frac{1}{\sin\omega_o}V_I\right) \\ &= \frac{1}{r^2}V_{\phi\omega_o} \end{aligned}$$

and

$$\begin{aligned} V_{yz} &= \frac{-1}{r^2\sin\omega_o}V_I + \frac{1}{r\sin\omega_o}V_{rI} \\ &= \frac{\partial}{\partial r}\left(\frac{1}{r\sin\omega_o}V_I\right) \\ &= \frac{-1}{r^2}V_\phi + \frac{1}{r}V_{r\phi}\ . \end{aligned}$$

In terms of the cross-track inclination functions $\bar{F}^{k\,*}_{lm}$ we now find for the gradients V_{xy} and V_{yz} the expressions listed in table 4.2. The other gradients are the same as in table 4.1.

Table 4.2 *New expressions for V_{xy} and V_{yz} in terms of the derivatives of the potential with respect to (r,ϕ,ω_o).*

	$H^{(\cdot\cdot)}_{lmk}$	A_{lmk}	B_{lmk}
xy	$-k\Gamma_l\bar{F}^{k\,*}_{lm}$	$\alpha_{lm}H^{(xy)}_{lmk}$	$\beta_{lm}H^{(xy)}_{lmk}$
yz	$-(l+2)\Gamma_l\bar{F}^{k\,*}_{lm}$	$\beta_{lm}H^{(yz)}_{lmk}$	$-\alpha_{lm}H^{(yz)}_{lmk}$

In table 4.2 the A_{lmk} and B_{lmk} are now the coefficients of respectively the cosine and sine of the argument $\psi_{km} = k\,\omega_o + m\,\omega_e$ where now $k = l - 2p - 1$ (compare appendix C). The summation over k runs for these gradients from $-(l-1)$ to $l-1$.

In the sequel we will use the expressions from table 4.2 for the gradients V_{xy} and V_{yz} and for the other gradients the expressions from table 4.1.

4.2 Colombo's method of error analysis

The idea of a gradiometric error analysis as it is presented here originated from (Colombo, 1987). As described in the previous section, it consists of the computation of the inverse of the normal matrix in the sense of a least squares adjustment, thereby using an expression for the gradients in terms of inclination functions. The

diagonal elements of this inverse represent the a–posteriori variances of the estimated potential coefficients. As we have seen, they can be computed without the use of actual observations. The latter implies that we are not restricted to some specified orbit and instrument characteristics. Given some mission goals in terms of resolution and accuracy requirements of derived gravity anomalies or geoid heights (as were given in section 2.2) one can search for the specific orbit and instrument characteristics matching these requirements. In turn, given some orbit and instrument characteristics (e.g. mission duration limitations, instrument accuracy limitations etc. imposed by technical or practical considerations) one can derive the resulting accuracy and resolution of the estimated potential coefficients or of certain derived quantities like gravity anomalies or geoid heights. Of course, the value of such computations is limited by the underlying assumptions of the method of analysis. For example, as already stated before, the present method does not include orbit or orientation uncertainties. We simply assume to be given a global, regularly distributed set of gradient observations relative to some known local coordinate system. For the sake of simplicity, however, we make several other assumptions. These will be listed below, see also (Colombo, 1987) and (Colombo, 1989).

1. We assume the orbit of the satellite to be circular. In reality the orbit will have an eccentricity different from zero, although very small. Furthermore, by assuming the orbit to be circular we neglect the orbital decay as a result of air drag. For non–circular orbits ($e \neq 0$), we would have to include in the model the so–called *eccentricity functions* $G_{lpq}(e)$ (Kaula, 1966) which in general are to be summed for $-\infty < q < \infty$. However, it appears that for nearly circular orbits the index q can be restricted, with sufficient accuracy, to three values: $-1 < q < 1$, cf. (Schrama, 1989) or (Wagner, 1989). In spite of the fact that inclusion of these eccentricity functions would not influence our analysis method fundamentally, we leave them out for simplicity. If required, the error analysis can be carried out at different values of r to reveal the influence of varying satellite altitude.
2. We consider the data to be distributed regularly along the orbit. During the measurement periods this assumption is very well met, but gaps occur during e.g. instrument failures and orbit maintenance maneuvers, since in those periods data is not useful or not even present. Also in a real mission gaps near the poles and, to a smaller extent, at the equator may occur as a result of excessive drag variations (Touboul et al., 1991). However, if the mission is long enough the overall data set may yield a very regular data distribution.
3. Regarding the mission length we assume that it consists of an integer number of complete revolutions and an integer number of nodal days. Furthermore we will assume that the mission duration equals that of one repeat period T_r, so that no ground–track repeat will occur during the mission. To this extent, the number of orbit revolutions N_r during the mission period and the number of (nodal) days N_d contained in it have to be relative prime integers. This

can be seen as follows. Consider equation 3.18, which constitutes a Fourier series as function of the argument ψ_{km}. This series can be regarded a time series if successive measurement points along the orbit are considered. The time parameter t enters the argument ψ_{km} if we write $\omega_o = \omega_o^0 + \dot{\omega}_o(t - t_0)$ and $\omega_e = \omega_e^0 + \dot{\omega}_e(t - t_0)$ where the dot denotes differentiation with respect to t and where ω_o^0 and ω_e^0 are evaluations of respectively ω_o and ω_e at t_0. Thus:

$$\begin{aligned}\psi_{km} &= k\,\omega_o + m\,\omega_e \\ &= k\,(\omega_o^0 + \dot{\omega}_o(t - t_0)) + m\,(\omega_e^0 + \dot{\omega}_e(t - t_0)) \\ &= k\,\omega_o^0 + m\,\omega_e^0 + (k\,\dot{\omega}_o + m\,\dot{\omega}_e)(t - t_0) \\ &\equiv \psi_{km}^0 + \dot{\psi}_{km}(t - t_0)\ . \end{aligned} \tag{4.6}$$

Whereas the term ψ_{km}^0 only represents a phase shift, the term $\dot{\psi}_{km}$ represents the actual measurement frequency. However, the frequency in terms of *cycles per revolution* (c.p.r.) is denoted $\beta_{km}(= \dot{\psi}_{km}/\dot{\omega}_o)$ and may be written as (cf. (Schrama, 1989) or (Schrama, 1990)):

$$\beta_{km} = k + m\frac{\dot{\omega}_e}{\dot{\omega}_o}$$

so that $\psi_{km} = \psi_{km}^0 + \beta_{km}\dot{\omega}_o(t - t_0)$. Now one nodal day is $2\pi/\dot{\omega}_e$ seconds and one revolution is $2\pi/\dot{\omega}_o$ seconds. If the mission (T_r seconds) takes exactly N_d nodal days and N_r orbit revolutions, there exists an integer ratio between $\dot{\omega}_e$ and $\dot{\omega}_o$, i.e.

$$\frac{\dot{\omega}_e}{\dot{\omega}_o} = \frac{N_d}{N_r}\ .$$

This means that after N_d nodal days and N_r orbit revolutions the ground track will exactly repeat. However, if N_d and N_r would have some common divisor d, this repeat would already occur after N_d/d days and N_r/d revolutions. Requiring N_d and N_r to be relative primes therefore ensures no orbit repeat during the mission period T_r. This also means that during this mission period the finest possible coverage of the earth's surface is obtained. More repeat periods, being nothing more than a "repetition of experiments", would, on the other hand, only result in a re–scaling of the variances.

4. The measurements which the instrument delivers are in general the result of an averaging process over some time interval $\Delta t'$, which has to be less or equal to the sample interval Δt. This averaging can be accounted for in the following manner.

   Consider equation 4.6. Let us, for the sake of simplicity, assume $t_0 = 0$ and the initial value ψ_{km}^0 to be zero. We then have $\psi_{km} = \dot{\psi}_{km}t$. Being a function of time t, the expression for the gradients can be integrated in order to account for the time–averaging mentioned above. The integration only applies

to the cosine and sine terms and takes place over the time interval $\Delta t'$. Upon introducing the integration variable τ, this integration becomes:

$$\frac{1}{\Delta t'}\int_{t-\frac{\Delta t'}{2}}^{t+\frac{\Delta t'}{2}} \cos\dot{\psi}_{km}\tau d\tau = \frac{1}{\Delta t'}\frac{1}{\dot{\psi}_{km}}\left[\sin\dot{\psi}_{km}(t+\frac{\Delta t'}{2}) - \sin\dot{\psi}_{km}(t-\frac{\Delta t'}{2})\right]$$
$$= \frac{2}{\dot{\psi}_{km}\Delta t'}\sin(\frac{\dot{\psi}_{km}\Delta t'}{2})\cos\dot{\psi}_{km}t$$
$$\frac{1}{\Delta t'}\int_{t-\frac{\Delta t'}{2}}^{t+\frac{\Delta t'}{2}} \sin\dot{\psi}_{km}\tau d\tau = \frac{2}{\dot{\psi}_{km}\Delta t'}\sin(\frac{\dot{\psi}_{km}\Delta t'}{2})\sin\dot{\psi}_{km}t\ .$$

The factor $J_{km} \equiv 2\sin(\frac{\dot{\psi}_{km}\Delta t'}{2})/(\dot{\psi}_{km}\Delta t')$ may be incorporated in the inclination function to obtain a kind of smoothed function $J_{km}\bar{F}^k_{lm}$. In the sequel we will assume all inclination functions to be multiplied by J_{km}. We furthermore take $\Delta t' = \Delta t$.

5. Non–gravitational effects will not be included in the present error analysis. We assume that, to first order, these effects are eliminated through common mode rejection (see chapter 2). Only second–order effects, due to e.g. non–symmetry in the instrument, remain. Following (Colombo, 1987) we assume a large part of these effects, as well as some instrument errors like thermal noise, to be present in a low frequency band, mainly below $\beta_{min} = 4$ c.p.r. (Schrama, 1990). Removing from the analysis the low frequencies $|\beta_{km}| < \beta_{min}$ may be considered appropriate to account for these non–gravitational effects. Other high–pass filters may also be used, but are not considered here.

6. For the moment the covariance matrix of the measurements, Q, will be considered to be a scaled unit matrix, i.e. $Q = \sigma^2 I$. This means that consecutive measurements are considered to be uncorrelated and of equal variance. The variance factor σ^2 may, however, have arbitrary values for different gradients. Spatial differences in the precision of the data due to phenomena like e.g. drag variations are therefore not included in the present analysis.

One of the most important consequences of the assumptions above is that the normal matrix (which is to be inverted) attains a block–diagonal structure. This will be explained in the next section. Each block of the normal matrix can be inverted separately, which means an enormous reduction of computation time. In this way it becomes possible to carry out the error analysis several times with only limited (computational) effort. Each time we may change the orbit and instrument characteristics (like the inclination or instrument noise level) thereby allowing the investigation of several different mission scenarios. These mission scenarios depend on the value of the following mission parameters:

h	satellite height above reference radius R, i.e. $r = R + h$
I	inclination of the orbit
Δt	sample interval, i.e. time between two successive measurements
T_r	mission duration (one repeat period)
σ	measurement precision
β_{min}	lowest measurement frequency included in the analysis
β_{max}	highest measurement frequency included in the analysis

It will be shown later (section 4.2.3) that only for a change in some of these parameters a new run of the error analysis program is needed (e.g. for the inclination I), but that a change in the other parameters (e.g. the height h) can be accounted for by simply scaling the *inverted* normal matrix.

4.2.1 Normal matrix

The normal matrix N of the least squares adjustment problem 4.2 is given by $A^T Q^{-1} A$. In this section we will derive the expressions for the elements of this matrix, under the assumptions given in the last section.

Inserting for the covariance matrix Q the expression $\sigma^2 I$ leaves for the normal matrix

$$N = \frac{1}{\sigma^2} A^T A \, .$$

We see that the elements of the normal matrix are computed by taking the inner product of columns of the design matrix A. Each column of A contains, for one specific combination of l and m, terms of the form 4.3 or 4.4, the argument ψ_{km} (being a function of time t) indicating consecutive observation points (epochs). If we, for simplicity, assume $t_0 = 0$ and $\psi^0_{km} = 0$, we have from equation 4.6 $\psi_{km} = k\dot{\omega}_o t + m\dot{\omega}_e t$ for all $t = 0, \ldots, T_r$ where T_r is one repeat period, in our case equal to the mission duration. Now we only have measurements in discrete points along the orbit with interval Δt, so we may write $\psi_{km} = k\dot{\omega}_o j\Delta t + m\dot{\omega}_e j\Delta t$ with $j = 0, \ldots, N_p - 1$ and $N_p = T_r/\Delta t$, the total number of measurements during the mission. In the previous section we have seen that N_d is the number of nodal days during the mission period T_r and N_r is the number of orbit revolutions. So we have:

$$T_r = N_d \frac{2\pi}{\dot{\omega}_e} \quad \Rightarrow \quad \dot{\omega}_e = N_d \frac{2\pi}{T_r} = \frac{N_d}{\Delta t} \frac{2\pi}{N_p}$$

$$T_r = N_r \frac{2\pi}{\dot{\omega}_o} \quad \Rightarrow \quad \dot{\omega}_o = N_r \frac{2\pi}{T_r} = \frac{N_r}{\Delta t} \frac{2\pi}{N_p}$$

which yields for ψ_{km}

$$\psi_{km} = \frac{2\pi j}{N_p}(\beta_{km} N_r) \qquad\qquad j = 0, \ldots, N_p - 1$$

with $\beta_{km} = k + m\frac{N_d}{N_r}$. Note that, since $-L \leq k \leq L$, β_{km} may take negative values too.

Taking the inner product of two columns 1 and 2 of A means a summation over the index j over all measurement points $j = 0, \ldots, N_p - 1$ for different values l_1, l_2, m_1 and m_2. Let us take as an example the gradient V_{zz} and the case $l - m$ even. For a $\bar{C}_{lm}$ coefficient (indicated by an upper index cc) the element of the normal matrix now becomes (leaving out, for the moment, the factor $(l+1)(l+2)\Gamma_l$):

$$\begin{aligned} h^{(zz)cc}_{l_1 m_1 l_2 m_2} &= \frac{1}{\sigma^2} \sum_{j=0}^{N_p-1} \left(\sum_{k_1} \bar{F}^{k_1}_{l_1 m_1} \cos \frac{2\pi j}{N_p}(\beta_{k_1 m_1} N_r) \right) \left(\sum_{k_2} \bar{F}^{k_2}_{l_2 m_2} \cos \frac{2\pi j}{N_p}(\beta_{k_2 m_2} N_r) \right) \\ &= \frac{1}{\sigma^2} \sum_{k_1, k_2} \bar{F}^{k_1}_{l_1 m_1} \bar{F}^{k_2}_{l_2 m_2} \sum_{j=0}^{N_p-1} \cos \frac{2\pi j}{N_p}(\beta_{k_1 m_1} N_r) \cos \frac{2\pi j}{N_p}(\beta_{k_2 m_2} N_r) \, . \end{aligned}$$

According to the orthogonality properties of trigonometric series, the summation over j in this equation reduces to:

$$\begin{array}{lll} 0 & \text{if} & |\beta_{k_1 m_1}| \neq |\beta_{k_2 m_2}| \\ \frac{N_p}{2} & \text{if} & |\beta_{k_1 m_1}| = |\beta_{k_2 m_2}| \neq 0 \\ N_p & \text{if} & \beta_{k_1 m_1} = \beta_{k_2 m_2} = 0 \, . \end{array}$$

In case N_p is even the result of the j-summation may also be equal to $N_p/2$ if $|\beta_{k_1 m_1}| = |\beta_{k_2 m_2}| \neq \frac{N_p}{2}$ and equal to N_p if $\beta_{k_1 m_1} = \beta_{k_2 m_2} = \frac{N_p}{2}$, but in a realistic mission with millions of measurement points this situation will never occur. The frequency β_{km} equals $k + m\frac{N_d}{N_r}$ where k may take values between $-L$ and L and m may take values from 0 to L. This means that there may be several k, m combinations corresponding to the same frequency β_{km}. This also implies that we may have

$$|\beta_{k_1 m_1}| = |\beta_{k_2 m_2}| \quad \text{for} \quad k_1 \neq k_2 \quad \text{and} \quad m_1 \neq m_2 \tag{4.7}$$

cf. (Schrama, 1990). This situation will not lead to a block-diagonal structure of the normal matrix, as will be explained later. Now the situation $\beta_{k_1 m_1} = \beta_{k_2 m_2}$ implies:

$$\frac{N_d}{N_r} = -\frac{k_1 - k_2}{m_1 - m_2}$$

where (as explained before) the numbers N_d and N_r are relative primes (i.e. they have no common divisor, so their fraction cannot be simplified any more). The

denominator on the right hand side will never be greater than L. If we now ensure that N_r is always greater than L there is no possible combination of the numbers k_1, k_2, m_1, m_2 for which 4.7 holds.

Furthermore, we have to avoid the situation where for certain k, m combinations frequencies β_{km} occur which are 180° out of phase, i.e. $\beta_{k_1 m_1} = -\beta_{k_2 m_2}$. A similar reasoning as above will, for this case, lead to the requirement that N_r be greater than $2L$. If now $N_r > 2L$ we only have $\beta_{k_1 m_1} = \beta_{k_2 m_2}$ if $m_1 = m_2 \equiv m$ *and* $k_1 = k_2 \equiv k$, at least for $m \neq 0$. Furthermore, in this case we never have $\beta_{km} = 0$. If $m_1 = m_2 \equiv m = 0$ we have a special situation since in that case $\beta_{k_1 0} = -\beta_{k_2 0}\ \forall\, k_1 = -k_2$. Also, for $m = 0$ we have $\beta_{km} = 0$ for $k_1 = k_2 \equiv k = 0$. The expression for an element of the normal matrix can now be simplified to:

$$h_{l_1 l_2 m}^{(zz)cc} = \frac{T_r}{2\sigma^2 \Delta t} \sum_{k=-min(l_1,l_2)[2]}^{min(l_1,l_2)} F_{+-}$$

where for $m \neq 0$

$$F_{+-} = \bar{F}_{l_1 m}^{k} \bar{F}_{l_2 m}^{k}$$

and for $m = 0$

$$F_{+-} = \bar{F}_{l_1 m}^{k} \bar{F}_{l_2 m}^{k} + \bar{F}_{l_1 m}^{-k} \bar{F}_{l_2 m}^{k} \,.$$

The notation $min(l_1, l_2)$ means the minimum of the two numbers l_1, l_2, where l_1 and l_2 have the same parity as a result of $k_1 = k_2 = k$.

For an $\bar{S}_{lm}$ coefficient, $h_{l_1 l_2 m}^{(zz)ss}$, the same result appears except that the case $m = 0$ is not present. The cross products for one $\bar{C}_{lm}$ and one $\bar{S}_{lm}$ coefficient, $h_{l_1 l_2 m}^{(zz)cs}$ and $h_{l_1 l_2 m}^{(zz)sc}$, are always zero due to the orthogonality properties of trigonometric functions. In an analogous manner the situation $l - m$ odd is treated. For the other gradients similar derivations hold, using the expressions from tables 4.1 and 4.2. Now it also becomes clear why we had to derive new expressions for the xy and yz derivatives. The $\sin \omega_o$ term in the denominator of the old expressions of those two gradients would have destroyed the orthogonality properties of the trigonometric functions. In general, the expressions for the elements of the normal matrix are:

$$h_{l_1 l_2 m}^{cc} = \frac{\Gamma_{l_1} \Gamma_{l_2} T_r}{2\sigma^2\, \Delta t} \sum_{k=-min(l_1,l_2)[2]}^{min(l_1,l_2)} Z_{l_1 l_2 m k} \qquad (4.8)$$

$$h_{l_1 l_2 m}^{ss} = h_{l_1 l_2 m}^{cc} \quad \text{(except for } m = 0\text{)}$$

$$h_{l_1 l_2 m}^{cs} = h_{l_1 l_2 m}^{sc} = 0$$

with $Z_{l_1 l_2 m k}$ taken from table 4.3 (with factors like $(l+1)(l+2)\Gamma_l$ restored).

Table 4.3 *Expressions used for the elements of the normal matrix.*

	$Z_{l_1 l_2 mk}$
xx	$(l_1+1+k^2)(l_2+1+k^2)F_{+-}$
xy	$k^2 F^*_{+-}$
xz	$(l_1+2)(l_2+2)k^2 F_{+-}$
yy	$((l_1+1)^2-k^2)((l_2+1)^2-k^2)F_{+-}$
yz	$(l_1+2)(l_2+2)F^*_{+-}$
zz	$(l_1+1)(l_1+2)(l_2+1)(l_2+2)F_{+-}$

In this table it is for $m \neq 0$

$$F^*_{+-} = \bar{F}^{k\,*}_{l_1 m} \bar{F}^{k\,*}_{l_2 m}$$

and for $m = 0$

$$F^*_{+-} = \bar{F}^{k\,*}_{l_1 m} \bar{F}^{k\,*}_{l_2 m} + \bar{F}^{-k\,*}_{l_1 m} \bar{F}^{k\,*}_{l_2 m} .$$

Note that the summation over k for the xy and yz gradient runs from $-min(l_1 - 1, l_2 - 1)$ to $min(l_1 - 1, l_2 - 1)$.

If several gradients are measured simultaneously at each observation point with equal variance and if they are used as independent measurements in the adjustment, the elements of the normal matrix can be found by summing expressions like 4.8 for each individual gradient. For example, if the gradients V_{yy}, V_{yz} and V_{zz} are measured, the expression for the elements of the normal matrix is equation 4.8 with in this case

$$\begin{aligned} Z^{(yy,yz,zz)}_{l_1 l_2 mk} = &\{ [(l_1+1)^2 - k^2][(l_2+1)^2 - k^2] + \\ &(l_1+1)(l_1+2)(l_2+1)(l_2+2) \} F_{+-} + \\ &(l_1+2)(l_2+2)F^*_{+-} . \end{aligned}$$

If some combination of gradients is used as a single observation, for example $2V_{yy} + V_{zz}$, the expression for the elements of the normal matrix is found by using equation 4.8 with

$$Z^{(2yy+zz)}_{l_1 l_2 mk} = [l_1(l_1+1) - 2k^2][l_2(l_2+1) - 2k^2]F_{+-} .$$

We see that with this method we can easily investigate various situations with respect to available measurement configurations.

But a much more interesting consequence of the derivations above is that, as a result of the orthogonality relations of the trigonometric series, through a suitable

choice of the ordering of the unknown potential coefficients, the normal matrix will attain a block–diagonal structure (Colombo, 1987) which is much more easy to invert. First of all, we have seen that the normal matrix elements connecting an arbitrary $\bar{C}_{lm}$ and an arbitrary $\bar{S}_{lm}$ coefficient are all zero. This means that if we order the coefficients in such a way that all $\bar{C}_{lm}$ coefficients come first, the normal matrix is divided into four blocks, the upper right and lower left one (the off–diagonal blocks) contain only zero's. Furthermore the two diagonal blocks are exactly the same. Also, since only elements for equal m differ from zero (at least if $N_r > 2L$), ordering the unknowns according to m causes the two main diagonal blocks to be divided into smaller blocks, one for each m. The size of these smaller blocks decreases from $L+1$ for $m = 0$ to 1 for $m = L$. Finally, as a consequence of the vanishing of the summation terms for which $k_1 \neq k_2$ (again only if $N_r > 2L$) only elements for which k is the same remain, which in turn implies that l_1 and l_2 have to have the same parity. So, the m–blocks are again divided into four smaller blocks if for each m the l are ordered according to their parity, first all even l and then all odd l. Of those four blocks only the diagonal blocks differ from zero, see figure 4.1. The size of these blocks is half the size of the m–blocks. The inversion problem is now reduced from that of a matrix of size L^2 to one of $2L$ times a matrix with a size of at most $\frac{1}{2}L$, which means an enormous reduction of computation time. The block–diagonal structure also implies that the error analysis can be carried out for certain groups of coefficients independently.

4.2.2 Presentation

With the expressions for the elements of the normal matrix N which are derived in the last section, we are now in a position to compute in an easy way the a–posteriori variances of the estimated potential coefficients. The assumptions stated in the beginning of section 4.2 may, to some extent, put limitations on the usefulness of the method. Nevertheless, the method is still very useful to gain a better understanding of the gradiometric analysis problem and at the same time shows possible problem areas.

The required a–posteriori variances are found to be the diagonal elements of the inverse normal matrix N^{-1}. A very common way of representation of these variances is by means of degree variances. The *error degree variance* (of the formally propagated error) is

$$\begin{aligned} \sigma_l^2 &= \sigma_{l0}^2(\bar{C}) + \sum_{m=1}^{l} \sigma_{lm}^2(\bar{C}) + \sigma_{lm}^2(\bar{S}) \\ &= \sigma_{l0}^2 + 2\sum_{m=1}^{l} \sigma_{lm}^2 \,, \end{aligned} \tag{4.9}$$

where use is made of the equality of the error variances for the $\bar{C}_{lm}$ and $\bar{S}_{lm}$ coefficients and where σ_{lm}^2 is a diagonal element of the inverse normal matrix N^{-1}

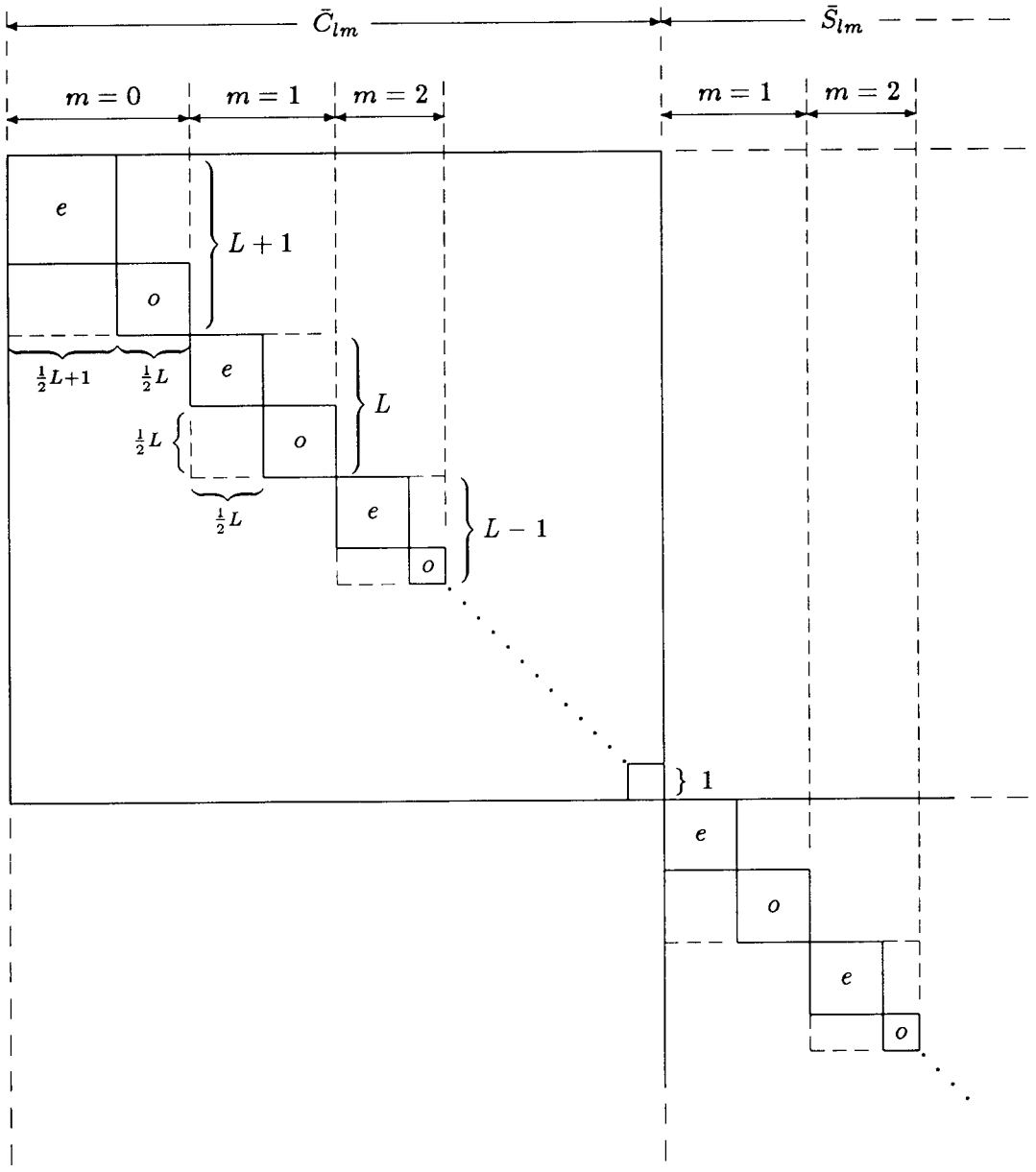


Figure 4.1 *Block-diagonal structure of the normal matrix. The notation "e" means even l and "o" odd l.*

for a specific value of l and m. From equation 4.9 we compute the error r.m.s. per coefficient per degree as

$$\bar{\sigma}_{lm} = \sqrt{\frac{\sigma_l^2}{2l+1}}\,, \tag{4.10}$$

the square of which is called *error degree-order variance*. Degree variances and degree–order variances can also be computed for the potential coefficients themselves, in which case they are called respectively *signal degree variances* and *signal degree–order variances*. They are defined as

$$c_l = \bar{C}_{l0}^2 + \sum_{m=1}^{l} (\bar{C}_{lm}^2 + \bar{S}_{lm}^2) \tag{4.11}$$

and

$$c_{lm} = \frac{c_l}{2l+1}\,. \tag{4.12}$$

For computational purposes, the potential coefficients $\bar{C}_{lm}$ and $\bar{S}_{lm}$ in eq. 4.11 may be taken from some a–priori known potential coefficient model, e.g. OSU86F (Rapp and Cruz, 1986). Another possibility is to compute the c_l directly using some degree variance model, like Kaula's rule (Kaula, 1966) or the Tscherning/Rapp model (Tscherning and Rapp, 1974). Note that, since the potential coefficients $\bar{C}_{lm}, \bar{S}_{lm}$ are dimensionless, so are σ_l^2, $\bar{\sigma}_{lm}$, c_l and c_{lm}.

Whereas the degree variance of some function (either signal or error) represents the power of the function per degree, one is also often interested in the total power of the function over all degrees. For example, the total signal power for potential coefficients is $\sum_{l=0}^{\infty} c_l$ with the c_l taken from equation 4.11. This total power also represents the average square value (or norm) of the signal over the unit sphere, cf. (Heiskanen and Moritz, 1967). The *global r.m.s.* is the square root of the total power. In the same way we may compute from the error degree variances (equation 4.9) the global error r.m.s. for potential coefficients by summing the σ_l^2 over all degrees. The mission goals for gradiometry are expressed in terms of global error r.m.s. values for gravity anomalies or geoid undulations. They can easily be computed from the global error r.m.s. for potential coefficients by multiplying the latter with the eigenvalues λ_l of the linear operator connecting the respective quantity with the potential coefficients (Rummel, 1991). The global error r.m.s. becomes

$$\sqrt{\sum_{l=0}^{\infty} \lambda_l^2 \sigma_l^2}$$

with for gravity anomalies $\lambda_l = \frac{GM}{R^2}(l-1)$ and for geoid undulations $\lambda_l = R$. The global r.m.s. of each arbitrary other gravitational quantity can be computed in a similar manner, as long as it is linearly related to the potential. Here we will only show the two mentioned quantities, namely the surface gravity anomalies and the geoid heights.

We see that the summation above includes all degrees from zero to infinity. The reason for this is that the gravitational potential is considered an element of an infinite dimensional Hilbert space of functions defined outside a convergence sphere which is approximately the surface of the earth. The spectrum of such functions includes all frequencies from zero up to infinity. In our error analysis, based on the availability of a finite number of measured samples, we can only estimate a limited part of this spectrum, e.g. up to some maximum degree L. As a consequence, the gravitational field can never be perfectly reproduced from the measurements, even in case of perfect measurements. There always remains a part of the spectrum which is neglected in the analysis. This neglected signal is called *omission error*. In case of regular sampling, it contains primarily the high frequency signal part for degrees $l > L$. As a result, the global error r.m.s. should consist of two parts: the propagated error from the analysis for degrees $l \leq L$ (called the *commission error*) and the omission error (neglected signal for degrees $l > L$). Since we do not know the true spectrum of the gravitational field, and since it is impossible in practice to really include all degrees up to infinity, the omission part is computed only up to some limited maximum degree $lmxs$ using a model representing the true spectrum. Thus, the global error r.m.s. values for gravity anomalies or geoid heights are computed as

$$\sqrt{\sum_{l=0}^{L_t} \sigma_l^2 \beta_l^2 \lambda_l^2 + \sum_{l=L_t+1}^{lmxs} c_l \beta_l^2 \lambda_l^2} \tag{4.13}$$

where the first summation indicates the commission error and the second the omission error, and where

β_l = $\frac{1}{1-\cos\psi_c} \frac{1}{2l+1} [P_{l-1}(\cos\psi_c) - P_{l+1}(\cos\psi_c)]$ smoothing operator

ψ_c : radius of a spherical cap with equal area as an equiangular block of size θ_c at the equator

λ_l = $\begin{cases} \frac{GM}{R^2}(l-1) & \text{for gravity anomalies} \\ R & \text{for geoid undulations} \end{cases}$

L_t : degree of truncation of the commission error which is less or equal to the maximum degree L of the performed error analysis

$lmxs$: truncation degree for the omission error .

The reason for including the smoothing operator β_l, see e.g. (Meissl, 1971), will be explained in section 4.2.7. A truncation degree for the commission error L_t smaller than L may be chosen, for example, if the error in the estimated coefficients for degrees $l > L_t$ exceeds 100 % of the signal, i.e. if for those degrees the signal to noise ratio $\sigma_l^2/c_l > 1$, cf. (Colombo, 1989) or (Rapp, 1989). Choosing L_t in this way is in fact comparable to "ideal" low-pass filtering. However, a more optimal way for

computing the global r.m.s. is using a minimum variance filter, like a Wiener filter, e.g. (Papoulis, 1965). In that case the expression becomes:

$$\sqrt{\sum_{l=0}^{L} \beta_l^2 \lambda_l^2 \left(\sigma_l^2 W_l^2 + c_l (1 - W_l)^2\right) + \sum_{l=L+1}^{lmxs} \beta_l^2 \lambda_l^2 c_l} \tag{4.14}$$

with

$$\text{Wiener filter:} \qquad W_l = \frac{c_l}{c_l + \sigma_l^2} \, .$$

The Wiener filter is derived as to ensure a *minimum total error* (commission + omission). For degrees where the signal power largely exceeds the noise power, the filter will approximately attain the value 1. In the estimation procedure this would imply that the data at these degrees are almost completely used. For the global error computation it therefore means that the error for those degrees comes almost exclusively from the propagated error σ_l^2 from the gradiometric analysis. If the noise power, however, is larger than the signal power, the filter tends to zero. For the global error this means that for those degrees the contribution comes largely from a–priori available gravity information, in our case the signal degree variances c_l, which can be seen from equation 4.14. In this equation, the term including σ_l^2 represents the commission part and both other terms the omission part. This means that, if one uses a Wiener filter, one may also have an omission error part for degrees below L, which in fact results from the imperfection of the measurements. We will, in our computations, exclusively use the latter equation 4.14.

The above derivations shall now be used to study the error behaviour of gradiometric experiments. Thereby we shall start with the most ideal situation. It is characterized by a gradiometer instrument which measures all six components of the gradient tensor (a so–called full tensor gradiometer). The full measurement signal (the complete spectrum of measurement frequencies, here indicated by the index k) is used with equal variances for all frequencies. Considering the orbit, an ideal situation is a polar orbit ($I = 90°$) in order to obtain complete global coverage. Afterwards we shall consider a number of restrictions (band limitation and non–polar orbits) and finally we discuss a scenario that resembles as closely as possible the current plans of the Aristoteles mission.

4.2.3 Ideal case

We consider here a full tensor gradiometer, which measures all six components of the gradient tensor. To obtain complete global coverage we assume the inclination of the orbit to be 90°. Finally we consider the instrument *not* to be band–limited. The latter implies that we include all measurement frequencies, low and high enough to recover the potential up to some specified maximum degree, in the analysis, and that we assume them to have equal accuracy. A discussion on a band–limited instrument will be given in the next section. For the moment we simply take β_{min} to be zero and

$\beta_{max} > 254$, which is high enough to recover potential coefficients up to $L = 240$, see section 4.2.4. These, and other characteristics of the ideal case gradiometer mission are listed in table 4.4.

Table 4.4 *Characteristics of the ideal case gradiometer mission.*

inclination	$I = 90°$
components	$V_{xx}, V_{xy}, V_{xz}, V_{yy}, V_{yz}, V_{zz}$
no band limitation	$\beta_{min} = 0$
satellite height	$h = 200$ km
sampling interval	$\Delta t = 4$ s
mission duration	$T_r = 6$ months
error spectrum	0.01 E/$\sqrt{\text{Hz}}$ white noise

In figure 4.2 the (dimensionless) error r.m.s. values for potential coefficients computed with equation 4.10 are shown for 10 different cases: for each of the six tensor components individually, for several combinations of components (namely all six together, $\{V_{yy}, V_{yz}, V_{zz}\}$ and $\{V_{yy}, V_{zz}\}$) and for the quantity $2V_{yy} + V_{zz}$. The latter quantity is chosen because it will be used in the potential coefficient recovery in section 4.3. The ordinates in these figures have a logarithmic scale. For all situations the overall pattern of the graphs is the same: after an initial decrease of the r.m.s. value (i.e. an increase of the a–posteriori accuracy) the lowest point is reached around degree $l \approx 60$, after which a steady increase occurs. This overall pattern can be explained by remembering the expressions for the elements of the normal matrix from table 4.3. As the result of two differentiations all the gradients are multiplied by a factor proportional to at least l^2. For the signal this means a higher contribution for higher degrees, as to be expected for second–order derivatives. For the a–posteriori error (inverse of the normal matrix) this means a decrease with increasing degree l. However, as already indicated in section 3.2.1, the natural attenuation effect of the gravitational potential with height is also present. For the gradients this factor is $(R/r)^{l+3}$. For the elements of the normal matrix this factor is again squared. For higher degrees, the signal will decrease, and the error r.m.s. will therefore increase. On a logarithmic scale this attenuation factor will result in a straight line, starting left below and reaching for the upper right side (see figure 4.3). Eventually this effect will be dominant, so the r.m.s. plots are likely to approximate this straight line for higher degrees, as can be seen in the figure.

Combination of the effects described above results in the lowest error for the part of the spectrum somewhere between degrees 30 and 100, slightly dependent on the observation type, making gradiometry especially suitable for determination of

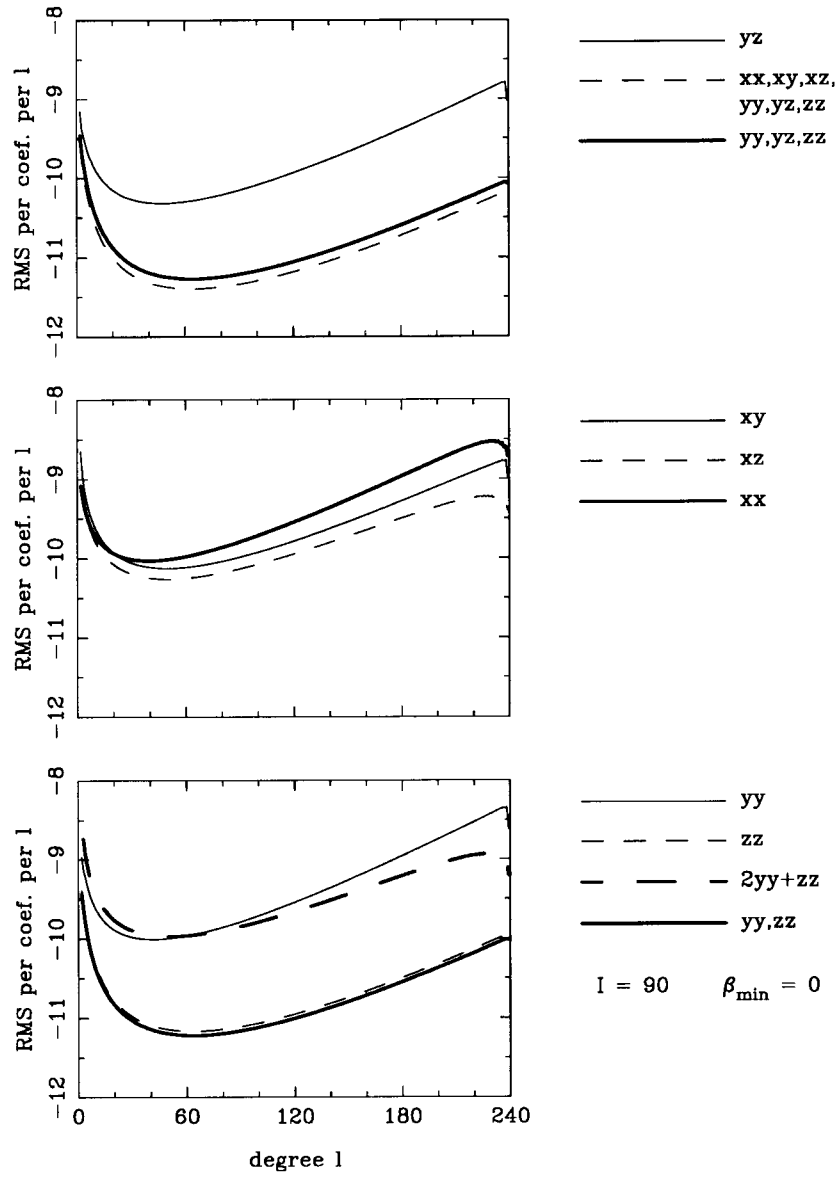


Figure 4.2 *Propagated error from gradiometry for various situations. Vertical scale is logarithmical. For all cases the satellite's altitude was 200 km, the mission duration T_r 6 months, the sample interval Δt 4 seconds and a 0.01 $E/\sqrt{Hz}$ white noise error spectrum was assumed.*

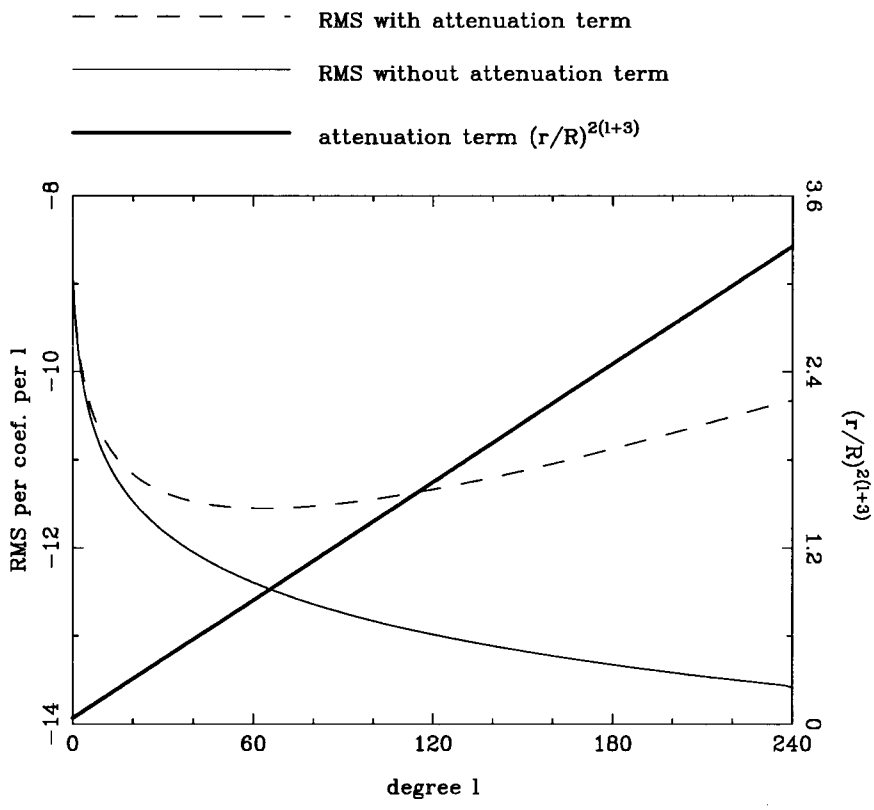


Figure 4.3 *Influence of attenuation factor on error r.m.s. plots. The figure shows error r.m.s. from V_{zz}. Left vertical scale is for r.m.s values, right vertical scale for attenuation factor. Both scales logarithmical.*

these degrees.

From figure 4.2 we see that the largest error is obtained for the V_{xx} and V_{yy} components, followed by that from respectively $2V_{yy} + V_{zz}$, V_{xy}, V_{xz} and V_{yz}. Out of the seven individual observations shown in the figure, V_{zz} gives the best result. Redundancy, however, ensures even better results for combinations of observations, the combination $\{V_{yy}, V_{zz}\}$ being slightly better than V_{zz} alone, followed by the combination $\{V_{yy}, V_{yz}, V_{zz}\}$ and the best results for the full tensor combination $\{V_{xx}, V_{xy}, V_{xz}, V_{yy}, V_{yz}, V_{zz}\}$. Note that for the six individual gradients, the level of the a–posteriori error relative to one another is not necessarily the inverse of the signal power level, cf. figure 3.4.

Once we have computed the r.m.s. values for a specific inclination and β_{min}, it is very easy to derive from them equivalent values for various altitudes, with various mission durations, sampling intervals and measurement precisions. We do not need to invert a new normal matrix in these cases. This can be understood as follows. Columns of the design matrix A are multiplied by the downward continuation factor $(R/r)^{l+3}$, cf. eq. 4.1, where $r = R + h$. So A can be written as $A = A'\,D$, where D is a diagonal matrix containing only the factors $(R/r)^{l+3}$ and A' equals A without these factors. Then, since $N = A^T Q^{-1} A$, we have

$$N = D^T A'^T Q^{-1} A'\, D \equiv D N' D$$

with $N' = A'^T Q^{-1} A'$. Computing N'^{-1} first, it can be scaled by pre and post multiplication with D^{-1} to obtain the results at satellite altitude, since $N^{-1} = D^{-1} N'^{-1} D^{-1}$. Furthermore, from eq. 4.8 we see that each element of N is multiplied by a factor $T_r/\sigma^2 \Delta t$. The square root of this factor can be included in D, and can thus be left out of the matrix inversion. If we now like to compute error r.m.s. values for other, arbitrary values of $h, \Delta t, T_r$ or σ^2, we do not need to carry out a new normal matrix inversion. The inverse N'^{-1}, computed only once, is simply scaled with an appropriate diagonal matrix D. Note that for combinations of gradients, like $\{V_{yy}, V_{zz}\}$, the scaling with σ^2 does not always work. If different gradients are given different σ's, a new inversion has to be established.

For the graphs of figure 4.2 scaling with T_r, σ^2 or Δt means a uniform shift of the complete curve. A different altitude means that the straight line representing the attenuation factor (see above) obtains another slope with subsequent effect on the r.m.s. curve. In figure 4.4 the effects mentioned above are illustrated for the error r.m.s. from V_{zz}.

As described in the previous section the error degree variances can be used to compute global error r.m.s. values for gravity anomalies and geoid undulations. The influence of the individual mission parameters $T_r, \Delta t, h$ and σ on this global r.m.s. is not so obvious any more. For a full tensor gradiometer, measuring all six tensor components, and for the case that all these six components are used in the error analysis with equal weights, this global r.m.s. is shown in figures 4.5 and 4.6.

Figure 4.5 shows the global r.m.s. for three different mission durations as a function of height. (Note that the straight lines are just added for clearness. They

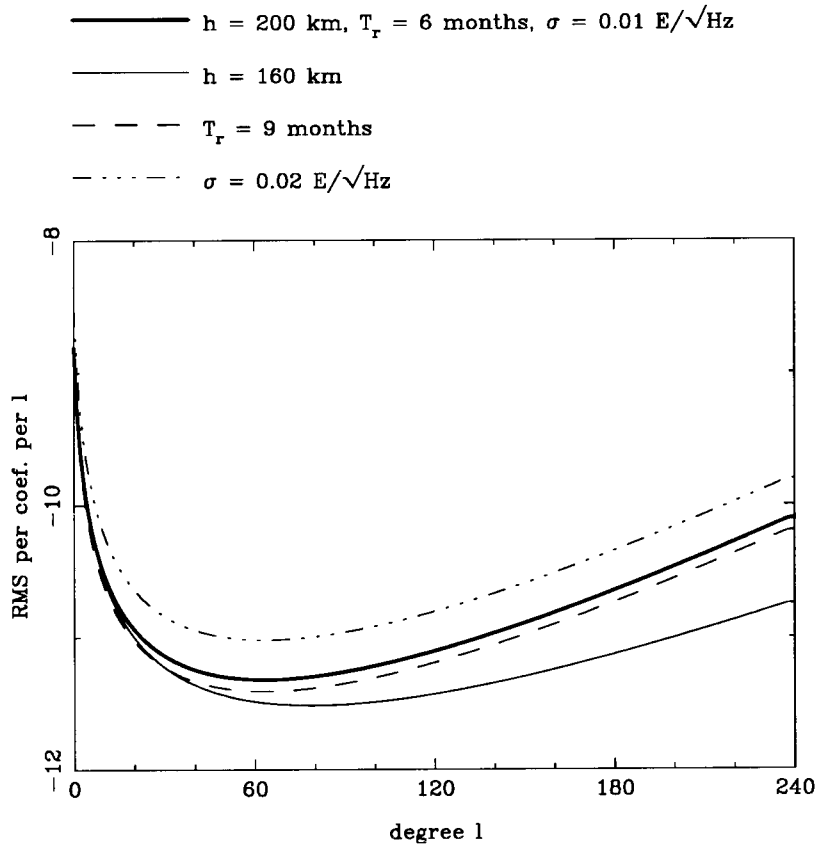


Figure 4.4 *Effect of a different altitude, mission duration and measurement precision on the a-posteriori error r.m.s. values from* V_{zz}.

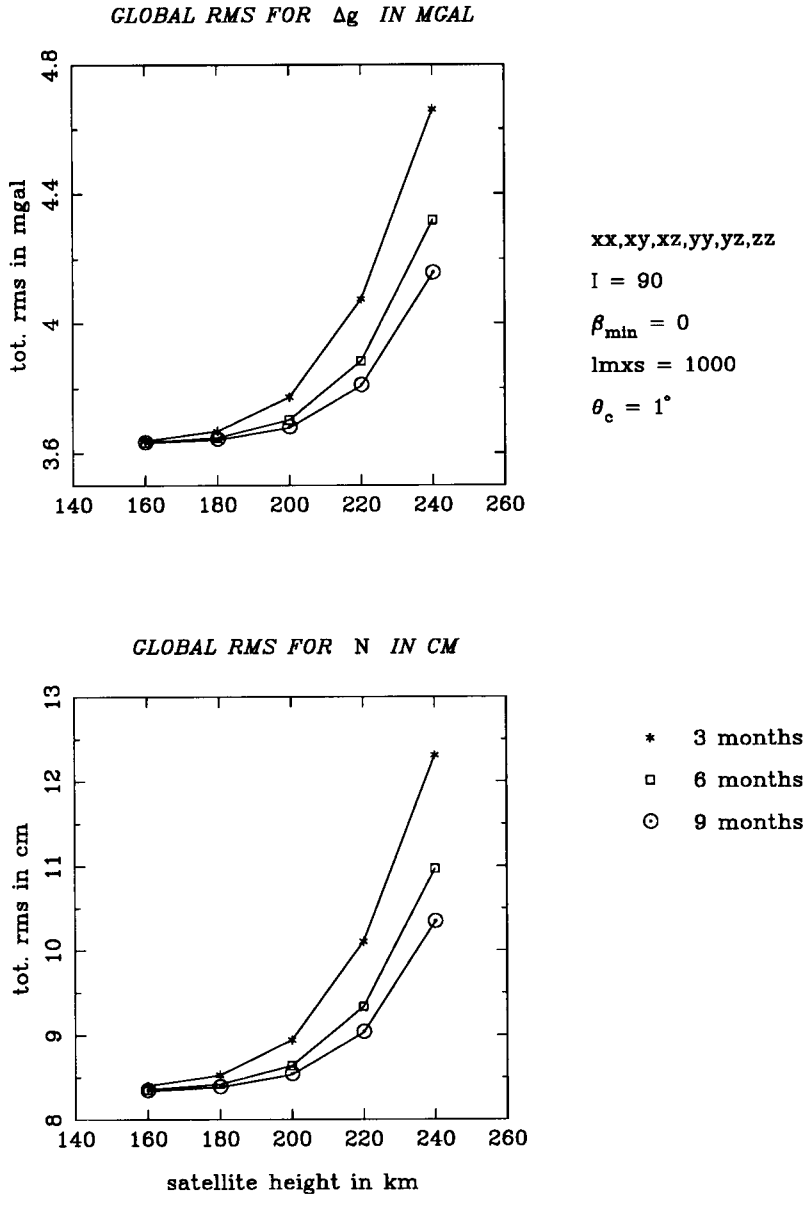


Figure 4.5 *Global r.m.s. (commission + omission error) for surface gravity anomalies Δg and geoid undulations N for three different mission durations as function of the satellite's altitude. Maximum degree for omission error was 1000. Numbers represent $1° \times 1°$ block averages.*

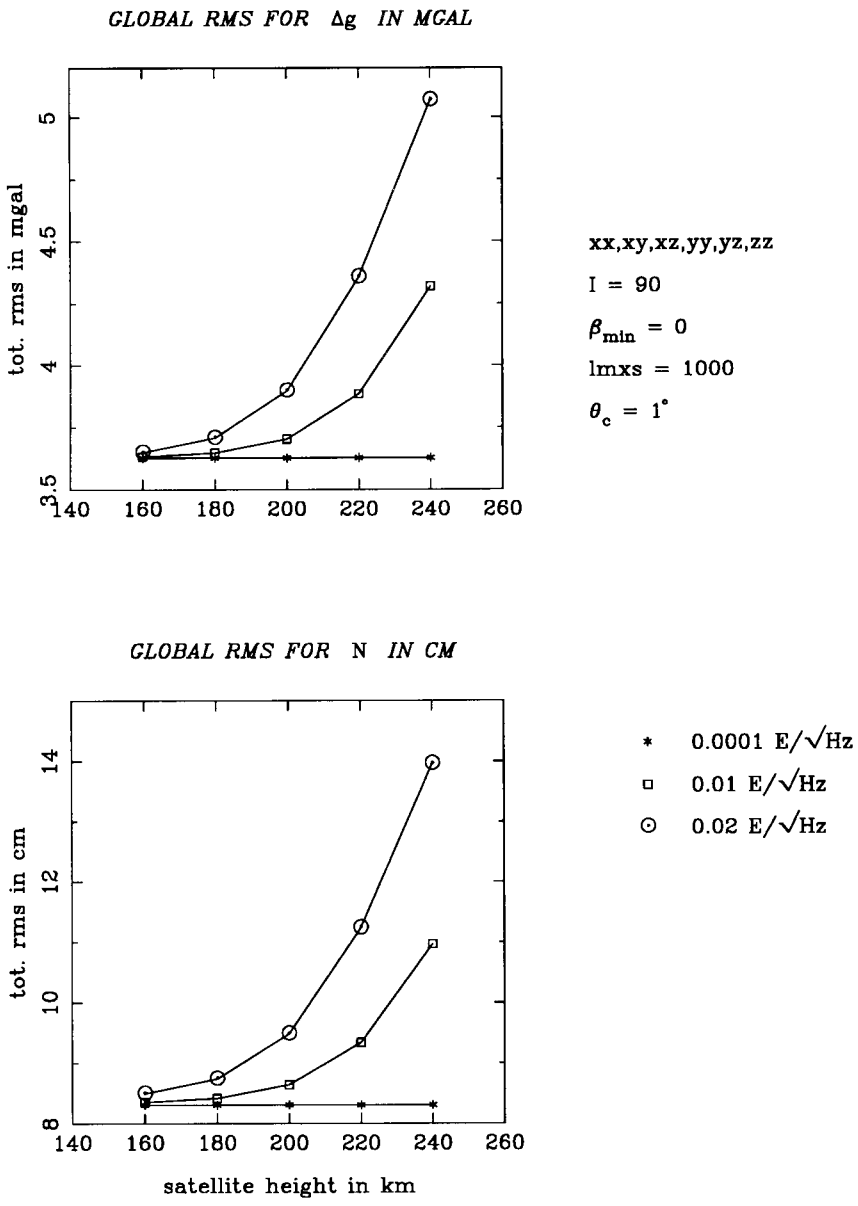


Figure 4.6 *Global r.m.s. (commission + omission error) for surface gravity anomalies Δg and geoid undulations N for three different measurement precisions as function of the satellite's altitude. Maximum degree for omission error was 1000. Numbers represent $1° \times 1°$ block averages.*

do not represent functional values since only the nodes (5 different altitudes) were really computed.) The numbers in this and the following figures of this type were all computed using equation 4.14, including the Wiener filter, where for the omission error (computed using the Tscherning/Rapp degree variance model) a maximum summation value $lmxs = 1000$ was taken. The numbers furthermore represent $1° \times 1°$ block averages since the smoothing operator β_l was included with $\theta_c = 1°$. The influence of $lmxs$ and θ_c on the results will be discussed in section 4.2.7. The most remarkable fact from this figure is that at the lowest altitude (160 km) tripling the mission duration gives no substantial improvement of the total r.m.s., cf. (Rapp, 1989).

