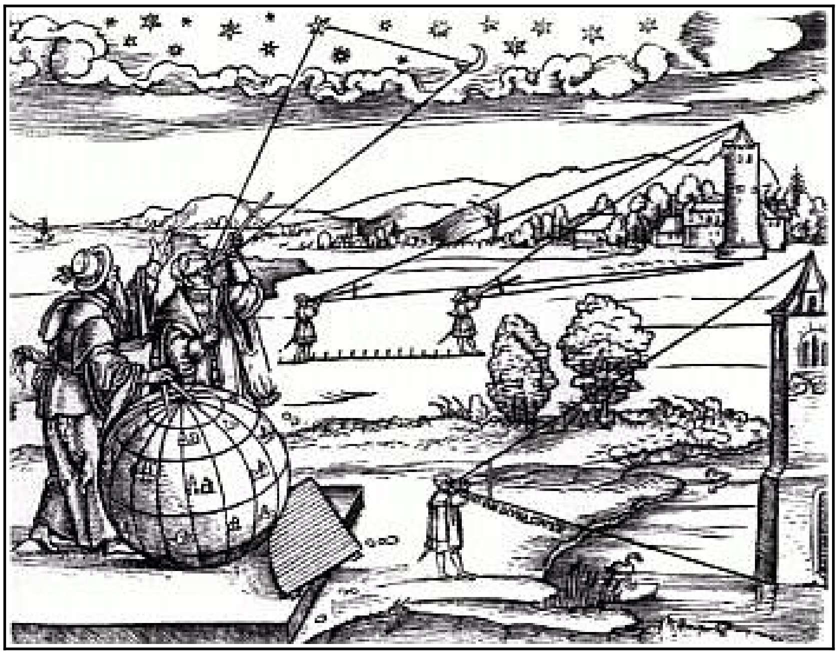
A Short Guide to
Celestial Navigation



Latest Revision: March 15th, 2019

First Published: May 20th, 1997

Much is due to those who first broke the way to knowledge,
and left only to their successors the task of smoothing it.

Samuel Johnson

Preface

Why should anybody still practice celestial navigation in the era of electronics and GPS? One might as well ask why some photographers still develop black-and-white photos in their darkroom instead of using a digital camera. The answer would be the same: because it is a noble art, and because it is rewarding. No doubt, a GPS navigator is a powerful tool, but using it becomes routine very soon. In contrast, celestial navigation is an intellectual challenge. Finding your geographic position by means of astronomical observations requires knowledge, skillfulness, and critical judgment. In other words, you have to use your brains. Everyone who ever reduced a sight knows the thrill I am talking about. The way is the goal.

It took centuries and generations of navigators, astronomers, geographers, mathematicians, and instrument makers to develop the art and science of celestial navigation to its present level, and the knowledge thus accumulated is a treasure that should be preserved. Moreover, celestial navigation gives us an insight into scientific thinking and creativeness in the pre-electronic age. Last but not least, celestial navigation may be a highly appreciated alternative if a GPS receiver happens to fail.

When I read my first book on navigation many years ago, the chapter on celestial navigation with its fascinating diagrams and formulas immediately caught my particular interest although I was a little intimidated by its complexity at first. As I became more advanced, I realized that celestial navigation is not nearly as difficult as it seems to be at first glance. Studying the literature, I found that many books, although packed with information, are more confusing than enlightening, probably because most of them have been written by experts and for experts. On the other hand, many publications written for beginners are designed like cookbooks, i. e., they contain step-by-step instructions but avoid much of the theory. In my opinion, one can not *really* comprehend celestial navigation and enjoy the beauty of it without knowing the mathematical background.

Since nothing really satisfied my requirements, I decided to write a compact manual for my personal use which had to include the most important definitions, formulas, diagrams, and procedures. As time went by, the project gained its own momentum, the text grew in size, and I started wondering if it might not be of interest to others as well. I contacted a few scientific publishing houses, but they informed me politely that they considered my work as dispensable ("Who is going to read this!"). I had forgotten that scientific publishing houses are run by marketing people, not by scientists. Around the same time, I became interested in the internet, and I quickly found that it is the ideal medium to share one's knowledge with others. Consequently, I set up my own web site to present my e-book to the public.

The style of my work may differ from other books on this subject. This is probably due to my different perspective. When I started the project, I was a newcomer to the world of navigation, but I had a background in natural sciences and in scientific writing. From the very beginning, it has been my goal to provide accurate information in a structured and comprehensible form. The reader may judge whether this attempt has been successful.

More people than I expected are interested in celestial navigation, and I would like to thank my readers for their encouraging comments and suggestions. However, due to the increasing volume of correspondence, I am no longer able to answer individual questions or to provide individual support. Unfortunately, I have still a few other things to do, e. g., working for a living. Nonetheless, I keep working on this publication at leisure, and I am still grateful for suggestions and error reports.

March 15th, 2019

Henning Umland

Web site:

https://celnav.de

Index

Preface

Chapter 1 The Basics of Celestial Navigation

Chapter 2 Altitude Measurement

Chapter 3 Geographic Position and Time

Chapter 4 Finding One's Position (Sight Reduction)

Chapter 5 Finding the Position of an Advancing Vessel

Chapter 6 Determination of Latitude and Longitude, Direct Calculation of Position

Chapter 7 Finding Time *and* Longitude by Lunar Distances

Chapter 8 Rise, Set, Twilight

Chapter 9 Geodetic Aspects of Celestial Navigation

Chapter 10 Spherical Trigonometry

Chapter 11 The Navigational Triangle

Chapter 12 General Formulas for Navigation

Chapter 13 Charts and Plotting Sheets

Chapter 14 Magnetic Declination

Chapter 15 Ephemerides of the Sun

Chapter 16 Navigational Errors

Chapter 17 The Marine Chronometer

Appendix

GNU Free Documentation License

Chapter 1

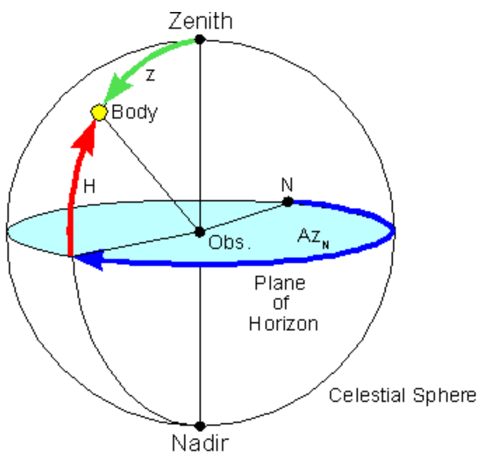
The Basics of Celestial Navigation

Celestial navigation, also called **astronomical navigation**, is the art and science of finding one's own geographic position through astronomical observations, mostly by measuring **altitudes** of celestial bodies – Sun, Moon, planets, or stars.

An observer watching the night sky without knowing anything about geography and astronomy might spontaneously get the impression of being on a horizontal plane located at the center of a huge hollow sphere with the celestial bodies attached to its inner surface. This naive concept of a spherical universe has probably been in existence since the beginning of mankind. Later, astronomers of the antiquity (Ptolemy et al.) developed it to a high degree of perfection. Still today, **spherical astronomy** is fundamental to celestial navigation since the navigator, like the astronomers of old, measures **apparent positions** of bodies in the sky without knowing their actual positions in space.

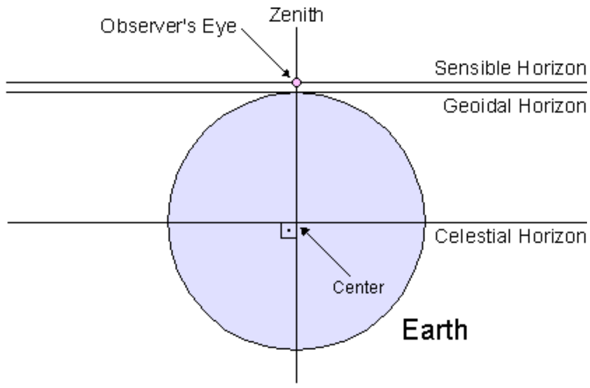
The apparent position of a body in the sky is defined by the **horizon system of coordinates** which is an example of a **spherical coordinate system**. In this system, an imaginary (!) observer is located at the center of the **celestial sphere**, a hollow sphere of infinite diameter, which is divided into two hemispheres by the plane of the **celestial (or geocentric) horizon** (*Fig. 1-1*). The center of the celestial sphere coincides with the center of the Earth which is also assumed to be a sphere. The first coordinate of the observed body is its **geocentric altitude**, **H**. H is the vertical angle between the plane of the celestial horizon and a straight line extending from the center of the celestial sphere to the body. H is measured from 0° through +90° above the horizon and from 0° through -90° below the horizon. The **geocentric zenith distance**, **z**, is the corresponding angular distance between the body and the **zenith**, an imaginary point vertically overhead. The zenith distance is measured from 0° through 180°. H and z are **complementary angles** (H + z = 90°). The point opposite to the zenith on the celestial sphere is called **nadir** (H = -90°, z = 180°). H and z are also arcs of the **vertical circle** going through zenith, nadir, and the observed body. The second coordinate of the body, the geocentric **true azimuth**, $\mathbf{Az_N}$, is the horizontal direction of the body with respect to the **geographic north point** on the celestial horizon, measured clockwise from 0°(N) through 360°. The geographic north point is the point where the vertical half circle going from the zenith through the **celestial north pole** to the nadir intersects the horizon (see chapter 3). The third coordinate of the body, its **distance** from the center of the celestial sphere, is not measured.

Fig. 1-1



In reality, the observer is **not** located on the plane of the celestial horizon but on or above the surface of the Earth. The imaginary horizontal plane passing through the observer's eye is called **sensible (or astronomical) horizon** (*Fig. 1-2*).

Fig. 1-2



The latter merges into the **geoidal horizon**, a plane tangent to the Earth at the observer's position, when the observer's eye is at sea level. The planes of celestial, geoidal, and sensible horizon are parallel to each other and perpendicular to the local **direction of gravity** which defines the positions of zenith and nadir on the celestial sphere.

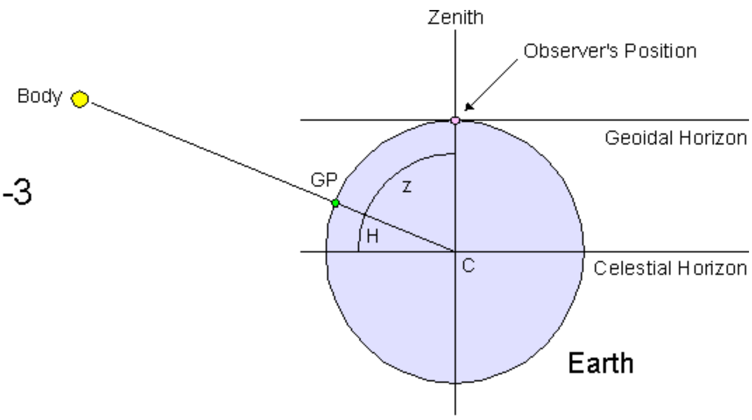
Since sensible and geoidal horizon are relatively close to each other (compared with the radius of the Earth), they can be considered as identical under most practical conditions. None of the above fictitious horizons coincides with the **visible horizon**, the line where the Earth's surface and the sky appear to meet.

Usually, the trigonometric calculations of celestial navigation are based on the geocentric altitudes (or geocentric zenith distances) of bodies. Since it is not possible to measure the geocentric altitude of a body directly, it has to be derived from its altitude with respect to the visible or sensible horizon (**altitude corrections**, chapter 2).

The altitude of a body with respect to the visible **sea horizon** is usually measured with a **marine sextant**. Measuring altitudes with respect to the (invisible) sensible horizon requires an instrument with an **artificial horizon**, e. g., a **theodolite** or a **bubble sextant** (chapter 2). An artificial horizon is a device that indicates a plane perpendicular to the local direction of gravity, for example by means of a **spirit level** or a pendulum.

Geocentric altitude and zenith distance of a celestial body are determined by the distance between the terrestrial observer and the **geographic position of the body**, **GP**. GP is the point where a straight line extending from the center of the Earth, C, to the celestial body intersects the Earth's surface (*Fig. 1-3*).

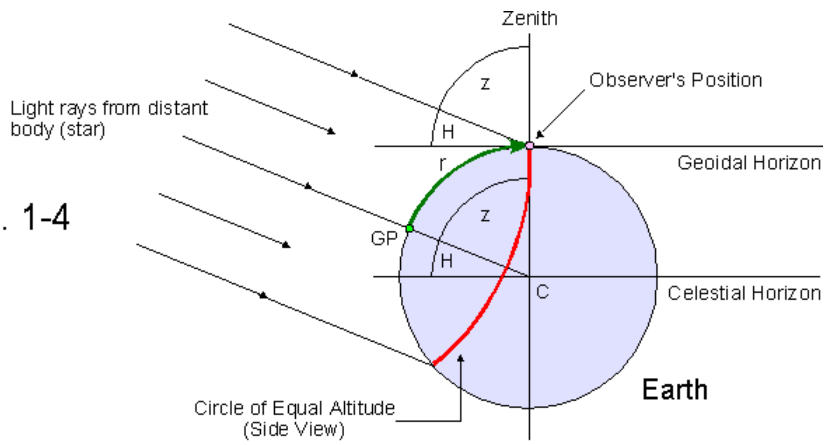
Fig. 1-3



A body is in the zenith (H = 90°, z = 0°) when GP is identical with the observer's position. A terrestrial (surface-bound) observer moving away from GP will experience that the geocentric zenith distanze of the body varies in direct proportion with the distance (measured along the surface) between himself and GP. The geocentric altitude of the body decreases accordingly. The body is on the celestial horizon (H = 0°, z = 90°) when the observer is one quarter of the circumference of the Earth away from GP. As soon as the observer moves farther away from GP, the body disappears below the horizon.

At a given instant, there is an infinite number of terrestrial positions from which the same altitude of a body is measured (unless the body is in the zenith). These positions are equidistant from GP and form a **circle of equal altitude** on the Earth's surface (Fig. 1-4).

Fig. 1-4



The **great circle distance** (chapter 3) of the observer from GP, r, is obtained through the following formula:

$$r[nm] = 60 \cdot z[°] \qquad \text{or} \qquad r[km] = \frac{Perimeter\ of\ Earth\ [km]}{360°} \cdot z[°]$$

One nautical mile (1 nm = 1.852 km) is the great circle distance of one minute of arc on the surface of the Earth. The mean perimeter of the Earth is 40031.6 km.

As shown in Fig. 1-4, light rays originating from a distant object (fixed star) are virtually parallel to each other when they arrive at the Earth. Therefore, the altitude of such an object with respect to the sensible horizon, called **topocentric altitude**, equals its geocentric altitude. In contrast, light rays coming from a relatively close body (Moon, Sun, planets) diverge significantly. This results in a measurable difference between topocentric and geocentric altitude, called **parallax in altitude** (chapter 2). The effect is greatest when observing the Moon, the body closest to the Earth.

The true azimuth of a body depends on the observer's position on the circle of equal altitude and can assume any value between 0° and 360°. Usually, the navigator is not equipped to measure the azimuth of a body with a precision meeting the requirements of celestial navigation (a compass bearing is too imprecise). However, there are methods to calculate the azimuth of a body with respect to the observer's actual or assumed position (chapter 4).

Whenever we measure the altitude or zenith distance of a celestial body, we have already gained some information about our own geographic position because we know we are somewhere on a circle of equal altitude defined by the center, GP (the geographic position of the body), and the great circle radius, r. Of course, the information available so far is still incomplete because we could be anywhere on the circle of equal altitude which comprises an infinite number of possible positions and is therefore also referred to as a **circle of position** (chapter 4).

We extend our thought experiment and observe a second body in addition to the first one. Logically, we are on two circles of equal altitude now. Both circles overlap, intersecting each other at two points on the Earth's surface. One of these two points of intersection is our own position (*Fig. 1-5a*). Theoretically, both circles could be tangent to each other. This case, however, is unlikely. Moreover, it is undesirable and has to be avoided (chapter 16).

