7 Space methods

7.1 Introduction

The subject of this chapter is the use of satellite observations for determining features of the gravity field and the figure of the earth. Only the barest essentials can be presented within the scope of a chapter. The reader will find more information in special textbooks such as Hofmann-Wellenhof et al. (2001), Montenbruck and Gill (2001), and Seeber (2003).

Historical remarks
Immediately after the first launch of artifical satellites (Sputnik 1957, Explorer 1958), their use for geodetic purposes was initiated, and by now the Global Positioning System (GPS) has become the most important method for a fast and precise determination of geodetic positions (see Sect. 5.3). Historically, the first observational methods were intended to determine the spatial direction and the distance to the satellite. Most of these methods are now obsolete, but some principles may be still useful.

Directions
They may be measured by photographing the satellite against the background of stars, or by means of radio waves transmitted from the satellite, using the principle of interference. Photography can only achieve an accuracy of about 0.2 arc seconds and is not used any more in its original sense. The principle of the photographic method was as follows. On the photographic plate, the image of the satellite is surrounded by images of stars. The directions to the surrounding stars are defined by their right ascensions α and declinations δ, which are known from astronomy. Therefore, by interpolation we find the right ascension and declination of the satellite representing the desired direction. This technique is now obsolete.

Ranges
They are measured by radar or by laser. Radar is used for measuring ranges to space probes orbiting in the solar system, which is important to space sciences rather than to geodesy. Lunar Laser Ranging (LLR) and Satellite Laser Ranging (SLR) are useful for determining the earth rotation parameters because of their high (subcentimeter) accuracy; however, their use is restricted to a limited number of fundamental stations.

Range rates

This measurement quantity is found by observing the Doppler effect with radio waves transmitted from a satellite. It is still used within GPS and in satellite-to-satellite tracking (SST).

Satellite altimetry

Here a short-wave electronic ray is sent, from a satellite flying over the oceans, vertically down to the ocean surface, reflected there and received by the satellite again. The measured travel time immediately gives the height H of the satellite above the ocean surface. Knowing the orbital position of the satellite with respect to the global reference system, we can compute the satellite height h above the ellipsoid. Then the difference $h - H$ is the geoidal height N. This is the case if the sea surface is assumed to coincide with the geoid. In reality, because of ocean currents, etc., both surfaces are separated by the "sea surface topography", which may reach the order of 1 m and is interesting to oceanography. It can be determined if an accurate ocean geoid is known from the gravitational field.

The principles of these methods are illustrated in Fig. 7.1, where **e** indicates the direction observation, s between tracking station and satellite refers to the range measurement, and, accordingly, ds/dt corresponds to Doppler observation, whereas ds/dt between the two satellites is obtained by SST; finally, H is measured by satellite altimetry.

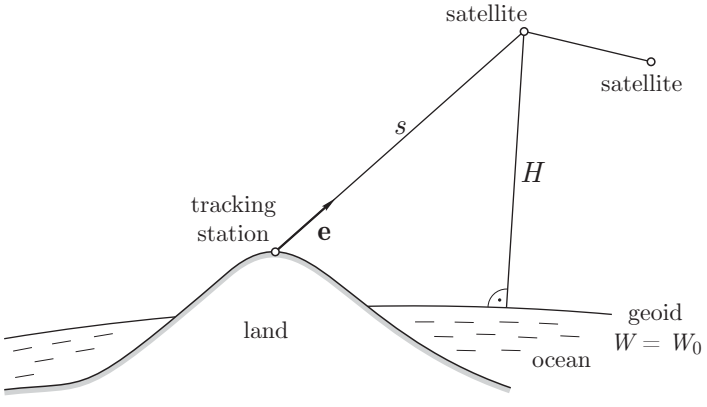

Fig. 7.1. Principles of satellite techniques

7.2 Satellite orbits

The first spectacular result from satellite observations, well advertised by NASA around 1960, was the "discovery" that "the earth is not an ellipsoid but rather shaped like a pear". This pear shape is caused by the spherical harmonic J_3. Its effect, at the North and South Poles, is on the order of 30 m, by three orders of magnitude less than the ellipticity coming from J_2, whose linear effect $a - b$ is about 20 km (!).

The first real result, also found around 1960, was a dramatic improvement in the accuracy of the flattening f itself, which lead to a change from 1/297.3, generally believed before, to 1/298.25, corresponding to a linear improvement of the earth size of about 70 m!

The earth's flattening causes the largest but not the only deviation of the earth gravitational field from that of a homogeneous sphere. Generally, the gravitational potential can be expanded into a series of spherical harmonics according to Sect. 2.5, Eq. (2–78):

$$V = \frac{G\,M}{r}\left\{1 - \sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^n J_n\,P_n(\cos\vartheta) + \sum_{n=2}^{\infty}\sum_{m=1}^{n}\left(\frac{a}{r}\right)^n \left[C_{nm}\,\cos m\lambda + S_{nm}\,\sin m\lambda\right] P_{nm}(\cos\vartheta)\right\}. \tag{7–1}$$

Here the terms containing J_n are the zonal harmonics, and those containing S_{nm} and C_{nm} are the tesseral harmonics.

The former notations $J_{nm} = -C_{nm}$ and $K_{nm} = -S_{nm}$ are not used any more for the *tesseral* harmonic coefficients; for the *zonal* harmonics, the use of J_n has prevailed so far, but also $C_{n0} = -J_n$ is being used.

Considering the moon, the only term of appreciable influence is J_2, which represents the flattening. Artificial satellites are, compared to the moon, much closer to the earth; typical heights above ground of a geodetically used satellite range from some 300 km up to 20 000 km. Hence, they are also influenced by harmonics other than J_2 and can, therefore, be used to determine harmonics of low degree. For this purpose, we must study the effect of gravitational disturbances on the orbits of close satellites.

Before we can do this, we must briefly review the theory of an undisturbed orbit, which means that the gravitational potential has the form

$$V = \frac{GM}{r}, \tag{7–2}$$

all C's and S's being zero. This represents the gravitational field of a point mass or a homogeneous sphere. Then the motion of a satellite is described

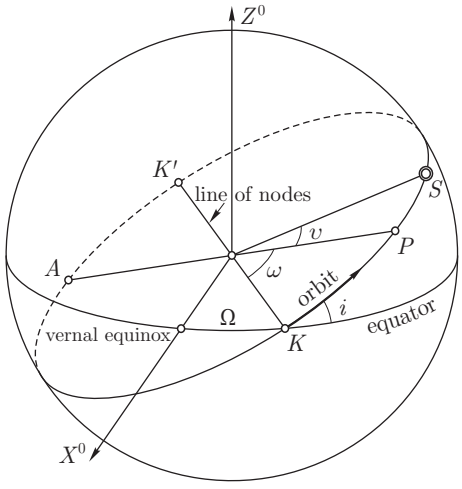

Fig. 7.2. Satellite orbit as projected onto a unit sphere

by *Kepler's three laws for planetary motion*. Satellites with parabolic or hyperbolic orbits are of no interest in this context.

According to *Kepler's first law*, the orbit is an ellipse of which the center of the earth occupies one focus. The position of the orbit in space is defined by the six *orbital elements*:

$$\begin{array}{ll} a & \text{semimajor axis,} \\ e & \text{eccentricity,} \\ i & \text{inclination,} \\ \Omega & \text{right ascension of the node,} \\ \omega & \text{argument of perigee,} \\ T & \text{time of perigee passage.} \end{array} \tag{7–3}$$

If a and b are the semiaxes of the orbital ellipse (there is no danger of confusion with those of the terrestrial ellipsoid!), then the eccentricity is defined by

$$e = \frac{\sqrt{a^2 - b^2}}{a}\,. \tag{7–4}$$

Figure 7.2 shows the projection of the orbit onto a geocentric unit sphere, where P is the perigee, A the apogee, K is the ascending node, K' the descending node, S is the instantaneous position of satellite. The line of nodes is the intersection of the orbital plane with the plane of the equator; it connects the ascending node K and the descending node K'. The right ascension of the node, Ω, is the angle between the line of nodes and the direction to the vernal equinox. The symbol Ω is also called longitude of

the node, but in conformity with astronomical terminology it is the right ascension of the (ascending) node. The major axis of the orbit intersects the orbital ellipse at the perigee P, the position where the satellite is closest to the earth, and at the apogee A, where the satellite is farthest away. The angle ω between the nodes and the major axis is the argument of perigee.

The angular distance of the satellite S from perigee is called *true anomaly* and denoted by v; it is a function of time. Note that this strange name comes from the history of astronomy; there is nothing anomalous with it!

The equation of the orbital ellipse may be written

$$r = \frac{p}{1 + e\,\cos v}\,, \tag{7–5}$$

where r is the distance of the satellite from the earth's center of mass and

$$p = \frac{b^2}{a} = a\,(1 - e^2) \tag{7–6}$$

is the length of the radius vector r for $v = 90°$. The radius vector r and the true anomaly v form a pair of polar coordinates in the orbital plane, and (7–5) is the well-known polar equation of an ellipse. See Fig. 7.3 for an illustration of these quantities, where F, the focal point, is the earth's center of mass.

According to *Kepler's second law*, the area of the elliptical sector swept by the radius vector r between any two positions of the satellite is proportional to the time it takes the satellite to pass from one position to the other. In other words, the time rate of change of the area swept by the radius vector is constant. Since the element of area of a sector in polar coordinates r and

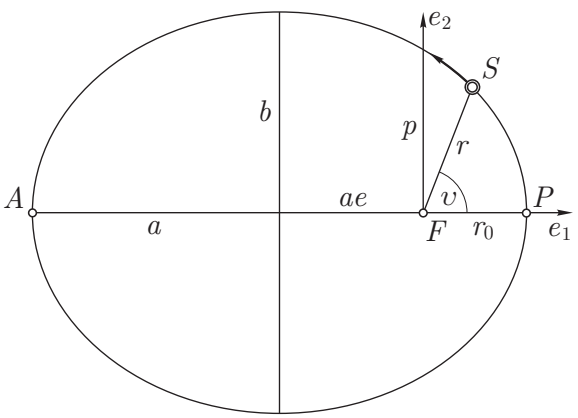

Fig. 7.3. Orbital ellipse

v is $\frac{1}{2}r^2 dv$, this law may be formulated mathematically as

$$r^2 \frac{dv}{dt} = \sqrt{GM\,a\,(1-e^2)}\,, \tag{7–7}$$

where the constant has already been given its proper value.