Figure 4.6 is of the same kind as the previous one, but now results are shown for three different measurement precisions. The case $\sigma = 0.0001$ E/$\sqrt{\text{Hz}}$ refers to the anticipated measurement precision of the superconducting gradiometer under development at NASA (Paik and Richard, 1986). What can be seen from this figure is that with this kind of high measurement accuracy, satellite altitude becomes of less importance, whereas for the lower accuracies it still plays an important role. At 160 km, however, all σ's give nearly the same result, cf. (Rapp, 1989). Apparently there is some lower limit for the total r.m.s. To understand this, we have to bear in mind that the r.m.s. values represent the total error, commission *and* omission error (see the previous section). Increasing the measurement precision changes the commission error part. For example, the commission error for gravity anomalies for the three increasing measurement precisions (0.02, 0.01 and 0.0001 E/$\sqrt{\text{Hz}}$) at 160 km altitude is respectively 0.28 mgal, 0.14 mgal and 0.0014 mgal. As for the omission error, only the contribution for degrees below L is changed (due to the Wiener filter) when increasing the measurement precision, whereas the part above L is not changed at all. The total error therefore has some constant base level determined by the omission error, especially the part above L. This part has to be decreased in order to fully benefit the improved measurement accuracy, which implies that we have to increase L. The maximum value for L is determined by the sampling density of the measurements, which is represented in our case by the combination of mission duration T_r and sampling interval Δt. So we see that in the end improving the measurement accuracy will not give better results if at the same time not also the sampling density is increased. A same reasoning explains the limited influence of the mission duration at 160 km altitude in figure 4.5.

In any case, both figures show that in the present ideal case the goals described in chapter 2 (5 mgal accuracy for gravity anomalies and 10 cm for geoid undulations with a resolution of better than 100 km) can easily be met, at least for a full tensor gradiometer. This conclusion is, however, a theoretical one since in reality the situation may and probably will not be so ideal. Nevertheless, the previous discussion gives more insight in the way in which various parameters influence the process and it shows the maximum obtainable result from gradiometry, of course under the present assumptions.

Singular normal matrix sub–blocks

When trying to invert the sub–blocks (m–blocks) of the normal matrix, it appears (and it will appear for other cases in subsequent sections) that some blocks are singular and can therefore not be inverted. The reason for this might be either a physical one or a numerical one. An example of a physical cause underlying such singularities can be found for the components V_{xy} and V_{yz} (both of which contain one differentiation in the cross–track direction). For a perfect polar orbit ($I = 90°$) zonal coefficients cannot be estimated from these components, the cross–track direction being always orthogonal to the direction of variation of the zonal coefficients (along the meridians). This is reflected in the model by means of vanishing of the cross–track inclination functions $\bar{F}_{lm}^{k\,*}$ for $m = 0$. Consequently, the $m = 0$ blocks for these components are singular.

We may also encounter numerical singularities. Especially for low orders (which constitute the largest sub–blocks, since the degrees l contained in them run from m to L) the difference between the largest and smallest eigenvalue may become very large, so the block will be ill–conditioned (see also section 4.2.6).

In either case, we do not obtain propagated error variances for the potential coefficients belonging to those blocks. Or stated otherwise, the potential coefficients belonging to the singular blocks apparently cannot be estimated from the observations. However, when computing error degree variances using equation 4.9, a summation is done over *all* orders m. Leaving out certain orders, due to the singularities described above, would therefore result in incorrect values for the degree variances. We would have, so to say, assumed zero error for those orders, where perhaps an infinite error would have been more appropriate. Therefore, the error r.m.s. plots included in this work have to be interpreted very carefully, always bearing in mind which m–blocks are left out. To this extent, table 4.5 shows which blocks appeared to be singular and were omitted from the analysis. For later reference, we included in this table all the cases yet to follow in subsequent sections.

What can we do about this? Probably the best (but certainly not the most pleasant) answer would be *not* to compute degree variances at all. The problems above first arise when computing the degree variances, whereas there is, of course, nothing wrong with the propagated error variances for each individual coefficient themselves, at least as far as they *can* be estimated (i.e. the non–singular blocks). There are, however, some reasons for using degree variances. First of all, they are generally found to be very illustrative, giving a good picture of the gravitational field and its main spectral characteristics. Secondly, if we do not compute degree variances, we are left with a very large amount of individual error variances to be compared for several cases. Whereas the number of degree variances "only" amounts to L (the maximum degree of the analysis), the number of individual coefficients is $(L + 1)(L + 2)$ minus the non–estimable ones. It is hard to make good pictorial or graphical representations for so much coefficients, at least for high maximum degrees, like $L = 240$ in our case.

An alternative would be to compute *order variances*. Then orders for which the sub–blocks of the normal matrix are singular appear in the graphs simply as gaps, but at least they would not interfere with other orders. But from table 4.5 we see that often only the even *or* odd degrees for a certain m appear to be singular. Then we are faced again with a similar problem as above. Furthermore, one is not used to interpret order variances, so we will not show this alternative here.

We already suggested above that instead of zero error, an infinite error for the coefficients of the singular blocks would perhaps be realistic. So we could try, in some way, to fill in the shortages occurring due to the singular blocks. For example, one could regard the non–estimable coefficients as contributing to the omission error and thus insert for them the signal degree–order variances from some degree variance model or the value of the coefficients from a potential coefficient model. But since for large parts of the spectrum the signal is much higher (partly by several orders) than the error, the results would become extremely bad (very high error r.m.s. values). Another possibility would be to insert the error variances from some existing potential coefficient model. But especially for the low degrees (where the influence of the singular blocks is relatively large) the error of the existing models is much smaller than that from gradiometry. Our results would therefore hardly change. Even a multiplication of the model error variances with a factor 3 before adding them to the degree variances did not have much influence.

Furthermore, both methods, inserting model signal or error variances for the singular blocks, lead to the problem of how to interpret the corresponding observational model belonging to such a strange combination of (error) variances. Concerning the additional error variances from some model, one could view upon them as a–priori information added to the estimation procedure. But those model error variances themselves are the result of another, former, estimation process, leading to a strange mixture of data, models and methods.

In the remaining of this chapter, we will therefore not pay any special attention to the phenomenon of the singular sub–blocks. For reference, the singular blocks are listed in table 4.5. In any case, the influence of the singular blocks on the computed degree variances is not so large, as might be illustrated by the following test. We computed the error r.m.s. values for the case $I = 90°$ and $\beta_{min} = 0$ for the six tensor components leaving out exactly the same orders for all components. From table 4.5 one can see that this means leaving out $m = 0$ (both even and odd degrees). Apart from a slight change in the individual r.m.s. curves, the ratios between the error r.m.s. curves, as compared to figure. 4.2, did not alter. The error from V_{xx} and V_{yy} remains largest, followed by that from V_{xy}, then V_{xz} and V_{yz} and finally V_{zz}.

4.2.4 Band limitation

In the previous section we assumed for the gradiometric measurements a 0.01 $\mathrm{E}/\sqrt{\mathrm{Hz}}$ white noise error spectrum over the full measurement bandwidth $0 \leq |\beta_{km}| \leq \infty$. In that case all measurements $V_{ij}(t)$ in the time domain will have equal precision, i.e.

Table 4.5 *Singular sub-blocks of the normal matrix. "e" means even degrees, "o" means odd degrees, + means both even and odd degrees. For the combinations* $\{V_{xx}, V_{xy}, V_{xz}, V_{yy}, V_{yz}, V_{zz}\}$, $\{V_{yy}, V_{yz}, V_{zz}\}$ *and* $\{V_{yy}, V_{zz}\}$ *never singular* m*-blocks appear.*

		order m														
I	β_{min}	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
							V_{xx}									
90°	0															
90°	4															
92°.3	0															
95°.3	0															
92°.3	4															
							V_{xy}									
90°	0	+														
90°	4	+	e	e	e	e		e		e						
92°.3	0															
95°.3	0															
92°.3	4	e	e	e	e	e										
							V_{xz}									
90°	0															
90°	4						o		o							
92°.3	0															
95°.3	0	e														
92°.3	4															
							V_{yy}									
90°	0															
90°	4				+	+	o	+	o		o		o		o	
92°.3	0															
95°.3	0	+														
92°.3	4	+														
							V_{yz}									
90°	0	+														
90°	4	+	e	e	e	e	+	e		e		e		e		e
92°.3	0															
95°.3	0	e	e													
92°.3	4	e	e	e	e	e										
							V_{zz}									
90°	0															
90°	4					+	o	e	o		o		o		o	
92°.3	0															
95°.3	0		e													
92°.3	4															
							$2V_{yy} + V_{zz}$									
90°	0															
90°	4														o	
92°.3	0															
95°.3	0	e														
92°.3	4															

the covariance matrix Q is a scaled unit matrix. As a result of Parseval's relation this measurement precision will in our case be 0.01 E. In reality the error spectrum of a gradiometer will only be white in a limited band, between some lower band limit β_{min} and an upper limit β_{max}. The measurement frequencies $|\beta_{km}|$ below β_{min} and above β_{max} will either not be present in the signal at all due to the sampling rate or will be distorted too much due to instrumental and environmental error sources (*coloured noise*).

An upper band limit β_{max} is caused by the sampling rate with which the measurements are taken along the satellite's orbit. For Aristoteles this sampling rate is foreseen at 4 s, so that the highest measurement frequency present is 0.125 Hz (twice the sampling rate or *Nyquist rate*). At an altitude of 200 km, the orbital period of the satellite is approximately 5310 s, so that $\beta_{max} = 5310 \times 0.125 \approx 664$ c.p.r. We limit our error analysis to $L = 240$, so also $|k|$ and m are limited to L. As a result, the highest measurement frequency included in the analysis is $|\beta_{km}| = 254$ c.p.r. which is well below β_{max}. An upper band limit as discussed here will therefore not cause any problems. Note that the choice of $L = 240$ means that frequencies above 254 c.p.r., although present in the signal, are not used in the analysis.

A lower band limit is the result of instrumental and environmental influences. For Aristoteles instrument stability (especially thermal stability) can only be maintained over a maximum period of 200 s. This limits the white noise error band to measurement frequencies with a period not exceeding 200 s, i.e. the lowest measurement frequency is 0.005 Hz, or $\beta_{min} \approx 27$ c.p.r. Frequencies with a period exceeding 200 s ($|\beta_{km}| < 27$) will be present in the signal, but with degraded accuracy. A $1/\beta$ error behaviour for these frequencies can be assumed for such coloured noise situation (Schrama, 1990). Therefore, an absolute lower band limit of 27 c.p.r. (with all frequencies below this limit removed from the analysis) will not be discussed here. Finally, non–gravitational orbital effects disturb the spectrum, mainly below 4 c.p.r. (see section 4.2). Removing these frequencies from the analysis can be considered appropriate to account for these effects. One could think of this as assuming an infinite error for these frequencies. A β_{min} of 4 c.p.r. should be considered as a minimum, also for the coloured noise situation.

In this section we will investigate the consequences of a lower band limit of 4 c.p.r., thereby assuming white noise for the full error spectrum above this limit. A coloured noise situation, with a $1/\beta$ error behaviour between 4 and 27 c.p.r. and white noise above 27, will be discussed in section 4.2.8.

One of the consequences of the *band limitation* described above is that the covariance matrix of the measurements is no longer a scaled unit matrix. In particular, the covariance matrix will be a full matrix, however, with a certain favourable structure. It will have constant diagonals (so–called *Toeplitz matrix*) and it can be shown that, with such covariance matrix, orthogonality properties of the trigonometric functions are preserved, so that the normal matrix will still be block–diagonal (compare section 4.2.1).

Another consequence of (lower) band limitation is that some degrees l below the

value of β_{min} cannot be estimated from the measurements any more, since for those degrees the total power content mainly (or solely) comes from the measurement frequencies below β_{min}. This can be seen as follows. Removing from the analysis the frequencies below β_{min} means that only the β_{km} for which $|\beta_{km}| \geq \beta_{min}$ are included. Let us first consider the case $\beta_{km} > 0$, so that we have $\beta_{km} \geq \beta_{min}$. With $\beta_{km} = k + m\alpha$ (where $\alpha = N_d/N_r$) and $k = l - 2p$, $0 \leq p \leq l$ we find that only those degrees l are present for which $l \geq \beta_{min}$. The other case, $\beta_{km} < 0$ leads to $l \geq (1-\alpha)^{-1}\beta_{min}$. If we take $N_r = 727$ and $N_d = -45$ (Schrama, 1990), which fulfil the conditions stated in section 4.2.1 and which refer to the Aristoteles mission, we find that e.g. for $\beta_{min} = 4$ degrees $l < 4$ are not present and for $\beta_{min} = 27$ degrees $l < 26$.

Figure 4.7 shows the a–posteriori error r.m.s. values for the case that $\beta_{min} = 4$. Frequencies below 4 c.p.r. were removed from the analysis and for the remaining frequencies a white noise error spectrum was assumed. A remarkable difference with figure 4.2 is the zig–zag pattern which occurs for certain components and combinations. This has to do with the step size of 2 for the index k, cf. equation 4.8. Given a certain value of m a β–limitation also implies a k–limitation, since $\beta_{km} = k + m\alpha$. This is especially clear for low orders m (since α is small), where a relatively large part of the total signal content is concentrated. The consequences of limiting k are different for even and odd degrees, due to the step size of 2. For example, consider the case $m = 0$. This is the easiest case since for $m = 0$ we have $\beta_{km} = k$. If we choose $\beta_{min} = 1$ this means all k–contributions for $k = 0$ are removed. But only the even degrees include a $k = 0$ part, so that the odd degrees remain unaffected. For other values of β_{min} a same reasoning can be applied. For higher values of m, limiting the β_{km} results in a difficult pattern of dropped k, m contributions. It is not so clear any more how this limits the k–summation for a certain degree l and how this is related to the parity of l. But since the contribution to the total signal content for higher orders is relatively small, the main effect of β–limitation is to be expected to come from the low orders.

It can be seen from the figure that, for certain observation types, the zig–zag pattern damps out with increasing degree. Two explanations can be given for this. As we have seen above, the main effect of a lower band limitation comes from the low orders. Low degrees are therefore more affected than high degrees since a relatively larger part of the total signal content comes from the low orders. Furthermore, as explained in the previous section, for higher degrees the r.m.s. curves approximate more and more a straight line induced by the attenuation factor $(R/r)^{l+3}$. It should be noted that (part of) the zig–zag pattern could be caused by the occurrence of singular m–blocks, as discussed in the previous section. Compared to the case $\beta_{min} = 0$ significantly more m–blocks are singular, however often only the even or the odd degrees, see table 4.5.

The singular m–blocks make it furthermore difficult to compare the overall level of the r.m.s. curves in the case $\beta_{min} = 0$ with that of $\beta_{min} = 4$, especially if the cause of the singularities is numerical instability. That the situation becomes worse

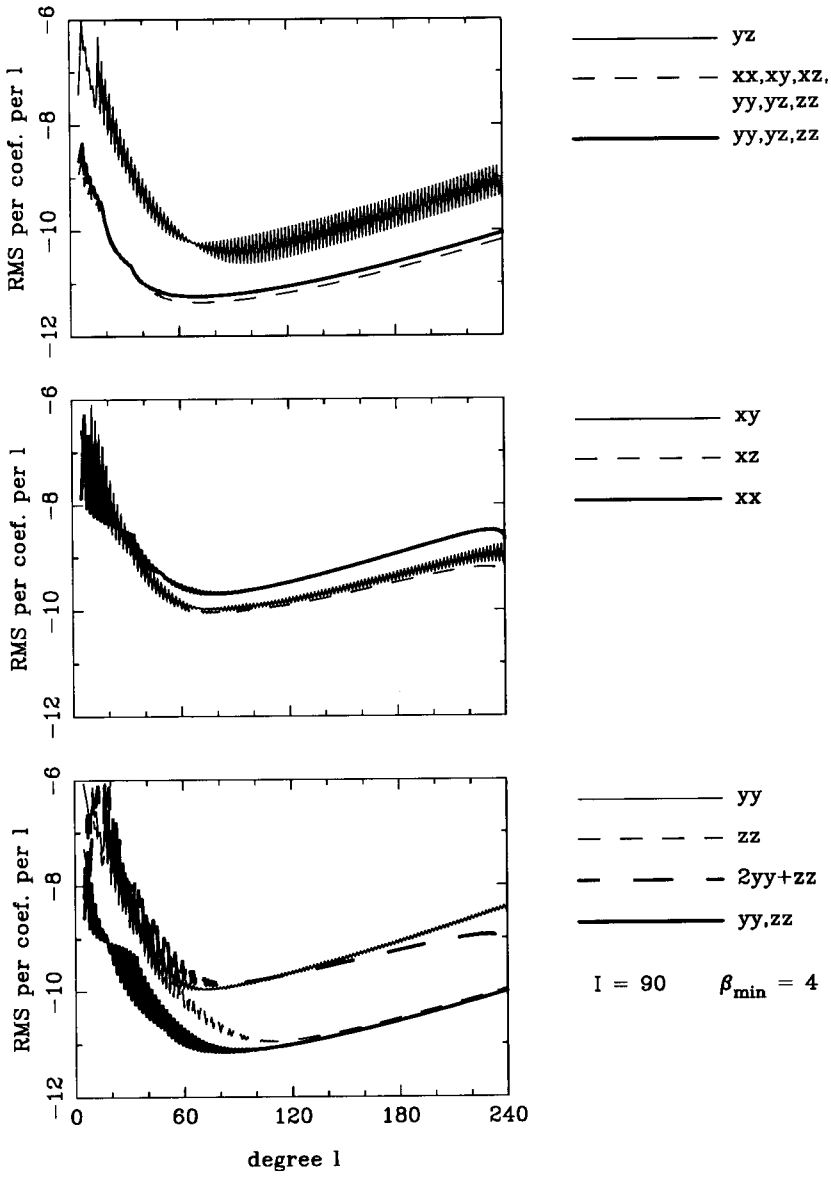


Figure 4.7 *Propagated error from gradiometry. Specifications as in figure 4.2 except now* $\beta_{min} = 4$.

in the latter case (i.e. larger propagated error) is not very surprising, but only a comparison for each coefficient individually would give a clear picture of the exact amount of distortion. If we, nevertheless, try to compare figures 4.2 and 4.7 we see that indeed mainly the lower degrees are affected. R.m.s. values reach up to two orders of magnitude higher values in the case $\beta_{min} = 4$. As a consequence, the minima of the curves are moved in the direction of increasing degree, around $l = 60$ for $\beta_{min} = 0$ but around $l = 120$ for $\beta_{min} = 4$.

A last remark concerns the discontinuous behaviour for some components for low degrees ($l < 20$, e.g. V_{yz}). This is a direct consequence of the singular m–blocks. As discussed in the previous section, the singular m–blocks are removed from the analysis so no contribution for those orders is included in the r.m.s. values, resulting in too optimistic error estimates. From table 4.5 we see that singular m–blocks mainly appear for low orders. Since low degrees are relatively more affected by low orders, this explains the bumpiness for low degrees.

Figures 4.8 and 4.9 present global r.m.s. values for surface gravity anomalies and geoid undulations for the present case $\beta_{min} = 4$, figure 4.8 for several mission durations and figure 4.9 for several measurement precisions. Both figures give the results for a full tensor gradiometer (i.e. the combination of all six gradients), the r.m.s. being computed with a maximum degree for the omission error of 1000 and a smoothing factor β_l representing $1° \times 1°$ block averages. Comparison of these figures with figures 4.5 and 4.6 reveals that the r.m.s. for gravity anomalies is hardly affected by this value of β_{min}, whereas the geoid undulations show clear differences. These differences not only concern the overall higher level of r.m.s. Also the influence of both the mission duration and the measurement precision is noticeable now, even for lower altitudes, where it was not in the previous section (see the differences at $h = 160$ km). Geoid undulations constitute a smoother signal than gravity anomalies. The latter are first–order derivatives of the gravitational potential, resulting in multiplication factors $(l + 1)$ etc., see chapter 3. It is therefore to be expected that the geoid undulations are more affected by changes in the lower degrees (as is the case if we choose $\beta_{min} \neq 0$) than gravity anomalies. Note that the high precision gradiometer ($\sigma = 0.0001$ E/$\sqrt{\text{Hz}}$) now shows its value, the r.m.s. of the geoid undulations being for this σ nearly the same for both $\beta_{min} = 0$ as well as $\beta_{min} = 4$, for all altitudes.

We have to bear in mind, however, that the figures present the results for a *full tensor* gradiometer. For the combinations $\{V_{yy}, V_{yz}, V_{zz}\}$ or $\{V_{yy}, V_{zz}\}$ (as for Aristoteles) the results are somewhat worse, especially for the geoid undulations. Global r.m.s. values from those combinations are respectively 15.6 cm and 97.0 cm for geoid undulations (at 200 km, 6 months mission and $\sigma = 0.01$ E/$\sqrt{\text{Hz}}$), whereas the values for the gravity anomalies are hardly changing. Individual components give even worse results. This shows the importance of combinations of components, i.e. the need for additional observations, in the band limited case. If other components are not available, like for Aristoteles (planar gradiometer, band limited, degraded lower spectrum), additional information from other measurement techniques (like GPS) is

necessary, see section 4.2.8. Furthermore, in section 4.2.6 we will discuss the possibility of improving the solution by adding prior information, i.e. stabilizing the solution by adding constraints in the sense of least squares collocation (Rummel et al., 1979).

4.2.5 Polar gaps

A realistic gradiometric mission not only differs from the ideal case sketched in section 4.2.3 through band limitation. In reality the satellite is also not likely to fly in a perfect polar orbit. For Aristoteles two different inclinations are considered: 92°.3 and 95°.3. Aristoteles is planned to fly for approximately six months in an orbit with inclination 95°.3 and another two weeks in an orbit with inclination 92°.3. A non–polar orbit gives rise to polar regions where no measurements are available. If the inclination of the orbit is I, polar caps, of size $2 \times (I - 90)°$, will arise which are not covered by ground tracks of the satellite's orbit. As a consequence an incomplete global coverage is obtained which is expected to influence a global recovery method as it is performed here. At first sight, one expects a polar gap, with size larger than half of the smallest wavelength to be recovered, to disturb the solution, at least for those short wavelengths (i.e. high degrees). For recovery, on a global scale, of gravitational features with very short wavelengths there are simply not enough (well distributed) data points. On the other hand, it is very difficult to exactly translate the missing of data in certain spatial regions to a possible distortion of recovered potential coefficients. This may become clear from figure 4.10, in which the propagated error for the case $I = 92°.3$ is shown.

Remarkably no substantial deterioration occurs compared to the case $I = 90°$, except for V_{yy}. Whereas the result for some components becomes only slightly worse, other components seem to become even better. In spite of the fact that this outcome is confirmed by other investigations (Schrama, 1990), it is difficult to give a clear explanation for it. Several aspects may play a role here. First consider the data distribution. In both cases, $I = 90°$ and $I = 92°.3$ we have the same amount of observations (namely a total of $T_r/\Delta t$ observations). Although for a 92°.3 inclination the polar gaps occur, the data density in the remaining part (the band $2°.3 \leq \theta \leq 177°.7$) becomes higher. Whereas the polar gaps may distort the recovery of some coefficients, the higher data density may improve the recovery of others. On the whole, a small improvement may occur if the inclination differs not too much from 90°. If the deviation from 90° becomes too large, the influence of the polar gaps will ultimately surpass that of the higher data density.

A second aspect which may play a role is the behaviour of the inclination functions. Besides on factors like Γ_l, $(l+1)$, k, etc. the total signal power contained in a specific degree depends on the inclination functions $\bar{F}^k_{lm}$ (or the cross–track inclination functions $\bar{F}^{k\,*}_{lm}$ for V_{xy} and V_{yz}). They make part of the so–called *sensitivity coefficients* $H^{(\cdot)}_{lmp}$ from tables 4.1 and 4.2. Let us focus here on one specific degree and order (i.e. fixed l and m). What remains is a summation over k of inclination functions and a possible factor k or k^2 (cf. tables 4.1 and 4.2). If we compute the

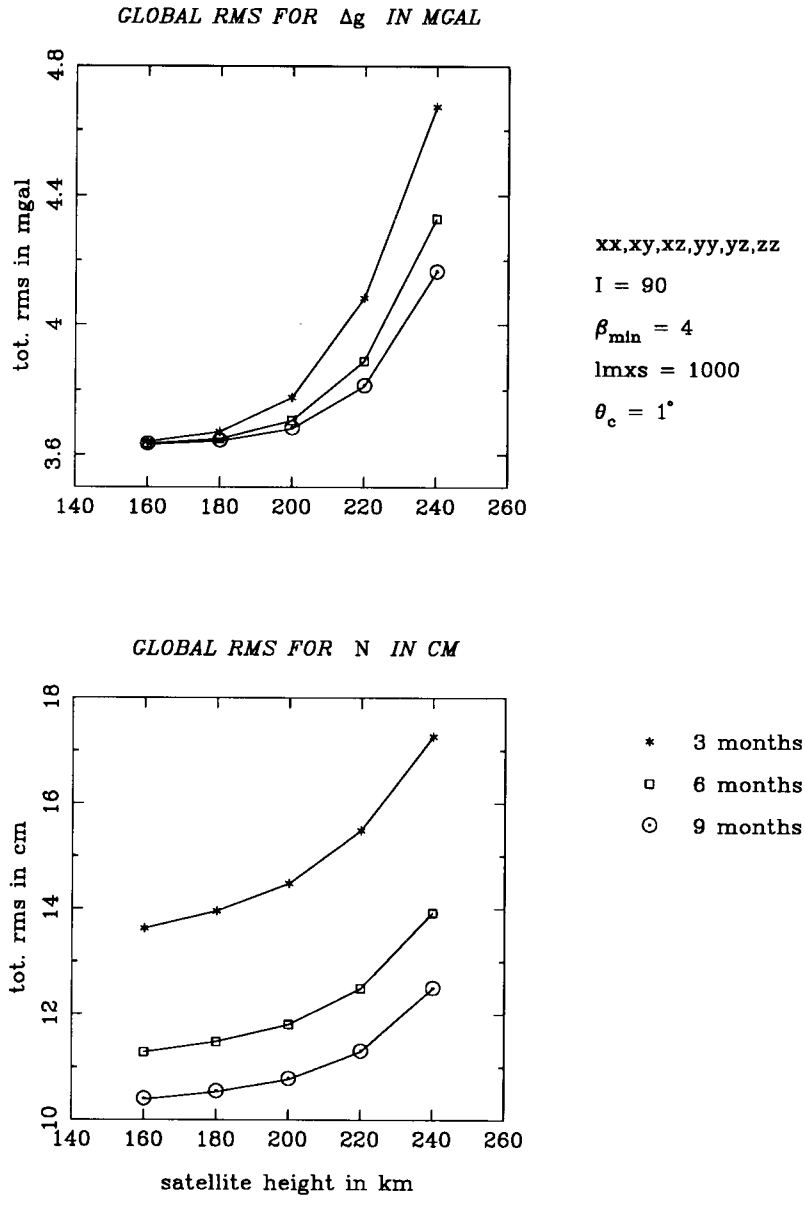


Figure 4.8 *Global r.m.s. values for several mission durations. Specifications as in figure 4.5 except now* $\beta_{min} = 4$.

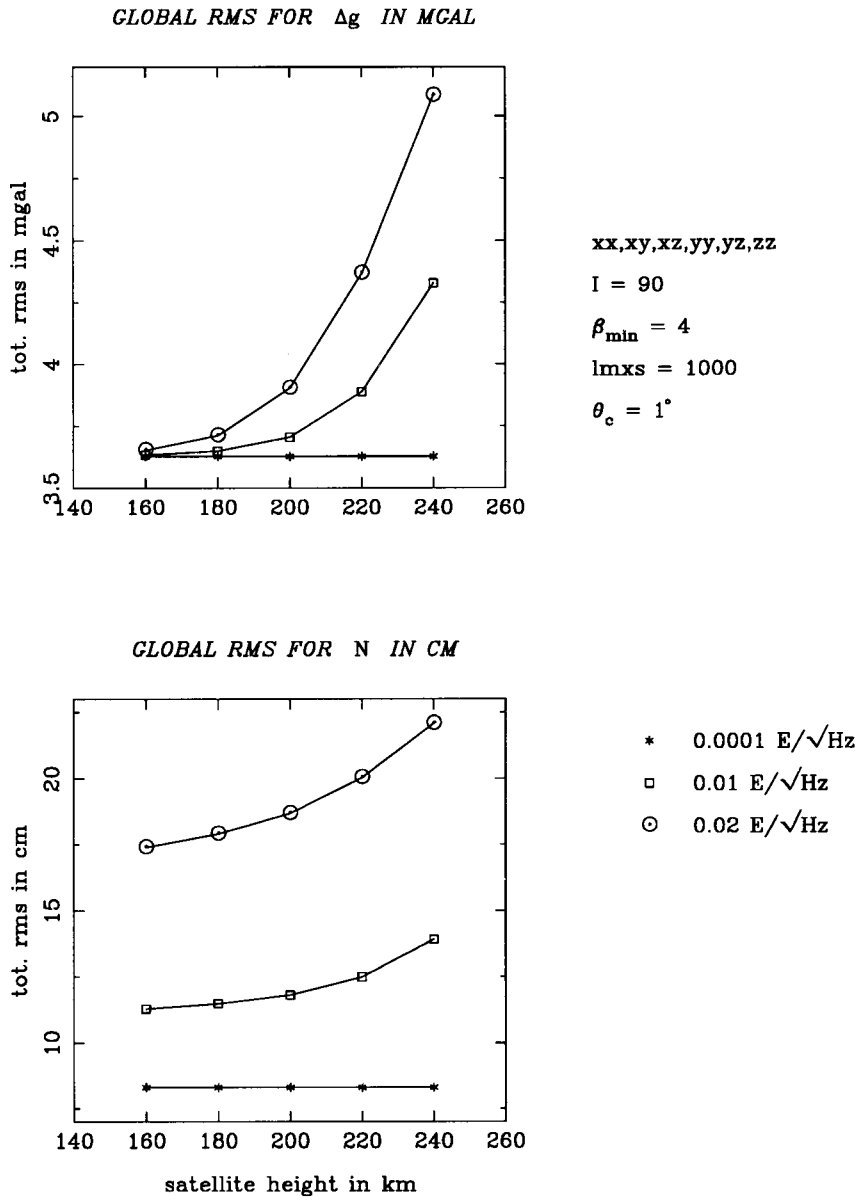


Figure 4.9 *Global r.m.s. values for several measurement precisions. Specifications as in figure 4.6 except now* $\beta_{min} = 4$.

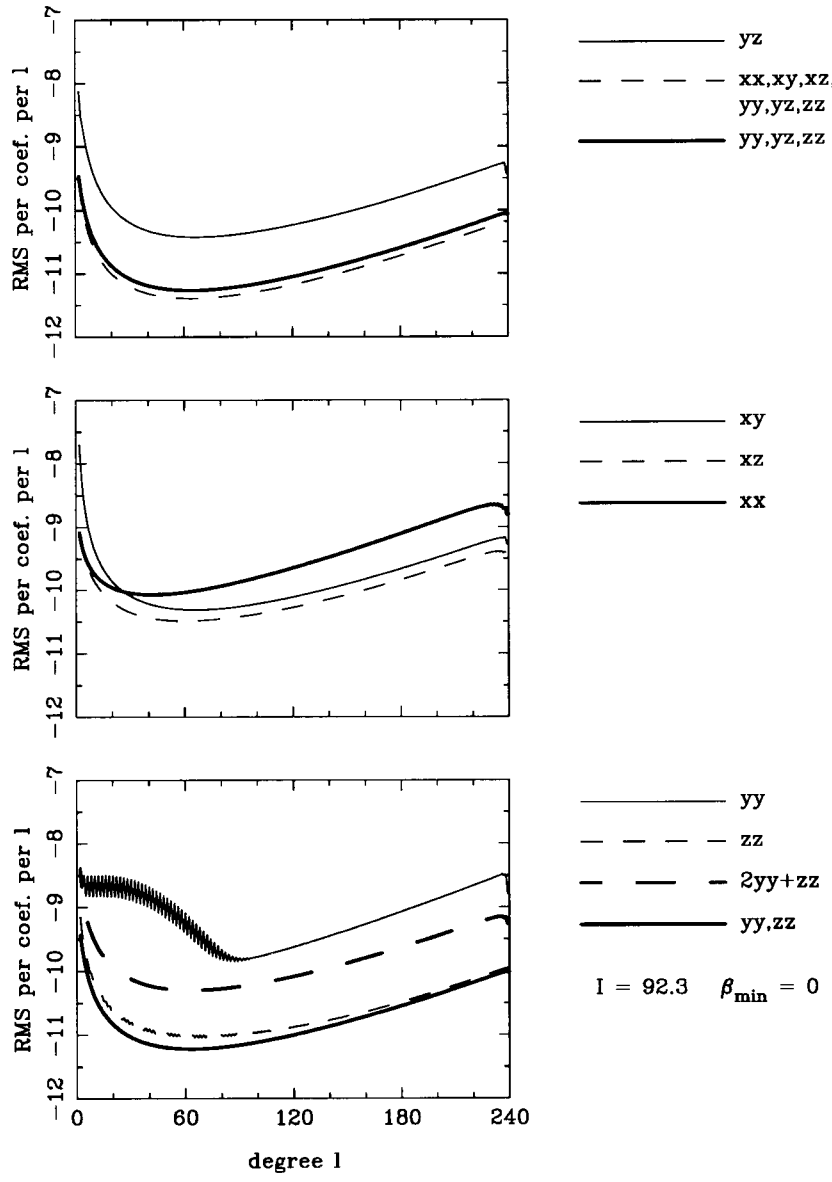


Figure 4.10 *Propagated error from gradiometry. Specifications as in figure 4.2 except now $I = 92°.3$.*

inclination functions for a specified l, m and k as function of I, not necessarily a maximum value is obtained for the $\bar{F}_{lm}^{k}$ at $I = 90°$. The value of some inclination functions will increase if $I > 90°$, some will perhaps decrease for higher inclinations. We did some test computations which confirmed this behaviour. If *all* inclination functions would obtain a maximum value at $I = 90°$, one would expect the signal amount to decrease for $I > 90°$, so that the propagated error would increase. Since this is not the case, the signal amount may also increase for higher inclinations (so that the propagated error decreases). Carrying out the k–summation for one specific l, m over all corresponding $\bar{F}_{lm}^{k}$ (some of which have a larger value, others a smaller) the result for $I = 92°.3$ might either be larger or smaller than the same sum for $I = 90°$, depending on the specific sensitivity coefficient (i.e. on the gradient).

A third aspect applies only to the V_{xy} and V_{yz} components. As mentioned in section 4.2.3 the cross–track inclination functions are all zero for $I = 90°$ and $m = 0$. This resulted in two singular m–blocks, cf. table 4.5. For $I = 92°.3$ this does not happen any more, i.e. the potential coefficients for $m = 0$ are now estimable from the observations. This means that the $m = 0$ contribution is included in the r.m.s. values shown in figure 4.10, whereas it was not in figure 4.2. This troubles a direct comparison of the two figures.

An inclination of $95°.3$ has a very strong impact on the results, as can be seen in figure 4.11. Not only does the level of the r.m.s. increase, also the shape of the curves changes drastically. Whereas it is expected that for this case the exchange between positive and negative influence of respectively data density and polar gaps will result in the negative, inclination function behaviour and singular m–blocks may play an important role here. In any way, without any stabilization method or without any additional information to account for the data gaps at the poles, these results are not acceptable.

4.2.6 Stabilization

The problem of solving gravitational information at the surface of the earth from observed data at satellite altitude is known to be improperly posed (also called ill–posed or ill–conditioned). A problem is called improperly posed if it does not meet at least one of the following three requirements: 1.) existence of the solution, 2.) uniqueness of the solution and 3.) stability of the solution (Rummel et al., 1979), (Moritz, 1980) or (Neyman, 1985). In our case, where we try to derive detailed gravitational information at the surface of the earth from a relatively smooth (attenuated) signal at satellite altitude (downward continuation), we encounter problems with the uniqueness and with the stability of the solution. The former case applies when some eigenvalues of a normal matrix sub–block become zero so that the block becomes singular, as we have seen e.g. for the V_{xy} and V_{yz} gradients for $m = 0$ at $I = 90°$ (cf. table 4.5). In the latter case, eigenvalues may become extremely small, so that the matrix is ill–conditioned and errors in the data, but also round–off errors of the computer, are greatly amplified during the inversion process. The block may

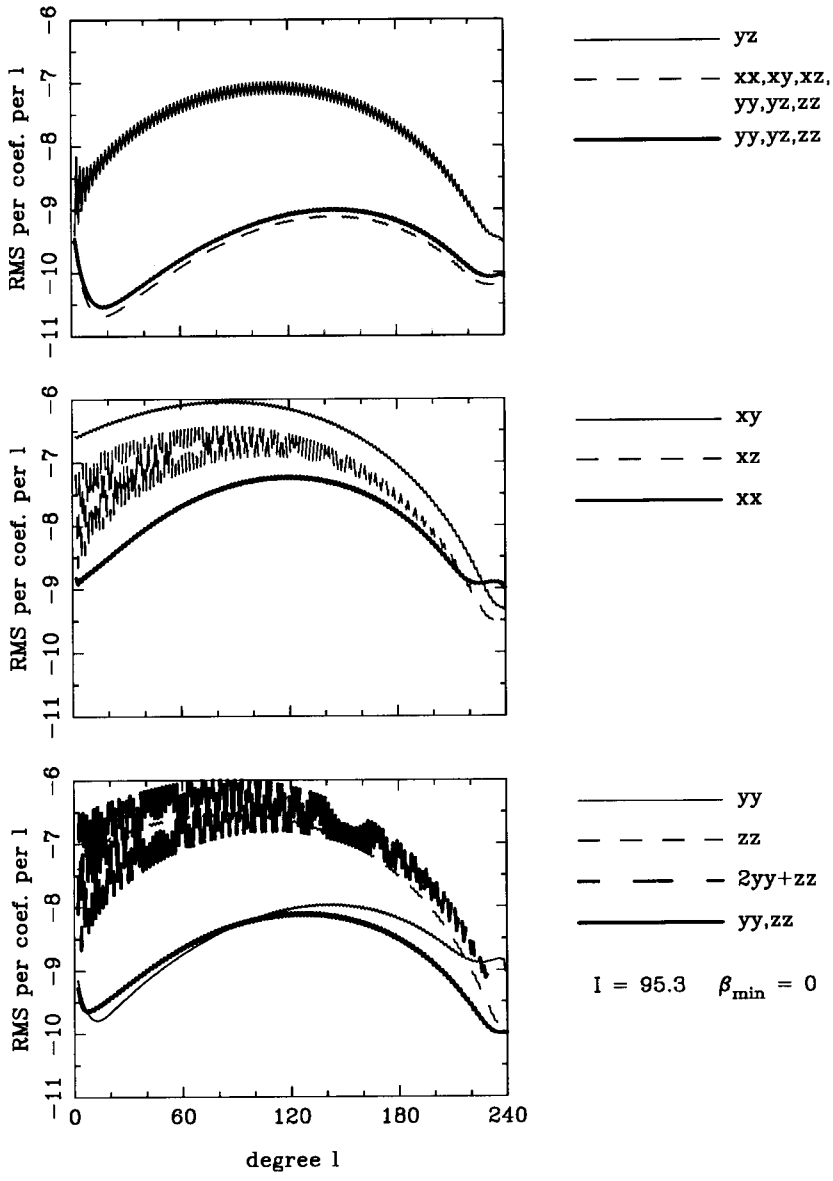


Figure 4.11 *Propagated error from gradiometry. Specifications as in figure 4.2 except now $I = 95°.3$.*

even become numerically singular in this case. We have already seen clear examples of such improperly posed problems in the foregoing sections, e.g. the case $\beta_{min} = 4$ or the case $I = 95°.3$.

For both cases the instability is again illustrated by means of the condition number of the individual blocks of the normal matrix. As we know, the normal matrix consists of blocks on the diagonal, decreasing in size. There is one block for each order m and degrees l of equal parity. Figures 4.12 and 4.13 show the condition number for the even degree blocks for the case $\beta_{min} = 4$ and $I = 95°.3$ respectively. The odd degree blocks give similar results. The condition number κ is defined here as the largest eigenvalue λ_{max} divided by the smallest eigenvalue λ_{min}:

$$\kappa = \frac{\lambda_{max}}{\lambda_{min}} .$$

The blocks which are expected to be numerically unstable can be traced by comparing the inverse of the condition number κ^{-1} for each block with the internal accuracy of the computer. In our case, all the computations are done using FORTRAN REAL*8 (double precision) numbers, i.e. some 15 or 16 significant digits. Relative errors of the order 10^{-15} or 10^{-16} are likely to occur. Blocks for which the condition number approaches (e.g. $\kappa^{-1} > 10^{-12}$) or exceeds this value are very ill–conditioned. Inspection of figure 4.13 leads to the conclusion that for almost all components and combination of components the blocks belonging to orders approximately below 20 tend to be ill–conditioned. Hence, also degrees below 20 are likely to give bad results. Also higher degrees are influenced by those ill–conditioned low orders. For the case $\beta_{min} = 4$, condition numbers for low–order blocks (approximately $m < 10$) reach even higher values, up to 10^{18} or higher, as can be seen from figure 4.12.

One of the possibilities to stabilize the solution is to make use of additional prior information about the unknowns. In terms of least squares collocation one adds the prior expectations of the unknowns together with their covariance matrix. The solution then becomes (compare equation 4.5):

$$\hat{c} = N^{-1}(A^T Q^{-1}\, \ell + P^{-1} c')$$

with the matrix N now defined as

$$N = A^T Q^{-1} A + P^{-1} \tag{4.15}$$

where c' are the prior expectations of the potential coefficients and P is their covariance matrix. In our case c' could contain the coefficients of one of the existing geopotential models, such as OSU86F, with P the corresponding covariance matrix. For the purpose of a global error analysis, as discussed here, a more common choice for c' is zero with P one of the existing signal degree variance models, e.g. Kaula's rule (Kaula, 1966). Here we take the degree variances from a Tscherning/Rapp (TR) model (Tscherning and Rapp, 1974). Such a choice implies that we consider the prior value of the potential coefficients to be zero and their variances to follow

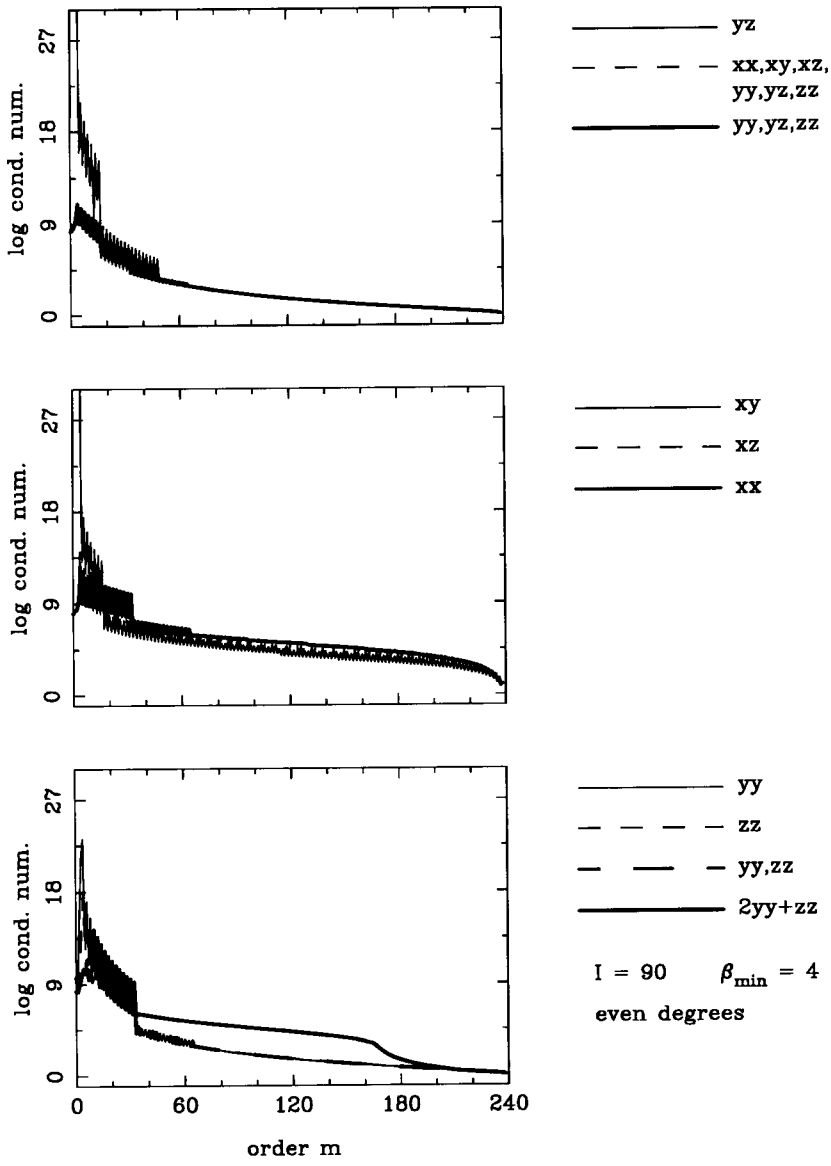


Figure 4.12 *Condition number for blocks of the normal matrix for even degrees. Specifications as in figure 4.2 except* $\beta_{min} = 4$.

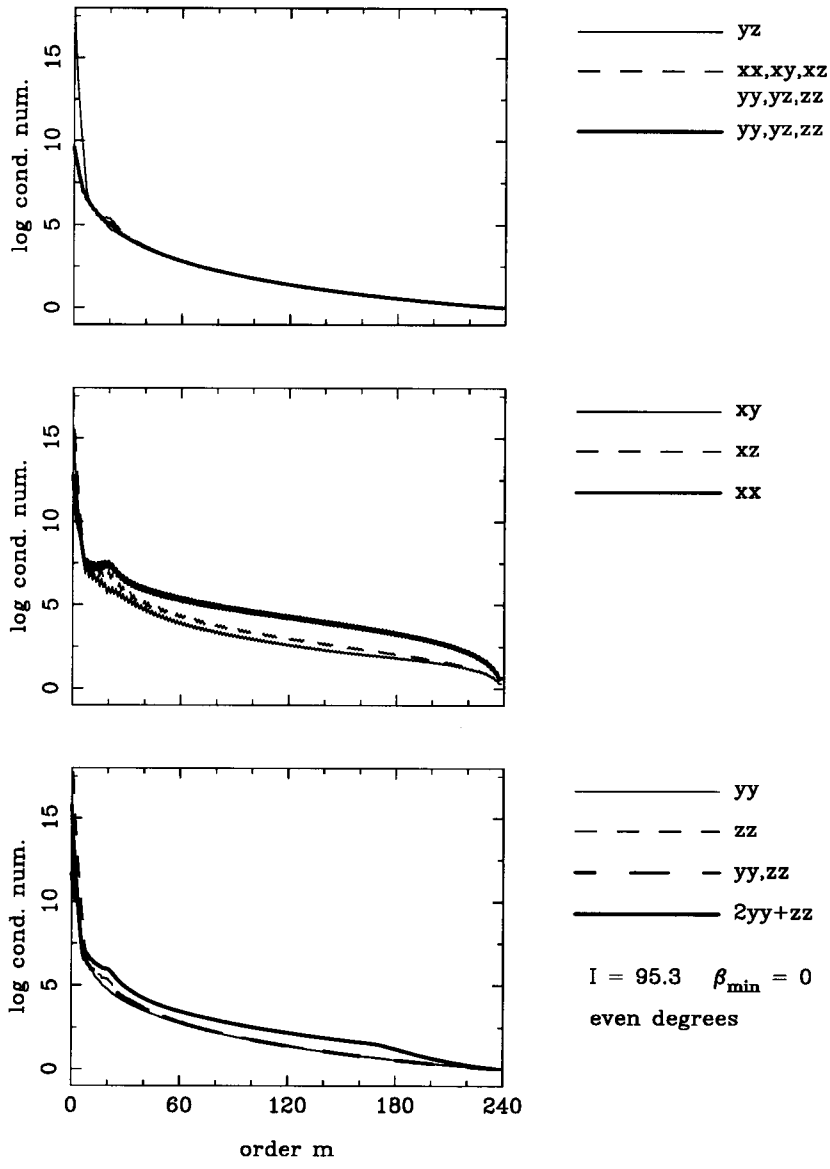


Figure 4.13 *Condition number for blocks of the normal matrix for even degrees. Specifications as in figure 4.2 except $I = 95°.3$.*

the TR model. The solution is in this case again given by equation 4.5, but now with the normal matrix as in equation 4.15. This means that the solution is a kind of weighted mean of a–priori information and gradiometry results.

Figure 4.14 shows the error r.m.s. for the case $\beta_{min} = 4$ where the solution was stabilized by adding to the normal matrix a diagonal matrix containing the TR degree variances. This figure should be compared with figure 4.7. For completeness the curve for the TR model is added. We see that the zig–zag pattern is damped for most solutions (clearly visible for V_{yz}). On the other hand, the curves tend to approximate the TR curve. Questions can be asked about the value of the, in this way, estimated coefficients, especially for higher degrees, since they obviously do not add much information to the a–priori known model degree variances. Above degree 240 all curves will follow almost exactly the TR model curve. This is a reason why we choose a maximum degree of 240 for our computations. Above this degree, potential recovery is not likely to give any new information, at least not for more realistic (and thus ill–conditioned) situations, like the one in figure 4.14.

The stabilized solution for $I = 95°.3$ is shown in figure 4.15. Compared with figure 4.11 the stabilization certainly has much influence, since for lower degrees the error r.m.s. curves now show a behaviour comparable to figure 4.2, though a little disturbed. For higher degrees the curves tend to follow that of the TR model.

Following (Xu and Rummel, 1991) one may view upon stabilization as biased estimation. As a purely mathematical technique, one of the purposes of biased estimation is to control instability. It is defined as:

$$\hat{\boldsymbol{c}}_{\boldsymbol{b}} = (A^T Q^{-1} A + K)^{-1} A^T Q^{-1} \boldsymbol{\ell}$$

where the index $\boldsymbol{b}$ means "biased estimate" and K is some arbitrary but positive definite matrix. This way of looking upon stabilization becomes important if the prior information added to the problem is not correct (Xu, 1991). In fact, this is exactly the situation in our stabilization method (where we take for K^{-1} the covariance matrix P of the prior expectations of the coefficients), since the degree variances of the TR model do not describe error variances of the potential coefficients, but are possible magnitudes of the coefficients themselves (signal variances).

Following (Xu and Rummel, 1991) the magnitude of the expected bias in this case can be computed as

$$Bia(\hat{\boldsymbol{c}}_{\boldsymbol{b}}) = -(A^T Q^{-1} A + P^{-1})^{-1} P^{-1} \boldsymbol{c}$$

where we should insert for $\boldsymbol{c}$, in principle, the true potential coefficients. In general one does not have the real coefficients, so the bias is approximated by replacing the real coefficients with either the biased estimates themselves, or some model coefficients. In (Xu and Rummel, 1991) it appeared that the former choice (biased estimates themselves) leads to too optimistic conclusions about the bias. Choosing coefficients computed from the model which was used for the K matrix appeared to give more conservative results.

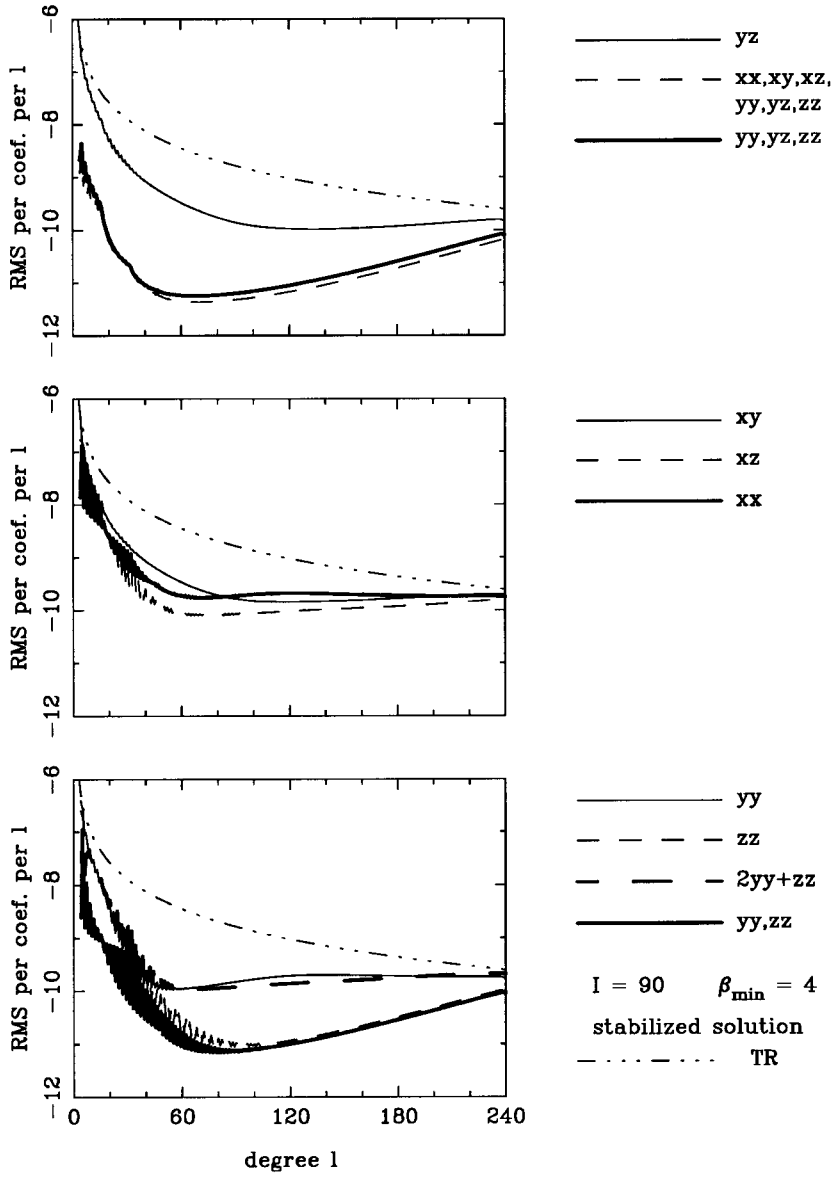


Figure 4.14 *Propagated error from gradiometry. Specifications as in figure 4.2, except $\beta_{min} = 4$, and now the solution is stabilized by means of a Tscherning/Rapp model added to the normal matrix.*

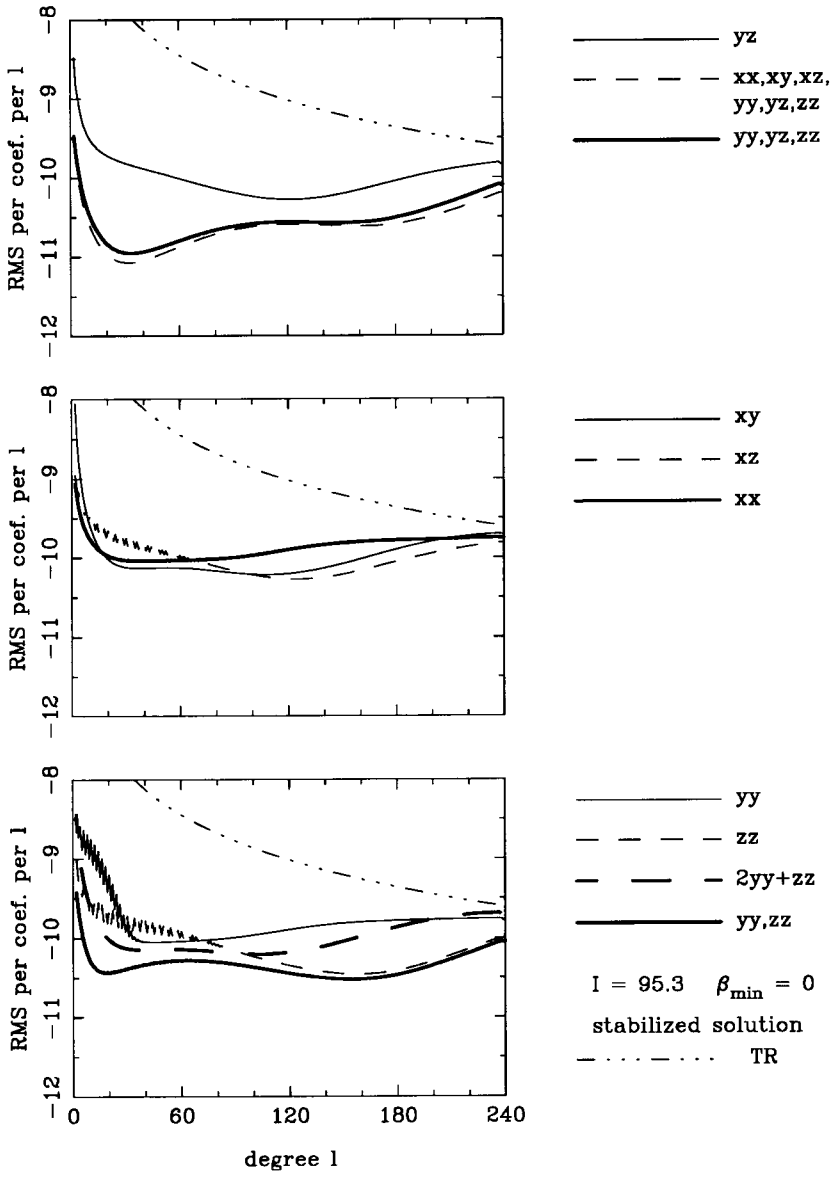


Figure 4.15 *Propagated error from gradiometry. Specifications as in figure 4.2, except $I = 95°.3$, and now the solution is stabilized by means of a Tscherning/Rapp model added to the normal matrix.*

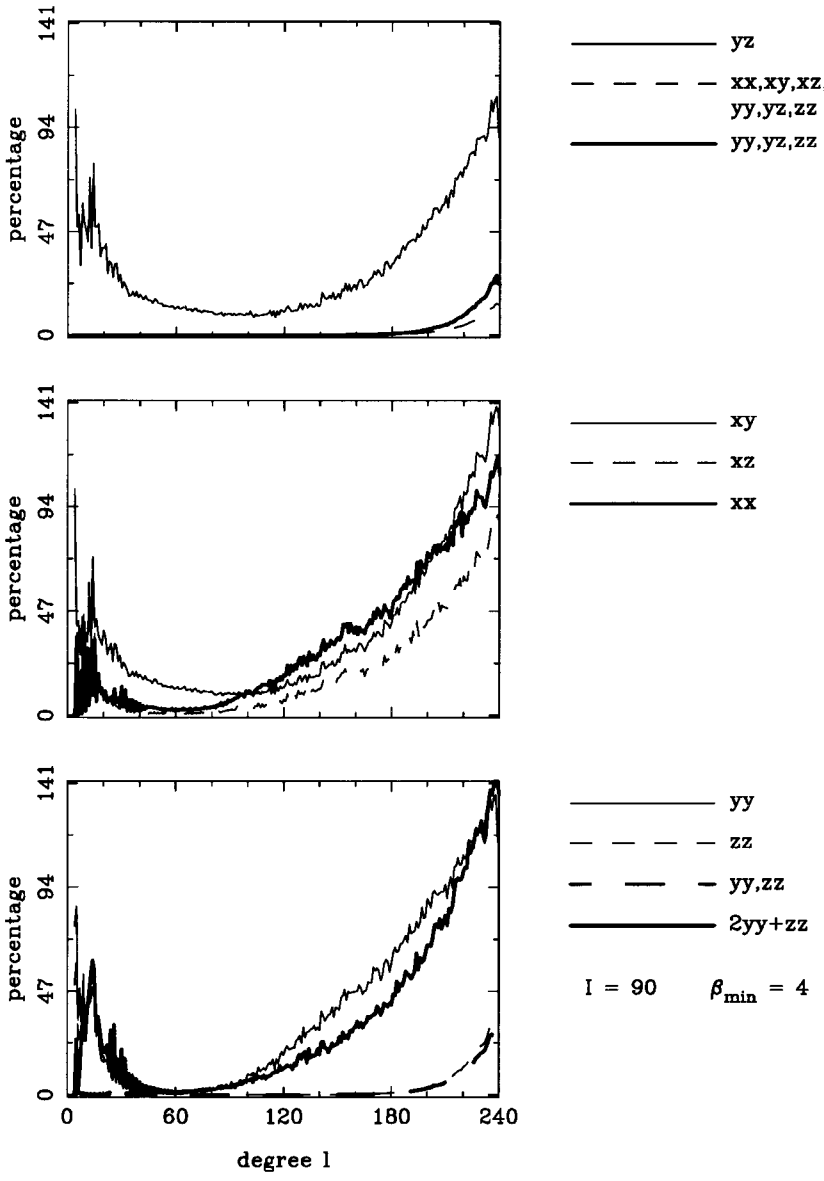


Figure 4.16 *Percentages of the biases with respect to the OSU86F model for the case* $\beta_{min} = 4$.

