CHAPTER 15

POINT POSITIONING

Note: This is a new version (nearly final version) of Chapter 15 of "Geodesy the Concepts"
Written by Spiros Pagiatakis with input from Petr Vanicek

The determination of the coordinates of a point on land, at sea, or in space with respect to an implied coordinate system is called point positioning or spatial positioning. The problem of point positioning may be stated as follows: Given the coordinates of observed extraterrestrial objects, such as stars or satellites, along with the measurements of quantities linking a terrestrial point to these objects, compute the coordinates of the point. Positioning of a point with respect to other terrestrial points, known as relative positioning, is treated in Chapter 16.

Because point positioning can be done in three different modes, three distinctly different classes of coordinate systems are needed: *terrestrial* coordinate systems for positioning points and objects in the immediate environment of the Earth (solid surface, oceans, and atmosphere); *celestial* coordinate systems for positions of the sighted stars and extragalactic radio sources; and *orbital* coordinate systems for the observed satellites. These coordinate systems have been linked through the orbital characteristics of the Earth as described by vernal point ♈ and the direction of the rotation axis of the Earth. This description is deemed to be *dynamical* and as such, these coordinate systems are affected by all the irregularities of the revolution of the Earth about the Sun and its spin about its own axis (cf., Chapter 5). These irregularities cause instability of the reference coordinate systems over time, a characteristic that is not desirable, particularly when high accuracy positioning is required.

Over the last three decades, the international scientific unions (IAU and IUGG) made a significant effort to remedy the instability of the classical *celestial* coordinate systems by fixing the coordinate axes directions relative to the distant matter of the universe (e.g., quasars) that have highly accurate and stable directions (viewed from the Earth) established by long-term astronomical measurements. Similarly, the *terrestrial* coordinate systems are fixed relative to selected stations on the surface of the Earth whose positions have been determined by a combination of different geodetic techniques over long periods of time. The generation of these modern celestial and terrestrial coordinate systems starts from a *conventional* snapshot of their classical equivalents at a particular point in time τ_0 in the past. Thereafter, as new astronomical and geodetic measurements are collected, these modern coordinate systems (origin and axes directions) are adjusted from time to time to fit the improved coordinates of the celestial reference points (e.g., quasars) or terrestrial stations (fiducial points). Thus, the modern systems have space-fixed axes whose directions are based on observations and they are thus only functions of time. As such, they are deemed to be *kinematical* in nature.

Whether the classical (*dynamical*) or new definition (*kinematical*) is followed, terrestrial coordinate systems are Earth-fixed: they both, spin and revolve with the Earth. Celestial coordinate systems do not spin but may revolve around the Sun with the same velocity as the Earth does. The orbital coordinate systems do not spin with the Earth but revolve with it.

In the current chapter, the classical (dynamical) coordinate systems are presented first because they form the basis for the definition of the new generation or kinematical ones. The first section presents the fundamentals of geodetic astronomy consisting of the definitions of the classical celestial, terrestrial and orbital coordinate systems and their transformations. The second section discusses the time systems, which are needed in the transformations among the various coordinate systems and in astronomical positioning. The third section treats the mathematical models used in the astronomical determination of position whereas the fourth section addresses the mathematical models for determining the position of a point from observations to satellites. The fifth section introduces the new generation of reference systems and the rationale for changing over from the classical to the modern. References are made to the similarities and differences with the classical coordinate systems. Finally, the sixth section addresses the concept of positioning on the reference ellipsoid. Transformation of positions from one ellipsoid onto another are also discussed as well as the mapping of the ellipsoid onto a plane. In the models shown here, we make no allowance for time deformations of the Earth. Throughout this chapter, the point to be positioned is considered to be stationary with respect to the Earth; positioning of a moving point or object, as part of navigation, is discussed in Chapter 16.

15.1. Fundamentals of geodetic astronomy

Let us begin by discussing the *celestial systems* of coordinates in the classical way. Since the distance from the Earth to the nearest star (excluding the Sun) is more than 10^9 times larger than the Earth's radius, the dimension of the Earth is negligible compared with the distance to stars. The stars in our galaxy are almost immobile; however, the galaxies themselves are believed to be moving at velocities comparable to the speed of light. To an observer on the Earth, though, even this motion is perceived to be very slow causing a displacement that rarely exceeds one second of arc per year. Therefore, one may consider the stars and galaxies to be located on a surface called the *celestial sphere* (see §1.1) the dimension of which is so large that the Earth can be considered dimensionless (as a point) at the centre of this sphere, as it was considered in §5.1. Directions on the Earth and in the solar system can then be extended to the celestial sphere. Points and curves on the celestial sphere obtained in this manner form the basis for the definition of all celestial coordinate systems. The Earth's (precessing and nutating) instantaneous spin axis (see §5.2) is extended outward to intersect the celestial sphere at the *north celestial pole* (NCP) and *south celestial pole* (SCP)–see FIG. 15.1. The Earth's equatorial plane (the plane perpendicular to the spin axis and containing the centre of mass C of the Earth) intersects the celestial sphere to form the *celestial equator.* A plane parallel to the celestial equator intersects the celestial sphere in a small circle called *celestial parallel.* Any great circle containing the poles is perpendicular to both, the celestial equator and parallels and is called *astronomical* (celestial) *meridian.* The gravity vector $\vec{g}$ of the observer extended upward intersects the celestial sphere at a point called the *zenith* (of the observer), and downward to define a point called the *nadir* (of the observer). The great circle made by a plane perpendicular to the observer's gravity vector is the *celestial horizon.* A small circle made by a plane parallel to the celestial horizon is called *almucantar.*

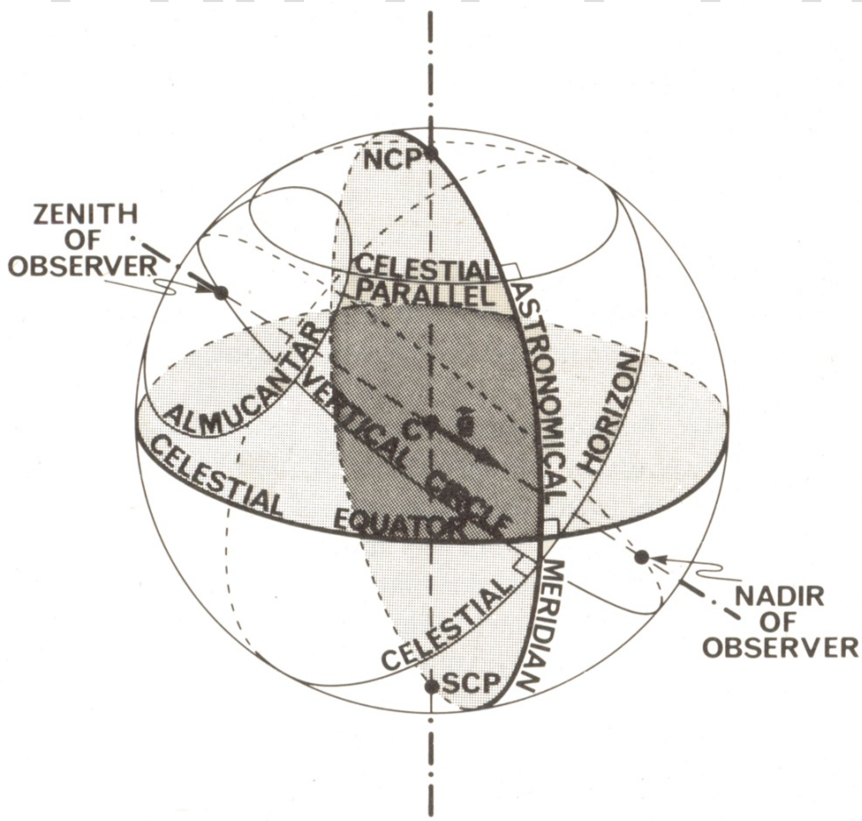

FIG. 15.1. Celestial sphere.

Any plane that contains the gravity vector is a vertical plane whose intersection with the celestial sphere is a *vertical circle.* The vertical plane normal to the astronomical meridian is called the *prime vertical* (or decuman); it intersects the celestial horizon to define the *east* and *west directions.*

All classical celestial coordinate systems will be defined by the location of their origin and the directions of the axes towards certain characteristic points on the celestial sphere (e.g., vernal point ♈), whose positions are defined dynamically using celestial mechanics. The designation '*apparent*' will signify that the origin of the celestial coordinate system is at the centre of mass of the Earth C (geocentre); when the adjective 'apparent' is not used, it will deem the system to be '*heliocentric*'. Celestial coordinate systems that are affected by all irregularities and at all frequencies of the rotation axis of the Earth (i.e., precession, nutation), will be designated as '*instantaneous*' or '*true*' and will be shown to be functions of time (τ). When a celestial system precesses but does not nutate, it will be called '*mean*'. Similarly, terrestrial coordinate systems defined with respect to the Earth's spin axis that wobbles will be called '*instantaneous*' and when the effect of polar wobble has been accounted for, they will be called '*conventional*'.

To define the position of a star anywhere on the celestial sphere, all one really needs to know is a direction. A direction is most simply defined as a unit vector in polar coordinates (r,θ,λ – see §3.3): since the first coordinate r always equals one, in effect the vector is specified by only the two angles (θ, λ).

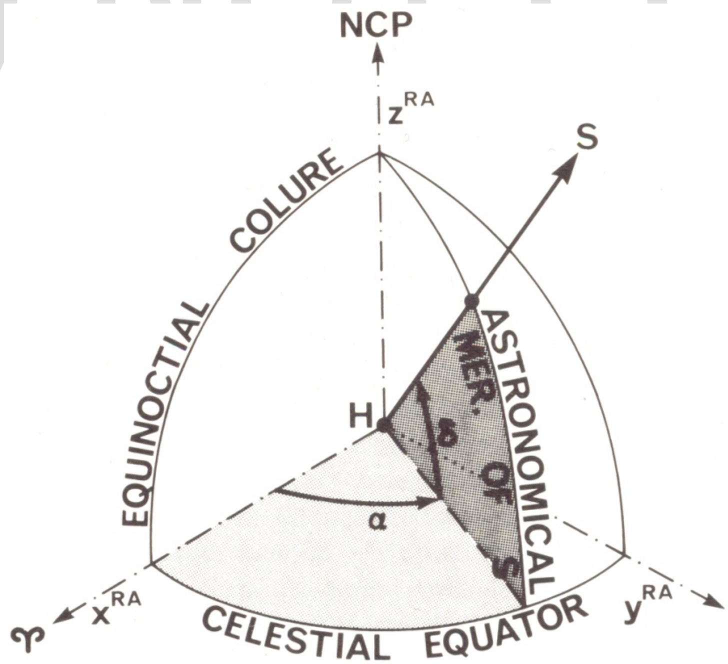

FIG.15.2. True right ascension system

Let us begin with the *true right ascension system* (TRA(τ)), the most important of the *celestial* systems (see §1.1). It is *heliocentric,* i.e., its origin is the Sun (H), the z^{RA}-axis is directed toward the NCP, the x^{RA}-axis points toward the vernal point ♈ (see §5.2), and the system is right-handed – cf. FIG. 15.2. The *declination* δ of the star (S) is the angle between the celestial equatorial plane and the direction from H to S, measured on the astronomical meridian plane of S. The *right ascension* α of S is the angle measured counterclockwise, as seen from the NCP, on the equatorial plane from ♈ to the astronomical meridian of S. Note that the astronomical meridian of ♈ is called the *equinoctial colure.* The unit vector describing the direction to S in this system is

$$\boxed{\vec{e}^{\,\mathrm{TRA}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\mathrm{TRA}} = \begin{bmatrix} \cos\delta\,\cos\alpha \\ \cos\delta\,\sin\alpha \\ \sin\delta \end{bmatrix},} \tag{15.1}$$

while the angles are related to the Cartesian components $(x,y,z)^{\mathrm{TRA}}$ by

$$\boxed{\begin{aligned} \delta &= \arcsin z^{\mathrm{TRA}}, \\ \alpha &= \arctan\left(y^{\mathrm{TRA}} / x^{\mathrm{TRA}}\right). \end{aligned}} \tag{15.2}$$

Since $\alpha \in [0, 2\pi]$, the second equation carries with it an uncertainty of π. It is thus sometimes preferable to use the equivalent equation for a half-angle,

$$\alpha = 2\arctan\frac{y^{\mathrm{TRA}}}{x^{\mathrm{TRA}} + \sqrt{(x^{\mathrm{TRA}})^2 + (y^{\mathrm{TRA}})^2}}\,, \tag{15.3}$$

which is unequivocal.

There are, however, some complications involved here: clearly, because of precession and nutation (cf., §5.2), the NCP, being defined through the Earth's instantaneous spin axis, moves with respect to the stars as a function of time. Thus, the TRA coordinate system changes with time as do the coordinates (a, δ) of the stars. When publishing the stars' positions in the TRA, it is necessary to specify the epoch τ_0 to which the coordinates refer. It usual for star catalogues to use an RA system that precesses but does not nutate. This RA system is called a *mean right ascension system* - MRA(τ_0) (see below).

The next most important coordinate system is the *local astronomical* – LA(τ), the one in which the observations to the stars are made. This system is *terrestrial* and is defined through the observer's gravity vector and the direction of the Earth's instantaneous spin axis. The observer's gravity vector can be realised for instance by setting and levelling a theodolite or total station and the *vertical* (altitude) *angle* ν, or *zenith distance* Z, and direction to the star(s) with respect to a reference direction on the surface of the Earth can be directly measured by these instruments (see FIG. 15.3). The direction to the instantaneous spin axis can only be sensed directly by special instruments (gyro-theodolites), with which the azimuth A of the star(s) and/or of a reference direction can be measured directly in addition to the measurement of angle ν (or Z). All these observations must be time-tagged very accurately that is, '*time*' is an essential part of the measurement. The gravity vector defines the negative z^{LA}-axis, and together with a parallel to the instantaneous spin axis (cf., §5.4) they define the xz^{LA}-plane of the system. The y^{LA}-axis completes the left-handed system. The origin of the LA(τ) is at the site of the observer on the surface of the Earth (T); this coordinate system is thus said to be *topocentric*. In this system, the unit vector in the direction of S is

$$\boxed{\vec{e}^{\,\mathrm{LA}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\mathrm{LA}} = \begin{bmatrix} \cos\nu \ \cos A \\ \cos\nu \ \sin A \\ \sin\nu \end{bmatrix},} \tag{15.4}$$

while the angles are related to the Cartesian coordinates through

$$\boxed{\nu = \frac{1}{2}\pi - Z = \arcsin z^{\mathrm{LA}}, \qquad A = 2\arctan\frac{y^{\mathrm{LA}}}{x^{\mathrm{LA}} + \sqrt{(x^{\mathrm{LA}})^2 + (y^{\mathrm{LA}})^2}}\,.} \tag{15.5}$$

Note that the system is not defined for those points whose gravity vector direction coincides with the direction of the conventional spin axis.

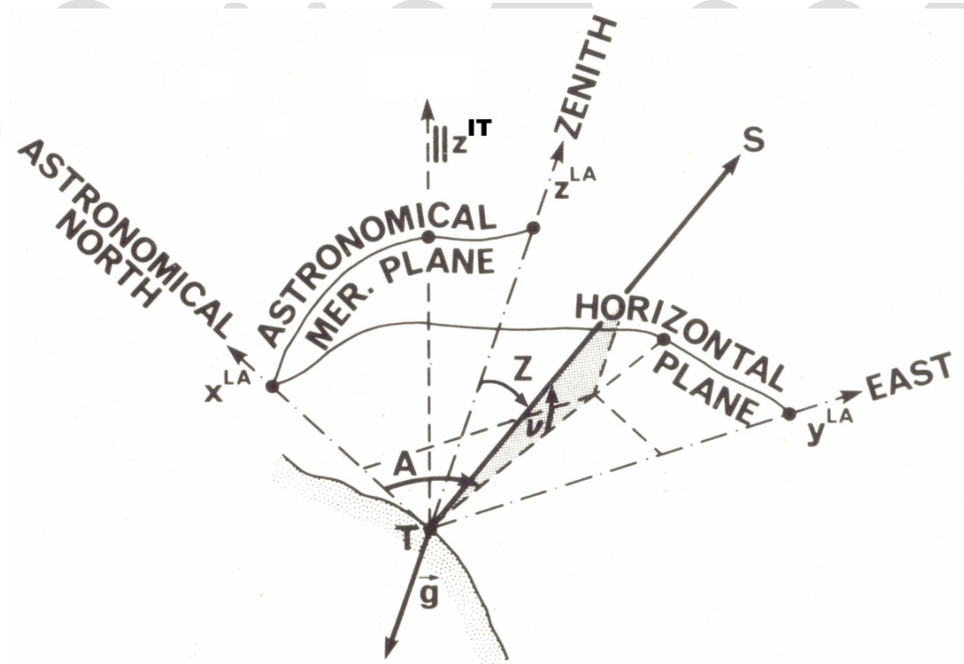

FIG. 15.3. Local astronomical system.

The observations to the stars are made in the topocentric LA(τ) system, which is spinning as well as revolving with the Earth. On the other hand, the star positions published in the MRA(τ_0) system that refers to an epoch (different from that of the observations) are motionless up to the precession. The problem faced in the astronomical determination of positions is in the reconciliation of the above two systems. This problem is normally solved through a series of transformations from one system to another. More accurately, the observations and the star coordinates are both transformed into a third coordinate system – the apparent right ascension system or system of apparent places – to be defined later. The position of the observer on the Earth's surface along with the azimuth of a reference line and the azimuth of the star at the time of observation τ are then determined as by-products of these transformations. Let us now go through these transformations, step by step ((a) to (f), see also FIG. 15.13), starting with the instantaneous LA(τ) system.

Step (a): The first coordinate system needed in transforming an LA(τ) system into the system of apparent places is the *instantaneous terrestrial system* – IT(τ), a natural system of coordinates. Its origin is at the centre of mass of the Earth (geocentre), the z^{IT}-axis coincides with the instantaneous (true) spin axis, the xz^{IT}-plane contains the *instantaneous Greenwich Observatory* [ROBBINS, 1976], and the y^{IT}-axis is selected to make the system right-handed (cf., FIG. 15.4). The unit vector in the direction of the local zenith is again given by formulae similar to those for the TRA and LA systems, which involve *astronomical latitude* Φ and *astronomical longitude* Λ (see FIG. 15.4). These two angles are the ones used to define the position of a point astronomically. The principles of transformation (§3.3) can now be applied to obtain the 'observed' unit vector in the direction of S (Eq. (15.4)) into the IT system. We get

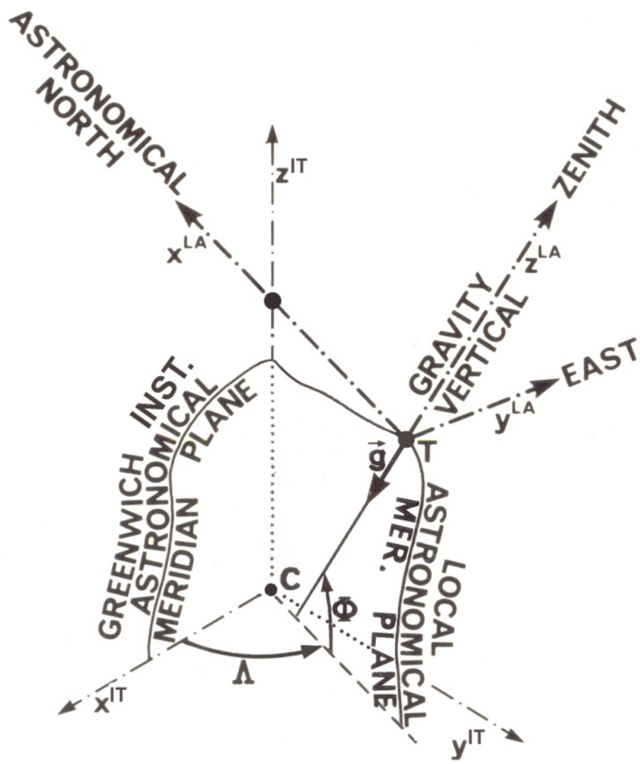

FIG. 15.4 Instantaneous terrestrial system

$$\bar{e}^{\,\mathrm{IT}} = \boldsymbol{R}_3(\pi-\Lambda)\boldsymbol{R}_2\left(\frac{1}{2}\pi-\Phi\right)\boldsymbol{P}_2\bar{e}^{\,\mathrm{LA}} = \begin{bmatrix} -\sin\Phi\,\cos\Lambda & -\sin\Lambda & \cos\Phi\,\cos\Lambda \\ -\sin\Phi\,\sin\Lambda & \cos\Lambda & \cos\Phi\,\sin\Lambda \\ \cos\Phi & 0 & \sin\Phi \end{bmatrix}\bar{e}^{\,\mathrm{LA}}. \tag{15.6}$$

Clearly, the IT(τ) system changes its position within the Earth with time because the spin axis moves with respect to the Earth, a motion that is known as polar wobble (polar motion). The x^{IT}-axis moves with time to allow the xz^{IT} plane to pass through the *instantaneous Greenwich Observatory* [ROBBINS, 1976], and its epoch should be that of the observations made.

Step (b): Next, the transformation is from the IT(τ) system into the *apparent right ascension system* or simply *apparent place system* (AP(τ)). The AP(τ) system is another geocentric system in which the z^{AP}-axis coincides with the z^{IT}-axis, and x^{AP}-axis points toward ♈, and y^{AP}-axis completes the system to make it right-handed. The situation is shown in FIG. 15.5. Clearly, the transformation from the IT(τ) to the AP(τ) system consists of rotating the IT system around the common z-axis by the angle known as *Greenwich Apparent Sidereal Time* (GAST), i.e.,

$$\bar{e}^{\,\mathrm{AP}} = \boldsymbol{R}_3(-\,\mathrm{GAST})\,\bar{e}^{\,\mathrm{IT}}. \tag{15.7}$$

GAST and other symbols appearing on this figure will be explained in the next section.

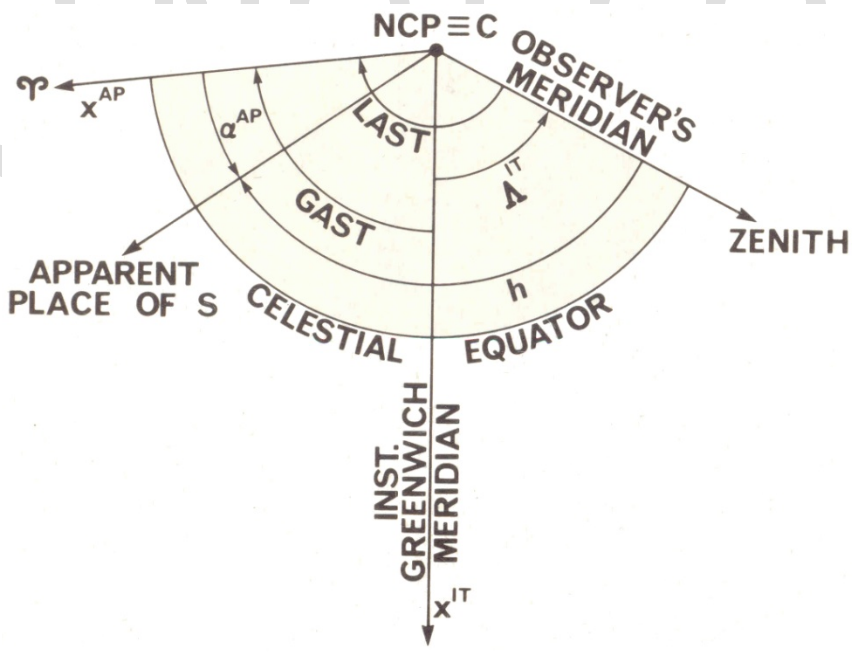

FIG. 15.5. Sidereal time, hour angle, right ascension, and longitude

It should be clear that all three coordinate systems, namely LA, IT, and AP are 'instantaneous' or 'true' systems because they change with time. Astronomical observations usually refer to the instantaneous spin axis of the Earth and as a consequence the astronomical coordinates Φ and Λ as well as azimuth A, derived from the above transformations (steps (a) and (b)) refer to the epoch of observation. This means that the position of a point on the surface of the Earth, as determined by astronomical means, will be time-dependent, which is highly undesirable for most geodetic positioning applications.