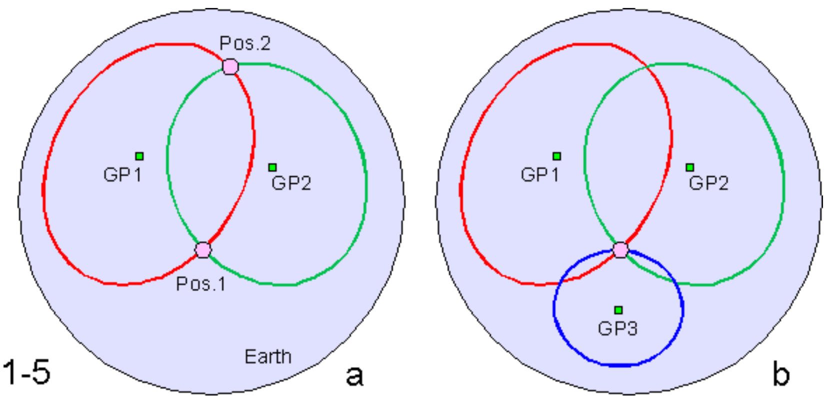


Fig. 1-5

In principle, it is not possible for the observer to know which point of intersection – Pos. 1 or Pos. 2 – is identical with his actual position unless he has additional information, e. g., a fair estimate of his position, or the **compass bearing** (approximate azimuth) of at least one of the bodies. The problem of **ambiguity** does not occur when three bodies are observed because there is only one point where all three circles of equal altitude intersect (*Fig. 1-5b*).

Theoretically, the observer could find his position by plotting the circles of equal altitude on a globe. Indeed, this method has been tried in the past but turned out to be impractical because precise measurements require a very big globe. Plotting circles of equal altitude on a map is possible if their radii are small enough. This usually requires observed altitudes of almost 90°. The method is rarely used since such altitudes are not easy to measure. In most cases, circles of equal altitude have diameters of several thousand nautical miles and do not fit on nautical charts. Further, plotting circles of such dimensions is very difficult due to geometric distortions caused by the respective map projection (chapter 13).

As a rule, the navigator has (and should have!) at least a rough idea of his position. It is therefore not required to plot a complete circle of equal altitude. In most cases only a short arc of the circle in the vicinity of the observer's estimated position is of interest.

If the curvature of the arc is negligible, depending on the radius of the circle and the map scale, it will suffice to plot a straight line (a secant or a tangent of the circle of equal altitude) instead of the arc. Such a line is called a **line of position** or **position line**.

In the 19th century, navigators developed very convenient mathematical and graphic methods for the construction of position lines on nautical charts. The point of intersection of at least two suitable position lines marks the observer's position. These methods, which are considered as the beginning of modern celestial navigation, will be explained in detail later.

In summary, finding one's geographic position by astronomical observations includes three basic steps:

1. Measuring the altitudes or zenith distances of two or more celestial bodies (chapter 2).

2. Finding the geographic position of each body at the instant of its observation by means of the Nautical Almanac or a suitable software almanac (chapter 3).

3. Deriving one's own position from the above data (chapter 4&5).

Chapter 2

Altitude Measurement

In principle, altitudes and zenith distances are equally suitable for navigational calculations. Traditionally, most formulas are based upon altitudes since these are easily measured using the visible sea horizon as a natural reference line. Direct measurement of the zenith distance requires an instrument with an artificial horizon, e. g., a pendulum or spirit level indicating the local direction of gravity (perpendicular to the plane of the sensible horizon) since a visible reference point in the sky does not exist.

Instruments

A **marine sextant** consists of a system of two mirrors and a telescope mounted on a sector-shaped metal frame (usually brass or aluminium alloy). Sextants with a plastic frame are also available. A schematic illustration of the optical components is given in *Fig. 2-1*. The horizon glass is a half-silvered mirror whose plane is perpendicular to the plane of the frame. The index mirror, the plane of which is also perpendicular to the frame, is mounted on the so-called index arm rotatable on a pivot perpendicular to the frame. The optical axis of the telescope is parallel to the frame and passes obliquely through the horizon glass. During an observation, the instrument frame is held upright, and the visible sea horizon is sighted through the telescope and horizon glass. A light ray originating from the observed body is first reflected by the index mirror and then by the back surface of the horizon glass before entering the telescope. By slowly rotating the index mirror on the pivot the superimposed image of the body is aligned with the image of the horizon line. The corresponding altitude, which is twice the angle formed by the planes of horizon glass and index mirror, can be read from the graduated limb, the lower, arc-shaped part of the sextant frame (*Fig. 2-2*). Detailed information on design, usage, and maintenance of sextants is given in [3] (see appendix).

Fig. 2-1

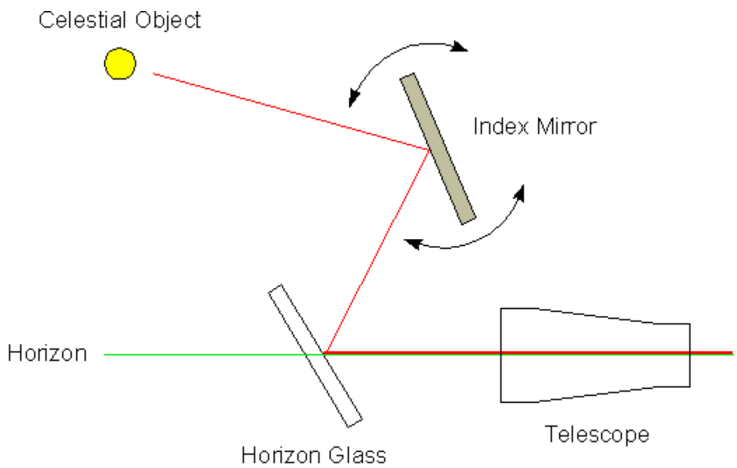


Fig. 2-2

On land, where the horizon is too irregular to be used as a reference line, altitudes have to be measured by means of instruments with an **artificial horizon**.

A **bubble attachment** is a special sextant telescope containing an **internal artificial horizon** in the form of a small spirit level the bubble of which (replacing the visible horizon) is superimposed on the image of the celestial body. Bubble attachments are expensive (almost the price of a sextant) and not very accurate because they require the sextant to be held absolutely still during an observation, which is rather difficult to manage. A sextant equipped with a bubble attachment is referred to as a **bubble sextant**. Special bubble sextants were used for air navigation before electronic navigation systems became standard equipment.

On land, a pan filled with water or, preferably, a more viscous liquid, e. g., glycerol, can be utilized as an **external artificial horizon**. As a result of gravity, the surface of the liquid forms a perfectly horizontal mirror unless distorted by movements or wind. The vertical angular distance between a body and its mirror image, measured with a marine sextant, is twice the altitude of the body. This very accurate method is the perfect choice for exercising celestial navigation in a backyard. *Fig. 2-3* shows a professional form of an external artificial horizon. It consists of a horizontal mirror (polished black glass) attached to a metal frame which is supported by three leg screws. Prior to an observation, the screws have to be adjusted with the aid of one or two detachable high-precision spirit levels until the mirror is exactly horizontal in every direction.

Fig. 2-3

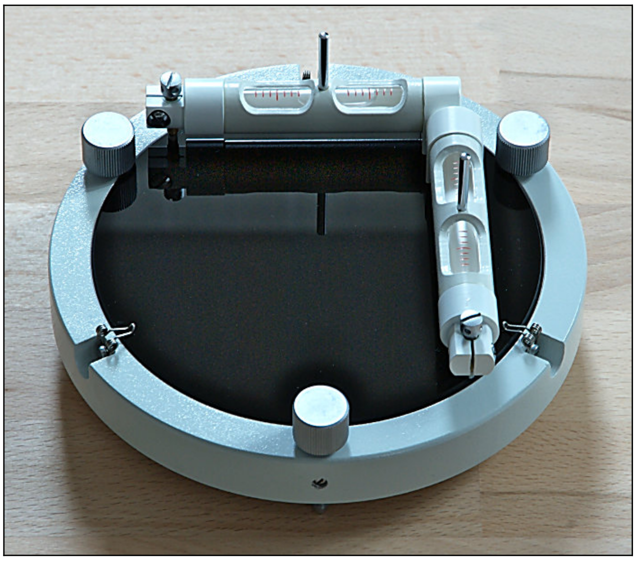


Fig. 2-4

A **theodolite** (*Fig. 2-4*) is basically a telescopic sight which can be rotated about a vertical and a horizontal axis. The angle of elevation (altitude) is read from the graduated vertical circle, the horizontal direction is read from the horizontal circle. The specimen shown above has vernier scales and is accurate to approx. 1'.

The vertical axis of the instrument is aligned with the direction of gravity by means of a spirit level (artificial horizon) before starting the observations. Theodolites are primarily used for surveying, but they are excellent navigation instruments as well. Some models can resolve angles smaller than 0.1' which is not achieved even with the best marine sextants. A theodolite is mounted on a tripod which has to rest on solid ground. Therefore, it is restricted to land navigation. Mechanical theodolites traditionally measure zenith distances. Electronic models can optionally measure altitudes. Most mechanical theodolites measure angles in **gradians** instead of degrees (100 gradians = 90°).

When viewing the Sun through an optical instrument, a proper shade glass must be used, otherwise the retina might suffer permanent damage! The sextant shown in *Fig. 2-2* has two sets of shade glasses (gray filters) attached to the frame which can be inserted into the respective optical path. Detachable shade glasses are available for most theodolites.

Altitude corrections

Any altitude measured with a sextant or theodolite contains errors. Altitude corrections are necessary to eliminate systematic altitude errors and to reduce the topocentric altitude of a body to the geocentric altitude (chapter 1). Altitude corrections do not remove random observation errors.

Index error (IE)

A sextant or theodolite may display a constant error (**index error, IE**) which has to be subtracted from every reading before the latter can be used for further calculations. The error is positive if the angle displayed by the instrument is greater than the actual angle and negative if the displayed angle is smaller. Errors which vary with the displayed angle require the use of an individual correction table if the error can not be eliminated by overhauling the instrument.

$$\text{1st correction:} \qquad H_1 = Hs - IE$$

The **sextant altitude**, **Hs**, is the altitude as indicated by the sextant before any corrections have been applied.

When using an external artificial horizon, H_1 (not Hs!) has to be divided by two.

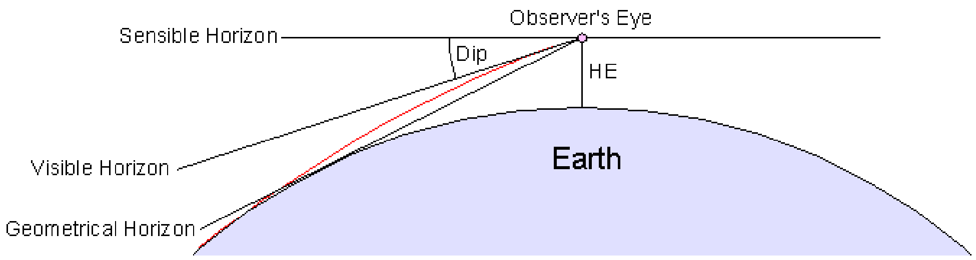
A theodolite measuring the zenith distance, z, requires the following formula to obtain H_1:

$$H_1 = 90° - (z - IE)$$

Dip of horizon

If the Earth's surface were an infinite plane, visible and sensible horizon would be identical. In reality, the visible sea horizon appears several arcminutes below the sensible horizon which is the result of two contrary effects, the curvature of the Earth's surface and atmospheric refraction. The **geometrical horizon** is a flat cone formed by an infinite number of straight lines tangent to the Earth and converging at the observer's eye. Since atmospheric refraction bends light rays passing along the Earth's surface toward the Earth, all points on the geometric horizon appear to be elevated, and thus form the **visible horizon**. Visible and geometrical horizon would be the same if the Earth had no atmosphere (*Fig. 2-5*).

Fig. 2-5



The vertical angular distance of the sensible horizon from the visible horizon is called **dip** (**of horizon**) and is a function of the **height of eye**, **HE**, the vertical distance of the observer's eye from the sea surface (the distance between sensible and geoidal horizon):

$$Dip['] \approx 1.76 \cdot \sqrt{HE[m]} \approx 0.97 \cdot \sqrt{HE[ft]}$$

The above formula is empirical and includes the effects of the curvature of the Earth's surface and of atmospheric refraction*. The influence of the height of eye should not be underestimated. Increasing HE from 2 m to 4 m, for example, causes the dip to change by approx. 1 arcminute.

*At sea, the dip of horizon can be obtained directly by measuring the angular distance between the visible horizon in front of the observer and behind the observer through the zenith. Subtracting 180° from the angle thus measured and dividing the resulting angle by two yields the dip of horizon. This very accurate method can not be accomplished with a sextant but requires a special instrument (prismatic reflecting circle) which is able to measure angles greater than 180°.

The correction for the dip of horizon has to be omitted (Dip = 0) if any kind of an artificial horizon is used since the latter is solely controlled by gravity and thus indicates the plane of the sensible horizon (perpendicular to the vector of gravity).

$$\text{2nd correction:} \qquad H_2 = H_1 - Dip$$

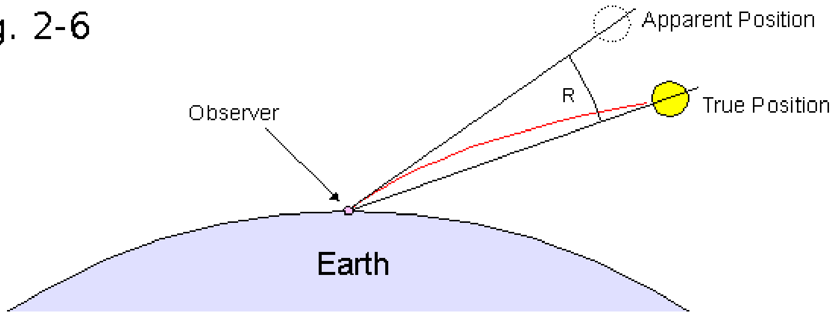
The altitude obtained after applying corrections for index error and dip is also referred to as **apparent altitude**, **Ha**.

$$Ha = H_2$$

Atmospheric refraction

A light ray coming from a celestial body is slightly deflected toward the Earth when passing obliquely through the atmosphere. This phenomenon is called **refraction**, and occurs always when light enters matter of different density at an angle smaller than 90°. Since the eye is not able to detect the curvature of the light ray, the body appears to be on a straight line tangent to the light ray at the observer's eye, and thus appears to be higher in the sky. R is the vertical angular distance between apparent and true position of the body measured at the observer's eye (*Fig. 2-6)*.

Fig. 2-6



Atmospheric refraction is a function of Ha (= H_2). Atmospheric **standard refraction**, $\mathbf{R_0}$, is zero at 90° altitude and increases progressively to approx. 34' as the apparent altitude approaches 0°:

Ha [°]	0	1	2	5	10	20	30	40	50	60	70	80	90
$\mathbf{R_0}$ **[']**	~34	~24	~18	9.9	5.3	2.6	1.7	1.2	0.8	0.6	0.4	0.2	0.0

There are several formulas to calculate R_0. *Smart's* formula yields very accurate results from 15° through 90° apparent altitude [2,9]:

$$R_0['] = \frac{0.96474}{\tan H_2} - \frac{0.00113}{\tan^3 H_2}$$

The constants used here are not exactly those given by *Smart* but have been slightly modified to match the results (within $\pm 10^{-4}$ ') obtained with *Saastamoinen's* highly accurate formula (see below) under the following conditions: T = 283.15 K, p = 1010 hPa, relative air humidity = 75%. For the purpose of marine navigation, *Smart's* formula can be used with apparent altitudes down to as low as 7° where the error reaches approx. -0.1'. Below 7°, the error increases rapidly and results become useless.

For altitudes between 0° and 15°, the following formula is suggested [10]. H_2 is measured in degrees:

$$R_0['] = \frac{34.133 + 4.197 \cdot H_2 + 0.00428 \cdot H_2^2}{1 + 0.505 \cdot H_2 + 0.0845 \cdot H_2^2}$$

A rather simple formula including the whole range of apparent altitudes from 0° through 90° was found by *Bennett*:

$$R_0[\,'\,] \;=\; \frac{1}{\tan\left(H_2[°] + \dfrac{7.31}{H_2[°] + 4.4}\right)}$$

Bennett's formula is sufficiently accurate for marine navigation. The maximum systematic error, measured at approx. 12° apparent altitude, is smaller than 0.1′ [2].

Atmospheric refraction is influenced by atmospheric pressure, p, air temperature, T, and, to a much lesser degree, by relative air humidity. Therefore the standard refraction, R_0, has to be multiplied with a correction factor, f, to obtain the actual refraction for a given combination of pressure and temperature.

$$R \;=\; f\cdot R_0 \qquad\qquad f \;=\; \frac{p[hPa]}{1010}\cdot\frac{283.15}{T[K]} \;=\; \frac{p[hPa]}{1010}\cdot\frac{283.15}{273.15+T[°C]} \;=\; \frac{p[\text{in. }Hg]}{29.83}\cdot\frac{510}{460+T[°F]}$$

Pressure and temperature are measured at the observer's position at the time of observation. The pressure must not be reduced to sea level when observing from an elevated position. By definition, **standard conditions** (f = 1) for marine navigation are **p = 1010 hPa** (29.83 in Hg) and **T = 283.15 K** (10°C, 50°F)*. The influence of air humidity on atmospheric refraction is usually ignored. Even the correction for atmospheric pressure and temperature is sometimes omitted (f = 1) since the resulting error is often tolerable.