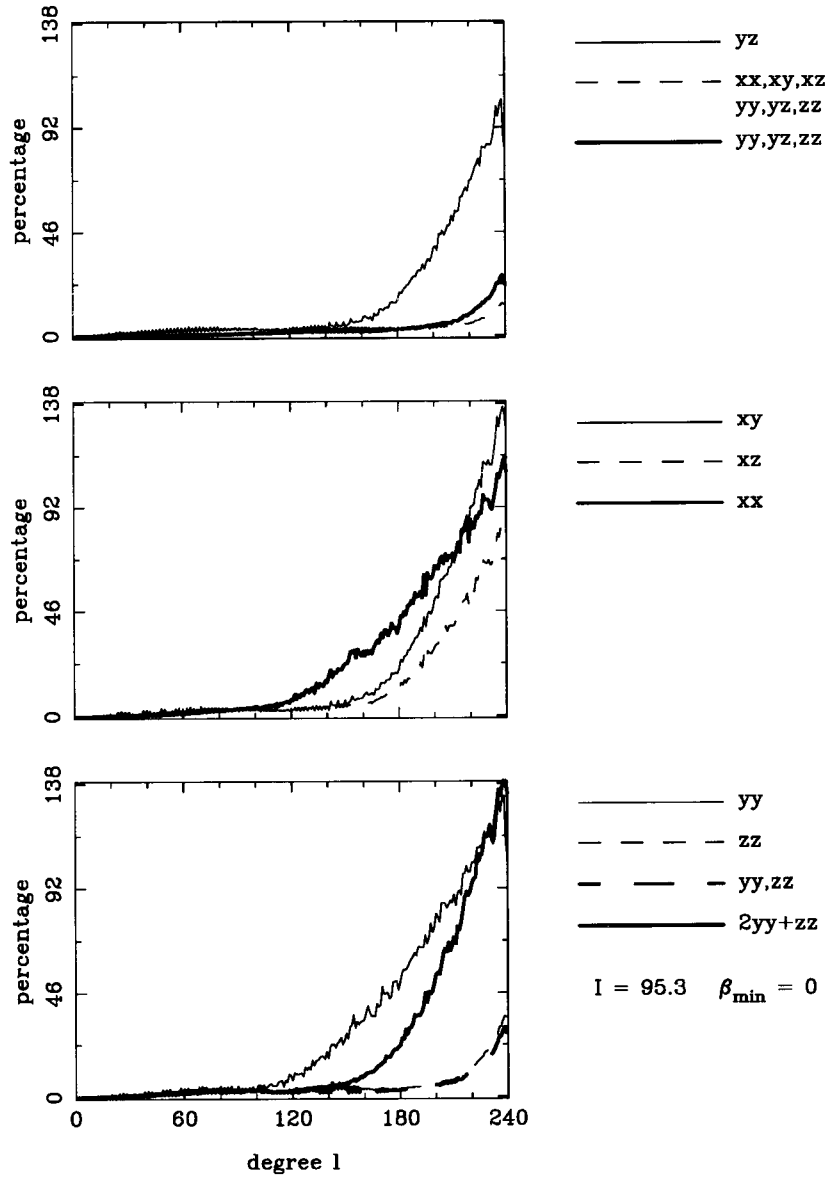


Figure 4.17 *Percentages of the biases with respect to the OSU86F model for the case $I = 95°.3$.*

For both stabilized solutions presented above ($\beta_{min} = 4$ and $I = 95°.3$) we computed the bias using the equation above. We chose as coefficients $\boldsymbol{c}$ the same values from the TR model which were used in the P matrix, but with the sign taken from the OSU86F potential coefficient set. Since we, at this moment, do not have "true" coefficients estimated from gradiometer measurements, we cannot compare the biases with such coefficients. But since using TR coefficients may give the most conservative result, cf. (Xu and Rummel, 1991), we may compare them to e.g. the OSU86F model. To this extent, the percentages of the degree variances of the biases with respect to those of the OSU86F model are computed and shown in figures 4.16 and 4.17. Depending on which component or combination of components is used for the analysis, large parts of the spectrum show biases which exceed 20 %, even 100 % of the model variances. However, for the case $I = 95°.3$ this only happens for higher degrees. Note that the results for all six components together, for V_{zz} and for the combinations $\{V_{yy}, V_{yz}, V_{zz}\}$ and $\{V_{yy}, V_{zz}\}$ are not so bad at all, the degree variances of the bias in these cases reaching approximately 25 % of the model values only for very high degrees ($l > 220$).

The accuracy of biased estimates is no longer given by the inverse of the normal matrix only. The bias term needs to be added to the a–posteriori error. The so–called mean squared error of $\hat{\boldsymbol{c}}_{\boldsymbol{b}}$, $MSE(\hat{\boldsymbol{c}}_{\boldsymbol{b}})$, obtained in this way is given by:

$$MSE(\hat{\boldsymbol{c}}_{\boldsymbol{b}}) = N^{-1} - N^{-1}P^{-1}N^{-1} + Bia(\hat{\boldsymbol{c}}_{\boldsymbol{b}})(Bia(\hat{\boldsymbol{c}}_{\boldsymbol{b}}))^T \qquad (4.16)$$

with N from eq. 4.15, see (Xu and Rummel, 1991). Figures 4.18 and 4.19 show the percentages of this MSE with respect to the propagated error (as coming from the inverse of the normal matrix only) from figures 4.14 and 4.15, for the cases $\beta_{min} = 4$ and $I = 95°.3$ respectively. In these figures the straight horizontal line indicates the 100 % line. If the bias would not have any effect, all percentages would have to lie on this 100 % line. Percentages smaller than 100 indicate degrees for which the propagated error (no bias term included) is larger than the MSE. It means that, for those degrees, the previously shown results (without the bias) were in fact too pessimistic. In this case we don't have to worry that stabilization leads to too optimistic error estimates. Higher percentages indicate degrees for which inclusion of the bias causes a higher a–posteriori error, so for those degrees the results from figures 4.14 and 4.15 were too optimistic.

Whereas in figure 4.18 some percentages reach high values of nearly 160, they show no substantial degree dependent systematic deviation from 100 %. On the average, the components V_{yz}, V_{xy} and V_{yy} seem to give less favourable results due to the stabilization (if the bias is included in the error), whereas the other components seem to become slightly better. For $I = 95°.3$, on the average, for most components the percentage is smaller than 100 (at least for large parts of the spectrum), indicating that if one accounts for the bias, estimates will become better.

So, although the estimates become biased, this bias stays limited, the a–posteriori error (which has to include the bias, so it is the MSE) being positively affected by

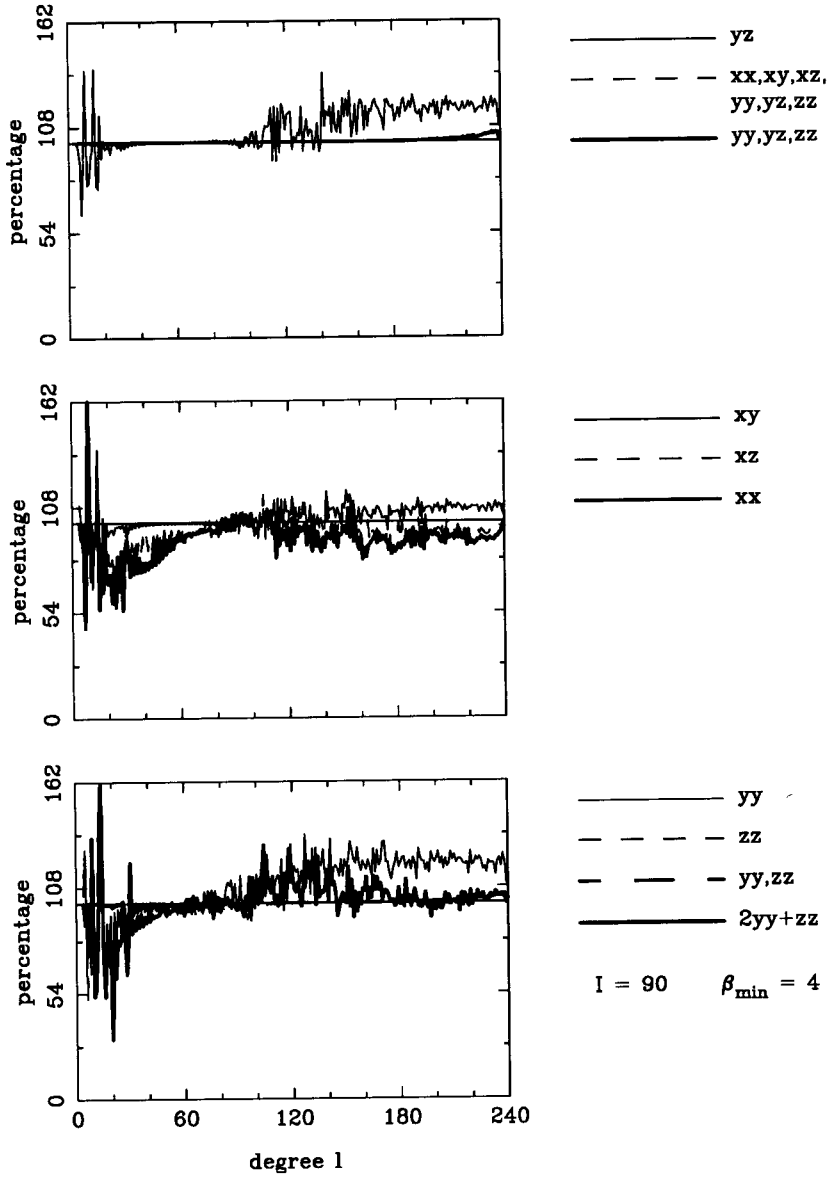


Figure 4.18 *Percentages of the MSE with respect to the a-posteriori error (N^{-1} only) for $\beta_{min} = 4$.*

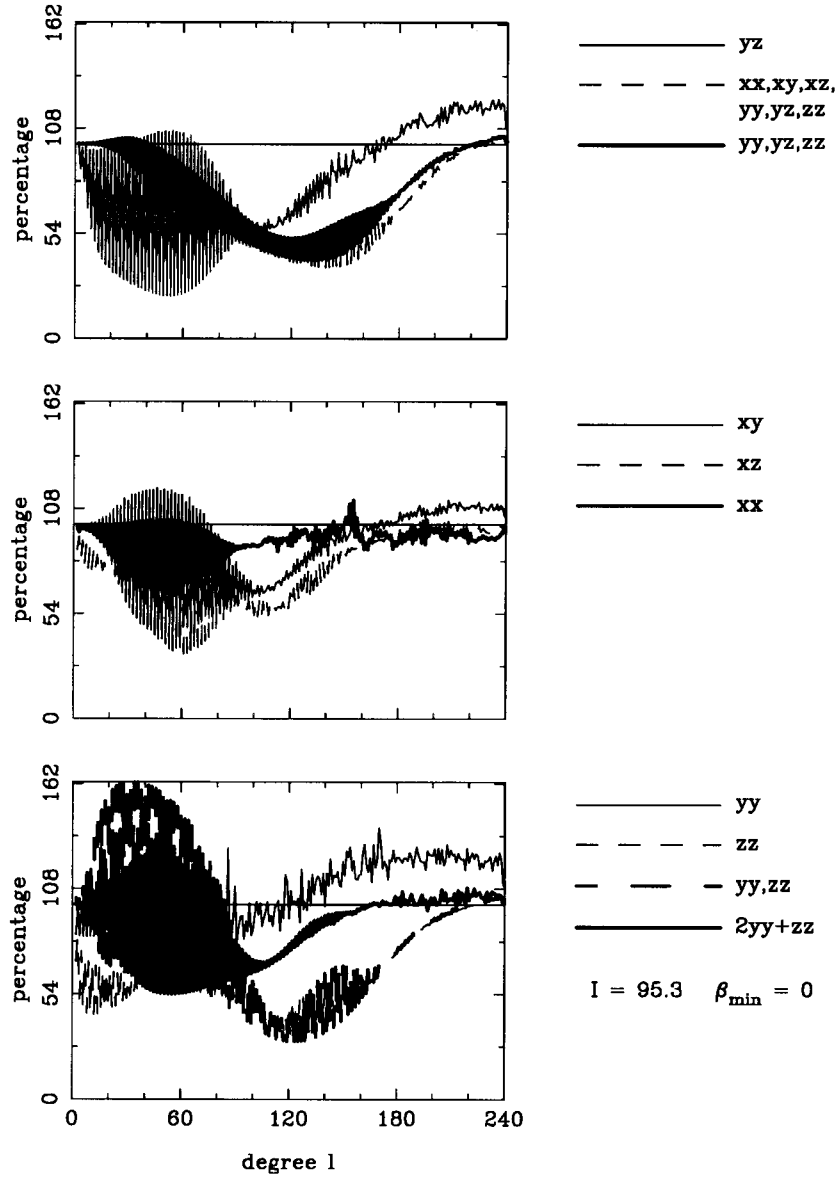


Figure 4.19 *Percentages of the MSE with respect to the a-posteriori error (N^{-1} only) for $I = 95°.3$.*

it on the average, especially for the case $I = 95°.3$. Thus, stabilization seems to be a satisfactory means of obtaining a stable and meaningful solution.

4.2.7 Omission, commission and smoothing

In sections 4.2.3 and 4.2.4 global r.m.s. values for gravity anomalies and geoid undulations were presented. They were computed using equation 4.14 and therefore included commission error and omission error. The commission error represents the propagated measurement error. It only concerns that part of the spectrum for which the potential coefficients can be estimated from the measurements, i.e. for degrees $l \leq L$ (see section 4.2.2). The omission error includes the neglected part of the spectrum above L, in principle up to infinity, but in practical computations truncated at some maximum degree $lmxs$. Furthermore, due to the Wiener filter, the omission error includes that contribution from degrees below L for which the signal to noise ratio is smaller than one. This part, in fact, stems from the imperfection of the measurement process. For the omission part we used the TR degree variance model. In the results presented above we truncated the summation for the omission part at $lmxs = 1000$.

The magnitudes of the degree variances computed with the TR model gradually decrease with increasing degree. The signal power contained in the upper parts of the spectrum therefore will be small. In order to see whether a maximum degree of 1000 is reasonable or not (i.e. the power above degree 1000 is negligible) we compute here the total error once again, but choose a maximum degree of 10,000. Thereby one should keep in mind that such a model is only based upon observational information up to a certain maximum degree, typically somewhere between 300 and 1000, above which it is speculation. This may become evident by comparing the high spectral part of various existing models, such as Kaula's rule (Kaula, 1966), TR (Tscherning and Rapp, 1974) or Jekeli/Moritz (Moritz, 1980) and (Jekeli, 1978).

Table 4.6 shows the global (commission, omission and total) error for surface gravity anomalies and geoid undulations in three different situations for the cases $lmxs = 1000$ and $lmxs = 10,000$. For reference also the case $lmxs = 240$ is added (i.e. omission error only due to Wiener filtering). It can be seen that the error for geoid undulations does not change if the omission error is summed up to degree 10,000 compared to the case where it is summed "only" up to degree 1000. Only if the numbers would have been presented with more decimal digits, differences could be detected, but then the numbers would pretend to be very accurate, which is not the case. Since the geoid signal is relatively smooth, this result is not surprising. The error for gravity anomalies only increases slightly (0.01 mgal), the gravity field being a little rougher than the geoid. In view of the present accuracy, this test confirms our procedure to carry out the omission error summation only up to a maximum degree $lmxs = 1000$.

The part of the omission error for degrees below 240 is relatively small, at least for the case $\beta_{min} = 0, I = 90°$ and $\beta_{min} = 4, I = 90°$. For the other case ($I = 95°.3$)

Table 4.6 *Commission, omission and total error for the observation combination* $\{V_{xx}, V_{xy}, V_{xz}, V_{yy}, V_{yz}, V_{zz}\}$ *and* $h = 200$ *km,* $\Delta t = 4$ *s,* $T_r = 6$ *months,* $\sigma = 0.01\ E/\sqrt{Hz}$ *,* $L = 240$ *and* $\theta_c = 1°$.

	gravity anomalies (mgal)			geoid undulations (cm)		
	commission	omission	total	commission	omission	total
$I = 90°$ and $\beta_{min} = 0$						
$lmxs = 240$	0.73	0.14	0.75	2.34	0.39	2.37
$lmxs = 1000$	0.73	3.63	3.70	2.34	8.32	8.64
$lmxs = 10,000$	0.73	3.64	3.72	2.34	8.32	8.64
$I = 90°$ and $\beta_{min} = 4$						
$lmxs = 240$	0.74	0.14	0.76	8.38	0.40	8.39
$lmxs = 1000$	0.74	3.63	3.71	8.38	8.32	11.81
$lmxs = 10,000$	0.74	3.64	3.72	8.38	8.32	11.81
$I = 95°.3$ and $\beta_{min} = 0$						
$lmxs = 240$	5.90	5.55	8.10	29.78	24.49	38.56
$lmxs = 1000$	5.90	6.63	8.87	29.78	25.86	39.44
$lmxs = 10,000$	5.90	6.64	8.88	29.78	25.87	39.44

Table 4.7 *Commission, omission and total error for the observation combination* $\{V_{xx}, V_{xy}, V_{xz}, V_{yy}, V_{yz}, V_{zz}\}$ *and* $h = 200$ *km,* $\Delta t = 4$ *s,* $T_r = 6$ *months,* $\sigma = 0.01\ E/\sqrt{Hz}$ *,* $L = 240$*,* $\theta_c = 1°$ *and a stabilized solution.*

	gravity anomalies (mgal)			geoid undulations (cm)		
	commission	omission	total	commission	omission	total
$I = 90°$ and $\beta_{min} = 0$						
$lmxs = 240$	0.72	0.13	0.73	2.31	0.37	2.34
$lmxs = 1000$	0.72	3.63	3.70	2.31	8.32	8.63
$lmxs = 10,000$	0.72	3.64	3.71	2.31	8.32	8.63
$I = 90°$ and $\beta_{min} = 4$						
$lmxs = 240$	0.73	0.13	0.74	8.37	0.38	8.38
$lmxs = 1000$	0.73	3.63	3.70	8.37	8.32	11.80
$lmxs = 10,000$	0.73	3.64	3.72	8.37	8.32	11.80
$I = 95°.3$ and $\beta_{min} = 0$						
$lmxs = 240$	0.84	0.14	0.85	3.08	0.39	3.11
$lmxs = 1000$	0.84	3.63	3.73	3.08	8.32	8.87
$lmxs = 10,000$	0.84	3.64	3.74	3.08	8.32	8.87

the propagated error is so high that a large part of the omission error comes from degrees below 240. Obviously, for this case, the gradiometer contributes little to the outcome.

The large errors in the case $I = 95°.3$ are due to the instability of the solution, as pointed out in the previous sections. For comparison we therefore also give a similar table (table 4.7), now computed from the stabilized solutions. The cases $\beta_{min} = 0, I = 90°$ and $\beta_{min} = 4, I = 90°$ are hardly affected. Compare this to the figures 4.14 and 4.15. For all cases, the difference between $lmxs = 1000$ and $lmxs = 10,000$ is still very small. On the other hand, for the case $I = 95°.3$, the contribution to the omission error from degrees below 240 decreases substantially, as does of course the commission part.

Another point which is important in this context, is the smoothing of the total error. If we look at equation 4.14, we see a smoothing operator β_l is included. This is done, because a global recovery of the gravitational potential always has some limited resolution as a result of the finite sampling distance of the measurements. This sampling distance limits the maximum obtainable degree, thereby introducing an error to the potential solution. In terms of equal angular blocks, the resolution to be obtained from a gradiometric analysis up to some maximum degree L is typically $180°/L$. If no smoothing is carried out, the global r.m.s. for gravity anomalies or any other quantity computed from a potential coefficient set which is derived from a global analysis, represents the error in discrete points in which all frequencies are present. This error would be very large, mainly due to the relatively poor determined higher frequencies and the frequencies in the omission part. This is not realistic, as we know that the analysis aims for, and has, a limited resolution. For certain applications (especially global gravity field investigations for the purpose of e.g. oceanography), the r.m.s. values should therefore be considered as block–averages over blocks with size induced by the maximum degree of the analysis. For other applications, e.g. regional geophysical investigations, the high frequency part should be obtained from a dense network of local gravity measurements. In order to obtain block–averages, a smoothing operator β_l is used which in fact damps out the high frequencies, aiming at a decrease of the sampling error.

In our analysis, we used a maximum degree $L = 240$ for the potential recovery, corresponding to a resolution of $0°.75 \times 0°.75$. But as we have seen that the higher degrees are only poorly determined and since the goal for gradiometry is a resolution of 100 km (corresponding to $1° \times 1°$) we introduced a block–size of $\theta_c = 1°$ for the smoothing operator. Nevertheless, in table 4.8 the influence of the smoothing is shown. There we computed the global r.m.s. for three cases: $\theta_c = 0°$ (i.e. no smoothing), $\theta_c = 0°.75$ and $\theta_c = 1°$, cf. (Rapp, 1989). From this table we see that the influence of smoothing is much larger than that of the maximum degree of the omission error. No smoothing gives unsatisfactorily error estimates. Again, geoid undulations are much less affected by both aspects than gravity anomalies. It can also be seen that the two aspects (inclusion of omission error above L on the one hand and smoothing on the other) have opposite effects on the global r.m.s. Inclusion of

Table 4.8 *Global r.m.s. for several block-sizes. Specifications as in table 4.6. $I = 90°$ and $\beta_{min} = 0$.*

	gravity anomalies (mgal)			geoid undulations (cm)		
	commission	omission	total	commission	omission	total
$\theta_c = 0°$						
$lmxs = 240$	1.43	0.29	1.46	4.32	0.81	4.40
$lmxs = 1000$	1.43	21.71	21.75	4.32	34.56	34.83
$lmxs = 10,000$	1.43	27.86	27.90	4.32	35.23	35.49
$\theta_c = 0°.75$						
$lmxs = 240$	1.00	0.19	1.02	3.09	0.55	3.14
$lmxs = 1000$	1.00	6.66	6.73	3.09	14.90	15.22
$lmxs = 10,000$	1.00	6.68	6.75	3.09	14.90	15.22
$\theta_c = 1°$						
$lmxs = 240$	0.73	0.14	0.75	2.34	0.39	2.37
$lmxs = 1000$	0.73	3.63	3.70	2.34	8.32	8.64
$lmxs = 10,000$	0.73	3.64	3.72	2.34	8.32	8.64

omission error above L increases the global r.m.s. because we have to account for the neglect of the higher spectrum parts. But smoothing decreases the global r.m.s. by gradually removing the influence of these higher degrees, which, nevertheless, remain present.

4.2.8 Aristoteles

In previous sections deviations from an ideal case were investigated by considering either a band limitation or a non–polar orbit. A combination of both effects will be investigated in this section, aiming at the situation of ESA's Aristoteles mission (see section 2.3). We do, however, not include in our analysis the possible contribution from GPS observations, we only give "gradiometry only" results. An error analysis for the combination "GPS + gradiometry" can be found in (Schrama, 1990) or (Visser, 1992).

Concerning the band limitation we refer to section 4.2.4. First of all we choose β_{min} equal to 4 in order to account for the non–gravitational effects. Secondly, Aristoteles will have a lower band limit of 0.005 Hz, corresponding approximately to

$\beta = 27$. Below 27, the frequencies will be available but with degraded accuracy. The 0.01 E/$\sqrt{\text{Hz}}$ white noise error spectrum applies to the frequency band between 0.005 and 0.125 Hz (the latter approximately corresponding to $\beta_{max} = 664$). Following (Schrama, 1990) a $1/\beta$ behaviour is assumed for the frequencies between 4 and 27. Between 4 and 27 the precision per frequency then becomes $(27/\beta_{km}) \cdot \sigma_0$ with $\sigma_0 = 0.01$ E/$\sqrt{\text{Hz}}$. Such a coloured noise situation will be assumed in the present section.

As for the inclination, we choose here $I = 92°.3$, despite the fact that Aristoteles will fly for 6 months in an orbit with inclination $I = 95°.3$. The reason is that in the two weeks after this 6 months phase, in which the orbit of the satellite will have an inclination of 92°.3, a so dense coverage of the polar regions is achieved (of course a small cap of 4°.6 excluded) that the complete mission may safely be assumed to be flown in this 92°.3 inclination orbit. The important mission parameters and specifications are listed in table 4.9.

Table 4.9 *Characteristics of the Aristoteles gradiometer mission.*

inclination	$I = 92°.3$
components	V_{yy}, V_{zz}
band limitation	$\beta_{min} = 4$
satellite height	$h = 200$ km
sampling interval	$\Delta t = 4$ s
mission duration	$T_r = 6$ months
error spectrum	0.01 E/$\sqrt{\text{Hz}}$ white noise above 27 c.p.r.
	$1/\beta$ behaviour between 4 and 27 c.p.r.

Figure 4.20 shows the error r.m.s. values for the present specifications. In this figure again the results for all 10 components and combinations are given. Since Aristoteles only measures V_{yy} and V_{zz} with sufficient accuracy the lowest of the three graphs is of special interest here. The others are included for reference and for preserving a uniform presentation. Comparing with figure 4.7 (in which $\beta_{min} = 4$ and $I = 90°$) it appears that the latter is little worse, despite the fact of complete global coverage when $I = 90°$. This fact was already observed and discussed in section 4.2.5. As a result of the lower band limit and the coloured noise the results for low degrees (approximately below 30) are not satisfactory.

For this reason also a stabilized solution was computed for the present case (shown in figure 4.21), which, however, appears to have mainly effect on high degrees. Comparing the unstabilized with the stabilized solution one can see that some

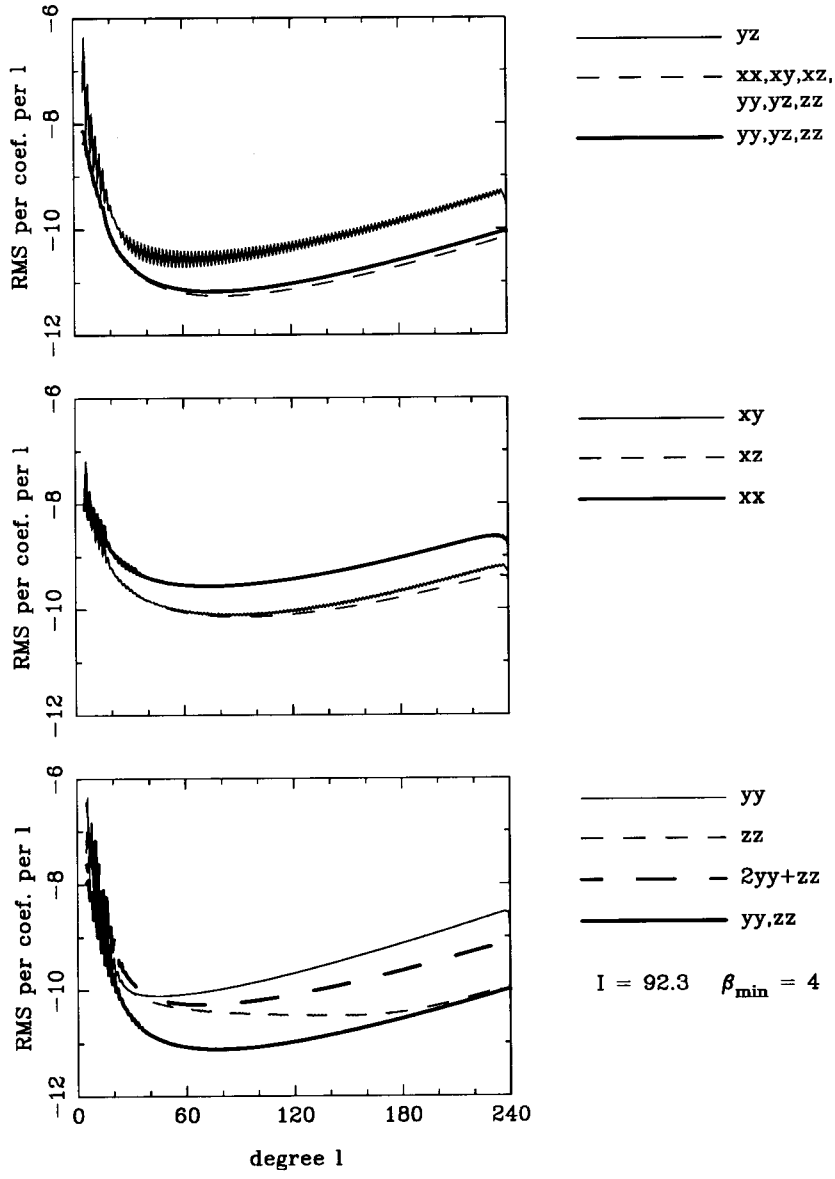


Figure 4.20 *Propagated error from gradiometry with the following specifications: $I = 92°.3$, $\beta_{min} = 4$, $T_r = 6$ months, $\Delta t = 4$ s, $h = 200$ km, $\sigma = 0.01\ E/\sqrt{Hz}$ above 27 c.p.r. and $1/\beta$ behaviour between 4 and 27 c.p.r.*

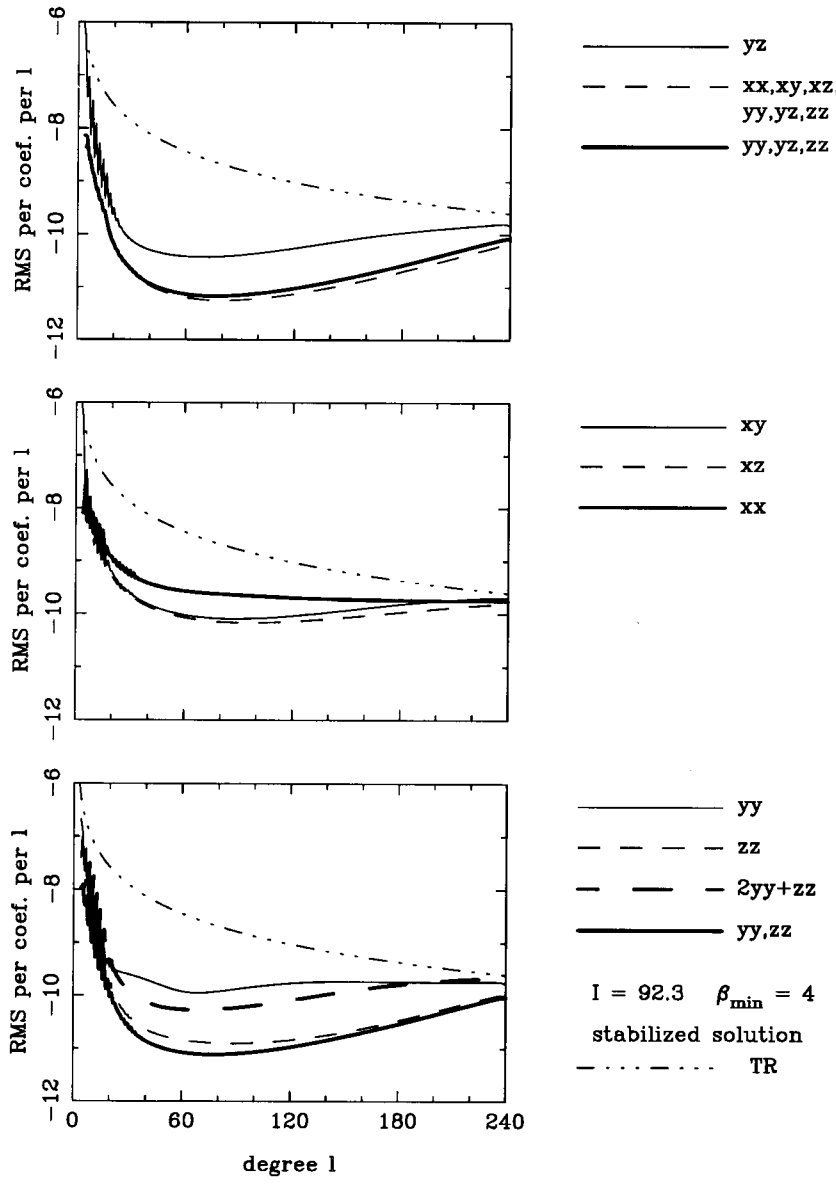


Figure 4.21 *Propagated error from gradiometry. Specifications as in figure 4.20, but now as a stabilized solution.*

Table 4.10 *Global r.m.s. for several observation types. Specifications as in figure 4.20.*

	gravity anomalies (mgal)			geoid undulations (cm)		
	commission	omission	total	commission	omission	total
unstabilized solution						
V_{yy}	5.53	7.54	9.35	613.67	396.20	730.46
$2V_{yy} + V_{zz}$	4.14	5.18	6.63	211.89	89.89	230.18
V_{zz}	2.97	4.61	5.48	388.86	453.81	597.62
$\{V_{yy}, V_{zz}\}$	1.15	3.64	3.82	44.11	8.49	44.92
stabilized solution						
V_{yy}	4.49	4.07	6.06	443.92	163.34	473.01
$2V_{yy} + V_{zz}$	3.52	4.00	5.33	162.47	32.61	165.71
V_{zz}	2.27	3.70	4.34	283.53	103.16	301.71
$\{V_{yy}, V_{zz}\}$	1.12	3.64	3.81	43.92	8.48	44.73

components become worse for lower degrees in the latter case (e.g. V_{yy}). However, this conclusion may be misleading, since contributions from singular blocks are excluded in the unstabilized case (compare table 4.5) and since we saw in section 4.2.6 that for a proper description of the error of a stabilized solution we should compute the MSE, equation 4.16, which includes the bias term.

Finally, table 4.10 gives global r.m.s. values for surface gravity anomalies and geoid undulations for the observation types V_{yy}, $2V_{yy} + V_{zz}$, V_{zz} and $\{V_{yy}, V_{zz}\}$. Also the results of the stabilized solution are included. The numbers in this table represent $1° \times 1°$ block averages (i.e. smoothing operator included). The maximum degree for the omission error is 1000.

The most important conclusion from this table is that if geoid undulations have to be derived from the present mission scenario, additional information is necessary since the results do not meet the goals from section 2.2. Even the stabilized solution gives too large errors for geoid undulations. The cause of this is the lower band limit of 4 c.p.r. and the coloured noise of the lower part of the measurement spectrum, which influence the geoid undulations rather heavily due to the relative smoothness of the signal. Gravity anomalies, on the other hand, already give satisfactory results, at least for the stabilized solution (although of course biased in this case). They are more sensitive to higher degrees than geoid undulations. Obviously, the higher part

of the spectrum is determined well enough by the gradiometer. In (Schrama, 1990) or (Visser, 1992) it is shown that additional GPS tracking information can be used to obtain long wavelength gravitational information, in which case also geoid undulations will be derived with the desired precision. The two techniques, GPS for the lower part of the spectrum and gradiometry for the higher part, are thus complementary, and the combination of the two on Aristoteles will meet the requirements of section 2.2.

Finally, is can also be seen from table 4.10 that stabilization in the present case mainly influences the omission part of the total error. This can be explained as follows. In the unstabilized case, the propagated error is relatively large, but the Wiener filter decreases its contribution to the commission part of the total error, thereby at the same time increasing the omission part. Since we have seen that the omission part, due to neglecting the degrees above L, is small, the large value of the omission part in the unstabilized case must be mainly due to the mentioned filter. In the stabilized situation, the propagated error gives significantly better results (compare figures 4.20 and 4.21), causing the filter to letting pass more commission error (however with less power). As a result the omission part decreases substantially.

4.2.9 Some computational aspects

The software for the error analysis runs on a CONVEX vector computer. This is a so-called SIMD machine (see section 3.3), nevertheless equipped with 4 CPU's which offer some possibilities of parallelization. Despite the fact that the optimalization capabilities of this computer (and especially the vectorization techniques) in general greatly reduce execution time, a single run of the error analysis program takes as much as 500 CPU seconds (for all 10 observation types). This, of course, depends mainly on the maximum degree L (and minimum degree) of the analysis, the mentioned time referring to an analysis up to degree 240. The number of degrees up to L determines the size of the blocks of the normal matrix, the range of the orders m between *mmin* and *mmax* determines the number of blocks in the normal matrix, though one usually takes *mmin* equal to 0 *mmax* equal to L.

Whereas one often expects matrix inversion to account for the major part of CPU time, it appeared here that setting up the normal matrix blocks is more costly. This is mainly due to the computation of the inclination functions $\bar{F}_{lm}^{k}$ and $\bar{F}_{lm}^{k\,*}$, which is responsible for approximately 46 % of the total CPU time. One could suggest to compute the inclination functions only once and store them for future reference. This costs, however, much memory space. Storing all inclination functions and cross-track inclination functions up to degree L requires the storage of $\frac{1}{6}\,(L\,+\,1)\,(L\,+\,2)\,(4\,L\,+\,3)$ numbers (Sneeuw, 1991a). If represented by double precision numbers this corresponds, for $L = 240$, to over 71 Mbyte, for one inclination only. For many inclinations, the memory costs may be too much.

We tried two methods for the computation of these inclination functions. The first one is described in appendix C and is based on an algorithm used by Wagner

(1983). It evaluates a kind of unit potential function at equidistant points along a great circle (which represents the satellite's orbit). With an FFT routine one obtains the Fourier coefficients of this function from which the inclination functions can be derived. For the evaluation of the unit potential we need to compute the Legendre functions. Even if one makes optimal use of the symmetry properties of these functions, cf. (Schrama, 1989), the largest part of CPU time goes to the recursive computation of the Legendre functions. The FFT is relatively fast (as it should be, considering its name).

The second method for computation of the inclination functions is by using recurrent relations. A subroutine package programmed by N. Sneeuw (Sneeuw, 1991a) and based on an algorithm by Emeljanov and Kanter (1989) was used for this purpose, with a supplementary routine for the cross–track inclination functions $\bar{F}_{lm}^{k\,*}$ (Sneeuw, 1991b). Since some of the recurrences for the inclination functions are unstable and overflows are likely to occur, extensive use is made of scaling techniques, which become especially important if the inclination functions are required for high degrees.

A few differences between the two methods are to be mentioned. First of all, the second (recurrent) method appeared to be slower. One of the main reasons for this is that recurrent relations are (almost) not suitable for vectorization. Furthermore, the method of scaling makes extensively use of conditional statements, which, especially if used inside inner DO–loops, prevent vectorization. For the FFT in the "unit potential" method a FORTRAN programmed subroutine from a CONVEX mathematical library was used, which is especially tailored for vectorization. A more conventional FFT routine, made for use on scalar computers, proved to be up to 6 times slower for FFT's of the order 1024 if run on the CONVEX.

Whereas the two methods give identical results in nearly all cases, small differences may occur in special situations. The latter are mainly the situations when the inclinations functions ought to be zero. The first method will almost always give zero values in these cases, but the second method may give small values of the order 10^{-15} or 10^{-16} due to numerical round–off errors during the recursions. For example, if $I = 90°$ and $m = 0$, all $\bar{F}_{lm}^{k\,*}$ are identically zero, resulting in singular blocks of the normal matrix for the components V_{xy} and V_{yz}. The recursive method, however, would result in $\bar{F}_{lm}^{k\,*}$ values of the order of 10^{-15}, so that not all elements of the $m = 0$ blocks will be equal to zero. If none of the diagonal elements of such block is exactly zero, the block may not be singular, so that, after inversion, very large values for the a–posteriori variances appear. Therefore, care has to be taken with the recursive method.

Since we do not need the complete inverse normal matrix, but only the diagonal elements, in reality the inverse is not computed at all. A Cholesky factorization is carried out on the normal matrix N, resulting in an upper triangular matrix R, such that $N = R^T R$. The i–th diagonal element of the inverse normal matrix, N_{ii}^{-1}, is then computed in the following way:

$$N_{ii}^{-1} = s_i^T s_i$$

in which the expression on the right hand side is the inner product of the vector s_i with itself. This vector s_i is computed by solving the triangular system

$$R^T s_i = e_i$$

where e_i is a vector containing only zero's except the i–th element, which equals 1. The computation of the diagonal elements of the inverse normal matrix in this way costs about 7 % of the total CPU time of the error analysis.

4.2.10 Conclusions

Due to the assumptions stated in section 4.2 the error analysis carried out here refers to a somewhat idealized situation. Of the six assumptions given there (circular orbit, regular data distribution, no ground-track repeat during mission period, measurements are considered to be averaged over sampling period, non–gravitational effects excluded, measurements are uncorrelated and of equal variance), the second one is probably the most critical one. As already mentioned, data gaps are likely to occur. A non–continuous data stream destroys the orthogonality properties on which in fact the whole analysis is based. Without orthogonality, the normal matrix becomes a full matrix, leading to an enormous amount of extra work for the inversion.

In an idealized situation (polar orbit and no band limitation) the mission goals in terms of r.m.s. values for derived gravimetric quantities like gravity anomalies and geoid undulations can be met in terms of accuracy level as well as resolution, at least for a full tensor gradiometer. Nevertheless, figure 4.2 suggests that also a planar gradiometer (of the type as will be used in the Aristoteles mission) may give satisfactory results in such an idealized situation, as numerical verification showed (although these results are not given here). Whereas satellite altitudes lower than 200 km seem to improve the results not very much, higher altitudes are certainly not desirable. The same applies for the mission duration. Finally, the measurement precision has somewhat more influence, although mainly for higher altitudes than 200 km.

We considered two major deviations from an idealized situation: (lower) band limitation and non–polar orbits. Of these two, lower band limitation has the largest impact on the results. The influence of a non–polar orbit can be greatly reduced by stabilization. A negative consequence of stabilization, biasedness of the estimates, appeared to have not so much influence, although it remains of course present and one has to bear that in mind.

Whereas stabilization may help for non–polar orbits, it is no remedy against band limitation, as can be concluded from section 4.2.6. Lower band limitation significantly degrades the lower degrees of the spectrum, whereas stabilization seems

to improve primarily the higher degrees. As a result too large biases occur for the lower degrees. For this problem, only the addition of supplementary observations, such as GPS tracking, can be of any real help (Schrama, 1990), (Visser, 1992).

For Aristoteles we assumed the total 6 months mission to be flown in a 92°.3 inclination orbit. Whereas this is not actually the case (only for an additional two weeks), the conclusions above indicate this assumption not to be very critical. As for the lower band limitation, this may really jeopardize the success of the mission, especially if we take a look to the r.m.s. values of geoid undulations (table 4.10). Additional GPS observations really become very important to obtain the long wavelength part of the spectrum with high enough precision. Two complementary techniques are necessary, GPS for the low degrees and gradiometry for higher degrees. Indeed, a GPS receiver is planned to be on board the Aristoteles satellite.

Inclusion of omission error and smoothing have a major influence on the presented global r.m.s. results. As for the omission error, neglecting the degrees above 1000 does not make much of a difference, but, especially in non–ideal cases, omission parts for degrees below L do have some influence. In fact, the omission error part determines some lower limit for the global r.m.s., see section 4.2.3.

Much more influence than the inclusion of the omission error has the introduction of a smoothing operator into the global r.m.s. computations. Without such an operator, r.m.s. values become extremely high. However, for our purposes, representation of the results in terms of block–averages is justified by taking into consideration the difficulty of deriving high degrees (although present in the true signal) globally from any measurement technique due to the inevitable limited sampling distance and mission altitude. If the goal of a gradiometric mission is a resolution of approximately 100 km, block–averages for $1° \times 1°$ blocks are appropriate.

4.2.11 Other error analysis methods

The error analysis presented so far was first used by O.L. Colombo (Colombo, 1987). It uses the actual gradiometric measurements as observations, related to the unknown potential coefficients by a linear model of the type 4.2. Another possibility is described in (Schrama, 1990). There, lumped coefficients A_{km}, B_{km} (in eq. 3.19) are used as observations. Equations of the type 3.20 are used as model equation, with for each gradient a different expression for H_{lmk}.

The lumped coefficients are determined from the observed gradients by means of a Fourier transformation. They serve so–to–say as pseudo observables for the next step, in which the potential coefficients are derived from them using equation 3.20 as observational model.

Both methods of error analysis, the one in which the observed gradients or the one in which lumped coefficients are taken as observations, agree in the sense that they both treat the gradient measurements, given in subsequent points along a satellite orbit, as a time series. The methods will therefore be called *time–like* (Koop et al., 1989). Furthermore, since the method used in this work takes the

gradient measurements themselves as observations, it is referred to as *time–like in the time domain.* The method which uses the lumped coefficients as observations may then be referred to as *time–like in the frequency domain.*

In (Koop et al., ibid.) also a different approach is mentioned. Instead of considering the measurements as being functions of time, one may also consider them as functions of space. One then assumes the measurements to be given on a sphere or in a spherical shell at satellite altitude. Such an approach is referred to as *space–like*[1]. It is equivalent to solving a geodetic boundary value problem. Depending on assumptions about the spatial distribution of the measurements and the a–priori variances, different strategies are possible.

First one may assume to have a complete, continuous global coverage of the earth with measurements. In this case an analytical solution for the inverse of the normal matrix exists (Rummel et al., 1989), at least for V_{zz}, $\{V_{xz}, V_{yz}\}$ and $\{2V_{xy}, V_{xx} - V_{yy}\}$. In reality, of course no continuous coverage is available. This fact is taken into account by weighting the a–priori covariance matrix in such a way that some assumed finite spatial resolution enters the solution. This method is referred to as *space–like continuous.*

The idea can be modified by assuming we are given a finite number of measurements randomly distributed inside some global spherical shell containing the satellite's orbit. This spherical shell is divided into equal–angular cells, inside of which the measurements are averaged per cell. Observations are now assumed to be regularly distributed point values (or block mean values), the observational noise being adapted accordingly, see e.g. (Rapp, 1989). Using a spherical harmonic expansion for the gradients, the normal matrix takes on a block–diagonal structure, due to the orthogonality of the sine and cosine series, and can be easily inverted. This method is referred to as *space–like discrete.* In (Koop et al., 1989) it is shown that all four mentioned methods (*time–like in the frequency domain*, *time–like in the time domain*, *space–like continuous* and *space–like discrete*) yield the same results.

4.3 Global recovery

Whereas the previous section (section 4.2) "only" discussed the quality of the potential coefficients to be derived from space–borne gradiometry (in terms of a–posteriori error r.m.s. values), we are n fact primarily interested in those estimated potential coefficients $(\bar{C}_{lm}, \bar{S}_{lm})$ themselves. As already explained in section 4.1 we try to solve potential coefficients from globally distributed gradiometric data. This process is indicated with the term *global recovery.* There are several methods of global recovery, one of them will be discussed in the present section. In order to carry out a global recovery of potential coefficients, we need a data set of measured gravity gradients. Such a set is not yet available, the first set has to be generated by

[1]The terms space–like and time–like used in this context have nothing to do with the same terms known from the theory of relativity.

the Aristoteles mission. Therefore, in this section, a *simulated* set of gradient data is used, computed at the Center for Space Research of the University of Texas at Austin, (Schutz et al., 1987) and (Schutz et al., 1988). A description of this simulation will be given in section 4.3.4. We will, for convenience, denote in the sequel the potential coefficients by $\{\bar{C}_{lm\alpha}|\alpha = 0, 1\}$, meaning $\bar{C}_{lm}$ if $\alpha = 0$ and $\bar{S}_{lm}$ if $\alpha = 1$.

4.3.1 Linear model

The observed gravity gradients are functions of the location of measurements and of the gravity field. In general, the location (e.g. the coordinates of the measurement point in some coordinate system) as well as the gravity field (e.g. expressed by means of a series expansion in terms of potential coefficients) are unknown. These are exactly the unknowns which we like to derive from the measurements. To this extent, the measurements are linked to a mathematical model, describing (approximately) the physical reality. In section 4.1 such a mathematical model (e.g. equation 4.1 for second–order radial derivatives) was the starting point for our discussion on (least squares) error analysis. In that section, we simply assumed to be given a global, regularly distributed set of gradient observations with some known measurement error. The given model equation was assumed to properly describe the physical reality and the unknown potential coefficients were assumed to be obtainable from the measurements by means of a least squares adjustment. All this meant that we implicitly assumed that the location of the measurement points was known, i.e. circular orbit with known radius and equal spaced measurements along the orbit to invoke orthogonality of the trigonometric series. Also the orientation of the local coordinate system in which the measurements are given was assumed to be known.

In reality, the coordinates of the measurement points are not exactly known, only approximately. Then the model equation which we used becomes non–linear. In this section we will show one possibility of dealing with the non–linear problem, a linearization procedure following (Rummel and Colombo, 1985) for gradiometry or in general following (Rummel, 1985a).

Consider an observation point P on the true (or actual) satellite orbit. Initially, the gradiometer output (i.e. the gradient observations) is given in an instrument or satellite coordinate system which has its origin in P. But as already mentioned in chapter 3 we assume that the gradients can be transformed to a local orbital or local north–oriented coordinate system (with origin also in P) by means of a known rotation matrix. However, since the actual orbit is unknown, the orientation of such local system in P is also unknown. Suppose now that we are given the coordinates of a nearby point P' on a known approximate (or nominal) orbit. Then we may take for the orientation of the local system the orientation of such a system in P', see figure 4.22. We will denote the local coordinate system situated in P but with the orientation of the system in P', with x^i, and assume we are given the components of the gradient tensor, $V_{ij}(P)$, in this system (the coordinates of P of course remaining unknown).

If we were given an initial state vector, the nominal orbit could be computed by means of integrating the dynamic equations of motion of the satellite, using as force model a gravitational field represented by some known set of potential coefficients (up to some maximum degree and order), and using the measurements from some satellite tracking system. The initial state vector is usually given in terms of the three location coordinates in a geocentric cartesian coordinate system (X, Y, Z) and the three components of the velocity vector $(\dot{X}, \dot{Y}, \dot{Z})$ at the initial time t_0. The differences between the true orbit and the nominal orbit are the orbit errors. The orbit errors contain a geometrically induced part (due to errors in the initial state vector) and a gravitationally induced part (due to the imperfect choice of a potential coefficient set for the orbit integration). This means that the observations are doubly dependent on the gravitational parameters, directly through the measurement type itself and indirectly by means of the dynamics of the satellite orbit, the latter being reflected in the orbit errors (Betti and Sansò, 1989). Furthermore, the orbit errors contain contributions from various other sources, of which atmospheric drag will be the most important. The latter error parts will, however, at the moment not be considered.

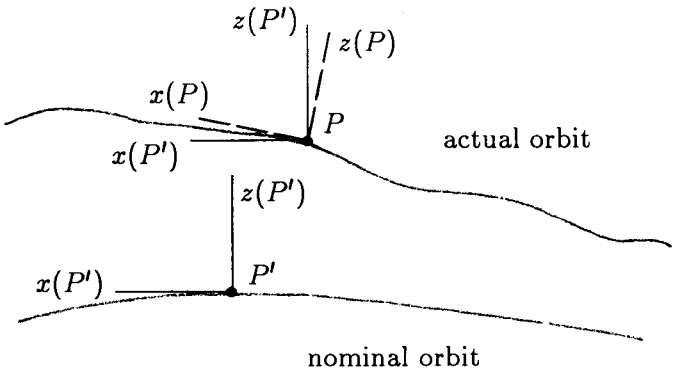


Figure 4.22 *Location of actual and approximate points P and P' and the local coordinate systems in those points.*

What follows is the introduction of a mathematical model to which the observations are linked, e.g. equations like the ones derived in section 4.1. In general we write:

$$\tilde{\boldsymbol{\ell}}(P,V) = \boldsymbol{\ell}(P,V) + \tilde{\varepsilon}$$

where $\tilde{\boldsymbol{\ell}}(P,V)$ are the observables as function of the point P and the gravitational potential V, $\boldsymbol{\ell}$ represents the mathematical model and $\tilde{\varepsilon}$ the (small) residuals between the observables $\tilde{\boldsymbol{\ell}}$ and the model $\boldsymbol{\ell}$. The "~" denotes stochastic quantities. In general, the residuals not only contain measurement noise but also model errors.

For our purpose, however, we assume that the stochastic behaviour of the residuals $\tilde{\varepsilon}$ after linearization is properly described by their first moment (expectation) to be zero, i.e. $E\{\tilde{\varepsilon}\} = 0$ (zero mean) so that $E\{\tilde{\boldsymbol{\ell}}\} = \boldsymbol{\ell}$ (unbiasedness), and its second moment (error variance) to be $E\{\tilde{\varepsilon}\tilde{\varepsilon}^T\} = Q$, with Q the error covariance matrix of the observations. In general, the mathematical model $\boldsymbol{\ell}(P, V)$ will be non–linear in P and V. The linearization procedure therefore consists of two parts, a linearization with respect to the gravitational potential V and one with respect to the coordinates $x^i(P)$ of the point P.

The former is done by splitting the gravitational potential V into a known normal (or reference) part U and an unknown disturbing part T:

$$V(P) = U(P) + T(P) \; .$$

Hence for the second–order derivatives we have

$$V_{ij}(P) = U_{ij}(P) + T_{ij}(P) \; ,$$

where ij may take any of the values xx, xy, xz, yy, yz or zz. The observation equation thus becomes:

$$\tilde{V}_{ij}(P) = U_{ij}(P) + T_{ij}(P) + \tilde{\varepsilon}_{ij} \; . \tag{4.17}$$

The second step of the linearization process is the linearization with respect to the location P, for which we expand $U_{ij}(P)$ into a Taylor series with respect to the coordinates of the known approximate point P', truncated after the linear term:

$$U_{ij}(P) \approx U_{ij}(P') + U_{ijk}(P')\Delta x^k$$

where $U_{ijk} = \frac{\partial U_{ij}}{\partial x^k}$ and $\Delta x^i = x^i(P) - x^i(P') = (\Delta x, \Delta y, \Delta z)$ the coordinate corrections between P and P', i.e. the orbit corrections between the actual and the nominal orbit. It is convenient (although not necessary) to compute the nominal orbit using the same potential U as is used above for the potential linearization.

Since the disturbing potential T is a small quantity of first order we assume $T_{ij}(P') \approx T_{ij}(P)$. Inserted into equation 4.17 this yields:

$$\tilde{V}_{ij}(P) = U_{ij}(P') + U_{ijk}(P')\Delta x^k + T_{ij}(P') + \tilde{\varepsilon}_{ij} \; ,$$

or

$$\begin{aligned} \Delta\tilde{\Gamma}_{ij} &\equiv \tilde{V}_{ij}(P) - U_{ij}(P') \\ &= U_{ijk}(P')\Delta x^k + T_{ij}(P') + \tilde{\varepsilon}_{ij} \end{aligned} \tag{4.18}$$

with $\Delta\tilde{\Gamma}_{ij}$ the gradient *anomaly*. In the simplest case the normal potential U is the potential of a spherically symmetric gravitational field (i.e. $U = \frac{GM}{r}$), but it may also be the potential of an ellipsoidal field or any other higher order approximation of the actual gravitational field. In general, the normal potential U is expanded into

a spherical harmonic series (or a series of inclination functions) represented by some (limited) set of harmonic coefficients $\bar{C}'_{lm\alpha}$. With the actual potential V expanded into a similar series with coefficients $\bar{C}_{lm\alpha}$ we find for the disturbing potential T an expansion in terms of the coefficient corrections $\Delta\bar{C}_{lm\alpha}$, where $\bar{C}_{lm\alpha} = \bar{C}'_{lm\alpha} + \Delta\bar{C}_{lm\alpha}$. In equation 4.18 the unknowns are the potential coefficient corrections $\Delta\bar{C}_{lm\alpha}$ in $T_{ij}(P')$, and the coordinate corrections Δx^i.