Kepler's third law reads

$$n^2 a^3 = GM\,, \tag{7–8}$$

where the satellite mass has been neglected and where

$$n = \frac{2\pi}{P} \tag{7–9}$$

is the "mean motion" (mean angular velocity) of the satellite, P being its period.

So far we have assumed that all J_n, C_{nm}, and S_{nm} in (7–1) are zero. This is not true because of the irregularities of the earth's gravitational field, even though these coefficients are small. Therefore, the satellite is subject to small perturbing forces. We may still consider the satellite orbit as an ellipse, but then the parameters of this ellipse, the orbital elements, will no longer be constant but will change slowly. At each instant, this *osculating ellipse* will be slightly different. It is defined as follows. Imagine that at the instant under consideration all perturbing forces suddenly vanish. Then the satellite will continue its motion along an exact ellipse; this is the osculating ellipse.

If we resolve the total perturbing force into rectangular components S, T, and W, where S is directed along the radius vector, W is normal to the orbital plane, and T is normal to S and W – note that this notation follows astronomical usage; there is no relation to the geodetic use of T and W for potentials! –, then the time rate of change of the orbital parameters can be expressed in terms of these components:

$$\begin{aligned}
\dot{a} &= \frac{2a^2}{b}\sqrt{\frac{a}{G\,M}}\left(e\,S\sin v + \frac{p}{r}\,T\right),\\
\dot{e} &= \frac{b}{a}\sqrt{\frac{a}{G\,M}}\left[S\sin v + \left(\frac{r+p}{r}\,\cos v + \frac{e\,r}{p}\right)T\right],\\
\dot{i} &= \frac{r}{b}\sqrt{\frac{a}{G\,M}}\,W\cos(\omega+v)\,,\\
\dot{\Omega} &= \frac{r}{b}\sqrt{\frac{a}{G\,M}}\,W\,\frac{\sin(\omega+v)}{\sin i}\,,\\
\dot{\omega} &= \frac{b}{a}\sqrt{\frac{a}{G\,M}}\left[-\frac{1}{e}\,S\,\cos v + \frac{r+p}{e\,p}\,T\,\sin v - \frac{r}{p}\,W\sin(\omega+v)\cot i\right].
\end{aligned} \tag{7–10}$$

As usual, $\dot{a}$ denotes da/dt, etc. The derivation of these equations may be found in any textbook on celestial mechanics, e.g., Plummer (1918: p. 151), Brouwer and Clemence (1961: p. 301), and Seeber (2003: Sect. 3.2.1.3), who uses the symbols K_1, K_2, K_3 instead of W, S, R.

7.3 Determination of zonal harmonics

The effect of the zonal harmonics on satellite orbits is much greater than that of the tesseral harmonics. Only zonal harmonics (J_2, J_3, J_4, ...) will give observable variations of the orbital elements themselves. The tesseral harmonics cause oscillatory disturbances that rapidly change their sign, whereas the effect of the zonal harmonics is cumulative. For this reason, we consider first the effect of zonal harmonics, that is, the effect of those independent of longitude λ. Hence we set

$$V = \frac{G\,M}{r} + R\,, \tag{7–11}$$

where the *perturbing potential*

$$R = -\frac{G\,M}{a_\mathrm{e}} \sum_{n=2}^{\infty} \left(\frac{a_\mathrm{e}}{r}\right)^{n+1} J_n\, P_n(\cos\vartheta) \tag{7–12}$$

is a function of r and ϑ only. Note that the main difference between the perturbing potential R of celestial mechanics and the disturbing potential T of physical geodesy is that R, but not T, also incorporates the effect of the flattening through J_2. There are also other perturbing forces acting on a satellite, such as the resistance of the atmosphere (atmospheric drag), radiation pressure exerted by the sunlight, etc. These nongravitational perturbances must be taken into account separately and will not be considered here.

Note that the equatorial radius of the earth (the semimajor axis of the terrestrial ellipsoid) has been denoted by a_e, in order to distinguish it from a, which now denotes the semimajor axis of the orbital ellipse. This notation will be used in what follows.

Since S is the component of the perturbing force along the radius vector, we have

$$S = \frac{\partial R}{\partial r}\,. \tag{7–13}$$

The components of the perturbing force along the meridian and the prime vertical are

$$-\frac{1}{r}\,\frac{\partial R}{\partial \vartheta} \quad \text{and} \quad \frac{1}{r\sin\vartheta}\,\frac{\partial R}{\partial \lambda}\,. \tag{7–14}$$

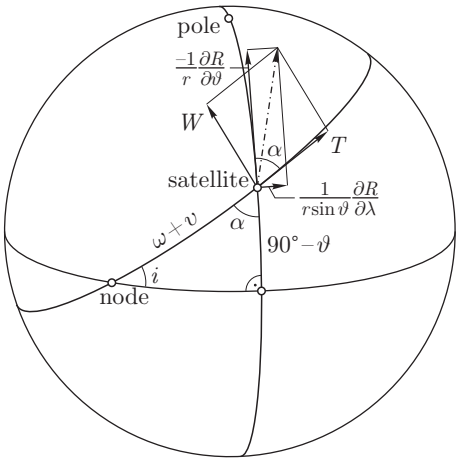

Fig. 7.4. Components of the perturbing force

The components T and W are obtained from them by a plane rotation (Fig. 7.4):

$$\begin{aligned} T &= -\frac{1}{r}\frac{\partial R}{\partial \vartheta}\cos\alpha + \frac{1}{r\,\sin\vartheta}\frac{\partial R}{\partial \lambda}\sin\alpha\,, \\ W &= -\frac{1}{r}\frac{\partial R}{\partial \vartheta}\sin\alpha - \frac{1}{r\,\sin\vartheta}\frac{\partial R}{\partial \lambda}\cos\alpha\,. \end{aligned} \tag{7–15}$$

From the rectangular spherical triangle in Fig. 7.4 it follows that

$$\cos\alpha = \frac{\cos(\omega+v)\,\sin i}{\sin\vartheta}\,, \qquad \sin\alpha = \frac{\cos i}{\sin\vartheta}\,, \tag{7–16}$$

so that finally

$$\begin{aligned} T &= -\frac{\cos(\omega+v)\,\sin i}{r\,\sin\vartheta}\frac{\partial R}{\partial \vartheta} + \frac{\cos i}{r\,\sin^2\vartheta}\frac{\partial R}{\partial \lambda}\,, \\ W &= -\frac{\cos i}{r\,\sin\vartheta}\frac{\partial R}{\partial \vartheta} - \frac{\cos(\omega+v)\,\sin i}{r\,\sin^2\vartheta}\frac{\partial R}{\partial \lambda}\,. \end{aligned} \tag{7–17}$$

We have included $\partial R/\partial\lambda$ because of the presence of longitude-dependent tesseral harmonics in the general case (see Sect. 7.5). In our present case, where R is given by (7–12), $\partial R/\partial\lambda$ is zero.

Now we must differentiate (7–12) with respect to r and ϑ, compute the components S, T, W from Eqs. (7–13) and (7–17), and substitute them into the system (7–10). In this way, we can express the rates of change $\dot{a}$, $\dot{e}$, ... of the orbital elements in terms of the coefficients J_2, J_3, J_4, ... We

cannot, however, observe these rates of change directly. Rather, we observe the changes of the orbital elements after several revolutions. The changes after *one* revolution, with period P, are

$$\Delta a = \int_{t_0}^{t_0+P} \dot{a}\, dt\,, \quad \Delta e = \int_{t_0}^{t_0+P} \dot{e}\, dt\,, \quad \Delta i = \int_{t_0}^{t_0+P} i\, dt\,, \quad \text{etc.} \quad (7\text{–}18)$$

The t_0 is an arbitrary "epoch" (instant of time). In order to perform these integrations, we must express $\dot{a}$, $\dot{e}$, ... in terms of one independent variable. For this independent variable, we may take the time t or the true anomaly v. The second possibility will be adopted here.

The polar distance ϑ is expressed as a function of v through the relation

$$\cos\vartheta = \sin(\omega + v)\, \sin i\,, \quad (7\text{–}19)$$

which follows from the rectangular spherical triangle in Fig. 7.4. The radius vector r is also a function of v according to (7–5). Finally, Kepler's second law (7–7) furnishes the relation between v and the time t:

$$\frac{dt}{dv} = \frac{r^2}{\sqrt{G\,M\,a\,(1-e^2)}}\,. \quad (7\text{–}20)$$

Hence, we may change the integration variable from t to v, obtaining, for instance,

$$\Delta a = \int_{t_0}^{t_0+P} \dot{a}\, dt = \int_{v=0}^{2\pi} \frac{da}{dv}\, dv\,, \quad (7\text{–}21)$$

where

$$\frac{da}{dv} = \frac{da}{dt}\,\frac{dt}{dv} = \frac{r^2}{\sqrt{G\,M\,a\,(1-e^2)}}\,\dot{a}\,. \quad (7\text{–}22)$$

Analogous formulas result for the other orbital elements.