The need therefore arises to define a new coordinate system that is attached solidly to the Earth (considered itself in this case solid) but be as close as possible to the IT. This system is the *conventional terrestrial system* (CT), it is the closest practical approximation of the *geocentric* natural system (IT) and it is probably the most important system in geodesy. Its origin is at the centre of mass of the Earth, the z^{CT}-axis points to the CIO (see §5.4), the xz^{CT}-plane contains the *mean Greenwich Observatory* [ROBBINS, 1976], and the y^{CT}-axis is selected to make the system right-handed (cf. FIG. 15.6). It should be noted that the astronomical meridian plane of the observer contains both, the gravity vector of the observer and the CIO; it is thus parallel to the conventional spin axis but generally does not contain the centre of mass of the Earth.

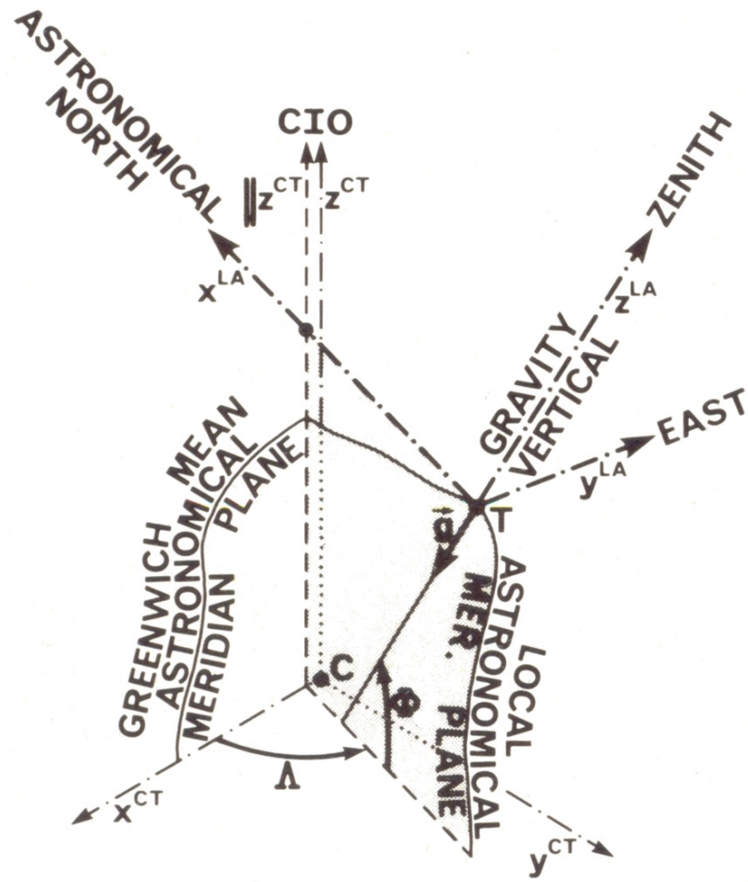

FIG. 15.6. Conventional terrestrial system

Step (c): We now seek a link between IT and CT systems to achieve the transformation of the instantaneous $\Phi(\tau)$, $\Lambda(\tau)$, and $A(\tau)$ into 'mean' quantities that refer to a certain reference epoch τ_0. The IT system is transformed into the *Conventional Terrestrial System* (CT) that differs from the IT system only in so far as its z^{IT}-axis coincides with the instantaneous rather than the conventional spin axis. Thus, the only difference between these two systems is that the z^{IT}-axis wobbles around the z^{CT}-axis, and this wobble is described by the two parameters x_p, y_p in angular units (cf. §5.4); this situation is shown in FIG. 15.7. The transformation is done through the following expression:

$$\vec{e}^{\,CT} = \boldsymbol{R}_2(-x_P)\,\boldsymbol{R}_1(-y_P)\;\vec{e}^{\,IT}. \tag{15.8}$$

Developing the trigonometric functions into power series and neglecting second and higher order terms (as shown already, x_p, y_p are of the order of a few tenths of second of arc), we get

$$\vec{e}^{\,CT} \approx \begin{bmatrix} 1 & 0 & x_p \\ 0 & 1 & -y_p \\ -x_p & y_p & 1 \end{bmatrix} \vec{e}^{\,IT}. \tag{15.9}$$

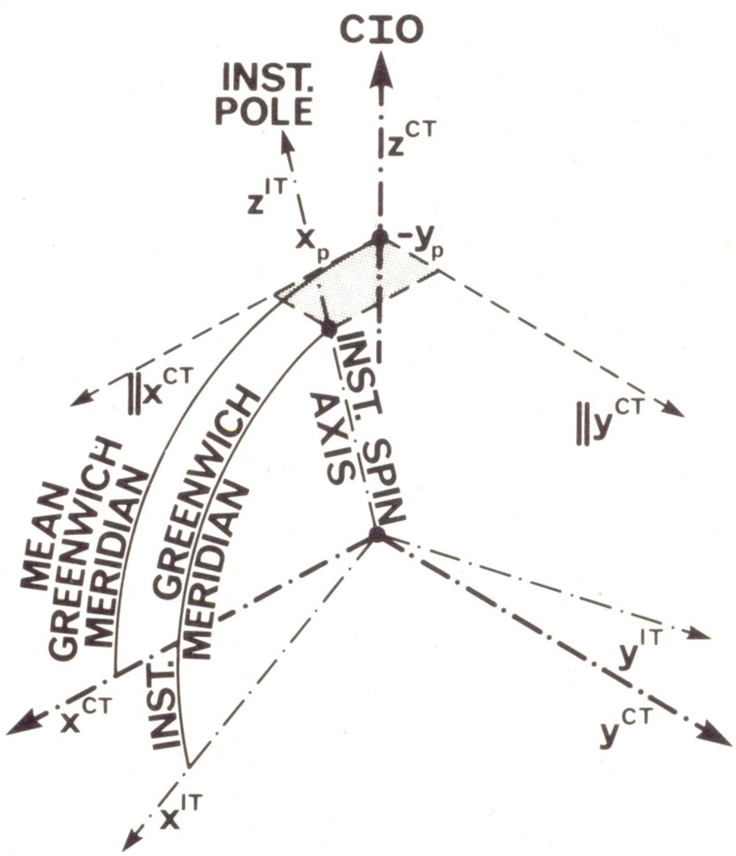

FIG. 15.7. Conventional and instantaneous terrestrial systems.

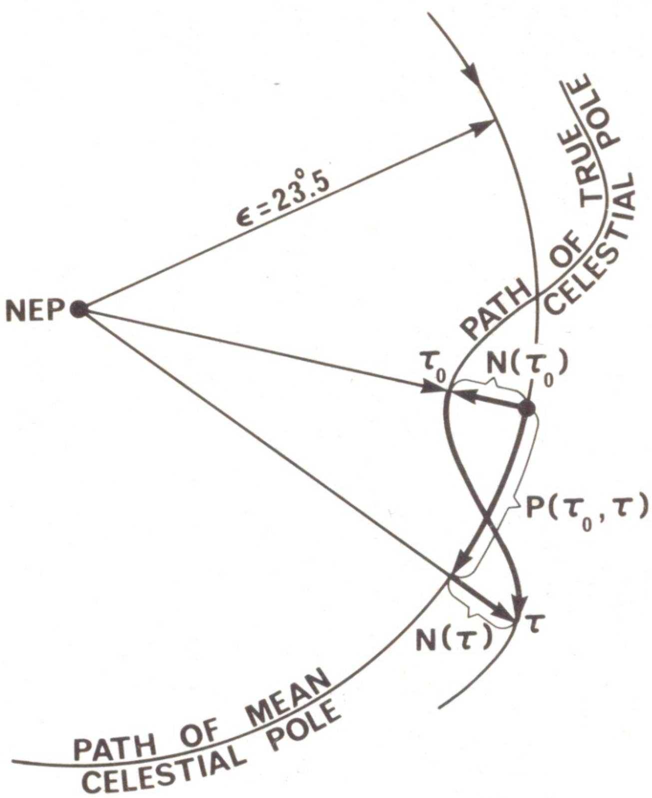

FIG. 15.8. Path of the celestial pole.

Step (d): Let us now return to the MRA(τ_0) system and again follow the path of transformations leading to the AP(τ) system but from the other side (cf., FIG. 15.13). The first step is to transform the MRA system from epoch τ_0 in which the catalogue of star coordinates is published, to epoch τ when the observations are made. To update the coordinates $\alpha^{\mathrm{MRA}}(\tau_0)$, $\delta^{\mathrm{MRA}}(\tau_0)$ to epoch τ, one must account for two effects: the effect of precession during the time interval $\tau - \tau_0$ (see FIG. 15.8), and the effect of proper motion of the stars during the same period.

The amount of precession $P(\tau_0, \tau)$ occurring in a time interval $[\tau_0, \tau]$ is usually spelled out in terms of three *precessional constants* (ζ_0, θ, z) as shown in FIG. 15.9. Expressions for these elements as functions of time were derived early in this century by NEWCOMB [1906]. More recent models have been developed over the past 20 years and they are discussed in the context of the new generation coordinate systems (see §15.5 for more details). The angles ($\tfrac{1}{2}\pi - \zeta_0$) and ($\tfrac{1}{2}\pi + z$) are the right ascensions of the ascending node of the mean equator at τ, measured respectively in the two mean systems (at τ_0 and τ). The angle θ is the inclination between the mean equators at τ and at τ_0. The transformation of α, δ from τ_0 to τ is made first by transforming α and δ into Cartesian components of the corresponding unit vector (cf., Eq. 15.1), then rotating this vector as follows:

$$\bar{e}^{\,\mathrm{MRA}(\tau)} = \mathbf{R}_3(-z)\mathbf{R}_2(\theta)\mathbf{R}_3(-\zeta_o)\bar{e}^{\,\mathrm{MRA}(\tau_0)}, \qquad (15.10)$$

and then back into α and δ by means of Eq. (15.2).

In addition to the apparent motion of the MRA system due to precession, the stars have a *proper motion* of their own. Because this motion appears, for all practical purposes to be linear, it is usually tabulated as constant rates of change in right ascension and declination for each star of interest and is taken care of within the transformation MRA(τ_0)⟶MRA(τ) [MUELLER, 1969].

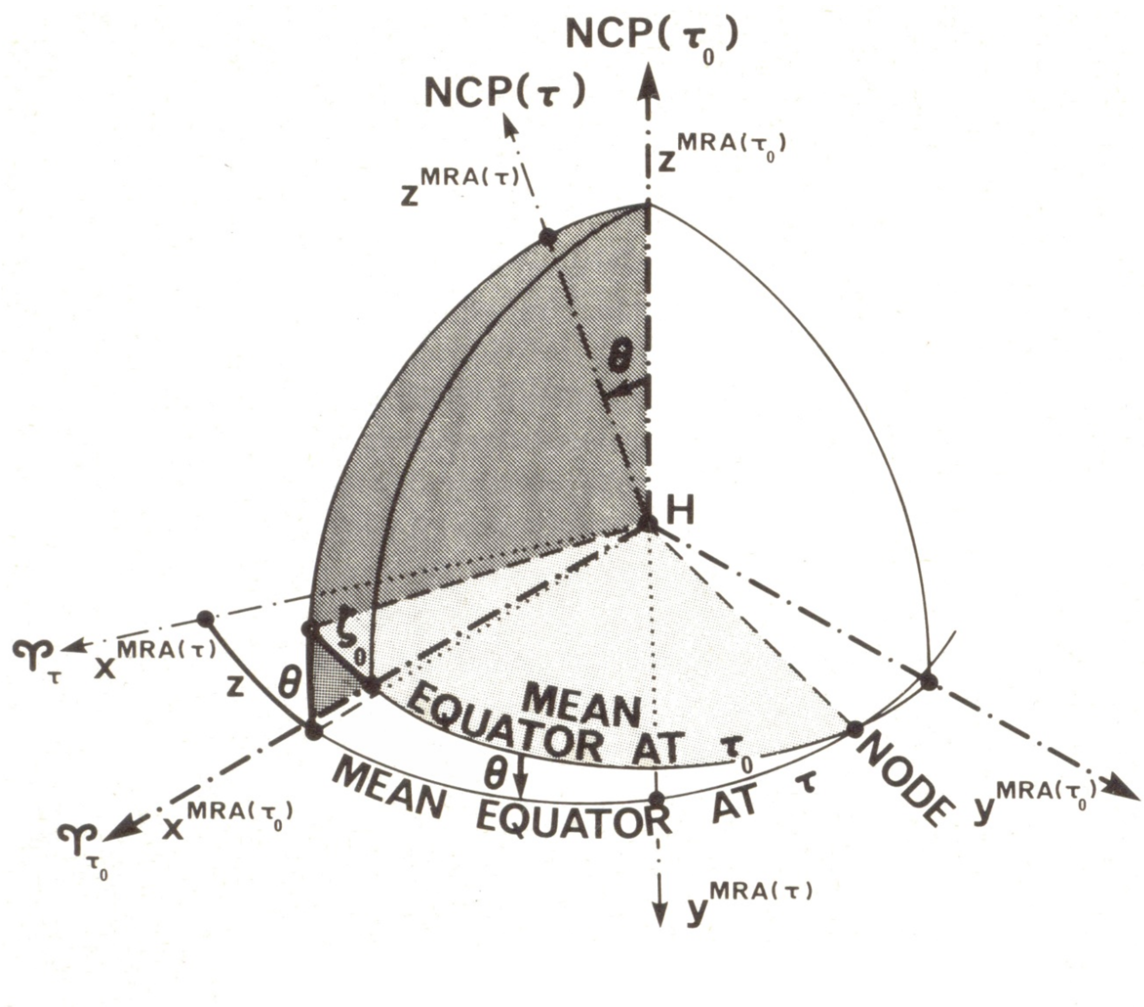

FIG. 15.9. Mean right ascension system and precession

Step (e): The next step in updating the star coordinates is to account for the nutation $N(\tau)$ (FIG. 15.7). This results in the *true right ascension system* at epoch τ, i.e., TRA(τ), whose z^{TRA}-axis coincides with the instantaneous spin axis of the Earth, while the true vernal point defines the direction of the x^{TRA}-axis. The effect of nutation $N(\tau)$ is usually spelled out in terms of *nutation in longitude* $\Delta\psi$ and *nutation in the obliquity* $\Delta\varepsilon$ (FIG. 15.10). The transformation of α and δ from the MRA(τ) to the TRA(τ) system is accomplished by first rotating the Cartesian components of the appropriate unit vector, i.e.,

$$\vec{e}^{\,\mathrm{TRA}(\tau)} = \boldsymbol{R}_1\left(-\varepsilon-\Delta\varepsilon\right)\boldsymbol{R}_3\left(\Delta\psi\right)\boldsymbol{R}_1\left(\varepsilon\right)\vec{e}^{\,\mathrm{MRA}(\tau)}, \tag{15.11}$$

and then transforming these back into α and δ by means of Eq. (15.2). The obliquity angle ε has already been defined in §5.2.

Step (f): The last step in the chain of transformations (updates) taking us to the AP(τ) system once more, is to account for the fact that the star is not observed from the origin H (i.e., from the centre of the Sun) of the RA system, but from the Earth. The heliocentric values of α and δ must thus receive a correction, called *annual parallax* that can be expressed as the parallactic angle of the radius of the Earth's orbit subtended at the star. The parallactic corrections to α and δ are obtained from simple expressions that reflect the position and distance of the star; for the nearest star, the correction is about 0.8″ [MUELLER, 1969].

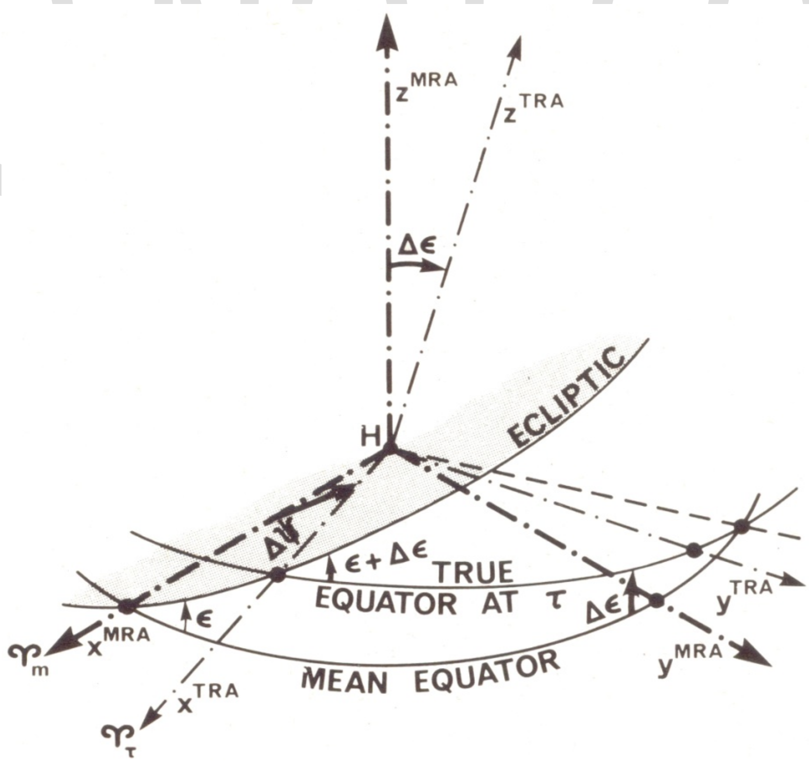

FIG. 15.10. True and mean right ascension systems.

Further, due to the fact that the observations are made from an orbiting (moving) Earth, the light from a star will appear to be coming from a slightly different direction than it actually does – see FIG. 15.11. This effect is called *annual aberration* and does not exceed 20"; it is calculated by means of the constant of (annual) aberration, υ/c, where υ stands for the speed of the Earth in its orbit, and c is the speed of light. In most cases, the above value is only a few seconds of arc as a result of the relative configuration of the Earth, the Sun, and the star in question [SMART, 1962]. It should be mentioned here that all the above corrections (updates) become unnecessary when a special star catalogue, called *Apparent Places of Fundamental Stars* (APFS) is used instead of the standard fundamental catalogues.

Now let us return to steps (a) and (b) of the transformation chain LA(τ)→AP(τ) (see FIG. 15.13) and consider what has to be done to the observations to make them compatible with the apparent places of the stars. There are three effects to be considered in this context: diurnal aberration, diurnal parallax, and refraction. *Diurnal aberration* is a result, again, of making measurements from an earthbound observing station spinning with the Earth. It is the translational velocity υ of the station that causes the star to undergo an apparent shift analogous to the one shown in FIG. 15.11. The constant of diurnal aberration υ/c ($\upsilon<2\pi R/\text{day}\approx 463\ \text{ms}^{-1}$) corresponds to a maximum correction of 0.3" on the equator. It is rather small, but as it is systematic and should be removed from the observations. Diurnal parallax is caused by the parallactic angle of the Earth's radius subtended at the star and it is always negligibly small. *Astronomical refraction*, the bending of the light rays from the stars as they enter the Earth's atmosphere (cf. §9.2), seriously affects zenith distance measurements and, to a lesser degree, even the azimuth measurements. This effect will be discussed in the context of the mathematical models developed in §15.3.

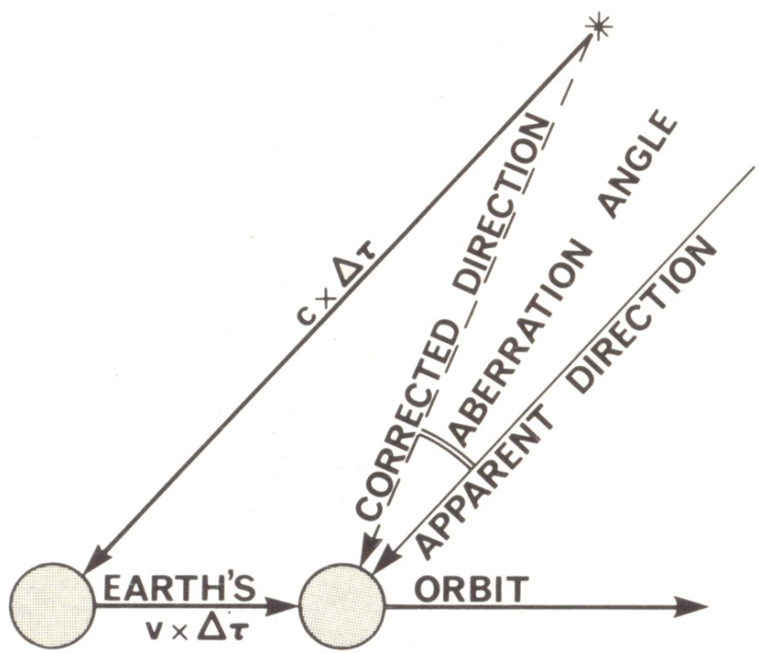

FIG. 15.11. Aberration

Sometimes there is a need for a coordinate system to be motionless with respect to the distant objects , outside our galaxy[1]. Such a system is called an *inertial system* and is characterised by the fact that Newton’s laws of motion apply in it: the system is at rest or moving with uniform rectilinear motion without rotation with respect to the distant stars and as such it has the property of generating no virtual acceleration on objects reckoned in that system. A good approximation of such an inertial system is the *ecliptical system* (E). It is heliocentric, its z^E-axis coincides with the Earth's precession axis (see §5.2), the x^E-*axis* points toward ♈, and y^E is chosen to make the system right handed (see FIG. 15.12). It uses *ecliptical latitude* β^E and *ecliptical longitude* λ^E as shown. This system is almost inertial except for the precessional shift of the *planetary precession* (see §5.2), i.e., the precession caused by the planets in the solar system, and the very slow movement with the galaxy (see §5.1). As we will see later (§15.5), the best approximation to the ‘true’ inertial system, when defining the new generation of coordinate systems, is achieved by a coordinate system whose origin is at the solar system barycentre[2] and with the coordinate axes fixed to the distant quasars; it is at present the best one to use for star coordinate publications.

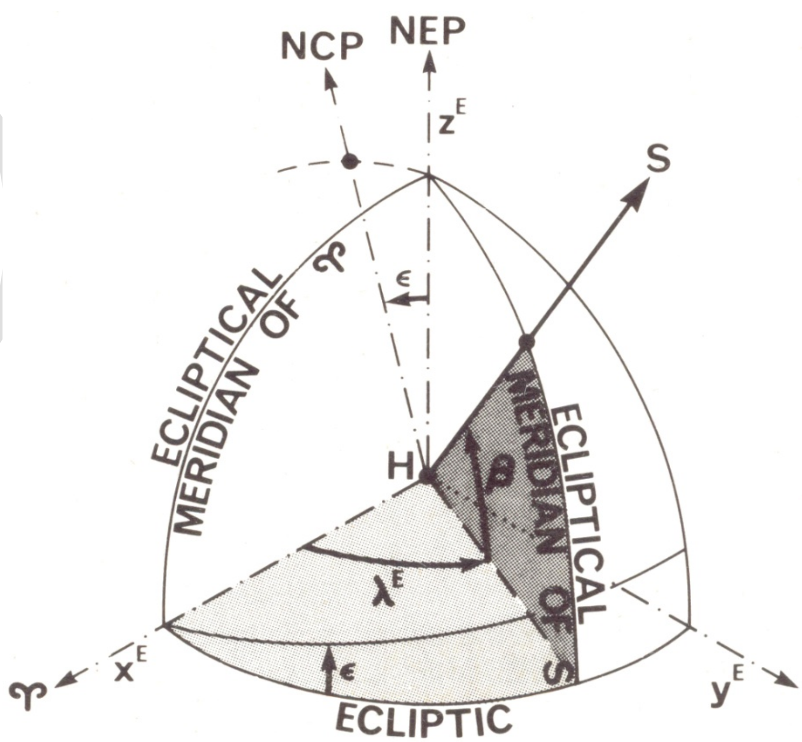

FIG. 15.12. Ecliptical system.