It is known from the theory of adjustment that, although the choice of approximate values does not influence the results (as long as they are inside the bounds of convergence so that the iterative estimation process will converge), it is important to compute these approximate values (in our case the normal potential U in the approximate point P') accurately, in order to obtain small anomalies and thus accelerating the iteration process. On the other hand, the coordinate corrections Δx^i will, in general, also be small so that the elements of the coefficient matrix $U_{ijk}(P')$ can be computed with less accuracy. Whereas one usually takes for U a series expansion up to some (high) maximum degree if it concerns the computation of the anomalies (i.e. for $U_{ij}(P')$), a spherical or ellipsoidal U suffices for computation of the coefficient matrix (i.e. for $U_{ijk}(P')$).

Let us take, for example, the potential of a homogeneous spherical mass distribution as simplified normal potential for the computation of the coefficient matrix U_{ijk}. This matrix can then be computed very easily, and if inserted into equation 4.18, it yields (for all six gradient tensor components separately):

$$\begin{bmatrix} \Delta\tilde{\Gamma}_{xx} \\ \Delta\tilde{\Gamma}_{xy} \\ \Delta\tilde{\Gamma}_{xz} \\ \Delta\tilde{\Gamma}_{yy} \\ \Delta\tilde{\Gamma}_{yz} \\ \Delta\tilde{\Gamma}_{zz} \end{bmatrix} = \frac{3GM}{r_{P'}^4} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + \begin{bmatrix} T_{xx} \\ T_{xy} \\ T_{xz} \\ T_{yy} \\ T_{yz} \\ T_{zz} \end{bmatrix} + \begin{bmatrix} \tilde{\varepsilon}_{xx} \\ \tilde{\varepsilon}_{xy} \\ \tilde{\varepsilon}_{xz} \\ \tilde{\varepsilon}_{yy} \\ \tilde{\varepsilon}_{yz} \\ \tilde{\varepsilon}_{zz} \end{bmatrix} . \qquad (4.19)$$

An equation like 4.19 can be set up in each observation point. If we have N_p observation points, we have $3N_p$ coordinate unknowns and $6N_p$ unknown T_{ij}'s. But the gravitational potential is common to all measurement points, so, expressing the potential by a series expansion up to some maximum degree L, the potential coefficient corrections $\Delta\bar{C}_{lm\alpha}$ are the same for all points, leaving us with $(L+1)^2 - 3$ potential unknowns (which is the total number of unknown coefficients in a series expansion up to degree L minus the three first–degree coefficients ($l = 1$) which are identically zero if the origin $r = 0$ of our geocentric coordinate system is the center of gravity of the earth). A system of observation equations arises if one joins the model equation 4.19 for all observation points together as function of the same set of potential coefficients.

Now one may proceed in different ways. Either one solves the system of equations for all unknowns simultaneously (i.e. position as well as potential unknowns). This method is used in the so–called theory of *integrated geodesy*, see e.g. (Hein, 1986) or (Moritz, 1980). Or one eliminates certain coordinate unknowns by forming appropriate linear combinations of anomalies, a method often used in the theory of the geodetic boundary value problem. One is then left with an adjustment problem with only potential coefficient corrections as unknowns. After they have been estimated, back substitution into the model equation gives estimates for the coordinate unknowns. Iteration of this procedure should converge to the right results. The next section will discuss such a procedure.

4.3.2 Iteration

In this section a possible iteration procedure will be sketched which can be used to solve potential coefficients (and coordinate unknowns) from gradiometric observations. The procedure used here was first presented in (Rummel and Colombo, 1985). The first step in this method is to carry out a pre–adjustment of the diagonal components of the gradient tensor using as model equation the Laplace equation (zero trace of the gradient tensor). In the case of Aristoteles this pre–adjustment cannot be carried out since not all diagonal tensor components are measured.

As starting point for the adjustment to follow we use the linearized model 4.19 derived in the previous section. The idea is to separate the solution of the potential coefficients from that of the coordinate unknowns. If we take a look at equation 4.19 we see that, in the present spherical approximation, only $\Delta\tilde{\Gamma}_{xz}$ depends on the coordinate correction Δx, $\Delta\tilde{\Gamma}_{yz}$ only on Δy but that the three diagonal components all depend on Δz. These three can therefore be used to eliminate the radial orbit correction Δz from the model equation. If we assume only $\Delta\tilde{\Gamma}_{yy}$ and $\Delta\tilde{\Gamma}_{zz}$ to be available (like with Aristoteles) then the linear combination

$$2\Delta\tilde{\Gamma}_{yy} + \Delta\tilde{\Gamma}_{zz} \tag{4.20}$$

eliminates Δz. If we take this linear combination in each observation point as new observables, there remains a system of linear equations with only the potential coefficient corrections as unknowns. Note that this is the reason we took this particular combination of gradients $2V_{yy} + V_{zz}$ as one of the combinations in the previous section on error analysis, section 4.2. Now we have separated the determination of the gravitational field from that of the orbit. This is only true up to the order of the earth's flattening, since we took the coefficient matrix of the coordinate corrections in equation 4.19 in spherical approximation.

On the left hand side of the resulting model equation we now have the linear combination $2\Delta\tilde{\Gamma}_{yy}+\Delta\tilde{\Gamma}_{zz}$ and on the right hand side the corresponding combination of the disturbing potential $2T_{yy} + T_{zz}$. This equation can be used to estimate the potential coefficient corrections $\Delta\bar{C}_{lm\alpha}$.

Also, from equation 4.19 we see that the tensor component $\Delta\tilde{\Gamma}_{xy}$ does (in spherical approximation) not depend on any coordinate correction at all. This equation can therefore be used for the estimation of the potential coefficient corrections together with the linear combination above (at least if this tensor component is measured, which is not the case in the Aristoteles mission).

The linear model is now of the same form as equation 4.2, and can be solved by means of a least squares estimation procedure, as described in section 4.1. This results in a set of potential coefficient correction estimates $\Delta\hat{\bar{C}}_{lm\alpha}$, from which we may compute an up–dated set of potential coefficients by adding them to the coefficients $\bar{C}'_{lm\alpha}$ of the normal potential U: $\bar{C}''_{lm\alpha} = \bar{C}'_{lm\alpha} + \Delta\hat{\bar{C}}_{lm\alpha}$.

What should follow is the computation of the coordinate corrections Δz by means of backward substitution of the solved potential coefficient corrections into the linear model eq. 4.19. We saw already that these coordinate corrections are in fact the radial orbit errors, the differences between the true and the nominal orbit in radial direction, and that they contain gravitationally and geometrically induced parts. The orbit errors reflect the unmodelled part of the motion of the satellite. This motion (being the time evolution of the state vector (Betti and Sansò, 1989)) can be described by an initial state vector (x^i and $\dot{x}^i$ at the initial time t_0) and some system of dynamical equations (equations of motion), e.g.: $\ddot{x}_i(t) = V_i(x^j(t)) + P_i$, where P_i are perturbing terms, which are neglected here for convenience, and t is time. In order to include the satellite motion (through the dynamical equation) into the estimation procedure, the determination of the coordinate corrections is divided into two parts. Assuming that the corrected set of potential coefficients $\bar{C}''_{lm\alpha}$ constitutes a better gravitational field representation than the set $\bar{C}'_{lm\alpha}$, we first compute an up–dated nominal orbit with these new coefficients $\bar{C}''_{lm\alpha}$ but with the old initial state vector. In this way we obtain a new set of up–dated approximate points P'' in which we compute new anomalies: $\Delta\tilde{\Gamma}'_{ij} \equiv \tilde{\Gamma}_{ij}(P) - U_{ij}(P'')$. Inserting these new, smaller, anomalies and the estimated potential coefficient corrections $\Delta\hat{\bar{C}}_{lm\alpha}$ into equation 4.19 and assuming $T_{ij}(P'') \approx T_{ij}(P') \approx T_{ij}(P)$, we obtain

$$\Delta\tilde{\Gamma}'_{ij} - T_{ij}(P'') = U_{ijk}(P'')\Delta x^k \tag{4.21}$$

where the U_{ijk} is still in spherical approximation and where the (also smaller) coordinate corrections $\Delta x^k = x^k(P) - x^k(P'')$ are now the orbit errors between the true orbit and the *up–dated* nominal orbit. They are therefore free from the influence of the $\Delta\hat{\bar{C}}_{lm\alpha}$ and are only due to uncertainties in the initial state vector, cf. (Rummel and Colombo, 1985). From this equation coordinate correction estimates $\Delta\hat{z}$ can be obtained.

With the resulting set of coordinate corrections $\Delta\hat{z}$ we may compute an up–dated initial state vector using some linear model. An example of such a linear model is the solution to the homogenous Hill equations (ibid.). The Hill equations are derived from linearizing the equations of motion of the satellite relative to some reference motion (e.g. a circular motion in a spherical gravitational field). For an

exact derivation of the Hill equations and their solution see e.g. (Kaplan, 1976) or (Colombo, 1986). For example, for the radial direction, the homogeneous solution of the Hill equations is given by:

$$\Delta\hat{z}(t) = -(3\Delta z_0 + \frac{2}{n}\Delta\dot{x}_0)\cos nt + \frac{\Delta\dot{z}_0}{n}\sin nt + 4\Delta z_0 + \frac{2}{n}\Delta\dot{x}_0 \tag{4.22}$$

where the initial conditions $\Delta z_0, \Delta\dot{z}_0, \Delta\dot{x}_0$ are considered at $t = 0$ and $n = \sqrt{\frac{GM}{r^3}}$ is the mean motion of the satellite. For each measurement epoch t an equation like 4.22 can be set up. All these equations are function of the same initial conditions $\Delta z_0, \Delta\dot{z}_0, \Delta\dot{x}_0$. Merging these equations into one system, estimates of $\Delta z_0, \Delta\dot{z}_0, \Delta\dot{x}_0$ can be obtained by solving the system of equations by least squares adjustment.

Now we have improved the two error parts contributing to the orbit errors: the gravitational part by improving the potential coefficient set and the geometrical part by improving the initial state vector. As a consequence we may again compute an up–dated nominal orbit with the coefficients $\bar{C}''_{lm\alpha}$ and the improved initial state vector. In fact, from here on, the whole process described above repeats itself, since with this new nominal orbit, up–dated gradient anomalies are derived, the linear combination is formed, potential coefficient corrections are estimated, etc. Assuming the process to converge to the right solution (cf. (Rummel and Colombo, 1985)), one repeats the above procedure until satisfactory accuracy is obtained.

In the case of Aristoteles the procedure ends here. If also the gradients V_{xz} and V_{yz} are measured the procedure may be continued by estimating the other two coordinate corrections Δx and Δy. The coordinate correction Δx can be obtained from $\Delta\tilde{\Gamma}_{xz}$ (which is computed along the final nominal orbit) by inserting the final potential coefficient set resulting from the repetition process above into equation 4.19. In the same way Δy is obtained from $\Delta\tilde{\Gamma}_{yz}$. The corresponding two solutions (cross–track and along–track directions) of the homogeneous Hill equations are:

$$\Delta\hat{x}(t) = \frac{2\Delta\dot{z}_0}{n}\cos nt + (6\Delta z_0 + \frac{4\Delta\dot{x}_0}{n})\sin nt - (6n\Delta z_0 + 3\Delta\dot{x}_0)t + \Delta x_0 - \frac{2\Delta\dot{z}_0}{n} \tag{4.23}$$

$$\Delta\hat{y}(t) = \Delta y_0 \cos nt + \frac{\Delta\dot{y}_0}{n}\sin nt \tag{4.24}$$

which, together with the radial solution, may be used to estimate a last improvement of the initial conditions and to compute a last nominal orbit. This best estimate of the orbit could be used to carry out a final potential coefficient estimation.

The whole estimation procedure is listed in table 4.11.

4.3.3 Space–like v. time–like

Whereas the iteration method described above may seem a nice example of gradiometric data analysis (since it separates the determination of the orbit from that

of the gravity parameters $\bar{C}_{lm\alpha}$) there remains one big problem, namely that of the enormous amount of data to be handled by the software. Two aspects have to be considered in this context.

First the number of observations. During a six months mission and with a data sampling rate of 4 seconds, almost 4 million measurement epochs become available. If we confine ourselves to the case of Aristoteles this means that at each epoch two observations are given, V_{yy} and V_{zz}, from which the linear combination $2\Delta\Gamma_{yy}+\Delta\Gamma_{zz}$ is derived. In terms of the model equation 4.2 this therefore means that the vector of observations $\boldsymbol{\ell}$ has over 4 million elements. In the least squares estimation procedure, eq. 4.5, we have to carry out the multiplication $\boldsymbol{b} = A^T\,\boldsymbol{\ell}$ (if we assume the covariance matrix Q to be a scaled unit matrix) which involves inner products of the columns of A with the observation vector $\boldsymbol{\ell}$. One such inner product thus consists of a summation over 4 million elements. Each column of A belongs to a certain l,m–combination and each row of A belongs to one individual observation. Each element of A is therefore of the type 4.3 or 4.4 for a unique set of (ω_o, ω_e, r). For each of the 4 million inner product elements the trigonometric functions $\cos\psi_{km}$ or $\sin\psi_{km}$ have to be evaluated for all necessary values of the index k. Even when using recursive type methods for the computation of these trigonometric functions, this part of the estimation process, i.e. computing $\boldsymbol{b}$, remains very time consuming.

Another aspect concerns the number of unknowns. Given some specified maximum degree L up to which one likes to estimate the potential coefficients, a total number of $(L+1)(L+2)$ unknown $\bar{C}_{lm\alpha}$ coefficients appear. E.g. if $L = 240$ (like it is usually the case in this thesis) we have over 58,000 unknowns. This means that the design matrix A has more than 58,000 columns and, as we have seen, almost 4 million rows. The normal matrix N has a size of $(58{,}000)^2$, but it is symmetric. Solving such a linear system is still an enormous effort, which under certain assumptions, however, can be drastically reduced by invoking a block–diagonal structure, as was done in section 4.2 with the error analysis.

If we are willing to accept some assumptions and approximations as far as it concerns the observation distribution, something can be done to reduce the computational efforts. Two methods are mentioned here: the *space–like* and the *time–like* method. The terms space–like and time–like are those used in section 4.2.11. The difference between the two is the way in which we look upon the data distribution. Both methods will be illustrated below.

Space–like

In the space–like approach we view upon the data set as being a set of observations distributed in some way (regular or irregular) in three dimensional space. The mutual relationship between the observations (they are all taken along one common satellite orbit) is not taken into consideration. The latter point is only reflected by the fact that during the iteration process (table 4.11) the initial state vector components are updated using some dynamical model for the satellite motion (i.e. Hill

equations), and by the fact that all measurements are considered to be located inside a spherical shell around the earth. This shell, concentric with the earth's center of mass, is just thick enough to contain all measurements. An observation is considered a function of its position coordinates r, θ, λ. An easy way of expressing the gradients in this case is by means of a spherical harmonic expansion, equation 3.16. When using this formulation, a block–diagonal structure of the normal matrix is achieved if the data, assumed to be situated on a sphere (i.e. constant radius approximation), is distributed regularly in θ–direction (observations along parallels) as well as in λ–direction (observations at equal intervals $\Delta\lambda$). Along each parallel the observations constitute a set of equally spaced samples. On such a set, the trigonometric functions $\sin m\lambda$ and $\cos m\lambda$ are orthogonal. This causes the elements of the normal matrix belonging to the coefficients $\bar{C}_{l_1 m_1 \alpha_1}$ and $\bar{C}_{l_2 m_2 \alpha_2}$ to be only different from zero if $\alpha_1 = \alpha_2$ and $m_1 = m_2$ (Rummel and Colombo, 1985). Compare this with the orthogonality in section 4.2.1.

Thus a regular, global grid of observations is needed in this case. But we do not have such a grid. If the actual data distribution shows some particular pattern at all, it will certainly not be of the type described above. However, in (Rummel and Colombo, 1985) it is suggested that such a grid structure can be obtained in the following way. The spherical shell in which all observations are contained, is divided into equal angular cells of thickness Δr (the thickness of the shell) and with size $\Delta\theta = \Delta\lambda$, the latter chosen at most half of the smallest gravitational wavelength to be recovered. All the gradient observations inside each cell are averaged. All equi–angular cells with their average gradient values formed in this way, together constitute a global grid of gradient observations.

A block–diagonal structure of the normal matrix is now obtained which means an enormous reduction of computation time. But also the number of "observations" (i.e. cell averages) has become less. If we, for example, like to recover the potential coefficients up to a maximum degree and order $L = 240$, the grid size $\Delta\theta = \Delta\lambda$ has to be at most 0°.75. The number of cells contained in a grid of this size is 115,200, which is much less than 4 million. So also the amount of time needed for the computation of $\boldsymbol{b}$ has been decreased.

The drastic decrease of computation time is a big advantage of gridding the data. On the other hand, averaging means a loss of precision due to the smoothing of the signal. Furthermore, we may put some questions about how to carry out the averaging process. A first problem is that in reality, we do not know the exact coordinates of the measurement points P. The decision in which cell an observation lies and where it is located inside the cell, can only be made on the basis of the coordinates of the approximate points P' along the nominal orbit. Due to orbit uncertainties in cross–track or along–track direction, an observation may therefore contribute to the cell average of the wrong cell (the radial orbit error will not lead to such problems). Since we have seen that the second–order derivatives have a large power content in higher degrees, these kinds of mistakes may have considerable influence. Furthermore, since the error variance of cell averages is computed from

the measurement error propagation through the averaging process, the resulting set of grid values will not have equal variances any more. Unfortunately, as long as we do not have the true coordinates, nothing can be done about the former problem, unless we could determine the true orbit with enough precision independently, e.g. by using GPS, which would imply that we are back at the earlier situation, where we don't have to estimate the orbit from the gradient measurements.

A related question is how to interpret the average values. One possibility is to first average the original gradient observations, and to attribute the average values to the center points of the equi–angular cells. The coordinates of those center points are known. The problem is reduced to that of just estimating potential coefficients, since the "orbit problem" is suppressed, compare (Vermeer, 1991). No linearization with respect to the coordinate unknowns is possible any more. Another possibility is to attribute the average gradient value to some imaginary average location inside the cell, not coinciding with the cell center. This average measurement point could be thought of as being computed by means of averaging the (unknown) coordinates of all the actual measurement points inside the cell. The cell center could then be chosen as approximate point P'. In this case there remains the problem of which kind of relation there is between on the one hand the averaged measurement points and the true orbit and on the other hand the cell centers and the nominal orbit.

Probably the best choice is to first linearize the problem and compute the gradient anomalies (eq. 4.18) and the linear combinations (eq. 4.20) for all observations individually. All observations inside the same cell are linearized with respect to the same approximate point P', namely the cell center, and for the normal gradient U_{ij} the value at the cell center P' is taken. The resulting set of linear combinations of gradient anomalies are now averaged on the basis of their approximate locations P'. In the estimation procedure from table 4.11 this means that the cell averages are formed between step 6 and 7, and that they only apply to step 7. Back substitution (step 10) is done with the original system of equations (one equation per observation, not the averages). Furthermore, one could use the cell averages only for computing and inverting the normal matrix $N = \frac{1}{\sigma^2} A^T A$, whereas the multiplication $A^T \boldsymbol{\ell}$ is carried out for all observations individually. In that case it is not even necessary to actually compute the cell averages. One simply chooses a global grid at a certain mean radius and with a certain step size in θ–direction. The λ–dependence is cancelled due to the trigonometric orthogonality relations. In this way however, we have to compute $\boldsymbol{b}$ again for almost 4 million observations. Note furthermore that, if we want to use the inverse of the present normal matrix (where cell averages are assumed) to describe the a–posteriori precision of the estimated potential coefficients, we have to adapt the scale factor σ^2 of the variance matrix of the observations Q to represent the variance of cell averages. In fact, in general the variance of a cell average becomes a function of the number of observations inside the cell and their location relative to the cell center. This results also in a θ–dependence of the variances.

Nevertheless, if one likes to compute cell averages, there remains the question

of the validity of the averaging process. Inside each cell the observations might be distributed rather irregularly. It should therefore be preferred to reduce the observations to the cell center using some higher order model, at least one which includes a dependency on the relative location of the measurement points with respect to the cell center. This becomes extremely important if the actual observations V_{ij} are averaged, since these numbers are relatively large and vary considerably throughout a cell. E.g. at 200 km altitude differences in the value of the gradients inside a cell with size $\Delta r = 10$ km and $\Delta\theta = 0°.5$ may be of the order of $10^{-10}\,s^{-2}$ due to the $\bar{C}_{20}$ contribution only.

It should be mentioned that a possible negative influence on the results of the averaging process may be overcome during the *iteration* procedure (see table 4.11, step 14), at least if the cell averages are formed between step 6 and 7 (as suggested above).

Time–like

As pointed out in section 4.2.11, the time–like approach considers the data set a time series, the observations given in subsequent points along a satellite orbit. The usual way of expressing the gradients in this case is by means of an expansion into a series of inclination functions, like eq. 3.18. The time dependency is expressed by the argument $\psi_{km} = \psi^0_{km} + \dot{\psi}_{km}t$, compare eq. 4.6.

In order to achieve a block–diagonal structure for the normal matrix in this case, we do not need an equi–angular global grid of observations. In section 4.2.1 we saw that such a structure is already obtained if the data is distributed regularly only *along the orbit*, under the condition that the number of nodal days N_d and the number of orbit revolutions N_r contained in the mission are relative primes and that $N_r > 2L$. For $L = 240$ the latter condition is fulfilled in a 6 months mission, like with Aristoteles. Also, with the foreseen constant sampling rate (4 seconds for Aristoteles) with which the measurements are taken, the requirement of a regular data distribution along the orbit is very well met. What remains as an assumption is the absence of data gaps. Unfortunately, data gaps *will* occur in a real mission (due to orbit maintenance manoeuvres and excessive drag variations at the poles). Note that in the space–like situation along–track data gaps play no role as far as it concerns the orthogonality requirements. Another problem that remains is the fact that not one reference orbit is computed for the full mission period. Instead, reference trajectories are integrated for some limited amount of time, e.g. every three or six days.

In the time–like approach there are no problems concerning the true orbit, nominal orbit and average values, because no averaging of the data is needed to invoke orthogonality. Except for the data gaps assumption, the time–like approach therefore seems more suitable for the gradiometric analysis (in the way described in section 4.3.2 above) than the space–like approach.

4.3.4 Simulated data

Whereas an error analysis can be carried out without the availability of a set of real measured quantities (like it was done in section 4.2), potential coefficient estimates can only be obtained if one has at ones disposal some set of observations. Up to now, however, no gradiometric satellite mission has been carried out, so no true data is available yet. In order to be able to investigate the gradiometric analysis process and to carry out test computations one depends on simulated data. The Center for Space Research (CSR) of the University of Texas at Austin computed a set of gravity gradient data along a simulated satellite orbit. Detailed information about this simulation can be found in (Schutz et al., 1987) and (Schutz et al., 1988). In this section we will only briefly discuss the main characteristics of this simulation.

In order to obtain a good global coverage, the CSR simulation spans 32 sidereal days after which the ground track of the satellite's orbit will repeat to within 10 km (i.e. $T_r = 32$ days). For a satellite, moving along a polar orbit ($I = 90°$) at a mean radius of approximately 6527 km, this results in 525 orbital revolutions. The force model used for the orbit computation consisted of the complete 360×360 OSU86F potential coefficient field (Rapp and Cruz, 1986) and the following values for the gravitational parameter GM and the reference radius R:

$$GM = 3.986004404 \cdot 10^{14}\,\mathrm{m^3\,s^{-2}}$$

$$R = 6378137\,\mathrm{m}\ .$$

Furthermore, the $\bar{C}_{21}$ and $\bar{S}_{21}$ coefficients were assigned the values $\bar{C}_{21} = -0.10 \cdot 10^{-9}$ and $\bar{S}_{21} = 0.102 \cdot 10^{-8}$. The force model did not include non-gravitational forces, luni-solar effects or temporal variations in the gravity field (Schutz et al., 1988). The computed data set contains the ephemeris of a low-orbiting GRM (Geopotential Research Mission) satellite at 4 second intervals. This resulted in a total amount of 691,210 measurement epochs.

At each epoch the elements of the gravity gradient tensor V_{ij} are evaluated. Despite the fact that only five elements of the gradient tensor are independent (V_{ij} is symmetric and traceless), all six elements comprising the upper (or lower) triangular part of V_{ij} are provided. The gradients were computed as the second-order partial derivatives of the gravitational potential in a geocentric cartesian coordinate system using a spherical harmonic expansion up to degree and order 360. The potential coefficients for this expansion were also taken from the OSU86F field, but the contribution from the zeroth order term ($\frac{GM}{r}$) was not included (i.e. $\bar{C}_{00} = 0$). All measurement errors were excluded from the simulation. For the purpose of the computations the observations were assumed to be uncorrelated, having a 0.01 $\mathrm{E}/\sqrt{\mathrm{Hz}}$ white noise error spectrum.

At this point already some remarks can be made about this simulation, as far as they concern our intended use. Referring to section 4.2.1 it is reminded here that for a good recovery of potential coefficients from satellite data up to a maximum degree L (avoiding a situation where different orbital frequencies, represented by some k, m

combination, are projected onto the same gravitational frequency, represented by an l, m combination), the number of orbital revolutions N_r in one repeat period has to be larger than $2L$. This in turn implies that for the present simulation (525 orbital revolutions) we can expect only a good recovery up to approximately $L = 262$.

The choice for repeating ground tracks after 32 days ensures a finest possible global coverage within this period. As a result the ground track nodes (cross over points between descending and ascending tracks) for the 525 orbital revolutions included in this repeat period have, at the equator, a spacing in λ–direction of approximately 0°.68. Simple reasoning tells us not to expect recovery of potential coefficients above degree $l = 262$ in this case, despite the fact that, due to the 4 second sampling interval in along–track direction, the spacing in θ–direction will be approximately 0°.27. Figure 4.23 shows for a small area ($50° \leq \lambda \leq 60°$ and $0° \leq \varphi \leq 10°$) the ground tracks (continuous lines) of the simulated orbit, as well as the grid lines of a 0°.5 × 0°.5 grid (broken lines). It can be seen from this figure that several blocks of the grid are not covered by ground tracks. An attempt to derive from the simulation potential coefficients up to degree 360 with the space–like method (for which such a 0°.5 × 0°.5 grid is needed) is therefore not likely to succeed, at least if the empty blocks are not accounted for in some manner.

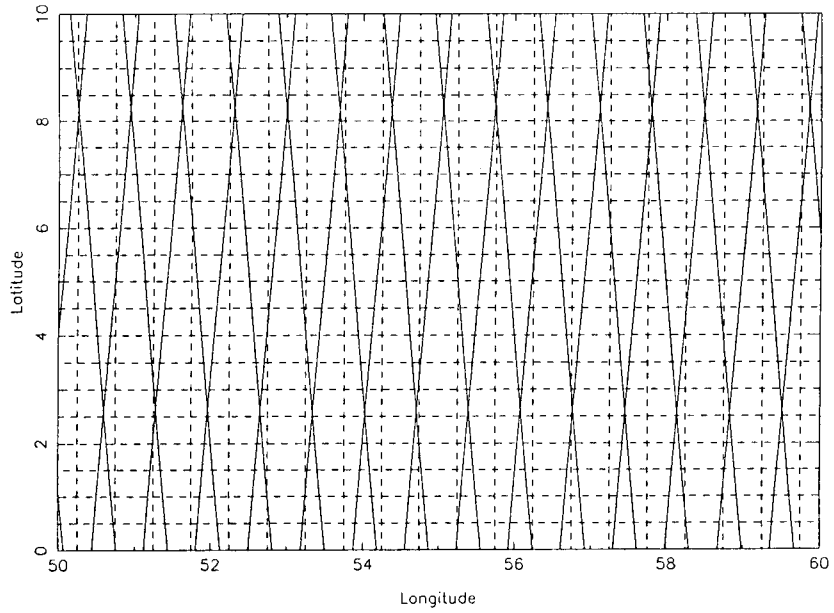


Figure 4.23 *Ground tracks (continuous lines) of the simulated CSR orbit for a small area. The broken lines are the grid lines of a 0°.5× 0°.5 grid.*

4.3.5 Space–like results

A computer program was written by D. Stelpstra (Stelpstra, 1990) to estimate potential coefficients from gradiometer measurements following the linearization procedure in the space–like approach, which was described in previous sections. As input data the satellite ephemeris and gravity gradients from the CSR simulation, described in the last section, was used. Test computations were done to demonstrate the possibilities of the present recovery method (Koop and Stelpstra, 1991).

To this extent, a global $1° \times 1°$ grid, being a commonly used and easily manageable grid type, was established from the gradient data set. Such a grid consists of 64,800 equi–angular blocks. For each measurement epoch, the given geocentric cartesian coordinates were transformed into curvilinear coordinates θ, λ. The latter coordinates determined in which block the measurement was assumed to be situated. Since there were 691,210 measurement epochs in the simulation, the result was an average number of $N_o = 10$ measurements per block. To simulate the elimination of the coordinate unknowns (i.e. the orbit errors), as is done in the linear model approach, at each measurement epoch the linear combination of observations $2\Delta\tilde{\Gamma}_{xx} + \Delta\tilde{\Gamma}_{zz}$ was computed. The resulting numbers were then averaged inside each grid block. For convenience we assumed all those block averages to have the same measurement precision (namely $\sigma\sqrt{5/N_o}$ with σ the precision of one gradient), which of course is not really true, since not each block contains the same number of measurements. At this stage we are only interested in estimates for the potential coefficients. Their a–posteriori error behaviour was already discussed in section 4.2. This means that the one important item concerning the error variances is the assumption of equal variance for block averages. This leads to a simple variance matrix, namely a scaled unit matrix. The value which we choose for the variance factor σ^2 is not important here. As can be seen from eq. 4.5 it will, in case that Q is a scaled unit matrix, drop out when computing $\hat{c}$.

From the $1° \times 1°$ grid one may estimate potential coefficients up to degree and order 180. We expected a good recovery of potential coefficients from the simulated 525 orbit revolutions only up to degree 262, so with $L = 180$ we are on the safe side. The estimated potential coefficients were used to compute signal degree variances. They are shown in figure 4.24, together with the degree variances from the original OSU86F model. Up to degree 140 the signal power seems to be recovered very well. Above degree 140 the estimated power is higher than the original one. This fact may be due to an *aliasing* effect, where the power of higher degrees, contained in the measurement signal but not solved for, is reflected in the lower ones (Jenkins and Watts, 1968). In time series analysis, this effect always appears when the chosen sampling interval is not small enough to detect the highest frequency which is present in the signal. The highest frequency which can be detected given some sampling interval Δ is the so–called *Nyquist frequency* $f_N = 1/(2\Delta)$. If f_N is smaller than the highest frequency present in the signal, aliasing occurs. When using spherical harmonics the aliasing effect cannot be expressed in a simple manner,

but it will be present. In our case, the OSU86F field contains potential coefficients up to degree and order 360, which were all used to simulate the gradient data. But we only solve for coefficients up to degree 180, so the power content for the degrees $181 \leq l \leq 360$ will be reflected in the spectrum between 0 and 180. Since in reality the gravitational field contains degrees up to infinity, an aliasing effect will always be present, even for higher degree solutions, although block–averaging of the data will decrease its influence to a large extent.

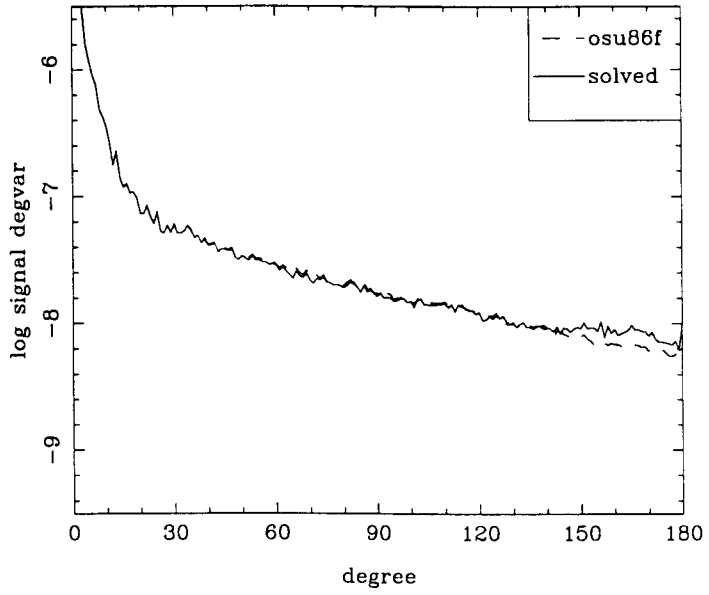


Figure 4.24 *Degree variances from solved potential coefficients from simulated gradient data using the space–like approach, as well as those from OSU86F.*

Nevertheless, the solution seems not bad at all. Especially if we remember that this estimation is in fact only one step of an iteration process in which one simultaneously solves for the gravitational field as well as the orbit. Furthermore, since we did not remove from the simulated data any other potential contribution apart from the central (zeroth order) term, the anomalies are relatively large.

The degree variances, however, show the spectral behaviour of the signal in terms of the power per degree. If this signal power per degree is well solved for, it does not necessarily mean that the distribution of the power over all orders inside this degree is properly solved for. One has to look into the individual potential coefficients themselves. To this extent the absolute values of the relative differences between the solved coefficients $\hat{\bar{C}}_{lm\alpha}$ and the original OSU86F coefficients $\bar{C}_{lm\alpha}$ were computed, i.e. $|(\hat{\bar{C}}_{lm\alpha} - \bar{C}_{lm\alpha})/\bar{C}_{lm\alpha}|$. If the resulting value is, for example, $10^0 = 1$ this means the relative error is 100 %. A value of 10^{-2} means 1 % relative error. We counted how much of these numbers were present in each $10^{0.1}$–class between 10^{-5}

and 10^5. A cumulative graph of these numbers, in percentages of the total amount of solved coefficients, is shown in figure 4.25. From this figure we see that over 80 % of the coefficients is solved with an error less than 100 %, however only about 20 % with an error less than 10 %. In order to see whether some particular pattern exists in the error or not, a grey-scale map of the same numbers in an l, m-scheme was created. This map showed no particular pattern, except for a slight increase with increasing degree.

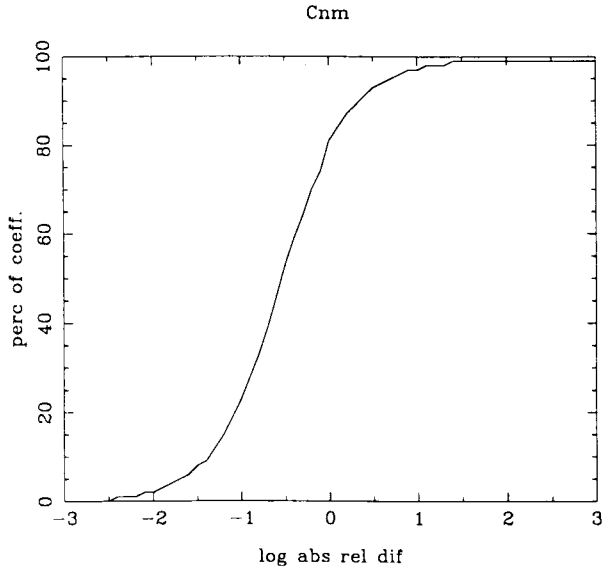


Figure 4.25 *Cumulative graph of the number of errors between solved and OSU86F coefficients in percentages of the total amount of solved coefficients. Horizontal axis: order of magnitude of the relative differences, logarithmical.*

It should be mentioned that the differences between the solved and original coefficients, as they are computed here, do not contain measurement errors, since we adopted a scaled unit matrix for the covariance matrix of the observations. The remaining error therefore consists of model errors and numerical errors. The latter are inevitable in each computational process.

4.3.6 Time–like results

In the time–like approach the observations are treated (as they are in reality) as measurements done in successive points along the satellite orbit. The parameters to indicate those measurement points are the orbital elements $a, e, I, \omega_o, \omega_e$. For the CSR simulation we have $I = 90°$ and for our convenience we assume $e = 0$. Furthermore, we assume all measurements to be situated on a sphere with radius equal to the mean radius of the simulated orbit. This leaves us with ω_o and ω_e as

variables to indicate the successive measurements. They are to be computed from the given geocentric cartesian coordinates x^I and the time t of the epoch. Care has to be taken with this transformation from x^I, t to ω_o, ω_e, especially with the identification of the correct quadrants.

No averaging of the measurements has to be done, so we are left with a data set of 691,210 records, each containing the measurement (or the linear combination $2\Delta\Gamma_{xx} + \Delta\Gamma_{zz}$) as well as the two coordinates ω_o and ω_e at the measurement time t, in double precision a storage requirement of a little more than 16 Mbyte. A computer program was written to estimate from this data set potential coefficients following the method described in sections 4.3.1 and 4.3.2. This estimation program is in fact an extended version of the error analysis program from section 4.2, now including the computation of $\boldsymbol{b} = A^T \boldsymbol{\ell}$ and the solution of the normal equations. Due to the enormous CPU time requirements for this program, we decided for the first run to average every two successive measurements, thereby halving the number of observations to be used in the estimation process. Since we only attempt to estimate coefficients up to a maximum degree $L = 180$ (like in the space–like approach) the data interval in θ–direction after this averaging (approximately 0°.54) will still be sufficient.

Whereas in a real mission like Aristoteles the lower part of the measurement spectrum will be too much distorted or not available with enough precision (band limitation, see section 4.2.4) we included in the present estimation the complete measurement spectrum (all β_{km} for $0 \leq m \leq L$ and $-L \leq k \leq L$). No stabilization technique was used (see section 4.2.6).

Degree variances, computed from the potential coefficients solved with the present time–like program, are shown in figure 4.26, together with the original OSU86F degree variances. Above degree 120 the solved spectrum diverges from the original one. However, the differences are larger than in the space–like approach (compare figure 4.24). Since, in both cases, we did not remove the contribution from the degrees above 180 from the measurements, the differences are due to an aliasing effect. In the space–like approach, where we first computed cell averages, this averaging acts as a smoothing process, decreasing the power content of the high degrees. Therefore the aliasing effect is expected to be less than here.

Also for the present solution we show (figure 4.27) the cumulative percentages of the relative errors in each coefficient. Somewhat less than 80 % of the coefficients is solved now with an error less than 100 %, but still about 20 % with an error less than 10 %.

It should be re-emphasized that, for both the space–like as well as the time–like approach, the results shown here are "only" a first step of what should be an iterative process in which at the same time the orbit is improved and the coefficients are estimated. A simplified simulation, based on zonal coefficients only (Rummel and Colombo, 1985), indicated that only a few steps are necessary for the iteration process to converge. A final judgement of the results and the estimation method can thus only be given if a complete iterative solution is computed. Nev-

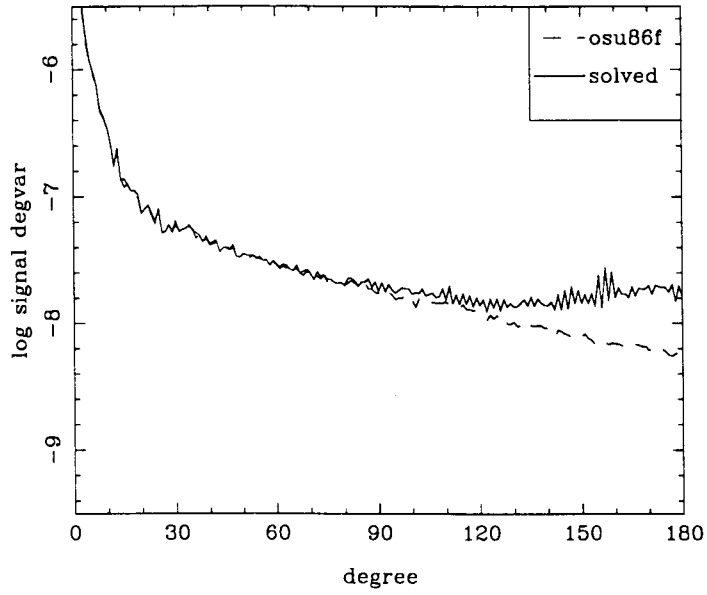


Figure 4.26 *Degree variances from solved potential coefficients from simulated gradient data using the time-like approach, as well as those from OSU86F.*

ertheless, the present results are promising, especially considering the fact that in both cases no stabilization technique was used.

4.3.7 Some computational aspects

From a computational point of view, the programs for the space–like and the time–like methods are very much alike, at least in broad outlines. Both programs do, in principle, nothing more than solving a system of the type $\boldsymbol{\ell} = A\,\boldsymbol{c}$ by means of a least squares adjustment for the same vector of unknowns $\boldsymbol{c}$ (which are the potential coefficients). The differences between the two approaches are found in the vector of observations $\boldsymbol{\ell}$ and the expression for the elements of the design matrix A. In case of the space–like method, the vector of observations $\boldsymbol{\ell}$ consists of block averages, a total of 64,800 elements in the case of a $1° \times 1°$ grid. In the time–like approach this vector consists of the original observations along the satellite orbit which are 691,210 elements in case of the one month CSR simulation.

Concerning the design matrix A, the difference between the two methods is that the space–like method uses spherical harmonics as base functions (for which we need to compute the Legendre functions $\bar{P}_{lm}$ and their derivatives), whereas the time–like method is set up in the rotated orbital system and is therefore based upon the inclinations functions $\bar{F}^k_{lm}$ and $\bar{F}^{k\,*}_{lm}$.

Both differences have far–reaching consequences as far as it concerns computation time. Before we illustrate this, we have to remind ourselves that the required

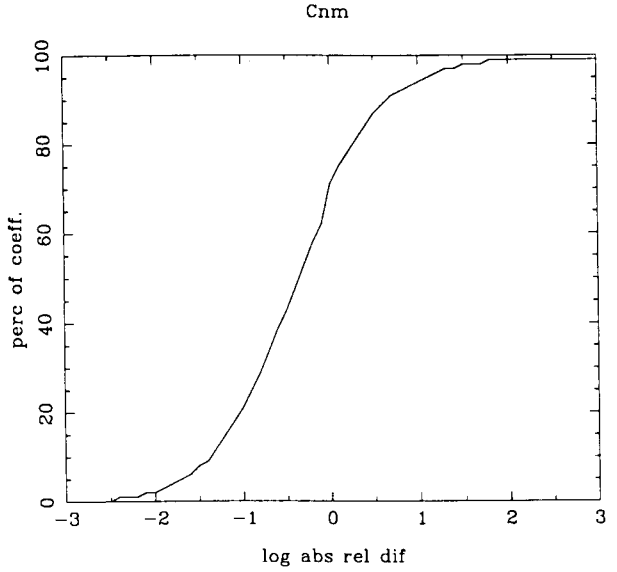


Figure 4.27 *Cumulative graph of the number of errors between solved and OSU86F coefficients in percentages of the total amount of solved coefficients. Horizontal axis: order of magnitude of the relative differences, logarithmical.*

computational effort strongly depends on the maximum degree L and the available hardware. All CPU results given in this section refer to software running on the CONVEX C240.

First of all, computing the inclination functions for all valid l, m, k up to some maximum degree L is much more time consuming than computation of the Legendre functions for all l, m up to the same degree. If the space–like program is used only for error analysis (i.e. setting up and inverting the normal matrix) it requires only 40 CPU seconds (for one observation type), against the time–like program approximately 800 seconds (for 10 observations types), both up to $L = 180$ and both on the same computer.

Secondly, and what is more important, the product $\boldsymbol{b} = A^T \boldsymbol{\ell}$ puts a heavy burden on the computations. Compared to the computations involved in $\boldsymbol{b}$ the error analysis contributes only little to the total CPU time for the adjustment process. For the space–like program the coefficient estimation up to $L = 180$ requires 370 CPU seconds (Stelpstra, 1990). For each of the 64,800 cell–averages we have to evaluate the sine and cosine of $m\lambda$, multiply them with the Legendre functions, some other factors and with the observation. But since the observations are regularly spaced on a grid, the amount of extra effort can be kept limited.

For the time–like program the discrepancy between "only" error analysis and adjustment is much larger. This is due to the fact that we have much more ob-

servations than in the space–like method. For each observation an element of A consists of a summation over k. For all l, m combinations we have to evaluate the sine and cosine of the argument $\psi_{km} = k\omega_o + m\omega_e$, multiply them with the inclination functions and some other factors, add them for all k and multiply them with the observation. As an example, consider the estimation of the zonal coefficients up to $L = 100$, which requires 717 CPU seconds, of which 95 % (681 sec.) is needed for setting up $\boldsymbol{b}$. Bearing in mind the total run–time for the error analysis for *all* orders and degrees up to $L = 240$ (500 CPU seconds, see section 4.2.9), one can imagine the computational effort to be delivered for a complete *estimation* up to such maximum degree. Indeed, the estimation of all potential coefficients up to degree and order 180 from the halved data set of 345,605 observations (as described in the previous section) required approximately 69 hours CPU time on the CONVEX. The time needed for setting up and inverting the normal matrix in this case becomes almost negligible.

A last remark concerns the system of normal equations. The solution of the system of normal equations, $N\,\boldsymbol{c} = \boldsymbol{b}$ is obtained in an equivalent manner as were the diagonal elements of the inverse normal matrix in the error analysis. With the upper triangular matrix R, resulting from a Cholesky factorization of N, the triangular system

$$R^T\,\boldsymbol{w} = \boldsymbol{b}$$

is solved for $\boldsymbol{w}$ by forward substitution. Afterwards, backward substitution of the system

$$R\,\boldsymbol{c} = \boldsymbol{w}$$

gives the required vector $\boldsymbol{c}$.

4.3.8 Conclusions

On basis of the presented results we cannot yet decide in favour of either the space–like or the time–like approach to estimate potential coefficients from gradiometer observations. However, the results shown in figures 4.24 to 4.27, make us believe that both methods bear the potential of converging to the right solution, if the iteration procedure is continued. At this moment we have to keep in mind that we only carried out the first step of this iteration process. Furthermore, in an actual mission one shall be faced with the presence of data gaps and non–polar orbits.

The space–like method is very attractive because it reduces computation time drastically, not only because of the block–diagonal structure of the normal matrix, but much more because of limitation of the length of the vector $\boldsymbol{b} = A^T\,\boldsymbol{\ell}$. On the other hand, the required averaging process in this case (to obtain cell averages) acts as a kind of smoothing operator, which may not be attractive if higher degrees are to be obtained from the measurements. Another point is that the space–like method does not offer the possibility of including the band limitation of the gradiometer into the procedure in a proper manner, whereas this band limitation is very likely to be present in a real mission.

The time–like method does not have those disadvantages due to its set–up in terms of measurement frequencies in c.p.r. (by using inclination functions). Furthermore, no averaging of the observations has to be done in this method to create a block–diagonal structure of the normal matrix. On the other hand, the influence of along–track data gaps may have more influence than in the space–like approach. A big disadvantage of this method is furthermore the required CPU time, which, in case of a real 6 months mission, will be large.

Both methods suffer from aliasing effects. Since in our simulation we know where and with which potential coefficients the observations were calculated, we could remove the contribution for degrees which we do not solve for from the observations. In reality, however, this can never be exactly done since we do not know these contributions, especially not the high degrees. A better test might be to set up and invert the m–blocks of the normal matrix (the step in the computations where the coefficients become correlated and aliasing finds its origin) for as much degrees as possible, but compute the vector $\boldsymbol{b}$ and solve the system only for the required degrees. For example, in our case we could have constructed and inverted the normal matrix for all degrees up to 360 (the maximum degree of the simulation) and solve only for degrees up to 180. In that case, at least theoretically, the aliasing effect must vanish.

Table 4.11 *Gradiometric potential coefficient estimation procedure, following (Rummel and Colombo, 1985).*

1	Pre–adjustment of the diagonal tensor components (if available) using Laplace' equation, including error propagation.
2	Choice of the normal potential field U.
3	Given U, tracking data and some first estimate of the initial state vector, compute a first nominal orbit.
4	Compute at approximate points P' the normal gradients U_{ij}, with which the gradient anomalies $\Delta\tilde{\Gamma}_{ij}$ are derived.
5	Evaluation of the coefficient matrix of the coordinate corrections U_{ijk} at the approximate points P' in spherical approximation.
6	Forming a linear combination, e.g. eq. 4.20 which means elimination of the coordinate corrections. The a–priori variance of the linear combination should be computed by error propagation.
7	Least squares adjustment to determine estimates of the potential coefficient corrections $\Delta\hat{\bar{C}}_{lm\alpha}$, based on the linear combination 4.20 and $\Delta\tilde{\Gamma}_{xy}$ (if available).
8	Up–date of the nominal orbit with the old initial state vector and the up–dated potential coefficients $\bar{C}''_{lm\alpha}$.
9	Compute at new approximate points P'' the normal gradients U_{ij}, with which up–dated gradient anomalies $\Delta\tilde{\Gamma}_{ij}$ are derived.
10	Back substitution to obtain coordinate corrections $\Delta\hat{z}$, eq. 4.21.
11	Improvement of the initial state vector elements $\Delta z_0, \Delta\dot{z}_0, \Delta\dot{x}_0$, eq. 4.22.
12	Up–date the nominal orbit again with improved initial state vector and improved potential coefficients $\bar{C}''_{lm\alpha}$.
13	Again up–date the gradient anomalies with new nominal orbit.
14	Repeat the process from step 5 on until it converges.
If the gradients V_{xz} and V_{yz} are available:	
15	Back substitution into eq. 4.19 to obtain coordinate corrections $\Delta\hat{x}$ and $\Delta\hat{y}$.
16	Final improvement of the initial state vector with equations 4.22, 4.23 and 4.24.
17	Compute final best estimate of the nominal orbit with this improved initial state vector and the final potential coefficient set.
18	Final potential coefficient estimation based on this best nominal orbit.

5

Relativistic view on gradiometry

As we have seen, a gradiometer measures the second–order partial derivatives of the gravitational potential. The nine elements $\{V_{ij}|i,j = 1,2,3\}$ (of which only 5 are independent) together built up the gradient tensor, also called tidal tensor or Eötvös tensor. The latter names point at a geometrical interpretation of the gradients in terms of curvature. Indeed, the elements of V_{ij} are related to the curvature of the equipotential surfaces and the lines of force in the following manner:

$$\{V_{ij}\} = -g \begin{bmatrix} \kappa_1 & t & f_1 \\ t & \kappa_2 & f_2 \\ f_1 & f_2 & H^* \end{bmatrix}$$

cf. (Hotine, 1969), (Marussi, 1982) or (Rummel, 1986), where κ_1 and κ_2 are the normal curvatures of the equipotential surface in x^1, respectively x^2 direction, t is the geodetic torsion, f_1 and f_2 are the components of the curvature vector of the line of force in x^1, respectively x^2 direction and H^* is the mean curvature of the equipotential surface. In this sense, gradiometry gives us insight in the geometrical structure of the space around the earth, being the result of the earth's gravitational field. However, these curvatures refer to either a two dimensional space (an equipotential surface) or a one dimensional space (the line of force). The embedding three dimensional space is just flat. This geometrical picture is appropriate in the Newtonian theory of gravitation. In Einstein's theory of gravitation, better known as the general theory of relativity, the viewpoint is different. The embedding space is no longer a three dimensional flat space (which of course is simply the space surrounding us, where we live in) but a four dimensional space, called spacetime, built from the usual three dimensional spatial part and the time as fourth dimension. This spacetime can hardly be visualised any more, but it has proven to be a useful mathematical concept in order to accurately describe physical features.

5. *Relativistic view on gradiometry*

There are several reasons why a relativistic description of gradiometry may be of interest. We will see later, that Einstein's general theory of relativity is in fact a geometrical theory of gravitation. Geometry and gravitation become two indistinguishable concepts. Geometry plays an essential part in the theory of relativity and, since gradiometry is so closely related to Newtonian geometry, it may be interesting to find out how gradiometry looks like in relativistic terms.

One of Einstein's conditions in the development of his gravitational theory was that, in the limit, it should reduce to Newton's theory. Indeed it can be shown that, after specifying this limit, Einstein's equations yield the Newtonian equations as an approximation. Conversely, we may expect that Newtonian concepts have a generalization in relativistic terms which, at least conceptually, must be very similar. So we may expect that in some way a relativistic description of gradiometry will include the geometrical structure, especially the curvature, of spacetime. This indeed will appear to be the case. The general relativistic counterpart of the gradiometric model equation is the so–called equation of geodesic deviation, in which relative accelerations are related to the curvature of spacetime. This curvature is fully described by the elements of the Riemann–Christoffel tensor and it appears that the non–vanishing elements of this Riemann tensor in the Newtonian limit, appear to be elements of the tidal tensor, the one shown above.

In this way relativity contributes to the gradiometric measurements themselves since the V_{ij} are part of the Riemann tensor. A second aspect which should be included in a relativistic description of satellite gradiometry, is the orbit of the satellite. One may be familiar with the ideas from special relativity (valid in the absence of gravitation) that features like time dilation or length contraction may play a role, at least if we consider particles moving with high velocities, such as earth orbiting satellites. However, satellites are in free fall in the gravitational field of the earth. In the general theory of relativity (with gravitation included) such motion is governed by the geodesic equation. Gravitation is not looked upon as some external force, but it is part of the geometry of spacetime. In fact, gravitation "curves" spacetime. Hence a satellite falling around the earth is regarded as a free particle, on which no forces act (disregarding disturbing forces like air drag etc.). Its path through spacetime is the "straightest" possible line, a geodesic, but spacetime, and thus the geodesic, are curved due to gravitation.

So, the equations of motion of a particle moving under the influence of gravitation, constitute a second object of study in which relativistic influences may play a role. A third point worth mentioning concerns the local reference system with respect to which the gradiometer measurements are taken. In the Newtonian case, as described in the foregoing chapters, we use a local cartesian (or orbital) coordinate system with its origin in the center of mass of the gradiometer and one axis always pointing radially outwards. Another possibility would be to keep the gradiometer "inertially" fixed (with respect to far stars), so that the coordinate system would not rotate. In both cases, the formulation implies the possibility of using some preferential coordinate system (in the latter case an inertial system). Newtonian theory

is based upon the existence of such inertial system(s). But in the general theory of relativity, no (global) inertial systems exist, only local ones, valid over some small spacetime region. Furthermore, each coordinate system should be equally well suitable for the formulation of physical features, there is no system to be preferred. It is therefore common to work with four dimensional local inertial systems, defined in the framework of relativity as moving with the particle along its worldline (its trajectory in spacetime). Their motion is determined by parallel transport of the system axes along the trajectory. Due to the curvature of the space, the local system may therefore move somewhat different than it would do in Newtonian physics.

In this chapter, we will pay some attention to the relativistic description of gradiometry. The first two aspects, (the relativistic contribution to the gradiometric measurements in terms of the Riemann tensor and the satellite's equations of motion) will be illustrated, the third aspect (parallel transport of the local inertial system) will not be treated separately. The reader is not supposed to be acquainted with the general theory of relativity. To this end, the chapter contains introductory sections on the theory of relativity, as far as needed for our purposes. The intention is to sketch the outlines of general relativity and the way in which it may play a role in gradiometry. However, this thesis is certainly not a textbook on relativity, and we do not try to give a complete treatment of all aspects of the theory. Therefore, only a few of them are treated, some of which only briefly. Many textbooks are written on the theory of relativity. As for the general theory of relativity, we mention (Bergmann, 1942), (Misner et al., 1973), (Ohanian, 1976) or (Foster and Nightingale, 1979). Extensive reference can be made everywhere to the literature mentioned above, as well as to others. Furthermore, recently several authors have published on the relativistic description of gradiometry. As for the three aspects mentioned above, we refer to (Theiß, 1984), (Soffel et al., 1987), (Soffel, 1989), (Ries et al., 1990), (Kopejkin, 1991) and (Gill et al., 1992). Furthermore, as gradiometry is expected to be a very sensitive technique (especially the superconducting gradiometer of Paik and Richard (1986)), gradiometric measurements could be used to test certain effects predicted by the general theory of relativity. See e.g. (Paik, 1989) or (Mashhoon et al., 1989).

5.1 Some aspects of the general theory of relativity

It is very customary to distinguish between the *special* and the *general* theory of relativity, abbreviated respectively by STR and GTR. The essential difference between the two is that the latter is capable of incorporating gravitation into the theory, whereas the former is only valid in the absence of gravitation. Despite this difference, and despite the fact that the special theory was developed first, the indication "general" and "special" suggest that the latter can be derived from the former by considering a special case (namely no gravitation), which indeed is true. We will therefore, in this chapter, focus on the GTR. Furthermore, since gravitation is the

major subject of this thesis, we will not specialize to the STR. Moreover, only those aspects of the GTR, which are either of special interest to us or are indispensable to a good understanding of the theory, are treated.

5.1.1 The spacetime of relativity

In Newtonian theory, any event (or happening) can be described by stating where (space) and when (time) it took place. "Where" is usually indicated by three coordinates, $\{x^i | i = 1,2,3\}$ (for example x, y, z). The chosen coordinate system thereby spans a three dimensional space. Analogously, we may look upon the time t as a coordinate spanning a one dimensional space, indicating the "when". In the theory of relativity, the three dimensional space and the one dimensional time are merged into a new, four dimensional space, called *spacetime* (Minkowski, 1952). Plotting the position of an event in an, in general, four dimensional diagram, gives its trajectory in spacetime, called the *worldline*. Usually a two dimensional subspace is plotted, one axis of which is always the time coordinate.

In order to put a geometry onto the space, we introduce a quadratic form ds^2, called the line element (see appendix B.1.3). This quadratic form is set up in terms of dx, dy, dz and dt, describing in a way the infinitesimal distance between two events. In the Newtonian case we have in fact two such "distances" (Ohanian, 1976):

$$dl^2 = dx^2 + dy^2 + dz^2$$

and $\quad dt$

but in relativistic spacetime there is only one:

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 \tag{5.1}$$

where the scalar c is the velocity of light in vacuum (multiplication of dt with c matches the dimensions of the separate terms; with c included, the unit of time is chosen such, that the velocity of light in vacuum is equal to unity). The "distance" between two events, expressed by the line element ds is an invariant, i.e. it is independent of the coordinate system used to express it. The specific form of equation 5.1 can be understood by recalling one of the fundamental postulates of relativity, which says that the speed of light c is the same in all inertial coordinate frames. Consider the motion of a light photon. If we have two coordinate systems x^i and x^I (with $\{x^i | i = 1,2,3\} \equiv (x, y, z)$ and $\{x^I | I = 1,2,3\} \equiv (X, Y, Z)$), its velocity is $c = dr/dt$ in the first coordinate system, and $c = dR/dT$ in the other, where $dr^2 = dx^2 + dy^2 + dz^2$ and $dR^2 = dX^2 + dY^2 + dZ^2$. The constantness of c can then be expressed as:

$$c^2dt^2 - dx^2 - dy^2 - dz^2 = c^2dT^2 - dX^2 - dY^2 - dZ^2 = 0$$

which is consistent with the assumption of the invariance of the line element ds in eq. 5.1, cf. (Foster and Nightingale, 1979). Eq. 5.1 shows that, for objects, other

than photons, the line element (also called sometimes *spacetime interval*) can be positive or negative. If ds^2 is positive, the line element is called *timelike*, if it is negative *spacelike*, and if it is zero (like for a photon) it is called *lightlike*.