After performing all these operations, which are lengthy but not too

difficult, we find

$$
\begin{aligned}
\Delta a &= 0\,,\\
\Delta e &= -\frac{1-e^2}{e}\,\tan i\;\Delta i\,,\\
\Delta i &= 3\pi\, e\left(\frac{a_\mathrm{e}}{p}\right)^3\left(1-\frac{5}{4}\,\sin^2 i\right)\cos i\,\cos\omega\;J_3\\
&\quad+\frac{45}{16}\,\pi\, e\left(\frac{a_\mathrm{e}}{p}\right)^4\left(1-\frac{7}{6}\,\sin^2 i\right)\sin 2i\,\sin 2\omega\;e\,J_4\;\cdots\,,\\
\Delta\Omega &= -3\pi\left(\frac{a_\mathrm{e}}{p}\right)^2\cos i\;J_2\\
&\quad+3\pi\left(\frac{a_\mathrm{e}}{p}\right)^3\left(1-\frac{15}{4}\,\sin^2 i\right)\cot i\,\sin\omega\;e\,J_3\\
&\quad+\frac{15}{2}\,\pi\left(\frac{a_\mathrm{e}}{p}\right)^4\left(1-\frac{7}{4}\,\sin^2 i\right)\cos i\;J_4\;\cdots\,,\\
\Delta\omega &= 6\pi\left(\frac{a_\mathrm{e}}{p}\right)^2\left(1-\frac{5}{4}\,\sin^2 i\right)\;J_2\\
&\quad+3\pi\left(\frac{a_\mathrm{e}}{p}\right)^3\left(1-\frac{5}{4}\,\sin^2 i\right)\sin i\,\sin\omega\;e\,J_3\\
&\quad-15\pi\left(\frac{a_\mathrm{e}}{p}\right)^4\left[\left(1-\frac{31}{8}\,\sin^2 i+\frac{49}{19}\,\sin^4 i\right)\right.\\
&\quad+\left.\left(\frac{3}{8}-\frac{7}{16}\,\sin^2 i\right)\sin^2 i\,\cos 2\omega\right]J_4\;\cdots\,.
\end{aligned}
\tag{7–23}
$$

Terms of the order of $e^2 J_3$ and $e^2 J_4$, which are very small, have been neglected in these equations. The proportionality of Δe and Δi is more or less accidental: it applies only with respect to long-periodic disturbances; $\dot{e}$ and di/dt themselves are not proportional. The quantity p is defined by (7–6); it is hardly necessary to repeat that a, p, e, etc., refer to the orbital ellipse and not to the terrestrial ellipsoid, of which a_e is the equatorial radius.

By integrating over one revolution, we have removed the *short-periodic* terms of periods P, $2P$, $3P$, ..., such as $\cos v$, $\cos 2v$, etc. What remains are *secular* terms, which are constant for one revolution and increase steadily with the number of revolutions, and the *long-periodic* terms, which change very slowly with time in a periodic manner. The argument of perigee ω increases slowly but steadily, so that the perigee of a satellite orbit also

rotates around the earth, but much slower than the satellite itself; a typical period of ω is two months. Therefore, terms containing $\cos\omega$, $\sin\omega$, or $\sin 2\omega$ are called long-periodic.

The first equation of (7–23) shows that the semimajor axis of the orbit does not change secularly or long-periodically. The eccentricity and the inclination undergo long-period, but not secular, variations, whereas Ω and ω change both secularly and long-periodically.

Equations (7–23) are linear in J_2, J_3, J_4, ... For practical applications, nonlinear terms containing J_2^2, $J_2 J_3$, $J_2 J_4$, etc., must also be taken into account, since J_2^2 is of the order of J_4. The derivation of these nonlinear terms is much more difficult, and their expressions are different in the various orbital theories that have been proposed. For these reasons, such expressions will not be given here.

Equations (7–23), supplemented by certain nonlinear terms, can be used to determine coefficients J_2, J_3, J_4, etc. Since the secular or long-periodic variations $\Delta\Omega$, $\Delta\omega$, Δe, Δi are known from observation for a sufficient number of satellites, we obtain equations of the form

$$\begin{aligned} a_2 J_2 + a_3 J_3 + a_4 J_4 + \cdots + a_{22} J_2^2 + a_{23} J_2 J_3 + \cdots &= A\,, \\ b_2 J_2 + b_3 J_3 + b_4 J_4 + \cdots + b_{22} J_2^2 + b_{23} J_2 J_3 + \cdots &= B\,, \\ \vdots \qquad \vdots \qquad \vdots \qquad\qquad \vdots \qquad\quad \vdots \qquad\quad & \vdots \end{aligned} \tag{7–24}$$

which can be solved for J_2, J_3, J_4, ... Since there can be only a finite number of these equations, we must neglect all J_n with n greater than a certain number n_0, which depends on the number of equations available, on their degree of mutual independence, etc. This used to be a difficulty with this method, but it has been overcome long ago by least-squares collocation (Moritz 1980 a: Sect. 21). For details see Schwarz (1976).

From (7–23) it is seen that the coefficients of the J_n depend essentially on the inclination i. It is, therefore, important to use satellites with a wide variety of inclinations, in order to get equations with a high mutual independence.

Now the question arises which orbital elements are to be used for determining the coefficients J_n. The semimajor axis a clearly cannot be used at all. As for the other elements, we must distinguish between coefficients of even and of odd degree n. The even coefficients J_2, J_4, ... can be determined well from the regression of the node, $\Delta\Omega$, and the rotation of perigee, $\Delta\omega$. To see this, inspect (7–23). The even harmonics cause secular disturbances of Ω and ω, which are much larger than the long-periodic effects of the odd coefficients, since J_3, J_5, ... are multiplied by the small eccentricity e.

On the other hand, in Δe and Δi the odd coefficients J_3, J_5, ... have a much larger effect than the even coefficients, which here appear with the small factor e. Therefore, the odd coefficients are determined from Δe or Δi, or from the change of perigee distance $r_0 = FP$ (Fig. 7.3). Since r_0 is the radius vector for $v = 0$, we have from (7–5) and (7–6)

$$r_0 = \frac{p}{1+e} = a\,(1-e)\,, \tag{7–25}$$

so that

$$\Delta r_0 = -a\,\Delta e \tag{7–26}$$

because $\Delta a = 0$. Thus, the variation of perigee distance is proportional to the variation of eccentricity and may be used instead of Δe.

Numerical values

Helmert (1884: p. 472) used the regression of the node of the moon's orbit to determine J_2, which is the only coefficient to have an appreciable effect on it. Note that for $e \doteq 0$ and $p \doteq a \gg a_\mathrm{e}$, the equation for $\Delta\Omega$ in (7–23) becomes

$$\Delta\Omega = -3\pi \left(\frac{a_\mathrm{e}}{a}\right)^2 J_2 \cos i\,. \tag{7–27}$$

Helmert found

$$J_2 = 1086.5 \cdot 10^{-6} \tag{7–28}$$

by averaging two widely different values. This corresponds to a flattening of

$$1/f = 297.8 \pm 2.2\,. \tag{7–29}$$

This value is quite close to the recent results but has a much larger uncertainty.

Reliable values by this method can only be obtained from close artificial satellites. Currently accepted values are, for example,

$$\begin{aligned} J_2 &= 1082.6359 \cdot 10^{-6}\,, \\ J_3 &= -2.5324 \cdot 10^{-6}\,, \\ J_4 &= -1.6198 \cdot 10^{-6}\,, \end{aligned} \tag{7–30}$$

whose standard errors are assumed to be better than $\pm 0.01 \cdot 10^{-6}$. The value for J_2 has been taken from the report of the IAG by Groten (2004), accessible from www.gfy.ku.dk/∼iag/HB2004/part5/51-groten.pdf. J_3 and J_4 are from the recent mission GRACE (see Sect. 7.5).

The most significant geodetic result is the reliable determination of J_2 and, therefore, of the flattening f, around 1/298.25. Already in 1964, the International Astronomical Union (IAU) adopted the value 298.25 corresponding to $J_2 = 1082.7 \cdot 10^{-6}$ (see Sect. 2.11), followed by the IAG International Geodetic Reference Systems 1967 and then 1980, which in the slightly different form of the World Geodetic System 1984 (WGS 84) is standard even today (2005).

7.4 Rectangular coordinates of the satellite and perturbations

We now describe how the rectangular coordinates of the satellite are computed from the orbital elements. Then we will outline how they are affected by the irregularities of the gravity field. These considerations are necessary for the determination of tesseral harmonics from satellite observations.

We introduce an equatorial coordinate system $X^0Y^0Z^0$ that is at rest with respect to the stars. The origin is at the earth's center of mass. The Z^0-axis coincides with its axis of rotation; the X^0Y^0-plane is the equatorial plane. The X^0-axis is the line of intersection of the equatorial plane and the ecliptic (the plane of the earth's orbit around the sun); according to astronomical terminology, it points to the *vernal equinox*. This coordinate system $X^0Y^0Z^0$ is fundamental in spherical astronomy. Note that the directions of the coordinate axes so defined are not completely constant in time. This fact requires certain refinements for which the reader is referred to Moritz and Mueller (1987: Chap. 7). In the present context, we consider the $X^0Y^0Z^0$-system as constant in time.