* These are different from standard conditions commonly used in chemistry and physics (p = 1013.25 hPa, T = 273.15 K).

Saastamoinen's refraction formula [10] includes corrections for temperature, atmospheric pressure, and relative air humidity:

$$R[\,''\,] \;=\; \frac{16.271\cdot Q\cdot\left(1+\dfrac{0.0000394\cdot Q}{\tan^2 H_2}\right)}{\tan H_2} \;-\; 0.0000749\cdot p[hPa]\cdot\left(\frac{1}{\tan H_2}-\frac{1}{\tan^3 H_2}\right)$$

$$Q \;=\; \frac{p[hPa]-0.156\cdot p_W[hPa]}{T[K]}$$

p_W is the partial pressure of water vapor in the atmosphere [10]. It is a function of relative air humidity, AH, and absolute temperature, T:

$$p_W[hPa] \;=\; 0.01\cdot AH[\%]\cdot\left(\frac{T[K]}{247.1}\right)^{\mathbf{18.36}}$$

Saastamoinen's formula assumes the observer to be at sea level and is most accurate at apparent altitudes greater than 20°. Note that in contrast to the aforementioned refraction formulas, results are measured in arcseconds here.

The common refraction formulas refer to a fictitious standard atmosphere with an average density gradient. The actual refraction may differ significantly from the calculated one if anomalous atmospheric conditions are present (temperature inversion, mirage effects, etc.). The influence of atmospheric anomalies increases strongly with decreasing altitude. Particularly refraction below ca. 5° apparent altitude may become erratic, and calculated or tabulated values in this range should not be blindly relied on. It should be mentioned that the dip of horizon, too, is influenced by atmospheric refraction and may become unpredictable under certain meteorological conditions.

$$\text{3rd correction:}\qquad H_3 \;=\; H_2 - R \;\approx\; H_2 - R_0$$

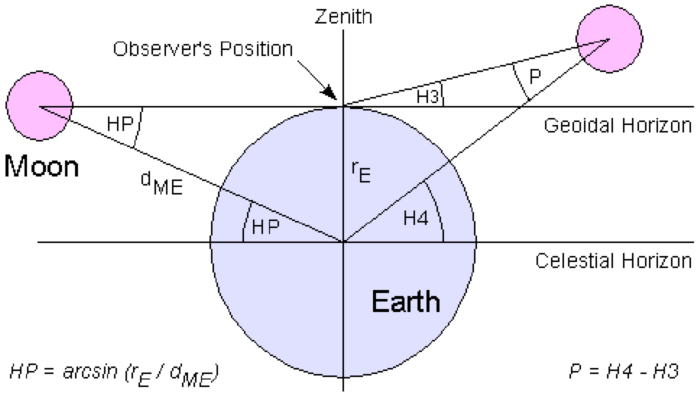
H_3 represents the topocentric altitude of the body, i. e., the altitude with respect to the sensible horizon.

Parallax

The trigonometric calculations of celestial navigation are based upon geocentric altitudes. *Fig. 2-7* illustrates that the geocentric altitude of an object, H_4, is always greater than its topocentric altitude, H_3. The difference H_4-H_3 is called **parallax in altitude**, **P**. P decreases as the distance between object and Earth increases. Accordingly, the effect is greatest when observing the Moon since the latter is the object nearest to the Earth. At the other extreme, P is too small to be measured when observing fixed stars (see chapter 1, *Fig. 1-4*). To be precise, the observed parallax refers to the sensible, not to the geoidal horizon.

However, since the height of eye is by several magnitudes smaller than the radius of the Earth, the resulting error is usually not significant.

Fig. 2-7



The parallax (in altitude) of a body being on the geoidal horizon is called **horizontal parallax**, **HP** (*Fig. 2-7*). The horizontal parallax of the Sun is approx. 0.15'. Current values for the HP of the Moon (approx. 1°) and the navigational planets are given in the **Nautical Almanac** [12] and similar publications, e.g., [13]. Tabulated values for HP refer to the equatorial radius of the Earth (equatorial horizontal parallax, see chapter 9). P is a function of topocentric altitude and horizontal parallax of a body.* It has to be added to H_3.

$$P = \arcsin\left(\sin HP \cdot \cos H_3\right) \approx HP \cdot \cos H_3$$

*To be exact, the parallax formula shown above is rigorous for the observation of the center of a body only. When observing the lower or upper limb, there is a small error caused by the curvature of the body's surface which is usually negligible. The rigorous formula for the observation of the respective limb is:

Lower limb: $$P = \arcsin\left[\sin HP \cdot \left(\cos H_3 + k\right)\right] - \arcsin\left(k \cdot \sin HP\right) \approx HP \cdot \cos H_3$$

Upper limb: $$P = \arcsin\left[\sin HP \cdot \left(\cos H_3 - k\right)\right] + \arcsin\left(k \cdot \sin HP\right) \approx HP \cdot \cos H_3$$

The factor k is the ratio of the radius of the observed body to the equatorial radius of the Earth (r_{Earth} = 6378 km).

4th correction: $$H_4 = H_3 + P$$

An additional correction for the oblateness of the Earth is recommended when observing the Moon (ΔP, see p. 2-8).

H_4 represents the geocentric altitude of the body, the altitude with respect to the celestial horizon.

Semidiameter

When observing Sun or Moon with a marine sextant or theodolite, it is not possible to locate the center of the body precisely. It is therefore common practice to measure the altitude of the **upper** or **lower limb** of the body and add or subtract the **apparent semidiameter**, **SD**. The latter is the angular distance of the respective limb from the center of the body (*Fig. 2-8*).

Fig. 2-8

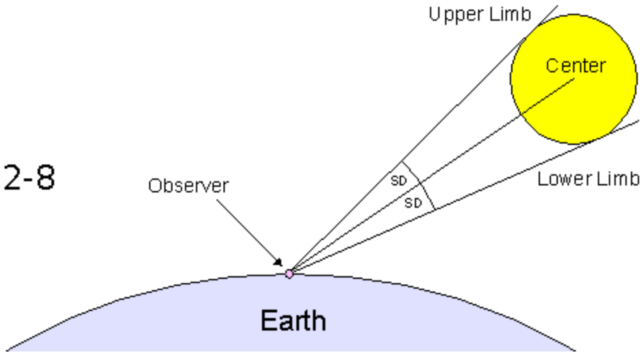


Fig. 2-8 illustrates the topocentric semidiameter of a body. However, we need the **geocentric** SD, the SD measured by a fictitious observer at the center the Earth, because H_4 is measured at the center the Earth (see *Fig. 2-4*). The geocentric semidiameters of Sun and Moon are given on the daily pages of the **Nautical Almanac** [12].

The geocentric SD of a body can be calculated from its tabulated horizontal parallax. This is of particular interest when observing the Moon.

$$SD_{geocentric} \; = \; \arcsin\left(k \cdot \sin HP\right) \; \approx \; k \cdot HP \qquad\qquad k_{Moon} \; = \; \frac{r_{Moon}}{r_{Earth}} \; = \; 0.2725$$

Although the semidiameters of the navigational planets are not quite negligible (the SD of Venus can increase to 0.5'), the apparent centers of these bodies are usually observed, and no correction for SD is applied. With a strong telescope, however, the limbs of the brightest planets can be observed. In this case the correction for semidiameter should be applied. Semidiameters of stars are much too small to be measured (SD = 0).

$$\text{5th correction:} \qquad H_5 \; = \; H_4 \pm SD_{geocentric}$$

(lower limb: add SD, upper limb: subtract SD)

When using a bubble sextant, we observe the center of the body and skip the correction for semidiameter (SD = 0).

The altitude obtained after applying the above corrections is called observed altitude, Ho. The latter represents the geocentric altitude of the center of the observed body.

$$Ho \; = \; H_5$$

Combined corrections for semidiameter and parallax (particularly recommended for observations of the Moon)

H_3 can be reduced to the observed altitude in one step. In spite of its simplicity, the following formula is rigorous. It includes the corrections for semidiameter **and** parallax in altitude (see Addendum at the end of this chapter):

$$Ho \; = \; H_3 + \arcsin\left[\sin HP \cdot \left(\cos H_3 \pm k\right)\right] \; \approx \; H_3 + HP \cdot \left(\cos H_3 \pm k\right)$$

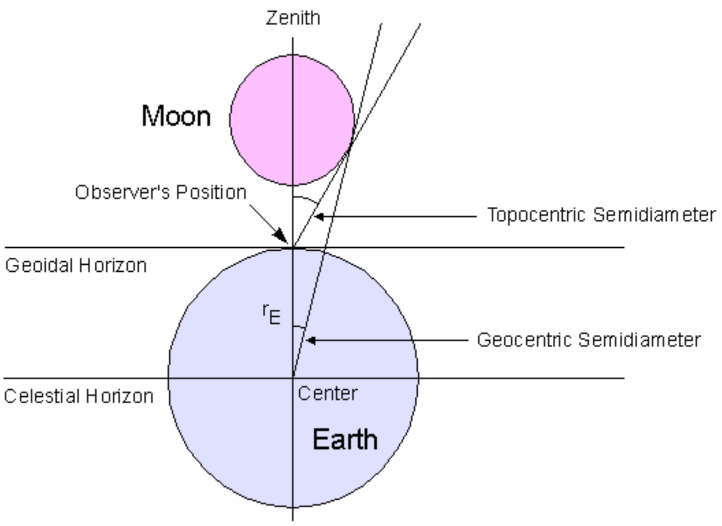
(lower limb: add k, upper limb: subtract k)

Alternative procedure for semidiameter and parallax

Correcting for semidiameter **before** correcting for parallax is also possible. In this case, however, we have to calculate with the **topocentric** semidiameter, the semidiameter of the respective body as seen from the observer's position on the surface of the Earth (see *Fig. 2-8*). With the exception of the Moon, the body nearest to the Earth, there is no significant difference between topocentric and geocentric semidiameter.

The topocentric SD of the Moon is only marginally greater than the geocentric SD when the Moon is on the sensible (geoidal) horizon but increases measurably as the altitude increases because of the decreasing distance between observer and Moon. The distance is smallest (decreased by about the radius of the Earth) when the Moon is in the zenith. As a result, the topocentric SD of the Moon being in the zenith is approximately 0.3' greater than the geocentric SD. This phenomenon is called **augmentation** (Fig. 2-9).

Fig. 2-9



The rigorous formula for the topocentric (augmented) semidiameter of the Moon is:

$$SD_{topocentric} = \arctan \frac{k}{\left[\sqrt{\dfrac{1}{\sin^2 HP} - \left(\cos H_3 \pm k\right)^2}\right] - \sin H_3}$$

(observation of lower limb: add k, observation of upper limb: subtract k)

The approximate topocentric semidiameter of the Moon can be calculated with a simpler formula given by *Meeus* [2]. It refers to the center of the Moon but is still accurate enough for the purpose of navigation (error < 1") when applied to the altitude of the upper or lower limb, respectively:

$$SD_{topocentric} \approx k \cdot HP \cdot \left(1 + \sin HP \cdot \sin H_3\right)$$

A very similar formula was proposed by *Stark* [14]:

$$SD_{topocentric} \approx \frac{k \cdot HP}{1 - \sin HP \cdot \sin H_3}$$

Thus, the alternative fourth correction is:

$$\text{4th correction (alt.):} \qquad H_{4,alt.} = H_3 \pm SD_{topocentric}$$

(lower limb: add SD, upper limb: subtract SD)

$H_{4,alt}$ represents the topocentric altitude of the center of the Moon.

Using the parallax formula explained above, we calculate P_{alt} from $H_{4,alt}$:

$$P_{alt} = \arcsin\left(\sin HP \cdot \cos H_{4,alt}\right) \approx HP \cdot \cos H_{4,alt}$$

Thus, the alternative fifth correction is:

$$\text{5th correction (alternative):} \qquad H_{5,alt} = H_{4,alt} + P_{alt}$$

$$Ho = H_{5,alt}$$

Since the geocentric SD is easier to calculate than the topocentric SD, it is usually more convenient to correct for the semidiameter in the last place or, better, to use the combined correction for parallax and semidiameter unless one has to know the augmented SD of the Moon for special reasons.

The topocentric semidiameter of the Moon can also be calculated from the observed altitude (the geocentric altitude of the center of the Moon), Ho:

$$SD_{topocentric} = \arcsin \frac{k}{\sqrt{1 + \dfrac{1}{\sin^2 HP} - 2 \cdot \dfrac{\sin Ho}{\sin HP}}}$$

Instead of Ho, the computed altitude, Hc, can be used (see chapter 4).

Correction for the oblateness of the Earth

The above formulas for parallax and semidiameter are rigorous for spherical bodies. In fact, the Earth is not exactly a sphere but rather resembles an **oblate spheroid**, a sphere flattened at the poles (chapter 9). In most cases, the navigator will not notice the difference. However, when observing the Moon, the flattening of the Earth may cause a small but measurable error (up to ±0.2') in the parallax, depending on the observer's position. Therefore, a small correction, ΔP, should be added to P if higher accuracy is required [12]. When using the combined formula for semidiameter and parallax, ΔP is added to Ho.

$$\Delta P \approx f \cdot HP \cdot \left(\sin(2 \cdot Lat) \cdot \cos Az_N \cdot \sin H^* - \sin^2 Lat \cdot \cos H^* \right) \qquad f = \frac{1}{298.257}$$

* Replace H with H3 or H4,alt, respectively.

$$P_{improved} = P + \Delta P$$

Lat is the observer's estimated or assumed latitude (chapter 4). Az_N, the true azimuth of the Moon, is either measured with a magnetic compass (compass bearing) or calculated using the azimuth formulas given in chapter 4.

Phase correction (Venus and Mars)

Since Venus and Mars show phases similar to the Moon, their apparent center may differ somewhat from the actual center. The coordinates of both planets tabulated in the **Nautical Almanac** [12] include the phase correction, i. e., they refer to the apparent center. The phase correction for Jupiter and Saturn is too small to be significant.

In contrast, coordinates calculated with **Interactive Computer Ephemeris** refer to the actual center. In this case, the upper or lower limb of the respective planet should be observed if the magnification of the telescope is sufficient, and the correction for semidiameter should be applied.

Altitude correction tables

The **Nautical Almanac** provides sextant altitude correction tables for Sun, planets, stars (pages A2 – A4), and the Moon (pages xxxiv – xxxv), which can be used instead of the above formulas if small errors (< 1') are tolerable (among other things, the tables cause additional rounding errors).

Other corrections

Sextants with an artificial horizon can exhibit additional errors caused by acceleration forces acting on the bubble or pendulum and preventing it from aligning itself with the direction of gravity. Such acceleration forces can be accidental (vessel movements) or systematic (coriolis force). The coriolis force is important to air navigation (high speed!) and requires a special correction formula.

In the vicinity of mountains, ore deposits, and other local irregularities of the Earth's crust, the vector of gravity may slightly differ from the normal to the reference ellipsoid, resulting in altitude errors that are difficult to predict (**local deflection of the vertical**, chapter 9). Thus, the **astronomical position** of an observer (resulting from astronomical observations) may be slightly different from his geographic (geodetic) position with respect to a **reference ellipsoid** (GPS position). The difference is usually small at sea and may be ignored there. On land, particularly in the vicinity of mountain ranges, position errors of up to 50 arcseconds (Alps) or even 100 arcseconds (Himalaya) have been found. Thus, surveying by astronomical observations requires local corrections for latitude and longitude, depending on the respective reference ellipsoid.

Addendum

The following is a detailed explanation of the formula used to reduce the topocentric altitude of the lower limb of the Moon to the geocentric altitude of the Moon's center in one step (Fig. 2-10). H is the altitude corrected for index error, dip, and atmospheric refraction (= H_3). The topocentric altitude of the upper limb is reduced in a similar way (see below).