It is convenient, and very customary in relativity, to introduce the following notation:

$$\begin{aligned}\{x^\mu | \mu = 0,1,2,3\} &= (x^{\mu=0}, x^{\mu=1}, x^{\mu=2}, x^{\mu=3}) \\ &= (x^0, x^1, x^2, x^3) \\ &\equiv (ct, x, y, z) \\ &= (ct, x^{i=1}, x^{i=2}x, ^{i=3}) \\ &= \{ct, x^i | i = 1,2,3\}\end{aligned} \tag{5.2}$$

where we adopt the convention to use Greek indices to denote the four relativistic spacetime coordinates and Latin indices to denote the three "usual" space coordinates. It is, furthermore, convenient to write

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{5.3}$$

where

$$\{\eta_{\mu\nu}\} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

is the so–called *Minkowski tensor*. Comparing equation 5.3 with equation B.12 from appendix B.3.1 we see that the Minkowski tensor corresponds to the metrical tensor, describing the metrical properties of the space. Furthermore, since it is diagonal and has constant elements, we learn from this appendix that it is the metrical tensor for flat space expressed in "cartesian" coordinates. This space is the spacetime from STR. The Minkowski metric replaces the three dimensional cartesian metric if we switch to four dimensions.

In the Newtonian formalism, in three dimensions, it is possible to introduce a cartesian coordinate system covering the entire space. Such a coordinate system is called (globally) *inertial*. All other coordinate systems, connected to such an inertial system by means of linear coordinate transformations (not involving time), so–called Galilean transformations, are also inertial systems. In relativistic spacetime, the coordinate system (ct, x, y, z), for which the line element 5.1 holds, is the generalization of a Newtonian inertial system. Different inertial systems are connected via the well-known *Lorentz transformations*, which are still linear coordinate transformations, but now the time t is transformed too.

Accelerated coordinate systems are not inertial. In the presence of a gravitational field, all coordinate systems are accelerated, so there is no privileged system (inertial

system) in which the metrical tensor takes on the simple Minkowski form. Therefore, gravitation cannot be included in the STR. Einstein solved this problem when setting up his GTR. We will see later that, in the GTR, we are no longer restricted to linear Lorentz coordinate transformations, but more general non–linear coordinate transformations are involved. The metrical tensor is no longer diagonal and its elements are in general functions of the coordinates, instead of constants. We will, furthermore, see that in this case the space is no longer flat, but curved, so that it is no longer possible to introduce into the space a cartesian (inertial) coordinate system which is "global in the sense that its coordinate neighbourhood is the whole of spacetime" (Foster and Nightingale, 1979). In a curved space such a coordinate system is only valid in an infinitesimal region of the space ("locally").

5.1.2 Equations of motion

Consider a particle freely moving in space. Since no forces act on the particle, it will move along a geodesic. A *geodesic* is the generalization of the concept of a straight line in three dimensional Euclidean space ($\mathbf{E}^3$) to more general curved spaces of arbitrary dimension. This generalization is done by requiring that there exists a parametrization u of the geodesic, such, that the tangent vectors dx^μ/du along the curve constitute a parallel vector field. The concept of parallelism is generalized by means of the *absolute derivative* $Dv^\mu (= v^\mu{}_{;\nu} dx^\nu$, see appendix B.2.3). The parameter u is called an *affine* parameter.

In the space under consideration the geodesic gives the shortest path between two points. Newton already stated (in his first law) that each particle at rest or in uniform motion along a straight line preserves its state if no forces are exerted on it. Considering a free particle to move along a geodesic is therefore nothing more than a generalization of this Newtonian law to arbitrary spaces. The equations of motion of the particle may thus be represented by the equation of a geodesic:

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\sigma} \dot{x}^\nu \dot{x}^\sigma = 0 \tag{5.4}$$

(see appendix B.2.3). In equation 5.4 the dot denotes differentiation with respect to the affine parameter u, i.e. $\dot{x}^\nu = dx^\nu/du$, and the $\Gamma^\mu_{\nu\sigma}$ are the Christoffel symbols (of the second kind), which may be computed from the metrical tensor $g_{\mu\nu}$ as

$$\Gamma^\mu_{\nu\sigma} = \frac{1}{2} g^{\mu\rho} \left(g_{\nu\rho,\sigma} + g_{\sigma\rho,\nu} - g_{\nu\sigma,\rho} \right) \tag{5.5}$$

(compare eq. B.15). Note that, if we would like to insert for the parameter u time t, we should bear in mind that, since time is in four dimensional spacetime one of the coordinates, it is no longer universal for the whole of space, so it is not an affine parameter. Therefore, we introduce the so–called *proper time* τ, which is measured by a clock, co–moving with the particle, and not by some kind of "inertial ticking" clock. This has also consequences for the velocity of the particle. In three dimensional non–relativistic space, the components of the velocity vector

are $v^i = dx^i/dt$. In four dimensional relativistic spacetime, we have to take the derivative with respect to the proper time τ. We obtain the so–called 4–velocity $u^\mu = dx^\mu/d\tau$, also called world velocity. We may of course also differentiate the coordinates with respect to the coordinate time t. Then we obtain the coordinate velocity $v^\mu = dx^\mu/dt = u^\mu/\gamma$. The quantity γ is given by $\gamma = 1/\sqrt{1 - v^2/c^2} = dt/d\tau$ (with $v^2 = v^i v_i$), which appears in the four dimensional Lorentz transformation of STR. With eq. 5.2 we may write the coordinate velocity as $v^\mu = (c, v^i)$.

Thus, if the elements of the metrical tensor are given we may compute the equations of motion of the particle by means of equation 5.4. Often, however, the metric has a complicated form, and one would have a hard time computing the equations of motion in this way.

One may therefore proceed in another manner (Foster and Nightingale, 1979). The equations of motion can also be obtained by using the *Euler–Lagrange* equations:

$$\frac{d}{du}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) - \frac{\partial L}{\partial x^\mu} = 0 \tag{5.6}$$

which can be found in many textbooks on classical mechanics, tensor analysis or relativity, e.g. (Goldstein, 1980), (Sokolnikoff, 1951) or (Ohanian, 1976). In this equation L is the Lagrangian, which is a function of the independent variables $\dot{x}^\mu$ and x^μ. Let us consider the following Lagrangian:

$$L^2 = g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu \,. \tag{5.7}$$

As we will show now, with this L the Euler–Lagrange equations reduce to the geodesic equation 5.4. (Note the resemblance of this L with the infinitesimal length (or *line element*) ds (equation B.12) from appendix B.3.1, which will appear to be more than just a coincidence.) From equation 5.7 we obtain:

$$\begin{aligned}
\frac{\partial L}{\partial \dot{x}^\mu} &= \frac{1}{2L}\frac{\partial}{\partial \dot{x}^\mu}\left(g_{\nu\sigma}\dot{x}^\nu \dot{x}^\sigma\right) \\
&= \frac{1}{2L}\left(g_{\nu\sigma}\delta^\nu_\mu \dot{x}^\sigma + g_{\nu\sigma}\dot{x}^\nu \delta^\sigma_\mu\right) \\
&= \frac{1}{L} g_{\mu\nu}\dot{x}^\nu \\
\frac{\partial L}{\partial x^\mu} &= \frac{1}{2L}\frac{\partial}{\partial x^\mu}\left(g_{\nu\sigma}\dot{x}^\nu \dot{x}^\sigma\right) \\
&= \frac{1}{2L} g_{\nu\sigma,\mu}\dot{x}^\nu \dot{x}^\sigma \,.
\end{aligned}$$

Inserting these in equation 5.6 yields:

$$\begin{aligned}
\frac{d}{du}\left(g_{\mu\nu}\dot{x}^\nu\right) - \frac{1}{2} g_{\nu\sigma,\mu}\dot{x}^\nu \dot{x}^\sigma = 0 \Leftrightarrow \\
g_{\mu\nu}\ddot{x}^\nu + g_{\mu\nu,\sigma}\dot{x}^\nu \dot{x}^\sigma - \frac{1}{2} g_{\nu\sigma,\mu}\dot{x}^\nu \dot{x}^\sigma = 0 \Leftrightarrow \\
g_{\mu\nu}\ddot{x}^\nu + \left(g_{\mu\nu,\sigma} - \frac{1}{2} g_{\nu\sigma,\mu}\right)\dot{x}^\nu \dot{x}^\sigma = 0
\end{aligned}$$

and on multiplying with the inverse (or associate) metrical tensor $g^{\sigma\rho}$ and using eq. 5.5 together with the identity

$$g_{\mu\nu,\sigma}\dot{x}^\nu\dot{x}^\sigma = \frac{1}{2}g_{\mu\nu,\sigma}\dot{x}^\nu\dot{x}^\sigma + \frac{1}{2}g_{\mu\sigma,\nu}\dot{x}^\nu\dot{x}^\sigma$$

we obtain the geodesic equation 5.4. This means that the Euler–Lagrange equations 5.6 are equivalent to the geodesic equations 5.4, at least if one uses a Lagrangian L of the form eq. 5.7. Evaluating equations 5.6 for deriving the equations of motion is often more convenient than straightforward application of the geodesic equation.

The variational principle and the equations of motion

Geodesics are the "shortest" paths between two points in the space under consideration (e.g. in Euclidean space a straight line). Consider a particle moving along such a geodesic and suppose we have parametrized the curve with the arc–length s, instead of some arbitrary parameter u. It can be shown (Foster and Nightingale, 1979) that the arc–length s is an affine parameter. The parameter s can be used as a measure of the length of the curve. It appears in the expression of the invariant line element ds:

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \ .$$

With the help of this line element ds, the distance S between two points on the curve P_1, with coordinates $x^\mu(s_1)$, and P_2, with coordinates $x^\mu(s_2)$, can be expressed as:

$$\begin{aligned} S &= \int_{s_1}^{s_2} ds \\ &= \int_{s_1}^{s_2} \sqrt{g_{\mu\nu}dx^\mu dx^\nu} \\ &= \int_{s_1}^{s_2} \sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}ds \end{aligned} \tag{5.8}$$

where $\dot{x}^\mu = dx^\mu/ds$. Since we are dealing with a geodesic we look for the "shortest" distance between the points P_1 and P_2. The integral in equation 5.8 therefore has to have a stationary (in our case minimal) value. This value can be found with the help of the so–called *calculus of variations*.

If the geodesic S has a minimal value, it will grow if we replace $x^\mu(s)$ with an arbitrary function

$$x^\mu(s) + \delta x^\mu(s) \tag{5.9}$$

in which the $\delta x^\mu(s)$ are small in the interval $s_1 < s < s_2$ (we call $\delta x^\mu(s)$ the ***variation*** of $x^\mu(s)$). So eq. 5.9 would represent another curve through the end points P_1 and P_2 in the neighbourhood of the geodesic. Since this other curve also passes through P_1 and P_2 it is

$$\delta x^\mu(s_1) = \delta x^\mu(s_2) = 0 \ . \tag{5.10}$$

If we now put $\sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu} = f(x^\mu, \dot{x}^\mu)$ we may write the change in distance between P_1 and P_2, being the result of replacing x^μ with $x^\mu + \delta x^\mu$, as:

$$\delta S = \int_{s_1}^{s_2} f(x^\mu + \delta x^\mu, \dot{x}^\mu + \delta\dot{x}^\mu)ds - \int_{s_1}^{s_2} f(x^\mu, \dot{x}^\mu)ds \, .$$

In linear approximation we have

$$f(x^\mu + \delta x^\mu, \dot{x}^\mu + \delta\dot{x}^\mu) = f(x^\mu, \dot{x}^\mu) + \frac{\partial f}{\partial x^\mu}\delta x^\mu + \frac{\partial f}{\partial \dot{x}^\mu}\delta\dot{x}^\mu$$

thus

$$\begin{aligned}\delta S &= \delta \int_{s_1}^{s_2} f(x^\mu, \dot{x}^\mu)ds \\ &= \int_{s_1}^{s_2} \left(\frac{\partial f}{\partial x^\mu}\delta x^\mu + \frac{\partial f}{\partial \dot{x}^\mu}\delta\dot{x}^\mu\right) ds \\ &= \int_{s_1}^{s_2} \left(\frac{\partial f}{\partial x^\mu}\delta x^\mu + \frac{\partial f}{\partial \dot{x}^\mu}\frac{d}{ds}\delta x^\mu\right) ds\end{aligned}$$

and upon integrating by parts the second term:

$$\delta S = \left[\frac{\partial f}{\partial \dot{x}^\mu}\delta x^\mu\right]_{s_1}^{s_2} + \int_{s_1}^{s_2} \left(\frac{\partial f}{\partial x^\mu} - \frac{d}{ds}\left(\frac{\partial f}{\partial \dot{x}^\mu}\right)\right)\delta x^\mu ds$$

in which, with eq. 5.10, the first term disappears. For the geodesic δS should be equal to zero for arbitrary δx^μ, so

$$\frac{d}{ds}\left(\frac{\partial f}{\partial \dot{x}^\mu}\right) - \frac{\partial f}{\partial x^\mu} = 0 \tag{5.11}$$

is the equation determining the geodesic. This equation has exactly the same form as the Euler–Lagrange equations 5.6 if we put $f = L$, so that the Lagrangian from eq. 5.7 is indeed identified as the line element.

Already at this point we may suspect an important concept which is used in the GTR. In classical mechanics, the Lagrangian L is related to the energy of the particle. If a Lagrangian of the form as in equation 5.7 properly describes the motion of the particle, then the expression for the line element ds must contain in some way the particle's energy. As the dx^μ are nothing but infinitesimal coordinate differences, the energy must be present in the elements of the metrical tensor! This notion is very important and we will come back to it in the next section.

5.1.3 The principle of equivalence

According to the *principle of equivalence* it is not possible to distinguish between gravitational and inertial accelerations. Consider for example an observer (astronaut) in a freely falling spacecraft. Suppose the spacecraft has no windows and the astronaut awakens after a long sleep. He does not know whether he is still in orbit

around the earth or has drifted away from the earth and from all other attracting masses. In order to find out he releases an object (proof–mass) and observes its motion. In both cases, the proof–mass will stay floating in front of him. If no forces act on either the astronaut or the proof–mass, this can be understood very easily. Newton already stated it in his first law (law of inertia). The spacecraft, astronaut and the proof–mass are in rest or in straight–line, uniform motion and will persist in doing so, as long as no forces are exerted on them. If a^i is the acceleration of the body with respect to some inertial coordinate system x^i, the equations of motion of the body can be described by:

$$a^i = 0$$

i.e. the motion is unaccelerated. If we would describe the motion of the spacecraft with respect to some other coordinate system x^r, which rotates with a constant angular velocity with respect to x^i, the acceleration a^r is no longer zero. Multiplying these accelerations by the masses of the bodies, ma^r, we obtain so–called apparent or inertial forces, which *seem* to act on the bodies (centrifugal and/or Coriolis forces). These are in fact no real forces, since with a linear coordinate transformation $x^i = x^i(x^r)$ we can always "transform them away". Let us call the mass m mentioned above the *inertial mass*. If there exists a coordinate system with respect to which all bodies not subjected to (real) forces are in rest or in uniform motion, it is called an inertial system. The x^i would be such inertial system, the x^r not, because of its rotation. However, all coordinate systems which are at rest or in uniform motion with respect to each other, are inertial systems.

If now the astronaut would find himself in free fall in a gravitational field, a force does act on him and the proof–mass. This gravitational force is proportional to the mass of the object it acts on:

$$f^i = Mg^i$$

with M the *gravitational mass* and g^i the gravitational acceleration. So if the motion of the astronaut is due to a gravitational force, it undergoes an acceleration

$$a^i = \frac{M}{m} g^i \, .$$

Currently we know that for all objects the ratio M/m is equal to 1 (accurately to 10^{-11}), i.e. the gravitational force results in the same acceleration for all objects, independent of their mass. This is exactly what the principle of equivalence says: there is no difference between inertial and gravitational mass. So also in this case the proof–mass will remain floating in front of the astronaut. Obviously, there is a coordinate system, co–moving with the spacecraft, relative to which the gravitational force can be "transformed away", as if it was an inertial force.

However, there is a solution to the astronaut's problem. Due to the convergence of the lines of force of the gravitational field of a spherical body, the proof–mass will very slowly move relative to the astronaut (often referred to as moving under the influence of tidal forces). Note that this is exactly the reason why satellite

gradiometry is possible. The relative motion as a result of the radial force field does not really violate the principle of equivalence, provided we limit the validity of the principle to infinitesimal small regions of space, where the tidal forces can be considered negligible. The co–moving coordinate system, extended over this small region (often referred to as local coordinate system) looks very much like a real inertial system, of which the metric is described by the Minkowski tensor.

In the STR (valid in the absence of gravitation) the laws of physics apply with respect to inertial coordinate systems. Non–inertial systems can always be transformed to inertial systems by suitable Lorentz transformations. According to the principle of equivalence, all frames of reference, including those in which a gravitational field is present, are equally well suited for formulating the laws of nature, since we have seen that gravitational fields can (at least locally) be transformed away by a suitable coordinate transformation. In this sense, the difference between the STR and the GTR is that in the latter all continuous, differentiable coordinate transformations are considered, whereas the STR "only" involves linear Lorentz transformations. In the literature one finds many different formulations of the principle of equivalence, in terms of mass, accelerations, coordinate systems, coordinate transformations, STR or GTR, validity of the laws of nature or physics, etc. etc. See e.g. (Bergmann, 1942), (Foster and Nightingale, 1979), (Ohanian, 1976).

We have seen that a free particle moves along a geodesic of which the equations are given by 5.4. This equation is a tensor equation, that is, it has the same form in all possible coordinate systems. If a gravitational field is present, the particle will no longer be a free particle since it moves under the influence of gravitation. However, according to the principle of equivalence, the gravitational field may be "transformed away" (at least locally) by a suitable coordinate transformation. Indeed, an observer moving along with the particle and carrying a coordinate frame with him, will consider the particle to be a free particle with respect to his coordinate frame, just like the astronaut in the example above. For this observer, equation 5.4 still holds for the motion of the particle. Then this equation should also hold in all other coordinate systems, including those not co–moving with the particle (in which the particle *does* accelerate). In that case it must be so that the gravitational effects are already present in equation 5.4. They must be included in the Christoffel symbols $\Gamma^{\mu}_{\nu\sigma}$. The Christoffel symbols in turn, are related to the metrical tensor $g_{\mu\nu}$ (see equation 5.6), so that we conclude that the gravitational field must be present in the elements of $g_{\mu\nu}$, or as Bergmann (1942) writes: "the $g_{\mu\nu}$ are the potentials of the gravitational field".

Since the gravitational potential is in general a function of the coordinates, the elements of the metrical tensor will also be functions of the coordinates. They describe the geometrical structure of the space. In flat space, in global cartesian coordinates, $g_{\mu\nu}$ has a diagonal form with constant elements. In such a coordinate system the $\Gamma^{\mu}_{\nu\sigma}$ are zero and equation 5.4 reduces to

$$\ddot{x}^{\mu} = 0 \,.$$

If a gravitational field would be present it would appear as if we could transform away the effects of gravitation globally by a coordinate transformation to global cartesian coordinates. We know, however, that the principle of equivalence only holds locally (in an infinitesimal region of space). Apparently we are not able to introduce a *global* cartesian coordinate system. In such a situation we infer from geometry (see appendix B.1.3) that the space is curved. The conclusion must be that gravitation curves the space. Note that mathematically the correct formulation for a curved space is a non–vanishing Riemann–Christoffel tensor.

The above conclusion implies that, in the GTR (i.e. in the presence of gravitation) gravitation is no longer considered a force, but a part of the geometry of the space. A free particle is now meant to move under the influence of gravitation alone, and it will still move along a geodesic, the geodesic, however, now being some curved trajectory, but still the "shortest" distance between two points, but in a curved space. This notion is the most important concept in the GTR. Gravitation and geometry are no longer two different concepts. The theory of gravitation becomes a geometrical theory in which the usual geometrical formalism can be applied. This is also one of the reasons that index notation and tensor analysis are much used in the GTR, since they are both especially suitable to geometrical problems. Note that in the above reasoning the crucial point was the validity of the geodesic equation in all coordinate systems. Since this is a natural result of the tensor formalism, one can understand Einstein's choice to use tensor analysis as mathematical tool in the development of the GTR.

5.1.4 The Einstein field equations

In the foregoing, we have seen that, in the GTR, based on the principle of equivalence, the gravitational potential is present in the elements of the metrical tensor. In Newtonian theory, the gravitational force field, being the gradient of a gravitational potential, finds its origin in the presence of matter. Poisson's equation

$$g^{\mu\nu}V_{\mu\nu} = -4\pi\rho G$$

(or in index free notation $\nabla^2 V = -4\pi\rho G$) where ρ is the mass density, V the gravitational potential and G the gravitational constant, describes the relation between this gravitational field and the matter distribution. This equation is referred to as *field equation*. When developing his GTR, Einstein was searching for an equivalent relation, with on one side the gravitational potential (in terms of (functions of) the metrical tensor $g_{\mu\nu}$) and on the other side some quantity expressing the matter (or in general: energy) distribution. Naturally, the equation had to be a tensor equation (since it must be invariant with respect to coordinate transformations), so the quantity expressing the energy distribution had also to be a symmetric, second–order tensor. It is called the energy tensor (also called: energy–momentum–stress tensor, stress tensor, stress–energy tensor, etc.), and it is denoted by $T^{\mu\nu}$. It may contain the matter density, momentum, as well as other kinds of energy, like electromagnetic

radiation. The energy tensor must satisfy

$$T^{\mu\nu}{}_{;\nu} = 0 \tag{5.12}$$

(i.e. vanishing divergence) which merely expresses conservation of energy. A last condition for the relativistic field equations is, that they have to reduce to Poisson's equation in the Newtonian limit. After a long search, Einstein came up with equations fulfilling all conditions, the famous *Einstein field equations*:

$$R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu} = \kappa T^{\mu\nu} \tag{5.13}$$

where $R^{\mu\nu}$ is the contravariant Ricci tensor, R the curvature scalar, equal to a contraction, $R^{\mu}{}_{\mu}$, of the Ricci tensor (the trace), and κ some constant, equal to $-8\pi G/c^4$. Equivalently one may write this equation in covariant form, with lower indices, or, after some manipulations, in the following form:

$$R_{\mu\nu} = \kappa(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}) \,, \tag{5.14}$$

with $T = T^{\mu}{}_{\mu}$. The left–hand side of equation 5.13 is called the Einstein tensor $G^{\mu\nu}$. Since the Ricci tensor is a contraction of the Riemann–Christoffel curvature tensor, i.e. $R_{\mu\nu} = R^{\sigma}{}_{\mu\nu\sigma}$, and since the latter contains the second–order derivatives of the metrical tensor, the field equation is indeed very analogous to the Newtonian situation, where we have also on the left hand side some function of the second–order derivatives of the gravitational potential, and on the right hand side the source term.

In general, the Einstein field equations are non–linear. Furthermore, since the Ricci tensor has 10 independent components, equation 5.14 is in fact a system of 10 non–linear equations. Solving these equations for $g_{\mu\nu}$ if some information or assumptions on the energy tensor are given, is therefore a very tedious job. Whereas the field equation is a tensor equation and therefore valid in all coordinate systems, one coordinate system may appear more convenient for finding a solution of the field equations than another. In order to simplify the search for a solution, one may, under certain assumptions, linearize the field equations and put some constraints on the matter distribution. This will be done in subsequent sections.

5.2 Weak field approximation

Suppose we have a coordinate system in which the metric tensor may be written as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{5.15}$$

where $\eta_{\mu\nu}$ is the Minkowski metric and the $h_{\mu\nu}$ are small quantities, such that products of them may be neglected. Since the elements of the metrical tensor include

the gravitational potential, the latter must now be present in the $h_{\mu\nu}$. Scaling with the velocity of light c to obtain dimensionless quantities, we have the condition

$$|h_{\mu\nu}| \sim \frac{GM}{rc^2} \ll 1 \,. \tag{5.16}$$

If this condition is fulfilled (which, in many applications, e.g. computations in our solar system, is usually the case) one often calls the linearization 5.15 the *weak field approximation*. But it is in fact not the above condition that makes the field weak. A weak field is obtained only if we assume the derivatives of $h_{\mu\nu}$ to be small too! The value of the gravitational potential itself does not tell us anything about the weakness or strongness of the field. We have to look at the amount in which the potential varies throughout space, since it is the potential difference which produce the gravitational force, i.e. the potential derivatives.

In general the metrical tensor can be used to raise and lower indices (see appendix B.3.2). In the linearized case of eq. 5.15, however, we shall, since the $h_{\mu\nu}$ are small, raise and lower indices using $\eta_{\mu\nu}$.

5.2.1 Newtonian limit

In the development of the GTR one of the conditions was, that, in the limit, the theory reduces to the Newtonian case. This limit is partly characterized by the weak field approximation, as described above. But, to arrive at Newton, some additional assumptions have to be taken. One of them is the assumption of a quasi–static gravitational field. For a quasi–static field, the time derivatives of the $h_{\mu\nu}$ are small compared to the space derivatives, i.e.

$$\frac{\partial}{\partial x^0} h_{\mu\nu} = \frac{1}{c}\frac{\partial}{\partial t} h_{\mu\nu} \ll \frac{\partial}{\partial x^i} h_{\mu\nu} \tag{5.17}$$

and are therefore neglected. Furthermore, we only consider particles which move with low velocity v, i.e. $v \ll c$. For such a particle $\gamma = 1/\sqrt{1 - v^2/c^2} \approx 1$ so that $u^\mu \approx v^\mu$. Using these assumptions in the geodesic equation 5.4 and neglecting terms quadratic in v, we obtain:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{00}\frac{dx^0}{d\tau}\frac{dx^0}{d\tau} = 0 \,.$$

For the Christoffel symbols we find with eq. 5.5

$$\begin{aligned}
\Gamma^\mu_{00} &= -\frac{1}{2} g^{\mu\sigma} g_{00,\sigma} \\
&= -\frac{1}{2} \eta^{\mu\sigma} h_{00,\sigma} \\
&= -\frac{1}{2} \eta^{\mu j} h_{00,j}
\end{aligned}$$

where we used the stationarity condition 5.17. The equation of motion for the $\mu = 0$ component will in this case read

$$\mu = 0 \quad : \quad c\frac{d^2t}{d\tau^2} \quad = \quad 0 \, ,$$

which exactly expresses our assumption $\gamma = dt/d\tau = 1$, and furthermore implies that $\dot{x}^0 = c$. Therefore, switching from τ to t, the spatial part of the equations of motion become

$$\mu = i \quad : \quad \frac{d^2x^i}{dt^2} \quad = \quad \tfrac{1}{2}c^2\eta^{ij}h_{00,j} \, .$$

Using well–known symbolic notation, the right hand side of this equation can be written as $-\frac{c^2}{2}\nabla h_{00}$, where we used $\eta^{ij} = -\delta^{ij}$ (Kronecker delta) and $\delta^{ij}\frac{\partial}{\partial x^j} \rightarrow \nabla$.

From Newtonian theory we know

$$\frac{d^2x^i}{dt^2} = \delta^{ij}V_j$$

with V the Newtonian gravitational potential, so that, in the present approximation, the relativistic theory will reduce to Newton if we identify:

$$h_{00} = -\frac{2V}{c^2} \, , \tag{5.18}$$

with which the g_{00} component becomes:

$$g_{00} = 1 - \frac{2V}{c^2} \, . \tag{5.19}$$

Note that this expression for g_{00} (Newtonian limit) is derived by comparing the relativistic equations of motion in the weak field, slow motion approximation, with the Newtonian equations. Nothing can be said at this point about the other elements of the metrical tensor (g_{0i} and g_{ij}), cf. (Misner et al., 1973). In fact, it does not matter how those other components look like, because, to the present level of approximation, they will not contribute to the equations of motion, because they are multiplied with small quantities v or v^2.

In the non–relativistic limit, the dominant component of the energy tensor is $T_{00} = \rho c^2$ (Ohanian, 1976) where ρ is the mass density. For this 00–component, the Einstein field equations 5.14 are:

$$R_{00} = \kappa\left(T_{00} - \frac{1}{2}Tg_{00}\right) \, .$$

Since $T_{00} = T = \rho c^2$ and, to the present level of approximation,

$$R_{00} \approx -\frac{\partial}{\partial x^i}\Gamma^i_{00} = \frac{1}{2}\eta^{ij}h_{00,ij} \rightarrow -\frac{1}{2}\nabla^2 h_{00}$$

this field equation reduces to

$$-\frac{1}{2}\nabla^2 h_{00} = \frac{1}{2}\kappa\rho c^2 \tag{5.20}$$

or with $h_{00} = -2V/c^2$ and $\kappa = -8\pi G/c^4$:

$$\nabla^2 V = -4\pi\rho G$$

which is the familiar Poisson equation. We see that, indeed, with the present assumptions and level of approximation, the relativistic theory reduces to Newtonian theory.

5.2.2 Linear approximation in relativistic terms

When working in terms of relativity, we no longer assume that we have a quasi–static field. Furthermore, we no longer deal with small velocities. We proceed, in this case, by inserting the metric 5.15 into the field equations 5.13. It can be shown, e.g. (Foster and Nightingale, 1979) or (Ohanian, 1976), that the field equations reduce to

$$\bar{h}_{\mu\nu,\sigma}{}^{\sigma} + (\eta_{\mu\nu}\bar{h}^{\sigma\rho}{}_{,\sigma\rho} - \bar{h}^{\sigma}{}_{\nu,\mu\sigma} - \bar{h}_{\mu\sigma,\nu}{}^{\sigma}) = 2\kappa T_{\mu\nu} \tag{5.21}$$

where

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu} \tag{5.22}$$

$$h = h^{\mu}{}_{\mu}\,.$$

If we would carry out a small coordinate transformation, $x^{\mu'} \to x^{\mu} + \xi^{\mu}$, a straightforward calculation shows (Foster and Nightingale, 1979) that

$$\bar{h}^{\mu'\nu'} \to \bar{h}^{\mu\nu} - \eta^{\nu\sigma}\xi^{\mu}{}_{,\sigma} - \eta^{\mu\sigma}\xi^{\nu}{}_{,\sigma} + \eta^{\mu\nu}\xi^{\sigma}{}_{,\sigma} \tag{5.23}$$

so that

$$\bar{h}^{\mu'\sigma'}{}_{,\sigma'} \to \bar{h}^{\mu\sigma}{}_{,\sigma} - \xi^{\mu}{}_{,\sigma}{}^{\sigma}\,.$$

If we choose the small quantities ξ^{μ} to be a solution of

$$\xi^{\mu}{}_{,\sigma}{}^{\sigma} = \bar{h}^{\mu\sigma}{}_{,\sigma}$$

it results (in the new coordinates) in

$$\bar{h}^{\mu'\sigma'}{}_{,\sigma'} = 0\,. \tag{5.24}$$

This is called the *gauge condition*. The quantities $\bar{h}_{\mu\nu}$ are the potentials of the field. The field equations will not determine these potentials uniquely. We are free to replace $\bar{h}_{\mu\nu}$ with the expression from eq. 5.23, because it will leave the field equations intact. Since we are free to choose the coordinates we like (because all equations are tensor equations), it is always possible to transform to a coordinate

system such that the gauge condition is fulfilled. This particular choice for ξ^μ, will simplify eq. 5.21 a great deal, because all three terms between brackets are zero. We obtain (we drop the primes) :

$$\bar{h}_{\mu\nu,\sigma}{}^{\sigma} = 2\kappa T_{\mu\nu} \; . \tag{5.25}$$

One often introduces the *d'Alembert operator* $\Box$, defined by

$$\Box = -\eta^{\sigma\rho}\frac{\partial}{\partial x^\sigma}\frac{\partial}{\partial x^\rho} = \nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} \; ,$$

which is the extension of the Laplace operator to four dimensional spacetime. With this operator, the simplified field equations 5.25 (in contravariant form) may be written as

$$\Box \bar{h}^{\mu\nu} = -2\kappa T^{\mu\nu} \tag{5.26}$$

which holds provided the quantities $\bar{h}^{\mu\nu}$ satisfy the gauge condition (cf. eq. 5.24):

$$\bar{h}^{\mu\nu}{}_{,\nu} = 0 \; .$$

Equation 5.26 is the extension to four dimensional relativistic spacetime of the Poisson equation in non–relativistic three dimensional space. The generalization to relativity of the Laplace equation would be

$$\Box \bar{h}^{\mu\nu} = 0$$

which is valid in empty space.

The d'Alembertian is known from the theory of electromagnetism. It is sometimes called the "wave operator", and an equation like 5.26 is referred to as a wave equation. Analogously to electromagnetism, one therefore assumes the existence of gravitational waves, generated by a source term $-2\kappa T^{\mu\nu}$, and propagating through space like radiation.

We will solve the field equation 5.26 for some special (simplified) situations. To do this, consider an isolated gravitating mass. We will (as we did in the Newtonian limit) assume this mass to be stationary, so that both $\bar{h}^{\mu\nu}$ and $T^{\mu\nu}$ are only functions of the space coordinates x^i. With this assumption the d'Alembertian reduces to the Laplacian ∇^2 so that we have to solve the equations:

$$\nabla^2 \bar{h}^{\mu\nu}(\mathbf{x}) = -2\kappa T^{\mu\nu}(\mathbf{x})$$

where $\mathbf{x}$ denotes the position vector in spatial coordinates. The general solution to this equation is

$$\bar{h}^{\mu\nu}(\mathbf{x}) = \frac{\kappa}{2\pi}\int_\Sigma \frac{T^{\mu\nu}(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d\Sigma$$

cf. (Ohanian, 1976), where the integral is taken over the region Σ containing the mass, and the prime denotes the integration variable. Expanding $1/|\mathbf{x}-\mathbf{x}'|$ into

a Taylor series and truncating after the second term (i.e. we neglect higher–order multipoles of the mass), we obtain:

$$\bar{h}^{\mu\nu}(\mathbf{x}) = \frac{\kappa}{2\pi r}\int_\Sigma T^{\mu\nu}(\mathbf{x}')d\Sigma + \frac{\kappa x^k}{2\pi r^3}\int_\Sigma x'_k T^{\mu\nu}(\mathbf{x}')d\Sigma\,. \tag{5.27}$$

With a few manipulations, one can show that, in this expression, the integrals

$$\int_\Sigma T^{i\mu}d\Sigma \quad , \quad \int_\Sigma x'_k T^{00}d\Sigma \quad \text{and} \quad \int_\Sigma x'_k T^{ij}d\Sigma$$

vanish. The vanishing of the first and last of these three integrals is a consequence of the conservation of energy (since the energy tensor satisfies the conservation law eq. 5.12), whereas the vanishing of the second integral expresses the choice for the origin of the coordinates to be in the center of mass.

In the following we shall give, for a few simple masses, the components of the energy tensor and derive, by means of the linearized field equations, the expression for the metrical tensor.

The field of a spherical, stationary mass

The elements of the energy tensor $T^{\mu\nu}$ for a spherical, stationary and non–rotating mass, are:

$$\begin{aligned} T^{00} &= \rho c^2 \\ T^{0i} &= 0 \\ T^{ij} &= 0 \end{aligned}$$

where ρ is the mass density, which can be obtained by comparing the energy content of spacetime with that of a perfect fluid without internal stress and pressure, of which the particles are motionless. Using these expressions in equation 5.27, we obtain for the $\bar{h}^{\mu\nu}$:

$$\bar{h}^{00} = \frac{\kappa c^2}{2\pi r}\int_\Sigma \rho d\Sigma = \frac{\kappa c^2 M}{2\pi r} = -\frac{4GM}{rc^2}\,,$$

while all other components vanish. From $h^{\mu\nu} = \bar{h}^{\mu\nu} - \frac{1}{2}\bar{h}\eta^{\mu\nu}$ (which follows from eq. 5.22) we obtain

$$h_{\mu\nu} = -\frac{2GM}{rc^2}\delta_{\mu\nu}\,.$$

With these disturbances, the elements of the metric tensor become:

$$\begin{aligned} g_{00} &= 1 - \frac{2U}{c^2} \\ g_{0i} &= 0 \\ g_{ij} &= -\delta_{ij}(1 + \frac{2U}{c^2}) \end{aligned} \tag{5.28}$$

where $U = \frac{GM}{r}$.

The field of a rotating, stationary mass

Consider now a rotating mass, however still stationary (i.e. negligible time derivatives). The mass particles, out of which the body is built up, have, in this case, some spatial velocity, v^i. The elements of the energy tensor are:

$$\begin{aligned} T^{00} &= \rho c^2 \\ T^{0i} &= \rho c v^i \\ T^{ij} &= \rho v^i v^j \ . \end{aligned}$$

Inserting these values into eq. 5.27 yields:

$$\begin{aligned} \bar{h}^{00} &= -\frac{4GM}{rc^2} \\ \bar{h}^{0i} &= \frac{\kappa c}{2\pi r^3} x^j \int_\Sigma \rho x'_j v^i d\Sigma = \frac{\kappa}{4\pi r^3} \varepsilon^{jik} x_j S_k \\ \bar{h}^{ij} &= 0 \end{aligned}$$

where

$$S_i = \int_\Sigma \varepsilon_{ijk} x'^j T^{k0} d\Sigma$$

is the spin angular momentum (sometimes denoted **J**). Proceeding in the same way as before, the elements of the metrical tensor become:

$$\begin{aligned} g_{00} &= 1 - \frac{2U}{c^2} \\ g_{0i} &= \frac{2G}{c^4 r^3} \varepsilon_{jik} x^j S^k \\ g_{ij} &= -\delta_{ij}(1 + \frac{2U}{c^2}) \ . \end{aligned} \tag{5.29}$$

Assume that the mass rotates around the x^3–axis with constant angular velocity ω_e. Then the elements g_{0i} become:

$$g_{0i} = \frac{4U}{5c^3} \left(\frac{R}{r}\right)^2 \omega_e \varepsilon_{ji3} x^j \ , \tag{5.30}$$

where R is the radius of the sphere.

5.3 Post–Newtonian approximation

Let us compare the metrical elements for the Newtonian limit, a spherically symmetric, stationary mass and a rotating mass, respectively eq. 5.19, 5.28 and 5.29. We see that the term proportional to c^{-2} in the g_{00} component is not a relativistic one. The factor c^2 arises from the fact that the g_{00} corresponds to the time coordinate,

which is $x^0 = ct$. All other terms in the three metrics that contain c, are relativistic contributions. These metrics were obtained by considering simplified situations. In general we have to do with more complicated forms of the energy tensor $T^{\mu\nu}$, with more types of energy included (e.g. internal stress and pressure) and with a non–spherical and non–homogeneous mass distribution. This may make it difficult to obtain an analytical solution of the field equations.

Following the line of construction of the metrics above, it seems a good approximation to expand the elements of the metrical tensor in terms of the small parameter U/c^2. The smallness of this quantity was one of the conditions for the weak field approximation, equation 5.16. Starting from the Newtonian limit, equation 5.19, the first step would be g_{00} up to $O(c^{-4})$, g_{ij} up to $O(c^{-2})$, and g_{0i} up to $O(c^{-3})$. Following steps would include terms of respectively $O(c^{-6})$, $O(c^{-4})$ and $O(c^{-5})$. This way of approximating the metric is called the *Post–Newtonian approximation* (PN approximation), and the metric is called PN–metric, see e.g. (Misner et al., 1973). Including, for all elements, only the first relativistic contribution, we obtain the so–called 1PN–metric, given by :

$$\begin{aligned} g_{00} &= 1 - \frac{2U}{c^2} + \frac{2U^2}{c^4} + O(c^{-6}) \\ g_{0i} &= \frac{4V_i}{c^3} + O(c^{-5}) \\ g_{ij} &= -\delta_{ij}(1 + \frac{2U}{c^2}) + O(c^{-4}) \end{aligned} \tag{5.31}$$

cf. (Soffel, 1990), (Mashhoon et al., 1989), or (Gill et al., 1992), where U is the gravitational potential of the mass and V_i is a vector potential[1] describing the field contributions of magnetic type, arising from moving masses (compare the g_{0i} in metric 5.29). Note that, whereas in the metrics 5.28 and 5.29 the potential $U = GM/r$ comes from a spherical mass distribution, the potential U and V_i in 5.31 (and also in 5.19) are generated by arbitrary masses and are determined respectively by the equations:

$$\nabla^2 U = 4\pi G\rho$$

$$\nabla^2 V_i = 4\pi G\rho v_i$$

cf. e.g. (Chandrasekhar, 1965), where v_i are the spatial velocity components of the mass points. So in general

$$U = G\int_\Sigma \frac{\rho(\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}d\Sigma$$

$$V_i = G\int_\Sigma \frac{\rho(\mathbf{x}')v_i'}{|\mathbf{x}-\mathbf{x}'|}d\Sigma\ .$$

[1] A note on notation: The use of the notation U for the Newtonian (scalar) gravitational potential and V_i for the vector potential is customary in PN theory. For this reason we adopt this notation here, in spite of the fact that we, in previous chapters, used V for the Newtonian potential and V_i for the first derivative (gradient) of the potential.

Expanding $1/|\mathbf{x}-\mathbf{x}'|$ into a Taylor series, we obtain, for a nearly spherical body, metric 5.29 as a first approximation, i.e. $U \approx GM/r$ and $V_i \approx \frac{G}{2cr^3}\varepsilon_{jik}x^j S^k$ (cf. (Soffel, 1990)). It must be noted that the expansion of the metrical elements in this way does in general not constitute an exact analytical solution of the linearized field equations. The PN formalism simply assumes the existence of a coordinate system where the metrical tensor has the form 5.31, cf. (Soffel, 1989).

5.4 Equations of motion revisited

If we are given some expression for the elements of the metrical tensor, we may compute the equations of motion, either by means of a direct evaluation of the geodesic equation 5.4 or using the Euler–Lagrange equations 5.6. In relativistic terms, the gravitational potential is included in the metric, so these equations describe the motion of a free particle (i.e. moving under the influence of gravitation alone), e.g. an earth orbiting satellite (if we neglect disturbing forces like drag or solar pressure).

The geodesic described by equation 5.4 is parametrized with an affine parameter u, in our case the proper time τ. For practical computations, one likes to obtain the equations of motion with respect to the coordinate time t. These can be obtained in the following way. We replace the differentiation $\frac{d}{d\tau}$ by $\frac{d}{dt}\frac{dt}{d\tau}$. Then we obtain:

$$\frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt}\frac{dt}{d\tau}$$

$$\frac{d^2x^\mu}{d\tau^2} = \frac{d^2x^\mu}{dt^2}\left(\frac{dt}{d\tau}\right)^2 + \frac{d^2t}{d\tau^2}\left(\frac{dt}{d\tau}\right)^{-1}\frac{dx^\mu}{d\tau}$$

which, if inserted in the geodesic equation, gives

$$\frac{d^2x^\mu}{dt^2} + \frac{d^2t}{d\tau^2}\left(\frac{dt}{d\tau}\right)^{-2}\frac{dx^\mu}{dt} + \Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{dt}\frac{dx^\sigma}{dt} = 0\,.$$

From the time component ($\mu = 0$) we obtain:

$$\begin{aligned}\frac{d^2x^0}{dt^2} &= -\Gamma^0_{\nu\sigma}\frac{dx^\nu}{dt}\frac{dx^\sigma}{dt} - \frac{d^2t}{d\tau^2}\frac{dx^0}{dt}\left(\frac{dt}{d\tau}\right)^{-2} \\ &= \frac{d^2ct}{dt^2} = 0 \\ \Rightarrow \quad \frac{d^2t}{d\tau^2}\left(\frac{dt}{d\tau}\right)^{-2} &= -\Gamma^0_{\nu\sigma}\frac{dx^\nu}{dt}\frac{dx^\sigma}{dt}c^{-1}\,,\end{aligned}$$

so that, if we use this in the equations of motion for the spatial components ($\mu = i$), we obtain:

$$\frac{d^2x^i}{dt^2} + \Gamma^i_{\nu\sigma}\frac{dx^\nu}{dt}\frac{dx^\sigma}{dt} - \Gamma^0_{\nu\sigma}\frac{dx^\nu}{dt}\frac{dx^\sigma}{dt}\frac{dx^i}{dt}c^{-1} = 0$$

or

$$\frac{d^2x^i}{dt^2} + \left(\Gamma^i_{\nu\sigma} - \Gamma^0_{\nu\sigma}\frac{dx^i}{dt}c^{-1}\right)\frac{dx^\nu}{dt}\frac{dx^\sigma}{dt} = 0 \tag{5.32}$$

as the spatial equations of motion in coordinate time.

For the computation of the equations of motion one needs the Christoffel symbols. They are computed using equation 5.5. In the following we will neglect the time derivatives of the metrical elements, since we assume they will be small compared to the spatial derivatives. Furthermore, we will put $c = 1$. Since c is constant, it acts as a kind of scaling parameter, so that putting it equal to 1 merely means a change of units. When numerical values have to be computed, one inserts in every term as many c's as needed, in order to match the dimensions of all terms, cf. (Ohanian, 1976). The PN–metric for which we will give the equations of motion is

$$\begin{aligned} g_{00} &= 1 - 2U + 2\beta U^2 \\ g_{0i} &= 2(\gamma + 1)V_i \\ g_{ij} &= -\delta_{ij}(1 + 2\gamma U) \end{aligned} \tag{5.33}$$

where the parameters β and γ are added for convenience. If $\beta = 1$ and $\gamma = 1$, we obtain the 1PN–metric 5.31. If $\beta = 0$ we obtain eq. 5.29, the metric of a rotating, stationary mass, and if $V_i = 0$, we obtain 5.28, the metric of a non–rotating mass. In the so–called *Parametrized Post–Newtonian* (PPN) theory, valid for more general mass (or in general energy) distributions, other types of potential functions are included in the metric, each with its own parameter as factor, see e.g. (Misner et al., 1973) or (Will, 1981). Nowadays it is common to include 10 such parameters in the PPN metric, two of which are β and γ, as above. Depending on the value of these parameters, different relativistic theories are obtained (also called metrical theories of gravitation). Einstein's theory of gravitation does not include all these extra PPN terms (i.e. they do not appear in the solution of the Einstein field equations). However, if we choose all PPN parameters to be zero except for β and γ which we choose to be 1, Einstein's theory is included in the PPN formalism. The latter is what we will do here, so we insert $\gamma = 1$. The parameter β is retained and will be propagated into the results so that the influence of the term containing it will be clear. For the 1PN–metric 5.33 the non–vanishing Christoffel symbols become:

$$\begin{aligned} \Gamma^i_{00} &= -U_{,i}(1 - 2(\beta + 1)U) \\ \Gamma^0_{0i} &= -U_{,i}(1 - 2(\beta - 1)U) \\ \Gamma^k_{ij} &= \delta_{ik}U_{,j} + \delta_{jk}U_{,i} - \delta_{ij}U_{,k} \\ \Gamma^j_{0i} &= 2(V_{i,j} - V_{j,i}) \\ \Gamma^0_{ij} &= 2(V_{i,j} + V_{j,i}) \end{aligned} \tag{5.34}$$

where the comma, as usual, denotes partial differentiation. In order to agree with much used (index free) notation, we introduce:

$\mathbf{x}$	$=$	(x, y, z)	particle's spatial position vector
$\mathbf{v}$	$=$	$(\dot{x}, \dot{y}, \dot{z})$	particle's spatial velocity vector (the dot means differentiation with respect to t)
∇U	$=$	(U_x, U_y, U_z)	potential gradient vector
$\mathbf{V}$	$=$	(V_x, V_y, V_z)	vector potential V_i
v^2	$=$	$\dot{x}^2 + \dot{y}^2 + \dot{z}^2$	squared magnitude of velocity
$\times$			cartesian outer product
$\cdot$			cartesian inner product .

With this notation, the equations of motion will become (cf. (Soffel, 1989), (Mashhoon et al., 1989) or (Ries et al., 1990)):

$$\frac{d\mathbf{v}}{dt} = \nabla U + \frac{1}{c^2}\left[-2(\beta+1)U\nabla U - 4(\nabla U \cdot \mathbf{v})\mathbf{v} + v^2\nabla U - 4\,\mathbf{v} \times (\nabla \times \mathbf{V})\right] , \quad (5.35)$$

where clearly the term between square brackets is the 1PN relativistic contribution to the equations of motion. The dominant relativistic contribution will come from a spherical, non–rotating mass, eq. 5.28. If we, therefore, insert $V_i = 0$, $\beta = 1$ and $U = \frac{GM}{r}$ (so that $\nabla U = -\frac{GM}{r^3}\mathbf{x}$), equation 5.35 will reduce to

$$\frac{d\mathbf{v}}{dt} = -\frac{GM}{r^3}\mathbf{x} + \frac{GM}{r^3c^2}\left[4\frac{GM}{r}\mathbf{x} + 4(\mathbf{x}\cdot\mathbf{v})\mathbf{v} - v^2\mathbf{x}\right] . \quad (5.36)$$

These equations may be solved for $\mathbf{x}$ to reveal the relativistic effects in the particle's motion, see e.g. (Soffel, 1989). Here, we are interested in the relativistic contributions to the motion of an earth orbiting satellite, in particular the Aristoteles satellite. So we take for M the mass of the earth and for $\{x^i | i = 1, 2, 3\} \equiv (x, y, z)$ a geocentric coordinate system. It should be noted that this coordinate system, despite the commonly used notation (x, y, z), is *not* a rectilinear (cartesian) coordinate system. In fact, the three coordinates are the spatial part of the four dimensional coordinate system (t, x, y, z) which labels points (events) in spacetime. And we have seen that spacetime is curved, compare the metric 5.33, so that no rectilinear coordinate system can be introduced on a global scale, only locally. In particular, for a spherical symmetric mass distribution, the curvature of spacetime manifests itself in the radial direction and in the time coordinate. This feature was already used by K. Schwarzschild in 1916 who found the first exact solution of the (non–linearized) Einstein field equations. His famous solution for a static spherically symmetric gravitational field (nowadays known as *The Schwarzschild solution*) was

put in terms of polar coordinates (t, r, θ, λ). In this case spacetime is curved in the two coordinates (t, r). As a consequence r does not measure the usual radial distance. However, the geometry for those parts of space which have fixed t and r is just that of a usual sphere. It can be shown that, by a proper choice of coordinates, the Schwarzschild solution gives the 1PN metric as a first approximation, at least the 1PN for a static (i.e. non–rotating) spherical symmetric mass. This means for the coordinates $\{x^i | i = 1,2,3\} \equiv (x, y, z)$, if one would relate them to some kind of polar coordinates $\{x^a | a = 1,2,3\} \equiv (r, \theta, \lambda)$ by means of a coordinate transformation $x^i = x^i(x^a)$ with $x = r\sin\theta\cos\lambda$, $y = r\sin\theta\sin\lambda$ and $z = r\cos\theta$ that r is not the usual radial distance. In order to deal with this problem, one sometimes introduces a new radial coordinate r', defined by $r' = r + \alpha\frac{GM}{c^2}$ where α acts as a gauge parameter for the choice of a definite coordinate system (Soffel, 1989). As a consequence all orbital parameters used to describe the satellite motion depend in this sense on the choice of the coordinate system.

Nevertheless, we will solve the equations of motion 5.36 by interpreting the satellite motion to be the solution of a perturbed Keplerian problem, where the second term on the right hand side (the one with the square brackets) is regarded as the perturbing force. If, in our simulation, one subtracts from the solution obtained in this way, that of a simple Keplerian problem with the same initial state vector but without the relativistic term, one obtains the relativistic contribution to the orbit of the satellite. We used a computer program developed by E. Schrama (1992) to carry out such computations. In reality, when working with real data, one solves, among other parameters, the elements of the initial state vector from the observations in the orbit determination process. In that case the relativistic effect cannot be determined in the way described above, since both orbits (with and without the relativistic perturbation) would have different initial state vectors and can thus not simply be subtracted. Furthermore we have, in that case, to bear in mind that existing values of parameters, which are used in the computations, could originally have been determined without relativistic models, so that possible relativistic effects are implied in those parameters.

It is well known that the dominant relativistic contribution in the motion of a particle in orbit around a spherical symmetric mass (whether it is an artificial satellite around the earth or a planet around the sun) is an advance of the periapse. This can be seen in figure 5.1. In this figure the relativistic orbit contributions for the Aristoteles satellite are shown, expressed in terms of radial, cross track and along track components. This figure was computed in the way described above, namely by solving equation 5.36 for $\mathbf{x}$ and subtracting the solution of the same equation without the relativistic term.

The figure reveals the dominant relativistic orbit contribution as a linearly growing along track effect to be interpreted as the perigeum advance mentioned above. It reaches over 38 m after two weeks. The figure also shows a periodic effect with a frequency of once per orbital revolution. This can be clearly seen if one removes the linear trend and computes the power density spectrum, as is done in figure 5.2 for

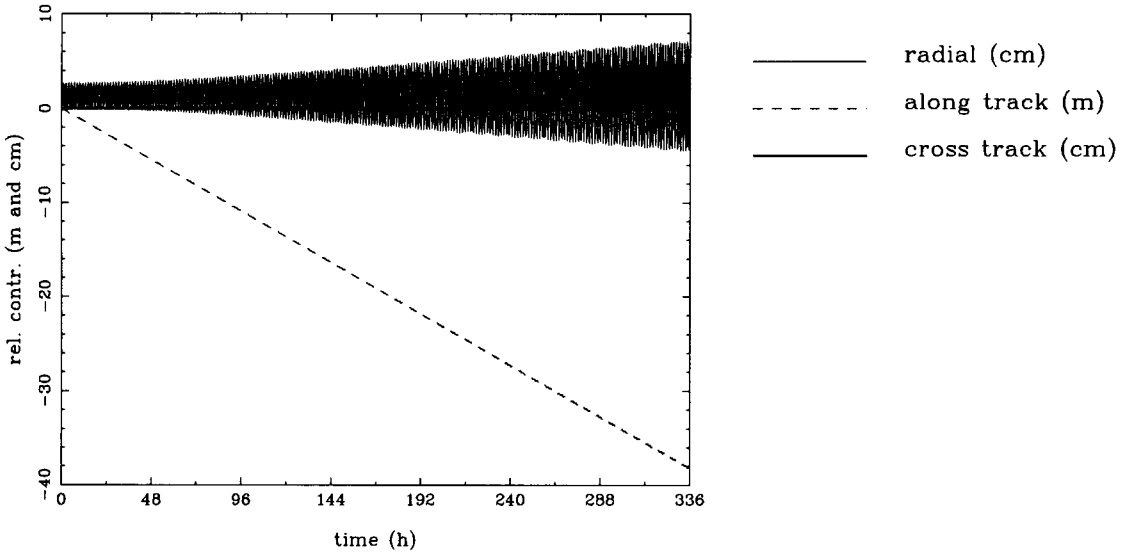


Figure 5.1 *Relativistic orbit contributions for the Aristoteles satellite where only the central term of the potential of the earth was included (so-called Schwarzschild effect). Satellite height was taken to be 200 km and orbit inclination 95°.3. The along track perturbation is plotted in m, the other perturbations in cm.*

the along track effect. For the radial perturbation this periodic effect, with a linearly growing envelope, can be seen in figure 5.1. Nevertheless the radial perturbation is much smaller then the along track effect, being of the order of a few cm after two weeks. The cross track contribution remains below the level of 10^{-4} m.

The along track relativistic effect can not be separated from a change in the value of the gravitational parameter GM, which relates the satellite's mean motion to the semi–major axis a and therefore merely acts as a scaling factor. Estimating the value of GM in the orbit determination process will therefore yield a value in which the dominant relativistic orbit effect is absorbed, cf. (Tapley, 1989) or (Schrama, 1992).

One of the most remarkable differences between the classical Newtonian gravitational theory and Einstein's relativistic theory of gravitation is the fact that in relativity not only a static mass distribution causes a gravitational field (generated by a scalar potential function), but also the *motion* of a mass causes a gravitational field (generated by a *vector* potential function). In analogy with electromagnetic theory the former gravitational field is said to be of electric type (sometimes called gravito–electric field), whereas the latter is of magnetic type (gravito–magnetic field). The vector potential for moving masses (for example rotating objects) appears in the metric in the g_{0i} components, the ones describing the interaction between time and space. In the 1PN metric it was included as the vector potential function V_i (or $\mathbf{V}$ in index free notation). The equations of motion 5.35 in their general form (with V_i included) allow us an easy check of the relativistic contribution to the satellite

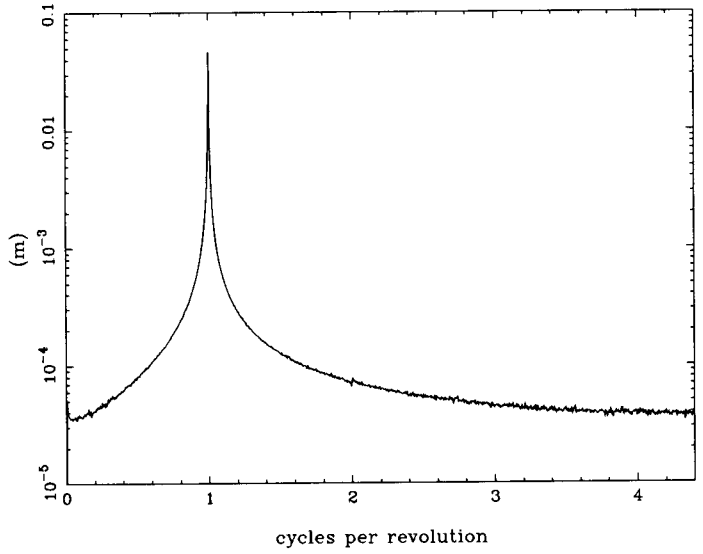


Figure 5.2 *Power density spectrum for the along track relativistic orbit effect for the Aristoteles satellite, remaining after removal of a linear trend.*

motion caused by moving masses. This gravito–magnetic effect on satellite orbits is also called the *Lense–Thirring effect*, first described by Lense and Thirring in 1918 (Thirring and Lense, 1918).