The relation between the rectangular coordinates of a satellite and the elements of its osculating ellipse (Sect. 7.2) at a certain time is found as follows. Consider Fig. 7.3 and the coordinate system $\mathbf{e}_1$, $\mathbf{e}_2$ defining the orbital plane. Assuming $\mathbf{e}_3$ orthogonal to this plane,

$$r \begin{bmatrix} \cos v \\ \sin v \\ 0 \end{bmatrix} \tag{7–31}$$

is the representation of the satellite in this system. This result may be transformed into the equatorial system $X^0Y^0Z^0$ by a rotation matrix $\mathbf{R}$ and results in a vector denoted as $\mathbf{X}^0 = [X^0,\ Y^0,\ Z^0]$. The transformation is

obtained by

$$\begin{bmatrix} X^0 \\ Y^0 \\ Z^0 \end{bmatrix} = \mathbf{R}\, r \begin{bmatrix} \cos v \\ \sin v \\ 0 \end{bmatrix}, \tag{7–32}$$

where the matrix $\mathbf{R}$ is composed of three successive rotation matrices (see Figs. 7.2 and 7.3) and is given by

$$\begin{aligned} \mathbf{R} &= \mathbf{R}_3\{-\Omega\}\, \mathbf{R}_1\{-i\}\, \mathbf{R}_3\{-\omega\} \\ &= \begin{bmatrix} \cos\Omega\cos\omega & -\cos\Omega\sin\omega & \sin\Omega\sin i \\ -\sin\Omega\sin\omega\cos i & -\sin\Omega\cos\omega\cos i & \\ \sin\Omega\cos\omega & -\sin\Omega\sin\omega & -\cos\Omega\sin i \\ +\cos\Omega\sin\omega\cos i & +\cos\Omega\cos\omega\cos i & \\ \sin\omega\sin i & \cos\omega\sin i & \cos i \end{bmatrix}, \end{aligned} \tag{7–33}$$

see Hofmann-Wellenhof et al. (2001: p. 43). The column vectors of the orthonormal matrix $\mathbf{R}$ are the axes of the orbital coordinate system represented in the equatorial system $\mathbf{X}_i^0$.

Substituting (7–33) into (7–32) and carrying out the multiplication (Montenbruck and Gill 2001: Eq. (2.51)) yields

$$\begin{aligned} X^0 &= r\,[\cos\Omega\,\cos(\omega+v) - \sin\Omega\,\sin(\omega+v)\cos i]\,, \\ Y^0 &= r\,[\sin\Omega\,\cos(\omega+v) + \cos\Omega\,\sin(\omega+v)\cos i]\,, \\ Z^0 &= r\,\sin(\omega+v)\,\sin i\,, \end{aligned} \tag{7–34}$$

where, according to (7–5),

$$r = \frac{a\,(1-e^2)}{1+e\,\cos v}\,. \tag{7–35}$$

This expresses the rectangular coordinates of the satellite in terms of the elements of its osculating orbit, the true anomaly v fixing its position as a function of time.

Since the osculating ellipse does not remain constant, it is convenient to use a fixed *reference orbit* – for instance, the osculating ellipse E_0 at a certain instant t_0, having the elements a_0, e_0, i_0, Ω_0, ω_0, T_0. At a later instant t, the orbital elements will have changed to $a_0 + \Delta_t a$, $e_0 + \Delta_t e$, $i_0 + \Delta_t i$, $\Omega_0 + \Delta_t \Omega$, $\omega_0 + \Delta_t \omega$, $T_0 + \Delta_t T$, which corresponds to an osculating ellipse E_t.

The orbital elements in (7–34) refer to this instantaneous osculating ellipse, so that $a = a_0 + \Delta_t a$, etc. Therefore, the coordinates X^0, Y^0, Z^0 depend on the time in two ways: *explicitly*, through the true anomaly v, and *implicitly*, through the variable elements of the osculating orbit. We eliminate the implicit dependence in the following way. We evaluate (7–34) using the elements a_0, etc., of the fixed reference ellipse. Then the coordinates so obtained depend on the time only explicitly and correspond to a Keplerian motion in space along a fixed ellipse. To convert them into true coordinates X^0, Y^0, Z^0, they must be corrected by $\Delta_t X^0$, $\Delta_t Y^0$, $\Delta_t Z^0$, for which the linear terms of a Taylor expansion of (7–34) give

$$\begin{aligned}
\Delta_t X^0 &= \\
&\frac{\partial X^0}{\partial a}\,\Delta_t a \frac{\partial X^0}{\partial e}\,\Delta_t e + \frac{\partial X^0}{\partial i}\,\Delta_t i + \frac{\partial X^0}{\partial \Omega}\,\Delta_t \Omega + \frac{\partial X^0}{\partial \omega}\,\Delta_t \omega + \frac{\partial X^0}{\partial v}\,\Delta_t v\,, \\
\Delta_t Y^0 &= \\
&\frac{\partial Y^0}{\partial a}\,\Delta_t a \frac{\partial Y^0}{\partial e}\,\Delta_t e + \frac{\partial Y^0}{\partial i}\,\Delta_t i + \frac{\partial Y^0}{\partial \Omega}\,\Delta_t \Omega + \frac{\partial Y^0}{\partial \omega}\,\Delta_t \omega + \frac{\partial Y^0}{\partial v}\,\Delta_t v\,, \\
\Delta_t Z^0 &= \\
&\frac{\partial Z^0}{\partial a}\,\Delta_t a \frac{\partial Z^0}{\partial e}\,\Delta_t e + \frac{\partial Z^0}{\partial i}\,\Delta_t i + \frac{\partial Z^0}{\partial \Omega}\,\Delta_t \Omega + \frac{\partial Z^0}{\partial \omega}\,\Delta_t \omega + \frac{\partial Z^0}{\partial v}\,\Delta_t v\,.
\end{aligned} \tag{7–36}$$

The partial derivatives are readily obtained by differentiating (7–34); note that r is a function of a, e, and v.

In these equations, we have used the perturbation of the true anomaly, $\Delta_t v$, instead of the perturbation of perigee epoch, $\Delta_t T$.

Perturbations expressed in terms of C_{nm} and S_{nm}

The perturbations of the orbital elements are found by integrating (7–10):

$$\Delta_t a = \int_{t_0}^{t} \dot{a}\,dt\,, \qquad \Delta_t e = \int_{t_0}^{t} \dot{e}\,dt\,, \qquad \ldots\,. \tag{7–37}$$

A similar expression can be written for $\Delta_t v$. The components S, T, W of the perturbing force are expressed in terms of J_n, C_{nm}, and S_{nm} using equations (7–12), (7–13), and (7–17), where the perturbing potential

$$\begin{aligned}
R = -\frac{G\,M}{a_\mathrm{e}} \sum_{n=2}^{\infty} \Big(\frac{a_\mathrm{e}}{r}\Big)^{n+1} \Big[& J_n P_n(\cos\vartheta) \\
& - \sum_{m=1}^{n} \left(C_{nm}\cos m\lambda + S_{nm}\sin m\lambda\right) P_{nm}(\cos\vartheta) \Big]
\end{aligned} \tag{7–38}$$

now also contains the tesseral harmonics.

By performing the integrations in (7–37), we obtain equations of the form

$$\begin{aligned} \Delta_t a &= \sum_{n,m} \left(A_{nm} C_{nm} + \bar{A}_{nm} S_{nm} \right) , \\ \Delta_t e &= \sum_{n,m} \left(B_{nm} C_{nm} + \bar{B}_{nm} S_{nm} \right) , \\ &\vdots \qquad \vdots \end{aligned} \tag{7–39}$$

where the coefficients A_{nm}, etc., are functions of the time t and are, as a rule, periodic. Zonal and tesseral harmonics have been combined in (7–39) by setting $J_n = -C_{n0}$ and admitting the value $m = 0$; this practice will be continued in what follows.

The substitution of (7–39) into (7–36) gives the perturbation of the rectangular coordinates X^0, Y^0, Z^0 as functions of the harmonic coefficients $C_{n0} = -J_n$, C_{nm}, and S_{nm} in the form

$$\begin{aligned} \Delta_t X^0 &= \sum_{n,m} \left(L_{nm} C_{nm} + \bar{L}_{nm} S_{nm} \right) , \\ \Delta_t Y^0 &= \sum_{n,m} \left(M_{nm} C_{nm} + \bar{M}_{nm} S_{nm} \right) , \\ \Delta_t Z^0 &= \sum_{n,m} \left(N_{nm} C_{nm} + \bar{N}_{nm} S_{nm} \right) , \end{aligned} \tag{7–40}$$

where again L_{nm}, $\bar{L}_{nm}$, M_{nm}, etc., are functions of the time t.

These perturbations are added to the coordinates computed from (7–34) using the orbital elements of the reference ellipse E_0. In this way, we obtain the rectangular coordinates of the satellite in the form

$$\begin{aligned} X^0 &= X^0(t;\, a_0, e_0, i_0, \Omega_0, \omega_0, T_0;\, C_{nm}, S_{nm}) , \\ Y^0 &= Y^0(t;\, a_0, e_0, i_0, \Omega_0, \omega_0, T_0;\, C_{nm}, S_{nm}) , \\ Z^0 &= Z^0(t;\, a_0, e_0, i_0, \Omega_0, \omega_0, T_0;\, C_{nm}, S_{nm}) \end{aligned} \tag{7–41}$$

as *explicit* functions of the time t, containing as *constant* parameters the orbital element of the reference ellipse E_0 and the gravitational coefficients C_{nm} and S_{nm}. This is the advantage of (7–41) over the system (7–34), which formally is much simpler but depends on the *variable* orbital parameters of the osculating ellipse.

The actual expressions for (7–41) are very complicated. Therefore, we have been satisfied with outlining the procedure, referring the reader for details to the pioneering book by Kaula (1966 a) and to his papers given there.

7.5 Determination of tesseral harmonics and station positions

Zonal harmonics give rise to secular and long-periodic perturbations of the orbital elements a, e, etc. Therefore, their influence can be detected in changes of orbital parameters obtained by integrating over many revolutions of the satellite.

The perturbations due to tesseral harmonics have a much shorter period. The longest period of a harmonic of the order $m = 1$ is one day, for $m = 2$ it is only half a day, etc. Therefore, we must look for another method, which is sensitive enough to detect even short-periodic effects and extracts as much information as possible from the observations.