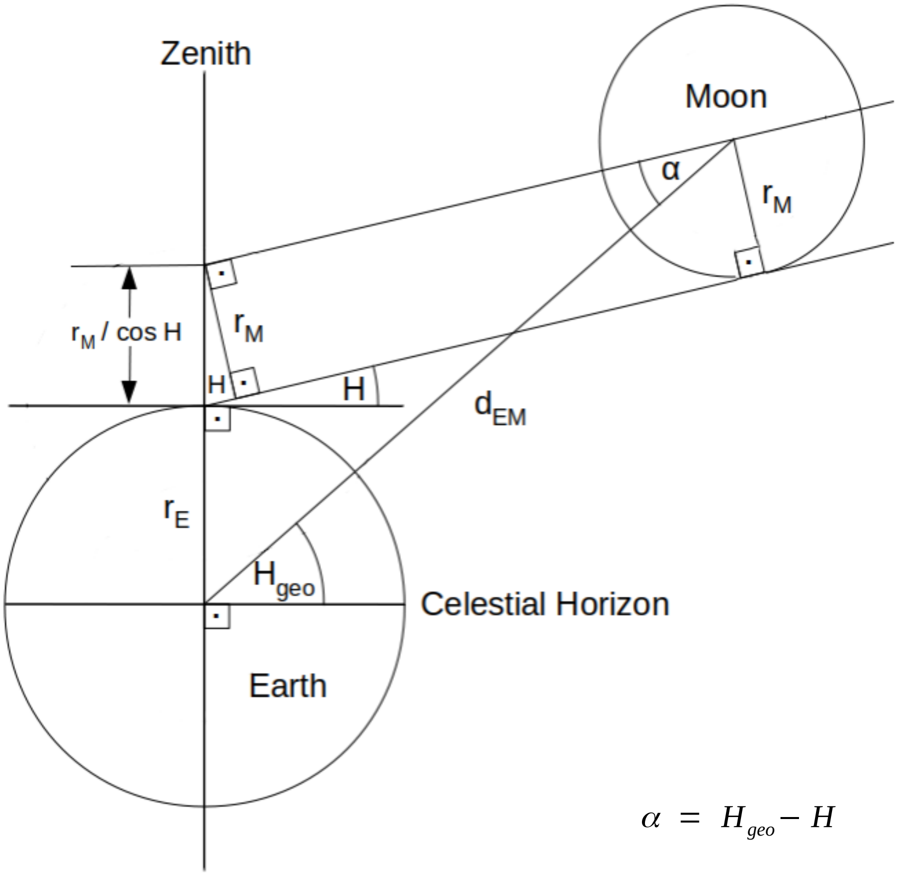


Fig. 2-10

First, we apply the law of sines for plane triangles:

$$\frac{\sin(H_{geo} - H)}{r_E + \dfrac{r_M}{\cos H}} \;=\; \frac{\sin(H + 90°)}{d_{EM}} \;=\; \frac{\cos H}{d_{EM}}$$

$$\sin(H_{geo} - H) \;=\; \frac{\left(r_E + \dfrac{r_M}{\cos H}\right) \cdot \cos H}{d_{EM}} \;=\; \frac{r_E \cdot \cos H}{d_{EM}} + \frac{r_M}{d_{EM}}$$

With $\sin HP = \dfrac{r_E}{d_{EM}}$, $k = \dfrac{r_M}{r_E}$, and $k \cdot \sin HP = \dfrac{r_M}{d_{EM}}$, we get

$$\sin(H_{geo} - H) \;=\; \sin HP \cdot (\cos H + k)$$

$$H_{geo} \approx\; H + HP \cdot (\cos H + k)$$

When observing the upper limb of the Moon, we get

$$\frac{\sin(H_{geo} - H)}{r_E - \dfrac{r_M}{\cos H}} \;=\; \frac{\cos H}{d_{EM}}$$ and, accordingly, $$\sin(H_{geo} - H) \;=\; \sin HP \cdot (\cos H - k)$$

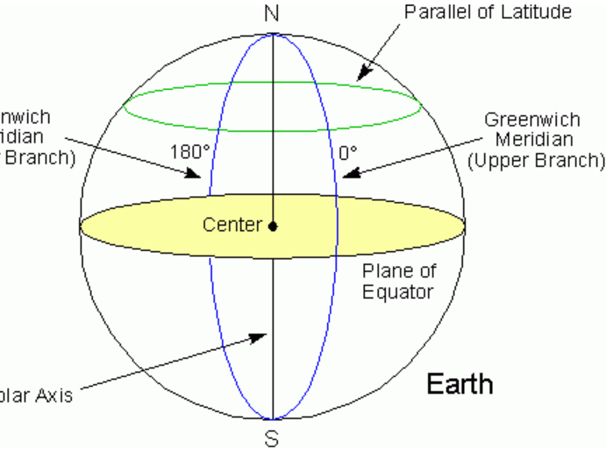
Chapter 3

Geographic Position and Time

Geographic and Astronomical Terms

In celestial navigation, the Earth is regarded as a sphere. This is an approximation only, but the errors caused by the flattening of the Earth are usually negligible (see chapter 9). A circle on the surface of the Earth whose plane passes through the Earth's center is referred to as a **great circle**. In contrast, a **small circle** is a circle on the surface of the Earth the plane of which does **not** pass through the Earth's center. The **equator** is the only great circle the plane of which is perpendicular to the **polar axis**, the rotation axis of the Earth. Further, the equator is the only **parallel of latitude** being a great circle. All other parallels of latitude are small circles whose planes are parallel to the plane of the equator. A **meridian** is a great circle going through the **geographic poles**, the two points where the polar axis intersects the Earth's surface. The **upper branch** of a meridian is the half from pole to pole passing through a given point, for example the observer's position. The **lower branch** is the opposite half. The meridian passing through the observer's position is called **local meridian**. The **Greenwich meridian**, the meridian passing through the center of the transit instrument at the **Royal Greenwich Observatory**, was adopted as the **prime meridian** at the **International Meridian Conference** in 1884. Its upper branch (0°) is the reference for measuring longitudes (0°... +180° to the east and 0°...–180° to the west), its lower branch (±180°) is the basis for the **International Date Line** (*Fig. 3-1*).

Fig. 3-1

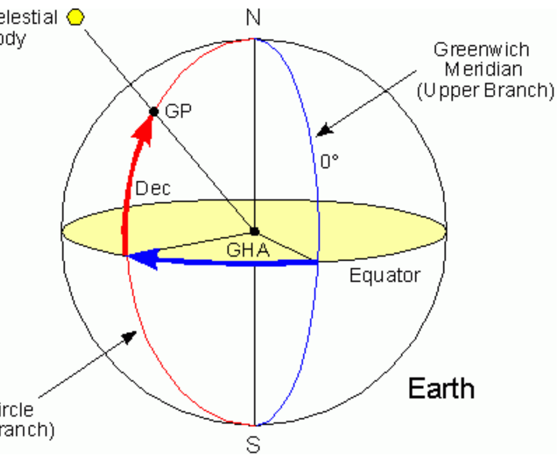


Each point of the Earth's surface has an imaginary counterpart on the surface of the celestial sphere obtained by central projection. The projected image of the observer's position, for example, is the **zenith**. Accordingly, there are two **celestial poles**, the **celestial equator**, **celestial meridians**, etc. The local celestial meridian is also a **vertical circle**, i. e., a great circle on the celestial sphere passing through the observer's zenith and nadir. Passing throught the celestial poles, the local celestial meridian marks the **north point** and the **south point** on the horizon. The vertical circle perpendicular to the local meridian, called **prime vertical**, marks the **west point** and **east point** on the horizon.

The Equatorial System of Coordinates

The **geographic position of a celestial body**, **GP**, is defined by the **equatorial system of coordinates**, a spherical coordinate system the origin of which is at the center of the Earth (*Fig. 3-2*).

Fig. 3-2



The **Greenwich hour angle** of a body, **GHA**, is the angular distance of the upper branch of the meridian passing through GP from the upper branch of the Greenwich meridian (Lon = 0°), measured **westward from 0° through 360°**. The meridian going through GP (as well as its projection on the celestial sphere) is called **hour circle**.

The **Declination** of a body, **Dec**, is the angular distance of GP from the plane of the equator, measured northward through +90° or southward through –90°. GHA and Dec are **geocentric coordinates** (measured at the center of the Earth). Although widely used, the term "geographic position" is misleading when applied to a celestial body since it actually describes a geocentric position in this case (see chapter 9). GHA and Dec are equivalent to **geocentric longitude and latitude** of a position with the exception that **longitudes are measured westward from the Greenwich meridian through −180° and eastward through +180°**.

Since the Greenwich meridian rotates with the Earth from west to east, whereas each hour circle remains linked with the almost stationary position of the respective body in the sky, the Greenwich hour angles of all celestial bodies increase by approx. 15° per hour (360° in 24 hours).

In contrast to stars (15° 2.46' /h), the GHAs of Sun, Moon, and planets increase at slightly different (and variable) rates. This is caused by the revolution of the planets (including the Earth) around the Sun and by the revolution of the Moon around the Earth, resulting in additional apparent motions of these bodies in the sky. For several applications it is useful to measure the angular distance between the hour circle of a celestial body and the hour circle of a reference point in the sky instead of the Greenwich meridian because the angle thus obtained is independent of the Earth's rotation. The **sidereal hour angle**, **SHA**, of a given body is the angular distance of its hour circle (upper branch) from the hour circle (upper branch) of the **first point of Aries** (also called **vernal equinox**, see below), measured **westward** from 0° through 360°. Thus, the GHA of a body is the sum of its sidereal hour angle and the GHA of the first point of Aries, $\mathbf{GHA_{Aries}}$:

$$GHA = SHA + GHA_{Aries}$$

(Subtract 360° if the resulting angle is greater than 360°.)

The angular distance of a celestial body **eastward** from the hour circle of the vernal equinox, measured in time units (24h = 360°), is called **right ascension**, **RA**. The latter is mostly used by astronomers whereas navigators prefer sidereal hour angles.

$$RA\,[h] = 24\,h - \frac{SHA\,[°]}{15} \qquad \Leftrightarrow \qquad SHA\,[°] = 360° - 15 \cdot RA\,[h]$$

Fig. 3-3 illustrates the various hour angles on the plane of the equator as seen from the celestial north pole (time diagram).

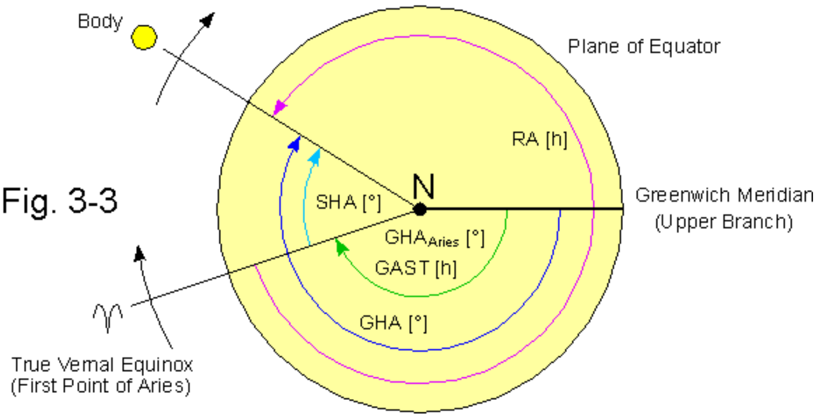


Fig. 3-3

Declinations are not affected by the rotation of the Earth. The declinations of Sun and planets change primarily due to the **obliquity of the ecliptic**, the inclination of the Earth's equator to the **ecliptic**. The latter is the **orbital plane of the Earth** and forms a great circle on the celestial sphere. The declination of the Sun, for example, varies periodically between ca. +23.5° (**summer solstice**) and ca. -23.5° (**winter solstice**) as shown in *Fig. 3-4*.

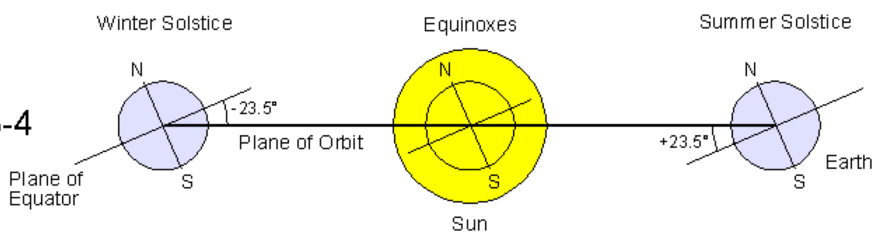


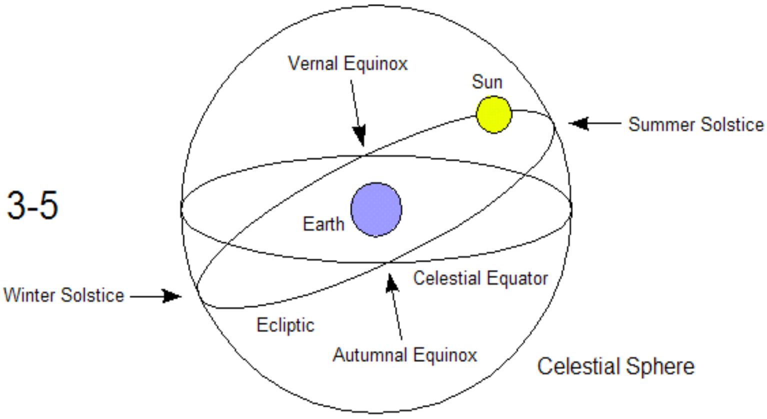
Fig. 3-4

The two points on the celestial sphere where the great circles of ecliptic and celestial equator intersect are called **equinoxes**. The term equinox is also used for the instants at which the apparent Sun, moving westward along the ecliptic during the course of a year, crosses the celestial equator, approximately on March 21 and on September 23. There is a **vernal equinox** (**first point of Aries, vernal point**) and an **autumnal equinox**.

The former is the reference point for measuring sidereal hour angles (*Fig. 3-5*). At the instant of an equinox (Dec ≈ 0°)*, day and night have roughly (!) the same length (12 h each), regardless of the observer's position (Lat. *aequae noctes* = equal nights).

*To be more precise, the equinoxes are defined as the instants at which the **ecliptic longitude** (λ) of the apparent Sun is either 0° (vernal equinox) or 180° (autumnal equinox) [10]. The actual declination of the Sun at such an instant may slightly differ from 0° since the Earth is not always exactly on the mean orbital plane (perturbations).

Fig. 3-5



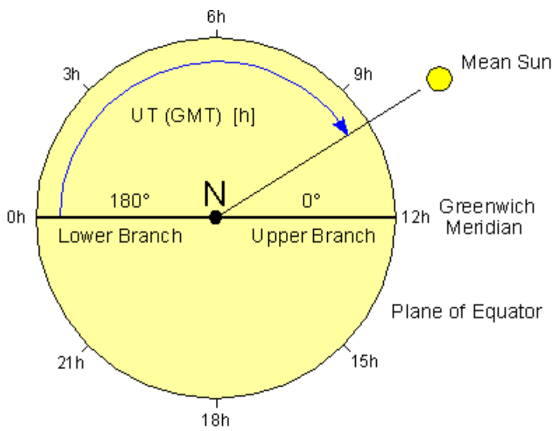
The declinations of the planets and the Moon are also influenced by the inclinations of their own orbits to the ecliptic. The plane of the Moon's orbit, for example, is inclined to the ecliptic by approx. 5° and makes a tumbling movement with a period of 18.6 years (Saros cycle). As a result, the declination of the Moon varies between approx. -28.5° and +28.5° at the beginning and at the end of the Saros cycle, and between approx. -18.5° and +18.5° in the middle of the Saros cycle.

Further, sidereal hour angles and declinations of all bodies change slowly due to the influence of the **precession** of the Earth's polar axis. Precession is a slow, tumbling movement of the polar axis along the surface of an imaginary double cone. One revolution lasts about 26000 years (Platonic year). As a result, the equinoxes move westward along the celestial equator at a rate of approx. 50" per year. Thus, the sidereal hour angle of each star decreases at about the same rate. In addition, there is a small elliptical oscillation of the polar axis with a period of 18.6 years (compare with the Saros cycle), called **nutation**, which causes the equinoxes to travel along the celestial equator at a periodically changing rate. Thus we have to distinguish between the ficticious **mean equinox of date** and the **true** (current) **equinox of date** (see time measurement). Accordingly, the declination of each body oscillates (nutation in obliquity). The same applies to the rate of change of the sidereal hour angle and right ascension of each body (see below). Even stars are not fixed in space but move individually, resulting in a slow drift of their respective declination and right ascension (**proper motion**). Finally, the apparent positions of bodies are influenced by other factors, e. g., the finite speed of light (**light time, aberration**), and **annual parallax**, the latter being caused by the Earth orbiting around the Sun [16]. The accurate prediction of geographic positions of celestial bodies requires complicated algorithms. The calculation of low-precision **ephemerides** of the Sun (sufficient for marine navigation) is described in chapter 15.