From eq. 5.35 we see that, leaving the vector potential $\mathbf{V}$ in, an additional relativistic term

$$-\frac{4}{c^2}\mathbf{v}\times(\boldsymbol{\nabla}\times\mathbf{V})$$

enters the disturbing acceleration. Let us evaluate this term for our spherical symmetric mass, which we now assume to rotate around the z–axis with angular velocity ω_e. Comparing eq. 5.30 with the 1PN metric 5.33 (with $\gamma = 1$) we find for $\mathbf{V}$:

$$V_i = \frac{1}{5}\frac{GM}{r}\left(\frac{R}{r}\right)^2 \omega_e \varepsilon_{ji3} x^j$$

or in vector notation

$$\mathbf{V} = \frac{1}{5}\frac{GM}{r}\left(\frac{R}{r}\right)^2 \omega_e \begin{bmatrix} -y \\ x \\ 0 \end{bmatrix} .$$

With this expression for **V** the Lense–Thirring part in the relativistic acceleration becomes

$$\frac{4}{5}\frac{GM}{r}\left(\frac{R}{r}\right)^2\frac{\omega_e}{r^2c^2}\begin{bmatrix}\dot{y}(x^2+y^2-2z^2)+3yz\dot{z}\\-\dot{x}(x^2+y^2-2z^2)-3xz\dot{z}\\3z(x\dot{y}-y\dot{x})\end{bmatrix}$$

cf. (Soffel, 1989). The effect of this term on the orbit of the earth orbiting Aristoteles satellite is shown in figure 5.3. It is computed in the same manner as figure 5.1 but now for the Lense–Thirring term only.

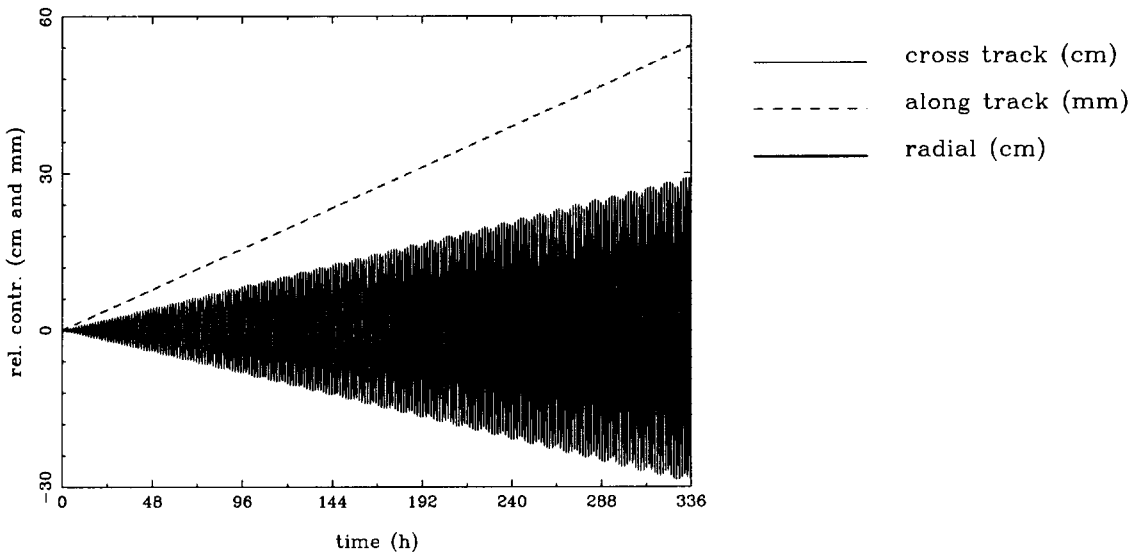


Figure 5.3 *Relativistic Lense–Thirring effect on the orbit of Aristoteles. The earth was assumed a spherical homogeneous mass, rotating with a constant angular velocity around the z–axis. Satellite height was taken to be 200 km and orbit inclination 95°.8.*

The figure mainly shows a once per revolution periodic effect due to this term, being largest in the cross track direction with a linearly growing envelope. In the along track direction the effect is smaller showing a secular trend. The radial effect remains below the level of 10^{-4} m.The cross track effect is caused by a drift of the right ascension of the ascending node Ω at a rate of approximately $3.8 \cdot 10^{-14}$ /s. Such a drift rate is well within the observable limits of modern tracking systems (Schrama, 1992).

5.5 Equation of geodesic deviation

As mentioned in the beginning of this chapter, the gradiometric model can be interpreted in relativistic terms by considering the so–called *equation of geodesic*

deviation. This equation describes the relative motion of two neighbouring particles moving on geodesics. We will derive and discuss it in this section. Consider to this extent the two geodesics, both affinely parametrized, e.g. with the proper time τ. The geodesics are close, and the particles move along them close together, such that, at any value of the parameter τ the coordinates of the first particle are $x^\mu(\tau)$ and of the second $x^\mu(\tau) + \xi^\mu(\tau)$, where ξ^μ is a small vector of coordinate differences. We will assume these coordinate differences to be infinitesimal, as well as their derivatives $d\xi^\mu/d\tau$, which means that not only are the particles close together, but they remain so for a long time. Higher–order terms in, or products of these quantities will therefore be neglected in the derivations. Note that, in general, the coordinates x^μ do not constitute a vector (rank 1 tensor). A tensor A^μ should transform to another coordinate system $x^{\mu'}$ with the transformation matrix $X^{\mu'}_{\ \mu} = \partial x^{\mu'}/\partial x^\mu$, i.e. $A^{\mu'} = X^{\mu'}_{\ \mu} A^\mu$. But coordinates transform via the coordinate transformation equations $x^{\mu'} = x^{\mu'}(x^\mu)$, from which the transformation matrix is derived by means of partial differentiation. However, if *differences* between coordinates are small (like we assume for the ξ^μ) they may be considered the elements of a vector (Foster and Nightingale, 1979), such that they transform as $\xi^{\mu'} = X^{\mu'}_{\ \mu} \xi^\mu$.

The equations of motion for the two particles become:

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma}(x) \frac{dx^\nu}{d\tau} \frac{dx^\sigma}{d\tau} = 0$$

$$\frac{d^2(x^\mu + \xi^\mu)}{d\tau^2} + \Gamma^\mu_{\nu\sigma}(x + \xi) \frac{d(x^\nu + \xi^\nu)}{d\tau} \frac{d(x^\sigma + \xi^\sigma)}{d\tau} = 0 \,,$$

where the Christoffel symbols are evaluated respectively in a point of the first geodesic (with coordinates x^μ) and a point of the second geodesic (with coordinates $x^\mu + \xi^\mu$). The latter is approximated as

$$\Gamma^\mu_{\nu\sigma}(x + \xi) \approx \Gamma^\mu_{\nu\sigma}(x) + \Gamma^\mu_{\nu\sigma,\rho}(x) \xi^\rho$$

so that, if we take the difference between the two geodesic equations, we obtain (neglecting terms quadratic in ξ^μ and $\dot{\xi}^\mu$):

$$\ddot{\xi}^\mu + 2\Gamma^\mu_{\nu\sigma} \dot{\xi}^\sigma \dot{x}^\nu + \Gamma^\mu_{\nu\sigma,\rho} \xi^\rho \dot{x}^\nu \dot{x}^\sigma = 0 \tag{5.37}$$

where the dot denotes differentiation with respect to τ. The first absolute derivative (see appendix B.2.3, eq. B.10) of ξ^μ is

$$\frac{D\xi^\mu}{d\tau} = \frac{d\xi^\mu}{d\tau} + \Gamma^\mu_{\nu\sigma} \xi^\sigma \frac{dx^\nu}{d\tau}$$

from which the second absolute derivative is computed as

$$\frac{D^2 \xi^\mu}{d\tau^2} = \ddot{\xi}^\mu + \Gamma^\mu_{\alpha\sigma,\beta} \xi^\sigma \dot{x}^\alpha \dot{x}^\beta + 2\Gamma^\mu_{\nu\sigma} \dot{\xi}^\nu \dot{x}^\sigma + \Gamma^\mu_{\beta\nu} \Gamma^\nu_{\alpha\sigma} \xi^\sigma \dot{x}^\alpha \dot{x}^\beta - \Gamma^\mu_{\sigma\nu} \Gamma^\nu_{\alpha\beta} \xi^\sigma \dot{x}^\alpha \dot{x}^\beta$$

or

$$\ddot{\xi}^\mu + 2\Gamma^\mu_{\nu\sigma}\dot{\xi}^\sigma\dot{x}^\nu = \frac{D^2\xi^\mu}{d\tau^2} - \left(\Gamma^\mu_{\alpha\sigma,\beta} + \Gamma^\mu_{\beta\nu}\Gamma^\nu_{\alpha\sigma} - \Gamma^\mu_{\sigma\nu}\Gamma^\nu_{\alpha\beta}\right)\xi^\sigma\dot{x}^\alpha\dot{x}^\beta$$

where we made use of the geodesic equation and relabelled dummy indices whenever convenient. Inserting this result in eq. 5.37, yields

$$\frac{D^2\xi^\mu}{d\tau^2} - \left(\Gamma^\mu_{\alpha\sigma,\beta} + \Gamma^\mu_{\beta\nu}\Gamma^\nu_{\alpha\sigma} - \Gamma^\mu_{\sigma\nu}\Gamma^\nu_{\alpha\beta}\right)\xi^\sigma\dot{x}^\alpha\dot{x}^\beta + \Gamma^\mu_{\alpha\beta,\sigma}\xi^\sigma\dot{x}^\alpha\dot{x}^\beta = 0$$

or

$$\begin{aligned}\frac{D^2\xi^\mu}{d\tau^2} &= \left(\Gamma^\mu_{\alpha\sigma,\beta} - \Gamma^\mu_{\alpha\beta,\sigma} + \Gamma^\mu_{\beta\nu}\Gamma^\nu_{\alpha\sigma} - \Gamma^\mu_{\sigma\nu}\Gamma^\nu_{\alpha\beta}\right)\xi^\sigma\dot{x}^\alpha\dot{x}^\beta \\ &= -R^\mu{}_{\alpha\sigma\beta}\xi^\sigma\dot{x}^\alpha\dot{x}^\beta \end{aligned} \tag{5.38}$$

with $R^\mu{}_{\alpha\beta\sigma}$ the Riemann–Christoffel curvature tensor. This is the equation of geodesic deviation, sometimes called *Jacobi equation*. It shows how two neighbouring geodesics deviate relative to each other. It is interesting, though not surprisingly, that the Riemann tensor enters this equation. The Riemann tensor describes the curvature of the space under consideration. If a global (i.e. covering the whole of space) coordinate system can be introduced, in which all components of the Riemann tensor vanish, the space is flat. In a flat space, in cartesian coordinates, the absolute derivative $D\xi^\mu/d\tau$ reduces to the ordinary total derivative $d\xi^\mu/d\tau$. So, with $R^\mu{}_{\alpha\beta\sigma} = 0$, the equation of geodesic deviation would reduce to

$$\frac{d^2\xi^\mu}{d\tau^2} = 0$$

of which the solution for ξ^μ is

$$\xi^\mu(\tau) = C^\mu\tau + D^\mu$$

where C^μ and D^μ are constants. We see that in this case, the separation between the geodesics grows linearly, as it should, since in a flat space we know that geodesics are straight lines. In a curved space, some components of the Riemann tensor will always be different from zero, whatever coordinate system we choose. The separation vector between the two geodesics will not change linearly as function of τ, but accelerated (the quantity $D^2\xi^\mu/d\tau^2$ is the "relative acceleration" vector), depending on the curvature of the space.

So the curvature of a space determines the relative acceleration between two free particles. On the other hand, measurement of this relative acceleration, tells us about the curvature of the space. Since in satellite gradiometry we measure such relative accelerations, this technique should give us the curvature of spacetime, which is curved due to the earth's gravitational field. If this is so, then the equation of geodesic deviation should reduce to the basic gradiometric model equation 2.1 in the Newtonian limit. This will be shown in the sequel.

First we will repeat the assumptions needed to arrive at Newton. The Newtonian limit is derived from the linearized (weak field) approximation. In this approximation we take for the metrical tensor $g_{\mu\nu}$ the expression $\eta_{\mu\nu} + h_{\mu\nu}$, with small $h_{\mu\nu}$ and small derivatives $h_{\mu\nu,\sigma}$. The latter also implies that the Christoffel symbols $\Gamma^{\sigma}_{\mu\nu}$ are small. Another assumption we made in order to arrive at Newton was $v << c$, so that $\gamma \approx 1$, thus $d/d\tau \approx d/dt$. This expresses the Newtonian situation of universal time t, i.e. all clocks read the same time. Furthermore, derivatives with respect to time are assumed small in comparison to spatial derivatives. In section 5.2.1 we compared the geodesic equation 5.4, under the above assumptions, with the Newtonian equations of motion, in order to show that the former will reduce to the latter. This appeared to be so if the only non–vanishing Christoffel symbols are

$$\Gamma^{i}_{00} = \frac{1}{2}\delta^{ij} h_{00,j} \; . \tag{5.39}$$

Since all clocks measure the same universal time t, so will the two clocks moving along with the two particles on the neighbouring geodesics, so that the first component of the separation vector ξ^{μ} will be zero, i.e. $\xi^0 = 0$. This simply means that the particle's accelerations are compared at equal times (Ohanian, 1976). With this result, and with eq. 5.39, the equation for the absolute derivative of the separation vector will become:

$$\begin{aligned}
\frac{D\xi^{\mu}}{d\tau} &= \frac{d\xi^{\mu}}{d\tau} + \Gamma^{\mu}_{\nu\sigma}\xi^{\sigma}\frac{dx^{\nu}}{d\tau} \\
&= \frac{d\xi^{\mu}}{d\tau} + \Gamma^{\mu}_{00}\xi^{0}\frac{dx^{0}}{d\tau} + \Gamma^{\mu}_{0i}\xi^{i}\frac{dx^{0}}{d\tau} + \Gamma^{\mu}_{i0}\xi^{0}\frac{dx^{i}}{d\tau} + \Gamma^{\mu}_{ij}\xi^{j}\frac{dx^{i}}{d\tau} \\
&= \frac{d\xi^{\mu}}{d\tau} \; .
\end{aligned}$$

In the same way, we find for the second absolute derivative $D^2\xi^{\mu}/d\tau^2 = d^2\xi^{\mu}/d\tau^2$. The latter is equal to $d^2\xi^{\mu}/dt^2$ since $\gamma = 1$. Thus, in the Newtonian limit, the left hand side of the equation of geodesic deviation 5.38 reduces to $d^2\xi^{\mu}/dt^2$. In order to find the Newtonian limit of the right hand side, we look at the expression for the Riemann–Christoffel tensor B.16. Products of the Christoffel symbols are neglected, and we are left with

$$R^{\mu}{}_{\alpha\sigma\beta} = \Gamma^{\mu}_{\alpha\beta,\sigma} - \Gamma^{\mu}_{\alpha\sigma,\beta} \; .$$

With eq. 5.39, the only non–vanishing components of this tensor are $R^{i}{}_{0k0} = -R^{i}{}_{00k} = \Gamma^{i}_{00,k} = \frac{1}{2}\delta^{ij}h_{00,jk}$. For $\mu = 0$ the equation of geodesic deviation becomes

$$\frac{d^2\xi^0}{dt^2} = 0$$

which indeed agrees with $\xi^0 = 0$. The spatial components yield

$$\frac{d^2\xi^{i}}{dt^2} = -R^{i}{}_{0k0}\xi^{k}\frac{dx^{0}}{dt}\frac{dx^{0}}{dt} - R^{i}{}_{00k}\xi^{0}\frac{dx^{0}}{dt}\frac{dx^{k}}{dt}$$

$$
\begin{aligned}
&= -R^{i}{}_{0k0}\xi^k c^2 \\
&= -\frac{1}{2}\delta^{ij}h_{00,jk}c^2\xi^k \\
&= \delta^{ij}\frac{\partial^2 V}{\partial x^j \partial x^k}\xi^k \,.
\end{aligned}
\tag{5.40}
$$

In the latter equation we inserted the h_{00} from eq. 5.18. Comparing the above equation with 2.1, we see that, indeed, the equation of geodesic deviation is the relativistic generalization of the gradiometric equation, at least if we take ξ^k as the coordinate differences between the two proof masses and $d^2\xi^i/dt^2$ as their acceleration difference. In relativistic terms, a gradiometer measures the curvature of the space, cf. (Misner et al., 1973), (Ohanian, 1976) or (Gill et al., 1992).

This curvature is completely determined by the elements of the Riemann–Christoffel tensor. Therefore, we will give those elements (or rather those of the covariant form of the curvature tensor, $R_{\nu\alpha\sigma\beta}$) for the 1PN metric 5.33. Inserting the Christoffel symbols from eq. 5.34 into the expression for the Riemann tensor B.16, we obtain for the non–vanishing elements of $R_{\nu\alpha\sigma\beta}$:

$$
\begin{aligned}
R_{0i0j} &= U_{,ij}(1-2\beta U) - (1+2\beta)U_{,i}U_{,j} + \delta_{ij}\delta^{kl}U_{,k}U_{,l} \\
R_{ijkl} &= \delta_{ik}U_{,jl} - \delta_{il}U_{,jk} + \delta_{jl}U_{,ik} - \delta_{jk}U_{,il} \\
R_{0ijk} &= 2(V_{k,ji} - V_{j,ki})
\end{aligned}
\tag{5.41}
$$

from which the non–vanishing elements of the Ricci tensor may be computed as:

$$
\begin{aligned}
R_{00} &= \delta^{kl}\{(1-2(\beta+1)U)U_{,kl} + 2(1-\beta)U_{,k}U_{,l}\} \\
R_{ii} &= \delta^{kl}U_{,kl} \\
R_{0i} &= 2\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{V})
\end{aligned}
$$

and the curvature scalar

$$
R = -2\delta^{kl}\{(1-(8-2\beta)U)U_{,kl} - (5-2\beta)U_{,k}U_{,l}\} \,.
$$

The elements of the Riemann tensor in eq. 5.41 are given with respect to the global (and curvilinear!) 4–dimensional coordinate system (t,x,y,z) spanning the whole of curved spacetime. They are not exactly the quantities which an actual satellite gradiometer directly measures. In order to find out what the relativistic contributions to the gradiometric measurements are we have to switch from the general expression 5.38 to some, more operational, equation in three dimensional space. To do this, some additional steps have to be taken, the most important of which is the definition of the so–called *proper reference frame.*

This is a local reference frame, moving with the observer along its worldline, and consisting of four orthonormal base vectors (also called an orthonormal tetrad), see e.g. (Synge, 1960) or (Misner et al., 1973). The base vectors may be chosen such as

to coincide in every point of the observer's worldline with the tangent vectors to the coordinate lines of some global coordinate grid. Such a frame is sometimes called a coordinate induced frame (Soffel, 1989). In this case we have, denoting the base vectors by $\mathbf{e}_{(\alpha)}$ (sometimes $\mathbf{e}_{\hat{\alpha}}$)

$$\mathbf{e}_{(\alpha)} = \frac{\partial}{\partial x^{\alpha}} .$$

The fact that these base vectors are orthonormal is expressed by

$$g^{\mu\nu} e_{(\alpha)\mu} e_{(\beta)\nu} = \eta_{(\alpha\beta)} . \tag{5.42}$$

In the notation $e_{(\alpha)\mu}$ the index (α) indicates which base vector it is and the index μ indicates the component of this base vector with respect to the coordinate system x^{μ}. In the sequel, the "indices between parentheses" (which in fact are labels to distinguish the base vectors) are treated as normal indices, which can be raised or lowered and summed over, cf. (Synge, 1960). This means we may define "contravariant" base vectors in the following way:

$$\mathbf{e}^{(\alpha)} = \eta^{(\alpha\beta)} \mathbf{e}_{(\beta)} .$$

Note that for the components of these base vectors with respect to some coordinate system the usual relations for raising and lowering of indices applies, i.e.

$$e_{(\alpha)\mu} = g_{\mu\nu} e^{\nu}_{(\alpha)} .$$

If the observer moves along a geodesic (i.e. he is not accelerated due to some external force or rotation) this proper reference frame is a local Lorentz frame, in which

$$g_{(\alpha\beta)} = \eta_{(\alpha\beta)} \qquad \text{and} \qquad \Gamma^{(\alpha)}_{(\beta\gamma)} = 0 . \tag{5.43}$$

Along a geodesic the base vectors of such orthonormal tetrad form a parallel vector field, i.e.

$$\frac{D e^{\mu}_{(\alpha)}}{du} = 0 . \tag{5.44}$$

Constructing an orthonormal tetrad in the way described above, we may resolve any vector or tensor into components along this tetrad, e.g.

$$\begin{aligned} v^{(\alpha)} &= v^{\mu} e^{(\alpha)}_{\mu} = v_{\mu} e^{(\alpha)\mu} \\ v_{(\alpha)} &= v^{\mu} e_{(\alpha)\mu} = v_{\mu} e^{\mu}_{(\alpha)} \\ t^{(\alpha\beta)} &= t^{\mu\nu} e^{(\alpha)}_{\mu} e^{(\beta)}_{\nu} = t_{\mu\nu} e^{(\alpha)\mu} e^{(\beta)\nu} \\ t_{(\alpha\beta)} &= t^{\mu\nu} e_{(\alpha)\mu} e_{(\beta)\nu} = t_{\mu\nu} e^{\mu}_{(\alpha)} e^{\nu}_{(\beta)} \end{aligned}$$

and the inverse relations

$$
\begin{aligned}
v^{\mu} &= v_{(\alpha)} e^{(\alpha)\mu} = v^{(\alpha)} e^{\mu}_{(\alpha)} \\
v_{\mu} &= v_{(\alpha)} e^{(\alpha)}_{\mu} = v^{(\alpha)} e_{(\alpha)\mu} \\
t^{\mu\nu} &= t^{(\alpha\beta)} e^{\mu}_{(\alpha)} e^{\nu}_{(\beta)} = t_{(\alpha\beta)} e^{(\alpha)\mu} e^{(\beta)\nu} \\
t_{\mu\nu} &= t^{(\alpha\beta)} e_{(\alpha)\mu} e_{(\beta)\nu} = t_{(\alpha\beta)} e^{(\alpha)}_{\mu} e^{(\beta)}_{\nu}
\end{aligned}
$$

and analogously for higher–order tensors. These kinds of components are invariants in the sense of tensor analysis (because they are not the components with respect to some coordinate system) but they obviously depend on the chosen orthonormal tetrad (Synge, 1960). We will now transform the equation of geodesic deviation from the tensorial expression 5.38 into an invariant expression with respect to a local orthonormal tetrad. Since the actual gradiometric measurements are performed with respect to such a local tetrad (namely a frame "bolted into the floor and walls of the satellite", (Misner et al., 1973, p. 327)), the resulting *invariant deviation equation* should reveal the relativistic contributions to the measured gradients. In this way we have constructed an operational description of relativistic gradiometry.

First, we have from equation 5.44

$$
\frac{De^{\mu}_{(\alpha)}}{du} = \frac{de^{\mu}_{(\alpha)}}{du} + \Gamma^{\mu}_{\nu\sigma} e^{\nu}_{(\alpha)} \frac{dx^{\sigma}}{du} = 0
$$

or

$$
\frac{de^{\mu}_{(\alpha)}}{du} = -\Gamma^{\mu}_{\nu\sigma} e^{\nu}_{(\alpha)} \frac{dx^{\sigma}}{du} \,. \tag{5.45}
$$

For the infinitesimal displacement vector ξ^{μ} between the two neighbouring geodesics from eq. 5.38 we have

$$
\frac{D\xi^{\mu}}{du} = \frac{d\xi^{\mu}}{du} + \Gamma^{\mu}_{\nu\sigma} \xi^{\nu} \frac{dx^{\sigma}}{du}
$$

and if we write $\xi^{\mu} = \xi^{(\alpha)} e^{\mu}_{(\alpha)}$ this yields

$$
\frac{D\xi^{\mu}}{du} = e^{\mu}_{(\alpha)} \frac{d\xi^{(\alpha)}}{du} + \xi^{(\alpha)} \frac{de^{\mu}_{(\alpha)}}{du} + \Gamma^{\mu}_{\nu\sigma} \xi^{(\alpha)} e^{\nu}_{(\alpha)} \frac{dx^{\sigma}}{du}
$$

which becomes if we insert eq. 5.45:

$$
\frac{D\xi^{\mu}}{du} = e^{\mu}_{(\alpha)} \frac{d\xi^{(\alpha)}}{du} \,.
$$

For the second absolute derivative we then find

$$
\frac{D^2\xi^{\mu}}{du^2} = \frac{D}{du}\left(\frac{D\xi^{\mu}}{du}\right) = \frac{D}{du}\left(e^{\mu}_{(\alpha)} \frac{d\xi^{(\alpha)}}{du}\right)
$$

$$
\begin{aligned}
&= \frac{De^{\mu}_{(\alpha)}}{du}\frac{d\xi^{(\alpha)}}{du} + e^{\mu}_{(\alpha)}\frac{D}{du}\frac{d\xi^{(\alpha)}}{du} \\
&= e^{\mu}_{(\alpha)}\frac{D}{du}\frac{d\xi^{(\alpha)}}{du} \qquad \text{(with eq. 5.44)} \\
&= e^{\mu}_{(\alpha)}\left(\frac{d}{du}\frac{d\xi^{(\alpha)}}{du} + \Gamma^{(\alpha)}_{(\beta\gamma)}\frac{d\xi^{(\beta)}}{du}\frac{dx^{(\gamma)}}{du}\right) \\
&= e^{\mu}_{(\alpha)}\frac{d^2\xi^{(\alpha)}}{du^2} \qquad \text{(with eq. 5.43)} \qquad (5.46)
\end{aligned}
$$

Insertion into eq. 5.38 gives

$$
e^{\mu}_{(\alpha)}\frac{d^2\xi^{(\alpha)}}{du^2} + R^{\mu}{}_{\nu\sigma\rho}\xi^{\sigma}\dot{x}^{\nu}\dot{x}^{\rho} = 0 \,.
$$

Multiplication of this equation with $e^{(\beta)}_{\mu}$ yields

$$
\begin{aligned}
e^{(\beta)}_{\mu}e^{\mu}_{(\alpha)}\frac{d^2\xi^{(\alpha)}}{du^2} + e^{(\beta)}_{\mu}R^{\mu}{}_{\nu\sigma\rho}\xi^{\sigma}\dot{x}^{\nu}\dot{x}^{\rho} &= 0 \\
\frac{d^2\xi^{(\alpha)}}{du^2} + e^{(\alpha)\mu}R_{\mu\nu\sigma\rho}\xi^{\sigma}\dot{x}^{\nu}\dot{x}^{\rho} &= 0 \,.
\end{aligned}
$$

If we now insert for the vectors $\dot{x}^{\nu}$, $\dot{x}^{\rho}$ and ξ^{σ} their transformations to the invariant components relative to the orthonormal tetrad (respectively $\dot{x}^{(\beta)}e^{\nu}_{(\beta)}$, $\dot{x}^{(\delta)}e^{\rho}_{(\delta)}$ and $\xi^{(\gamma)}e^{\sigma}_{(\gamma)}$) we obtain

$$
\frac{d^2\xi^{(\alpha)}}{du^2} + \eta^{(\alpha\epsilon)}R_{(\epsilon\beta\gamma\delta)}\xi^{(\gamma)}\dot{x}^{(\beta)}\dot{x}^{(\delta)} = 0
$$

where we defined the invariant components

$$
R_{(\alpha\beta\gamma\delta)} = R_{\mu\nu\sigma\rho}e^{\mu}_{(\alpha)}e^{\nu}_{(\beta)}e^{\sigma}_{(\gamma)}e^{\rho}_{(\delta)} \,.
$$

As a final step we define

$$
K_{(\epsilon\gamma)} \equiv R_{(\epsilon\beta\gamma\delta)}\dot{x}^{(\beta)}\dot{x}^{(\delta)}
$$

with which the invariant deviation equation becomes, cf. (Synge, 1960):

$$
\frac{d^2\xi^{(\alpha)}}{du^2} + \eta^{(\alpha\beta)}K_{(\beta\gamma)}\xi^{(\gamma)} = 0
$$

or in "covariant" form (obtained after multiplication with $\eta_{(\alpha\delta)}$)

$$
\frac{d^2\xi_{(\alpha)}}{du^2} + K_{(\alpha\beta)}\xi^{(\beta)} = 0
$$

of which the spatial part is

$$
\frac{d^2\xi_{(i)}}{du^2} + K_{(ij)}\xi^{(j)} = 0 \qquad (5.47)
$$

which closely resembles the Newtonian form of the gradiometric equation. The quantity $K_{(ij)}$ takes the place of the gradient tensor and is therefore sometimes referred to as tidal matrix. Its components are found as

$$K_{(ij)} = R_{(i\beta j\delta)}\dot{x}^{\beta}\dot{x}^{\delta}$$

which approximately equals (at least for low spatial velocities)

$$\begin{aligned} K_{(ij)} &= c^2 R_{(i0j0)} \\ &= c^2 R_{\mu\nu\sigma\rho} e^{\mu}_{(i)} e^{\nu}_{(0)} e^{\sigma}_{(j)} e^{\rho}_{(0)} \,. \end{aligned} \tag{5.48}$$

The dominant relativistic contributions to the gradients are again expected to result from a static spherical symmetric mass distribution. For such a mass distribution it is convenient to use the 4–dimensional polar coordinates $\{x^{\mu'}|\mu' = 0,1,2,3\} \equiv (ct, r, \theta, \lambda)$ with the primes dropped in the sequel. In these coordinates, the non–vanishing components of the Riemann tensor, computed from eq. 5.41 with c's restored, are (up to 1PN order in c^{-2})

$$\begin{aligned} R_{0101} &= \frac{2GM}{r^3c^2}\left(1 - 3\frac{GM}{rc^2}\right) \\ R_{0202} &= -\frac{GM}{rc^2}\left(1 - 3\frac{GM}{rc^2}\right) \\ R_{0303} &= -\sin^2\theta\frac{GM}{rc^2}\left(1 - 3\frac{GM}{rc^2}\right) \\ R_{1212} &= \frac{GM}{rc^2} \\ R_{1313} &= \sin^2\theta\frac{GM}{rc^2} \\ R_{2323} &= 2\,r\sin^2\theta\frac{GM}{c^2} \end{aligned} \tag{5.49}$$

where we have put $\beta = 1$. Finally we need, in order to evaluate eq. 5.48, expressions for the components of the base vectors of the orthonormal tetrad. The tangent vectors to the coordinate lines of the coordinate grid (ct, r, θ, λ) are respectively $\partial/c\partial t$, $\partial/\partial r$, $\partial/\partial\theta$ and $\partial/\partial\lambda$. In order to fulfil eq. 5.42 (so that the base vectors are orthonormal) we have (up to first order in U/c^2)

$$\begin{aligned} \mathbf{e}_{(0)} &= \left(1 + \frac{GM}{rc^2}\right)\frac{\partial}{c\partial t} \\ \mathbf{e}_{(1)} &= \left(1 - \frac{GM}{rc^2}\right)\frac{\partial}{\partial r} \\ \mathbf{e}_{(2)} &= \frac{1}{r}\left(1 - \frac{GM}{rc^2}\right)\frac{\partial}{\partial\theta} \\ \mathbf{e}_{(3)} &= \frac{1}{r\sin\theta}\left(1 - \frac{GM}{rc^2}\right)\frac{\partial}{\partial\lambda} \,. \end{aligned}$$

This system of base vectors can be constructed in every point of the observer's worldline. But each of those frames is at rest with respect to the spatial part of spacetime. The proper reference frame of an observer co–moving with a satellite is moving through space, and if it is earth pointing, it also undergoes a spatial rotation. Without any loss of generality (as a result of the spherical symmetry of the mass) we will, for convenience, consider here an equatorial satellite orbit, i.e. $\theta = \pi/2$. The motion of the proper reference frame then takes place in the $\mathbf{e}_{(3)}$ direction, and we assume it to be circular. In order to obtain the co–moving tetrad $\mathbf{e}_{(\alpha')}$, we apply a Lorentz transformation to the $\mathbf{e}_{(\alpha)}$ frame where we take for the velocity the satellite's mean orbital motion n_0. Thus $v^{(i)} = n_0 r k^{(i)}$ where $k^{(i)}$ is a spatial unit vector pointing in the along track direction, which in our case is the $\mathbf{e}_{(3)}$ direction, so $\{k^{(i)}\} = (0,0,1)$. Such a transformation is also called a *Lorentz boost* in $\mathbf{e}_{(3)}$ direction. So

$$\mathbf{e}_{(\alpha')} = \Lambda_{(\alpha')}^{(\beta)} \mathbf{e}_{(\beta)}$$

where $\Lambda_{(\alpha')}^{(\beta)}$ is the Lorentz transformation matrix (e.g. (Misner et al., 1973) or (Soffel, 1989)) given (up to 1PN order) by

$$\{\Lambda_{(\alpha')}^{(\beta)}\} = \begin{bmatrix} 1+\frac{1}{2}\frac{GM}{rc^2} & 0 & 0 & \frac{n_0 r}{c} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{n_0 r}{c} & 0 & 0 & 1+\frac{1}{2}\frac{GM}{rc^2} \end{bmatrix} .$$

With this matrix the base vectors of the co–moving frame $\mathbf{e}_{(\alpha')}$ become

$$\begin{aligned} \mathbf{e}_{(0')} &= \left(1+\frac{3}{2}\frac{GM}{rc^2}\right)\frac{\partial}{c\partial t} + \frac{n_0}{c}\frac{\partial}{\partial\lambda} \\ \mathbf{e}_{(1')} &= \left(1-\frac{GM}{rc^2}\right)\frac{\partial}{\partial r} \\ \mathbf{e}_{(2')} &= \frac{1}{r}\left(1-\frac{GM}{rc^2}\right)\frac{\partial}{\partial\theta} \\ \mathbf{e}_{(3')} &= \frac{n_0 r}{c}\frac{\partial}{c\partial t} + \frac{1}{r}\left(1-\frac{1}{2}\frac{GM}{rc^2}\right)\frac{\partial}{\partial\lambda} . \end{aligned}$$

Now we are in a position to compute the components of the tidal matrix $K_{(ij)}$ (omitting the primes). We obtain from eq. 5.48:

$$\begin{aligned} K_{(11)} &= 2\frac{GM}{r^3} - 3\frac{(GM)^2}{r^4c^2} \\ K_{(22)} &= -\frac{GM}{r^3} \\ K_{(33)} &= -\frac{GM}{r^3} + 3\frac{(GM)^2}{r^4c^2} \end{aligned} \qquad (5.50)$$

of which the terms with c^{-2} are obviously the first relativistic contributions to the gradients. The first terms on the right hand side coincide with the Newtonian expressions for the gradients of a spherical symmetric mass distribution. For example, since the $x^{(\alpha)=1}$ component points in the radial direction, $K_{(11)}$ is equivalent to V_{zz} in the local orbital (cartesian) coordinate system which we used in the non–relativistic approach. In the same way $K_{(33)}$ coincides with V_{xx} and $K_{(22)}$ with V_{yy}. Estimating the size of the relativistic contributions, we see that they are of the order of $3 \cdot 10^{-6}$ E for a satellite at 200 km altitude, cf. (Soffel et al., 1987) or (Gill et al., 1992). For the present Aristoteles mission scenario (200 km, 0.01 E/$\sqrt{\text{Hz}}$ measurement precision) such small contribution can be neglected. Other relativistic effects on the gradients due to e.g. the gravito–magnetic field or higher–order multipole expansions are not discussed here. The reader is referred to (Theiß, 1984) and (Gill et al., 1992).

Conclusions

By means of a technique called *differential accelerometry* it is possible to measure the second–order partial derivatives of the earth's gravitational potential. An instrument which measures these so–called *gravity gradients* is called a *gradiometer*. If a gradiometer is on board an earth orbiting satellite, a global set of gravity gradients can be obtained in a few months time. From this set of observations the gravitational potential (represented by a set of *potential coefficients*) can be derived by means of an iterative least squares adjustment, with which at the same time the satellite's trajectory can be estimated. The linear model used for the adjustment is usually obtained from partial differentiation of a series expansion of the gravitational potential. If, in the future, more accurate gradiometers and/or satellite tracking systems become available, it might be necessary to include relativistic effects in the model both for the satellite orbit and the gradients.

It is customary, and for our purposes also appropriate, to expand the gravitational potential function, using geocentric polar coordinates $x^A = (r, \theta, \lambda)$, into a series of spherical harmonic functions. In such a series the *Legendre functions* appear. For satellite applications, however, it is more suitable to use a coordinate system related to the satellite orbit. In that case, the gravitational potential is written as function of orbital coordinates $x^a = (r, \omega_o, \omega_e)$ or $x^{a'} = (r, \omega_o, I)$. Due to the rotation of an equatorial coordinate system to an orbital system, so–called *inclination functions* appear in the series. Expressions of the first– and second–order derivatives of the potential can be easily derived by differentiating the series expansion with respect to the coordinates. In the case of satellite gradiometry, expressions for the second–order derivatives are needed with respect to a local cartesian coordinate system. Therefore, the partial derivatives of the potential series expansion have to be transformed from any of the geocentric curvilinear coordinate systems x^A, x^a or $x^{a'}$ to this local cartesian system. With the help of *index notation* and *tensor analysis* a compact, general algorithm can be formulated for the transformation of potential derivatives between arbitrary coordinate systems.

The transformation equations, in which series expressions for the potential derivatives are inserted, then serve as observation equations in a gradiometric analysis procedure. By means of a *least squares adjustment* approach, the unknown parameters (in our case the potential coefficients) can be estimated. An advantage of such least squares estimation is that an *error analysis* can be carried out prior to an actual

experiment. The inverse of the normal matrix provides the formal variances and covariances of the unknown parameters. Thus, given some mission scenario (in terms of satellite altitude, orbital inclination, mission duration, sampling interval and measured tensor components) and some a–priori error model for the observations, the precision of the estimated potential coefficients can be computed and compared with the requirements. Reversely, given such requirements, the error analysis can be used to put demands on the mission scenario as to meet these requirements.

The requirements for a gradiometer mission typically are less than 100 km spatial resolution (half–wavelength) and a precision of less than 10 cm global r.m.s. for geoid heights or 5 mgal for gravity anomalies. The mentioned resolution corresponds to a spherical harmonic expansion up to at least 180 or more (somewhere between 200 or 300). From the error analysis appeared that, for most mission scenarios, a signal to noise ratio of 1 was reached near degree 240, which will be sufficient for the required resolution.

If *global r.m.s.* values for geoid heights or gravity anomalies have to be computed from the a–posteriori error estimates of the potential coefficients (as they result from the error analysis), we have to bear in mind that in reality the gravitational potential is a continuous function and should be represented by an infinite series (including all frequencies from zero to infinity). The error analysis, however, is limited to some maximum degree and order, in our case 240. The error resulting from this part is called *commission error*. The neglected part of the spectrum, above degree 240, also contributes to the global mean error, and should in fact be taken into account. This neglected part is called *omission error*. Since we do not know the true gravitational spectrum, the omission part can only be approximated by using some prior knowledge of the average behaviour of the gravity field. The global error r.m.s. values presented in this work always represent the sum of commission and omission error. The omission part is taken into account up to a maximum degree of 1000. Above this degree no substantial contribution was found. Furthermore, we always show *smoothed* global error values, representing $1° \times 1°$ block averages. Due to the mission altitude and the sampling distance it is difficult to estimate higher degrees so that smoothing is justified for our purposes. Depending on further applications, un–smoothed values (representing point values) can be used too.

A spherical harmonic expansion up to a maximum degree and order L of 240 contains more than 58,000 potential coefficients. If these are to be estimated in a least squares adjustment, the normal matrix will have a size of $(58,000)^2$. Even with large computers, the solution of a system of linear equations of this size will be a major task. Under certain assumptions, and by ordering the unknowns in a specific manner, the normal matrix attains a *block–diagonal* structure. The largest block to be inverted will have a size of $(240/2+1)^2$, so that computation time is drastically decreased.

The assumptions to arrive at the block–diagonal structure are: 1.) circular orbit; 2.) regularly distributed data along the orbit (i.e. no data gaps); and 3.) the number of orbital revolutions and the number of nodal days contained in one

repeat period are relative prime integers, where the number of revolutions should be larger than $2L$. The second of these assumptions seems to be the most critical one. *Data gaps* are likely to occur, not only due to instrumental failures, but also due to orbit maintenance maneuvers and as a result of excessive drag variations near the poles and the equator. A non–continuous data stream destroys the orthogonality properties on which the block–diagonal structure is based.

In an ideal situation, an SGG mission will be flown with a full tensor gradiometer, measuring at a sampling rate of 4 seconds all nine gradients with a 0.01 $\mathrm{E}/\sqrt{\mathrm{Hz}}$ white noise error spectrum. The satellite will fly in a circular, polar orbit at e.g. 200 km altitude for at least 6 months. Global error r.m.s. values derived from such an ideal mission will be approximately 8.6 cm for geoid undulations and 3.6 mgal for gravity anomalies at a resolution of 0.°75 half–wavelength. These results easily fulfil the requirements given above, even without having additional information, either from a GPS receiver on board the satellite or in terms of a–priori information in a kind of least squares collocation set–up.

However, a more realistic mission like *Aristoteles* differs from the ideal situation on several aspects. The three most important differences are: 1.) band limitation and coloured noise measurement error spectrum; 2.) non–polar orbit; and 3.) planar gradiometer measuring only the out of plane tensor components with high enough precision. All three limitations have a rather large effect on the results. Apart from numerical singularities (resulting from ill–conditioned normal matrix sub–blocks) other normal matrix sub–blocks will be singular too as a result of the limitations. The occurrence of such *singularities* shows that certain unknown parameters are not estimable from the observations any more.

The planar gradiometer is a penalty resulting from the non drag free concept of Aristoteles. The along track disturbance accelerations due to air drag are too large to allow for very sensitive accelerometer axes in this direction.

Aristoteles is planned to fly for approximately six months in an orbit with inclination 95.°3. During additional two weeks the orbit will have an inclination of 92.°3. In both cases, *polar gaps* occur, i.e. polar areas which are not covered by ground tracks of the satellite. If a global recovery of potential coefficients is intended, problems might occur due to these gaps. Small deviations from a 90° inclination do not deteriorate the results significantly. Remarkably, the results slightly improve for a 92.°3 inclination. It is very likely that this phenomenon results from a higher data density in the remaining part of the earth. Furthermore, the behaviour of the inclination functions may play a role, as well as the fact that with a polar orbit, zonal coefficients are not estimable from gradients involving one cross–track derivative. For a 95.°3 inclination large polar gaps occur and the problem to derive a complete global set of potential coefficients from the data becomes improperly posed. The solution can be *stabilized* by adding prior information (e.g. prior expectations of the potential coefficients and their covariance matrix) and solve the system in the sense of least squares collocation. The solution becomes stable, but the estimates will be *biased.* However, investigations showed that, for large parts of the spectrum, a

stabilized solution does not lead to too optimistic error estimates.

More than the polar gaps, the *band limitation* of the Aristoteles gradiometer will be a problem. Only in a frequency band between 0.005 Hz and 0.125 Hz a 0.01 $E/\sqrt{Hz}$ error level can be achieved. The upper limit (resulting from the sampling rate of 4 seconds) will not cause any problems if the maximum degree and order of the analysis is 240. However, a lower limit (resulting from instrumental instability and non–gravitational orbital effects), has influence on all degrees in the solution. Whereas a frequency of 0.005 Hz would correspond to approximately 27 cycles per revolution (c.p.r.), we took an absolute lower band limit of 4 c.p.r. (to account for non–gravitational orbital effects). Furthermore, we investigated, apart from a flat error spectrum above 4 c.p.r., a $1/\beta$ *coloured noise* error behaviour between 4 and 27 c.p.r.

With a band limited gradiometer, certain potential coefficients are no longer estimable from the measurements due to a lack of information, resulting in singular normal matrix sub–blocks. Especially the lower part of the spectrum, where the contribution of the low orbital frequencies is relatively large, will cause problems. The requirements for an SGG mission are no longer met. Additional information is necessary. Given the present Aristoteles–gradiometry specifications, a combined SGG/SST mission seems to be the best alternative. For an SGG–only mission one has to use a gradiometer with either a much better precision (e.g. 0.0001 $E/\sqrt{Hz}$) or one which measures more components of the gradient tensor.

The results of the error analysis are presented in terms of error *degree variances*. The degree variance is the sum over all orders of a degree of the individual error estimates per coefficient. In some of the above situations, however, we do not obtain error estimates for all coefficients, viz. those cases where singular normal matrix sub–blocks appear. The degree variances computed in such cases thus do not represent the whole spectrum. Adding prior information in the sense of a stabilized solution overcomes this problem, but it is difficult to interpret the resulting "mixed" degree variances. In fact, the problem of singularities on certain degrees or orders is inherent to satellite methods. Degree variances are not the appropriate representation of the error situation in such cases. "Order variances" might be an alternative. Perhaps it gives more insight to plot all individual error estimates per coefficient in a perspectively plotted l, m–scheme, and find ways to identify the contribution of the prior information.

Whereas an error analysis learns us much about the influence of mission scenarios on the precision of the estimated unknowns, the final goal of a SGG mission is a set of estimated potential coefficients, together with the precision of all coefficients. The *recovery* of potential coefficients can be done in an iterative least squares adjustment scheme, in which at the same time the satellite's trajectory is determined. Depending on the way in which the data is handled there are, in the present set–up, two possible methods for the recovery. In the *space–like* method one transforms the data along the orbit into a global equi–angular grid of mean values. The observation equations are based on a series expansion of the gravitational potential in geocentric

polar coordinates. Equal step size in λ–direction and orthogonality of trigonometric series give rise to the block–diagonal structure of the normal matrix. The number of "observations" in the least squares estimation process is limited (it equals the number of blocks in the grid) so computation time is relatively short. The averaging process, required to obtain the gridded data, on the one hand acts as a smoothing operator so that high frequencies are lost, but on the other hand aliasing is reduced. A disadvantage of this method is the impossibility to include band limitation of the gradiometer in an easy manner.

In the *time–like* method the data does not need to be transformed into an equi–angular grid. A series expansion of the potential in orbital coordinates is used and with equal data step size along the orbit a block–diagonal normal matrix is obtained. Data is not averaged, so higher degrees are preserved, as a result of which however aliasing will play a larger role. Due to the very large number of observations the computation time will be rather large. Since possible data gaps may destroy orthogonality they have to be accounted for. With this method it is easy to include the effect of band limitation of the gradiometer. A test was done with both methods. We used a set of *simulated* gradients, computed along a circular, polar orbit for a 32 days mission. Only a first step of the iteration process was implemented, without updating the orbit. Potential coefficients were estimated up to degree and order 180 (approx. 33,000 unknowns). Although the test should be considered preliminary, the results look promising.

The requirements for an SGG mission are very high. Already a mission like Aristoteles puts high demands on the technology of the instrument and the spacecraft. Nevertheless, with ever improving technological developments gradiometers with much higher precision may become available, e.g. superconducting gradiometers. Also satellite orbit determination techniques are still improving. In order to fully benefit from such future perspectives it shall be necessary to improve the mathematical models as to include relativistic effects. Even better, gradiometry might become a technique such accurate, that aspects of the theory of relativity can be tested. Finally, from a theoretical (geometrical) point of view, gradiometry perfectly lends itself to be formulated in terms of curvature of spacetime.

We mention three aspects of an SGG mission where *relativistic theory* may play a role: 1.) the satellite orbit; 2.) the observed gradient tensor; and 3.) the moving local reference frame. The third one is important for the orientation of the instrumental frame with respect to which measurements are taken, but this aspect is not treated here. In the general theory of relativity a satellite can be considered a freely falling particle, moving along a four dimensional spacetime *geodesic*. Gravitation is not considered an external force, but is translated into the geometry of the space. Due to gravitation, four dimensional spacetime is *curved* and so are the geodesics. The equation of the geodesic therefore constitute the equations of motion of the satellite. Extracting the spatial part of these equations reveals the relativistic contribution to the satellite motion.

Viewing upon gradiometry as the relative movement of two or more proof masses,

we find the relativistic correspondence in the so–called *equation of geodesic deviation.* This equation is governed by the Riemann tensor, in which the second–order derivatives of the gravitational potential appear. Thus relativistic effects are also present in the observed gradients, and in this sense, a gradiometer measures the curvature of four dimensional spacetime.

In relativistic terms, the gravitational field of the earth can be considered a weak field. The equations of the satellite motion and the gradients are therefore derived in the so–called *weak field approximation.* In particular, we use the so–called first Post–Newtonian (1PN) approximation of the spacetime metric, keeping only the first–order relativistic contributions. The dominant relativistic orbit contribution is an effect in along track direction, which, however, in reality cannot be separated from a change in the value of the gravitational parameter GM. For a spherical, non–rotating earth and for a satellite like Aristoteles, other relativistic orbit effects are, at the moment, small enough to be neglected. Effects due to the rotation of the earth are even smaller. Furthermore, in the case of a spherical, non–rotating earth, relativistic contributions to the gradients at 200 km are of the order of 10^{-6} E. Although this is much smaller than the Aristoteles measurement precision, the question whether the relativistic effects in the gradients can be determined from such satellite missions is a difficult one, depending not only on the measurement precision but also on the mission duration and the principle character of the instrument.

appendix **A**

Coordinate systems

Several coordinate systems are used in this thesis, curvilinear as well as rectilinear, with different orientations and located with the origin in different points. Furthermore, the derivatives of the gravitational potential function with respect to most of these coordinate systems are needed at different stages during the derivations. The choice for a particular coordinate system often depends on the specific application, either because the coordinate system is particularly well suitable for the geometry of the problem, or because the practical or instrumental implementation of the problem prescribes the use of a specific coordinate system. In principle one could choose any coordinate system one likes, in practice only a few coordinate systems will have favourable characteristics for that specific problem. On the other hand, the introduction of a coordinate system is only artificial, it has in fact nothing to do with the physical reality of the problem itself. .

In this appendix we introduce some special, much used, coordinate systems. Some of them are applied in this thesis, others are just given for illustration. Referring to section B.3.1, we will only consider metrical spaces, in most cases also linear, i.e. Euclidean spaces $\mathbf{E}^n$, in particular for $n = 3$.

In the first section the coordinate systems will be defined, in most cases relative to a cartesian geocentric coordinate system, by giving the coordinate transformation equations between the geocentric cartesian and the new coordinates. Each coordinate system will be addressed by a special set of indices which indicate the particular system. Note that in appendix B the choice of indices was arbitrary, since at that stage no connection to some particular coordinate systems was given.

In the second section the metric tensor and the Christoffel symbols for some of these coordinate systems will be listed. They are needed when deriving expressions for the transformation of potential derivatives by using tensor analysis, as is done in section 3.1.

A.1 Definition

Possibly the most familiar coordinate system is the *cartesian* coordinate system. This is a rectilinear coordinate system of which the coordinate axes intersect in one point (the origin), are mutually orthogonal and along which the scale is the same in all directions. If the origin of such a cartesian coordinate system is in the geocenter, the coordinates are denoted with x^I, the indices coming from the set $\{I, J, K, L\}$. This system is called the *geocentric cartesian coordinate system.* As we will do for all coordinate systems, the separate coordinates are given special names (kernel letters), which, in this case, are

$$\{x^I | I = 1,2,3\} = (x^{I=1}, x^{I=2}, x^{I=3}) \equiv (X, Y, Z) .$$

The X–axis points to the Greenwich meridian, the Z–axis to the North–Pole and the Y–axis completes the set to a right handed coordinate system, see figure A.1.

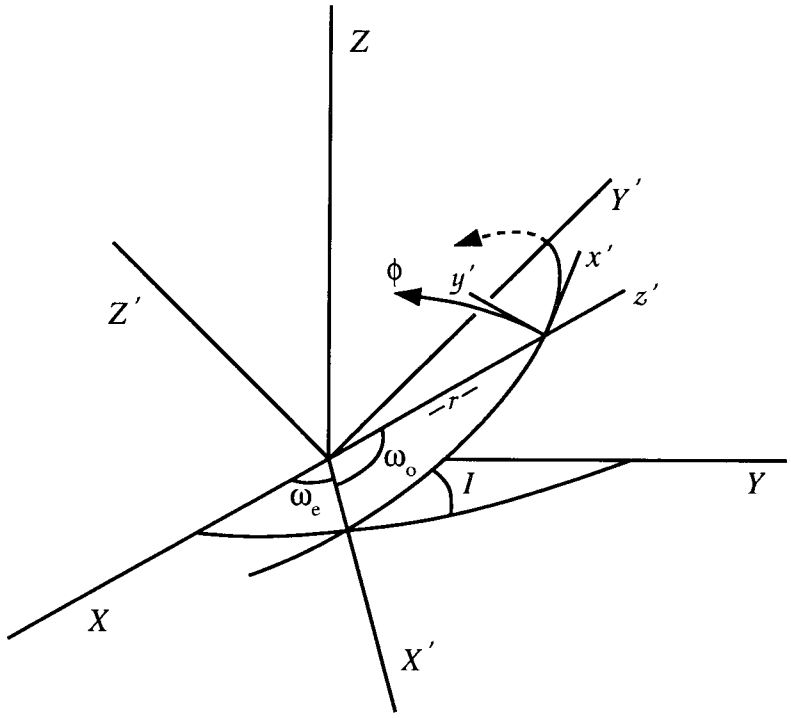


Figure A.1 *Geocentric cartesian and orbital coordinate systems.*

In cartesian coordinates, the metrical tensor takes on the diagonal form with all

diagonal elements having the value 1, so:

$$\{g_{IJ}\} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{A.1}$$

In this case we also have $\{g^{IJ}\} = \{g_{IJ}\}$. Also in every other cartesian system (with a different origin and/or a different orientation) the metrical tensor has this form. For example, if we rotate the x^I system in such a way that the XY–plane coincides with the orbital plane of a satellite in orbit around the Earth, we obtain a new geocentric cartesian coordinate system, denoted by $\{x^{I'} | I' = 1,2,3\} \equiv (X', Y', Z')$, which has the same origin but a different orientation, see figure A.1. The X'–axis of this system is directed towards the ascending node. This system is called the *orbital cartesian coordinate system.* The coordinate transformation between the two systems, $x^I = x^I(x^{I'})$, is as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \cos\omega_e & -\sin\omega_e \cos I & \sin\omega_e \sin I \\ \sin\omega_e & \cos\omega_e \cos I & -\cos\omega_e \sin I \\ 0 & \sin I & \cos I \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} \tag{A.2}$$

which is obtained through successive rotations about the angles ω_e and I. The matrix in this equation is the transformation matrix $\frac{\partial x^I}{\partial x^{I'}}$.

A third cartesian coordinate system we will frequently use is a local cartesian system, denoted $\{x^i | i = 1,2,3\} \equiv (x, y, z)$, which has its origin in a point on a satellite's orbit and which is oriented with the z–axis radially outwards, the x–axis directed along track and the y–axis cross track such that it is a right–handed system (see also figure A.1). This system is called the *local orbital coordinate system.* The transformation between the geocentric cartesian system and the local orbital system, $x^I = x^I(x^i)$, can be obtained through successive rotations about the angles ω_e, I and ω_o and a translation in radial direction by r. This will yield:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -\cos\omega_e \sin\omega_o - \sin\omega_e \cos\omega_o \cos I & \sin\omega_e \sin I & \cos\omega_e \cos\omega_o - \sin\omega_e \sin\omega_o \cos I \\ -\sin\omega_e \sin\omega_o + \cos\omega_e \cos\omega_o \cos I & -\cos\omega_e \sin I & \sin\omega_e \cos\omega_o + \cos\omega_e \sin\omega_o \cos I \\ \cos\omega_o \sin I & \cos I & \sin\omega_o \sin I \end{bmatrix} \begin{bmatrix} x \\ y \\ z + r \end{bmatrix} . \tag{A.3}$$

Another cartesian coordinate system to be used is a local north–oriented cartesian system, denoted by $\{x^{i'} | i' = 1,2,3\} \equiv (x', y', z')$, which has its origin in some terrestrial or space point and is oriented with the z'–axis radially outwards, the x'–axis directed north and the y'–axis directed west. This system is called *local north–oriented coordinate system.* The transformation $x^I = x^I(x^{i'})$ is obtained through

successive rotations about the angles θ and λ and a translation in radial direction (see figure A.2):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -\cos\lambda\cos\theta & \sin\lambda & \cos\lambda\sin\theta \\ -\sin\lambda\cos\theta & -\cos\lambda & \sin\lambda\sin\theta \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z'+r \end{bmatrix} . \qquad \text{(A.4)}$$

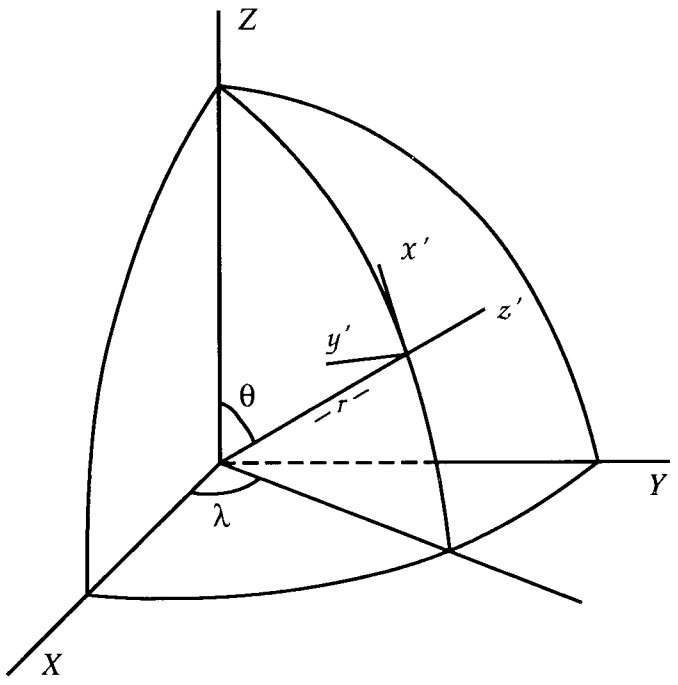


Figure A.2 *Geocentric cartesian, local north-oriented and geocentric spherical coordinate systems.*

Whereas in the transformation equations above the quantities $I, \omega_e, \omega_o, r, \theta, \lambda$ act as parameters to fix the orientation of the $x^{I'}, x^{i'}$ and x^i coordinate systems with respect to the x^I system, they themselves can also be used as coordinates. From the mentioned quantities we select the following coordinate sets:

$\{x^A | A = 1,2,3\} \equiv (r, \theta, \lambda)$ geocentric polar coordinates

$\{x^a | a = 1,2,3\} \equiv (r, \omega_o, \omega_e)$ I–orbital coordinates

$\{x^{a'} | a' = 1,2,3\} \equiv (r, \omega_o, I)$ ω_e–orbital coordinates .