The observed elements are essentially spatial polar coordinates of the satellite with respect to the observing station: the distance s and the direction as determined by two angles. Corresponding to our coordinate system X^0, Y^0, Z^0 introduced in the preceding section, these two angles are the *right ascension* α and the *declination* δ, whose definition may be seen in Fig. 7.5. The angles α and δ are polar coordinates in three-dimensional space and were obtained by photographing the satellite against the background of stars, as outlined in Sect. 7.1. They are outdated nowadays but retained for geometrical intuition and symmetry. Most important are distances s measured by GPS, radar, or laser. Note that the measurement of the range rate ds/dt of the satellite by means of the Doppler effect is also important for the determination of tesseral harmonics and station positions.

Denoting in the equatorial system $X^0Y^0Z^0$ the rectangular coordinates of the terrestrial station P by X_P^0, Y_P^0, Z_P^0 and of the satellite S by X_S^0, Y_S^0, Z_S^0,

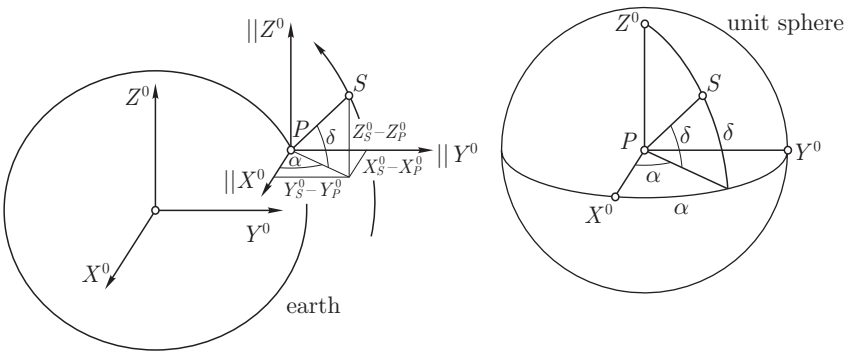

Fig. 7.5. Direction to the satellite defined by right ascension α and declination δ

we find by inspecting Fig. 7.5

$$\begin{aligned} X_S^0 - X_P^0 &= s\,\cos\delta\,\cos\alpha\,, \\ Y_S^0 - Y_P^0 &= s\,\cos\delta\,\sin\alpha\,, \\ Z_S^0 - Z_P^0 &= s\,\sin\delta\,, \end{aligned} \tag{7–42}$$

so that

$$\begin{aligned} \alpha &= \tan^{-1}\frac{Y_S^0 - Y_P^0}{X_S^0 - X_P^0}\,, \\ \delta &= \tan^{-1}\frac{Z_S^0 - Z_P^0}{\sqrt{(X_S^0 - X_P^0)^2 + (Y_S^0 - Y_P^0)^2}}\,, \\ s &= \sqrt{(X_S^0 - X_P^0)^2 + (Y_S^0 - Y_P^0)^2 + (Z_S^0 - Z_P^0)^2}\,. \end{aligned} \tag{7–43}$$

We now compute the rectangular coordinates X_P^0, Y_P^0, Z_P^0 of the observing station P. The system $X^0Y^0Z^0$, being fixed with respect to the stars, rotates with respect to the earth. The coordinates of P in this system are, therefore, functions of time. Let X_P, Y_P, Z_P be the coordinates of P in the usual geocentric coordinate system fixed with respect to the earth. In this system, the Z-axis, coinciding with the Z^0-axis, is the earth's axis of rotation; the X-axis lies in the mean meridian plane of Greenwich, corresponding to the longitude $\lambda = 0°$; and the Y-axis points to $\lambda = 90°$ east. Figure 7.6 shows that

$$\begin{aligned} X_P^0 &= X_P\cos\theta_0 - Y_P\sin\theta_0\,, \\ Y_P^0 &= X_P\sin\theta_0 + Y_P\cos\theta_0\,, \\ Z_P^0 &= Z_P\,. \end{aligned} \tag{7–44}$$

The angle θ_0 is called *Greenwich sidereal time*; its value is

$$\theta_0 = \omega\,t\,, \tag{7–45}$$

where ω is the angular velocity of the earth's rotation. It is proportional to the time t and, in appropriate units, measures it. Thus, absolute Greenwich time is needed to convert the terrestrial coordinates X_P, Y_P, Z_P to the celestial coordinates X_P^0, Y_P^0, Z_P^0 that are required in (7–42) and (7–43).

As a final step, we substitute the station coordinates, as given by (7–44), and the satellite coordinates, as symbolized by (7–41), into (7–43), obtaining expressions of the form

$$\begin{aligned} \alpha &= \alpha(X_P, Y_P, Z_P;\ t;\ a_0, e_0, i_0, \Omega_0, \omega_0, T_0;\ C_{nm}, S_{nm})\,, \\ \delta &= \delta(X_P, Y_P, Z_P;\ t;\ a_0, e_0, i_0, \Omega_0, \omega_0, T_0;\ C_{nm}, S_{nm})\,, \\ s &= s(X_P, Y_P, Z_P;\ t;\ a_0, e_0, i_0, \Omega_0, \omega_0, T_0;\ C_{nm}, S_{nm})\,. \end{aligned} \tag{7–46}$$

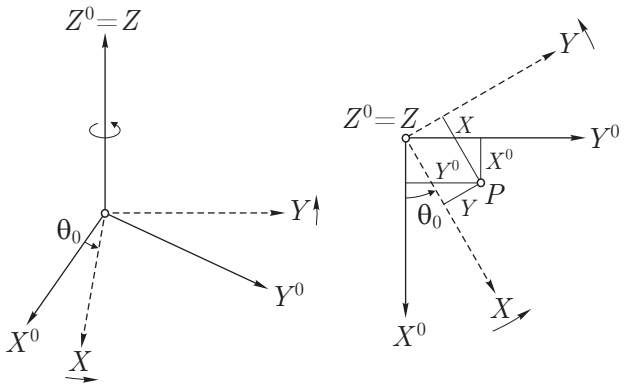

Fig. 7.6. Geocentric coordinate systems $X^0Y^0Z^0$ (celestial) and XYZ (terrestrial)

Besides depending on the station coordinates and the time, they also contain the orbital and gravitational parameters.

Every observation furnishes an equation of type (7–46). Provided we have a sufficient number of such observation equations, we can solve them for the station coordinates X_P, Y_P, Z_P, for the orbital parameters a_0, e_0, etc., of the reference ellipse, and for a certain number of gravitational parameters C_{nm} and S_{nm}. This is the principle of the *orbital method*. In practice, differential formulas will be applied to determine corrections to assume approximate values by means of a least-squares adjustment. Therefore, the actual analytical developments are from the outset directed toward obtaining differential formulas corresponding to (7–46). The substitutions indicated above are, thus, consistently performed in terms of the corresponding differential expressions. In this way we are able to operate with linear equations and to employ that efficient tool of linear analysis, matrix calculus. Simple though the principle of this procedure is, the details when written out are nevertheless so complicated that the reader must again be referred to the literature, e.g., Kaula (1966 a), Montenbruck and Gill (2001). Computer formula manipulation is also used.

Besides these analytical problems, which have been satisfactorily solved, the geodetic application of (7–46) raises difficulties similar in principle to those involved in the determination of zonal harmonics by means of (7–24), but even more serious in practice. Strictly speaking, an infinite number of unknowns, C_{nm}, S_{nm}, etc., are to be determined from a finite number of observations. In order to get a definite solution, it must be assumed that the effect of higher-degree terms is negligibly small. But even then there are very many unknowns: coordinates of the observing stations, parameters of the

reference orbit, and gravitational parameters; in addition, other unknowns must be included to take into account nongravitational forces acting on the satellite, such as air drag. An appropriate computational tool is least-squares collocation with parameters (Moritz 1980 a: Sect. 16).

To get a strong solution, observations should be evenly distributed both in space (with respect to the inclination of the satellites used) and in time.

Present results

At present (2005), several determinations of tesseral harmonics up to the degree 360 are available from a combination of satellite and terrestrial data. Soon the degree 1800 will be achieved. These coefficients represent the large-scale features of the disturbing potential T and, hence, of the geoid, since the geoidal height is given by $N = T/\gamma$. There is a general agreement between the essential aspects of these determinations as expressed in geoidal maps, although the details of these maps, and even more so the individual coefficients, are rather different.

As an example we take the first nonzonal coefficients, C_{22} and S_{22}, which, according to Sect. 2.6, Eq. (2–95), express the inequality of the earth's principal equatorial moments of inertia or, somewhat loosely speaking, its triaxiality. According to Groten (2004), we have $C_{22} = (1574.5 \pm 0.7) \cdot 10^{-9}$ and $S_{22} = (-903.9 \pm 0.7) \cdot 10^{-9}$.

Concerning the order of magnitude, J_2 is on the order of magnitude of 10^{-3}, where all the other coefficients are of order 10^{-6}. This is why the earth can be approximated by an ellipsoid so well.

7.6 New satellite gravity missions

7.6.1 Motivation and introductory considerations

Accuracy requirements in geodesy, geophysics, and oceanography for detailed gravity field information amount to 1 mgal for gravity anomalies. The related accuracy for the geoid ranges from 1 to 2 cm. In the presatellite era, the earth's gravity field was known with high accuracy only in a few regions of the world. Primarily, the available accurate gravity field information was based on terrestrial and airborne measurements. This implied that in large parts of the world there were virtually no gravity data available.

Why do we need the earth's gravity field at all? Following Pail (2003), first, the gravity field reflects the mass inhomogeneities in the earth's interior and on the earth's surface. Second, it is fundamental for the determination of the geoid (see Chap. 11) which, in its turn, may be regarded as a physical

reference surface for a number of geodynamic processes (subject to continents, oceans, ice masses, atmosphere, etc.) and their interaction. The mass inhomogeneities are a necessary prerequisite to understand convection motions in the earth's mantle which are responsible for plate tectonics. Some large and many small lithospheric plates with a thickness of some 100–200 km move with a relative velocity of some centimeters per year. At the edges of the plates, seismic zones and volcanoes are situated.