[1] The classical coordinate systems were defined in a period prior to the introduction of radio astronomical observations (VLBI) and thus only distant stars in other galaxies were visible with optical astronomy.
[2] Centre of mass of the Sun and planets taken together.

15.2. Time Systems

The last concept essential for the transformation among the different coordinate systems in astronomical positioning, but most importantly in the definition of the new coordinate systems as used today internationally, is the concept of time. Odd as it may seem, time in coordinate system transformations may be interpreted as merely an angle between two corresponding axes of two particular coordinate systems, as we have already seen in the case of GAST (cf., Fig. 15.7). Similar to the three different classes of coordinate systems (celestial, terrestrial, and orbital), there are three different classes of time systems: a) the *Earth rotation time system* based on the spin of the Earth with respect to the Sun (solar time) or with respect to the stars (sidereal time); b) the *atomic time system* that is based on the ticking of an electronic oscillator that is kept steady by comparing it to the frequency of the emitted microwave signal from an atom as its spinning electrons change energy levels, for example cesium-133; and c) the *Global Positioning System (GPS) time system* that is used by the United States Naval Observatory (USNO) to monitor the timing of the GPS satellites. A detailed presentation of the astronomical time scales can be found in Seidelmann and Fukushima [1992] and in the Explanatory Supplement to the Astronomical Almanac [Seidelmann et al., 1992; Chapter 2].

Before we define the above time systems, let us begin with a time scale that is associated with all three classes of time systems, the *Julian Date* (JD). The JD has been in use for centuries by astronomers, geophysicists, and others as unambiguous dating system that uses a day count. The JD day count starts at 12:00 noon, January 1, of year -4712 (4713 BC) of the Julian proleptic calendar, i.e., of one Julian Period of exactly 7980 Julian years of 365.25 days. In the late 1950's the *Modified Julian Date* (MJD) was introduced as MJD = JD – 2400000.5 [Seidelmann et al., 1992; Chapter 2].

The oldest of the three time systems, i.e., the *Earth rotation time system*, may be divided into two families, namely the *solar time system* and the *sidereal time system*. The Earth spins about its rotation axis with respect to the Sun once in a *solar day* and the *solar second* was historically defined as 1/86400 of a mean solar day. With the development of precise clocks in the early 1900, it was soon discovered that the length of the solar day was not constant and thus not adequate for high precision measurements. In 1956, the International Committee for Weights and Measures, defined the '*second' of time* as an appropriate fraction of the period of revolution of the Earth around the Sun for a particular epoch based on Newcomb's '*Tables of the Sun*' and astronomical observations made in the eighteenth and nineteenth centuries. This time scale was called '*ephemeris time*' or ET and the '*ephemeris second*' was defined as '*the 1/31556925.9747 of the time between the passage of the Sun through successive vernal points (tropical year) for 1900 January 0 12h ephemeris time*'. ET was the time scale that was used in the *Astronomical Almanac* for over 20 years until it was replaced by the terrestrial dynamical time (see below) in 1984, after the implementation of the IAU 1976 System of Astronomical Constants. In this family of solar times there are three time scales [e.g., Seidelmann et al., 1992; Chapter 2]:

(a) UT0: *Universal Time Zero* that reflects the actual non-uniform spin of the Earth with respect to the Sun. It is burdened with the effect of polar motion in as much as the local astronomical meridians defining the UT (through Λ) are slightly displaced.

(b) UT1: *Universal Time One* (or just UT – formerly called Greenwich Mean Time) again depicts the actual non-uniform rotation of the Earth, but the effect of polar motion is subtracted.

(c) UT2: *Universal Time Two* is the smoothest UT with all known corrections applied. UT2 is not useful for geodetic or geophysical applications and is mostly of historical interest. The

Explanatory Supplement to IERS Bulletins A and B [SEIDELMANN ET AL., 1992; p. 50] contains details on the corrections applied. The UT0 and UT1 are related through the following relation

$$\mathrm{UT0} = \mathrm{UT1} + \tan\Phi\left(x_p \sin\Lambda + y_p \cos\Lambda\right), \tag{15.12}$$

where x_p and y_p are the pole offsets published in IERS Bulletin A and Φ and Λ are the observer's coordinates in the CT system.

The second family in the Earth rotation time systems is the *sidereal time* that considers the spin rate of the Earth with respect to the 'fixed' stars instead of the Sun. The corresponding spin period is called 's*idereal day*', which is on average four minutes shorter than a solar day, a difference that accumulates to one full day in one solar year. In context with the coordinate system transformations, and in particular between AP and IT, we have seen that '*sidereal time*' of a location on the surface of the Earth is the time elapsed since the transit of the local meridian across the true vernal point and is directly related to the right ascension (cf., Eq. 15.9; FIG. 15.5). In this system we often use the notion of the *hour angle h* of a star *S*, which is the angle between the astronomical meridian of *S* and that of the observer. In the family of sidereal times we recognize the following time scales [SEIDELMANN ET AL., 1992; p. 50]:

(a) GAST: *Greenwich apparent sidereal time* is the time elapsed since the transit of the Greenwich meridian across the true vernal point (true equinox of date), i.e., it is the hour angle of the true vernal point as seen at Greenwich (Greenwich Hour Angle – GHA). GAST is affected by precession and nutation.

(b) GMST: *Greenwich mean sidereal time* is the time elapsed since a transit of the Greenwich meridian across the mean vernal point (mean equinox of date). The GMST is GAST corrected for nutation effects, i.e., it is affected only by precession; it is linked directly to UT1 through the following equation [SEIDELMANN ET AL., 1992; p. 50]:

$$\mathrm{GMST} = 24110.54841 + 8640184.812866T + 0.093104T^2 - 0.0000062T^3 \quad , \tag{15.13}$$

where T is time in Julian centuries of 36525 days from year 2000, January 1, at 12:00 UT1, (epoch J2000.0). The above formula gives GMST in seconds. GAST and the GMST are related through the approximate (with an error less than 2μs) equation of equinoxes reckoned in time units (e.g. ASTRONOMICAL ALMANAC, 2012):

$$\begin{aligned}\mathrm{GAST} &= \mathrm{GMST} + \text{Equation of equinoxes} \approx \\ &\mathrm{GMST} + \frac{1}{15}\left(\Delta\psi\cos\varepsilon + 0''.00264\sin\Omega + 0''.000063\sin 2\Omega\right),\end{aligned} \tag{15.14}$$

where $\Delta\psi$ is the total nutation in longitude, ε is the mean obliquity of the ecliptic and Ω is the mean longitude of the ascending node (to be defined in §15.5) of the Moon. Note the factor 1/15 that translates angular to time units). Precise GAST, GMST and the equation of equinoxes are also tabulated in the Astronomical Almanac at 0^h UT1 for each day and should be interpolated to the required time if full precision is required.

(c) LAST: *Local apparent sidereal time* is the time elapsed since the transit of the local meridian (observer's meridian) across the true vernal point or equivalently, it is the hour angle of the true vernal point as seen by the observer.

(d) LMST: *Local mean sidereal time* is the time elapsed since the transit of the local meridian across the mean vernal point. Clearly (cf., FIG. 15.5),

$$
\begin{aligned}
\mathrm{LAST} &= \mathrm{GAST} + \Lambda^{\mathrm{IT}} \\
\mathrm{LMST} &= \mathrm{GMST} + \Lambda^{\mathrm{CT}}
\end{aligned}
\qquad (15.15)
$$

The second class of time systems is the class of *atomic times* that was introduced as a reproducible time scale independent from the irregularities of the Earth's rotation. This time scale is defined by the 'ticking' of an electronic oscillator that is kept steady by constraining it to the frequency of the emitted microwave signal from an atom as its spinning electrons change energy levels. The atomic second (SI unit) is defined as 9192631770 times the period of the microwave signal emitted by the cesium-133 atom [BIPM, 1998]. The Hydrogen atom allows a more accurate measurement of the 'atomic second' than cesium-133 because the period of the microwave signal emitted by it is much shorter. However, the 'hydrogen maser' has worse long-term stability than cesium-133.

There are two distinct families of atomic time scales in use, namely the '*coordinate time*' and '*dynamic time*'. The former scale is defined by a clock whose position is fixed with respect to the coordinate system in question and in absence of any gravitational potential that is, it has been corrected for general and special relativistic effects. On the other hand, the '*dynamic*' time scale is defined by a clock that moves with respect to the considered coordinate system in the presence of a gravitational potential, and it is thus a relativistic time scale[3]. Coordinate time ticks faster than its dynamic counterpart at a rate of about one-half second per year.

Coordinate and dynamic time scales can be considered and used on the surface of the Earth or other celestial bodies, on spacecraft, or at other locations of interest in space, such as the solar system barycenter [Kaplan, 2005]. The need to define such location of the 'clock' arises because atomic time appears as an independent argument in many theoretical developments of dynamical astronomy [e.g., Moisson and Bretagnon, 2001]. As expected, conversions between coordinate and dynamic time scales and transformations between Earth rotation and atomic time scales are necessary; but such conversions are generally rather complicated and they are beyond the scope of this book. The interested reader should refer to Kaplan [2005]. It is important to note however that these transformations are inherently four-dimensional, i.e., they are space-time transformations that depend on the relative trajectory of the clock and the observer as well as the gravitational fields involved [Kaplan, 2005]. This makes the atomic time scales an integral part of the modern reference coordinate systems in a more involved way than the Earth rotation time scales (e.g., UT, GAST) played in the transformations among the classical reference coordinate systems as simple rotations.

International Atomic Time (known under its French name "Temps Atomique International" –TAI) is the primary time standard in the world today and it is defined as the average time kept by more than 200 atomic clocks around the world. Atomic clocks are influenced by a number of environmental effects and appropriate corrections are applied by the BIPM. TAI is referred to the mean sea level (geoid) through its conventionally accepted geopotential value of W_0 and as such TAI is relativistic terrestrial time: the 'hypothetical' atomic clock is spinning with the Earth under the influence of the Earth's gravitational field. Finally, TAI is the weighted average of all participating atomic clocks and the results are disseminated monthly in *Circular T* of the BIPM. The zero point of TAI was somewhat arbitrarily defined so that UT1-TAI was approximately equal to zero on January 1, 1958. Its offset from the ephemeris time was precisely defined as 32.184s for January 1, 1977 [Astronomical Ephemeris and Nautical Almanac, 2006].

[3] The clock is affected by Special and General Relativity

The following atomic time scales are used today:

(a) UTC: *Coordinated Universal Time* is broadcast by the National Institute of Standards and Technology (NIST) radio station WWV (USA), the National Research Council radio station CHU (Canada) and others. The UTC and TAI have the same rate, but the UTC is kept within about 0.9 seconds of mean solar time UT1 by adding to it leap seconds, typically once per year but at times more often, to keep pace with the slowing of the Earth's rotation [KAPLAN, 2005]. Leap seconds are announced in the *IERS Bulletin C*. The UTC is the civilian time we use today that differs from time zones by an integer number of hours.

(b) TDT or TT: *Terrestrial dynamic time* or simply *terrestrial time* is a relativistic time scale used today in the apparent astronomical ephemerides for astronomical observations taken from the Earth's surface; TT is also referred to the geoid (at geopotential W_0 = const.). The TT replaced the ephemeris time in January 1, 1984 and has the same rate as TAI but differs from it by a constant offset of 32.184s in order to maintain continuity between the old system (ET) and TT. The TT is related to TAI and UTC by the following formulae [KAPLAN, 2005]

$$\mathrm{TT} = \mathrm{TAI} + 32.184\mathrm{s} = \mathrm{UTC} + (\text{number of leap seconds}) + 32.184\mathrm{s}. \tag{15.16}$$

(c) TCG: *Geocentric coordinate time* is referred to the geocentre C and differs from TT only by relativistic effects that amount to about 0.7 parts per billion. The TCG is used as the independent variable of time for all calculations pertaining to precession, nutation, and artificial satellites of the Earth.

(d) TDB: *Barycentric dynamic time* is a relativistic time referred to the barycenter of the solar system but measured by atomic clocks on Earth! This scale was introduced by the IAU in 1976 but obviously in an erratic manner; it ceased being used in 1991 when it was replaced by the barycentric coordinate time (see below) and TCG.

(e) TCB: *Barycentric coordinate time* is the time scale which would be kept by an atomic clock at rest with respect to the solar system barycenter, but moving with it in the galaxy in the absence of any gravitational potential. Therefore, TCB and TT differ only by relativistic effects. The TCB is used in theoretical calculations in dynamical astronomy, such as those of planetary motion [MOISSON AND BRETAGNON, 2001].

The third class of time systems is the *Global Positioning System time* – GPS time, which is markedly different from the other two classes in that it is based on the so called '*composite clock*' or '*paper clock*' that is a combination of all time and frequency standards used at all GPS monitoring stations. The GPS time is monitored and managed by the U.S. Naval Observatory (USNO) for the purpose of providing reliable and stable coordinated time reference for the whole GPS system, and its epoch is 00:00h UTC on January 6, 1980. The GPS time is not adjusted and thus differs from UTC by the number of leap seconds introduced since 1980. This offset is part of the navigation message transmitted by the GPS satellites (see §15.4) that is used by the receivers automatically to display UTC rather than GPS time. On the 1st of January 1999, GPS time was ahead of UTC by 13s, TAI was ahead of GPS time by 19s and of UTC by 32s.

It may now be useful to close this section by recapitulating the coordinate systems introduced in this section and the time systems associated with them as well as the transformations among them. This is done in FIG. 15.13.

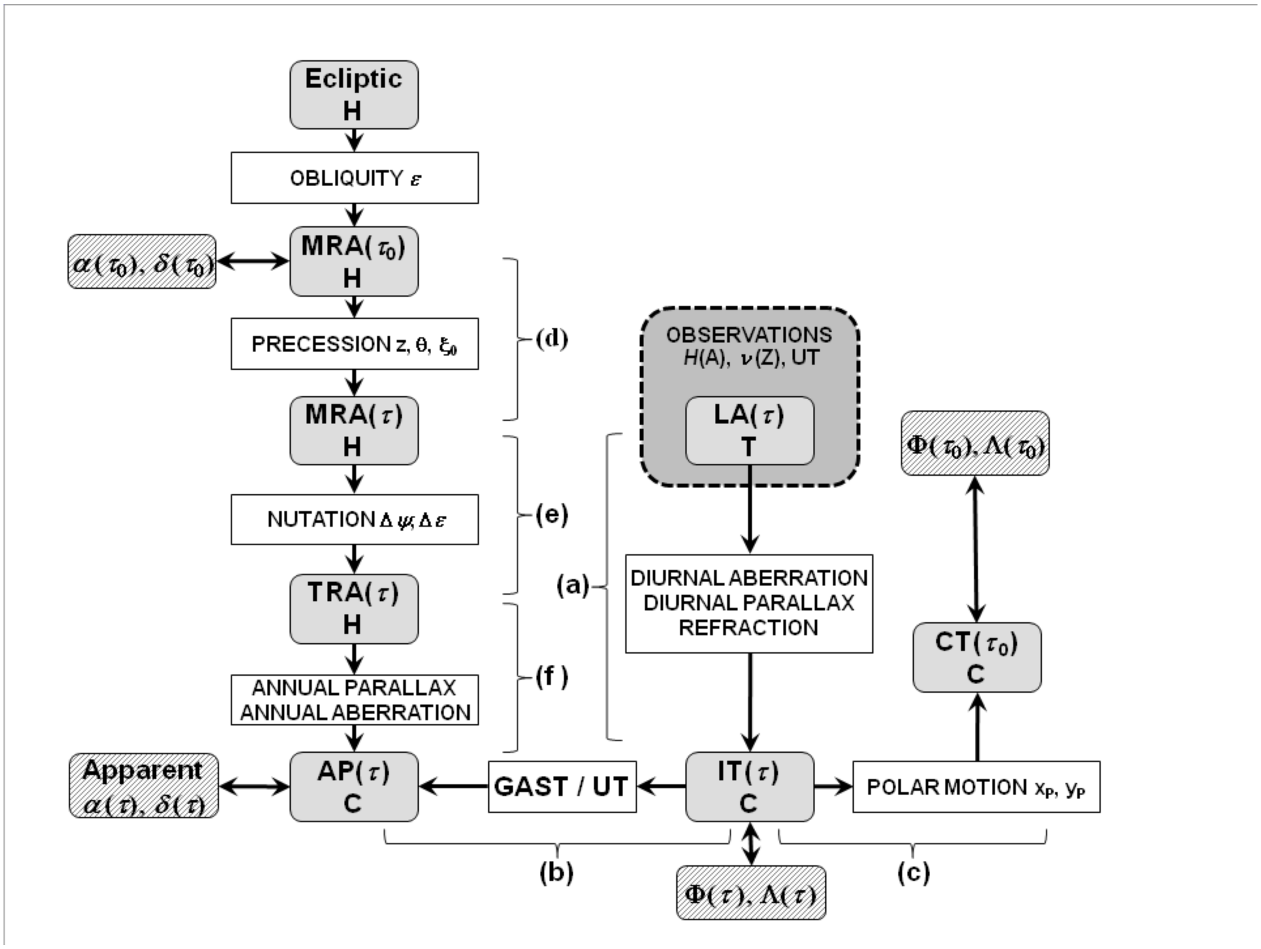

FIG. 15.13. Overview of the classical coordinate systems and their transformations. The different transformation steps described in the text are indicated with lower case letters (a) through (f). The T, C and H indicate topocentric, geocentric and heliocentric systems, respectively. Shaded areas show coordinate systems, hatched blocks indicate the coordinates resulting from the system, and open boxes show the transformation parameters.

15.3. Astronomical positioning

Geodetic astronomical positioning (at optical wavelengths) has been diminished significantly over the last 20 years due to the widespread availability and ease of use of satellite positioning systems. However, geodetic astronomy is still applied locally for the determination of the direction of the gravity vector and astronomical azimuths. More importantly, geodetic radio-astronomy using Very Long Baseline Interferometry–VLBI (to be discussed in Chapter 16) plays a leading role in the definition of the celestial coordinate systems by making observations to the distant quasars, while it provides highly accurate relative positioning capability to continental and transcontinental scales. In view of the above, only the fundamentals of the optical astronomical positioning will be given here.

Astronomical positioning is understood to be the determination of the astronomical latitude Φ and longitude Λ of a point by means of specific observations ℓ to stars. The determination of the astronomical azimuth A to another point (by astronomical means) has traditionally been considered an integral part of this task, and it is thus treated here as well. The most widespread mathematical models linking the observables ℓ with the astronomical positions or azimuth fall into three classes:

(a) latitude models: $\boldsymbol{f}_1(\Phi,\ell)=0$,

(b) longitude models: $\boldsymbol{f}_2(\Lambda,\ell)=0$,

(c) azimuth models: $\boldsymbol{f}_3(\mathrm{A},\ell)=0$.

Clearly, Φ and Λ can be determined separately or together depending on the form of the model and the observables ℓ involved. Observations can be collected with different accuracies, depending upon the instrument(s) and methodology used. Implicitly involved in all the above models are the coordinates of the observed stars; throughout this section, we will consider these coordinates (α, δ) to be known in the AP system, obtained for instance from the Astronomical Almanac (see §15.1 and §15.2).

Most commonly used latitude models require the observations of two quantities: the zenith distance Z (or vertical angle v), and the hour angle h of a star. The latter however, cannot be measured directly; it is determined as the sum of α, obtained for the particular star from a star catalogue, and of measured time (cf., FIG. 5). Longitude models need only the knowledge of h, i.e., the time and the α of the star used. Simultaneous models for Φ and Λ are characterised by requiring the measurements of the zenith distances (Z_1 and Z_2) and hour angles (h_1 and h_2) to at least two stars. Azimuth models use either Z or h and the horizontal angle between the selected star and the desired point. Models also exist which do not fit into this classification; for these, the reader is advised to consult, e.g., MUELLER [1969].

The basic instrument required for measuring zenith distances is the universal theodolite. Specialised forms of theodolites are employed for very high accuracies. Other instruments may also be used; these are mentioned within the context of the mathematical models. For precise timing, one needs a chronometer (timepiece) and an HF radio receiver equipped with an amplifier and chronograph. Auxiliary equipment to measure air temperature and pressure is also used because these are needed for determining the vertical refraction correction to zenith distances. What follows is the development of the mathematical models used in astronomical positioning based on the traditional coordinate systems. We will then extend these models to apply to the new generation of coordinate systems (Section 15.5).

(a) The *latitude mathematical model* is obtained from the transformation between the AP and (instantaneous) LA systems. One gets after combining Eqs. (15.6) and (15.7) (inverse transformations)

$$\vec{e}^{\,\mathrm{LA}} = \boldsymbol{R}_2^{-1}\left(\frac{\pi}{2} - \Phi\right)\boldsymbol{R}_3^{-1}(\pi - \Lambda)\boldsymbol{P}_2\,\boldsymbol{R}_3(\mathrm{GAST})\,\vec{e}^{\,\mathrm{AP}}\ . \tag{15.17}$$

By expressing the Cartesian components of $\vec{e}^{\,\mathrm{AP}}$ and $\vec{e}^{\,\mathrm{LA}}$ as functions of (α, δ) and (ν, A) from Eqs. (15.1) and (15.4), respectively, and after some development, the equation for the third component of Eq. (15.4) is obtained in the following form:

$$\boxed{\sin\Phi\sin\delta + \cos\Phi\cos\delta\cos h - \cos Z = 0.} \tag{15.18}$$

Before the observed zenith distance Z is inserted in the above model, or any model for that matter, it should be corrected for the astronomical refraction effect. The *zenith distance astronomical refraction correction* ΔZ follows from Eq. (9.12) namely,

$$\Delta Z = -\tan Z \int_1^{n_0} \frac{dn}{n}, \tag{15.19}$$

where the integration is made from outside the atmosphere (where n = 1) to the observer (where $n = n_0 > 1$). Since n as a function of time and location, it is not known exactly and the above integral cannot be evaluated exactly either. A practical solution can be arrived at by adopting a model for the composition of the atmosphere. Using this model, the integral can be approximated, usually in the form of a series. For details see, e.g., SAASTAMOINEN [1972; 1973] and KOVALEVSKY, [1995]. The final result of this approximation, valid for $Z \le 75^0$, is [SEIDELMANN, 1992]

$$\Delta Z = 0^\circ\!.00452\, P \tan Z \,/\, (273 + T)\ , \tag{15.20}$$

where P is the barometric pressure (millibar) and T is the temperature (ºC) at the time of observation. The accuracy of the refraction correction determined from Eq. (15.20) depends on the deviation of the actual atmospheric conditions from the model (standard) atmosphere but it is at the sub-arcsecond level. For Z>75º other approximate formulae exist [see e.g., SEIDELMANN, 1992]. We can now derive the expression for the azimuth A of a star as a function of Φ, δ, and h. This expression is obtained from Eqs. (15.1), (15.4), and (15.6) as

$$\boxed{\tan \mathrm{A} = \frac{\sin h}{\sin\Phi\cos h - \tan\delta\cos\Phi}\ .} \tag{15.21}$$

The remaining troublesome effect is that of the unaccounted residual astronomical refraction on the zenith distance, after correction ΔZ has been applied (cf., Eq. (15.20)). It can nearly be eliminated by observing pairs of stars symmetrical with respect to and close to the zenith. Residual refraction in zenith distances of such stars nearly cancels out; the cancellation occurs because of the symmetrical composition of the atmosphere with respect to the zenith. The method based on this fact is called the *method of latitude determination by meridian zenith distances* and is in widespread use. For a star-pair transiting the meridian in upper culmination (north and south of the zenith), the model becomes [MUELLER, 1969]

$$\boxed{\Phi = \frac{1}{2}\left(\delta_S + \delta_N\right) + \frac{1}{2}\left(Z_S - Z_N\right),} \tag{15.22}$$

while for a star-pair in lower culmination (north, denoted by subscript N, and south, denoted by S, of the zenith), the model is

$$\boxed{\Phi = \frac{1}{2}\left(\delta_S - \delta_N\right) + \left(Z_S - Z_N\right) + \frac{1}{2}\pi \, .} \tag{15.23}$$

High accuracy latitude determination ($\sigma_\Phi = 0.2''$) can be achieved by employing the above models (Eqs. (15.22) and (15.23)) and by using accurate instrumentation and measuring techniques.