The transformation equations $x^I = x^I(x^A)$ are given by:

$$\begin{aligned} X &= r\sin\theta\cos\lambda \\ Y &= r\sin\theta\sin\lambda \\ Z &= r\cos\theta\ . \end{aligned} \tag{A.5}$$

The x^a and $x^{a'}$ systems both have the same transformation equations:

$$\begin{aligned} X &= r(\cos\omega_e\cos\omega_o - \sin\omega_e\sin\omega_o\cos I) \\ Y &= r(\sin\omega_e\cos\omega_o + \cos\omega_e\sin\omega_o\cos I) \\ Z &= r\sin\omega_o\sin I \end{aligned} \tag{A.6}$$

where in the case of the x^a system ω_e is the third coordinate ($x^{a=3}$) and I acts as a parameter fixing the orientation of x^i and vice versa for the $x^{a'}$ system. Note that equation A.5 can be obtained from equation A.4 if we put $x' = y' = z' = 0$ and that equation A.6 can be obtained from equation A.3 by putting $x = y = z = 0$. Note furthermore that the x^a system is not an orthogonal coordinate system so we expect some off–diagonal components of the metrical tensor in these coordinates to be unequal to zero.

All of the above coordinate systems, except one, can be used to label all points of the 3–dimensional space under consideration, regardless of the arbitrary values of the possible parameters. The exception is the x^a system. This system has the inclination I as parameter. If $I \neq 90°$ parts of the space are not "covered" by the coordinate system. These parts are cones with the top in the geocenter and with the Z–axis as symmetry axis and with the top angle with respect to the Z–axis equal to I. However, this system will only be used to describe points along a satellite orbit with inclination I and for that purpose the system can be used very well.

A fourth curvilinear coordinate system which will be used is also a polar coordinate system but not relative to the geocentric cartesian coordinate system x^I (like x^A), but relative to the orbital cartesian system $x^{I'}$. It is called *orbital polar coordinate system* and is denoted by $\{x^{A'}|A' = 1,2,3\} \equiv (r,\phi,\omega_o)$. We have the following relations (see figure A.1):

$$\begin{aligned} X' &= r\cos\phi\cos\omega_o \\ Y' &= r\cos\phi\sin\omega_o \\ Z' &= r\sin\phi\ . \end{aligned} \tag{A.7}$$

The transformation $x^I = x^I(x^{A'})$ can be found by inserting equation A.7 into equation A.2.

The final two coordinate systems to be introduced are two ellipsoidal systems (see figure A.3):

$\{x^{\bar{A}}|\bar{A} = 1,2,3\} \equiv (h,\varphi,\lambda)$ geodetic coordinates

$\{x^{\bar{a}}|\bar{a} = 1,2,3\} \equiv (u,\beta,\lambda)$ ellipsoidal coordinates ,

of which the transformation equations to geocentric cartesian coordinates are:
$x^I = x^I(x^{\bar{A}})$:

$$\begin{aligned} X &= v\cos\beta\cos\lambda \\ Y &= v\cos\beta\sin\lambda \\ Z &= u\sin\beta \end{aligned} \tag{A.8}$$

$x^I = x^I(x^{\bar{a}})$:

$$\begin{aligned} X &= (N+h)\cos\varphi\cos\lambda \\ Y &= (N+h)\cos\varphi\sin\lambda \\ Z &= (N(1-e^2)+h)\sin\varphi\ , \end{aligned} \tag{A.9}$$

with $v = \sqrt{u^2+E^2}, E^2 = a^2 - b^2, N = a(1-e^2\sin^2\varphi)^{-1/2}, e^2 = (a^2-b^2)/a^2$, a is the semi–major axis of the ellipsoid and b the semi–minor axis. For problems expressed in ellipsoidal coordinates it is often convenient to have a local cartesian coordinate system directed along the normal to the ellipsoidal surface. This cartesian system is called *local ellipsoidal coordinate system* and it is denoted by $\{x^{\bar{i}} | \bar{i} = 1,2,3\} \equiv (\bar{x},\bar{y},\bar{z})$. The $\bar{z}$–axis is directed outwards, normal to the ellipsoid, the $\bar{x}$–axis directed north, tangent to the ellipsoidal surface and the $\bar{y}$–axis is directed west (see figure A.3). We will not actually use these ellipsoidal coordinate systems in this work, but they are given here just for reference.

In table A.1 all mentioned coordinate systems are listed.

A.2 Metric and Christoffel symbols

As already mentioned, the metrical tensor in all cartesian coordinate systems is the same, so:

$$\{g_{IJ}\} = \{g_{I'J'}\} = \{g_{ij}\} = \{g_{i'j'}\} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

The Christoffel symbols are therefore zero in all these coordinates. This is not true in curvilinear coordinates. From equations A.5 – A.9 we may compute the transformation matrices $\frac{\partial x^I}{\partial x^A}$, etc. These can be used in equation B.13 to compute the metrical tensor in the systems $x^A, x^a, x^{a'}, x^{A'}, x^{\bar{A}}$ and $x^{\bar{a}}$ from the one given above. For each of the coordinate systems the Christoffel symbols are then computed using equation B.15. In the sequel all the metrical tensors (table A.2) and the Christoffel symbols (tables A.3 to A.8) are listed.

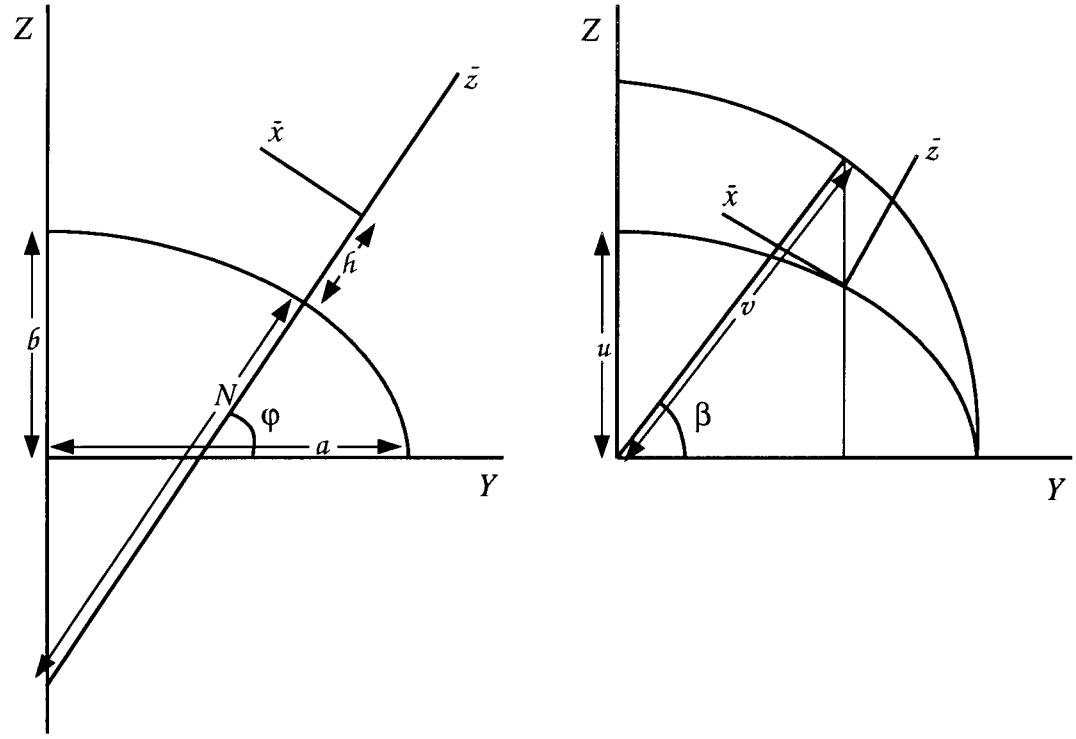


Figure A.3 *Ellipsoidal and geodetic coordinate systems.*

Table A.1 *List of coordinate systems and their notation*

x^I	(X, Y, Z)	geocentric cartesian coordinates
x^A	(r, θ, λ)	geocentric polar coordinates
$x^{I'}$	(X', Y', Z')	orbital cartesian coordinates
$x^{A'}$	(r, ϕ, ω_o)	orbital polar coordinates
$x^{i'}$	(x', y', z')	local north–oriented coordinates
x^i	(x, y, z)	local orbital coordinates
x^a	(r, ω_o, ω_e)	I–orbital coordinates
$x^{a'}$	(r, ω_o, I)	ω_e–orbital coordinates
$x^{\bar{A}}$	(h, φ, λ)	geodetic coordinates
$x^{\bar{a}}$	(u, β, λ)	ellipsoidal coordinates
$x^{\bar{i}}$	$(\bar{x}, \bar{y}, \bar{z})$	local ellipsoidal coordinates

Table A.2 *Elements of the metrical tensor for several coordinate systems.* $(M = (1-e^2)N^3/a^2$ *and* $L^2 = u^2 + E^2 \sin^2\beta)$

geocentric polar	1	0	0
(r, θ, λ)	0	r^2	0
	0	0	$r^2 \sin^2\theta$
orbital polar	1	0	0
(r, ϕ, ω_o)	0	r^2	0
	0	0	$r^2 \cos^2\phi$
I–orbital	1	0	0
(r, ω_o, ω_e)	0	r^2	$r^2 \cos I$
	0	$r^2 \cos I$	$-r^2(\sin^2 I \sin^2\omega_o - 1)$
ω_e–orbital	1	0	0
(r, ω_o, I)	0	r^2	0
	0	0	$r^2 \sin^2\omega_o$
geodetic	1	0	0
(h, φ, λ)	0	$(h+M)^2$	0
	0	0	$(h+N)^2 \cos^2\varphi$
ellipsoidal	$\frac{L^2}{v^2}$	0	0
(u, β, λ)	0	L^2	0
	0	0	$v^2 \cos^2\beta$

Table A.3 *Christoffel symbols in geocentric polar coordinates* (r, θ, λ)

Γ^1_{AB}	0	0	0
	0	$-r$	0
	0	0	$-r\sin^2\theta$
Γ^2_{AB}	0	$\frac{1}{r}$	0
	$\frac{1}{r}$	0	0
	0	0	$-\sin\theta\cos\theta$
Γ^3_{AB}	0	0	$\frac{1}{r}$
	0	0	$\cot\theta$
	$\frac{1}{r}$	$\cot\theta$	0

Table A.4 *Christoffel symbols in orbital polar coordinates* (r, ϕ, ω_o)

$\Gamma^1_{A'B'}$	0	0	0
	0	$-r$	0
	0	0	$-r\cos^2\phi$
$\Gamma^2_{A'B'}$	0	$\frac{1}{r}$	0
	$\frac{1}{r}$	0	0
	0	0	$-\sin\phi\cos\phi$
$\Gamma^3_{A'B'}$	0	0	$\frac{1}{r}$
	0	0	$\tan\phi$
	$\frac{1}{r}$	$\tan\phi$	0

Table A.5 *Christoffel symbols in I-orbital coordinates (r, ω_o, ω_e)*

Γ^1_{ab}	0	0	0
	0	$-r$	$-r\cos I$
	0	$-r\cos I$	$r(\sin^2\omega_o \sin^2 I - 1)$
Γ^2_{ab}	0	$\frac{1}{r}$	0
	$\frac{1}{r}$	0	$\cos I \tan\omega_o$
	0	$\cos I \tan\omega_o$	$-\tan\omega_o(\sin^2\omega_o \sin^2 I - 1)$
Γ^3_{ab}	0	0	$\frac{1}{r}$
	0	0	$-\tan\omega_o$
	$\frac{1}{r}$	$-\tan\omega_o$	$-\cos I \tan\omega_o$

Table A.6 *Christoffel symbols in ω_e-orbital coordinates (r, ω_o, I)*

$\Gamma^1_{a'b'}$	0	0	0
	0	$-r$	0
	0	0	$-r\sin^2\omega_o$
$\Gamma^2_{a'b'}$	0	$\frac{1}{r}$	0
	$\frac{1}{r}$	0	0
	0	0	$-\sin\omega_o\cos\omega_o$
$\Gamma^3_{a'b'}$	0	0	$\frac{1}{r}$
	0	0	$\cot\omega_o$
	$\frac{1}{r}$	$\cot\omega_o$	0

Table A.7 *Christoffel symbols in geodetic coordinates* (h, φ, λ)

$\Gamma^1_{\bar{A}\bar{B}}$	0	0	0
	0	$-(h+M)$	0
	0	0	$-(h+N)\cos^2\varphi$
$\Gamma^2_{\bar{A}\bar{B}}$	0	$\frac{1}{h+M}$	0
	$\frac{1}{h+M}$	$\frac{3e^2M^2}{N(h+M)(1-e^2)}\sin\varphi\cos\varphi$	0
	0	0	$\frac{h+N}{h+M}\sin\varphi\cos\varphi$
$\Gamma^3_{\bar{A}\bar{B}}$	0	0	$\frac{1}{h+N}$
	0	0	$-\frac{h+M}{h+N}\tan\varphi$
	$\frac{1}{h+N}$	$-\frac{h+M}{h+N}\tan\varphi$	0

Table A.8 *Christoffel symbols in ellipsoidal coordinates* (u, β, λ)

$\Gamma^1_{\bar{a}\bar{b}}$	$\frac{uE^2}{v^2L^2}\cos^2\beta$	$\frac{E^2}{L^2}\sin\beta\cos\beta$	0
	$\frac{E^2}{L^2}\sin\beta\cos\beta$	$-\frac{uv^2}{L^2}$	0
	0	0	$-\frac{uv^2}{L^2}\cos^2\beta$
$\Gamma^2_{\bar{a}\bar{b}}$	$-\frac{E^2}{v^2L^2}\sin\beta\cos\beta$	$\frac{u}{L^2}$	0
	$\frac{u}{L^2}$	$\frac{E^2}{L^2}\sin\beta\cos\beta$	0
	0	0	$\frac{v^2}{L^2}\sin\beta\cos\beta$
$\Gamma^3_{\bar{a}\bar{b}}$	0	0	$\frac{u}{v^2}$
	0	0	$-\tan\beta$
	$\frac{u}{v^2}$	$-\tan\beta$	0

appendix **B**

Index notation and tensor analysis

A large part of the time spent on the solution to problems is often devoted to, on the one hand, the language in which the problem and its solution are written (i.e. *notation*) and on the other hand the mathematical formalism or tools used for solving the problem. It is said that without the use of index notation and tensor analysis Einstein could not have made his general theory of relativity to become such a success, or even that this theory would never have been developed at all without its use. This may illustrate the influence which the choice of a proper notation and formalism may have on the theory and it therefore justifies the amount of effort put into the question of notation and mathematical formalism.

This appendix will therefore contain a short treatise on the two mentioned subjects: notation and mathematical tools. However, not all possibilities will be treated here. For the notation we choose here the (kernel–) index notation and as mathematical tool the tensor analysis. Of course there are other possibilities, which are perhaps more modern or more commonly used. There are, however, two reasons that we choose here for index notation and tensor analysis. At first index notation and tensor analysis are very elegant tools and easy to work with. Of course it requires, as always, some experience to work efficiently with them, but as soon as one is getting acquainted with it, one discovers that they are very suitable for almost all our problems and that they therefore create a single general framework in which all parts of the subject fit, that formulas become very short and easy to read and write and (last but not least) that these formulas can be very easy converted into programmable code for computational purposes. The second reason is that this thesis contains a chapter on relativistic aspects. At present index notation and tensor analysis are still very commonly used for relativistic purposes.

Nevertheless, the use of vectors and matrices and their corresponding notation is perhaps even more widespread. Also in this thesis it is used, since it is often convenient and it suffices our needs. However, sometimes people like to switch

between index notation and matrix notation. To this extent, the present appendix will also contain a section on the relation between them.

B.1 Index notation

The terms index notation and tensor analysis are often mentioned in one breath. One has to bear in mind, however, that they are two distinct topics. As already illustrated above, index notation is only a kind of language for writing something down. It can be used for all kinds of quantities, including the ones which are not tensors (Moon and Spencer, 1986). Tensor analysis is a kind of mathematical tool which can be used to solve certain problems. It can be expressed using index notation but also other notations may be used. The two topics will therefore be treated separately. In this section some fundamental concepts of index notation will be pointed out, in the next section some concepts of tensor analysis.

B.1.1 Kernel letters

In index notation an object (or entity) is a set (or system or array) of numbers which are in some sense related to each other. The relationship between the numbers may be based on geometrical or physical properties or may be a purely mathematical one. In the latter case we may also use the word holor (Moon and Spencer, 1986), in the former case the word *quantity* is often used. Since in this work almost all objects have some physical or geometrical background, we will use the word quantity throughout the work. The numbers may also be called *elements* or merates. Usually the numbers out of which a quantity is built up are taken relative to some coordinate system. Then they may be called coordinates or *components*. We will use the word components throughout this work. The meaning of the word component used in this sense must not be confused with the meaning of the same word customary in vector analysis, although the two may in special cases coincide (Moon and Spencer, 1986).

A quantity is represented by a letter. For example, the gravitational potential is denoted by V, the metric of a space by $\mathbf{g}$, the Christoffel symbols are Γ, a base vector is $\mathbf{e}$, coordinates are x, etc. As one can see, letters to indicate quantities may be roman or greek, small or capital, even boldfaced. It is not recommendable to add other symbols or characters, like bars, accents, numbers or tildes, to the kernel letter. This is to avoid confusion with indices and to keep formulas visually as clear as possible. One should, furthermore, never assume that the reader is familiar with what the writer thinks is a commonly accepted notation. It is better to state always explicitly the meaning of all kernel letters and the quantities they represent.

In the kernel–index notation of Schouten (Schouten, 1954) a quantity is always denoted by the same letter, independent of the coordinate system with respect to which the components of the quantity are expressed. In this case we call the letter the *kernel letter*. Not all authors using index notation follow this convention, so one often

sees that if the same quantity is expressed with respect to another coordinate system, another kernel letter is used, or that a bar or some other symbol is added to the kernel letter (e.g. $g \rightarrow \bar{g}$ or $V \rightarrow V'$). In the kernel–index notation a change of coordinates is indicated by a change of the type of indices (see next section). So the word "kernel letter" not only has the visual meaning of the central character representing a quantity but it also has the fundamental implication that the introduction of an arbitrary coordinate system is only artificial and does not change anything of the physical or geometrical nature of the quantity. The numerical value of the components of the quantity may change when transforming to another coordinate system, but the physical or geometrical characteristics remain, of course, the same.

Furthermore, we will try to leave the kernel letter intact as much as possible, also when performing other operations on it then coordinate transformations. So if the effect of any operation can be indicated by a change in the type, number or place of the indices, the kernel letter will remain the same. These kinds of operations are, for example, raising and lowering of indices, contraction, covariant differentiation, transpose.

B.1.2 Indices

A quantity which consists of more than one component (so it is not a scalar) is indicated by a kernel letter with one or more subscripts or superscripts, also called (lower or upper) indices, attached to it. For example, the components of the gravitational acceleration vector are indicated by V_A, the components of the metrical tensor by g_{ij}, the components of the Christoffel symbols by $\Gamma^{\alpha}_{\beta\gamma}$, etc. Indices may be roman or greek letters and small letters or capitals. The index is in fact a short hand notation for the separate components, for example in 3–dimensional space we have

$$\{V_A | A = 1,2,3\} = (V_{A=1}, V_{A=2}, V_{A=3}) .$$

If this is simply abbreviated by V_A we have to state explicitly the range (dimension) and the meaning of the index A, for example: the index A represents polar coordinates in 3–dimensional space:

$$\{x^A | A = 1,2,3\} = (x^{A=1}, x^{A=2}, x^{A=3}) ,$$

the index i represents cartesian coordinates

$$\{x^i | i = 1,2,3\} = (x^{i=1}, x^{i=2}, x^{i=3}) .$$

It is often convenient to give the various components special symbols like

$$\{x^A | A = 1,2,3\} \equiv (r, \theta, \lambda)$$

and

$$\{x^i | i = 1,2,3\} \equiv (x, y, z) ,$$

so that we also have

$$\{V_A | A = 1,2,3\} \equiv (V_r, V_\theta, V_\lambda)$$

and

$$\{V_i | i = 1,2,3\} \equiv (V_x, V_y, V_z) \ .$$

The latter notation is of course not according to the conventions in the last paragraph but in most cases this kind of notation will not lead to any confusion and the meaning of the symbols will always have to be evident from the context. The following notation, however, is ambiguous:

$$(V_1, V_2, V_3)$$

because this notation does not show with respect to which coordinate system the components are given.

From now on we will adopt a much used convention which is to confuse a quantity with its components (Foster and Nightingale, 1979). For example, we refer to a quantity V_A rather than a quantity $\mathbf{V}$ with components V_A. Note that this convention becomes important if we follow the rule given in the last section to leave the kernel letter intact as much as possible, even under operations changing the number of indices such as contraction or covariant differentiation. When referring explicitly to the quantity and not to its components, the kernel letter is often printed boldface, like in the example above: quantity $\mathbf{V}$ with components V_A.

So the type of the indices indicates the coordinate system. In the kernel–index notation we address to each coordinate system a special (limited) set of successive letters from the alphabet to be used as indices. For example, the set of indices $\{i, j, k, l\}$ can be used for a local cartesian coordinate system, $\{A, B, C, D\}$ for curvilinear polar geocentric coordinates, etc. We allow for indices the use of other symbols attached to the letters, like $\{i', j', k', l'\}, \{\bar{A}, \bar{B}, \bar{C}, \bar{D}\}$. In combination with the kernel letter we then have for example V_A the components of the gravitational acceleration vector with respect to the polar coordinates x^A, V_i the same quantity but expressed with respect to local cartesian coordinates x^i, etc.

As one can see, indices may be placed high or low. Upper indices represent contravariant components and lower indices represent covariant components, at least if we are dealing with tensors. In section B.2.2 the meaning of these terms will be discussed.

Any index appearing twice (one upper and one lower index) in a quantity or in a product (in general in a term of an expression) is, according to Einstein's summation convention, summed over all values the index can take (i.e. over its range or dimension), for example

$$R^\alpha{}_\beta \, T^\beta = \sum_{\beta=1}^{n} R^\alpha{}_\beta \, T^\beta \ ,$$

where n is the dimension. Such indices are called *dummy* or summation indices because after evaluating the expression they are cancelled. Summing over equal

indices in a quantity or in a product is also called *contraction*. Indices which are not dummy indices are called *free* indices. After summation over dummy indices in an equation the remaining (free) indices on the left hand side and on the right hand side should be the same and in the same place (upper or lower), for example

$$a^i{}_j\, b^j = c^i$$

$$a_j\, b^j{}_k + c_k = d_k$$

are correct expressions, but

$$a^i{}_j\, b^k = c_j$$

is not. We will call this rule here *(weak) index balance.*

Since dummy indices are summed over, we may always replace them by other letters, as long as they belong to the same coordinate system and as long as they are not already in use as free indices.

The number of indices attached to a kernel letter indicates the rank (or valence or order) of the quantity. If a quantity is of second rank it is represented by a kernel letter with 2 indices, for example a_{ij}. If the i–index is raised we could write the result as a^i_j, $a^i{}_j$ or $a^{\cdot i}_{j}$. When working only with index notation, the first form is always clear. The indices may always be placed as near as possible to the kernel letter. If one likes to convert certain equations at the end of a derivation to matrix notation, the second (or third) form is to be preferred. In this case not only the place of an index upper or lower matters but also the place right or left: when raising or lowering an index it must keep its place in horizontal direction relative to the kernel letter (see next section). It may help to place a dot under each upper index if other indices follow (third form), but this makes the expression visually unattractive.

B.1.3 Matrices

Despite the fact that the language of index notation is sufficient for writing down all kinds of problems, many people still like to work with matrices because matrices and matrix notation (including vector notation, sometimes called symbolic or abstract notation) are found to be more illustrative than index notation. Now the conversion between the two is not always clear. Problems occur with the ordering of the quantities in a term, with the transpose and with the inverse of a matrix. In this section we will discuss these problems. Furthermore, additional conventions will be given, which, if strictly followed, will prevent these problems to occur.

One of the most important differences between the two languages is that in matrix notation only quantities up to rank 2 can be expressed adequately, whereas in index notation there is no limitation on the rank of quantities. So in this section we will restrict ourselves to scalars (rank 0), vectors (rank 1) and matrices (rank 2).

In index notation the ordering of the quantities in a term is of no importance:

$$a^i{}_j b^j{}_k$$

is the same as

$$b^j{}_k a^i{}_j$$

(i.e. multiplication is commutative). But in matrix notation it is known that $AB \neq BA$ (where A and B are matrices). So if we derive in index notation an expression and we like to convert it later to matrix notation, we have to find a way to discover the ordering of the quantities in a term. Furthermore we have to find out if we are dealing with the original quantity, its transpose or its inverse. In index notation the latter two play no explicit role. An expression can always be evaluated correctly if one strictly follows the rules for dummy and free indices, without having to know anything about the transpose or the inverse. It is not even necessary to define these forms (although this is possible). Since in matrix notation the two are very important, we have to find a way to see whether the quantity is a transpose, inverse or the original one.

The problems mentioned can be dealt with by following some additional conventions than the ones mentioned already earlier.

1. First of all we connect to the first index (when reading from left to right) the rows of the matrix and to the second index the columns, for example in A_{ij}, i indicates rows and j columns. Also x^i is a one column ("standing") vector, the index i indicating the rows. A "lying" vector has, in this sense, to be written as $x^{\cdot i}$, i denoting the columns now, the dot (in the place of the rows) being added for clearness. This implies that indices should always maintain their place in horizontal direction. This is especially important during the processes of raising and lowering the indices. A quantity with one upper and one lower index belonging to the same set, should therefore always be written with shifted indices: $A_i{}^j$ and not A^j_i. In an equation the free indices on the left and right hand side should now also be in the same place left or right, not only upper and lower. For example

   $$x^\alpha = R_i{}^\alpha x^i$$

   should be written as

   $$x^\alpha = R^\alpha{}_i x^i \ .$$

   We call this *strong index balance*. In terms of a matrix product this is easy to understand: x^i and x^α are "standing" vectors (the indices indicating rows) of which the lengths have to agree with respectively the number of columns and the number of rows of the matrix R. This also means that in a product equal indices always appear twice: once as row and once as column index.

2. Secondly, when reading the indices from left to right, the alphabetical order indicates which form of the quantity we have: alphabetical order indicates the original form and reverse order the transpose, for example A_{ij} is the original

quantity and A_{ji} the transpose. Also: $b_j{}^k$ original, $b^k{}_j$ transpose or $d^j{}_k$ original, $d_k{}^j$ transpose. This rule only applies if the indices belong to the same coordinate system.

In the case of coordinate transformation matrices we are dealing with quantities of which the indices belong to different coordinate systems, for example $R^i{}_\alpha$. In that case we have to look into the definition of R. If R (or any other kernel letter used for the coordinate transformation) is defined (or introduced) by means of an equation like

$$x^i = R^i{}_\alpha x^\alpha$$

the normal order of the indices is $\{i, \alpha\}$ and if it is defined using

$$x^\alpha = R^\alpha{}_i x^i$$

the order is $\{\alpha, i\}$. So the indices indicating the "new" coordinates (left hand side of the transformation equation) come first.

Strictly speaking the transpose should be expressed by a different kernel letter. But since the two are very strongly related (they are built up out of the same (numerical) values only in a different ordering), since the difference between the transpose and the original quantity can be shown without ambiguity by means of the indices and since it is commonly accepted, we will use the same kernel letter for the original quantity and the transpose. This is also common in matrix notation.

3. In matrix notation the inverse A^{-1} of a matrix A is defined as $A^{-1}A = I$ where I is the unit matrix. In index notation the inverse a_{rs} of a quantity A^{rs} is defined through

$$A^{rs}a_{st} = \delta^r{}_t \ .$$

where $\delta^r{}_t$ is the *Kronecker delta* whose value is 1 if $r = t$ and zero if $r \neq t$. We see another kernel letter has to be used for the inverse, since the inverse is in fact another quantity. The definition equation relates the inverse with the original quantity. We try to pronounce this relationship a little bit more by using for the original quantity and the inverse the capital and small version of the same letter (or vice versa) but this is not strictly necessary. Using a different kernel letter becomes extremely important if we are dealing with mixed quantities (quantities having both upper and lower indices) like $R^i{}_j$. Its inverse $r^i{}_j$ would be defined by $R^i{}_j r^j{}_k = \delta^i{}_k$. If we would have used the same kernel letter here the notation would have been very ambiguously: $R^i{}_j R^j{}_k = \delta^i{}_k$. One has to customize oneself to take for the inverse always a different kernel letter and state explicitly the definition equation. So if $R^i{}_j$ is in matrix notation R and $r^i{}_j$ is its inverse R^{-1} we have the following four possibilities:

	index notation	→ matrix notation
original quantity	$R^i{}_j$	$\rightarrow R$
transpose	$R_j{}^i$	$\rightarrow R^T$
inverse	$r^i{}_j$	$\rightarrow R^{-1}$
transpose of inverse	$r_j{}^i$	$\rightarrow R^{-T}$.

There are however two situations in which we may use the same kernel letter for the inverse and the original quantity. That is if the quantity has either two upper or two lower indices or if the indices do not belong to the same coordinate system:

original quantity	A^{ij} , B_{ij} , $C^i{}_a$	$\rightarrow A$, B , C
transpose	A^{ji} , B_{ji} , $C_a{}^i$	$\rightarrow A^T$, B^T , C^T
inverse	A_{ij} , B^{ij} , $C^a{}_j$	$\rightarrow A^{-1}$, B^{-1} , C^{-1}
transpose of inverse	A_{ji} , B^{ji} , $C_j{}^a$	$\rightarrow A^{-T}$, B^{-T} , C^{-T} .

It is repeated here once again that for mixed quantities the latter does not apply so a different kernel letter *has* to be used.

4. The last new convention concerns the ordering of the quantities in a product. From convention 1 it follows that this ordering is such that equal indices (dummy indices) are as close as possible together. So if a derivation in index notation leads to a product like

$$a^{st}b_{rs}$$

it has to be written as

$$b_{rs}a^{st} ,$$

the dummy index s appearing in both terms close together. In the same way: $V_{ab}R_i{}^aR^b{}_j$ has to be written as $R_i{}^aV_{ab}R^b{}_j$ and $g_{ij}x^iy^j$ as $x^ig_{ij}y^j$.

Example

To illustrate the use and consequences of the new conventions we show an example. Imagine two coordinate systems of which the base vectors are respectively $\mathbf{e_i}$ and $\mathbf{e_a}$. We assume the dimensions of **i** and **a** are equal. The components of a vector $\mathbf{x}$ in the two systems are respectively x^i and x^a:

$$\mathbf{x} = x^i\mathbf{e_i} = x^a\mathbf{e_a} .$$

Suppose the two systems are connected by a transformation:

$$\mathbf{e_i} = R_i{}^a \mathbf{e_a} \ .$$

Then we have

$$\mathbf{x} = x^i \mathbf{e_i} = x^i R_i{}^a \mathbf{e_a} = x^a \mathbf{e_a} \ .$$

Inspection of the last equation shows that the transformation of the components is

$$x^a = R^a{}_i x^i \ ,$$

where we have interchanged the indices i and a of R based on convention 1 and reversed the order of x and R based on convention 4. The inverse of the last equation is

$$x^j = r^j{}_a x^a$$

with

$$R_i{}^a r_a{}^j = \delta_i{}^j \ .$$

So if, in matrix notation, we substitute R for $R_i{}^a$ we see that if the base vectors transform with R, the coordinates transform with R^{-T}, based on convention 3. If a vector v_a transforms like $v_i = A_i{}^a v_a$, a matrix V_{ab} will transform like

$$V_{ij} = A_i{}^a V_{ab} A^b{}_j$$

(see section B.2.1). In matrix notation this becomes (if we substitute $V_{ij} \rightarrow \Gamma$, $V_{ab} \rightarrow V$):

$$\Gamma = A \, V \, A^T \ ,$$

a familiar result.

There remains one thing to be explained. Sometimes we have the situation where the numerical values of two sets of components (of the same rank and dimension) are the same, but the two sets are not represented by kernel letters with the same indices in the same places. According to the new conventions above we may only equate quantities if free indices are of the same type and in the same place (upper and lower as well as right and left). For example, in "ordinary" index notation, the symmetry of a matrix is expressed as $a^{ij} = a^{ji}$. This would not be correct based on the strong index balance. Another situation where this occurs is with orthogonal rotation matrices, where the transpose equals the inverse. Expressing this as $R^j_i = r_i{}^j$ is not correct.

These kinds of equalities will be expressed with brackets[1]:

$$\{a^{ij}\} = \{a^{ji}\}$$

[1]This notation is chosen following (Moon and Spencer, 1986). In this book a very beautiful and extensive treatment on index notation and tensor theory is given, which, however, does not fully coincide with the conventions presented here.

and

$$\{R^j{}_i\} = \{r_i{}^j\} .$$

The number and type of the indices on the left and right hand side are still the same, but the place may be different. This indicates numerical equality for all values the indices can assume, so $\{a^{ij}\} = \{a^{ji}\}$ means $a^{11} = a^{11}, a^{12} = a^{21}, a^{13} = a^{31}, a^{23} = a^{32}$, etc.

Example

Suppose that, in the previous example, the matrix V is symmetric: $\{V_{ab}\} = \{V_{ba}\}$. Now let us derive the transpose of Γ, which is in index notation V_{ji}:

$$\begin{aligned}
\{V_{ji}\} = \{R_j{}^a V_{ab} R^b{}_i\} &= \{R_j{}^b V_{ba} R^a{}_i\} && \text{(dummy indices)} \\
&= \{R_j{}^b V_{ab} R^a{}_i\} && \text{(symmetry of } V\text{)} \\
&= \{R^a{}_i V_{ab} R_j{}^b\} && \text{(convention 4)} \\
&= \{R_i{}^a V_{ab} R^b{}_j\} && \text{(convention 1)} \\
&= \{V_{ij}\} && \text{(expression for } V_{ij}\text{)} ,
\end{aligned}$$

so we see Γ is also symmetric.

As already said, the additional conventions given in this section are only needed if one likes to convert from index notation to matrix notation. If this is not the case (what is to be recommended) "ordinary" index notation can be used (Moon and Spencer, 1986).

B.2 Tensor analysis

We use tensor analysis in some chapters of this work because it is a suitable mathematical tool, as stated in section B.1. Another reason we use it, follows from the fact that this work deals with some aspects of gravitation. And since Einstein we know that gravitation is closely connected to geometry, in particular the geometry of curved spaces. Especially in the latter case, the tensor analysis, as it is based on the absolute differential calculus of Ricci and Levi–Civita, is found to be a very suitable mathematical concept.

In the same sense as index notation and tensor analysis are two different topics, though strongly connected in practice, also geometry and tensor analysis are not the same. If we, in this work, talk about geometry, we are dealing with metrical spaces, curvature, surfaces, metric, curves, etc., all in a more or less concrete sense. The geometrical objects are the objects of study themselves. If we look upon geometry in this way, tensor analysis is again (only) a mathematical tool to help us solve our geometrical problems. The two topics, being distinct, are therefore treated

here separately. First we deal, in the present section, with tensor analysis. We discuss the tensor concept, the terms contravariant and covariant and the process of differentiation. In the next section, some geometrical aspects are discussed: spaces, metric and curvature.

The division of the topics, treated in this appendix, into three parts (i.e. index notation, tensor analysis and geometry) based on the considerations given above and in the last section, may seem artificial to some readers, but it is just a way of ordering the broad range of topics related to the theory and it suits the present work.

B.2.1 Tensor

A quantity is defined as a set of elements which are in some sense mutual connected. This connection depends in its turn on the definition of the quantity and it will in most cases be a physical or a geometrical one. Such a relationship between the elements of a quantity is one of the characteristics of *tensors*. But a more important aspect of a tensor is that its elements are the components of the quantity with respect to some coordinate system. This does not mean that all sets of components are tensors. To be a tensor, the set needs to have some additional, special properties, which will be explained in this section. The important point here is that there is a distinction between on the one hand the quantity itself, which has some physical meaning, and on the other hand its representation by a set of numbers with which we can carry out computations. These numbers *may* be the components of the quantity relative to a coordinate system and this *has* to be so as one of the conditions for the set to be a tensor. One can imagine that the value of the numbers changes if another coordinate system is chosen to represent the same quantity. Consider for example the components of a displacement vector in $\mathbf{E}^3$ in two different cartesian coordinate systems having the same origin but a different orientation. Since the introduction of a coordinate system is only artificial and has nothing to do with the quantity itself, we like to have a mathematical framework underlying our computations which is independent of this arbitrary choice of coordinates. This is exactly the essential characteristic of tensor analysis and it is captured in the word "absolute" in the absolute differential calculus of Ricci and Levi-Civita: "The tensor calculus is said to be absolute because it is independent of the details of the choice of coordinates, that is, the equations have the same form in all coordinate frames" (Ohanian, 1976, p.221).

To see what this means mathematically we discuss the transformation between two coordinate systems. Let us denote the "old" coordinate system with x^a and the "new" with x^i and let the two be connected by the relationship

$$x^i = x^i(x^a) \,, \tag{B.1}$$

expressing the fact that the new coordinates are some functions of the old. The coordinate transformation B.1 may be chosen at will, but for convenience we will

impose some limitations on it, namely that the transformation function is single–valued (which means that there is a one–to–one correspondence between x^i and x^a), that it is analytic (i.e. its derivatives up to a sufficient order exist and are continuous) and the Jacobian of the transformation is different from zero in every point, cf. (Moon and Spencer, 1986). In this case also the inverse of the transformation exists:

$$x^a = x^a(x^i) \; .$$

For the rest the transformation may be completely arbitrary, non–linear as well as linear. Manipulations of equation B.1 tend to be complicated if the transformation is non–linear. But if we are willing to deal with infinitesimal coordinate changes (which is the case in tensor *analysis*), we may linearize the transformation B.1.

Differential changes in the coordinates are then transformed as

$$dx^i = \frac{\partial x^i}{\partial x^a} dx^a$$

or inverse

$$dx^a = \frac{\partial x^a}{\partial x^i} dx^i \; ,$$

where the partial derivatives $\frac{\partial x^i}{\partial x^a}$ are computed from B.1. These partial derivatives are in general functions of the coordinates. Only in the case of linear coordinate transformations they are constants.

Now we *define* a contravariant tensor (under general coordinate transformations) as a quantity which transforms according to

$$A^{ij\cdots p} = \frac{\partial x^i}{\partial x^\alpha} \frac{\partial x^j}{\partial x^\beta} \cdots \frac{\partial x^p}{\partial x^\vartheta} A^{\alpha\beta\cdots\vartheta} \; , \tag{B.2}$$

a covariant tensor as a quantity which transforms according to

$$A_{ij\cdots p} = \frac{\partial x^\alpha}{\partial x^i} \frac{\partial x^\beta}{\partial x^j} \cdots \frac{\partial x^\vartheta}{\partial x^p} A_{\alpha\beta\cdots\vartheta} \tag{B.3}$$

and a mixed tensor as a quantity transforming according to

$$A^{i\cdots l}{}_{m\cdots p} = \frac{\partial x^i}{\partial x^\alpha} \cdots \frac{\partial x^l}{\partial x^\delta} \frac{\partial x^\epsilon}{\partial x^m} \cdots \frac{\partial x^\vartheta}{\partial x^p} A^{\alpha\cdots\delta}{}_{\epsilon\cdots\vartheta} \; , \tag{B.4}$$

where the partial derivatives $\frac{\partial x^i}{\partial x^\alpha}$ and $\frac{\partial x^\alpha}{\partial x^i}$ (defined above) are the inverses of one another according to

$$\frac{\partial x^i}{\partial x^\alpha} \frac{\partial x^\alpha}{\partial x^j} = \delta^i_j \; .$$

Tensors of rank 0 are called scalars, tensors of rank 1 are called vectors. We will in general not call tensors of rank 2 matrices, because not each set of numbers which can be written as a system with rows and columns (a matrix) has to be a tensor. A

quantity is a tensor only if it fulfils one of the transformation laws B.2, B.3 or B.4. A tensor of rank 2, however, can always be represented as a matrix, but that does not make the two identical. As a concession to matrix notation we will, in the sequel, refer to $\frac{\partial x^i}{\partial x^a}$ as the transformation matrix and to $\frac{\partial x^a}{\partial x^i}$ as the inverse transformation matrix, thereby keeping in mind the remark about second–rank tensors made above.

We see that the tensor character of a quantity is only of interest if we are dealing with different coordinate systems or with coordinate transformations. This does not always necessarily have to be the case. But *if* different coordinate systems are involved and a quantity is a tensor, its representation in each coordinate system is the same except for the type of the indices, for example A^i versus A^α, B_{abc} versus $B_{i'j'k'}$, etc. Transforming from one coordinate system to another may never cause the tensor representation to change from, for example, A^i to a_α, A^i to A_{ab} or A^i to $A^a + B^a$, etc. This means that, if an equation consists only of tensors, its form is the same in any coordinate system, except for the type of the indices. This fact is one of the most important and most powerful concepts of tensor analysis. We have for example

$$v^i{}_j = w^i{}_k z^k{}_j + y^i x_{.j}$$

$$S = g_{ij} x^i y^j$$

in one coordinate system, which is written as

$$v^a{}_b = w^a{}_c z^c{}_b + y^a x_{.b}$$

$$S = g_{ab} x^a y^b$$

in another. But also if all components of a tensor A_{ij} are zero in one coordinate system, i.e. $A_{ij} = 0$ (or $A_{ij} = 0_{ij}$, the right hand side being the "null tensor") then they are zero in any other coordinate system.

In the examples above, we see that tensors may be added or multiplied with other tensors. Addition of tensors in only possible with tensors of the same rank and the same type (remember: dummy indices are to be considered cancelled after evaluation). The result is again an tensor of the same rank and the same type. Multiplication of tensors will result in a new tensor of which (*a*) the rank is the sum of the rank of the multiplied tensors minus the number of dummy indices, and (*b*) the type of the indices is the type of the remaining free indices in the product. It can be shown (cf. (Hotine, 1969), (Moon and Spencer, 1986)) that these operations indeed result in new tensors. Reversely, each tensor can be written as the product or sum of other tensors, with or without contraction, as long as they obey the rules above.

B.2.2 Co– and contravariant

In the previous section covariant and contravariant tensors were defined as quantities of which the components transform in a certain manner. This way of introducing

tensors is based on the classical approach to tensor algebra and tensor analysis (cf. (Ohanian, 1976), (Sokolnikoff, 1951), (McConnell, 1957), (Hotine, 1969) and many other textbooks). This classical way of introducing tensors is a geometrical one and it in fact goes back to Euclides. He, for the first time, gave a meaning to the concept of a *geometrical* space when trying to describe the world (or space) surrounding us. Later the connection to some specific geometrical structure was abandoned and a space was considered just a set of points with some arbitrary structure. Whereas the space considered by Euclides was a linear, 3–dimensional, geometrical space (Euclidean space or $\mathbf{E}^3$), a space in a purely mathematical sense may be of any dimension, linear or non–linear and with or without any special geometrical structure.

Later, when Descartes suggested to project the collection of real–numbers onto a straight line, the concept of coordinates was introduced into the space. The elements, out of which a space was built up (i.e. points) could now be represented by a set of numbers ($\{x^i | i = 1, \ldots, n\}$,$n$ being the dimension of the space). The points are said to be labelled by n real coordinates. If the structure of the space is reduced to zero, we are merely considering a set of labelled points and the space is said to be *arithmetic*, denoted by $\mathbf{X}^n$. Whereas Descartes used linear coordinates to label the points of $\mathbf{E}^3$, in general also non–linear coordinates may be introduced. A linear space (a space in which linear coordinates can be introduced) without metrical structure (see section B.3.1) is called an *affine* space: $\mathbf{A}^n$. In such a space, various coordinate systems are connected through linear (or affine) coordinate transformations. A next step is the introduction of an orthonormal coordinate system into an affine space (i.e. a system of which the coordinate axes have the same origin, are mutual orthogonal (independent) and have the same scale). In doing so, we endow the space with a specific metric and the space is called Euclidean.

As already mentioned before, the tensor character is in fact only of importance if different coordinate systems and the transformations between them are concerned. A tensor is a quantity which behaves as an *invariant* under such coordinate transformations. In this sense, a tensor is an extension of a *vector*, a concept originally coming from physics. Gibbs introduced a geometrical visualization of a vector, namely an arrow, characterized by its direction and its length. A point, having a fixed location, is the most elementary (geometrical) invariant under coordinate transformations. A vector, fixed by means of its begin and end points, is also an invariant. Of course the numerical values of the components of the vector may change if we carry out a coordinate transformation, but the vector itself, having some definite physical or geometrical meaning, remains unchanged. Its components express, however, the connection between the invariant vector and the arbitrarily introduced coordinate system.

The components of a vector $\boldsymbol{\lambda}$ are the scalars λ^a relative to a basis $\{\mathbf{e_a} | a = 1, \ldots, n\}$ such that $\boldsymbol{\lambda} = \lambda^a \mathbf{e_a}$. The set of vectors $\mathbf{e_a}$ can be considered a basis of the space if it spans the space (i.e. every vector in the space may be written as a linear combination of the members of the set $\mathbf{e_a}$) and if its members are linearly

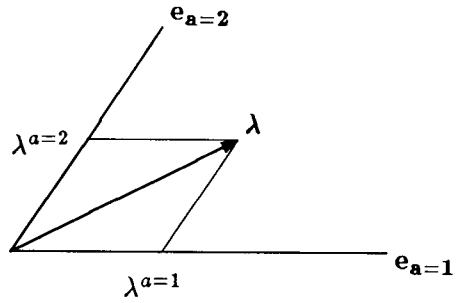


Figure B.1 *Contravariant components.*

independent. The base vectors may also be visualized by arrows, however of unit length, lying along the coordinate axes, pointing in the positive coordinate direction and with the begin points in the origin of the coordinate system. In a linear space (affine space) the components of an arbitrary vector with begin point not in the origin of the coordinate system, may be obtained by parallel transport of the basis from the origin to the begin point of the vector. If we now consider a vector field, defined throughout the whole space, such transport of the basis will result in a basis–field, i.e. in each point of the space we have a set of base vectors.

If we use curvilinear coordinates in a linear space, or if the space is not linear itself, we generalize the concept of a basis to the set of tangent vectors to the coordinate curves in each point. The tangent vectors now span in each point the (local) tangent space. The basis now depends on the coordinates, just as the vectors of a vector field may depend on the coordinates. (In tensor *algebra* we deal with problems either restricted to the local tangent space in a point or concerning the whole space in which case the space has to be linear (so that all tangent spaces coincide with it). Tensor *analysis* studies problems either dealing with the connection of different tangent spaces in a non–linear space or dealing with curvilinear coordinates in linear space, in both cases forcing us to linearize and to consider infinitesimal coordinate changes.)

The components of a vector relative to a basis in the sense described above (i.e. in general the tangent vectors to the coordinate curves in a point) are now called *contravariant components*, denoted with upper indices as in $\boldsymbol{\lambda} = \lambda^a \mathbf{e_a}$. This definition is, at this point, completely arbitrary, but we have to make a start somewhere. The geometrical interpretation of these contravariant components is probably well known to everybody, at least in Euclidean 2 or 3–dimensional space. There, those components are the parallel projections of the end point of the vector onto the coordinate lines, see figure B.1. In the case of non–linear coordinates the picture is essentially the same being only differential now.

Sets of *coordinate lines* in an n–dimensional space, along each of which one particular coordinate varies and the other $n-1$ coordinates are constant, are not the only way in which a coordinate system reflects the geometrical structure of a space.

A coordinate system provides us also with a dual structure of sets of *coordinates planes* on each of which one particular coordinate is constant and the other $n-1$ coordinates vary. In the former structure (sets of coordinate lines) we may find the base vector in the x^b–direction by taking the derivatives of all the coordinates $x^a, a = 1, \ldots, n$ with respect to x^b:

$$\frac{\partial x^a}{\partial \{x^b\}}$$

for $a = 1, \ldots, n$ and fixed b. This gives a unit vector (tangent vector) in x^b–direction. The complete basis is found by repeating this for all values b can take (i.e. $b = 1, \ldots, n$).

In the case of sets of coordinate planes (dual structure) we take the derivative of one particular x^a with respect to all coordinates $x^b, b = 1, \ldots, n$:

$$\frac{\partial \{x^a\}}{\partial x^b}$$

for $b = 1, \ldots, n$ and fixed a. This also gives a unit vector which is to be regarded as the gradient vector of the scalar field x^a. Repeating this for all values of a we obtain the so–called *dual basis* $\mathbf{e^a}$. Gradient vectors of arbitrary scalar fields $\varphi = \varphi(x^a)$ may now be written with their components relative to this dual basis. So if the vector $\boldsymbol{\mu}$ is the gradient of φ then $\boldsymbol{\mu} = \mu_a \mathbf{e^a}$. The components with respect to the dual basis are called *covariant components* and they are written with lower indices[2]. In the same sense as the gradient of a scalar field reflects the rate of change of it in a certain direction (i.e. the density of the $\varphi = constant$ surfaces) the covariant components are related to the density of the coordinate planes in the direction perpendicular to those planes. Only in an Euclidean space (with a metric defined) we may visualize these components as the orthogonal projections of the end point of a vector onto the coordinate lines, see figure B.2 (Hotine, 1969). Note that in this figure the vectors are not of the same type as in figure B.1, which fact we tried to indicate by using dashed lines in figure B.2. In other spaces no satisfactory visualization is possible.

So we have at this moment two types of vectors. At first we have ordinary *vectors* (contravariant vectors) with (contravariant) components relative to the basis. Examples of this type of vectors are displacement vectors, velocity vectors, etc. These are the vectors everybody is customary to work with. Secondly we have *covectors* with covariant components relative to the dual basis. These are, for example, gradient vectors. Only if in the space a metric is defined the two sets of components are related by means of the metrical tensor (see section B.3). In this case each vector may be written with covariant or contravariant components. The distinction "vector" and "covector" looses its usefulness and we will simply talk about vectors. For the components, however, the distinction between covariant and contravariant remains.

[2]Sometimes quantities which can be written with covariant components relative to the dual basis are called 1–forms, cf. (Misner et al., 1973).

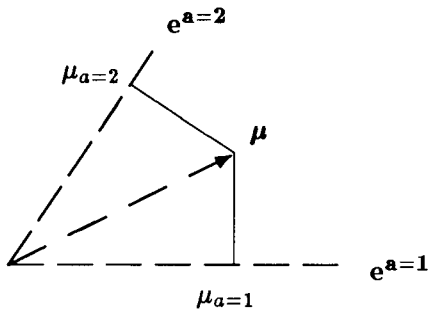


Figure B.2 *Covariant components.*

Tensors, being an extension of vectors, may also have covariant or contravariant components. For tensors of rank 2 or higher these components are, however, difficult to visualize geometrically, the extension being a purely mathematical one.

Now consider a general coordinate transformation $x^i = x^i(x^a)$. From this transformation we may compute the transformation matrix $\frac{\partial x^i}{\partial x^a}$ and the inverse transformation matrix $\frac{\partial x^a}{\partial x^i}$ such that $\frac{\partial x^i}{\partial x^a}\frac{\partial x^a}{\partial x^j} = \delta^i_j$. In a point P we now have two sets of coordinate lines, one set for the x^i–coordinates and one for the x^a–coordinates. The tangent vectors along all the coordinate lines x^a (together forming the basis $\mathbf{e_a}$) were found above through $\frac{\partial x^a}{\partial\{x^b\}}, a = 1, \ldots, n$ for each b. In the same manner we may find a set of tangent vectors along the x^i coordinate lines through $\frac{\partial x^a}{\partial\{x^i\}}, a = 1, \ldots, n$ for each i. This set forms a new basis $\mathbf{e_i}$ in P, spanning the same tangent space as does $\mathbf{e_a}$. The quantities $\frac{\partial x^a}{\partial x^i}$ may be interpreted as the components of the base vectors $\mathbf{e_i}$ relative to the basis $\mathbf{e_a}$, just as the quantities $\frac{\partial x^a}{\partial x^b} = \delta^a_b$ are the components of the base vectors $\mathbf{e_a}$ relative to this basis itself:

$$\mathbf{e_i} = \frac{\partial x^a}{\partial x^i}\mathbf{e_a}$$

and

$$\mathbf{e_a} = \frac{\partial x^b}{\partial x^a}\mathbf{e_b} = \delta^b_a\mathbf{e_b} \ .$$

We see that under a coordinate transformation $x^i = x^i(x^a)$ the contravariant base vectors $\mathbf{e_a}$ transform with the inverse transformation matrix $\frac{\partial x^a}{\partial x^i}$. Then we have also the inverse relation:

$$\mathbf{e_a} = \frac{\partial x^i}{\partial x^a}\mathbf{e_i} \ .$$

For the dual basis we have in a similar manner

$$\mathbf{e^i} = \frac{\partial x^i}{\partial x^a}\mathbf{e^a}$$

$$\mathbf{e^a} = \frac{\partial x^a}{\partial x^i}\mathbf{e^i} \ .$$

In order to see how the contravariant components of a vector transform, we proceed as follows:

$$\boldsymbol{\lambda} = \lambda^a \mathbf{e_a} = \lambda^a \frac{\partial x^i}{\partial x^a} \mathbf{e_i} = \lambda^i \mathbf{e_i}$$

$$\Rightarrow \lambda^i = \frac{\partial x^i}{\partial x^a} \lambda^a$$

so the contravariant components transform with the transformation matrix $\frac{\partial x^i}{\partial x^a}$. In a similar manner we may show that covariant components transform with the inverse transformation matrix $\frac{\partial x^a}{\partial x^i}$. Whereas we already defined the contravariant and covariant components at the moment of introduction of the two sets of base vectors (basis and dual basis) we could have equally well defined them at this stage as the components transforming under a general coordinate transformation $x^i = x^i(x^a)$ with respectively the transformation matrix $\frac{\partial x^i}{\partial x^a}$ or the inverse transformation matrix $\frac{\partial x^a}{\partial x^i}$. Extension of this reasoning to quantities of higher rank forms the background of the tensor definition given in section B.2.1.

Besides the classical theory of tensors described above (which is strongly connected with geometry due to historical developments) there is also a more modern approach for introducing tensors, which uses the abstract–algebraic concepts of space, vector space, sets, linear and multi–linear functionals, mappings, manifolds, groups, charts, atlases, etc. Examples of this approach can be found in (Foster and Nightingale, 1979), (Bishop and Goldberg, 1968) and others. Whereas in the classical approach the first step was a special metrical 3–space (Euclidean space $\mathbf{E^3}$) which was later generalized to non–metrical spaces of any dimension, the modern approach starts with general spaces of arbitrary dimension and may end up (after imposing more and more conditions onto the space) with the $\mathbf{E^3}$ as a very special case. In spite of the $\mathbf{E^3}$ being in fact the traditional "working space" of geodesists, the modern approach must also have some appeal to them if interpreted as a framework in which we work from "large" (most general) to "small" (specific case).

B.2.3 Derivatives

In the last section we saw that tensor analysis deals either with problems concerning non–linear spaces or problems in linear spaces expressed in curvilinear coordinates. (It is obvious that the use of curvilinear coordinates in a non–linear space is inevitable.) Consider now one of these two possible situations. In the space, a tensor field is defined which, in general, will be a function of the coordinates. In a point of this space we may construct a basis spanning the local tangent space in that point. In curvilinear coordinates this basis will also depend on the coordinates so we are in fact considering a field of coordinate dependent base vectors and tangent spaces (O'Neill, 1966). A change of coordinates will in general imply a change of the local basis with respect to which the tensor field is defined. Study of the tensor field therefore requires knowledge about the way in which neighbouring tangent spaces

are connected. The situation of continuously changing base vectors (which is the case when using curvilinear coordinates) forces us to consider differential coordinate changes. This brings us to examining the process of differentiation of tensors. If we like to continue our computations using tensor analysis, we require that this differentiation process yields again a tensor. Partial differentiation of a tensor, as known from "ordinary" analysis, does, however, in general not lead to a tensor, as can be shown easily. Consider an arbitrary contravariant vector v^a which, under a general coordinate transformation $x^i = x^i(x^a)$ will transform as $v^i = \frac{\partial x^i}{\partial x^a} v^a$. The partial derivative of this equation with respect to x^i yields:

$$\begin{aligned} \frac{\partial v^i}{\partial x^j} &= \frac{\partial}{\partial x^j}\left(\frac{\partial x^i}{\partial x^a}\right) v^a + \frac{\partial x^i}{\partial x^a}\frac{\partial v^a}{\partial x^j} \\ &= \frac{\partial^2 x^i}{\partial x^a \partial x^b}\frac{\partial x^b}{\partial x^j} v^a + \frac{\partial x^i}{\partial x^a}\frac{\partial x^b}{\partial x^j}\frac{\partial v^a}{\partial x^b} \,. \end{aligned} \tag{B.5}$$

The partial derivative is denoted by a comma preceding the differentiation index: ${v^i}_{,j} = \frac{\partial v^i}{\partial x^j}$. If the term containing the second–order partial derivative was absent, ${v^i}_{,j}$ would transform as a mixed tensor. We conclude that obviously the ordinary partial derivative is not a tensor. One exception to this conclusion is the partial derivative of a scalar field S, which, according to the chain rule, transforms as

$$\frac{\partial S}{\partial x^i} = \frac{\partial x^a}{\partial x^i}\frac{\partial S}{\partial x^a} \,.$$

So for each scalar field S, $S_{,a}$ is a tensor.