Time Measurement in Navigation and Astronomy

Since the Greenwich hour angle of any celestial body changes rapidly, many tasks of celestial navigation require accurate time measurement, and the instant of each observation should be measured to the second if possible. This is usually done by means of a chronometer and a stopwatch (chapter 17). The effects of time errors are discussed in chapter 16. On the other hand, the Earth's rotation with respect to celestial bodies provides an important basis for astronomical time measurement. Coordinates tabulated in the **Nautical Almanac** refer to **Universal Time**, **UT**. UT has replaced **Greenwich Mean Time**, **GMT**, the traditional basis for civil time keeping. Conceptually, UT (like GMT) is the hour angle of the fictitious **mean Sun**, expressed in hours (24h = 360°) with respect to the **lower branch** of the Greenwich meridian (mean solar time, *Fig. 3-6*).

Fig. 3-6



UT is calculated with the following formula which refers to the upper branch of the Greewich meridian:

$$UT[h] = \frac{GHA_{Mean\,Sun}[°]}{15} + 12$$

(If UT is greater than 24 h, subtract 24 hours.)

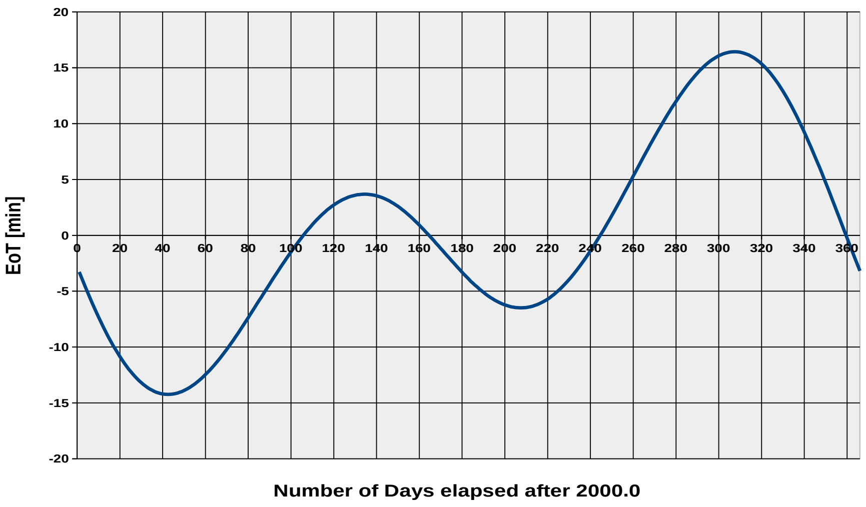
By definition, the GHA of the mean Sun increases by exactly 15° per hour, completing a 360° cycle in 24 hours. The unit for UT is 1 solar day, the time interval between two consecutive meridian transits of the mean Sun.

The rate of change of the GHA of the **apparent (observable) Sun** varies periodically and is sometimes slightly greater, sometimes slightly smaller than 15°/h during the course of a year. This behavior is caused by the eccentricity of the Earth's orbit and by the obliquity of the ecliptic. The time measured by the hour angle of the apparent Sun with respect to the lower branch of the Greenwich meridian is called **Greenwich Apparent Time**, **GAT**. A **sundial** located at the Greenwich meridian would indicate GAT. The difference between GAT and UT at a given instant is called **equation of time, EoT**:

$$EoT = GAT - UT$$

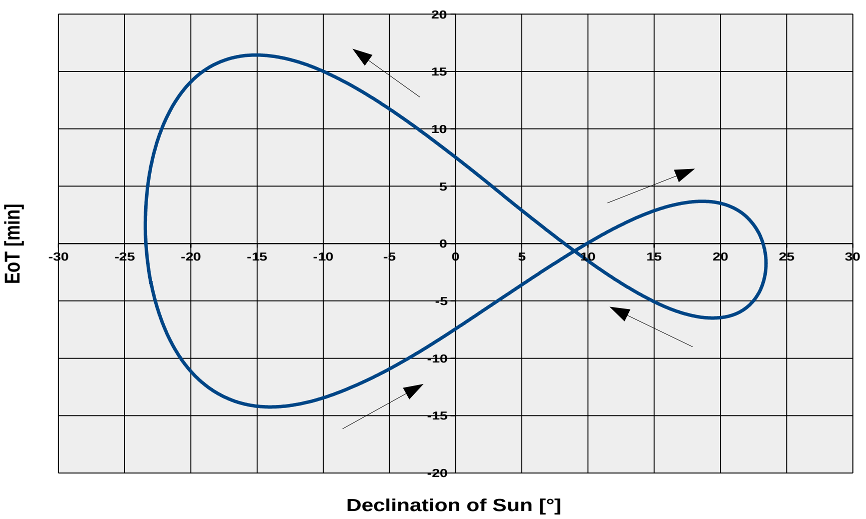
EoT varies periodically between approx. −14 and +17 minutes (*Fig. 3-7*). Predicted values for EoT for each day of the year (at 0:00 and 12:00 UT) are given in the Nautical Almanac (grey background indicates negative EoT). EoT is needed when calculating times of sunrise and sunset, or determining a **noon longitude** (with the exception of using the method described in chapter 6). Formulas for the calculation of EoT are given in chapter 15.

Fig. 3-7



Plotting EoT versus the declination of the Sun yields a figure-8 shaped curve which is called **analemma**. This curve is, for example, needed for the construction of certain types of sundials which compensate for EoT. *Fig. 3-8* shows the analemma for the year 2000.

Fig. 3-8



The hour angle of the mean Sun with respect to the lower branch of the local meridian (the upper branch going through the observer's position) is called **Local Mean Time**, **LMT**. LMT and UT are linked through the following formula:

$$LMT\,[h] \;=\; UT\,[h] + \frac{Lon\,[\circ]}{15}$$

The instant of the mean Sun passing through the upper branch of the local meridian is called **Local Mean Noon**, **LMN**.

A **zone time** is usually the Local Mean Time at a geographic longitude being a multiple of ±15°. Thus, zone times differ by an integer number of hours (with few exceptions). In the US, for example, Eastern Standard Time (UT −5h) is LMT at −75° longitude, Pacific Standard Time (UT−8h) is LMT at −120° longitude. Central European Time (UT+1h) is LMT at +15° longitude. Nowadays, zone times in civil life are no longer based upon UT but on UTC (see below). **Time zones** are areas where civil life is controlled by an adopted zone time. The boundaries of existing time zones have been established on various grounds (political, economical, geographical, or historical).

The hour angle of the apparent Sun with respect to the lower branch of the local meridian is called **Local Apparent Time**, **LAT**:

$$LAT\,[h] \;=\; GAT\,[h] + \frac{Lon\,[\circ]}{15}$$

The instant of the apparent Sun crossing the upper branch of the local meridian is called **Local Apparent Noon**, **LAN**.

Time measurement by the Earth's rotation does not necessarily require the Sun as the reference point in the sky. **Greenwich Apparent Sidereal Time**, **GAST**, is a time scale based upon the Greenwich hour angle (upper branch) of the **true vernal equinox of date, $\mathbf{GHA_{Aries}}$** (see Fig. 3-3). Hourly values of GHA_{Aries} are tabulated in the Nautical Almanac.

$$GAST\,[h] \;=\; \frac{GHA_{Aries}\,[\circ]}{15}$$

In the past, GAST has been measured by the Greenwich meridian transit of stars since GAST and the right ascension of the observed star are numerically equal at the moment of meridian transit.

The Greenwich hour angle (measured in hours) of the imaginary **mean vernal equinox of date** (travelling along the celestial equator at a constant rate) is called **Greenwich Mean Sidereal Time**, **GMST**. The difference between GAST and GMST at a given instant is called **equation of the equinoxes**, **EQ**, or **nutation in right ascension**. EQ can be predicted precisely. It varies periodically within approx. ±1s.

$$EQ \;=\; GAST \,-\, GMST$$

Due to the Earth's revolution around the Sun, a mean sidereal day (the time interval between two consecutive meridian transits of the mean equinox) is slightly shorter than a mean solar day:

$$24\,h \;\; Mean\,Sidereal\,Time \;=\; 23\text{h}\,56\text{m}\,4.090524\text{s} \;\; Mean\,Solar\,Time$$

UT is no longer measured by the hour angle of the Sun but is calculated from GMST, to which it is linked through a formula [10]. Nowadays, sidereal time (and thus, UT) is obtained by observation of extragalactical radio sources (quasars) which can be regarded as rigidly fixed to the imaginary celestial sphere since they do not exhibit any measurable **proper motion**. Their apparent diurnal motions are measured through **Very Long Baseline Interferometry (VLBI)**. This technology, which involves a global network of observation stations, produces much more accurate results than, e. g., the observation of any meridian transit.

By analogy with LMT and LAT, there is a **Local Mean Sidereal Time, LMST**, and a **Local Apparent Sidereal Time, LAST**:

$$LMST\,[h] \;=\; GMST\,[h] + \frac{Lon\,[\circ]}{15} \qquad\qquad LAST\,[h] \;=\; GAST\,[h] + \frac{Lon\,[\circ]}{15}$$

Solar time and sidereal time are both linked to the Earth's rotation. The Earth's rotation speed, however, decreases slowly (tidal friction). Moreover, it is subject to small fluctuations due to random movements of matter within the Earth's body (magma) and on the surface (water, air). Therefore, neither of the two time scales is strictly linear. Many applications in astronomy and physics, however, do require a linear time scale. One example is the calculation of ephemerides since the motions of celestial bodies in space are independent of the Earth's rotation.

International Atomic Time, **TAI**, is the most accurate time standard presently available. It is obtained by statistical analysis of data supplied by a world-wide network of several hundred atomic clocks. TAI forms the basis for other important time scales.

Civil life is mostly determined by **Coordinated Universal Time, UTC** (= **Zulu Time, Z**), which is the basis for time signals broadcast by radio stations, e. g., WWV. UTC is further available (2019) through the National Institute of Standards and Technology, NIST [14]. UTC is controlled by TAI. Due to the variable (in total decreasing) rotation speed of the Earth, UT tends to lag behind UTC. This is deemed undesirable since the cycle of day and night is linked to UT, not to UTC. To ensure that the difference, ΔUT, remains within the specified range of ±0.9s, UTC is synchronized to UT by introducing leap seconds at certain dates (June 30, December 31) whenever ΔUT approaches the upper or lower limit. Up to now (2019), only positive leap seconds in the form of minutes of 61 seconds each have been introduced (retarding UTC).

$$UT \;=\; UTC + \Delta UT \qquad\qquad UTC \;=\; TAI - N$$

N is the cumulative number of leap seconds introduced since 1972 (N = 37 in 2019.0). As a consequence of the occasional leap second, UTC is **not** a continuous time scale. Current and predicted values for ΔUT (= UT1−UTC) and DUT1 (= ΔUT rounded to the next 0.1s) are published weekly by the **IERS Rapid Service** (IERS Bulletin A) [15]. The IERS further gives prior notice of every upcoming insertion of a leap second (IERS Bulletin C). Note that UT is generally used as a synonym for **UT1** which is obtained by correcting **UT0** for the small effect of **polar motion** [17].

GPS Time, GPST, and Terrestrial Time, TT (≈ **Terrestrial Dynamical Time**, **TDT**), are continuous <u>and</u> linear time scales derived from TAI:

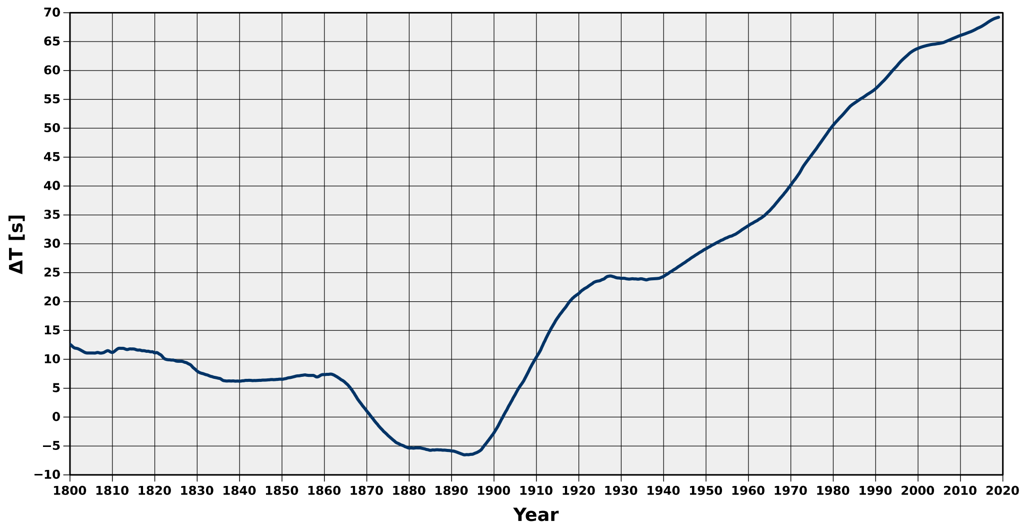
$$GPST = TAI - 19\,s \qquad\qquad TT = TAI + 32.184\,s$$

TT has replaced **Ephemeris Time**, **ET**. The offset of 32.184 s with respect to TAI is necessary to ensure a seamless continuation of ET. TT is widely used in astronomy (calculation of ephemerides) and space flight. The difference between TT and UT is denoted as ΔT.

$$TT \;=\; UT + \Delta T$$

Fig. 3-9 shows the development of ΔT since January 1800. At the beginning of the year 2019, ΔT was +69.2s [15].

Fig. 3-9



ΔT is of some interest to the navigator since computer almanacs require UT <u>and</u> TT (TDT) as time arguments (programs using only UT calculate on the basis of interpolated or extrapolated ΔT values). Since ΔT is affected by the earth's rotation, a precise long-term prediction is not possible. Therefore, computer almanacs using only UT as time argument will become less accurate in the long run. ΔT can be calculated with the following formula:

$$\Delta T \;=\; 32.184\,s + N - \Delta UT \;=\; 32.184\,s + (TAI - UTC) - (UT1 - UTC)$$

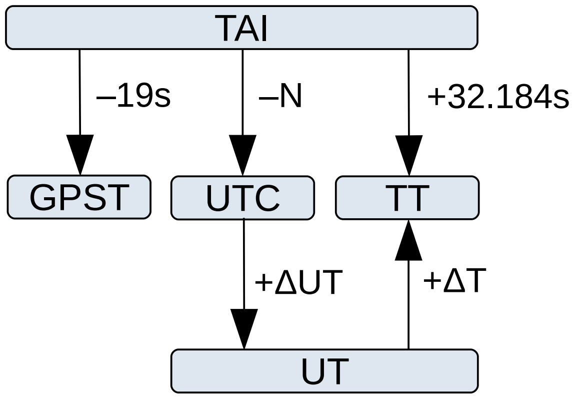
In addition to current UT1-UTC values (see above), the IERS Bulletin A includes values for TAI-UTC (= cumulative number of leap seconds, N).

At present (2019) there is an ongoing discussion among scientists about the usefulness of leap seconds. Computer operating systems (mostly running on UTC), for example, would be less prone to errors when linked to a continuous time scale. One potential error source with UTC, for example, is the creation of accurate time stamps for files. Another argument against UTC is that the drift of UT with respect to atomic time is so slow that its effect on civil life will not become obvious in the foreseeable future. Assuming that UT and UTC drift apart at an average rate of 1 second per year, it would take about 3600 years to reach a difference of 1 hour.

Compare this with the effects of the 1-hour jumps from standard time to summer time (daylight saving time) and back imposed by many governments. Up to now, no decision has been made. The practice of celestial navigation would not be affected by an abolition of leap seconds since the calculation of ephemerides (RA, Dec) is based upon TT whereas the Greenwich hour angle of a celestial body results from its right ascension and sidereal time at Greenwich (the latter being measured by the Earth's rotation).

The formal interrelations between TAI and other time scales are summarized in *Fig. 3-10*.

Fig. 3-10



The GMT Problem

The term GMT has become ambiguous since it is often used as a synonym for UTC now. Moreover, astronomers used to reckon GMT from the upper branch of the Greenwich meridian until 1925 (the time thus obtained is sometimes called Greenwich Mean Astronomical Time, GMAT). Therefore, the term GMT should be avoided in scientific publications, except when used in a historical context.