(b) The *longitude mathematical model* is based on the use of Eqs. (15.15), where the GAST is obtained through the UTC by synchronising a local timepiece (e.g., a quartz crystal chronometer) with a time standard by means of HF radio time signals, and the LAST is obtained from observed quantities. Specifically, from FIG. 15.5,

$$\mathrm{LAST} = h + \alpha \, , \tag{15.24}$$

and h is obtained from the third Eq. (15.17) as

$$\cos h = \frac{\cos Z - \sin\delta \sin\Phi}{\cos\delta \cos\Phi} \, , \tag{15.25}$$

so that

$$\boxed{\Lambda = \cos^{-1} \frac{\cos Z - \sin\delta \sin\Phi}{\cos\delta \cos\Phi} + \alpha - \mathrm{GAST}.} \tag{15.26}$$

Clearly, latitude Φ has to be known and Z observed.

Different observational techniques can be applied to reduce the systematic errors in the longitude termination, such as the *method of longitude determination by zenith distances*, and the *method of longitude determination by transit times*. The interested reader can find more details in THORSON [1965] and MUELLER [1969].

(c) The *astronomical azimuth mathematical model* uses Eq. (15.21). There, latitude Φ must be known, and the hour angle h is obtained from Eqs. (15.15) and (15.24), taking Λ^{IT} to be known and the GAST as derived from the UTC. The horizontal angle between the star S and the desired terrestrial point is measured directly by a high accuracy universal theodolite or a geodetic theodolite (low accuracy) and added to the azimuth A of S.

To finish this section, let us point out that all the methods shown above give astronomical quantities (Φ, Λ, and A of a terrestrial point) in the IT coordinate system, because the observations made in the LA system were transformed directly into the AP system using Eqs. (15.6) and (15.7). If their counterparts, Φ^{CT}, Λ^{CT}, A^{CT} in the CT system are needed, the transformation given by Eq. (15.9) is used. Realising that a unit vector in the CT system (and similarly for the IT system) is given as

$$\bar{e}^{\mathrm{CT}} = \begin{bmatrix} \cos\Phi^{\mathrm{CT}}\cos\Lambda^{\mathrm{CT}} \\ \cos\Phi^{\mathrm{CT}}\sin\Lambda^{\mathrm{CT}} \\ \sin\Phi^{\mathrm{CT}} \end{bmatrix}, \tag{15.27}$$

yields, for instance for the third component,

$$\sin\Phi^{\mathrm{CT}} = \sin\Phi^{\mathrm{IT}} + \cos\Phi^{\mathrm{IT}}(y_p \sin\Lambda^{\mathrm{IT}} - x_p \cos\Lambda^{\mathrm{IT}}). \tag{15.28}$$

Division by $\cos\Phi^{\mathrm{IT}}$ and development of $\sin\Phi^{\mathrm{CT}}$ into a Taylor series around Φ^{IT} yields

$$\boxed{\Phi^{\mathrm{CT}} \approx \Phi^{\mathrm{IT}} + y_p \sin\Lambda - x_p \cos\Lambda.} \tag{15.29}$$

The equation for the longitude follows similarly from the transformation of the first two components:

$$\boxed{\Lambda^{\mathrm{CT}} \approx \Lambda^{\mathrm{IT}} - \left(x_p \sin\Lambda + y_p \cos\Lambda\right) \tan\Phi\ .} \tag{15.30}$$

The corresponding equation for the azimuth is derived from the reasoning that the observer's astronomical meridian is displaced by the polar motion. This leads to the following simple result [MUELLER, 1969]:

$$\boxed{A^{\mathrm{CT}} \approx A^{\mathrm{IT}} - \left(x_p \sin\Lambda + y_p \cos\Lambda\right) \sec\Phi\ .} \tag{15.31}$$

Note that the superscripts CT and IT have been dropped in some terms in the above three equations as it is immaterial which values are used to evaluate the corrective terms. In the rest of this chapter, the superscript CT will be similarly dropped.

15.4. Satellite positioning

Satellite methods utilise artificial Earth satellites to which ranges, range differences, directions, and combinations thereof, are measured for the purpose of determining the coordinates of the observing, also known as tracking station. *Satellite point positioning*, in contrast with the other satellite methods to be discussed in §16.1 and §17.3, requires the coordinates (position) of the satellite to be known. These are usually given in the orbital coordinate system.

In defining this system, we imagine that in the first approximation the satellite obeys Kepler's laws (cf. §5.1) and moves along an *orbital ellipse* with the Earth's centre of mass at one of the foci (cf. FIG. 15.14). When the satellite's orbit is close to the Earth, it is perturbed by the irregularities of the Earth's gravitational field and deviates from a plane ellipse – see Chapter 23. Nevertheless, it is convenient to regard the orbit to be, in the first approximation, Keplerian, i.e., planar and elliptical, and treat the perturbations as temporal variations of the six elements describing such a *Keplerian motion*. Thus, in our development, these orbital elements will be functions of time.

The point of the satellite's closest approach to the Earth is called the *perigee,* and the point of farthest recession the *apogee.* Both, the perigee and the apogee lie at the ends of the major axis of the orbital ellipse, called the *line of apsides.* The size and shape of the orbital ellipse are usually defined using the major semi-axis a_0 and eccentricity e, much the same way as for the meridian ellipse in §7.3. The relation between these quantities and b, the minor semi-axis of the ellipse, is given by Eq. (7.11).

Now, consider the satellite to be at a point S on the orbital ellipse. The angular distance between the perigee and S is called the satellite *anomaly.* There are three kinds of anomalies in use. The *true anomaly* f (cf., FIG. 15.14) is the angle between the line of apsides and the line joining the centre of mass of the Earth with the satellite, reckoned from the perigee in an anticlockwise direction when viewed from the NCP or from the vernal point for orbital planes that contain the NCP and SCP (polar orbits). The *eccentric anomaly* E is the angle between the line of apsides and the line joining the geometrical centre of the ellipse with the projection of the satellite S' on the concentric circle of radius a_0. The *mean anomaly* μ (not shown in FIG. 15.14) is the true anomaly corresponding to the motion of an imaginary satellite of uniform angular velocity; $\mu = 0$ at the perigee and then increases linearly with time at a rate of 2π per revolution.

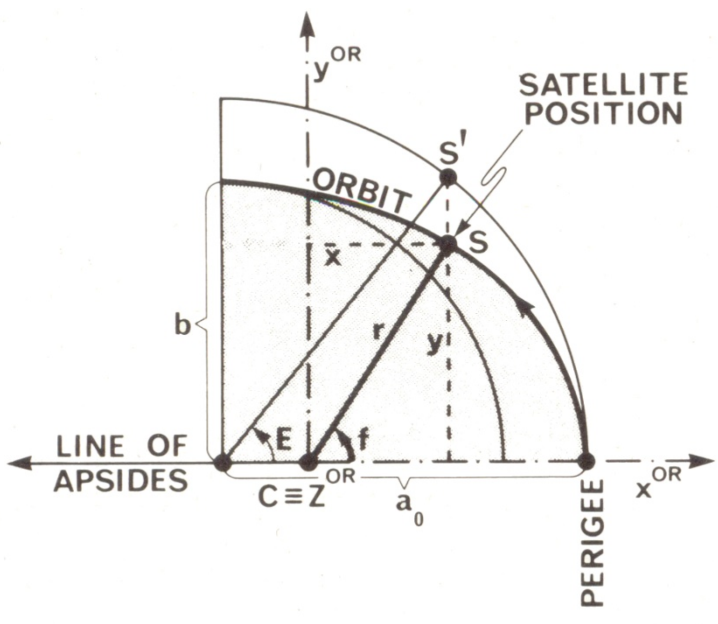

FIG. 15.14. One-quarter of a satellite orbital ellipse

After some development, the relation between the true and eccentric anomalies is obtained from FIG. 15.14 as

$$\boxed{\tan f = \frac{(1-e^2)^{1/2}\sin E}{\cos E - e}.} \tag{15.32}$$

The relation between the eccentric anomaly E and the mean anomaly μ is given by *Kepler's equation* [KAULA, 1966]:

$$\boxed{\mu = E - e\sin E.} \tag{15.33}$$

Often, the mean anomaly μ is given, and it is necessary to find the eccentric anomaly E from Eq. (15.44). This can be done by iterations or by developing sin E into a power series and inverting the equation [BROUWER AND CLEMENCE, 1961; SEEBER, 2003].

Now we can define the *orbital coordinate system* (OR) as follows (cf., FIG. 15.14): the origin is at the Earth's centre of mass C, the x^{OR}-axis coincides with the line of apsides, the y^{OR}-axis corresponds to $f=\pi/2$, and the z^{OR}-axis completes the right-handed system. Thus, the instantaneous position vector of a satellite is given by

$$\boxed{\vec{r}^{\,\mathrm{OR}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\mathrm{OR}} = r\begin{bmatrix} \cos f \\ \sin f \\ 0 \end{bmatrix} = \begin{bmatrix} a_0(\cos E - e) \\ a_0(1-e^2)^{1/2}\sin E \\ 0 \end{bmatrix},} \tag{15.34}$$

where r, f, a_0, e, E generally vary with time.

The orientation of the OR system with respect to another system has to be defined to enable us to transform one into the other, and for this, three more parameters are needed. The usual way of selecting these parameters is shown in FIG. 15.15. The orbital plane is extended to intersect the celestial sphere, and its trace, the projected orbit, intersects the celestial equator at the *ascending node*, the point where the satellite crosses the equator from south to north. Analogously, the *descending node* is the point of crossing from north to south (cf. §5.2). The angle between the celestial equator and the orbital half plane that contains the part of the orbit stretching from the ascending to the descending nodes is the *inclination i*. The angle between the ascending node branch of the nodal line and the line of apsides, reckoned anticlockwise looking toward the origin, is the *argument of perigee* ϖ. The angle between x^{AP} (true vernal equinox) and the nodal line, measured anticlockwise from the $+z^{\mathrm{OR}}$-axis on the equatorial plane, is the *right ascension of the ascending node* Ω.

These three quantities, along with those defining the orbital ellipse and the motion of the satellite in the orbit, constitute the six *Keplerian orbital elements.* Because they represent a very important parameterisation of the orbit, we list them below [e.g., LEICK, 2004]:

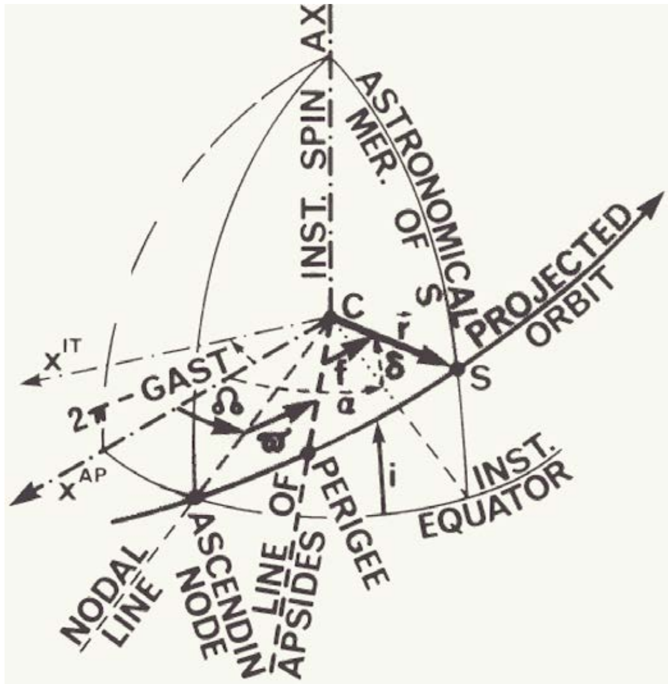

FIG. 15.15. Keplerian orbital elements.

a_0 major semi-axis e eccentricity	size and shape of the orbit.
ϖ argument of perigee Ω right ascension of ascending node i inclination	position of the orbit in the AP system
μ mean (or other) anomaly	position of the satellite in orbit

There are other alternative and equivalent parameterisations of the orbit in use. For example, the geocentric Cartesian representation $(x,\ y,\ z,\ \dot{x},\ \dot{y},\ \dot{z})$, which describes the position of the satellite $(x,\ y,\ z)$ along with its velocity vector $(\dot{x},\ \dot{y},\ \dot{z})$ at a given epoch, will be used in §17.3. Another Cartesian system is shown in §23.2.

Point positioning requires the transformation of positions of the satellite, usually predicted ahead of time and broadcast by the satellite, from the OR system to a terrestrial system (usually the CT), where they are used to compute the coordinates of the observing station. It is helpful to realise that the orbital plane does not rotate with the Earth but remains, in the first approximation, fixed in the AP system, and that the OR and AP systems both have their origin at the Earth's centre of mass (cf., FIG. 15.15). The transformation from the OR to CT system proceeds in three steps: OR → AP, AP → IT, and IT → CT. The first and second steps should be clear from FIG. 15.15. The third step is given by Eq. (15.9). When put together, the complete transformation reads

$$\vec{r}^{\,\mathrm{CT}} = \boldsymbol{R}_2(-x_p)\boldsymbol{R}_1(-y_p)\boldsymbol{R}_3(\mathrm{GAST})\boldsymbol{R}_3(-\Omega)\boldsymbol{R}_1(-i)\boldsymbol{R}_3(-\varpi)\vec{r}^{\,\mathrm{OR}}. \tag{15.35}$$

Note that $\vec{r}^{\,\mathrm{OR}}$ is given by Eq. (15.34). It is thus not difficult to see that even the CT coordinates of the satellite are functions of time τ, and we speak about the time-varying position, or *ephemeris*, of the satellite. The way the ephemeris is generated – through prediction or interpolation – will be shown in Chapter 23.

With the satellite position expressed in the CT coordinate system, we can formulate the various models for positioning:

(a) The *range mathematical model* can be written, for example, as (see FIG. 15.16)

$$\boxed{\bar{e}_i^{\,j}\vec{r}_i = \bar{e}_i^{\,j}\vec{r}^{\,j} - \rho_i^j\ ,} \tag{15.36}$$

where $\rho_i^j = \rho(\vec{r}_i, \vec{r}^{\,j})$ is the measured range from tracking station P_i to satellite position S_j; $(x^j, y^j, z^j) = \vec{r}^{\,j}$ (obtained from Eq. (15.35)) are the known Cartesian coordinates of the satellite at time τ_j; and $(x_i, y_i, z_i) = \vec{r}_i$ are the unknown Cartesian coordinates of the tracking station, all in the CT system. For each range measurement, one such vector equation can be written, and three such (linearly independent) equations give a unique solution for the three unknown coordinates $\vec{r}_i$. Configurations of satellite positions that do not give a unique solution, called *critical configurations*, were studied extensively by BLAHA [197lb]. The interested reader can find more information in LEICK [2004]. Observations of more than three ranges lead to an overdetermined set of equations, which can be treated by the method of least squares, as discussed in Part III.

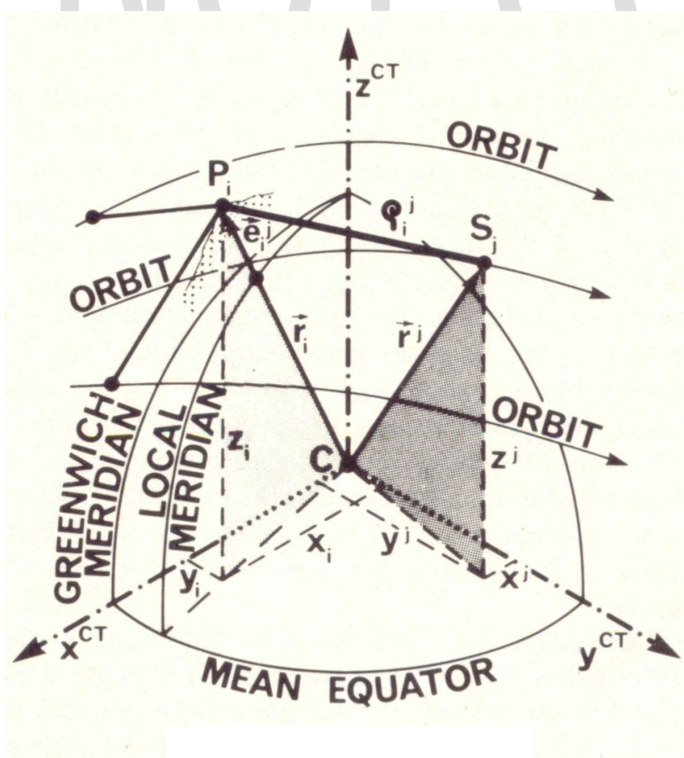

FIG. 15.16. Ranging to satellites.

It is of interest to observe that when the satellite ephemeris is specified in the OR (rather than the CT system), then the coordinates x_p, y_p of the instantaneous pole or the GAST or both of them may also be treated as unknown parameters and may be determined together with $\vec{r}_i$.

Range measurements are obtained by timing the travel time of electromagnetic waves between the tracking station and the satellite (see §9.2). The source can be either earthbound while the satellite carries a retro-reflector, or satellite borne. To date, the most accurately timable source of an electromagnetic signal devised is laser, which emits short pulses of monochromatic light of a duration of a few tenths of nanosecond (ns); this method is known as *satellite laser ranging* (SLR) and it is normally earthbound. Presently the round trip travel time of a laser pulse to a reflecting satellite, such as the LAGEOS (LAser GEOdynamic Satellite) can be determined electronically to an accuracy at the pico-second (ps) level, which corresponds to an accuracy of about 1 to 3 mm in the range ρ [e.g., BENDER, 1992; PEARLMAN ET AL., 2002]. The limiting factor is the inability of the electronic circuitry to more accurately detect either the ends or the centre of the returned pulse.

It should be pointed out that ranging is by no means confined to artificial satellites. When several retro-reflectors were placed on the lunar surface, *lunar laser ranging–LLR* became a practical and viable alternative. Presently, there are over 32 satellite missions that carry on-board retro-reflectors for precise positioning and geodynamics [see e.g., IAG's *International Laser Ranging Service* on line at http://ilrs.gsfc.nasa.gov/satellite_missions/ with many more planned missions to be launched in the next few to several years. It is noted that the SLR (and LLR) is not an all-weather system (requires clear sky) and is rather restricted to specially equipped observatories because of the size and cost of facilities required. Thus, it is of limited use to ordinary users. More details on the SLR observing systems, techniques and mathematical models can be found for instance in SEEBER [2003]. The IAG's *International Laser Ranging Service–ILRS* provides global satellite and lunar laser ranging data and their related products to support geodetic and geophysical research activities and develops the necessary global standards and specifications that contribute to the definition and maintenance of the International Terrestrial Reference Frame – ITRF (to be discussed in §15.5).

A second mode of ranging, used extensively by a variety of users nowadays, times radio signals. The radio signal used for the *satellite radio ranging* (SRR) must be of a frequency higher than 30 MHz to penetrate through the ionosphere (cf., §9.2). It also should be as coherent as possible thus, requiring the use of highly stable oscillators. Cesium (Ce) and rubidium (Rb) atomic oscillators are used for this purpose in the satellites of the *NAVSTAR Global Positioning System* (GPS). Because the emitted radio-wave is continuous, the timing of the travel time must be arranged differently from the timing of discrete laser pulses. The timing is done by the ground *satellite receiver* whose clock (oscillator) must be not only highly accurate and stable but also synchronised with the satellite oscillators so that it can associate the reception of timing marks with the time elapsed between their emission and reception. This synchronisation usually cannot be done accurately enough, and the resulting *clock offset* Δt_i, in reality the (approximately) constant difference between the satellite time and the ground time, has to be included as an additional unknown in the model (Eq. (15.36)), which then acquires a term $-c\Delta t_i$, on the left-hand side, where c is the speed of light. Under these conditions, clearly more than three ranges have to be observed to yield a solution $(\vec{r}_i, \Delta t_i)$ of the model [e.g., KAPLAN AND HEGARTY, 2006].

The GPS has become the most important geodetic positioning system accessible to all users and for many applications. It provides nearly continuous positioning and timing information globally under all weather conditions. The current GPS constellation comprises 32 active satellites and many spares distributed in six orbital planes. The orbits are nearly circular at an altitude of about 20 200km with inclination i=55° and orbital period of nearly 12h. The GPS was officially fully operational on July 17, 1995 and due to its configuration more than four and up to 10 satellites are visible to the user at (almost) any time thus, providing redundant observations to achieve accurate positioning [see for instance LEICK, 2004]. The GPS satellites emit highly coherent circularly-polarised carrier signals on frequencies of L_l=1575.42 MHz (λ=19cm) and L_2=1227.60 MHz (λ=24.4cm). The L_1 carrier frequency is modulated by two pseudo-random (PRN) sequences of zeros and ones, called P-code (precision code) and C/A-code (coarse/acquisition code). The L_2 carrier frequency is modulated only by the P-code, with the exception of new satellites (they also have A/C code). The last modulation of both carrier signals of each of the satellites is achieved by a low frequency (50 Hz) stream of data describing the ephemeris of that particular satellite, known as the *broadcast ephemeris*. The user's receiver is designed to decipher this message, which also contains information about the clock and the general state of health of the transmitting satellite and convert it to the appropriate satellite position $\vec{r}^{\,j}$ at the time τ_j of the range measurement. Alternatively, a more *precise ephemeris* may be used instead if higher precision is required.

The P-code repeats every 266 days (~2.35×10^{14} bits) and is split in 38 segments (7 days each), each segment being uploaded to a different satellite which is thus characterised uniquely (PRN number). The access to P-code can be denied to the users if the GPS Mission Operations Segment decides to activate the anti-spoofing (AS) mode or W-code; when W-code is activated, the satellites transmit the Y-code (or P(Y)-code which is P-code encrypted by W-code). The C/A-code (1023 bits of 1ms duration each) repeats itself approximately once per second and it has no access restriction for the users. However, it provides only low accuracy positioning capability. In the future, under the GPS modernisation program, the C/A-code will be added to the L_2 carrier and a military code (M-code) to both L_1 and L_2 carriers [SHAW ET AL., 2000]. In order to satisfy the aviation requirements, a third carrier frequency (L_5=1176.45MHz) will be added to the next generation of GPS satellites; L5 will have more power and wider bandwidth. At the time of writing, two newer generation GPS satellites transmit L_5 carrier wave and it is anticipated that by the end of the decade all GPS satellites will be transmitting this third carrier wave. The reader can find more details on the GPS space segment and signal structure for instance in LANGLEY, [1990], KAPLAN, [1996], SHAW ET AL., [2000], SEEBER [2003] and LEICK, [2004].