Partial differentiation obviously is not the right method to account for the connection of neighbouring tangent spaces. In order to find this connection, we look at the way in which the basis $\mathbf{e_a}$ changes under a differential change of coordinates dx^a in each of the coordinate directions. Expressing this change of the base vectors $\frac{\partial \mathbf{e_a}}{\partial x^b}$ relative to the basis $\mathbf{e_a}$ itself, we may write

$$\frac{\partial \mathbf{e_a}}{\partial x^b} = \Gamma^c_{ab}\, \mathbf{e_c} \tag{B.6}$$

where the Γ^c_{ab} are called the *coefficients of the linear connection* or for short "linear connection", "connection coefficients", "affine connection", etc. The way of introducing the connection coefficients as in equation B.6 may be interpreted as imposing a certain structure on an arithmetic space (although this structure is not a metric). This structure consists of nothing more then requiring the base vectors to change in a *linear* way under a differential coordinate change and not completely arbitrarily. The space is now said to be a linearly connected space (Moon and Spencer, 1986). The Γ^c_{ab} may in general be functions of the coordinates. Sometimes they are introduced just by giving their transformation equation (ibid.), which is:

$$\Gamma^c_{ab} = \frac{\partial x^i}{\partial x^a}\frac{\partial x^j}{\partial x^b}\frac{\partial x^c}{\partial x^k}\Gamma^k_{ij} + \frac{\partial x^c}{\partial x^i}\frac{\partial^2 x^i}{\partial x^a \partial x^b} \,. \tag{B.7}$$

This equation (which we give without derivation) shows that the connection coefficients do not form a tensor. It furthermore shows that the linear connection is symmetric in its two subscripts, at least if $\frac{\partial^2 x^i}{\partial x^a \partial x^b} = \frac{\partial^2 x^i}{\partial x^b \partial x^a}$. We will in the sequel assume that the latter is always the case. If this is not true, the basis field is called ***anholonomic*** and the space is a so–called *torsional* space in which differentiation is possible, but integration is not.

With the linear connection we try to find a process of differentiation of a tensor which gives again a tensor. From equation B.7 we solve for $\frac{\partial^2 x^i}{\partial x^a \partial x^b}$ through multiplication with $\frac{\partial x^j}{\partial x^c}$:

$$\frac{\partial^2 x^i}{\partial x^a \partial x^b} = \frac{\partial x^i}{\partial x^c}\Gamma^c_{ab} - \frac{\partial x^j}{\partial x^a}\frac{\partial x^k}{\partial x^b}\Gamma^i_{jk} \; .$$

Substitution of this expression into equation B.5 gives after some manipulations (cf. (Moon and Spencer, 1986)):

$$\frac{\partial v^i}{\partial x^j} + \Gamma^i_{jk}\, v^k = \frac{\partial x^i}{\partial x^a}\frac{\partial x^b}{\partial x^j}\left(\frac{\partial v^a}{\partial x^b} + \Gamma^a_{bc}\, v^c\right) \; .$$

This equation shows that the quantity on the left hand side transforms as a mixed tensor. It is called the *covariant derivative* of a contravariant vector and is denoted by a semi–colon preceding the differentiation index: ${v^i}_{;j}$ (or $\nabla_j v^i$):

$${v^i}_{;j} \equiv {v^i}_{,j} + \Gamma^i_{jk}\, v^k \; . \tag{B.8}$$

We see the covariant derivative consists of two parts. One part is the familiar partial derivative. The additional term, containing the connection coefficients, accounts for the change of the basis in the differentiation process. Sometimes it is no longer important that a tensor was originally created through the process of covariant differentiation. In that case we simply write v^i_j. This also means that every covariant index of a tensor of rank r can be viewed upon as originating from a process of covariant differentiation of another tensor of rank $r-1$, namely one covariant index less. For example: ${A^i}_{jk} = {A^i}_{;jk} = {A^i}_{;jk}$ or $V_{rs} = V_{r;s} = V_{;rs}$ or $\varphi_a = \varphi_{;a} = \varphi_{,a}$.

Consider now the scalar field $S = v_a w^a$. Since for a scalar field the partial derivative transforms as a tensor, it is $S_{,a} = S_{;a}$. So we have:

$$(v_a w^a)_{,b} = (v_a w^a)_{;b}$$

or

$$\begin{aligned} v_{a,b}\, w^a + v_a\, {w^a}_{,b} &= v_{a;b}\, w^a + v_a\, {w^a}_{;b} \\ &= v_{a;b}\, w^a + v_a({w^a}_{,b} + \Gamma^a_{bc}\, w^c) \end{aligned}$$

yielding (with a change of dummy indices)

$$v_{a,b}\, w^a = v_{a;b}\, w^a + v_c \Gamma^c_{ab}\, w^a$$

which has to be true for arbitrary w^a, so that we find for the covariant derivative of a covariant vector:

$$v_{a;b} = v_{a,b} - \Gamma^c_{ab}\, v_c \ . \tag{B.9}$$

The covariant derivative of a contravariant tensor of rank 2, t^{ab} can be found in a similar manner by writing $t^{ab} = v^a w^b$. The extension to tensors of arbitrary rank or type follows analogously. For each superscript one *adds* to the partial derivative a term with Γ, and for each subscript one *subtracts* a term with Γ (like the ones in equations B.8 and B.9).

The covariant derivative is the tensor analogy of the partial derivative of classical analysis. For the total derivative $dv^a = \frac{\partial v^a}{\partial x^b} dx^b$ the analogy in tensor analysis is the *absolute derivative*, defined as:

$$Dv^a \equiv v^a_{\ ;b}\, dx^b = dv^a + \Gamma^a_{bc}\, v^c dx^b \ . \tag{B.10}$$

Again only for a scalar field the absolute derivative of tensor analysis equals the total derivative of ordinary analysis. The absolute derivative can be used to extend the notion of a field of "parallel" vectors along a straight line in $\mathbf{E}^3$ (which is $\frac{dv^a}{du} = 0$) to a parallel field of vectors along a general curve in a general space (see section B.3.1). For the latter we have

$$\frac{Dv^a}{du} = 0 \tag{B.11}$$

along the curve parametrized by the parameter u. If we take for v^a in this equation the tangent vector dx^a/du along the curve, equation B.11 becomes

$$\frac{d^2x^a}{du^2} + \Gamma^a_{bc} \frac{dx^b}{du} \frac{dx^c}{du} = 0$$

which is the equation of a *geodesic*. So geodesics are those curves in space along which the tangent vectors in all points form a parallel vector field. Note that the meaning of the word parallel in this sense is much more general and abstract than the visual meaning it has in Euclidean space.

B.3 Geometry

In the previous sections some aspects of index notation and tensor analysis were treated. As explained in section B.2 purely geometrical subjects were left out. In the present section attention will be paid to geometry. First the notion and classification of several spaces will be treated, which was already initiated in section B.2.2. Furthermore, we saw in section B.2.3 that the introduction of the linear connection Γ already imposes some kind of structure to a space. The step from such non–Riemannian (non–metrical) spaces to Riemannian (metrical) spaces consists of the introduction of a metrical tensor. This tensor will be treated in section B.3.2. Special attention will be drawn to curvature and curvature related quantities in the last part of this section. Many general textbooks exist on geometry. As an example see (O'Neill, 1966).

B.3.1 Spaces

In section B.2.2 a space was considered a set of points with some structure. Such a definition still remains vague because no explanation of the words "point" and "structure" is given (Moon and Spencer, 1986). The reader, probably familiar with these concepts, will have an intuitive understanding of these ideas. When proceeding, the meaning will become clearer through the description of some special cases. Instead of "space", the word *manifold* is also used sometimes, but in the literature one encounters various definitions of a manifold, strongly dependent on further applications. Most frequently it is used when introducing the tensor concept in a modern, abstract–algebraic manner (see section B.2.2). In that case, the structure imposed on the set of points (which has to have a finite dimension) has to be such that coordinate transformations involve differentiable functions. Or simply put: the space has to have certain smoothness properties. The classical, geometrical treatment of tensor analysis usually uses the word space. The introduction of a coordinate system (viewed upon as a mapping of the elements of the space (the points) onto the set of real numbers) allows labelling of the points in a unique manner. If the structure of the space is reduced to zero, we call it an *arithmetic space*, denoted by $\mathbf{X^n}$.

A very primitive structure can be imposed on the arithmetic space by giving a certain meaning to concepts like "in", "out", "near", etc. The space is then called *topological*. As we already saw, another way of giving some kind of structure to a space is the introduction of the connection coefficients Γ^a_{bc}. In general these are functions of the coordinates, i.e. $\Gamma^a_{bc}(x^a)$. The space is now a *linearly connected non–metrical space*, denoted by $\mathbf{L^n}$. If, in such a space, linear coordinates may be introduced, we have a linearly connected, non–metrical *linear* space, which we already encountered before as an *affine space*, denoted by $\mathbf{A^n}$. Calling the latter space affine or linear may tempt us to call the former "curved", but since curvature is a concept which only has a meaning in metrical spaces, this may be misleading. As a compromise a linearly connected, non–metrical and non–linear space is sometimes called "pseudo–curved". Different coordinate systems labelling a pseudo–curved space are connected via non–linear transformations. The "point" is one of the few invariants of such a space. In affine space (in which coordinate systems are connected through linear transformations) there are more invariants (like proportionalities of lengths and surfaces or all kinds of intersection properties) giving already the possibility of solving some elementary geometrical problems.

What remains are concepts like distance, angle, orthogonality and others, which first obtain a meaning if a *metric* is introduced in the space. This idea originated from Riemann, who proposed to introduce the concept of distance by means of a quadratic differential form

$$ds^2 = g_{ab}dx^a dx^b \ , \tag{B.12}$$

the infinitesimal distance ds being an invariant and the quantities g_{ab} the metrical coefficients (or *metrical tensor*), which, in general, are functions of the coordinates. The metrical tensor will be discussed in the next section. A space endowed with a

metric by means of equation B.12 is called a *Riemannian space*[3], denoted by $\mathbf{R^n}$.

In general a Riemannian space is curved. We may use this word here, since in a metrical space we can give a meaning to the word curvature, as we will see in section B.3.3. Only if it is possible to introduce in the space globally (i.e. covering the whole space) a linear orthogonal coordinate system, the space is called linear, flat or *Euclidean* and is denoted by $\mathbf{E^n}$. The complete classification of spaces as it is presented here is given schematically in figure B.3.

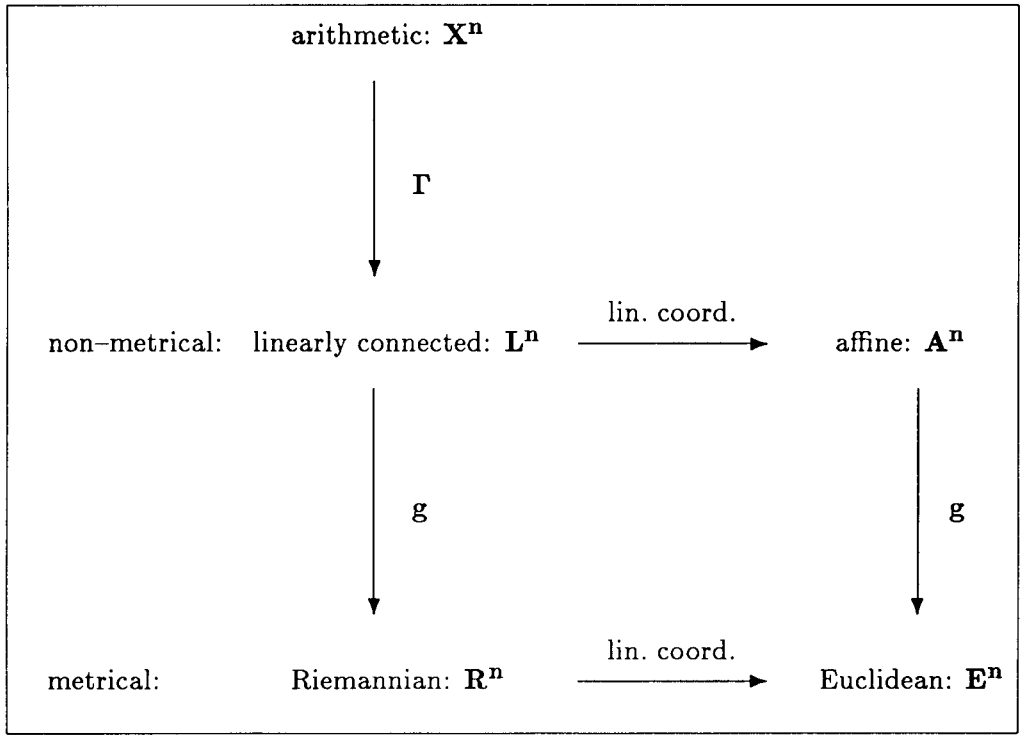


Figure B.3 *Spaces*

B.3.2 Metric

The metric, imposed on a space by means of the introduction of the metrical tensor g, allows us to give a meaning to geometrical concepts like length, angle, volume, etc. From the definition of the invariant ds (equation B.12) it follows that g_{ab} is symmetric. In general g_{ab} is a full matrix of which the components are functions of the coordinates. Since g_{ab} is a tensor (which we will not prove here), its transformation

[3]In fact, Riemann introduced the quadratic form directly into an arithmetic space. The connection coefficients Γ were afterwards expressed in terms of the metrical tensor, as we shall see later.

equation is

$$g_{ab} = \frac{\partial x^i}{\partial x^a}\frac{\partial x^j}{\partial x^b} g_{ij} \ . \tag{B.13}$$

This equation is extremely useful for computation of the components of g_{ab} in various curvilinear coordinate systems from g_{ij} in cartesian coordinates, at least if the transformation equations $x^i = x^i(x^a)$ are known. The metric tensor provides the space with an inner product. So analogously to equation B.12, which gives the infinitesimal length ds belonging to a differential change of coordinates dx, the length v of a vector v^a is given by

$$v^2 = g_{ab} v^a v^b \ ,$$

and the inner product of two vectors v^a and w^a is

$$g_{ab} v^a w^b \ .$$

If the latter quantity is zero, the two vectors are orthogonal. The angle θ between the two vectors is now defined as

$$\cos\theta \equiv \frac{g_{ab} v^a w^b}{vw}$$

with $v = \sqrt{g_{ab} v^a v^b}$ and $w = \sqrt{g_{ab} w^a w^b}$. If we have an orthogonal coordinate system and we take for v^a and w^a differential changes of the coordinates in each of the coordinate directions, we can show that the metrical tensor g_{ab} takes on the diagonal form, cf. (Moon and Spencer, 1986). If, furthermore, the coordinates are linear, the diagonal components of g_{ab} are constants. In a linear, non–orthogonal coordinate system, g_{ab} has not the diagonal form, but its components are still constants.

The *contravariant metric tensor* g^{ab} (also called associated or conjugate metric tensor) is defined through

$$g^{ab} g_{bc} = \delta^a{}_c$$

and it can be considered the inverse of g_{ab}. The contravariant metric tensor provides the dual space with an inner product:

$$g^{ab} V_a W_b \ .$$

Now let us define the components of a covariant vector $\boldsymbol{\mu}$ as the tensor product

$$\mu_a = g_{ab} \lambda^b$$

where λ^b are the components of an arbitrary contravariant vector $\boldsymbol{\lambda}$. Using the definition of the contravariant metric tensor we find the inverse relation as

$$\lambda^a = g^{ab} \mu_b \ .$$

The lengths of the two vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are respectively

$$\mu = \sqrt{g^{ab} \mu_a \mu_b}$$

and

$$\lambda = \sqrt{g_{ab}\lambda^a\lambda^b} \, .$$

We may now write:

$$\mu^2 = g^{ab}\mu_a\mu_b = \lambda^a\mu_a$$
$$\lambda^2 = g_{ab}\lambda^a\lambda^b = \lambda^a\mu_a$$

from which we conclude that $\mu^2 = \lambda^2$. We therefore identify the two vectors and write for the components

$$\lambda_a = g_{ab}\lambda^b$$

or

$$\lambda^a = g^{ab}\lambda_b \, .$$

These operations (which involve the metrical tensor) are called respectively the *lowering* and *raising* of indices. They give us the possibility to compute from a vector either the covariant or the contravariant components. In a space where a metric is defined, the distinction between covariant and contravariant therefore only applies to components, not any more to the vectors (or in general tensors) themselves (compare section B.2.2).

Whereas the introduction of a metric offers us the possibility of solving a large number of geometrical problems *more* than without a metric, it is in fact a limitation on the generality of the space. As we shall see, the introduction of the metric limits the choice for the connection coefficients Γ. These have now a specified form to be computed from the metrical tensor. This relationship will be derived now.

Consider two arbitrary parallel vector fields λ^a and μ^a along a curve parametrized by a parameter u. According to equation B.11 we have for the two vector fields

$$\frac{D\lambda^a}{du} = 0$$

and

$$\frac{D\mu^a}{du} = 0 \, .$$

In $\mathbf{E}^3$ the inner product of two parallel vector fields is constant along the curve, so we require analogously that in curved space of any dimension

$$\frac{d(g_{ab}\lambda^a\mu^b)}{du} = \frac{D(g_{ab}\lambda^a\mu^b)}{du} = 0 \, .$$

(For scalars the total derivative equals the absolute derivative, see section B.2.3.) This yields:

$$\begin{aligned}\frac{D(g_{ab}\lambda^a\mu^b)}{du} &= \left(\frac{Dg_{ab}}{du}\right)\lambda^a\mu^b + g_{ab}\left(\frac{D\lambda^a}{du}\right)\mu^b + g_{ab}\lambda^a\left(\frac{D\mu^b}{du}\right) \\ &= \left(\frac{Dg_{ab}}{du}\right)\lambda^a\mu^b\end{aligned}$$

$$= g_{ab;c}\frac{dx^c}{du}\lambda^a\mu^b$$
$$= 0\ .$$

This must hold for arbitrary vector fields λ^a and μ^a and for arbitrary tangent vectors $\frac{dx^c}{du}$ so we have

$$g_{ab;c} = 0$$

(which is a tensor equation, so it holds in any coordinate system) or if we use equation B.9

$$g_{ab,c} = \Gamma^d_{ac}\, g_{bd} + \Gamma^d_{bc}\, g_{ad}\ .$$

Cyclic permutation of the indices gives

$$g_{ca,b} = \Gamma^d_{bc}\, g_{ad} + \Gamma^d_{ab}\, g_{cd}$$

$$g_{bc,a} = \Gamma^d_{ab}\, g_{cd} + \Gamma^d_{ac}\, g_{bd}\ .$$

Adding the second and the third equation and subtracting the first and also dividing by 2 yields

$$\Gamma^d_{ab}\, g_{cd} = \frac{1}{2}\left(g_{ac,b} + g_{bc,a} - g_{ab,c}\right) \tag{B.14}$$

or

$$\Gamma^c_{ab} = \frac{1}{2}g^{cd}\left(g_{ad,b} + g_{bd,a} - g_{ab,d}\right)\ . \tag{B.15}$$

Christoffel already derived this quantity (equation B.15) some time before the development of the tensor analysis, and it is therefore called *Christoffel symbol of the second kind.* The Christoffel symbol of the first kind is equation B.14:

$$\Gamma_{abc} \equiv g_{cd}\Gamma^d_{ab}\ .$$

The explicit form of the components of the metric tensor and the Christoffel symbols are given for several coordinate systems in section A.2.

We should mention that, with the use of the metric tensor, we may also give a definite meaning to the concepts of divergence and the Laplacian. Consider a scalar field φ. The gradient of this scalar field is $\varphi_{;a} = \varphi_a$, which forms a covariant vector. The contravariant components are $\varphi^a = g^{ab}\varphi_b$. Covariant differentiation of the latter and contraction of the indices gives the divergence of φ^a:

$$\varphi^a{}_{;a} = \varphi^a{}_a = g^{ab}\varphi_{ab}\ ,$$

which may also be considered the Laplacian of the scalar field φ.

B.3.3 Curvature

In a curved space no linear coordinates can be introduced, at least not globally. Take for example the 2–dimensional surface of a sphere. It may be described by curvilinear polar coordinates (θ, λ) but not with cartesian coordinates (x, y). In a flat space, however, linear coordinates as well as curvilinear coordinates may be introduced. Take for example in $\mathbf{E}^3$ cartesian coordinates (x, y, z) and polar coordinates (r, θ, λ). In general, the components of the metrical tensor in curvilinear coordinates are functions of the coordinates. Only in linear coordinates, those components are constants. Suppose now, that we are given a metrical tensor of which the components are functions of the coordinates. How can we find out if this metric belongs to a curved space, so that it by no means can be transformed to a coordinate system in which the components are constants, or that it is just the metrical tensor of flat space but expressed in curvilinear coordinates? Or stated otherwise: what is the condition for a space to be flat?.

In order to get an answer to this question we look at the way in which the components of the metrical tensor change under a differential change of coordinates, i.e. we compute the derivatives of the metrical tensor. In linear coordinates the components of the metrical tensor are constants, so their partial derivatives will be zero. But the partial derivative is not a tensor, so we may not conclude that in that case the components of the metrical tensor will be constant in every other coordinate system. We have to look to the covariant derivative. However, this does not help us since we already saw in the last section that the first covariant derivative of the metrical tensor is zero in all coordinate systems, independent of the structure of the space. We may try to look to the Christoffel symbols. They are some function of the first partial derivatives of the metrical tensor. But they also are no tensors, so if they are zero in linear coordinates, this does not mean that they are zero in other coordinate systems.

The answer to the question above has to be found by considering the second partial derivatives of the metrical tensor. Now the first partial derivatives of the metrical tensor appear, via the Christoffel symbols, in the expression for the first covariant derivative of an arbitrary vector v_a (in general of arbitrary tensors). For the second partial derivatives we therefore compute the second covariant derivative of v_a. We find:

$$\begin{aligned} v_{a;bc} &= (v_{a;b})_{,c} - \Gamma^d_{ac}\, v_{d;b} - \Gamma^d_{bc}\, v_{a;d} \\ &= (v_{a,b} - \Gamma^d_{ab}\, v_d)_{,c} - \Gamma^d_{ac}(v_{d,b} - \Gamma^e_{bd}\, v_e) - \Gamma^d_{bc}(v_{a,d} - \Gamma^e_{ad}\, v_e) \\ &= v_{a,bc} - \Gamma^d_{ab,c}\, v_d - \Gamma^d_{ab}\, v_{d,c} - \Gamma^d_{ac}\, v_{d,b} + \\ &\quad + \Gamma^d_{ac}\, \Gamma^e_{bd}\, v_e - \Gamma^d_{bc}\, v_{a,d} + \Gamma^d_{bc}\, \Gamma^e_{ad}\, v_e \,. \end{aligned}$$

We also have, on interchanging the indices b and c:

$$\begin{aligned} v_{a;cb} = v_{a,cb} &- \Gamma^d_{ac,b}\, v_d - \Gamma^d_{ac}\, v_{d,b} - \Gamma^d_{ab}\, v_{d,c} + \\ &+ \Gamma^d_{ab}\, \Gamma^e_{cd}\, v_e - \Gamma^d_{bc}\, v_{a,d} + \Gamma^d_{bc}\, \Gamma^e_{ad}\, v_e \,. \end{aligned}$$

Subtraction yields:

$$v_{a;bc} - v_{a;cb} = v_d \left(\Gamma^d_{ac,b} - \Gamma^d_{ab,c} + \Gamma^e_{ac}\,\Gamma^d_{be} - \Gamma^e_{ab}\,\Gamma^d_{ce}\right)$$

where we used $v_{a,bc} = v_{a,cb}$ and the symmetry of the Christoffel symbols and we also changed dummy indices. The left hand side of this equation is obviously a tensor, so is v_d, which implies that the term between brackets is also a tensor. This is the *Riemann–Christoffel tensor* $R^d{}_{abc}$, also called the *curvature tensor*:

$$R^d{}_{abc} \equiv \Gamma^d_{ac,b} - \Gamma^d_{ab,c} + \Gamma^d_{be}\,\Gamma^e_{ac} - \Gamma^d_{ce}\,\Gamma^e_{ab}\ . \tag{B.16}$$

Now, in flat space and in linear coordinates, the components of the metrical tensor are constant. The components of the Christoffel symbols are in this case zero, and so are all the components of the Riemann–Christoffel tensor, as follows from equation B.16. Since the latter is a tensor, its components will in this case also be zero in all other coordinate systems. The vanishing of the Riemann–Christoffel tensor is therefore a condition for the space to be flat, and vice versa. It can be shown that this is a necessary and sufficient condition (see for example (Ohanian, 1976)). The curvature tensor contains all the information about the curvature of the space in its n^4 components (where n is the dimension of the space). However, it possesses a number of symmetries and its components satisfy certain identities, which reduce the number of independent components to $\frac{1}{12}n^2(n^2-1)$ (see for example (Moon and Spencer, 1986)). Nevertheless, it is sometimes useful to consider contractions of the curvature tensor. The *Ricci tensor* is defined as the contraction of the curvature tensor on its first and last index:

$$R_{ab} \equiv R^c{}_{abc}$$

and the *curvature scalar* is defined as

$$R \equiv g^{ab} R_{ab} = R^a{}_a\ .$$

We mention one last contracted form of the curvature tensor, especially valuable in the general theory of relativity. It is the *Einstein tensor*, defined as

$$G^{ab} \equiv R^{ab} - \frac{1}{2} R\, g^{ab}\ .$$

It can be shown that this tensor is symmetric and has zero divergence, i.e.

$$G^{ab}{}_{;b} = 0\ .$$

appendix **C**

Inclination functions

In chapter 4 we used, as observation equations, expressions for the local cartesian second–order potential derivatives in terms of derivatives with respect to the orbital polar coordinates (r, ϕ, ω_o). The latter coordinate system has the satellite's orbit as equator. The expressions are derived in section 3.1. We see that in this case we have to know the expression for V_ϕ, the potential derivative with respect to ϕ, for example for the computation of the gradients V_{xy} and V_{yz}. However, in section 3.2 we had the potential given as function of either (r, θ, λ) or $(r, \omega_o, \omega_e, I)$ and not of the coordinates (r, ϕ, ω_o). In order to find V_ϕ we proceed as follows.

Let us compare the expressions for V_y in terms of $x^{a'}$ and in terms of $x^{A'}$ (cf. (Betti and Sansò, 1989)):

$$V_y = \frac{1}{r \sin \omega_o} V_I$$
$$V_y = \frac{1}{r} V_\phi \ .$$

From these expressions we see that obviously

$$V_\phi = \frac{1}{\sin \omega_o} V_I \ . \tag{C.1}$$

For V_I a series expansion exists (see section 3.2) so a similar expansion has to exist for V_ϕ in which the term $\sin^{-1} \omega_o$ has already been included. This expansion can be found in an analogous manner as the expansion of the potential in inclination functions. The latter expansion is derived in (Kaula, 1966) and we will only briefly show here the derivation of the series expansion for V_ϕ.

We start with formula (3.58) from (ibid.) which we write in the following manner

$$V_{lm} = \frac{GM}{R} \left(\frac{R}{r} \right)^{l+1} \mathrm{Re} \left[\{ (C_{lm} \cos m\omega_e + S_{lm} \sin m\omega_e) +$$

$$+ j\,(C_{lm}\sin m\,\omega_e - S_{lm}\cos m\,\omega_e)\}\sum_{t=0}^{k} T_{lmt}\sin^{l-m-2t} I \times$$
$$\times\sum_{s=0}^{m}\begin{pmatrix} m \\ s \end{pmatrix} j^s \cos^s I \cos^{m-s}\omega_o \sin^{l-m-2t+s}\omega_o\Bigg]$$

where $j=\sqrt{-1}$, Re means the real part and

$$T_{lmt} = \frac{(-1)^t(2l-2t)!}{2^l t!(l-t)!(l-m-2t)!}\,.$$

When differentiating V_{lm} with respect to I and dividing by $\sin\omega_o$ we obtain (compare equation C.1)

$$\frac{\partial V_{lm}}{\partial\phi} = \frac{GM}{R}\left(\frac{R}{r}\right)^{l+1}\mathrm{Re}\left[\sum_{t=0}^{k} T_{lmt}\sum_{s=0}^{m}\begin{pmatrix} m \\ s \end{pmatrix} j^s f(I)g(\omega_e)h(\omega_o)\right] \tag{C.2}$$

where

$$f(I) = \sin^{l-m-2t-1} I\cos^{s-1} I\left[(l-m-2t)\cos^2 I - s\sin^2 I\right]$$
$$g(\omega_e) = (C_{lm} - jS_{lm})\,e^{jm\omega_e}$$
$$h(\omega_o) = \sin^{l-m-2t+s-1}\omega_o\cos^{m-s}\omega_o\,.$$

Equation C.2 can be compared with formula (3.59) from (Kaula, 1966). Proceeding in the same way as (ibid.), we arrive at

$$\frac{\partial V_{lm}}{\partial\phi} = \frac{GM}{R}\left(\frac{R}{r}\right)^{l+1}\sum_{p=0}^{l-1} F^*_{lmp}(I)\left\{\begin{bmatrix} S_{lm} \\ C_{lm}\end{bmatrix}_{l-m:\mathrm{odd}}^{l-m:\mathrm{even}}\cos\psi_{lmp} + \right.$$
$$\left. + \begin{bmatrix} -C_{lm} \\ S_{lm}\end{bmatrix}_{l-m:\mathrm{odd}}^{l-m:\mathrm{even}}\sin\psi_{lmp}\right\} \tag{C.3}$$

where

$$\psi_{lmp} = (l-2p-1)\,\omega_o + m\,\omega_e$$

$$F^*_{lmp}(I) = \sum_{t=0}^{tmax} T_{lmt}\sum_{s=0}^{m}\begin{pmatrix} m \\ s \end{pmatrix} f(I)2^{2t-l+1}(-1)^{k+t}\times$$
$$\times\sum_{c=cmin}^{cmax}\begin{pmatrix} l-m-2t+s-1 \\ c \end{pmatrix}\begin{pmatrix} m-s \\ p-t-c \end{pmatrix}(-1)^c$$

$$
\begin{aligned}
f(I) &\quad \text{as in equation C.2} \\
k &= \text{integer part of } (l-m)/2 \\
tmax &= \min(k,p) \\
cmax &= \min(l-m-2t+s-1, m-s) \\
cmin &= \max(0, p-t)\ .
\end{aligned}
$$

Methods of computing of the inclination functions and their derivatives with respect to I can be found in e.g. (Wagner, 1983) , (Schrama, 1989) and (Goad, 1987). The cross-track inclination functions $F^*_{lmp}(I)$ as derived above can be computed in a similar manner as will be shown hereafter. For the description of a *recursive* algorithm for computation of the inclination functions or the cross-track inclination functions, we refer to respectively (Sneeuw, 1991a) and (Sneeuw, 1991b).

Following (Wagner, 1983) the method for computation of the inclination functions described in (Schrama, 1989) makes use of a "unit potential function" developed along a great circle with inclination I. Along this circular unit orbit (defined by $R=1, \Omega=\theta_G=e=0$ (so also $\omega_e=0$) and $M+\omega=\omega_o \equiv u$) we evaluate the function $\frac{\partial V_{lm}}{\partial \phi}$ which can be obtained by applying the chain rule of differentiation

$$
\frac{\partial V_{lm}}{\partial \phi} = \frac{\partial V_{lm}}{\partial \theta}\frac{\partial \theta}{\partial \phi} + \frac{\partial V_{lm}}{\partial \lambda}\frac{\partial \lambda}{\partial \phi} \tag{C.4}
$$

with

$$
\begin{aligned}
\frac{\partial V_{lm}}{\partial \theta} &= \bar{P}'_{lm}(\cos m\lambda + \sin m\lambda) \\
\frac{\partial V_{lm}}{\partial \lambda} &= m\bar{P}_{lm}(\cos m\lambda - \sin m\lambda)
\end{aligned}
$$

where we use a unit potential function defined by $GM=R=r=\bar{C}_{lm}=\bar{S}_{lm}=1$. Now we have to find the partial derivatives $\frac{\partial \theta}{\partial \phi}$ and $\frac{\partial \lambda}{\partial \phi}$. From appendix A we have on the unit sphere

$$
\begin{aligned}
X &= \sin\theta\cos\lambda \\
Y &= \sin\theta\sin\lambda \\
Z &= \cos\theta
\end{aligned} \tag{C.5}
$$

and

$$
\begin{aligned}
X' &= \cos\phi\cos\omega_o \\
Y' &= \cos\phi\sin\omega_o \\
Z' &= \sin\phi\ .
\end{aligned} \tag{C.6}
$$

The transformation between the coordinate systems x^I and $x^{I'}$ consists only of a rotation about the X–axis by an angle I (since $\omega_e = 0$), or:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos I & -\sin I \\ 0 & \sin I & \cos I \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} . \tag{C.7}$$

Combination of eq. C.5, C.6 and C.7 leads to:

$$\sin\theta\cos\lambda = \cos\phi\cos\omega_o \tag{C.8}$$

$$\sin\theta\sin\lambda = \cos\phi\sin\omega_o\cos I - \sin\phi\sin I \tag{C.9}$$

$$\cos\theta = \cos\phi\sin\omega_o\sin I + \sin\phi\cos I . \tag{C.10}$$

Differentiating eq. C.10 yields for $\phi = 0$:

$$\frac{\partial\theta}{\partial\phi} = -\frac{\cos I}{\sin\theta}$$

and differentiating the quotient of C.9 and C.8

$$\frac{\partial\lambda}{\partial\phi} = -\frac{\sin I\cos\lambda}{\sin\theta} .$$

Now we have from eq. C.3 for the unit potential

$$\frac{\partial V_{lm}}{\partial\phi} = \sum_{p=0}^{l-1} \bar{F}^*_{lmp} \left\{ \cos(l-2p-1)u + \begin{bmatrix} -1 \\ 1 \end{bmatrix} \sin(l-2p-1)u \right\}$$

where the bar over F indicates that the cross–track inclination functions are now normalized. This series can be regarded as a Fourier–type series. Introducing $k = l - 2p - 1$ yields

$$\frac{\partial V_{lm}}{\partial\phi} = \sum_{k=0,1[2]}^{l-1} \left\{ \left(\bar{F}^*_{lm(l-1+k)/2} + \bar{F}^*_{lm(l-1-k)/2} \right) \cos ku + \right.$$

$$\left. + \left(\bar{F}^*_{lm(l-1-k)/2} - \bar{F}^*_{lm(l-1+k)/2} \right) \begin{bmatrix} -1 \\ 1 \end{bmatrix} \sin ku \right\}$$

where

$$k = \begin{cases} 0,2,4,\ldots,l-1 & l : \text{odd} \\ 1,3,5,\ldots,l-1 & l : \text{even} \end{cases} .$$

Comparison with an ordinary Fourier series

$$\frac{\partial V_{lm}}{\partial \phi} = \sum_{i=0}^{l-1} a_i^{\phi} \cos iu + b_i^{\phi} \sin iu$$

yields

$$\begin{aligned}
\bar{F}^*_{lm(l-1)/2} &= a_0^{\phi} && \text{for odd } l \\
\bar{F}^*_{lm(l-1+k)/2} &= (a_k^{\phi} + b_k^{\phi})/2 && \text{for } l-m \text{ even} \\
\bar{F}^*_{lm(l-1-k)/2} &= (a_k^{\phi} - b_k^{\phi})/2 && \text{for } l-m \text{ even} \\
\bar{F}^*_{lm(l-1+k)/2} &= (a_k^{\phi} - b_k^{\phi})/2 && \text{for } l-m \text{ odd} \\
\bar{F}^*_{lm(l-1-k)/2} &= (a_k^{\phi} + b_k^{\phi})/2 && \text{for } l-m \text{ odd} .
\end{aligned}$$

The Fourier coefficients a_k^{ϕ} and b_k^{ϕ} are derived by computing the unit potential at discrete points along a great circle. With an FFT routine these potential values (time domain) are transformed to the coefficients a_k^{ϕ} and b_k^{ϕ} (frequency domain).

References

Arabelos, D. and C.C. Tscherning (1990), Simulation of Regional Gravity Field Recovery from Satellite Gravity Gradiometer Data using Collocation and FFT, *Bulletin Géodésique*, **64**, 363–382

Balmino, G., F. Barlier, A. Bernard, C. Bouzat, R. Rummel and P. Touboul (1985), Proposal for a Satellite Gravity Gradiometer Experiment for the Geosciences, Bureau Gravimétrique International, Toulouse

Balmino, G. and J.P. Barriot (1990), Methods of Global Recovery of Harmonic Coefficients from SGG in the General case, Workpackage 520, Study on Precise Gravity Field Determination Methods and Mission Requirements, Final Report (Phase 2), ESA Contract no. 8153/88/F/FL

Balmino, G., J. Barriot, R. Koop, B. Middel, N.C. Thong and M. Vermeer (1991), Simulation of Gravity Gradients: a Comparison Study, *Bulletin Géodésique*, **65**, 218–229

Betti, B. and F. Sansò (1989), The Integrated Approach to Satellite Geodesy, in: *Lecture Notes in Earth Sciences*, **25**, *Theory of Satellite Geodesy and Gravity Field Determination*, F. Sansò, R. Rummel (Eds.), Springer–Verlag, Berlin Heidelberg New York

Bergmann, P.G. (1942), *Introduction to the Theory of Relativity*, Dover Publications, New York

Bishop, R.L. and S.I. Goldberg (1968), *Tensor Analysis on Manifolds*, Macmillan, New York

Carroll, J.J. and P.H. Savet (1959), Space Navigation and Exploration by Gravity Difference Detection, *Aero/Space Engineering*, 44–47

Chandrasekhar, S. (1965), The Post–Newtonian Equations of Hydrodynamics in General Relativity, *The Astrophysical Journal*, **142**, 1499–1512

Colombo, O.L. (1981), Numerical Methods for Harmonic Analysis on the Sphere, Department of Geodetic Science, report **310**, Ohio State University, Columbus

Colombo, O.L. (1986), Notes on the Mapping of the Gravity Field using Satellite Data, in: *Lecture Notes in Earth Sciences*, **7**, *Mathematical and Numerical Techniques in Physical Geodesy*, H. Sünkel (Ed.), Springer–Verlag, Berlin Heidelberg New York

Colombo, O.L. (1987), The Global Mapping of the Gravity Field with an Orbiting Full–Tensor Gradiometer: an Error Analysis, presented at XIX Assembly of the IUGG, Vancouver, Canada

Colombo, O.L. (1989), Advanced Techniques for High–Resolution Mapping of the Gravitational Field, in: *Lecture Notes in Earth Sciences*, **25**, *Theory of Satellite Geodesy and Gravity Field Determination*, F. Sansò, R. Rummel (Eds.), Springer–Verlag, Berlin Heidelberg New York

Dornier (1990), Aristoteles Additional Study, Executive Summary Report **5**, ESTEC Contract no. 8355/89/NL/JS

Dziewonski, A.M. (1984), Mapping the Lower Mantle: Determination of Lateral Heterogeneity in P velocity up to Degree and order 6, *Journal of Geophysical Research*, **89**, B7, 5929–5952

Emeljanov, N.V. and A.A. Kanter (1989), A Method to Compute Inclination Functions and their Derivatives, *Manuscripta Geodaetica*, **14**, 77–83

Eötvös, R. (1953), *Gesammelte Arbeiten*, Akadémiai Kiadó, Budapest

ESA (1991), Report of the Earth Observation User Consultation Meeting, ESA SP-1143

Fischbach, E., D. Sudarsky, A. Szafer, C. Talmadge and S. Aronson (1986), Reanalysis of the Eötvös Experiment, *Physical Review Letters*, **56**, 1, 3–6

Forward, R.L. (1974), Review of Artificial Satellite Gravity Gradiometer Techniques for Geodesy, in: *Use of Artificial Satellites for Geodesy and Geodynamics*, G. Veis (Ed.), 157-192

Forward, R.L. (1981), Gravity Sensors and the Principle of Equivalence, *IEEE Transactions on Aerospace and Electronic Systems*, **17**, 4, 511–519

Forward, R.L. (1982), Flattening Spacetime near the Earth, *Physical Review D*, **26**, 4, 735–744

Foster, J. and J.D. Nightingale (1979), *A Short Course in General Relativity*, Longman Group, Harlow, Essex

Gerstl, M. (1980), On the Recursive Computation of the Integrals of the Associated Legendre Functions, *Manuscripta Geodaetica*, **5**, 181–199

Gill, E., M. Soffel, H. Ruder and M. Schneider (1992), Relativistic Motion of Gyroscopes and Space Gradiometry, DGK **A–107**, München

Goad, C.C. (1987), An Efficient Algorithm for the Evaluation of Inclination and Eccentricity Functions, *Manuscripta Geodaetica*, **12**, 11-15

Goldstein, H. (1980), *Classical Mechanics*, 2nd edition, Addison–Wesley, Reading, Massachusetts

GRADIO (1989), Accelerometer Predevelopment, Final Report, ESTEC Contract no. 7631/88/NL/JS

Hager, B.H. (1983), Global Isostatic Geoid for Plate and Boundary Layer Models of the Lithosphere, *Earth and Planetary Science Letters*, **63**, 97–109

Hein, G.W. (1986), Integrated Geodesy – State–of–the–Art 1986 reference text, in: *Lecture Notes in Earth Sciences*, **7**, *Mathematical and Numerical Techniques in Physical Geodesy*, H. Sünkel (Ed.), Springer–Verlag, Berlin Heidelberg New York

Heiskanen, W.A. and H. Moritz (1967), *Physical Geodesy*, Freeman and Co., San Francisco

Hotine, M. (1969), *Mathematical Geodesy*, ESSA Monograph 2, U.S. Department of Commerce, Washington, D.C.

Ilk, K.H. (1983), Ein Beitrag zur Dynamik ausgedehnter Körper – Gravitationswechselwirkung –, DGK **C–288**, München

Ilk, K.H. (1987), Regional Gravity Field Mapping: Satellite Gravity Gradiometry versus Satellite–to–Satellite Tracking Techniques, presented at XIX Assembly for the IUGG, Vancouver, Canada

Ilk, K.H., R. Sigl and M. Thalhammer (1990), Regional Gravity Field Recovery From Gradiometer Measurements, Workpackage 440, Study on Precise Gravity Field Determination Methods and Mission Requirements, Final Report (Phase 2), ESA Contract no. 8153/88/F/FL

Jekeli, C. (1978), An Investigation of two Models for the Degree Variances of Global Covariance Functions, Department of Geodetic Science, report **275**, Ohio State University, Columbus

Jenkins, G.M. and D.G. Watts (1968), *Spectral Analysis and its Applications*, Holden–Day, Inc., Oakland, California

Jung, K. (1961), *Schwerkraftverfahren in der Angewandten Geophysik*, Akademische Verlaggesellschaft Geest & Portig K.–G, Leipzig

Kaplan, M.H. (1976), *Modern Spacecraft Dynamics & Control*, John Wiley & Sons, New York

Kaula, W.M. (1966), *Theory of Satellite Geodesy*, Blaisdell Publishing Company, Waltham, Massachusetts

Koop, R. and D. Stelpstra (1989), On the Computation of the Gravitational Potential and its First and Second Order Derivatives, *Manuscripta Geodaetica*, **14**, 373–382

Koop, R., E.J.O. Schrama, R. Rummel and M. van Gelderen (1989), Gravity Field Recovery Performance, Aristoteles Add-On Study, Part 1, Dornier W.P. no. 2410/1

Koop, R. and D. Stelpstra (1991), Potential Coefficient Recovery from the CSR Set of Simulated Satellite Gradiometry Observations, *Geophysical Research Letters*, **18**, 10, 1897–1900

Kopejkin, S.M. (1991), Relativistic Manifestations of Gravitational Fields in Gravimetry and Geodesy, *Manuscripta Geodaetica*, **16**, 301–312

Lambeck, K. (1990), Aristoteles: An ESA Mission to Study the Earth's Gravity Field – Impact of a High–Resolution Gravity–Field Mission for Planet Earth, *ESA Journal*, 1, **14**, 1–21

Marsh, J.G., F.J. Lerch, B.H. Putney, T.L. Felsentreger, B.V. Sanchez, S.M. Klosko, G.B. Patel, J.W. Robbins, R.G. Williamson, T.E. Engelis, W.F. Eddy, N.L. Chandler, D.S. Chinn, S. Kapoor, K.E. Rachlin, L.E. Braatz and E.C. Pavlis (1989), The GEM–T2 Gravitational Model, NASA Technical Memorandum 100746

Marussi, A. (1982), On the Structure of the Tidal Field, in: Geodesy and Global Geodynamics, Lectures of the 3rd Int. Summer School on Geodesy and Global Geodynamics, H. Moritz, H. Sünkel (Eds.), Institut für Theoretische Geodäsie, Graz, Austria

Mashhoon, B., H.J. Paik and C.M.Will (1989), Detection of the Gravitomagnetic Field using an Orbiting Superconducting Gravity Gradiometer. Theoretical Principles., *Physical Review D, Particles and Fields*, **39**, 10, 2825–2838

McConnell, A.J. (1957), *Applications of Tensor Analysis*, Dover Publications, New York

Meissl, P. (1971), A Study of Covariance Functions related to the Earth's Disturbing Potential, Department of Geodetic Science, report **151**, Ohio State University, Columbus

Minkowski, H. (1952), Space and Time, in: *The Principle of Relativity*, by A. Einstein, H.A. Lorentz, H.Weyl and H. Minkowski, Dover Publications, New York

Misner, C.W., K.S. Thorne and J.A. Wheeler (1973), *Gravitation*, W.H. Freeman, New York

Moritz, H. (1968), Kinematical Geodesy, DGK **A–59**, München

Moritz, H. (1980), *Advanced Physical Geodesy*, H. Wichman Verlag, Karlsruhe

Moon, P. and D.E. Spencer (1986), *Theory of Holors. A Generalization of Tensors*, Cambridge University Press, Cambridge

NASA (1987), Geophysical and Geodetic Requirements for Global Gravity Field Measurements 1987–2000, Report of a Gravity workshop, Colorado Springs

Neyman, Y.M. (1985), Improperly posed problems in geodesy and methods of their solution, in: *Local Gravity Field Approximation*, proceedings of the Beijing International Summer School, Ed. K.P. Schwarz, Publication 60003, Calgary, Canada

Ohanian, H. (1976), *Gravitation and Spacetime*, W.W. Norton & Company, New York

O'Neill, B. (1966), *Elementary Differential Geometry*, Academic Press, New York

Oppenheim, A.V., A.S. Willsky and I.T. Young (1983), *Signals and Systems*, Prentice–Hall International, London

Paik, H.J. and J–P. Richard (1986), Development of a Sensitive Superconducting Gravity Gradiometer for Geological and Navigational Application, NASA Contractor Report 4011

Paik, H.J. (1989), Tests of General Relativity in Earth Orbit using a Superconducting Gravity Gradiometer, *Adv. Space Res.*, **9**, no. 9, 41–50

Papoulis, A. (1965), *Probability, Random Variables and Stochastic Processes*, McGraw–Hill Kogakusha, Tokyo

Pond, S. and G.L. Pickard (1983), *Introductory Dynamical Oceanography*, 2nd ed., Pergamon Press, Oxford

Rapp, R.H. (1977), Potential Coefficient Determinations from 5° Terrestrial Gravity Data, Department of Geodetic Science, report **251**, Ohio State University, Columbus

Rapp, R.H. (1981), The Earth's Gravity Field to Degree and Order 180 using Seasat Altimeter Data, Terrestrial Gravity Data, and other Data, Department of Geodetic Science, report **322**, Ohio State University, Columbus

Rapp, R.H. and J.Y. Cruz (1986), Spherical Harmonic Expansion of the Earth's Gravitational Potential to Degree 360 using 30' Mean Anomalies, Department of Geodetic Science, report **376**, Ohio State University, Columbus

Rapp, R.H. (1989), Signals and accuracies to be expected from a satellite gradiometer mission, *Manuscripta Geodaetica*, **14**, 36–42

Reed, G.B. (1973), Application of Kinematical Geodesy for Determining the Short Wave Length Components of the Gravity Field by Satellite Gradiometry, Department of Geodetic Science, report **201**, Ohio State University, Columbus

Ries, J.C., C. Huang, M.M. Watkins and B.D. Tapley (1990), The Effects of General Relativity on Near–Earth Satellites, in: *Astrodynamics 1989; Proceedings of the AAS/AIAA Astrodynamics Conference*, held August 7–10, 1989, Stowe, Vermont, *Advances in the Astronautical Sciences*, **71**, 85–93

Robbins, J.W. (1985), Least Squares Collocation applied to Local Gravimetric Solutions from Satellite Gravity Gradiometer Data, Department of Geodetic Science, report **368**, Ohio State University, Columbus

Rummel, R., K.P. Schwarz and M. Gerstl (1979), Least Squares Collocation and Regularization, *Bulletin Géodésique*, **53**, 343–361

Rummel, R. (1985a), From the Observational Model to Gravity Parameter Estimation, in: *Local Gravity Field Approximation*, proceedings of the Beijing International Summer School, Ed. K.P. Schwarz, Publication 60003, Calgary, Canada

Rummel, R. (1985b), Satellite Gradiometry – a Promising Concept since 25 Years, presented at Spring Meeting of AGU, Baltimore, Maryland

Rummel, R. (1985c), Satellitengradiometrie, *ZFV*, **6**, 242–257

Rummel, R. and O.L. Colombo (1985), Gravity Field Determination from Satellite Gradiometry, *Bulletin Géodésique*, **59**, 233–246

Rummel, R. (1986), Satellite Gradiometry, in: *Lecture Notes in Earth Sciences*, **7**, *Mathematical and Numerical Techniques in Physical Geodesy*, H. Sünkel (Ed.), Springer–Verlag, Berlin Heidelberg New York

Rummel, R. (1989a), Aristoteles: Surface Gravity from Space Geometry, in: Proceedings of the Italian Workshop on the European Solid–Earth Mission Aristoteles, Aeritalia, Trevi

Rummel, R. (1989b), SGG Principles and State of the Art, Workpackage 130, Study on Precise Gravity Field Determination Methods and Mission Requirements, Final Report, ESA Contract no. 7521/87/F/FL

Rummel, R., P. Teunissen and M. van Gelderen (1989), Uniquely and Overdetermined Geodetic Boundary Value Problems by Least Squares, *Bulletin Géodésique*, **63**, 1–33

Rummel, R. (1991), Physical Geodesy, Lecture Notes, Delft University of Technology, Faculty of Geodetic Engineering, Delft

Rummel, R. and M. van Gelderen (1992), Spectral Analysis of the Full Gravity Tensor, *Geophysical Journal International*, **111**, 159–169

Savet, P.H. (1969), Gravity Field Exploration by a new Gradient Technique, *J. Spacecraft*, **6**, 710–716

Schouten, J.A. (1954), *Ricci Calculus*, 2nd ed., Springer–Verlag, Berlin

Schrama, E.J.O. (1984), Orbit Integration based upon Interpolated Gravitational Gradients, graduate thesis, Delft University of Technology, Faculty of Geodetic Engineering, Delft

Schrama, E.J.O. (1989), The Role of Orbit Errors in Processing of Satellite Altimeter Data, Netherlands Geodetic Commission, Publications on Geodesy, New Series, **33**

Schrama, E.J.O. (1990), Gravity Field Error Analysis: Application of GPS Receivers and Gradiometers on Low Orbiting Platforms, NASA Technical Memorandum 100769

Schrama, E.J.O. (1992), Relativistic effects on motion of near–Earth satellites, internal report

Schutz, B.E., B.D. Tapley, J.B. Lundberg and P. Halamek (1987), Simulation of a Geopotential Mission for Gravity Studies, *Manuscripta Geodaetica*, **12**, 51–63

Schutz, B.E., J.B. Lundberg, L.K. White and P.G. Antreasian (1988), Developments in the Simulation of a Geopotential Research Mission, proceedings of the AIAA/AAS Astrodynamics Conference, Minneapolis, Minnesota

Schwarz, K.P. (1985), Data types and their spectral properties, in: *Local Gravity Field Approximation*, proceedings of the Beijing International Summer School, Ed. K.P. Schwarz, Publication 60003, Calgary, Canada

SESAME (1986), Proceedings of an ESA Special Workshop on Solid Earth Science & Application Mission for Europe, ESA SP–1080

Sneeuw, N.J. (1991a), Inclination Functions. Group theoretical Background and a Recursive Algorithm, Faculty of Geodetic Engineering/Dept. of Mathematical and Physical Geodesy, report **91.2**, Delft University of Technology

Sneeuw, N. (1991b), Non–singular Cross–track Derivatives of the Gravitational Potential using Rotated spherical Harmonics, proceedings of the XX General Assembly of the IUGG, IAG Symposium G3, Vienna

Soffel, M., E. Gill, H. Ruder and M. Schneider (1987), Relativistic Gradiometry, proceedings of the International Association of Geodesy Symposia, Vancouver, Canada, August 10–22, 1987

Soffel, M.H. (1989), *Relativity in Astrometry, Celestial Mechanics and Geodesy*, Springer–Verlag, Berlin New York Heidelberg

Soffel, M.H. (1990), Lectures on Post–Newtonian Theory, Editorial de la Universidad de Costa Rica, San José

Sokolnikoff, I.S. (1951), *Tensor Analysis. Theory and Applications.*, John Wiley & Sons, New York

Spakman, W. (1988), Upper Mantle Delay Time Tomography, *Geologica Ultraiectina*, **53**, Utrecht

Stelpstra, D. (1990), Gravity Recovery from Satellite Gradiometry, graduate thesis, Delft University of Technology, Faculty of Geodetic Engineering, Delft

Stewart, R., L.–L. Fu and M. Lefebvre (1986), Science Opportunities from the Topex/ Poseidon Mission, JPL Publication 86–18, NASA, Pasadena, California

Synge, J.L. (1960), *Relativity: The General Theory*, North–Holland Publ. Comp., Amsterdam

Tapley, B.D. (1989), Fundamentals of Orbit Determination, in: *Lecture Notes in Earth Sciences*, **25**, *Theory of Satellite Geodesy and Gravity Field Determination*, F. Sansò, R. Rummel (Eds.), Springer–Verlag, Berlin Heidelberg New York

Theiß, D.S. (1984), Neue gravitative Effekte rotierender Massen. Möglichkeiten für weitere Tests der Allgemeinen Relativitätstheorie, Inaugural–Dissertation, Universität Köln, Köln

Thirring, H. and J. Lense (1918), Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie, *Phys. Z.*, **19**, 156–163

Tikhonov, A.N. and V.Y. Arsenin (1963), *Solutions of Ill–posed Problems*, John Wiley & Sons, New York

Touboul, P., A. Bernard, F. Barlier and C. Berger (1991), Air Drag Effect on Gradiometer Measurements, *Manuscripta Geodaetica*, **16**, 73–91

Tscherning, C.C. and R.H. Rapp (1974), Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations and Deflections of the Vertical implied by Anomaly Degree Variances, Department of Geodetic Science, report **208**, Ohio State University, Columbus

Tscherning, C.C. (1976), Comparison of the Second–Order derivatives of the Normal Potential Based on the Representation by a Legendre series, *Manuscripta Geodaetica*, **2**, 71–92

Tscherning, C.C., R. Forsberg and M. Vermeer (1990), Methods for Regional Gravity Field Modelling from SST and SGG Data, Report of the Finnish Geodetic Institute **90:2**, Helsinki

Vermeer, M. (1991), Geoid Recovery at 0.5 Degree Resolution from Global Satellite Gradiometry Data Sets, in: Determination of the Geoid. Present and Future, IAG Symposium No. 106, Eds. R.H. Rapp, F. Sansò, Springer–Verlag, New York Berlin Heidelberg

Visser, P.N.A.M. (1992), The Use of Satellites in Gravity Field Determination and Model Adjustment, PhD Thesis, Delft University Press, Delft

Wagner, C.A. (1983), Direct Determination of Gravitational Harmonics from low–low GRAVSAT Data, *Journal of Geophysical Research*, **88**, 309–321

Wagner, C.A. (1989), Summer School Lectures on Satellite Altimetry, in: *Lecture Notes in Earth Sciences*, **25**, *Theory of Satellite Geodesy and Gravity Field Determination*, F. Sansò, R. Rummel (Eds.), Springer–Verlag, Berlin Heidelberg

Wichiencharoen, C. (1985), Recovery of 1°–mean Anomalies in a Local Region from a Low–Low Satellite to Satellite Tracking Mission, Department of Geodetic Science, report **363**, Ohio State University, Columbus

Will, C. (1981), *Theory and Experiment in Gravitational Physics*, Cambridge University Press, Cambridge

Woodhouse, J.H. and A.M. Dziewonski (1984), Mapping the Upper Mantle: Three–Dimensional Modelling of Earth Structure by Inversion of Seismic Waveforms, *Journal of Geophysical Research*, **89**, B7, 5953–5986

Wunsch, C. (1992), Physics of the Ocean Circulation, Lecture Notes of the International Summerschool of Theoretical Geodesy on "Satellite Altimetry in Geodesy and Oceanography", Trieste

Xu, P.L. (1991), Least Squares Collocation with Incorrect Prior Information, *ZFV*, **116**, 266–273

Xu, P.L. and R. Rummel (1992), The Value of Minimum Norm Estimation of Geopotential Fields, *Geophysical Journal International*, **111**, 170–178