The Nautical Almanac

Predicted values for GHA and Dec of Sun, Moon and the navigational planets are tabulated for each integer hour (UT) of a calendar year on the daily pages of the **Nautical Almanac**, **N.A.**, and similar publications [12, 13]. GHA_{Aries} is tabulated in the same manner.

Listing GHA and Dec of all 57 fixed stars used in navigation for each integer hour of the year would require too much space in a book. Therefore, sidereal hour angles are tabulated instead of Greenwich hour angles. Since declinations and sidereal hour angles of stars change only slowly, tabulated values for periods of 3 days are accurate enough for celestial navigation. GHA is obtained by adding the SHA of the respective star to the current value of GHA_{Aries}.

GHA and Dec for each second of the year are obtained using the **interpolation tables** at the end of the N.A. (printed on tinted paper), as explained in the following directions.

1.

We note the exact time of observation (UT), determined with a chronometer and a stopwatch. If UT is not available, we can use UTC. The resulting error is tolerable in most cases.

2.

We look up the day of observation in the N.A. (two pages cover a period of three days).

3.

We go to the nearest integer hour preceding the time of observation and note GHA and Dec of the observed body. In case of a fixed star, we form the sum of GHA Aries and the SHA of the star, and note the tabulated declination. When observing planets, we note the ***v*** and ***d*** factors given at the bottom of the appropriate column. For the Moon, we take *v* and *d* for the nearest integer hour preceding the time of observation.
The quantity *v* is necessary to apply an additional correction to the following interpolation of the GHA of Moon and planets. It is not required for stars. The Sun does not require a *v* factor since the correction has been incorporated in the tabulated values for the Sun's GHA.

The quantity *d*, which is negligible for stars, is the rate of change of Dec, measured in arcminutes per hour. It is needed for the interpolation of Dec. The sign of *d* is critical!

4.

We look up the minute of observation in the interpolation tables (1 page for each 2 minutes of the hour), go to the second of observation, and note the **increment** from the respective column.

We enter one of the three columns to the right of the increment columns with the *v* and *d* factors and note the corresponding **corr**(ection) values (*v*-corr and *d*-corr).
The sign of *d*-corr depends on the trend of declination at the time of observation. It is positive if Dec at the integer hour following the observation is greater than Dec at the integer hour preceding the observation. Otherwise it is negative.

v -corr is negative for Venus. Otherwise, it is always positive.

5.

We form the sum of Dec and *d*-corr (if applicable).

6.

We form the sum of GHA (or GHA Aries and SHA in case of a star), increment, and *v*-corr (if applicable).
SHA values tabulated in the Nautical Almanac refer to the true vernal equinox of date.

Interactive Computer Ephemeris

Interactive Computer Ephemeris, **ICE**, is a computer almanac developed by the U.S. Naval Observatory (successor of the **Floppy Almanac**) in the 1980s. ICE is FREEWARE (no longer supported by USNO), compact, easy to use, and provides a vast quantity of accurate astronomical data for a time span of almost 250 (!) years.

In spite of its archaic design and cumbersome handling (DOS program), ICE is still a useful tool for navigators and amateur astronomers.

Among many other features, ICE calculates GHA and Dec for a given body and time as well as altitude and azimuth of the body for an **assumed position** (see chapter 4) and, moreover, sextant altitude corrections. Since the navigation data are as accurate as those tabulated in the **Nautical Almanac** (error ≦ 0.1'), the program makes an adequate alternative, although a printed almanac (and sight reduction tables) should be kept as a backup in case of a computer failure. The following instructions refer to the final version (0.51). Only program features relevant to navigation are explained.

1. Installation

Copy the program files to a chosen directory on the hard drive, floppy disk, USB stick, or similar storage device. ICE.EXE is the executable program file.

2. Getting Started

DOS users: Change to the program directory and enter "ice" or "ICE ". Windows users can run ICE in a DOS box.
Linux users: Install the DOS emulator "DOSBox". Copy the ICE files to a directory of your choice in your personal folder. ICE is started through the command “dosbox“, followed by a blank space and the path to the program file:

dosbox /home/<user name>/<program directory>/ICE.EXE   (Note that Linux is case-sensitive.)

After ICE has started, the main menu appears. Use the function keys F1 to F10 to navigate through the submenus. The program is more or less self-explanatory. Go to the submenu INITIAL VALUES (F1). Follow the directions on the screen to enter date and time of observation (F1), assumed latitude (F2), assumed longitude (F3), and your local time zone (F6). Assumed latitude and longitude define your assumed position. Use the correct data format, as shown on the screen (decimal format for latitude and longitude).

After entering the above data, press F7 to accept the values displayed. To change the default values permanently, edit the file ice.dft with a text editor (after making a backup copy) and make the desired changes. Do not change the data format. The numbers have to be in columns 21-40. To create an output file to store calculated data, go to the submenu FILE OUTPUT (F2) and enter a chosen file name, e.g., OUTPUT.TXT.

3. Calculation of Navigational Data

From the main menu, go to the submenu NAVIGATION (F7). Enter the name of the body. The program displays GHA and Dec of the body, GHA and Dec of the Sun (if visible), and GHA of the vernal equinox for the date and time (UT) stored in INITIAL VALUES.

Hc (computed altitude) and Zn (true azimuth) mark the apparent position of the body as observed from the assumed position. Approximate altitude corrections (refraction, SD, PA), based upon Hc, are also displayed (for lower limb of body). The semidiameter of the Moon includes augmentation.

The coordinates calculated for Venus and Mars do **not** include the phase correction. Therefore, the upper or lower limb (if discernable) should be observed.

ΔT is TT(TDT)-UT, the predicted difference between **terrestrial time** and UT for the given date. The ΔT value for 2019.0 predicted by ICE is 84.8s, the actual value at 2019.0 is 69.2s which demonstrates that the extrapolation algorithm used by ICE is outdated.

Horizontal parallax and semidiameter of a body can be extracted from the submenu POSITIONS (F3). Choose APPARENT GEOCENTRIC POSITIONS (F1) and enter the name of the body (Sun, Moon, planets). The last column shows the distance of the center of the body from the center of the Earth, measured in astronomical units (1 AU = $149.6 \cdot 10^6$ km). HP and SD are calculated as follows:

$$HP = \arcsin \frac{r_E[km]}{distance[km]} \qquad SD = \arcsin \frac{r_B[km]}{distance[km]}$$

r_E is the equatorial radius of the Earth (6378 km). r_B is the radius of the respective body (Sun: 696260 km, Moon: 1378 km, Venus: 6052 km, Mars: 3397 km, Jupiter: 71398 km, Saturn: 60268 km).

The apparent geocentric positions refer to TT (TDT). However, the difference between TT and UT has no significant effect on HP and SD.

To calculate the times of rising and setting of a body, go to the submenu RISE & SET TIMES (F6) and enter the name of the body. The columns on the right display the time of rising, meridian transit, and setting for the assumed location (UT+x hours, according to the time zone specified).

The increasing error of ΔT values predicted by ICE may lead to reduced precision when calculating navigation data in the future. The coordinates of the Moon are particularly sensitive to errors of ΔT. Unfortunately, ICE has no option for editing and modifying the internal ΔT algorithm. The high-precision part of ICE, however, is not affected since TT (TDT) is the time argument.

To circumvent the ΔT problem, extract GHA and Dec using the following procedure:

1. Compute GAST using SIDEREAL TIME (F5). Here, the time argument is UT (!).

2. Edit date and time at INITIAL VALUES (F1). Now, the time argument is TT (= UT+ΔT). Compute RA and Dec using POSITIONS (F3) and APPARENT GEOCENTRIC POSITIONS (F1).

3. Use the following formula to calculate GHA from GAST and RA (RA refers to the true vernal equinox of date):

$$GHA[°] = 15 \cdot (GAST[h] + 24\text{h} - RA[h])$$

(If the resulting angle is greater than 360°, subtract 360°.)

High-precision GHA and Dec values thus obtained can be used as an internal standard to cross-check medium-precision data obtained through NAVIGATION (F7).

Chapter 4

Finding One's Position (Sight Reduction)

Lines of Position

Any geometrical or physical line passing through the observer's (still unknown) position and accessible through measurement or observation is called a **position line** or **line of position**, **LOP**. Examples are circles of equal altitude, meridians, parallels of latitude, bearing lines (compass bearings) of terrestrial objects, coastlines, rivers, roads, railroad tracks, power lines, etc. **A single position line indicates an infinite series of possible positions.** The observer's actual position is marked by the point of intersection of at least two position lines, regardless of their nature. A position thus found is called **fix** in navigator's language. The concept of the position line is essential to modern navigation.

Sight Reduction

Finding a line of position by observation of a celestial object is called **sight reduction**. Although some background in mathematics is required to comprehend the process completely, knowing the basic concepts and a few equations is sufficient for most practical applications. The mathematical background is given in chapter 10 and chapter 11. In the following, we will discuss the semi-graphic methods developed by *Sumner* and *St. Hilaire*. Both methods require relatively simple calculations only and enable the navigator to plot lines of position on a nautical chart or **plotting sheet** (chapter 13).

Knowing altitude and geographic position of a body, we also know the radius of the corresponding circle of equal altitude (our circular line of position) and the position of its center. As mentioned in chapter 1 already, plotting circles of equal altitude on a chart is usually impossible due to their large dimensions and the distortions caused by map projection. However, *Sumner* and *St. Hilaire* showed that only a short arc of each circle of equal altitude is needed to find one's position. Such a short arc can be represented by a secant or a tangent of the circle.

Local Meridian, Local Hour Angle and Meridian Angle

The meridian passing through a given position, usually that of the observer, is called **local meridian**. In celestial navigation, the angular distance between the hour circle of the observed body (upper branch) and the local meridian (upper branch) plays a fundamental role. On the analogy of the Greenwich hour angle, we can measure this angle westward from the local meridian (0°...+360°). In this case, the angle is called **local hour angle**, **LHA**. It is also possible to measure the angle westward (0°...+180°) or eastward (0°...−180°) from the local meridian in which case it is called **meridian angle**, **t**. In most navigational formulas, LHA and t can be substituted for each other since the trigonometric functions return the same results. For example, the cosine of +315° equals the cosine of −45° due to the periodicity of trigonometric functions.

Like LHA, t is the algebraic sum of the Greenwich hour angle of the body, GHA, and the observer's geographic longitude, Lon. To make sure that the obtained angle is in the desired range, the following rules have to be applied when forming the sum of GHA and Lon:

$$LHA = \begin{cases} GHA + Lon & \text{if} \quad 0° < GHA + Lon < 360° \\ GHA + Lon + 360° & \text{if} \quad GHA + Lon < 0° \\ GHA + Lon - 360° & \text{if} \quad GHA + Lon > 360° \end{cases}$$

$$t = \begin{cases} GHA + Lon & \text{if} \quad GHA + Lon < 180° \\ GHA + Lon - 360° & \text{if} \quad GHA + Lon > 180° \end{cases}$$

In all calculations, the sign of Lon and t, respectively, has to be observed carefully. The sign convention* used throughout this publication is:

Eastern longitude:	**positive**
Western longitude:	**negative**
Eastern meridian angle:	**negative**
Western meridian angle:	**positive** (same as LHA)

*There are other sign conventions found in the literature. These require modified formulas and a different set of conversion rules.

For reasons of symmetry, we will operate with the meridian angle throughout this text. Provided Lat and Dec are constant, the meridian angle +t results in the same altitude as the meridian angle −t.*

*Accordingly, LHA leads to the same altitude as 360°-LHA.

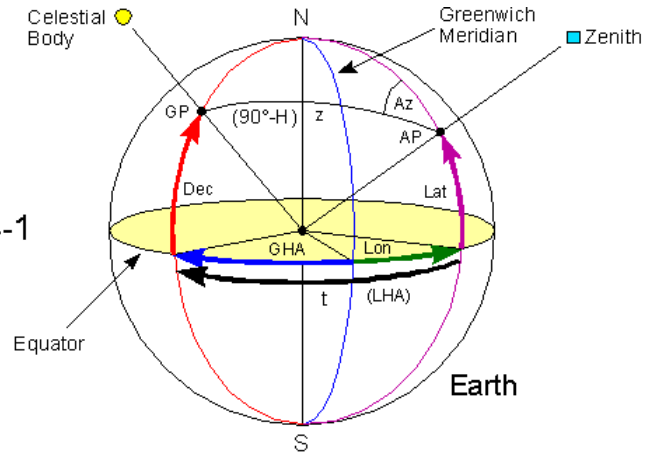


Fig. 4-1

Fig. 4-1 illustrates the various angles involved in the sight reduction process. The spherical triangle formed by GP, AP, and the geographic north pole is called **navigational** (or **nautical**) **triangle** (chapter 11).** AP is the observer's position, be it real, estimated, or assumed (see intercept method).

Other sources define the nautical triangle as the spherical triangle formed by GP, AP, and the **elevated pole (the celestial pole above the horizon). In this case, other rules than those used in this publication have to be used.

Sumner's Method

In December 1837, *Thomas H Sumner*, an American sea captain, was on a voyage from South Carolina to Greenock, Scotland. When approaching *St. George's Channel* between Ireland and Wales, he managed to measure a single altitude of the Sun after a longer period of bad weather. Using the time sight formula (see chapter 6), he calculated a longitude from his estimated latitude. Since he was doubtful about his estimate, he repeated his calculations with two slightly different latitudes. To his surprise, he was able to draw a straight line through the three positions thus located on his chart. As it happened, the line passed through the position of a lighthouse off the coast of Wales (*Small's Light*). By intuition, Sumner steered his ship along this line and soon after, *Small's Light* came in sight. Sumner concluded that he had found a "line of equal altitude". The publication of his method in 1843 marked the beginning of “modern“ celestial navigation [18]. Although rarely used today, it is still an interesting alternative to the intercept method (see below). It is easy to comprehend and the calculations to be done are quite simple.

Fig. 4-2 shows the points where a circle of equal altitude (a **small circle** unless Ho=0) intersects two chosen parallels of latitude.

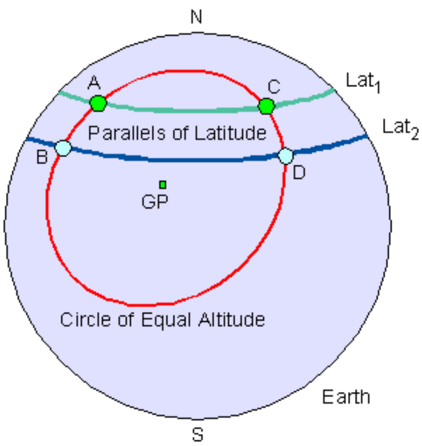


Fig. 4-2

An observer being between Lat_1 and Lat_2 is either on the arc A-B or on the arc C-D. With a rough estimate of the longitude of his position, the observer can easily find on which of the two arcs he actually is, for example, A-B. The arc thus found is the relevant part of our line of position, the other arc is discarded. On a chart, we can approximate the line of position by drawing a straight line through A and B. This line, called **Sumner line**, is a secant of the circle of equal altitude. The error caused by the curvature of the arc usually remains tolerable, with the exception of high-altitude observations. Before plotting the Sumner line on our chart, we have to calculate the respective longitude of each intersection point, A, B, C, and D. How to do this is described in the following.

Procedure:

1.

We choose a parallel of latitude (Lat_1) north of our estimated latitude. Preferably, Lat_1 should be marked by the nearest horizontal grid line on our chart or plotting sheet. We call this an **assumed latitude**.

2.

From Lat_1, Dec, and the observed altitude, Ho, we calculate the meridian angle, t, using the following formula:

$$t \;=\; \pm\,\arccos\frac{\sin Ho - \sin Lat \cdot \sin Dec}{\cos Lat \cdot \cos Dec}$$

The equation is derived from the navigational triangle (chapter 10 & chapter 11). It has two solutions, +t and –t, because cos(+t) = cos(–t). Geometrically, this corresponds to the fact that the circle of equal altitude intersects the parallel of latitude at two points with a longitude difference of $2 \cdot t$. Using the following formulas and rules, we obtain the longitudes of these two points of intersection, Lon and Lon' :

$$Lon \;=\; t - GHA$$

$$Lon' \;=\; 360° - t - GHA$$

$$\begin{array}{lll} \text{If} \;\; Lon < -180° & \rightarrow & Lon + 360° \\ \text{If} \;\; Lon' < -180° & \rightarrow & Lon' + 360° \\ \text{If} \;\; Lon' > +180° & \rightarrow & Lon' - 360° \end{array}$$

Comparing the two longitudes thus obtained with our estimated longitude, we select the most probable one and discard the other. This method of finding one's longitude is called **time sight** (chapter 6).