The user's receiver is capable of generating at least the C/A-code, and if known, the P-code, and, by matching this generated sequence with the incoming sequence, it can measure the difference between the emission time τ and the reception time t. This difference, corrected for the above mentioned clock offset and multiplied by the speed of electromagnetic propagation, gives the desired range. This ranging method that is based on one-way propagation of the signal (satellite to receiver) and two clocks is different from the traditional two-way ranging using electromagnetic distance measuring (EDM) equipment; it is called *pseudoranging*. Ranges can also be determined by performing phase measurements on the carrier waves of λ=19cm and λ=24.4cm. In this approach, the carrier wave arriving from the satellite is Doppler shifted by the relative motion of the satellite and receiver. The receiver itself generates a replica of the carrier wave, which is then added to the incoming Doppler shifted carrier wave from the satellite, producing a beat signal. The range then is determined by the number of full beat cycles plus a fraction of a cycle using the beat wavelength. The problem with this approach is that the receiver cannot measure the number of full cycles (only a fraction of one wavelength) but once this *cycle* *ambiguity* is resolved, the range to the satellite can be determined much more accurately than with the pseudorange approach. This is because phase measurements performed to determine the fraction of the beat cycle are far more accurate than timing. There exist different methods for the *ambiguity resolution* and the reader is referred to the GPS literature, for instance, WELLS ET AL., [1987], KAPLAN, [1996], SEEBER [2003] and LEICK, [2004]. Let us just mention in passing that ranges are not the only observables obtained from the GPS. Other modes can be used as will be discussed in §16.1.

Before the observed ranges ρ are introduced into the mathematical model (Eq. 15.36), they must be corrected for the effect of atmospheric refraction. The general formula for the *range refraction correction,* also called the *(propagation) delay correction,* takes on the form (cf., Eq. (9.9)):

$$O\rho = \int_{e} (n-1) \mathrm{d}S \,, \tag{15.37}$$

where n is the index of refraction along the path $\mathcal{C}$ between P_i and S_j (Eq. (9.5)). Since the indices of refraction for the troposphere and the ionosphere are completely different, the integration is normally broken into two separate parts. It is then usual to speak of *tropospheric delay correction* $O\rho^{\mathrm{T}}$ within about the lower 50-60km of the atmosphere and *ionospheric delay correction* $O\rho^{\mathrm{I}}$ above the troposphere to about 1000 km. The above integral requires that n once more be known all along the path. Several researchers have contributed toward developing a formula for the tropospheric delay correction (see, e.g., HOPFIELD [1969] and YIONOULIS [1970]), and a good review of their work can be found in WELLS [1974] and WELLS ET AL., [1986]. Generally, the tropospheric correction can be written as

$$O\rho^{\mathrm{T}} = \sum_{i=d,w} k_i / \sin[(\nu^2 + \theta_i^2)^{1/2}] \,, \tag{15.38}$$

where ν is the vertical angle to the satellite. Subscript d stands for the dry (also known as 'hydrostatic') component of the air; and K_d is a function of temperature T, pressure P, and the height H of the tracking station. Subscript w stands for the 'wet' component, and K_{w} is again a function of T, P, H, and, in addition, the vapour pressure e. The corrections to the vertical angle ν are θ_{d}=2.5°and θ_{w}=1.5°. Average values may be adopted for K_d and K_w; for example, MOFFETT [1971] gives the following average values for a maritime climate: K_d =2.31m and K_w =0.20m. For a formula different from the above, see SAASTAMOINEN [1973]; any of these formulae are generally valid only for $\nu > 10°$ and are accurate to about 0.2m.

The ionospheric delay depends upon the free electron content along the signal path and the frequency of the wave. The latter indicates that the ionosphere is a dispersive medium which means that different ionospheric delay corrections may be obtained by using waves of different frequencies. By expressing n in form of a series:

$$n = 1 + \frac{c_1}{f^2} + \frac{c_2}{f^4} + \cdots, \tag{15.39}$$

where f is the frequency of the signal, and c are functions of the ray's trajectory and time, independent of f. Equation (15.37) then yields

$$O\rho^{\mathrm{I}} = \int (n-1)dS = \frac{b_1}{f^2} + \frac{b_2}{f^4} + \cdots, \tag{15.40}$$

where the b's are again independent of f. Note that this correction, being a function of frequency, can be evaluated if ranges are measured simultaneously at two different frequencies, as is the case with the NAVSTAR/GPS system. We shall return to this correction later.

The error in position determination using range data can be broken down into two main components: the error in range and the error in ephemeris, also called the orbital error or error in the satellite position. As an example, for the GPS, the error in the broadcast ephemeris is about 2-5m [e.g., LANGLEY ET AL., 2000; LEICK, 2004] and the errors in ranges depend on which 'code' is used for ranging. P-code ranges have standard deviations at the level of several decimetres, while C/A-code ranges are accurate to several metres. Ranges from carrier beat phase measurements are accurate to a few millimeters. A position determined by P-code when ranging for a few hours should be accurate to about 0.5m (one standard deviation, internal consistency) [WELLS ET AL., 1986]. On the other hand, the standard deviation of the positions determined by satellite laser ranging observations spread over many months, is reported to be about 2–3mm [BENDER, 1992].

(b) Logically, the next model that should be treated here is the *direction mathematical model.* To obtain directions to a satellite, the satellite is photographed against a background of known stars. Then star and satellite positions are identified on the photographic plate, and measurements are made of the location of the satellite image relative to the images of known stars. With a and δ of the stars known, the a and δ of the satellites can be estimated. This seemingly simple procedure has several problems that limit its accuracy and make it, in fact, quite complicated. The most troublesome is the uncertainty in modelling the camera lens distortion which can only be done by introducing many redundant observations. The accuracy of the catalogued star positions presents additional limitations. The effect of atmospheric refraction (Eq. (15.23)) is also a problem [MUELLER, 1964]. As in the case of the range model, the orbital errors contribute further uncertainties.

It is not possible to determine the three-dimensional position of a single tracking station from direction observations alone. At least one distance either on the ground or to a satellite has to be measured to supply the missing scale. Directions will thus be considered only as part of the next model that uses directions and ranges measured simultaneously. This approach is conceptually much simpler. Further uses of satellite directions will be dealt with in §16.1 and §17.3. However, direction measurements to satellites using photographic cameras (Baker-Nunn, BC4) have been long discontinued.

(c) The *simultaneous direction and range mathematical model* can be written as (see FIG. 15.17)

$$\boxed{\vec{r}_i = \vec{r}^{\,j} - \vec{\rho}_i^{\,j}\ ,} \tag{15.41}$$

where $\vec{r}_i$ and $\vec{r}^{\,j}$ are again radial vectors of the observing point and satellite respectively. The *topocentric vector* $\vec{\rho}_i^{\,j}$ in the CT system is given by

$$\vec{\rho}_i^{\,j\,\mathrm{CT}} = \boldsymbol{R}_2(-x_p)\boldsymbol{R}_1(-y_p)\boldsymbol{R}_3(\mathrm{GAST})\rho_i^j \begin{bmatrix} \cos\delta_j \cos\alpha_j \\ \cos\delta_j \sin\alpha_j \\ \sin\delta_j \end{bmatrix}. \tag{15.42}$$

Clearly, even with only one observation $\vec{\rho}_i^{\,j}$ it is possible to determine uniquely the position $\vec{r}_i$ of the tracking station. When more than one satellite position is observed, then Eq. (15.41) becomes one of several observation equations and averaging has to be used to give the three components of $\vec{r}_i$. In this instance, we note that the model is of the explicit-in-parameters $(x_i,\, y_i,\, z_i)$ variety the handling of which was discussed in Chapter 11. Since this method of point positioning is not widely used, there is no good estimate available of the achievable point position accuracy.

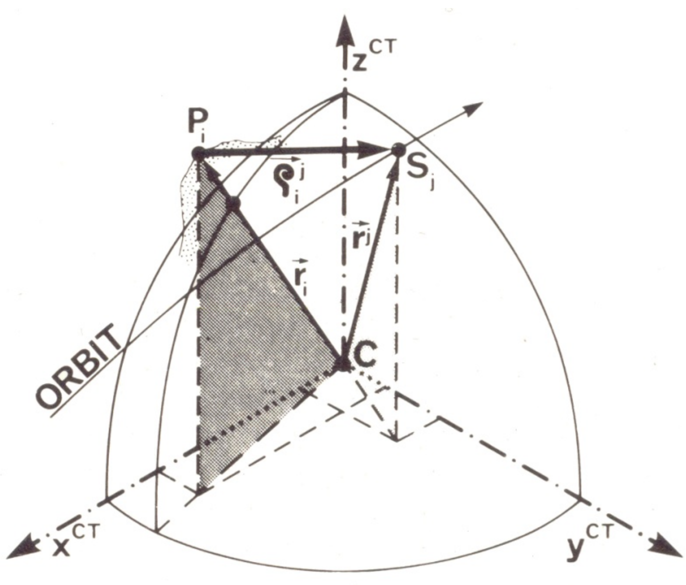

FIG. 15.17. Simultaneous ranging and directions to a satellite.

(d) The next model of point positioning using satellites is the *range difference mathematical model*. When a moving object, such as an orbiting satellite, transmits an electromagnetic signal of a certain (constant) frequency $f_{\mathrm{T}} = 2\pi / \lambda_{\mathrm{T}}$ the observer receives a signal of frequency $f_{\mathrm{R}} = 2\pi / \lambda_{\mathrm{R}}$, which varies with the velocity of the transmitter with respect to the receiver (see FIG. 15.18). Mathematically, the received wavelength λ_{R} is related to the constant transmitted wavelength λ_{T} through the *Doppler equation* [MENZEL, 1955]

$$\lambda_{\mathrm{R}} = \lambda_{\mathrm{T}}\left(1+\frac{\dot{\rho}}{c}\right)\left(1+\frac{\dot{\rho}^2}{c^2}\right)^{1/2}, \tag{15.43}$$

where c is the speed of light and $\dot{\rho}$ is the rate of change of the range with time (see FIG. 15.19). Equation (15.43) is said to describe the *Doppler effect*.

From Eq. (15.43) we see that $\lambda_{\mathrm{R}} = \lambda_{\mathrm{T}}$ when and only when $\dot{\rho}=0$. This situation occurs only when the satellite moves with a velocity $\dot{\vec{r}}$ normal to the range vector $\vec{\rho}$ (cf. FIG. 15.19). The point of the satellite orbit, where the satellite velocity is normal to the satellite range vector, is the *point of closest approach* (PCA). Only the frequency (f_{T}) transmitted at the PCA is later received by the receiver undistorted.

The *TRANSIT* satellite system was designed to exploit the Doppler effect in a novel way and although it has been superseded by the GPS and is not operational anymore, it is instructive to present here at least its concepts. The system employed five satellites with nearly circular (e≈0) polar ($i \approx \pi/2$) orbits of a mean altitude of about 1074 km (i.e., a_0≈6371km + 1074km ≈7445km). This altitude implied an orbital period of about 107 minutes (see Eq. (5.1)) and the translational velocity $\dot{r}$ of about 7.3 kms^{-1}. With $\dot{\rho}$ being always smaller than $\dot{r}$, the relativistic effect $(\dot{\rho}/c)^2$ in Eq. (15.54) accounts for less than 3×10^{-10} of the wavelength λ and can be neglected.

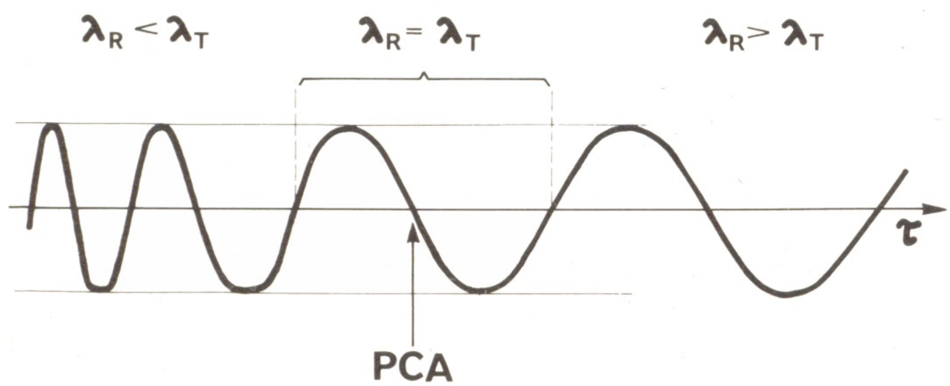

FIG. 15.18. Variation of received signal wavelength

Now, using λ=c / f and substituting frequencies for wavelengths in Eq. (15.43), we obtain

$$f_{\mathrm{R}} \approx \frac{2\pi}{\lambda_{\mathrm{T}}\left(1+\dot{\rho}/c\right)} \approx f_{\mathrm{T}}\left(1-\dot{\rho}/c\right). \tag{15.44}$$

Then it is easy to express the range rate $\dot{\rho}$ as a function of the two frequencies:

$$\dot{\rho} \approx c\left(1-f_{\mathrm{R}}/f_{\mathrm{T}}\right). \tag{15.45}$$

Finally, the *range difference* $\nabla\rho$ between two satellite positions $S(\tau_j)=S^j$ and $S(\tau_k)=S^k$ is given simply as

$$\nabla\rho \approx \frac{c}{f_{\mathrm{T}}}\int_{\tau_j}^{\tau_k}(f_{\mathrm{T}}-f_{\mathrm{R}})\,\mathrm{d}\tau\,. \tag{15.46}$$

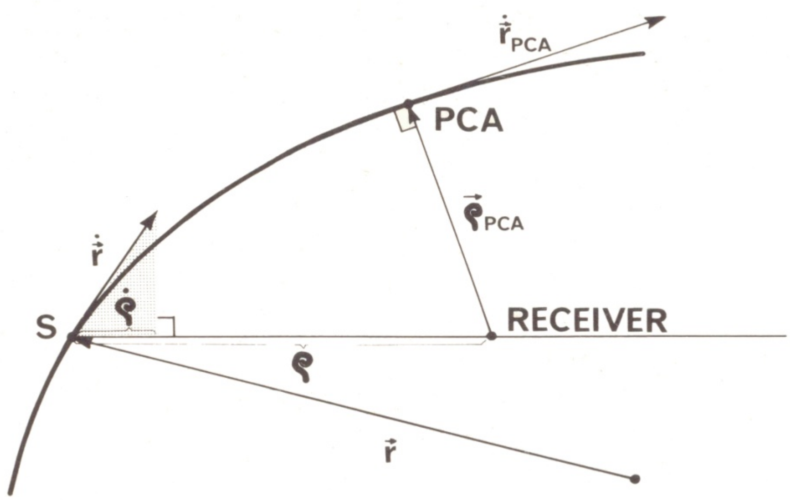

FIG. 15.19. The rate of change of range ρ.

Assuming that the beat frequency $f_{\mathrm{T}}-f_{\mathrm{R}}$ can be observed and integrated, how can the range differences $\nabla\rho$ be used to obtain the position of the receiver? The observed $\nabla\rho$ can be used in a scheme known as *hyperbolic positioning,* the concept of which is as follows. Suppose we know two positions S^1, S^2 of the satellite and the $\nabla\rho_i^{12}$ observed at the point P_i. From elementary geometry we know that P_i must lie on one of the hyperbolic surfaces H_1 , H_2 shown (in two dimensions) in FIG. 15.20, because each hyperbolic surface is a locus of a constant difference in ranges reckoned from the two foci S^1, and S^2. If we then get another range difference, e.g., $\nabla\rho_i^{23}$, relating P_i to another pair of hyperbolic surfaces, then P_i must lie on one of the curves made by intersecting the appropriate hyperbolic surfaces. A third range difference should then provide us with only one point at the intersection of two such curves.

The TRANSIT system was designed so that a whole string of satellite positions on one *orbital arc,* called a *satellite pass,* could be used. One satellite pass was usually taken as the visible part of the orbit that spanned two successive passages through the almucantar of Z = 82°. The satellite positions on each pass were spaced equidistantly in time and the used constant time interval $\Delta\tau$ depended upon the design of the receiver, the optimum having been about 30s.

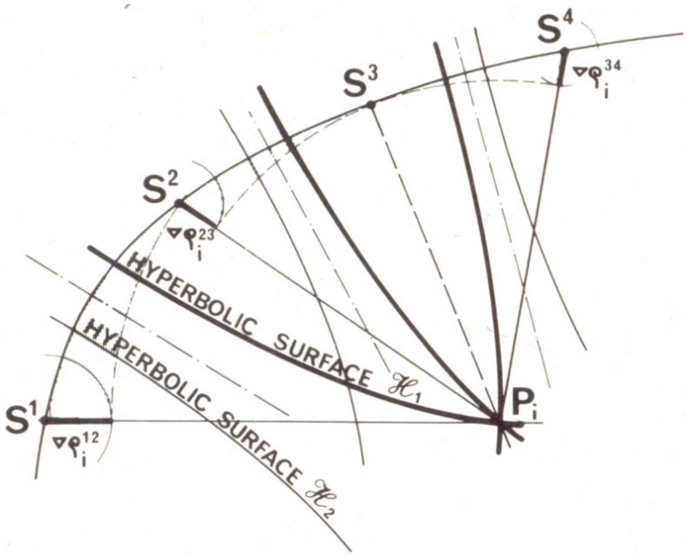

FIG. 15.20. Hyperbolic positioning

The *range difference mathematical model* needed for the determination of the receiver antenna's position is obtained simply by formulating two range mathematical models (Eq. (15.36)) for P_i and two different satellite positions S^j, S^k. Subtracting the first from the second, we get

$$\left(\vec{e}_i^{\,k} - \vec{e}_i^{\,j}\right)\vec{r}_i = \left(e_i^k \vec{r}^{\,k} - \vec{e}_i^{\,j}\vec{r}^{\,j}\right) - \nabla\rho_i^{jk}. \tag{15.47}$$

Substituting for the range difference, we obtain finally

$$\boxed{\left(\vec{e}_i^{\,k} - \vec{e}_i^{\,j}\right)\vec{r}_i - \frac{c}{f_{\mathrm{T}}}\nabla\tau\ \nabla f_i^l = \left(\vec{e}_i^{\,k}\vec{r}^{\,k} - \vec{e}_i^{\,j}\vec{r}^{\,j} - \frac{c}{f}D_i^{jk}\right).} \tag{15.48}$$

where Δf_i^l refers to the l^{th} pass observed at P_i and D_i^{jk} is the integrated Doppler count. The *Doppler count* is the number of cycles of the Doppler signal, i.e., the Doppler frequency tracked in time. and the integral of the Doppler count accumulated between the two-time marks τ_j and τ_k transmitted by the satellite is the *integrated Doppler count* D_i^{jk}**.**

In order to evaluate the position $\vec{r}_i$ and the m-tuple of frequency offsets for m passes, we have not only to observe the integrated Doppler counts D_i^{jk} but also have to know the positions $\vec{r}^{\,j}$, $\vec{r}^{\,k}$ of the satellites at the instants of transmission of the integration time marks. These positions are evaluated either from the predicted orbital information broadcast by the satellite or from the post-fitted orbital information. The broadcast ephemeris is given in terms of coordinates very close to Keplerian orbital elements and their rates of change. Satellite positions computed from the broadcast information were good to about 25 m along the track, 15 m radially, and 5 m across the track. The other alternative, the precise ephemeris, was about twice as accurate as the broadcast ephemeris.

It should be mentioned here that the same observable, i.e., the integrated Doppler count, can be obtained from GPS. With the GPS satellites moving at a considerably lower angular velocity, however, there is no geometrical advantage in using the GPS in this mode [VANÍČEK ET AL., 1984].

15.5. Modern reference coordinate systems

In the previous sections, the key attribute in the definition of the celestial and terrestrial coordinate systems as well as of the transformations among them was the time-varying position and orientation of the Earth. The position of the centre of mass, the orientation of the Earth's spin axis, with respect to both, the crust (affected by polar wobble) and the stars (affected by precession and nutation), and the orbital plane (ecliptic) of the Earth around the Sun define fundamental points and directions that were used to orient the coordinate axes in space. It is said that these coordinate systems were *dynamically* defined by taking into consideration the complex motion of the Earth and planets around the Sun and the motion of the solar system within the galaxy (cf., Chapter 5) in a mechanical approach (through celestial mechanics). The astronomical catalogues used for astronomical positioning were based on a series of *Catalogues of Fundamental Stars FKn,* where *n* is a sequential number, for short. These catalogues contained positions of hundreds of thousands of stars in the AP-system of coordinates. The latest such catalogue that was used in conjunction with the classical coordinate systems was the FK5.

Astronomical ephemerides and Catalogues of Fundamental Stars constituted the *reference frame* that is, the set of identifiable reference points, on the celestial sphere along with their estimated coordinates and their precision. Alternatively, a reference frame can be viewed as a coordinate system embedded in the real geometrical space that can only be defined indirectly by the coordinates of selected reference points known as *fiducial points*. These fiducial points that essentially imply the definition of a metric within the coordinate system can directly be observed with respect to the LA coordinate system using various observing systems, such as theodolites, astronomical telescopes, radio telescopes and others. A reference frame is said to be the practical realisation of a more complex entity known as '*reference system*' to be discussed in Chapter 16. A *coordinate system*, as used in the classical approach (cf., previous sections), is merely a vector space with an orthogonal basis and an origin and as such should be viewed as one of the fundamental components of a reference system. The above distinction between reference systems, reference frames and coordinate systems has been used in practice since the early 1990's and we will be following this approach here.

In recent years, significant improvements in the observing systems and mathematical models, driven by the need for higher positioning accuracy:

1) for engineering applications,

2) for the study of the Earth system[1] dynamics and

3) the universe at large (cosmology),

have necessitated the revision of the classical reference coordinate systems to provide, highly accurate, multipurpose and global referencing for all applications. The process of improving the classical coordinate systems to respond to more complex and more demanding point positioning needs was initiated by the IAU during its 1985 General Assembly in New Delhi, India. In 1987, the IAU and the IUGG established the IERS, (cf., §4.2), to serve the astronomical, geodetic and geophysical communities by defining a new generation of reference systems. These reference systems can be viewed as consisting of theories, mathematical models, standards, methodologies, reference (fiducial) points and data for precise point positioning with respect to a new suite of coordinate systems. Since then, a series of important IAU resolutions, particularly in the period

[1] The Earth system includes the solid Earth, hydrosphere, cryosphere and atmosphere and all physical processes involved.

1997-2000, provided the foundation for the creation of the new reference systems. What follows is the rationale for the need of redefinition of the classical coordinate systems as components of two major modern reference systems namely, the *International Celestial Reference System* (ICRS) and the *International Terrestrial Reference System* (ITRS), to be duly defined later.

Starting in the mid-1980s, high precision VLBI astrometry[2] (wide-angle[3] or all-sky astrometry) has been conducted not only in the optical but notably in radio wavelengths. Radio-astronomy revolutionised astrometry in three fundamental ways. It reaches far beyond the optical telescopes to enable us to "observe" extragalactic radio sources (quasars) that are so far that they show very small, practically immeasurable proper motion. This is the feature that allows us to realise a nearly inertial frame of reference as already mentioned

a) it achieves significantly higher precision (as much as one order of magnitude higher) in the determination of directions (α, δ, as derived from the LA coordinates) of the extragalactic objects distributed widely across the sky. Radio wavelengths are much less influenced by astronomical refraction (bending) compared to visible wavelengths. But, let us be fair and state that optical astrometry has also achieved milliarcsecond accuracy in wide-angle direction measurements from European Space Agency's satellite mission Hipparcos [e.g., KOVALEVSKY ET AL., 1997; KAPLAN 2005], because this orbiting telescope operates outside the Earth's atmosphere which fact reduces significantly the astronomical refraction;

c) it attains high precision in the determination of the Earth's orientation parameters (length-of-day (LOD) and polar motion, cf., §5.4) from which UT1 (cf., §15.2) is derived.