Many time-dependent earth-related processes can be regarded as changes of the mass distribution and, thus, influence the gravity field, e.g., ocean circulation, ice mass variations, sea level change, tides, volcanism, post-glacial rebound. These variations may be categorized according to their periodicity. Some of these effects are extremely long-periodic or secular, e.g., plate tectonics with about 100 million years. In contrast, changes of the ice masses may amount to some 10 years only; even immediate events like earthquakes may occur.

These variations are referred to a global physical reference surface, the geoid. Therefore, the more accurately we know the geoid, the better we accurately understand the previously mentioned effects. Referring to various disciplines, the earth's gravity field is important for, e.g., geodesy, geophysics, oceanography, and climatology.

Geodesy

As mentioned in Sect. 5.3, GPS has revolutionized geodesy in many respects. Despite the tremendous importance of GPS, in Sect. 5.4 it was shown that the user of GPS gets only *geometric* quantities: WGS 84 coordinates, i.e., geocentric rectangular coordinates X, Y, Z or, computed from them, ellipsoidal coordinates φ, λ, h (see Sect. 5.6.1). Therefore, the height obtained by GPS, i.e., the ellipsoidal height h, is purely geometric. To transform these heights into orthometric heights H by $H = h - N$, the geoidal undulation N is required. Using satellites to determine the earth's gravity field, a globally uniform height system will result.

Additionally, an accurate knowledge of the earth's gravity field improves the orbit determination of satellites.

Oceanography

The sea surface topography (SST), i.e., the difference between the geoid and the mean sea surface, can be determined when combining satellite altimetry data and the earth's gravity field data. From Fig. 7.7 we obtain the relation

$$h = N + \mathrm{SST} + \Delta H + a\,, \tag{7–47}$$

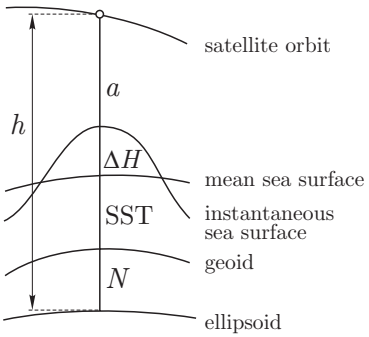

Fig. 7.7. Satellite altimetry

where h is the ellipsoidal height of the altimeter satellite (based on orbit computations), N is the geoidal height, SST is the sea surface topography to be derived, ΔH is caused by the instantaneous tidal effect, and a is the altimeter measurement. Note that (7–47) is a simplified representation, since, e.g., usually SST is split into a dynamic and a constant part. Refer to Seeber (2003: Sect. 9.3.1) for more details.

Knowing the sea surface topography, ocean currents and circulations may be explained, which is highly interesting for our understanding of the global energy transport. Ocean currents together with their time variations are an important indicator for climatic changes.

This method suffers from different accuracy influences in the results: when referring the mean sea surface to the ellipsoid, centimeter accuracy could be achieved. Involving the gravity model and referring the sea surface topography to the geoid as in Fig. 7.7, an improved geoid is required for a consistent accuracy level.

Geophysics

As mentioned earlier, the earth's gravity field reflects the mass inhomogeneities in the interior of the earth. Knowing gravity values on the earth's surface and, in addition, complementary data (e.g., magnetic and seismic data), improved models for the structure and processes in the earth's interior may be obtained. These processes may cause the movement of tectonic plates which are responsible for earthquakes. Thus, we see that the gravity field is the fundamental link in a chain of interactive processes. Using more descriptive terms, an improved knowledge on the gravity field may yield more accurate methods to predict earthquakes. This justifies any effort on the determination of the earth's gravity field.

7.6.2 Measurement concepts

From the introduction above, the need for an accurate determination of the earth's gravity field becomes evident. Three different measurement concepts evolved, leading to three different gravity field satellite missions:

- satellite-to-satellite (SST) tracking in high-low mode being realized by the "Challenging Minisatellite Payload" (CHAMP) mission,
- satellite-to-satellite tracking in low-low mode being realized by the "Gravity Recovery and Climate Experiment" (GRACE), and
- satellite gravity gradiometry, the objective of the "Gravity Field and Steady State Ocean Circulation Explorer" (GOCE) mission.

Before giving some details on the objectives and payloads of the missions, the different concepts are briefly described.

Satellite-to-satellite tracking in high-low mode

The principle is shown in Fig. 7.8. The orbit of the low earth orbit (LEO) satellite is continuously determined by satellites of global systems such as GPS, GLONASS or, in the future, Galileo. Note that the term "high-low mode" is not really appropriate because the satellites of GPS, GLONASS, and Galileo belong to the mean earth orbit (MEO) satellites and not to the high earth orbit (HEO) satellites. However, we keep the notation as used in Seeber (2003: Sect. 10.1). Apart from satellite-to-satellite tracking, the LEO satellite uses an accelerometer. In principle, three-dimensional perturbing accelerations caused by the earth's gravity field are measured. These accelerations correspond to first derivatives of the gravitational potential V.

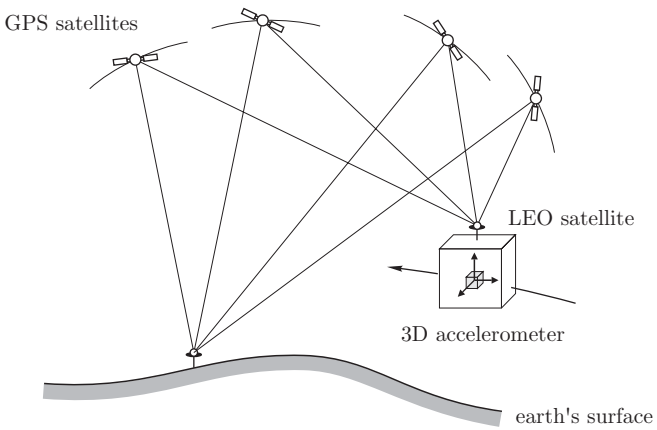

Fig. 7.8. Satellite-to-satellite tracking in high-low mode

The gravity field is derived by inverting (in the sense of inverse problems, cf. the remark on inverse problems at the end of Sect. 1.13) the information obtained from the satellite orbit

Satellite-to-satellite tracking in low-low mode

The principle is shown in Fig. 7.9. Two LEO satellites are placed in the same orbit but separated by some hundreds of kilometers (about 220 km in the case of GRACE). Ranges and range rates between the satellites are measured to utmost accuracy. Individually, the orbit of each LEO satellite is affected by perturbing accelerations which correspond to the first derivatives of the gravitational potential. In combination, differences of accelerations result. In addition, the position of the LEOs is determined by GPS satellites. This means that inherently satellite-to-satellite tracking in high-low mode is also implied. The effect of nongravitational forces on the satellite, e.g., due to air drag, must either be compensated or measured by an accelerometer.

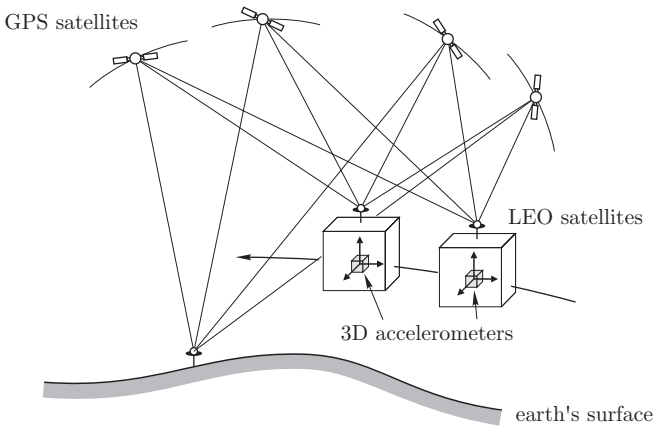

Fig. 7.9. Satellite-to-satellite tracking in low-low mode

Satellite gravity gradiometry

Compared to the just decribed low-low mode of satellite-to-satellite tracking with a long baseline between the two LEOs, the baseline between the accelerometer units tends to zero in case of satellite gravity gradiometry. This is achieved by placing both units into a single satellite (Fig. 7.10). Therefore, satellite gradiometry is the measurement of acceleration differences in three spatial orthogonal directions between the test masses of the six accelerometer units (two on each of the three axes) inside the satellite. In other words, the measured signal is the difference in gravitational acceleration at the satellite,

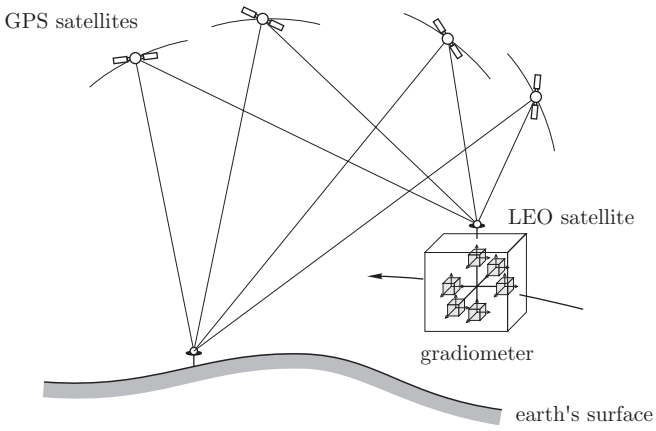

Fig. 7.10. Satellite gravity gradiometry with a three-axis gradiometer

where the gravitational signal arises from the attracting masses of the earth. Thus, the measured signal corresponds to the gradients of the component of the gravity acceleration, i.e., the second derivatives of the gravitational potential. For instance, in obvious notation we read from Fig. 7.11

$$\frac{V_{x2} - V_{x1}}{\Delta z} = \frac{\Delta V_x}{\Delta z} \doteq \frac{\partial V_x}{\partial z} = V_{xz}\,. \tag{7–48}$$

Summarizing the briefly described three methods, satellite-to-satellite tracking in high-low mode, satellite-to-satellite tracking in low-low mode, and satellite gravity gradiometry, we may say that the basic observable is gravitational acceleration. Following Rummel et al. (2002), the case of satellite-to-satellite tracking in high-low mode corresponds to a three-dimensional position, velocity or acceleration determination of a LEO satellite. The three-dimensional accelerometry corresponds to gravity acceleration. Mathematically, this is expressed by the first derivatives of the gravitational potential.