3.

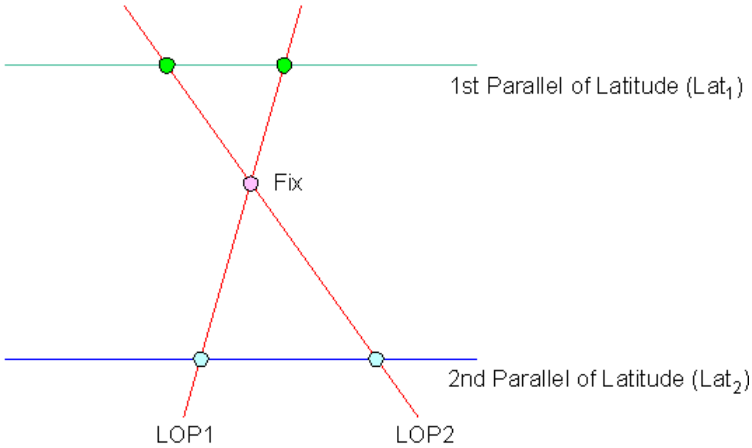
We choose an assumed parallel of latitude (Lat_2) south of our estimated latitude. The difference between Lat_1 and Lat_2 should not exceed a few degrees. We repeat steps 1 and 2 with the second latitude, Lat_2.

4.

On our plotting sheet, we mark each longitude thus obtained on its corresponding parallel of latitude, and plot the Sumner line through the points thus located (LOP1, see *Fig. 4-3*).

Using the same parallels of latitude, we repeat steps 1 through 4 with the declination and observed altitude of a second body. The point where the Sumner line thus obtained, LOP2, intersects LOP1 is our fix.

Fig. 4-3



If we have only a very rough estimate of our latitude, the resulting point of intersection may be slightly outside the interval defined by both parallels, but the fix is still correct. As already said, any fix obtained with *Sumner's* method has a small error caused by ignoring the curvature of the circles of equal altitude. We can improve the fix by iteration. For this purpose, we take a chart with a larger scale, choose a new pair of assumed latitudes, nearer to the fix, and repeat the procedure with the same altitudes. Alternatively, we can choose three parallels of latitude, mark the resulting longitudes on them, and plot the position line through these points with a flexible curve ruler. Ideally, the azimuth difference between the observed bodies should be 90° (30°...150° is tolerable). Otherwise, the fix would become indistinct. When the azimuth of a body is close to 0° or 180°, an almost horizontal LOP will be obtained on the plotting sheet. This has to be observed when choosing the parallels of latitude and the scale of the plotting sheet.

Sumner's method has the advantage that we do not have to measure angles on a chart. Therefore, we can find the fix by plotting the position lines on ordinary squared paper available at any stationary shop. Special small area plotting sheets (chapter 13) are not needed here unless we want to advance or retire a position line (chapter 5).

The Intercept Method

This procedure was developed by the French navy officer *St. Hilaire* and others and was first published in 1875. Afterwards, it gradually became the standard procedure for sight reduction since it avoids some of the restrictions of the time sight and *Sumner's* method. Although the theoretical background is more complicated than with *Sumner's* method, the practical application is very convenient.

Theory:

For any given position of the observer, the geocentric altitude of a celestial body is solely a function of the observer's latitude, the declination of the body, and the meridian angle (or local hour angle).

$$H = f\left(Lat, Dec, t\right)$$

The **altitude formula** is obtained by applying the law of cosines for sides to the navigational triangle (chapter 10 & 11):

$$H = \arcsin\left(\sin Lat \cdot \sin Dec + \cos Lat \cdot \cos Dec \cdot \cos t\right)$$

An alternative form* of the altitude formula is based on the haversine formula described in chapter 10 and chapter 11:

$$H = \arcsin\left(\cos\left(Lat - Dec\right) - \cos Lat \cdot \cos Dec \cdot \left(1 - \cos t\right)\right)$$

*Only four trigonometric functions instead of five have to be calculated when using the latter version.

We choose an arbitrary point on our nautical chart which is not too far from our estimated position. Preferably this is the nearest point where two grid lines on the chart intersect. This point is called **assumed position**, **AP** (*Fig. 4-4*). Using one of the above formulas, we calculate the altitude of the body resulting from Lat_{AP} and Lon_{AP}, the geographic coordinates of AP. The altitude thus obtained is called **computed** or **calculated altitude**, **Hc**.

Usually, Hc will slightly differ from the actually **observed altitude**, **Ho** (chapter 2). The difference, ΔH, is called **intercept**.

$$\Delta H = Ho - Hc$$

Ideally (no observation errors), Ho and Hc would be identical if the observer were exactly at AP.

In the following, we will discuss which possible positions of the observer would result in the same intercept, ΔH. For this purpose, we assume that the intercept is an infinitesimal quantity and denote it by dH. The general formula is:

$$dH = \frac{\partial H}{\partial Lat} \cdot d\,Lat + \frac{\partial H}{\partial t} \cdot dt$$

This differential equation has an infinite number of solutions. Since dH and both differential coefficients are constant, it can be reduced to a linear equation of the general form:

$$d\,Lat = a + b \cdot dt$$

Thus, the graph is a straight line, and it is sufficient to dicuss two special cases, dt=0 and dLat=0, respectively.

In the first case, the observer is on the same meridian as AP, and the small change dH is solely caused by a small variation of latitude, dLat, whereas t is constant (dt = 0). We differentiate the altitude formula with respect to Lat:

$$\sin H = \sin Lat \cdot \sin Dec + \cos Lat \cdot \cos Dec \cdot \cos t$$

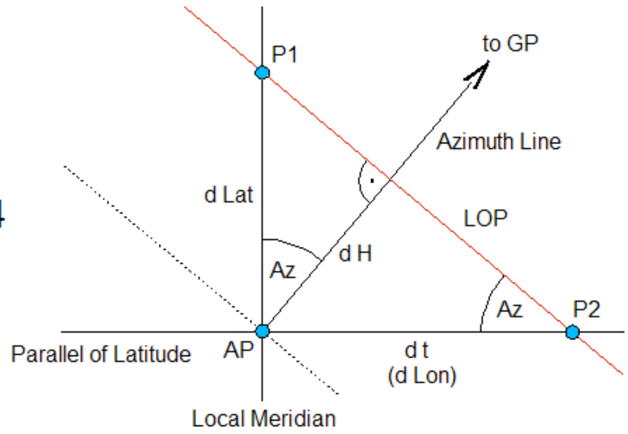
$$d\left(\sin H\right) = \left(\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t\right) \cdot d\,Lat$$

$$\cos H \cdot dH = \left(\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t\right) \cdot d\,Lat$$

$$d\,Lat = \frac{\cos H}{\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t} \cdot dH$$

Adding dLat to Lat_{AP}, we obtain the point P1, as illustrated in *Fig.4-4*. P1 is on the circle of equal altitude.

Fig. 4-4



In the second case, the observer is on the same parallel of latitude as AP, and dH is solely caused by a small change of the meridian angle, dt, whereas Lat is constant (dLat=0). Again, we begin with the altitude formula:

$$\sin H = \sin Lat \cdot \sin Dec + \cos Lat \cdot \cos Dec \cdot \cos t$$

Differentiating with respect to t, we get

$$d(\sin H) = -\cos Lat \cdot \cos Dec \cdot \sin t \cdot dt$$

$$\cos H \cdot dH = -\cos Lat \cdot \cos Dec \cdot \sin t \cdot dt$$

$$dt = -\frac{\cos H}{\cos Lat \cdot \cos Dec \cdot \sin t} \cdot dH$$

Adding dt (corresponding with an equal change of longitude, dLon) to Lon_{AP}, we obtain the point P2 which is on the same circle of equal altitude. Thus, we would measure Ho at P1 and P2, respectively. Knowing P1 and P2, we can now plot a straight line passing through these positions. This line, a tangent of the circle of equal altitude, is our approximate **line of position**, **LOP**. The great circle passing through AP and GP is represented by a straight line perpendicular to the line of position, called **azimuth line**. The arc between AP and GP is the radius of the circle of equal altitude. The distance between AP and the point where the azimuth line intersects the line of position is the intercept, dH. The angle formed by the azimuth line and the local meridian of AP is called **azimuth angle**, **Az**. The same angle is measured between the line of position and the parallel of latitude passing through AP (*Fig. 4-4*).

There are several ways to derive Az and the **true azimuth**, **Az_N**, from the right (plane) triangle defined by the vertices AP, P1, and P2:

Time-Altitude Azimuth:

$$\cos Az = \frac{dH}{d\,Lat} = \frac{\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t}{\cos H}$$

$$Az = \arccos \frac{\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t}{\cos H}$$

Alternatively, this formula can be derived from the navigational triangle (law of sines and cosines, chapter 10 & chapter 11). Az is not necessarily identical with the true azimuth, Az_N, since the arccos function returns angles between 0° and +180°, whereas Az_N is measured from 0° to +360°. To obtain Az_N, we have to apply the following rules after calculating Az with the formula for time-altitude azimuth:

$$Az_N = \begin{cases} Az & \text{if} \quad t < 0° \quad (180° < LHA < 360°) \\ 360° - Az & \text{if} \quad t > 0° \quad (0° < LHA < 180°) \end{cases}$$

Time Azimuth:

$$\tan Az = \frac{d\,Lat}{\cos Lat \cdot dt} = \frac{\sin t}{\sin Lat \cdot \cos t - \cos Lat \cdot \tan Dec}$$

The factor cos Lat is the relative circumference of the parallel of latitude on which AP is situated (equator = 1).

$$Az = \arctan \frac{\sin t}{\sin Lat \cdot \cos t - \cos Lat \cdot \tan Dec}$$

Alternatively, the time azimuth formula can be derived from the navigational triangle (law of cotangents, chapter 10 & chapter 11). Knowing the altitude is not necessary. This formula requires a different set of rules to obtain Az_N:

$$Az_N = \begin{Bmatrix} Az & \text{if} & t < 0 \ \left(180° < LHA < 360°\right) & \text{AND} & denominator < 0 \\ Az + 360° & \text{if} & t > 0 \ \left(0° < LHA < 180°\right) & \text{AND} & denominator < 0 \\ Az + 180° & \text{if} & denominator > 0 & & \end{Bmatrix}$$

Calculating the time azimuth is much more convenient with the arctan2 (= atan2) function. The latter is part of many programming languages and spreadsheet programs and eliminates the quadrant problem. Thus, no conversion rules are required to obtain Az_N. In a LibreOffice spreadsheet, for example, the equation would have the following format:

$$Az_N = 180° + degrees\left(\text{atan2}\left(\sin(Lat) * \cos(t) - \cos(Lat) * \tan(Dec), \sin(t)\right)\right)$$

(Some programming languages and spreadsheet programs use a semicolon as a separator, not a comma.)

Altitude Azimuth:

This formula is directly derived from the navigational triangle (cosine law, haversine formula, chapter 10 & chapter 11) without using differential calculus.

$$\cos Az = \frac{\sin Dec - \sin H \cdot \sin Lat}{\cos H \cdot \cos Lat} = 1 - \frac{\cos(H - Lat) - \sin Dec}{\cos H \cdot \cos Lat}$$

$$Az = \arccos \frac{\sin Dec - \sin H \cdot \sin Lat}{\cos H \cdot \cos Lat} = \arccos\left(1 - \frac{\cos(H - Lat) - \sin Dec}{\cos H \cdot \cos Lat}\right)$$

Az_N is obtained through the same rules as used with the time-altitude azimuth:

$$Az_N = \begin{Bmatrix} Az & \text{if} & t < 0 \ \left(180° < LHA < 360°\right) \\ 360° - Az & \text{if} & t > 0 \ \left(0° < LHA < 180°\right) \end{Bmatrix}$$

Azimuth by the Law of Sines:

The azimuth can also be obtained by application of the law of sines for spherical triangles (chapter 10 & chapter 11). The formula is quite simple and does not require the latitude to be known. However, this comes at the cost of ambiguity since the sine of any given angle, α, equals the sine of its supplement, 180°-α. Therefore, the formula should be used with care.

$$Az = \arcsin \frac{\cos Dec \cdot \sin t}{\cos Ho}$$

To obtain the true azimuth, we apply the following rules:

$$Az_N = \begin{Bmatrix} 180° + Az & \text{or} & -Az & \text{if} & t < 0° & \left(180° < LHA < 360°\right) \\ 180° + Az & \text{or} & 360° - Az & \text{if} & t > 0° & \left(0° < LHA < 180°\right) \end{Bmatrix}$$

To find out which of the two solutions for either case is the one we are looking for, we have to compare them with the compass bearing of the observed body. The difference between both solutions becomes very small as Az approaches +90° or –90°. In such a case, a clear distinction will be impossible, and one of the azimuth formulas described further above should be chosen instead.

Fig. 4-5 shows a macroscopic view of the line of position, the azimuth line, and the circles of equal altitude. In contrast to dH, ΔH is a measurable quantity. Further, the position line is curved.

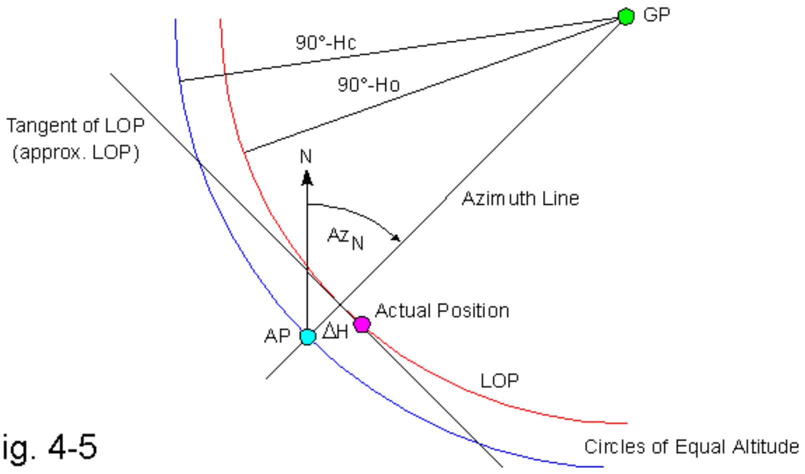


Fig. 4-5

Procedure for the Intercept Method:

Although the theory of the intercept method may look complicated at first glance, the practical application is very simple and does not require any background in differential calculus. The procedure comprises the following steps:

1.

We choose an **assumed position**, **AP** (see Fig. 4-1), which should be near to our **estimated position**. Preferably, AP should be defined by an integer number of degrees or arcminutes for Lat_{AP} and Lon_{AP}, respectively, depending on the scale of the chart we are using. Our estimated position itself may be used as well, but plotting a position line is easier when putting AP on an intersection point of two grid lines.

2.

We calculate the meridian angle, t_{AP}, (or the local hour angle, LHA_{AP}) from GHA and Lon_{AP}, as shown earlier.

3.

We calculate the geocentric altitude of the observed body as a function of Lat_{AP}, t_{AP}, and Dec (computed altitude):

$$Hc = \arcsin\left(\sin Lat_{AP} \cdot \sin Dec + \cos Lat_{AP} \cdot \cos Dec \cdot \cos t_{AP}\right)$$

4.

We calculate the **true azimuth** of the body, $\mathbf{Az_N}$, for example with the altitude azimuth formula:

$$Az = \arccos \frac{\sin Dec - \sin Hc \cdot \sin Lat_{AP}}{\cos Hc \cdot \cos Lat_{AP}}$$

$$Az_N = \begin{cases} Az & \text{if} \quad t < 0 \quad (180° < LHA < 360°) \\ 360° - Az & \text{if} \quad t > 0 \quad (0° < LHA < 180°) \end{cases}$$

5.

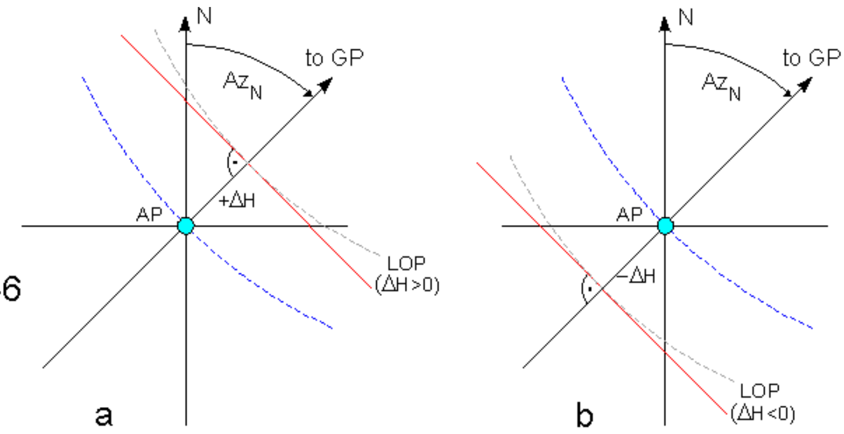
We calculate the **intercept**, **ΔH**, the difference between observed altitude, Ho (chapter 2), and computed altitude, Hc. The intercept, which is directly proportional to the difference between the radii of the corresponding circles of equal altitude, is usually expressed in nautical miles:

$$\Delta H[nm] = 60 \cdot \left(Ho[°] - Hc[°]\right)$$

6.