In the realm of mathematical modelling, three key improvements that contributed to higher precision of the definition of the coordinate systems have been introduced:

a) Significant improvement of the accuracy of precession-nutation models. Since the mid-1970s, a series of new precession-nutation models have been developed to replace Newcomb's [1906] old tables. These include the International Astronomical Union – IAU 1976 precession model [LIESKE ET AL., 1976], and the IAU 1980 precession-nutation model [WAHR, 1981; SEIDELMANN, 1982]. The latest IAU2000A (high precision) and IAU2000B (reduced precision) precession models based on the theory developed by MATHEWS ET AL., [2002] have been adopted since 1st January 2003 by IAU and the IERS [PETIT AND LUZUM, 2010];

b) improvements in the prediction, using celestial mechanics, of instantaneous position and velocity vectors (state vectors) of the moon, the planets of our solar system, and asteroids referenced to the solar system *barycenter*. The standard model for these predictions used today that forms the basis of the astronomical ephemerides and of the definition of the

[2] Astrometry deals with the *positions* and *motions* of celestial objects as Astronomy did until about the 19th century. Since then, Astronomy grew wider in scope to include more than just determining positions and motions of stars and as such Astrometry has now become a section of astronomy.

[3] Narrow-field or small-angle, Astrometry uses large optical telescopes whose observational accuracy is severely limited by Earth's atmosphere, thermal and structural telescope instrumental errors, and the inability of large terrestrial telescopes to offer all-sky visibility. VLBI (with its multitude of antennas and using different radio frequencies), Hipparcos satellite (optical telescope orbiting outside the Earth's atmosphere and therefore had no blind spots) and other modern space observing systems can measure parallax and proper motions of quasars and stars from large angle, or from space therefore with increased visibility and resolution. These are called wide-angle or all-sky astrometry systems.

modern celestial reference systems as well as the support of space missions is the Jet Propulsion Laboratory's *Planetary and Lunar Ephemeris DE-405* [Standish, 1998];

c) the consideration of space-time coordinates and their 4D space-time transformation, for the Earth and the solar system within the framework of general relativity [Kaplan, 2005].

The definition of the coordinate systems associated with the ICRS and the ITRS follows the '*kinematical*' approach (cf. §15.1) that fixes the triad of axes to specified points in space at a specific time epoch τ_0, meaning that the axes are "kinematically fixed". Similar to the realm of the classical dynamical coordinate systems seen in §15.1, and §15.2, we distinguish between two classes of kinematical coordinate systems, namely the *celestial* and *terrestrial* coordinate systems. The generic kinematical definition of both classes uses the positions of well-known objects, such as extragalactic radio sources (e.g., quasars), for the celestial coordinate systems, or stable terrestrial stations, for the terrestrial coordinate systems, to fix the coordinate axes of the coordinate systems at epoch τ_0, rather than using the continuously changing spin axis of the Earth and the vernal point as reference directions.

It is important to emphasise that in order for the modern coordinate systems to be as close as possible to the classical ones and make the transition as seamless as possible, the positions of the quasars and other extragalactic sources that are used to define the directions of the axes of the modern coordinate systems had been those determined originally with respect to the classical coordinate systems (e.g., MRA(τ_0) and CT(τ_0)) at an accepted reference epoch τ_0 (to be discussed later in this section). The kinematical definition of the coordinate systems is a natural evolution of the dynamical systems whose conceptual treatment (§15.1) and comprehension are absolutely essential for understanding the modern systems. In the sequel, we provide the definition of the new coordinate systems and their transformations (see also Feissel and Mignard, 1998; Seidelmann and Kovalevsky, 2002]), to parallel those presented in §15.1. An interested reader can find a comprehensive review of the new concepts and definitions in [Kaplan, 2005] and in the IERS Conventions 2010 [Petit and Luzum, 2010].

As with the classical coordinate systems (cf., §15.1), let us begin with the definition of the *barycentric celestial*–BC(τ_0) coordinate system that is associated with, or is a fundamental component of the *Barycentric Celestial Reference System* (BCRS), where the term "barycenter" refers to the solar system's barycenter. BCRS belongs to the broader and more abstract class of the barycentric *International Celestial Reference Systems* – ICRS [see definition in Petit and Luzum, 2010]. Here, and in order to be consistent with the definitions of the classical coordinate systems, we will define the coordinate systems only as main components of the reference systems, along with their corresponding transformations and therefore, we will drop the 'RS' designation when defining them. Accordingly, the coordinate system associated with the BCRS will be designated as BC and the coordinate system associated with the ICRS will be designated as IC and so forth.

The BC(τ_0) coordinate system is the system in which the directions of the stars and distant extragalactic radio sources, such as quasars, are determined and published to provide a stable inertial reference frame at time epoch τ_0 [Kaplan, 2005]. Such a coordinate system that may be paralleled with the classical MRA(τ_0), is inertial in character, up to the negligible proper motion of the radio sources, at least with respect to the precision of the observations. The coordinates (directions) of these 'defining' or 'fiducial' extragalactic objects, namely α and δ have been obtained from many astrometric observations using radio-astronomy [Ma and Feissel 1997; Ma et al., 1998] and they are known to better than one-half of a milliarcsecond [Petit and Luzum, 2010; p. 23]. For the definition of this coordinate system we do not need to use the direction of the axis of rotation of the Earth, nor do we need to know the direction to the NCP or the vernal point ♈ to orient the axes. The coordinate axes are kinematically fixed with respect to the quasars in such a way so as to provide the coordinates α, and δ of all defining objects (currently 212 in total) in the least squares solution of over-determined system of observation equations, see §14.3. More specifically, this coordinate system has the following characteristics:

(a) It is an equatorial system with origin at the solar system barycenter;

(b) the xy^{BC}-plane (principal plane) is very close to the mean equator of J2000.0 thus, it provides the reference plane for the declination ($\delta=0$) of all defining quasars in the mean sense at reference epoch τ_0=J2000.0

(c) the x^{BC}-*axis* (reference direction for the right ascension) is fixed on the mean equator of J2000.0 and in such a direction so as to provide the reference for right ascensions ($\alpha=0$) of all defining quasars in the mean sense; therefore it is pointing very close to the dynamical equinox of J2000.0. The direction of the x^{BC}-*axis* is known as the *Celestial Intermediate Origin*[4] – CIO (not to be confused with the Conventional International Origin that defines the *z-axis* of the classical CT system and ensures that the x^{BC}-*axis* is a non-rotating origin (NRO[5]) for the right ascensions [Guinot, 1979] (IAU resolution B1.8, 1997). The characterisation 'intermediate' will become obvious later when discussing the coordinate system transformations;

(d) the reference pole is defined by the z^{BC}-*axis* that is perpendicular to the principal plane (see (b) above) and thus, fixed in space. This pole is known as the *Conventional Reference Pole* (CRP) and it is consistent with the fifth fundamental star catalogue FK5 (within the uncertainty of the FK5 – IAU resolution B1.8, 1997);

(e) the time scale is the barycentric coordinate time (TCB), and the coordinates of the stars and extragalactic objects are the least squares estimates of $\alpha(\tau_0)$ and $\delta(\tau_0)$ as explained above;

The BC(τ_0) coordinate system is similar (up to the time epoch τ_0) to the MRA(τ_0) as defined in §15.1. As such, Eqs. (15.1), (15.2) and (15.3), except the superscript 'TRA', can be used to establish the relations between Cartesian coordinates and right ascension and declination in the BC-system.

The next most important coordinate system is the *international terrestrial*–IT(τ_0) coordinate system that is associated with the ITRS to which all modern geodetic positions are referred (not to be confused with the instantaneous terrestrial–IT defined in §15.4). This coordinate system is equivalent to the classical conventional terrestrial (CT) coordinate system. Following, once again, the kinematical approach, the coordinate axes are implied by the coordinates of selected stable fiducial terrestrial stations. These fiducial stations cover the Earth's surface uniformly, and their

[4] The term "origin" is confusing from the mathematical perspective. In fact, the x^{BC}-*axis* is the reference direction (α=0) for the right ascensions.

[5] Similar to the previous footnote, a better name for NRO would be Fixed Reference Direction. However, we follow here the IAU/IERS nomenclature for consistency

coordinates have been derived from many, more or less independent high precision geodetic observing systems[6] discussed by PETIT AND LUZUM [2010; Ch. 4]. It is conceivable that, due to improvements in the observational techniques, crustal deformation and refinements in the positions of the extragalactic objects, the positions of the defining terrestrial stations will change with time albeit very slowly, most likely at the rate of tectonic plate motions (a few centimeters per year, i.e., a few arc-msec/a). Therefore, the orientation of the axes of the IT(τ_0) will have to be upgraded from time to time as required. The orientation of the axes of the IT(τ_0) was initially given by the *Bureau International de l'Heure* (BIH) to be consistent with the *BIH Terrestrial System* of τ_0=J1984.0 (BTS84) within about 0.005″ (~ 15cm on the Earth's surface). Clearly, the IT(τ_0) coordinate system is fixed to the Earth and thus, it resembles the classical CT system already seen in §15.1. The characteristics of IT(τ_0) can be summarised as follows [MCCARTHY, 1996; PETIT AND LUZUM, 2010; Section 4.1.1]:

(a) It is a geocentric equatorial system with the centre of mass C being defined for the whole Earth, including oceans and atmosphere;

(b) The xy^{IT}*-plane* was initially coincident with the mean equatorial plane of BTS84, within about 0.005";

(c) The z^{IT}*-axis* is perpendicular to the equatorial plane and its intersection with the Earth's surface defines the *International* (or IERS) *Reference Pole* (IRP), something analogous to CIO of the classical CT coordinate system; The xz^{IT}*–plane* (prime meridian) does not coincide with any Greenwich meridian (mean, true, or any other) and defines the *Terrestrial Intermediate Origin* (TIO) for longitudes; this is in accordance with the IAU resolution B1.8 of 1997 that recommends the use of the '*non-rotating origin*' (NRO) for both, ICRS and ITRS;

(d) The unit of length is the metre (SI) and the time scale is the TDT[7] (or TT) in the year 2000.

The time evolution of the orientation is ensured by using a no-net rotation condition with respect to horizontal tectonic motions over the whole Earth. Observations needed for geodetic point positioning (terrestrial) can be either astronomical or, most commonly, satellite. Astronomical positioning has already been discussed in §15.1 and the relevant mathematical models were presented in §15.3. Satellite positioning was discussed in §15.4. As we have seen, satellite positioning is possible through observations to satellites whose positions along the orbit are known, for instance in the OR system. To be suitable for positioning, the satellite orbital coordinates should be transformed into IT to furnish the desired point positions. Similar to the classical coordinate systems, we need to link the two systems, OR and IT through a series of transformations (cf., Fig. 15.21) starting, once again with LA. The interested reader can find more details on these transformations in the IERS Conventions, 2010 [PETIT AND LUZUM, 2010; Section 5.4).

Here is the series of transformations needed (cf., Fig. 15.21):

Step (a), LA(τ)→TI(τ)**:** The first intermediate coordinate system needed in the series of transformations of the (instantaneous) local astronomical–LA(τ) system into BC(τ_0), in which the

[6] Very Long Baseline Interferometry (VLBI), Satellite Laser Ranging (SLR), Lunar Laser Ranging (LLR), Global Positioning System (GPS), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), Precise Range And Range-Rate Equipment (PRARE), Doppler/TRANSIT observations, optical astrometry, and tide gauges and meteorological sensors

[7] Older definitions of IT used the TCG time scale (see § 15.2)

positions of stars are published, is the *terrestrial intermediate*–TI(τ) coordinate system. TI is very similar to the classical IT system already treated in §15.1. The origin of TI is at C, the z^{TI}-*axis* is aligned with the Earth's spin axis that is affected by all motions up to quasi-diurnal terms with amplitudes under 0.01″ (ocean tides, short-term nutations) and thus, it is *quasi-instantaneous.* The z^{TI}-*axis* defines the CI pole on the celestial sphere known as the *Celestial Intermediate Pole* (CIP). The xz^{TI}-*plane* is fixed with respect to the Earth and defines the non-rotating *Terrestrial Intermediate Origin* (TIO), with "non-rotating" meaning with respect to the Earth. TIO is very close to the Greenwich meridian; it was originally defined to be as close as possible to BTS84[8] meridian. The transformation LA→TI is identical to the classical LA→IT (cf., Step (a) in the classical transformations) that involves a "translation" of the topo-centre to geocentre C via diurnal aberration and diurnal parallax and subsequently, a reflection and two rotations given by Eq. (15.6), by replacing superscript IT with TI.

Step (b), TI(τ)→CI(τ)**:** The next step concerns the transformation of TI into the *celestial intermediate* (CI) coordinate system, a system somewhat analogous to the well-known apparent places (AP) system. The CI, like TI, is also quasi-instantaneous. It is geocentric with its z^{CI}-*axis* pointing to CIP defined in step (a), the x^{CI}-*axis* pointing towards the *Celestial Intermediate Origin*–CIO, the same non-rotating origin (NRO) as that of BC, and the y^{CI}-*axis* completes a right-handed coordinate system. Since x^{CI}-*axis* is fixed on CIO while the other two axes are not, explains its quasi-instantaneous character. The time scale of CI is TCG.

The transformation of TI into CI consists of only a simple rotation around z^{TI} by the *Earth Rotation Angle* – θ, an operation similar to rotation by GAST in the transformation of the IT system into the AP in the classical systems (Eq.(15.7)). The Earth rotation angle is defined as the angle reckoned along the equator of the CIP between CIO and TIO. Therefore, the transformation is simply analogous to Eq. (15.7) (cf., Step (b) in the classical transformations)

$$\vec{e}^{\,CI} = \mathbf{R}_3(-\theta)\ \vec{e}^{\,TI}. \qquad (15.49)$$

Angle θ can be calculated as a function of UT1, e.g., [CAPITAINE ET AL., (2000)] or retrieved from the Astronomical Almanac [ASTRONOMICAL EPHEMERIS AND NAUTICAL ALMANAC, 2006]for a specific TT time.

Step (c), TI(τ)→IT(τ_0)**:** Parallel to the classical systems, we now seek the transformation of the 'quasi-instantaneous' astronomical latitude $\Phi(\tau)$ and longitude $\Lambda(\tau)$ reckoned in TI(τ) into the corresponding 'conventional' quantities, reckoned in IT(τ_0), that refer to a particular reference epoch τ_0. This is analogous to the classical IT→CT transformation (cf., Step (c) in §15.1). The difference between IT and TI is that the z^{TI}-*axis* is not affected by polar motion while z^{IT} is. Consequently, in addition to the misalignment of the *z-axes*, the equatorial plane of TI differs from the one of the IT by a small angle caused by polar motion; we note here that this misalignment of the two equatorial planes was omitted in the classical transformation IT→CT.

[8] The BIH terrestrial System–BTS of 1984 was established using station coordinates derived from VLBI, LLR, SLR and Doppler/TRANSIT observations (Boucher and Altamimi, 1985). The reference meridian of BTS84 was not coincident with the Royal Observatory of Greenwich Prime Meridian but 5.3" east of it. This was accepted by BIH to be the prime meridian of BTS84 (for more details see McCarthy and Petit (2004, Section 4.2.2)

This omission was considered inconsequential for applications in the past. Now, in addition to the two well-known rotations about the *x*- and *y-axes* (cf. Eq. (15.8)), TI must be rotated clockwise about its *z-axis* by a small angle s' to locate x^{IT} (TIO) on the equatorial plane of the TI; s' is known as the *TIO locator*. The transformation is given by [PETIT AND LUZUM, 2010; Eq. (5.3)]

$$\bar{e}^{\,IT} = \mathbf{R}_2(-x_P)\ \mathbf{R}_1(-y_P)\ \mathbf{R}_3(-s')\ \bar{e}^{\,TI}, \tag{15.50}$$

and s', in angular units, is given by [IBID, 2010; Eq. (5.4)],

$$s' = \frac{1}{2}\int_{\tau_0}^{\tau}\left(x_p \dot{y}_p - \dot{x}_p y_p\right)\mathrm{d}\tau. \tag{15.51}$$

In the above formulae, x_p and y_p define the pole position of TI (i.e., the CIP) with respect to the IT pole (IRP), τ_0 is the reference epoch of IT and τ is the epoch of transformation. The '*IERS Bulletin A*' provides the position (x_p, y_p) of the CIP with respect to the IRP. Since the TIO locator is very small it suffices to use the following linear approximation [LAMBERT AND BIZOUARD, 2002] for its evaluation

$$s' \approx -47(\tau - 51544.5)/36525, \tag{15.52}$$

where τ is in modified Julian days (MJD) and s' in μarc-seconds; s' has an average amplitude of about 47 μarc-seconds. Both, IT and TI use the same atomic time scale (TT) thus, no additional terms in Eq.(15.50) are needed.

Step (d), BC(τ_0)→GC(τ_0)**:** Let us now turn to the *barycentric celestial*–BC(τ_0) coordinate system and again follow the path of transformations leading to the CI(τ) system but from the "celestial" side (cf., FIG. 15.21), as we did with the classical coordinate systems (MRA(τ_0)→AP(τ)). The first step is to transform BC(τ_0) to the *geocentric celestial*–GC(τ_0). The GC is the geocentric counterpart of BC defined previously i.e., GC differs from BC in the origin (BC uses the barycenter of the solar system, while GC, uses C) and in the time scale (BC uses TCB whereas GC uses TCG (cf., §15.2)). Therefore, the transformation between them includes: a) a translation using annual aberration and annual parallax and light bending (general relativity) and b) time system difference transformation (TCB→TCG) that involves special and general relativity terms [PETIT AND LUZUM, 2010]. The coordinates of the stars and extragalactic objects in GC will be *mean apparent* $\alpha(\tau_0)$ and $\delta(\tau_0)$. We must note here that there is no corresponding transformation in the classical coordinate systems.

Step (e), GC(τ_0)→CI(τ)**:** This is the last step in the chain of transformations that will take us to the *celestial international* (CI) coordinate system, defined in Step (b) above which is, as stated already, analogous to the classical AP(τ). Both, GC and CI are geocentric celestial coordinate systems and use the same time scale (TCG). To transform GC to CI, we will need a series of rotations to account for precession, nutation and general relativity effects, such as *geodesic* *precession-nutation*[9], and *Lense-Thirring precession* (rotational dragging of the reference frame), the details of which are considered outside the scope of this book. In addition, we will need a small rotation angle s about z^{GC}-*axis* to account for the difference between the two x-axes, because CIP (the pole of CI) is affected by precession and nutation, whereas CRP (pole of GC) is fixed. The transformation GC→ CI is thus given by [PETIT AND LUZUM, 2010; Eq. (5.6)]:

$$\vec{e}^{\,\mathrm{CI}} = \boldsymbol{R}_3(E)\,\boldsymbol{R}_2(-d)\,\boldsymbol{R}_3(-E)\;\boldsymbol{R}_3(s)\,\vec{e}^{\,\mathrm{GC}}, \tag{15.53}$$

where E and d are given by [CAPITAINE, ET AL., 2003; Eq. (2)]

$$\begin{aligned} E &= arc\tan(Y/X) \\ d &= arc\tan\left[\left((X^2+Y^2)/(1-X^2-Y^2)\right)^{1/2}\right] \end{aligned} \tag{15.54}$$

and $X(\tau)$ and $Y(\tau)$ are the Cartesian coordinates of the offset of the CIP (pole of CI) with respect to the GC due to both, precession and nutation. This representation differs from Eqs. (15.10) and (15.11) that treat precession and nutation separately, using the classical precession and nutation angles. We note here that $X(\tau)$ and $Y(\tau)$ are given by the newest precession-nutation model IAU 2006/2000A [see, e.g., IERS Conventions, 2010; Section 5.5.4], that also includes the geodesic precession-nutation [FUKUSHIMA, 1991]. The small angle s, known as the *CIO locator*, is also given as a function of $X(\tau)$ and $Y(\tau)$ [IBID, 2010; Eq. 5.8]. It is instructive to mention that the transformation CI(τ)→GC(τ_0) given by the inverse of Eq. (15.53), is very similar to TRA(τ)→MRA(τ)→ MRA(τ_0) in the classical transformation steps (e) and (d) respectively, with the only difference being in the origin of the systems and the presence of CIO locator s. In closing, the complete set of transformations among the new coordinate systems, as they have been endorsed by IAU and IERS, can be seen in FIG. 15.21. The reader can compare the traditional systems with the modern ones by direct comparison of FIGS. 15.13 and 15.21.

Step (f), OR(τ)→IT(τ_0)**:** This step concerns transformations from the orbital coordinate system (OR) where the positions of the satellites are estimated (e.g., GNSS orbits), into the IT(τ_0) coordinate system where they are used to compute the 'conventional' coordinates of the observing station. This transformation is very similar to the classical system transformations (cf., Eq. 15.35) and includes two steps: 1) OR(τ)→TI(τ) using the satellite orbital elements and the Earth rotation angle θ, and 2) TI(τ)→IT(τ_0) via TIO locator s' and polar motion to transform to IT(τ_0) (cf., Eq. (15.50)). The complete transformation reads (cf., Eq. (15.35)) that involves the equivalent transformations of the classical systems):

$$\vec{r}^{\,\mathrm{IT}} = \boldsymbol{R}_2\left(-x_p\right)\boldsymbol{R}_1\left(-y_p\right)\boldsymbol{R}_3(-s')\,\boldsymbol{R}_3(\theta)\,\boldsymbol{R}_3(-\Omega)\,\boldsymbol{R}_1(-i)\,\boldsymbol{R}_3(-\varpi)\,\vec{r}^{\,\mathrm{OR}}. \tag{15.55}$$

[9] Relativistic effect due to the rotation of the geocentric coordinate system GC with respect to the solar-system barycentric coordinate system and within the gravitational field of the solar system [e.g., BRUMBERG, 1991])

We note here that the GNNS satellite orbits are routinely provided in the World Geodetic System of 1984–WGS84 or equivalently (up to a few millimeters) in the ITRF. The WGS84 has undergone many mutations since its original conception to match the different realisations of the ITRS, i.e., to match the corresponding reference frame ITRF. Once again, current satellite-derived positions can only be referenced to the ITRF rendering the classical CT system obsolete. We will discuss ITRF and WGS84 in full detail in Chapter 16.

We now return to the astronomical positioning using the new coordinate systems noting their similarity to the classical ones as already presented in §15.3. The starting equation used to develop the mathematical models is Eq. (15.17) but modified as follows:

$$\vec{e}^{\,\mathrm{LA}} = \boldsymbol{R}_2^{-1}\left(\frac{\pi}{2} - \Phi\right)\boldsymbol{R}_3^{-1}(\pi - \Lambda)\,\boldsymbol{P}_2\,\boldsymbol{R}_3(\theta)\,\vec{e}^{\,\mathrm{CI}}\ , \tag{15.56}$$

where $\vec{e}^{\,\mathrm{AP}}$ has been replaced by $\vec{e}^{\,\mathrm{CI}}$ and GAST by the Earth rotation angle θ. If the positions of the stars in the modern systems are published in BC, the chain of transformations BC→GC→CI must be followed to transform the published α and δ of the stars into the CI. This chain of transformations is given in Steps (d) and (e) above. Similar to the classical systems (cf., 15.17), Eq. (15.56) provides the quasi-instantaneous Φ^{TI} and Λ^{TI}. Further transformation of the astronomical latitude and longitude to the IT is needed through polar motion and TIO locator s'. However, given that s' is usually much smaller that the achievable observational accuracy, at least for ordinary astronomical point positioning, Eqs. (15.29), (15.30) and (15.32) can be used directly to provide Φ^{TI} and Λ^{TI} by replacing superscripts CT and IT with IT and TI, respectively. We must note here that any astronomical position (Φ, Λ, A) can be referenced directly to the IT since the available astronomical ephemerides and polar motion components are readily available in the new systems. Transforming classical CT coordinates to the new IT or vice versa, will require the transformation parameters between the two. These include: a) the position of the IRP with respect to the old CIO, b) the transformation of the old time scale (UTC) to the new one (TT) and (c) the use of the TIO locator. These transformation parameters are available in the IERS conventions [PETIT AND LUZUM, 2010] and relevant references therein.