Considering the low-low mode, the principle corresponds to the line-of-

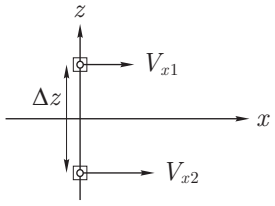

Fig. 7.11. Measuring the second derivative V_{xz}

sight measurement of the range, range rate or acceleration difference between the two low-orbiting satellites. The intersatellite link corresponds to acceleration differences between the two LEO satellites. Mathematically, this is expressed by the difference of first derivatives of the gravitational potential over a long baseline (i.e., the distance between the two LEO satellites).

In the case of satellite gradiometry, three-dimensional acceleration differences referring to the very short baseline realized by the gradiometer are measured. The gradient of gravity components corresponds to the acceleration gradient. Mathematically, this is expressed by the second derivatives of the gravitational potential.

Another feature inherent to satellite gravity missions should be kept in mind: the amplification of errors by the factor $(r/R)^{n+1}$ when transferring the measurement comprising the signal and noise from satellite altitude to the earth's surface. The factor $(r/R)^{n+1}$ describes the field attenuation with altitude. This error amplification effect is minimized by using an orbit as low as possible and by not measuring the potential V itself or its gradient but rather its second-order derivatives as in gravity gradiometry.

7.6.3 The CHAMP mission

The information on the challenging minisatellite payload (CHAMP) mission has been extracted primarily from http://op.gfz-potsdam.de/champ.

The Geoforschungszentrum Potsdam initiated the CHAMP idea and has the main responsibility. The primary CHAMP objectives are the following:

- mapping of the global gravity field, or, more specifically, to accurately determine the long-wavelength features of the static earth gravity field and its temporal variations (caused, e.g., by atmospheric mass redistributions, ocean circulation, sea level changes resulting from polar ice melting);
- mapping of the global magnetic field, or, more specifically, to accurately determine the main and crustal magnetic field of the earth and its space-time variations;
- profiling of the ionosphere and the troposphere, or, more specifically, to derive from GPS signal refraction data information on the temperature, water vapor, and electronic content of the atmosphere.

The CHAMP mission was launched on July 15, 2000 from the Russian Plesetsk cosmodrome. The main mission parameters of the respective satellite are the following:

- almost circular (eccentricity $e < 0.004$) and near-polar ($i = 87°$) orbit,

- initial altitude of 454 km,
- designed lifetime of five years for the mission (but the life expectation is much higher!),
- weight of 522 kg, length of 8.3 m (including a "boom" of 4 m length), width of 1.6 m, height of 0.75 m.

This initial altitude may be regarded as a compromise between gravity and magnetic field measurements. Considering the gravity field, a lower altitude would be desirable. Primarily due to atmospheric drag, the altitude will decrease to about 300 km and even less which is important because of an increasing sensitivity with respect to gravity field coefficient determination.

The reason for the curious 4 m boom is that the magnetometry assembly must be separated from the main body of the satellite ("magnetic cleanliness reasons", see http://op.gfz-potsdam.de/champ).

To achieve the mission goals, the following payload is on board of the satellite:

- dual-frequency GPS receiver connected to a multiple antenna system to determine the orbit of the CHAMP satellite using code and phase pseudoranges;
- three-axis accelerometer to measure the nongravitational accelerations acting on the spacecraft (air drag, solar radiation pressure, albedo, etc.);
- laser retroreflector for backup tracking to measure two-way ranges between ground stations and the satellites with 1–2 cm accuracy; these measurements support the precise orbit determination;
- fluxgate magnetometer to measure the vector components of the magnetic field of the earth (this instrument is supported by a scalar magnetometer to provide a calibration capability of the fluxgate magnetometer);
- equipment to determine the ion density and temperature, the drift velocity, and the electric field;
- two advanced star trackers to provide high-precision attitude information as required for the three-axis accelerometer, the digital ion drift meter, but also for the attitude control of the satellite.

Typical other equipment required for a proper operation of the satellite but with no specific relation to the scientific objectives of the mission is not detailed here, such as the cold gas propulsion system, the thermal control system, the power generation, the data handling, the telemetry, tracking and command system. Furthermore, we do not list items of the control segment

of the CHAMP mission, but refer the reader to the previously mentioned homepage.

As explained before, the measuring principle for CHAMP is satellite-to-satellite tracking in high-low mode. The gravity field of the earth perturbes the CHAMP satellite orbit. These perturbing accelerations correspond to first derivatives of the gravitational potential V. This implies that the gravity field of the earth may be derived from observed gravitational satellite orbit perturbations applying numerical orbit integration (Montenbruck and Gill 2001) or using the energy balance principle (Ilk 1999, Jekeli 1999, Sneeuw et al. 2002).

For further reading see Reigber et al. (2003), Seeber (2003: Sect. 10.2.2).

7.6.4 The GRACE mission

The information on the gravity recovery and climate experiment (GRACE) mission has been extracted primarily from http://op.gfz-potsdam.de/grace.

The GRACE mission is a joint project between the U.S. National Aeronautics and Space Administration (NASA) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The primary objectives of the mission are the following:

- determination of the global high-resolution gravity field of the earth,
- temporal gravity variations.

In addition, another task is the determination of the total electron content by GPS measurements to get knowledge on the refractivity in the ionosphere and troposphere. The two satellites of this mission were launched simultaneously on March 17, 2002 from the Russian Plesetsk cosmodrome. The main mission parameters of the two satellites are the following:

- almost circular (eccentricity $e < 0.005$) and near-polar ($i = 89°$) orbit,
- initial altitude between 485 km and 500 km,
- the two satellites are some 220 km apart (this requires orbit maneuvers every one or two months to maintain the separation between the two spacecraft),
- design lifetime of the mission is five years (but extended operation is envisaged),
- the weight of each satellite is about 480 kg and the length about 3 m.

As with CHAMP, also the altitude of the GRACE satellites will decrease in the course of their lifetime primarily because of atmospheric drag. The

amount of this decrease depends on the solar activity cycle and may accumulate in the mission lifetime to some 50 km on low activity, and up to 200 km on high activity, see http://op.gfz-potsdam.de/grace.

The range between the two satellites must be determined extremely accurately. Its range rate must be known to better than $1\,\mu\mathrm{m\,s^{-1}}$, which is achieved by intersatellite microwave measurements. The basic idea is that variations in the gravity field cause variations in the range between the two satellites; areas of stronger gravity will affect the lead satellite first and, therefore, accelerate it away from the following satellite (Seeber 2003: p. 479).

GRACE will not only provide a static global gravity field but also its temporal variations.

To achieve the mission goals, the following payload is on board of the two satellites:

- The K-band ranging system is the key instrument of GRACE to measure the range changes between both satellites using dual-band microwave signals (i.e., two one-way ranges) with a precision of about $1\,\mu\mathrm{m\,s^{-1}}$. The ranges are obtained at a sampling rate of 10 Hz.
- The GPS receiver serves for the precise orbit determination of the GRACE spacecraft and provides data for atmospheric and ionospheric profiling. To achieve this, satellite-to-satellite tracking between the GRACE satellites and the GPS satellites is realized. A navigation solution comprising position, velocity, and a time mark is derived on board. The navigation solution is required for the attitude control system. The precise orbit based on code and carrier pseudoranges is determined on ground.
- The attitude and orbit control system comprises a cold gas propulsion system, three magnetic torque rods, star trackers, a three-axis inertial reference unit to measure angular rates, and a three-axis magnetometer.
- The accelerometer measures all nongravitational accelerations on the GRACE spacecraft, e.g., due to air drag or solar radiation pressure.
- The laser retroreflector is a passive payload instrument used to reflect short laser pulses transmitted by ground stations. The distance between a ground station and a GRACE satellite can be measured with an accuracy of 1–2 cm. The laser retroreflector data are primarily used together with the GPS receiver data for the precise orbit determination.

In 2004, the GRACE science team released to the public a first version of a new earth gravity field model complete to degree and order 150. The resulting improved geoid together with satellite altimetry will advance the

knowledge on oceanographic, geodetic, and solid earth issues such as oceanic heat flux, change of sea level, ocean currents, precise positioning, orbit determination, and leveling.

The GRACE concept can be regarded as a one-dimensional gradiometer with a very long baseline of 220 km (Seeber 2003: p. 480). In contrast to this concept, GOCE uses very short baselines (50 cm) in three directions.

7.6.5 The GOCE mission

The main sources of this section are www.esa.int/export/esaLP/goce.html, ESA (1999), Müller (2001), Drinkwater et al. (2003), and Pail (2003).

The gravity field and steady-state ocean circulation explorer (GOCE) mission is a Core Mission of the ESA Living Planet Programme. The primary objectives of the GOCE mission are to measure the earth's stationary gravity field and to model the geoid with extremely high accuracy. More specifically:

- to determine the gravity anomalies with an accuracy of 1 mgal,
- to determine the geoid with an accuracy of 1–2 cm,
- to achieve these results at a spatial resolution better than 100 km.

According to the above mission requirements, GOCE is intended for a representation of the gravity potential by spherical harmonics complete at least to degree and order 200 (corresponding to the spatial resolution of 100 km), but 250 is envisaged.