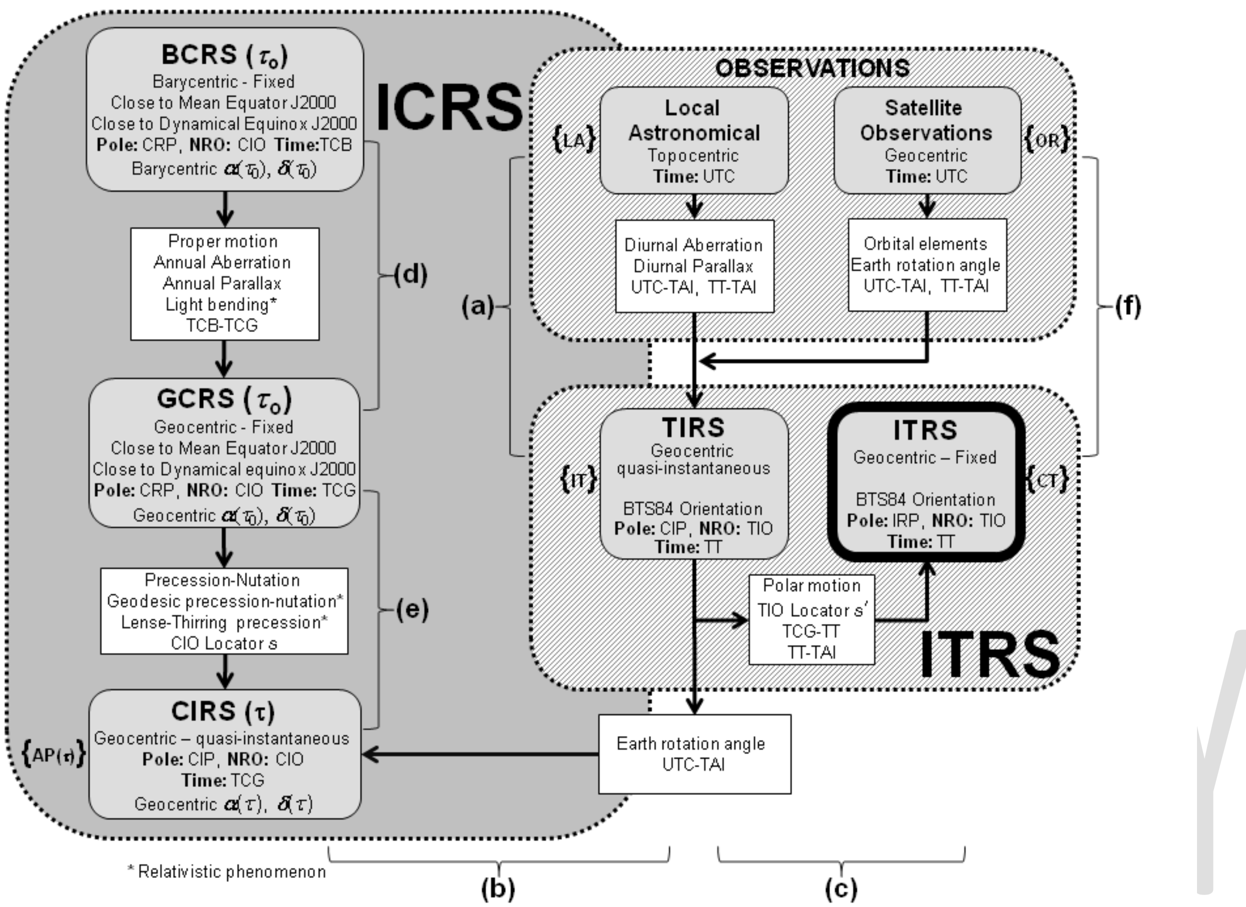

FIG. 15.21. Transformations between ICRS and ITRS associated coordinate systems. Coordinate systems in braces indicate the classical (dynamic) systems that are closest to the kinematic ones. The different transformation steps described in the text are indicated with lower case letters (a) through (d). The LA and OR systems are identical in both classical and modern systems.

15.6. Transformations of terrestrial positions

It is often necessary to transform the position of a point, normally located on the Earth's surface, from one coordinate system into another. It is natural to formulate a mathematical model and solve for the coordinates in one particular coordinate system. Later, one may want to refer the coordinates of the same point to another coordinate system, and that is where the necessity for a point transformation arises. In §15.3, we encountered one such situation in which the astronomically determined position Φ, Λ was solved for in the IT system and later transformed into the CT system. Central to the point transformations treated here is the concept of the reference (biaxial) ellipsoid, with its geodetic curvilinear coordinates ϕ, λ, h (cf. §7.1). The following topics are included in this section:

(a) The transformation of the geodetic curvilinear coordinates $(\phi,\lambda,h)^G$ into their representative Cartesian coordinates $(x, y, z)^G$ – cf. §3.3 – and vice versa.

(b) The transformation of the CT (or ITRS) coordinates $(x,y,z)^{CT}$ into non-geocentric geodetic coordinates $(\phi, \lambda, h)^G$ and vice versa.

(c) The transformation of the astronomical coordinates (Φ, Λ) into geodetic curvilinear coordinates (ϕ, λ, h), along with the transformation of the astronomical azimuth (A) into geodetic azimuth (a), and the orthometric height (H) into geodetic height (h) and vice versa. The ways of positioning of a geodetic reference ellipsoid within the Earth are discussed within these transformations.

(d) The transformation of one triplet of geodetic curvilinear coordinates $(\phi,\lambda,h)_1$ referred to an ellipsoid (a_1, b_1) into another triplet $(\phi, \lambda, h)_2$ referred to another ellipsoid (a_2, b_2).

(e) The transformation of horizontal, geodetic curvilinear coordinates (ϕ,λ) into map coordinates $(x, y)^M$ and vice versa.

In the above list, there are two classes of transformation. The first class, consisting of transformation (a), is the transformation within one family of coordinate systems (see §3.3). The second is the class of transformation between families of coordinate systems with different locations and orientation

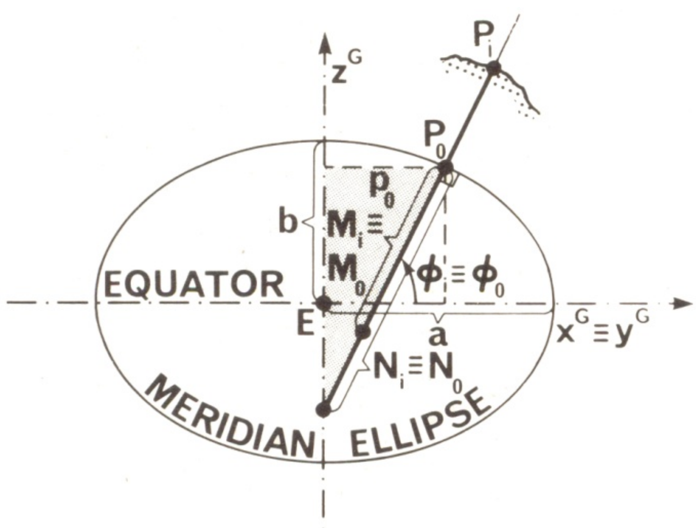

FIG. 15.22. Geometry of a biaxial ellipsoid.

Before getting into the first transformation, let us derive the expression for one more entity related to the biaxial ellipsoid. From Eq. (3.90) we obtain, for any point P_0 on the ellipsoid (see FIG. 15.22).

$$N_0 \cos\phi_0 = p_0, \tag{15.57}$$

where N_0 is called the *prime vertical radius of curvature* at P_0 and can be derived from Eqs. (7.10) and (7.12) as

$$N_0 = \frac{a^2}{\left(a^2\cos^2\phi_0 + b^2\sin^2\phi_0\right)^{1/2}}. \tag{15.58}$$

It is clearly akin to M given by Eq. (7.14).

(a) In the first transformation, realising that in the representative Cartesian *geodetic system* (G) of coordinates (corresponding to the curvilinear system) one has

$$x_0^{\mathrm{G}} = p_0 \cos\lambda_0, \qquad y_0^{\mathrm{G}} = p_0 \sin\lambda_0\,, \tag{15.59}$$

we can write the position vector of the normal projection P_0 of P_i on the ellipsoid simply as

$$\begin{aligned} \vec{r}_0^{\,\mathrm{G}} &= \vec{r}^{\,\mathrm{G}}(\phi_0,\lambda_0) = \vec{r}^{\,\mathrm{G}}(\phi_i,\lambda_i) \\ &= N_0 \begin{bmatrix} \cos\phi_0 \cos\lambda_0 \\ \cos\phi_0 \sin\lambda_0 \\ (b^2/a^2)\sin\phi_0 \end{bmatrix} = N_i \begin{bmatrix} \cos\phi_i \cos\lambda_i \\ \cos\phi_i \sin\lambda_i \\ (b^2/a^2)\sin\phi_i \end{bmatrix} \end{aligned} \tag{15.60}$$

where the z^{G}-component is given by Eq. (7.10). To obtain the position vector of a P_i located above point P_0 on the ellipsoid, the two constituent vectors are added as follows (see FIG. 15.23):

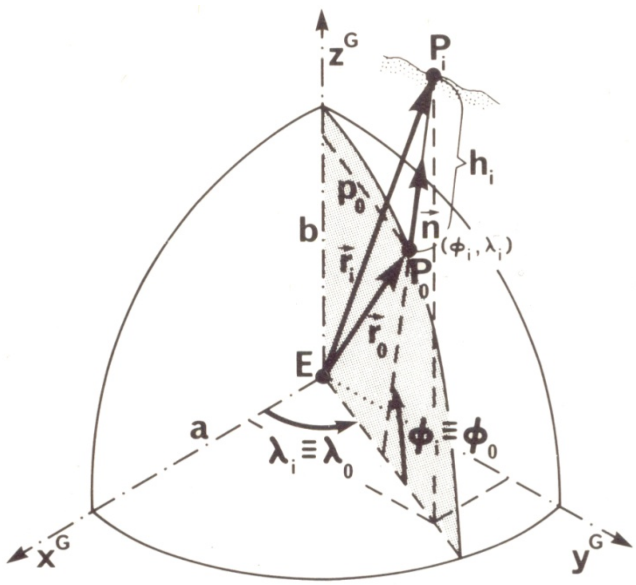

FIG. 15.23. Points on and above the reference ellipsoid.

$$\vec{r}_i^{\,G} = \vec{r}^{\,G}(\phi_i, \lambda_i) + h_i \vec{n}^{\,G}(\phi_i, \lambda_i), \tag{15.61}$$

where

$$\vec{e}^{\,G} = \vec{n}^{\,G}(\phi_i, \lambda_i) = \begin{bmatrix} \cos\phi_i \cos\lambda_i \\ \cos\phi_i \sin\lambda_i \\ \sin\phi_i \end{bmatrix} \tag{15.62}$$

is the unit vector normal to the ellipsoid at P_0, while h_i is the geodetic height of P_i (cf. §7.1). The resultant position vector then is equal to

$$\boxed{\vec{r}_i^{\,G} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} (N_i + h_i)\cos\phi_i \cos\lambda_i \\ (N_i + h_i)\cos\phi_i \sin\lambda \\ (N_i b^2/a^2 + h_i)\sin\phi_i \end{bmatrix}.} \tag{15.63}$$

This is the transformation equation of (ϕ, λ, h) into the G system.

The inverse transformation can be solved by iterations or in closed form. Both approaches employ the distance p from the minor axis which, for any point P (after dropping the subscript i), equals

$$p = \left(x^2 + y^2\right)^{1/2}, \tag{15.64}$$

or from Eq. (15.63),

$$p = (N + h)\cos\phi. \tag{15.65}$$

From Eqs. (15.63) and (7.10) we have

$$z = \left(N + h - e^2 N\right)\sin\phi, \tag{15.66}$$

and finally

$$\frac{z}{p} = \tan\phi\left(1 - \frac{e^2 N}{N+h}\right). \tag{15.67}$$

This equation is the point of departure for both approaches.

The iterations are usually initiated by solving first for ϕ from the above equation [HEISKANEN AND MORITZ, 1967]. Putting $h = 0$, we get

$$\phi^{(0)} = \arctan\left(\frac{z}{p}\left(1-e^2\right)^{-1}.\right) \tag{15.68}$$

The kth iteration then consists of evaluating successively $N^{(k)} = N(\phi^{(k\text{-}1)})$ from Eq. (15.58); $h^{(k)} = h(\phi^{(k\text{-}1)}, N^{(k)})$, from Eq. (15.65); and $\phi^{(k)} = \phi(N^{(k)}, h^{(k)})$ from Eq. (15.68). The iterations are repeated until the following inequalities are satisfied:

$$\left|h^{(k)} - h^{(k-1)}\right| < a\varepsilon \quad \text{and} \quad \left|\phi^{(k)} - \phi^{(k-1)}\right| < \varepsilon\,, \tag{15.69}$$

for some a priorily chosen value of ε. Once ϕ and h are found, λ is evaluated from either of the first two Eqs. (15.63) or

$$\lambda = 2\arctan\frac{y}{x + \sqrt{x^2 + y^2}}. \tag{15.70}$$

The closed form solution uses Eqs. (15.65) and (15.66) to obtain

$$p\ \tan\phi - z = e^2 N\ \sin\phi\ . \tag{15.71}$$

In this equation, the only unknown is ϕ, N being a function of ϕ as well. Substituting N from Eq. (15.58), Eq. (15.59) changes to

$$p\ \tan\phi - z = \frac{ae^2\ \sin\phi}{\left(\cos^2\phi + \left(b^2/a^2\right)\sin^2\phi\right)^{1/2}}. \tag{15.72}$$

Dividing the numerator and denominator of the right-hand side by $\cos\phi$ and squaring the whole equation yields

$$p^2\tan^4\phi - 2pz\ \tan^3\phi + \left(z^2 + \frac{p^2 - a^2e^4}{1-e^2}\right)\tan^2\phi - \frac{2pz}{1-e^2}\ \tan\phi + \frac{z^2}{1-e^2} = 0\,. \tag{15.73}$$

This is a quartic (biquadratic) equation in $\tan\phi$ in which the values of all the coefficients are known. Standard procedures for solving quartic equations exist (e.g., KORN AND KORN [1968]). Once a solution for $\tan\phi$ is obtained, N and h are computed from Eqs. (15.58) and (15.65) respectively. Longitude λ follows directly from Eq. (15.63) or Eq. (15.70) thus, completing the inverse transformation. PAUL [1973] showed that the closed form approach is about 25% faster than the iterative. It should be noted that since the (ϕ, λ, h) system is a two-parametric system of coordinates (cf. §3.3), the two parameters a, b (or a, e, or some other combination) play a role in all the above transformations.

(b) The second transformation from the CT into the G system requires knowledge of the position and orientation of the reference ellipsoid within the Earth. This task of positioning and orienting the reference ellipsoid is known as the *establishment of a horizontal geodetic datum* [YEREMEHEF AND YURKINA, 1969; MATHER, 1970; PICK AT AL., 1973]: the reference ellipsoid, defined by the values of its parameters a, b, then becomes the datum–a specific coordinate surface (cf. §3.3). The positioning of the ellipsoid requires six more parameters to eliminate its six degrees of freedom, i.e., the six ways in which the ellipsoid can move relative to the Earth. These six bring the total number of *datum parameters* to eight. Since we are interested in the transformation of the (ϕ, λ, h) into the CT system, it is natural to specify the six datum position parameters at the Earth's centre of mass – the *geocentric set of datum position parameters* – as the three CT coordinates of the ellipsoid's centre, called *datum translation components* x_E, y_E, z_E, and the three *datum misalignment angles,* ε_x, ε_y, ε_z, required to define the misalignment between the two sets of axes (see FIG. 15.24). The ways datums are positioned in practice will be shown in §17.1 and §18.1. The derivation of the inverse transformation is left to the reader.

The transformation from $(\phi, \lambda, h)^G$ into $(x,y,z)^{CT}$ is done in two steps: first $(\phi, \lambda, h)^G \rightarrow (x,y,z)^G$ using Eq. (15.63), and then $(x,y,z)^G \rightarrow (x,y,z)^{CT}$ employing the following formula

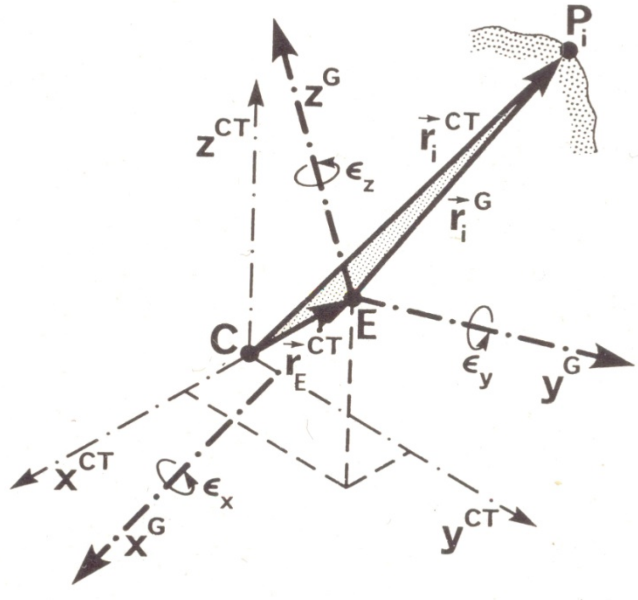

FIG. 15.24. Geocentric set of datum position parameters.

$$\vec{r}^{\,\mathrm{CT}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\mathrm{CT}} = \boldsymbol{R}_1(\varepsilon_x)\boldsymbol{R}_2(\varepsilon_y)\boldsymbol{R}_3(\varepsilon_z)\begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\mathrm{G}} + \begin{bmatrix} x_E \\ y_E \\ z_E \end{bmatrix}^{\mathrm{CT}} . \tag{15.74}$$

It is obviously advantageous to have the two sets of axes parallel, i.e., $\varepsilon_x = \varepsilon_y = \varepsilon_z = 0$, so that the above equation simplifies to

$$\vec{r}^{\,\mathrm{CT}} = \vec{r}^{\,\mathrm{G}} + \vec{r}_E^{\,\mathrm{CT}} . \tag{15.75}$$

In converting the CT coordinates (obtained, e.g., from satellite positioning) to the G system, the accuracy of the results depends on the accuracy of $\vec{r}_i^{\,\mathrm{CT}}$. The reader can verify that the above transformations are also valid when CT is replaced by ITRS. Currently, the origin of the GRS80 reference ellipsoid associated with the ITRS is at the geocentre as determined by SLR observations, whereas its orientation is determined by VLBI. Present day point positioning is done globally on the GRS80 reference ellipsoid therefore, $(x,y,z)^{\mathrm{ITRS}}=(x,y,z)^{\mathrm{GRS80}}$, and the transformation defined by Eq. (15.74) is not necessary.

(c) The next transformation should enable us to transform the astronomically determined positions (Φ, Λ) to geodetic positions (ϕ, λ). It is not as clear geometrically as the transformations treated above and requires some preliminary cognizance. In particular, the position of the G system with respect to the Earth's gravity field, the framework for the Φ, Λ coordinates, has to be properly understood. To explain the concepts involved, one has to use two more coordinate systems: the LA system, introduced in §15.1, and the local geodetic system. It is important to note that the following transformations refer only to the traditional definition of the coordinate systems and the traditional astronomical definition of the position and orientation of the (local) reference ellipsoid with respect to some point on the surface of the Earth.

The *local geodetic system* (LG) is defined as follows (see FIG. 15.25): it is topocentric (T); the z^{LG}-axis is the outward ellipsoid normal passing through T; the x^{LG}-axis is directed toward *geodetic north,* i.e., it lies in the *geodetic meridian* plane defined by the ellipsoid normal at T and the minor axis (z^{G}) of the reference ellipsoid; and the y^{LG}-axis is chosen so that the system is left-handed. Note that the LG system makes angles ϕ and λ with the G system; clearly the relation of the LG system to the G system is analogous to the relation of the LA system to the CT system, as the reader can see by comparing FIGS. 15.25 and 15.6. The analogy between the LG and LA systems goes further: analogous to the astronomical quantities (see §15.1), here we define the *geodetic vertical angle* ν', the *geodetic zenith distance* Z', and the *geodetic azimuth* α, all as shown in FIG. 15.25. We note, however, that while the LA system is a natural system dictated by the physical properties of the Earth, the LG system is not.

Let us now examine the relations among the four coordinate systems (CT, LA, G, LG) shown in FIG. 15.26. A unit vector $\vec{e}^{\,\mathrm{LA}}$ can be rotated into the CT system as follows. From Eq. (15.74) we have

$$\vec{e}^{\,\mathrm{CT}} = \boldsymbol{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z)\,\vec{e}^{\,\mathrm{G}} . \tag{15.76}$$

Recalling the analogy between the pairs CT, LA and G, LG of systems, an equation analogous to (15.6) is used to transform $\vec{e}^{\,\mathrm{LA}}$ to $\vec{e}^{\,\mathrm{G}}$. Finally, from FIG. 15.26, and keeping in mind that both systems are left-handed, we get

$$\vec{e}^{\,\mathrm{LG}} = \boldsymbol{R}_3(\Delta\alpha)\,\boldsymbol{R}_2(-\xi)\,\boldsymbol{R}_1(\eta)\,\vec{e}^{\,\mathrm{LA}} = \boldsymbol{R}^{\mathrm{T}}(\Delta\alpha, -\xi, n)\,\vec{e}^{\,\mathrm{LA}}, \tag{15.77}$$

so that

$$\vec{e}^{\,\mathrm{CT}} = \boldsymbol{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z)\,\boldsymbol{R}_3(\pi-\lambda)\,\boldsymbol{R}_2\!\left(\frac{\pi}{2}-\phi\right)\boldsymbol{P}_2\;\boldsymbol{R}^{\mathrm{T}}(\Delta\alpha, -\xi, \eta)\,\vec{e}^{\,\mathrm{LA}}. \tag{15.78}$$

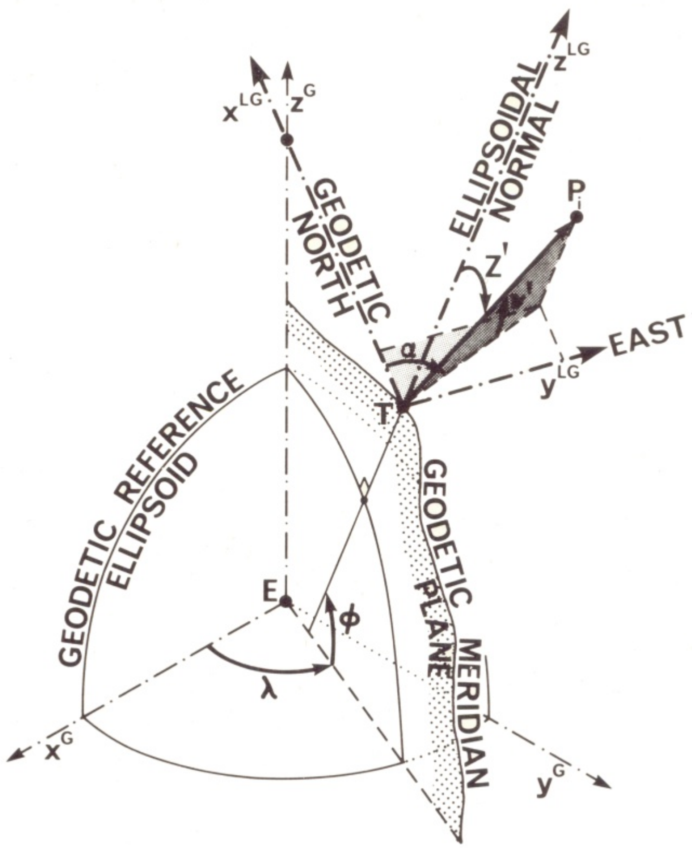

FIG. 15.25. Local geodetic system.

To maintain consistency among the four systems, this equation must be equivalent to Eq. (15.6), which describes the same transformation only using different transformation parameters. The following equation, hence, must be valid:

$$
\begin{aligned}
&\boldsymbol{R}_3(\pi-\Lambda)\,\boldsymbol{R}_2\!\left(\frac{\pi}{2}-\Phi\right)\boldsymbol{P}_2 \\
&\quad=\boldsymbol{R}(\varepsilon_x,\,\varepsilon_y,\,\varepsilon_z)\,\boldsymbol{R}_3(\pi-\lambda)\,\boldsymbol{R}_2\!\left(\frac{\pi}{2}-\phi\right)\boldsymbol{P}_2\;\boldsymbol{R}^T(\Delta\alpha,-\xi,\eta).
\end{aligned}
\tag{15.79}
$$

In this equation, ε_x, ε_y, ε_z, $\Delta\alpha$, ξ, η, $\Lambda-\lambda$, $\Phi-\phi$ would normally be very small quantities. Thus, their trigonometric functions can be developed into power series and only the first terms retained. This results in the following condition [VANÍČEK AND CARRERA, 1985]:

$$
\begin{bmatrix}\Delta A\\ \xi\\ \eta\end{bmatrix}=\begin{bmatrix}(\Lambda-\lambda)\sin\phi\\ \Phi-\phi\\ (\Lambda-\lambda)\cos\phi\end{bmatrix}-\begin{bmatrix}\cos\phi(\varepsilon_x\cos\lambda+\varepsilon_y\sin\lambda)+\varepsilon_z\sin\phi\\ \varepsilon_x\sin\lambda-\varepsilon_y\cos\lambda\\ -\sin\phi(\varepsilon_x\cos\lambda+\varepsilon_y\sin\lambda)+\varepsilon_z\cos\phi\end{bmatrix},
\tag{15.80}
$$

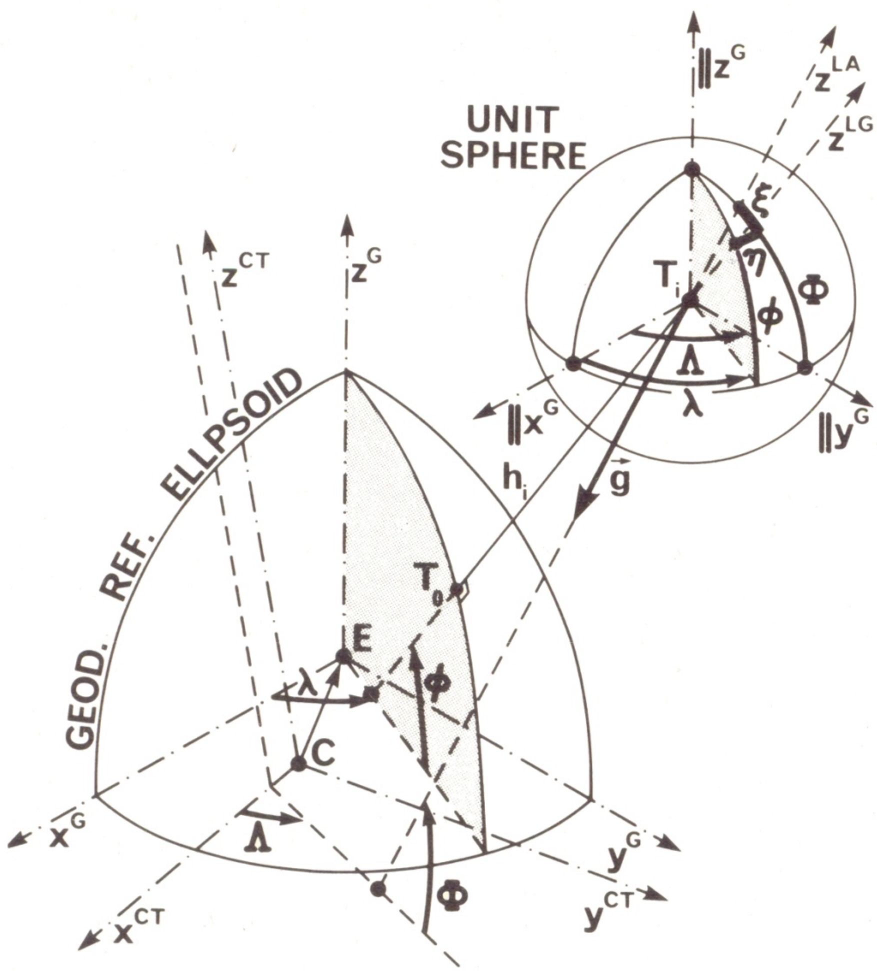

FIG. 15.26. Parallelism conditions.

which must be satisfied by all the quantities involved: geodetic, ε_x, ε_y, ε_z, ϕ , λ , α; astronomical, Φ, Λ, A; and the surface deflection components ξ, η. It is interesting to note that Eq. (15.80) can be simplified to

$$\boxed{\begin{bmatrix} \Delta A \\ \xi \\ -\eta \end{bmatrix} = \begin{bmatrix} (\Lambda-\lambda)\sin\phi \\ \Phi-\phi \\ -(\Lambda-\lambda)\cos\phi \end{bmatrix} - \mathbf{R}_2(\phi-\pi)\,\mathbf{R}_3(\lambda-\pi) \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{bmatrix}.} \tag{15.81}$$

It should also be pointed out that if the G system (and the reference ellipsoid with it) is positioned and oriented with respect to the CT system locally at a point T_0, a *topocentric set of datum position parameters* is needed. These six independent parameters can be, for instance, ϕ_0 , λ_0, α_0, ξ_0, η_0, N_0 [VANÍČEK AND WELLS, 1974]. If this mode of datum positioning (and orientation) is used, then Eqs. (15.81) must take a special form, namely:

$$\begin{bmatrix} \Delta A \\ \xi \\ -\eta \end{bmatrix} = \begin{bmatrix} (\Lambda-\lambda)\sin\phi \\ \Phi-\phi \\ -(\Lambda-\lambda)\cos\phi \end{bmatrix} - \boldsymbol{R}_2(\phi-\pi)\,\boldsymbol{R}_3(\lambda-\pi) \begin{bmatrix} \cos\phi_0\ \cos\lambda_0 \\ \cos\phi_0\ \sin\lambda_0 \\ \sin\lambda_0 \end{bmatrix} \Delta_0 , \tag{15.82}$$

where Δ_0 is the misalignment angle reckoned around the ellipsoidal normal at T_0. We note that if the G system was properly aligned with the CT system, i.e., if $\varepsilon_x = \varepsilon_y = \varepsilon_z = 0$, the second term on the right-hand side of Eqs. (15.81) and (15.82) would vanish. We would be left with a system of three very simple, but very important, equations:

$$\boxed{\Delta A = A-\alpha = (\Lambda-\lambda)\sin\phi ,} \tag{15.83}$$

known as the *Laplace equation* for azimuths, and

$$\boxed{\Phi-\phi = \xi} \tag{15.84}$$

$$\boxed{(\Lambda-\lambda)\cos\phi = \eta} \tag{15.85}$$

which are the defining equations for the Meridian and prime vertical components of the deflection of the vertical (cf. §6.4) in terms of astronomical and geodetic coordinates in the absence of any misalignment. It should be clear that, since the equations are formulated for a point on the Earth's surface, the deflections are of the surface type. If the point happens to be on the geoid, then we have the geoid deflections.

The above three equations constitute the topocentric *conditions for parallelism* for the G and CT systems. If the parallelism is desired, then Eqs. (15.83) to (15.85) must be satisfied for all points on the Earth's surface (including the origin T_0), bearing in mind that Φ, Λ, A are directly observable.

It is of interest to realise that ΔA is the angle between the x^{LG}- and x^{LA}-axes. Using Eq. (15.85), we can write the Laplace equation in yet another form: namely,

$$A-\alpha = \Delta A = \eta\ \tan\phi . \tag{15.86}$$

An alternative set of parallelism condition equations can be obtained by considering the observable quantities A and Z. Rotating a unit vector in the LA system into the LG system (see FIGS. 15.3, 15.25, and 15.26), we obtain

$$\vec{e}^{\,\mathrm{LG}} = \boldsymbol{R}_3(\Delta A)\,\boldsymbol{R}_2(-\xi)\,\boldsymbol{R}_1(\eta)\,\vec{e}^{\,\mathrm{LA}}, \tag{15.87}$$

or

$$\begin{bmatrix} \cos\alpha\,\sin Z' \\ \sin\alpha\,\sin Z' \\ \cos Z' \end{bmatrix} \approx \begin{bmatrix} 1 & \Delta A & \xi \\ -\Delta A & 1 & \eta \\ -\xi & -\eta & 1 \end{bmatrix} \begin{bmatrix} \cos A\,\sin Z \\ \sin A\,\sin Z \\ \cos Z \end{bmatrix}.$$

Expanding the left-hand side into a Taylor series at (A, Z), we get

$$\begin{gathered} (A-\alpha)\begin{bmatrix} \sin A\,\sin Z \\ -\cos A\,\sin Z \\ 0 \end{bmatrix} + (Z-Z')\begin{bmatrix} -\cos A\,\cos Z \\ -\sin A\,\cos Z \\ \sin Z \end{bmatrix} \\ \approx \begin{bmatrix} 0 & \Delta A & \xi \\ -\Delta A & 0 & \eta \\ -\xi & -\eta & 0 \end{bmatrix} \begin{bmatrix} \cos A\,\sin Z \\ \sin A\,\sin Z \\ \cos Z \end{bmatrix}. \end{gathered} \tag{15.88}$$

The third equation gives directly

$$\boxed{Z - Z' \approx -\xi\,\cos A - \eta\,\sin A\,.} \tag{15.89}$$

Multiplication of the first equation by sin A, the second by cos$A$, subtraction of the second from the first, and substitution for ΔA from Eq. (15.86) yields:

$$\boxed{A - (\xi\,\sin A - \eta\,\cos A)\cot Z - \alpha \approx \eta\,\tan\Phi\,.} \tag{15.90}$$

These equations can be used when precise zenith distances are observed in the network, i.e., in the three-dimensional approach (see § 17.1). We note that the second term on the left-hand side of Eq. (15.81) is simply a correction to the observed astronomical azimuth A to relate it to the same ellipsoidal normal (see FIG. 15.24) as the geodetic azimuth α. We will return to these two sets of conditions for parallelism in the context of three-dimensional (see §17.1) and horizontal (§18.1) networks.

We can now finally formulate the transformation of the natural astronomical (physically meaningful) quantities Φ, Λ, A, Z, H into geodetic (conventional) quantities (ϕ, λ, α, Z', h). If the parallelism conditions are satisfied, the transformations $\phi \leftrightarrow \Phi$, and $\lambda \leftrightarrow \Lambda$ are given by Eqs. (15.84) and (15.85) and the transformations $\alpha \leftrightarrow A$ and $Z' \leftrightarrow Z$ by Eqs. (15.90) and (15.89), respectively. The accuracy of ϕ, λ, α, and Z' so obtained is dictated by the accuracy of Φ, Λ, A, and Z (about 0.1″ to 0.2″ as seen in §15.2) and by the accuracy of the known values of ξ and η (about 1", see §24.3). The inverse transformations are hardly ever used. It should be noted that the geodetic coordinates so derived refer to the same reference ellipsoid (the same position within the Earth as well as the same shape) as the one used for ξ and η. Also, it is clear that if the reference ellipsoid is not aligned to the CT system, in which Φ and Λ are given, then the misalignment angles must be taken into account, as should be clear from the foregoing explanations.

The height above the ellipsoid is related to the height above the sea level H and geoid height N simply through Eq. (7.3). The accuracy of the transformations $h \leftrightarrow H$ is limited by the accuracy of N, presently about a few decimetres, and that of h (cf. §16.1 and §17.1) and H (cf. §16.4 and Chapter 19).

(d) The next transformation to be considered is that between coordinates referred to two different datums. In this transformation it is necessary to account for the difference in the location of the geometrical centre of each reference ellipsoid, the difference in size and shape of the two ellipsoids, and the difference in orientation.

Consider the ellipsoids with sizes and shapes defined by (a_1, f_1) and (a_2, f_2) or alternatively (a_1, b_1) and (a_2, b_2), and the locations of their geometrical centres with respect to the geocentre defined by $\vec{r}_E^{\,1}$ and $\vec{r}_E^{\,2}$. Their misalignment angles with respect to the CT system are (ε_{x1}, ε_{y1}, ε_{z1}), and (ε_{x2}, ε_{y2}, ε_{z2}), respectively. Let us also denote the coordinates of a point referred to the first datum by (ϕ_1, λ_1, h_1); we wish to find coordinates (ϕ_2, λ_2, h_2) of the same point referred to the second datum.

There are two techniques for obtaining (ϕ_2, λ_2, h_2) as functions of (ϕ_1, λ_1, h_1). The first, a direct approach, is to find CT coordinates from Eqs. (15.63) and (15.64) and then find (ϕ_2, λ_2, h_2) using either the iterative method or the closed form inverse solution, as shown earlier under (a) in this section. The second technique, which we show here, is a differential technique that can be applied when the parameter differences $\delta a = a_2 - a_1$, $\delta f = f_2 - f_1$, $\delta x_E = x_{E2} - x_{E1}, \ldots, \delta\varepsilon_z = \varepsilon_{z2} - \varepsilon_{z1}$, for the two datums are sufficiently small. This method is as follows [Krakiwsky and Wells, 1971; p.47]:

Let the CT coordinates of a point referred to a geodetic datum be given by Eq. (15.74). This can be rewritten for small misalignment angles as

$$\begin{aligned}\vec{r}^{\,\mathrm{CT}} &= \begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\mathrm{G}} + \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}^{\mathrm{G}} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{bmatrix} + \begin{bmatrix} x_E \\ y_E \\ z_E \end{bmatrix}^{\mathrm{CT}} \\ &= \vec{r}^{\,\mathrm{G}} + \boldsymbol{T}^{\mathrm{G}} \vec{\varepsilon} + \vec{r}_E^{\,\mathrm{CT}} \ ,\end{aligned} \tag{15.91}$$

as the reader can check for himself. An identical equation can be written for the other datum. Distinguishing between the two datums by added subscripts and subtracting the first from the second equation, we get

$$\vec{r}_2^{\,\mathrm{G}} - \vec{r}_1^{\,\mathrm{G}} + T^{\mathrm{G}}\left(\vec{\varepsilon}_2 - \vec{\varepsilon}_1\right) + \vec{r}_{E2}^{\,\mathrm{CT}} - \vec{r}_{E1}^{\,\mathrm{CT}} \approx \vec{0}. \tag{15.92}$$

Expressing now the geodetic Cartesian coordinates $\vec{r}_1^{\,\mathrm{G}}$, $\vec{r}_2^{\,\mathrm{G}}$ in terms of corresponding curvilinear geodetic coordinates from Eqs. (15.63), we get, after lengthy development,

$$\vec{r}_2^{\,\mathrm{G}} - \vec{r}_1^{\,\mathrm{G}} = \boldsymbol{J}\begin{bmatrix} \phi_2 & - & \phi_1 \\ \lambda_2 & - & \lambda_1 \\ h_2 & - & h_1 \end{bmatrix} + \boldsymbol{B}\begin{bmatrix} \delta a \\ \delta f \end{bmatrix}, \tag{15.93}$$

where, using spherical approximations ($f = 0$, $N = M = a$, $h = 0$),

$$\boldsymbol{J} \approx \begin{bmatrix} -a\sin\phi\cos\lambda & -a\cos\phi\sin\lambda & \cos\phi\cos\lambda \\ -a\sin\phi\sin\lambda & a\cos\phi\cos\lambda & \cos\phi\sin\lambda \\ a\cos\phi & 0 & \sin\phi \end{bmatrix}, \tag{15.94}$$

$$\boldsymbol{B} \approx \begin{bmatrix} \cos\phi\cos\lambda & a\sin^2\phi\cos\phi\cos\lambda \\ \cos\phi\sin\lambda & a\sin^2\phi\cos\phi\sin\lambda \\ \sin\phi & a(\sin^2\phi - 2)\sin\phi \end{bmatrix}. \tag{15.95}$$

Substitution of Eq. (15.93) in Eq. (15.92) yields

$$\boldsymbol{J}\left(\begin{bmatrix} \phi_2 \\ \lambda_2 \\ h_2 \end{bmatrix} - \begin{bmatrix} \phi_1 \\ \lambda_1 \\ h_1 \end{bmatrix}\right) + \boldsymbol{B}\begin{bmatrix} \delta a \\ \delta f \end{bmatrix} + \begin{bmatrix} \delta x_E \\ \delta y_E \\ \delta z_E \end{bmatrix} + \boldsymbol{T}\begin{bmatrix} \delta\varepsilon_x \\ \delta\varepsilon_y \\ \delta\varepsilon_z \end{bmatrix} \approx \vec{0}\,, \tag{15.96}$$

and the desired transformation equation finally reads

$$\begin{bmatrix} \phi_2 \\ \lambda_2 \\ h_2 \end{bmatrix} \approx \begin{bmatrix} \phi_1 \\ \lambda_1 \\ h_1 \end{bmatrix} - \boldsymbol{J}^{-1}\left(\begin{bmatrix} \delta x_E \\ \delta y_E \\ \delta z_E \end{bmatrix} + \boldsymbol{T}\begin{bmatrix} \delta\varepsilon_x \\ \delta\varepsilon_y \\ \delta\varepsilon_z \end{bmatrix} + \boldsymbol{B}\begin{bmatrix} \delta a \\ \delta f \end{bmatrix}\right), \tag{15.97}$$

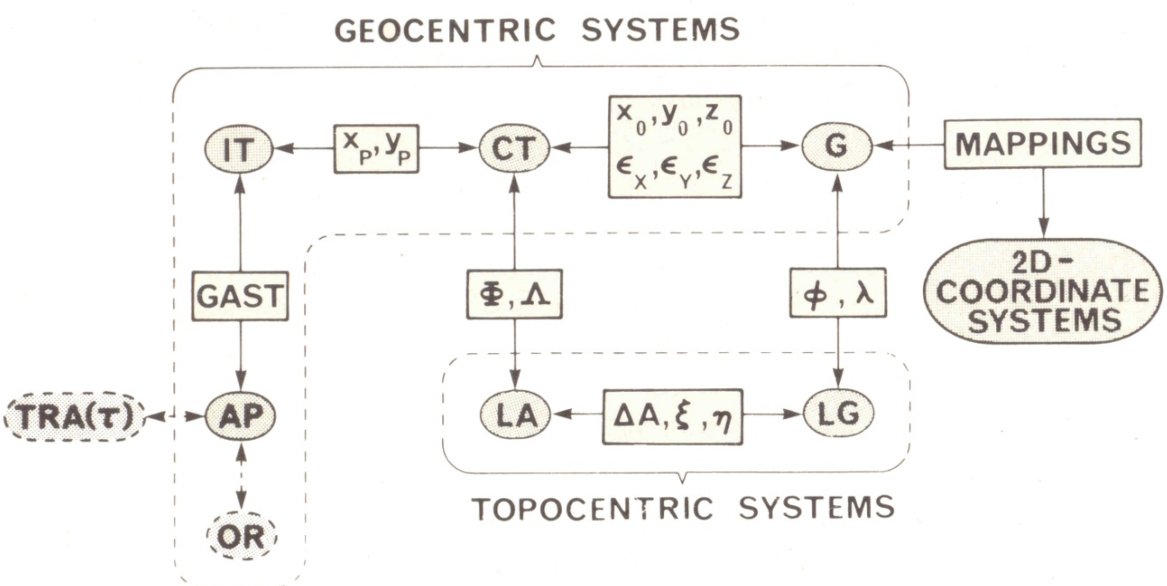

FIG. 15.27. Commutative diagram of transformations of terrestrial positions

where

$$\boldsymbol{J}^{-1} \approx \begin{bmatrix} -\sin\phi\,\cos\lambda/a & -\sin\phi\,\sin\lambda/a & \cos\phi/a \\ -\sin\lambda/(a\,\cos\phi) & -\cos\lambda/(a\,\cos\phi) & 0 \\ \cos\phi\,\cos\lambda & \cos\phi\,\sin\lambda & \sin\phi \end{bmatrix}. \tag{15.98}$$

Matrices $\boldsymbol{B}$, $\boldsymbol{J}$, and $\boldsymbol{T}$ can be evaluated on either of the two datums, since the differences in the ellipsoids are assumed to be small. Matrix $\boldsymbol{T}$ is best written as

$$\boldsymbol{T} = a \begin{bmatrix} 0 & -\sin\phi & \cos\phi\,\sin\lambda \\ \sin\phi & 0 & -\cos\phi\,\cos\lambda \\ -\cos\phi\,\sin\lambda & \cos\phi\,\cos\lambda & 0 \end{bmatrix}. \tag{15.99}$$

(e) The direct and inverse transformations of the geodetic latitude and longitude (ϕ,λ) into two-dimensional Cartesian *map coordinates* $(x, y)^{\mathrm{M}}$, and vice versa, can be written in terms of mapping equations usually simply called *mappings:*

$$x = x(\phi, \lambda), \qquad y = y(\phi, \lambda), \tag{15.100}$$

$$\phi = \phi(x, y), \qquad \lambda = \lambda(x, y). \tag{15.101}$$

As the general theory of mapping lies within the realm of mathematical cartography, mappings will not be dealt with in this book. Suffice it to say here that a multitude of mappings exist, which can be used to represent the horizontal geodetic coordinates, and the interested reader is referred to the literature (e.g., HOTINE [1946, 1947]; RICHARDUS AND ADLER [1972]; MALING [1973]). Of particular importance to geodesy are the conformal mappings that are used in various geodetic computations; these will be treated, more appropriately, in §16.3.

In closing, a commutative diagram showing all the transformations treated in this section is presented in FIG. 15.27. It is instructive to note that the AP system is the link between the terrestrial and the other systems. The symmetry of the (CT, LA) and (G, LG) pairs, mentioned earlier in this chapter, is also clearly visible. Finally we note the fact that the transformations between these four systems make a closed diagram, which explains why a datum (G system) can be positioned, as shown above under (c), in two different ways: globally and locally.

Based on the definition of the modern reference coordinate systems (cf. §15.5; FIGS. 15.13 and 15.21), FIG. 15.27 can be modified by the following identities: CT≡ITRS, IT≡TIRS, AP≡CIRS and G≡GRS80. Since GRS80 is coincident with the coordinate system associated with the ITRS, then $(x, y, z)^{\mathrm{ITRS}} \equiv (x, y, z)^{\mathrm{GRS80}}$. In addition, since modern geodetic point positioning does not use a local datum any longer, then LG does not exist and the Laplace equations (cf. Eqs. (15.83), (15.84) and (15.85)) are only used to directly transform $(\phi, \lambda) \leftrightarrow (\Phi, \Lambda)$. Clearly, $(\phi, \lambda)^{\mathrm{GRS80}}$ can directly be transformed to $(x, y)^{\mathrm{M}}$. Clearly, transformation (d) from one datum to another is only applied today on legacy data referenced to different datums.