From the *geodetic point of view*, a global geoid of 1–2 cm accuracy and a gravity field model accurate to 1 mgal at about 100 km spatial resolution may be used – among many other important applications – for the following purposes:

- Control (or replacement) of traditional leveling by leveling with GPS. In Sect. 4.6 we have learned the basic equation (4–72), $H = h - N$, relating the orthometric height H (above the geoid), the ellipsoidal height h (above the ellipsoid), and the geoidal undulation N. With N accurately known from GOCE and h measured by GPS (Sects. 5.5, 5.6.1), the orthometric height H is readily obtained.
- Worldwide unification of height systems so as to refer to one height datum which allows for comparison of different sea levels (e.g., in the North Sea and in the Mediterranean) and sea-level changes (which may be caused by melting continental ice sheets). Remember that the geoid is defined as an equipotential surface which follows a hypothetical ocean surface at rest (in the absence of tides and currents and other smaller influences). Consequently, a precise geoid is crucial in deriving accurate measurements of ocean currents and sea-level changes.

- Providing a significant improvement in satellite orbit determination and prediction. This especially applies to low-orbiting satellites. The highly accurate gravity field will enable a better separation of the perturbations caused by the static gravity field and other perturbing forces (not only the nongravitational forces caused by air drag and solar radiation pressure but also perturbations caused by the solid earth and ocean tides).

The duration of the mission is scheduled with nominally 20 months, including a 3-month commissioning and calibration phase and two measurement phases, each lasting six months and separated by a long eclipse period. The other main mission parameters are the following:

- due for launch in 2007 from Plesetsk in Russia,
- sun-synchronous orbit, inclination 96.5°,
- measurement altitudes: approximately 250 km,
- single ground station in Kiruna, Sweden, to exchange data and commands; the European Space Operations Center (ESOC) at Darmstadt will be used for mission and satellite control.

The main payload components are the following:

- three-axis gravity gradiometer based on three pairs of electrostatic servo-controlled accelerometers to measure gravity gradients in three spatial orthogonal directions: the desired signal is the difference in gravitational acceleration (between a pair of accelerometers separated by 0.5 m) at the test mass location inside the spacecraft caused by gravity anomalies from attracting masses of the earth;
- geodetic dual-frequency (to compensate for ionospheric delays) multichannel GPS receiver with codeless tracking capability to (1) determine the orbit of the GOCE satellite and (2) derive gravity information from this orbit (the first task is performed by satellite-to-satellite tracking in the high-low mode: this provides knowledge of the precise position of the [low] spacecraft relative to [high] reference satellites such as the GPS satellites; the second task is solved by orbit perturbation analysis yielding gravity information);
- laser retroreflector to enable tracking by ground-based laser stations;
- attitude control accomplished by actuators comprising an ion thruster assembly, star trackers, a three-axis magnetic torquer, and some other sensors;
- length of the satellite about 5 m, cross section of $1\,m^2$, weight about 1000 kg.

Referring to the results, the main output of this mission will be the following:

- spherical-harmonic coefficients for the gravitational potential, see, e.g., (2–80),
- corresponding variance-covariance matrix.

Derived products from this main output are geoidal heights, gravity anomalies, and also oceanographic data.

It is important to mention that the GPS orbit analysis of GOCE will rather yield long-wavelength information of the gravity field, while the satellite gravity gradiometry will yield the short-wavelength information.

GOCE is the first "drag-free" mission, which implies that the satellite moves in free fall around the earth. Therefore, a drag compensation and attitude control system is required to compensate for drag forces and torques.

This and more information may be found in Rebhahn et al. (2000), Drinkwater et al. (2003), Pail (2003), www.esa.int/livingplanet/goce.

Measurements

The basic principle of gradiometry in GOCE is the measurement of acceleration differences for a very short baseline. Considering two accelerometers separated by 50 cm on one axis, Müller (2001) and Pail (2003) write the two observation equations as

$$\begin{aligned} \mathbf{a}_1 &= \left[\mathbf{M} + \dot{\mathbf{\Omega}} + \mathbf{\Omega}\mathbf{\Omega}\right]\Delta\mathbf{x} + \mathbf{f}_{\mathrm{ng}}\,, \\ \mathbf{a}_2 &= -\left[\mathbf{M} + \dot{\mathbf{\Omega}} + \mathbf{\Omega}\mathbf{\Omega}\right]\Delta\mathbf{x} + \mathbf{f}_{\mathrm{ng}}\,, \end{aligned} \tag{7–49}$$

where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the measured accelerations of the two accelerometers on the axis, and $\mathbf{M}$ is the Marussi tensor,

$$\mathbf{M} = \begin{bmatrix} \dfrac{\partial^2 V}{\partial x^2} & \dfrac{\partial^2 V}{\partial x\,\partial y} & \dfrac{\partial^2 V}{\partial x\,\partial z} \\ \dfrac{\partial^2 V}{\partial x\,\partial y} & \dfrac{\partial^2 V}{\partial y^2} & \dfrac{\partial^2 V}{\partial y\,\partial z} \\ \dfrac{\partial^2 V}{\partial x\,\partial z} & \dfrac{\partial^2 V}{\partial y\,\partial z} & \dfrac{\partial^2 V}{\partial z^2} \end{bmatrix}, \tag{7–50}$$

which comprises the second derivatives of the gravitational potential (our target quantity!). Furthermore, the skewsymmetric matrix

$$\mathbf{\Omega} = \begin{bmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{bmatrix} \tag{7–51}$$

comprises the components of the angular velocity and is used to describe the orientation of the gradiometer. Since $\mathbf{\Omega}$ is skewsymmetric, the tensor $\mathbf{\Omega\Omega}$ is symmetric. Finally, $\Delta\mathbf{x}$ in (7–49) is the vector from the intersection of the three coordinate axes to the respective accelerometer (where the same length is assumed), and $\mathbf{f}_{\mathrm{ng}}$ comprises all nongravitational effects (air drag, solar radiation pressure, etc.).

Now we once add ("common mode") and once subtract ("differential mode") the two accelerations in (7–49) and obtain

$$\begin{aligned} (\mathbf{a}_1 + \mathbf{a}_2)/2 &= \mathbf{f}_{\mathrm{ng}}\,, \\ (\mathbf{a}_1 - \mathbf{a}_2)/2 &= \big[\mathbf{M} + \dot{\mathbf{\Omega}} + \mathbf{\Omega\Omega}\big]\Delta\mathbf{x}\,, \end{aligned} \tag{7–52}$$

where we can extract the nongravitational effects $\mathbf{f}_{\mathrm{ng}}$ in the common mode. Introducing the quantity

$$\mathbf{\Gamma} = \mathbf{M} + \dot{\mathbf{\Omega}} + \mathbf{\Omega\Omega} \tag{7–53}$$

and assuming a known geometry of the gradiometer, i.e., $\Delta\mathbf{x}$ may safely assumed to be known, then the remaining task is to extract the gravity gradient tensor $\mathbf{M}$ from $\mathbf{\Gamma}$. This can be achieved by the two relations

$$\begin{aligned} (\mathbf{\Gamma} - \mathbf{\Gamma}^T)/2 &= \dot{\mathbf{\Omega}}\,, \\ (\mathbf{\Gamma} + \mathbf{\Gamma}^T)/2 &= \mathbf{M} + \mathbf{\Omega\Omega}\,, \end{aligned} \tag{7–54}$$

where the superscript T denotes transposition. To verify these relations, a little matrix calculus is needed. If, generally, $\mathbf{K}$ is a symmetric matrix, then we have $\mathbf{K} = \mathbf{K}^T$. If $\mathbf{K}$ is a skewsymmetric matrix, then we have $\mathbf{K} = -\mathbf{K}^T$.

Referring now to (7–53), we know that $\mathbf{M}$ is symmetric, $\dot{\mathbf{\Omega}}$ is skewsymmetric, and $\mathbf{\Omega\Omega}$ is symmetric. Therefore, transposing (7–53) yields

$$\mathbf{\Gamma}^T = \mathbf{M} - \dot{\mathbf{\Omega}} + \mathbf{\Omega\Omega}\,. \tag{7–55}$$

Using now (7–53) and (7–55), we get immediately

$$\mathbf{\Gamma} - \mathbf{\Gamma}^T = 2\dot{\mathbf{\Omega}} \tag{7–56}$$

and, finally,

$$(\mathbf{\Gamma} - \mathbf{\Gamma}^T)/2 = \dot{\mathbf{\Omega}}\,, \tag{7–57}$$

which completes our proof for the first relation of (7–54). To prove the second relation of (7–54), we add Eqs. (7–53) and (7–55):

$$\mathbf{\Gamma} + \mathbf{\Gamma}^T = 2\mathbf{M} + 2\mathbf{\Omega\Omega} \tag{7–58}$$

or

$$(\mathbf{\Gamma} + \mathbf{\Gamma}^T)/2 = \mathbf{M} + \mathbf{\Omega\Omega}\,, \tag{7–59}$$

which concludes our proof.

Since we have determined $\dot{\mathbf{\Omega}}$ in (7–57), we can get $\mathbf{\Omega}$ by integration:

$$\mathbf{\Omega}(t) = \mathbf{\Omega}(t_0) + \int_{t_0}^{t} \dot{\mathbf{\Omega}}\, dt\,, \tag{7–60}$$

where the initial orientation $\mathbf{\Omega}(t_0)$ is obtained from the star trackers. Squaring the result for $\mathbf{\Omega}(t)$ yields $\mathbf{\Omega\Omega}$, which is needed in (7–59) so that we find

$$\mathbf{M} = (\mathbf{\Gamma} + \mathbf{\Gamma}^T)/2 - \mathbf{\Omega\Omega} \tag{7–61}$$

as final result for the desired Marussi tensor $\mathbf{M}$. Many more details may be found in Rummel (1986).