We take a chart or plotting sheet with a convenient scale (depending on the respective scenario), and draw a suitable length of the azimuth line through AP (*Fig. 4-6*). On this line, we measure the intercept, ΔH, from AP (towards GP if ΔH>0, away from GP if ΔH<0) and draw a perpendicular through the point thus located. This perpendicular to the azimuth line is our approximate line of position (the red line in *Fig. 4-6*).

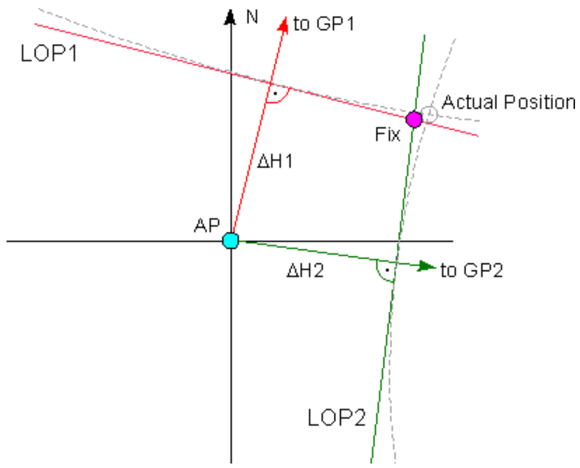
Fig. 4-6



7.

To obtain our position, we need at least one additional position line. We repeat the procedure with altitude and GP of a second celestial body or of the same body at a different time of observation (*Fig. 4-7*). The point where both position lines (tangents) intersect is our fix. The second observation does not necessarily require the same AP to be used.

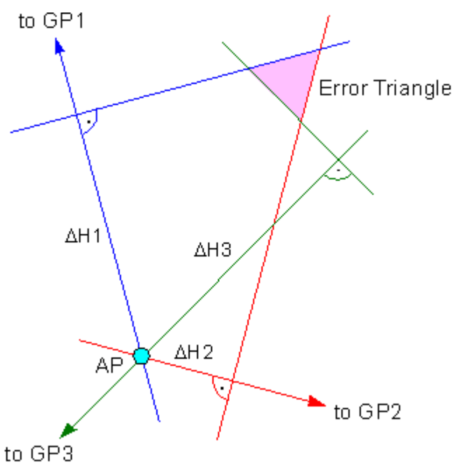
Fig. 4-7



Since the intercept method ignores the curvatures of the actual position lines, the obtained fix is not our exact position but rather an **improved position** (compared with AP). The residual error remains tolerable as long as the radii of the circles of equal altitude are not too small and AP is not too far from our actual position (chapter 16). The geometric error inherent to the intercept method can be decreased by **iteration**, i.e., substituting the obtained fix for AP and repeating the calculations (same altitudes and GP's). This will result in a more accurate position. If necessary, we can reiterate the procedure until the obtained position remains virtually constant (rarely needed).

Accuracy is also improved by observing three bodies instead of two. Theoretically, the position lines should intersect each other at a single point. Since no observation is entirely free of errors, we will usually obtain three points of intersection forming an **error triangle** (*Fig. 4-8*).

Fig. 4-8



Area and shape of the error triangle give us a rough estimate of the quality of our observations (chapter 16). Our **most probable position**, **MPP**, is approximately (!) represented by the “center of gravity“ of the error triangle (the point where the bisectors of the three angles of the error triangle meet).

When observing more than three bodies, the resulting position lines will form the corresponding polygons.

Direct Computation

If we do not want to plot lines of position to determine our fix, we can calculate the most probable position directly from n observations (n > 1). The Nautical Almanac provides an averaging procedure based on statistical methods. First, the auxiliary quantities A, B, C, D, E, and G have to be calculated.

$$A = \sum_{i=1}^{n} \cos^2 Az_i$$

$$B = \sum_{i=1}^{n} \sin Az_i \cdot \cos Az_i$$

$$C = \sum_{i=1}^{n} \sin^2 Az_i$$

$$D = \sum_{i=1}^{n} (\Delta H)_i \cdot \cos Az_i$$

$$E = \sum_{i=1}^{n} (\Delta H)_i \cdot \sin Az_i$$

$$G = A \cdot C - B^2$$

In the above formulas, Az_i denotes the true azimuth of the respective body. The ΔH values are measured in degrees (same unit as Lon and Lat). The geographic coordinates of the observer's MPP are then obtained as follows:

$$Lon = Lon_{AP} + \frac{A \cdot E - B \cdot D}{G \cdot \cos Lat_{AP}} \qquad Lat = Lat_{AP} + \frac{C \cdot D - B \cdot E}{G}$$

The method does not correct for the geometric errors caused by the curvatures of position lines. Again, these are eliminated, if necessary, by iteration. For this purpose, we substitute the calculated MPP for AP. For each body, we calculate new values for t (or LHA), Hc, ΔH, and Az_N. With these values, we recalculate A, B, C, D, E, G, Lon, and Lat.

Upon repeating this procedure, the resulting positions will converge rapidly. In the majority of cases, less than two iterations will be sufficient, depending on the distance between AP and the true position.

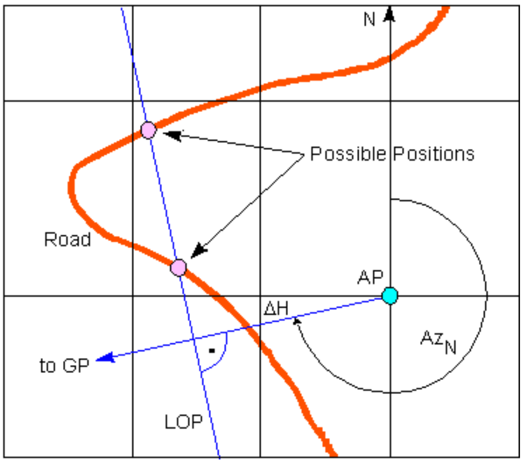
We should keep in mind, however, that this procedure is sensitive to outliers which may lead to a distorted result. Using the methods of **robust statistics** (see chapter 16) might be the better choice when outliers are suspected to exist.

Combining Different Lines of Position

Since the point of intersection of any two position lines, regardless of their nature, marks the observer's geographic position, one celestial LOP may suffice to find one's position if another LOP of a different kind is available.

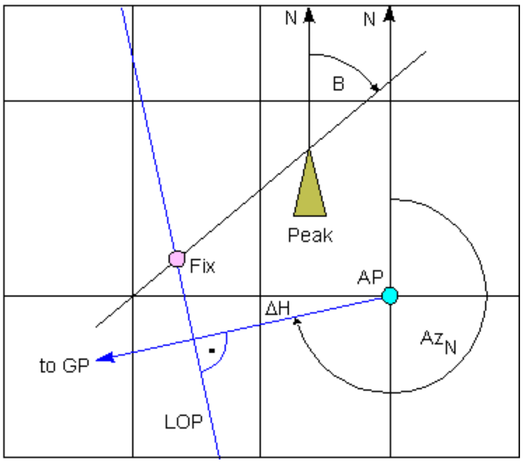
In a desert devoid of any landmarks, for instance, we can determine our current position by finding the point on the map where a position line obtained by observation of a celestial object intersects the dirt road we are travelling on (*Fig. 4-9*).

Fig. 4-9



We can as well find our position by combining our celestial LOP with the bearing line of a distant mountain peak or any other prominent landmark (*Fig. 4-10*). B is the compass bearing of the terrestrial object (corrected for magnetic declination).

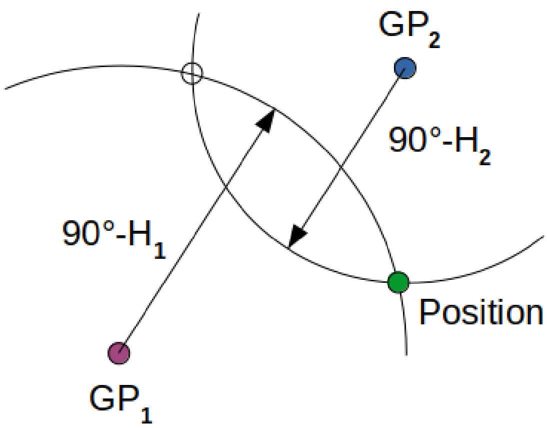
Fig. 4-10



High-altitude plots

In rare cases when a body near the zenith (H > 80°) is observed, we can not ignore the curvature of the position line. In such a case, the position should be improved by iteration. Alternatively, we can plot the whole circle of equal altitude or the relevant part of it with a compass (*Fig. 4-11*). The radius on the chart is 90°-Ho. We should never forget, however, that small circles of equal altitude increase the risk of mistaking the second point of intersection for the true position (ambiguity).

Fig. 4-11



Chapter 5

Finding the Position of an Advancing Vessel

Celestial navigation on an advancing vessel requires a correction for the change of position between subsequent observations unless the latter are performed in rapid succession or, better, simultaneously by two observers.

To apply this correction, the navigator has to know two parameters, **course made good, CMG**, and **speed made good, SMG**. The former is the actual direction (measured clockwise from true north) in which the vessel is moving. The latter is the (average) speed over ground.

Assuming that we make our first observation at the instant T_1 and our second observation at T_2, the distance, d, traveled during the time interval T_2-T_1 is

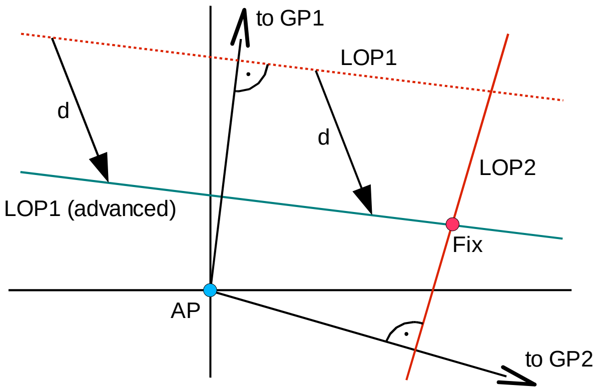
$$d[nm] = \left(T_2[h] - T_1[h]\right) \cdot SMG[kn]$$

1 kn (knot) = 1 nm/h

Although we have no knowledge of our absolute position yet, we know our second position relative to the first one now. On a chart, this shift of position is represented by an arrow (vector) defined by CMD and d.

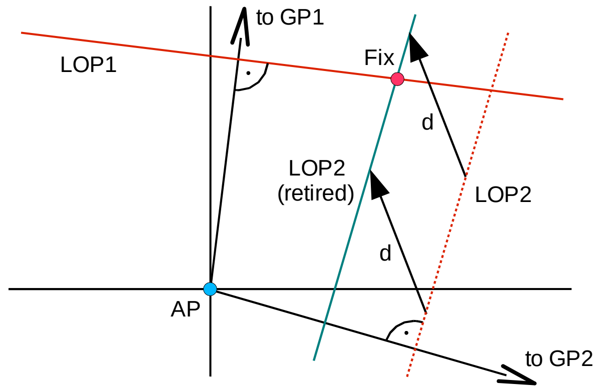
To find the absolute position, we plot both position lines in the usual manner, as illustrated in chapter 4. Next, we choose an arbitrary point on the first position line, LOP1 (resulting from the observation at T_1), and translate this point along the (free) vector defined by d and CMG. Finally, we draw a parallel of the first position line through the point thus located. The point where this **advanced position line** intersects the second line of position (resulting from the observation at T_2) marks our position at the instant T_2. A position obtained in this fashion is called **running fix** (*Fig. 5-1*).

Fig. 5-1



In a similar manner, we can obtain our position at $\mathbf{T_1}$ by **retiring** the second position line, LOP2. In this case we have to increase or decrease CMD by 180° (*Fig. 5-2*).

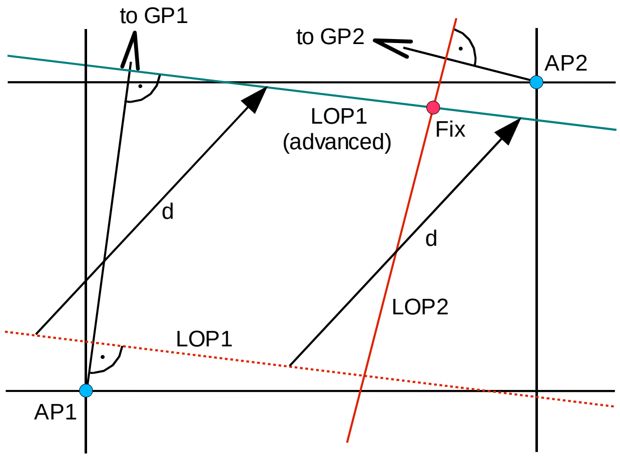
Fig. 5-2



Terrestrial lines of position may be advanced or retired in the same way as astronomical position lines.

It is also possible to choose two different assumed positions. AP1 should be close to the estimated position at T_1, AP2 close to the estimated position at T_2 (*Fig. 5-3*).

Fig. 5-3



A running fix is not as accurate as a stationary fix. For one thing, course and speed over ground can only be estimated since current and wind (drift) are not exactly known in most cases.

Further, there is a geometrical error inherent to the method. The latter is based on the assumption that each point of the circle of equal altitude, representing a possible position of the vessel, travels the same distance, d, along the **rhumb line** (chapter 12) defined by CMG. The result of such an operation, however, is a slightly distorted circle. Consequently, an advanced or retired LOP is not exactly parallel to the original LOP. The resulting position error usually increases as the distance, d, increases [19]. Thus, we should not travel hundreds of miles before making the second observation when fairly accurate results are required.

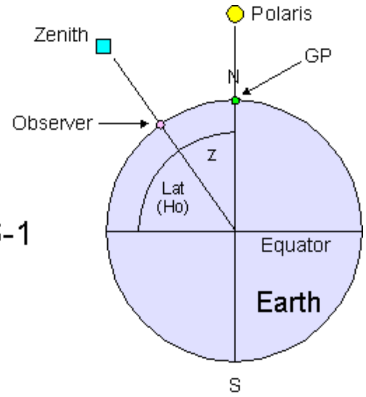
Chapter 6

Determination of Latitude and Longitude, Finding a Position by Direct Calculation

Latitude by Polaris

The geocentric altitude of a celestial object being at the celestial north pole would be numerically equal to the latitude of the observer's position (*Fig. 6-1*).

Fig. 6-1



This is nearly the case with Polaris, the pole star. However, since the declination of Polaris is not exactly 90° (89° 16.0' in 2000.0), the altitude of Polaris depends on its local hour angle. The altitude of Polaris is also affected, to a lesser degree, by nutation. To obtain the accurate latitude from the observed altitude, several corrections are necessary:

$$Lat = Ho - 1° + a_0 + a_1 + a_2$$

The corrections a_0, a_1, and a_2, respectively, depend on LHA_{Aries} (estimated), the observer's estimated latitude, and the number of the current month. They are given in the Polaris Tables of the Nautical Almanac [12]. To extract the data, the observer has to know his approximate position and the approximate time.

Unfortunately, the Nautical Almanac does not provide GHA and Dec for Polaris. When using a computer almanac, however, we can find Lat with the following simple procedure. Lat_E is our estimated latitude, Dec is the declination of Polaris, and t_E is the estimated meridian angle of Polaris (calculated from GHA and our estimated longitude). Hc is the computed altitude, Ho is the observed altitude (chapter 4).

$$Hc \approx \arcsin\left(\sin Lat_E \cdot \sin Dec + \cos Lat_E \cdot \cos Dec \cdot \cos t_E\right)$$

$$\Delta H = Ho - Hc$$

Adding the altitude difference, ΔH, to the estimated latitude, we obtain the improved latitude:

$$Lat_{improved} \approx Lat_E + \Delta H$$

The improved latitude is accurate to 0.1' when Lat_E is smaller than ±70° and when the error of Lat_E is smaller than 2°, provided the exact longitude is known. At higher latitudes, the algorithm becomes less accurate and is not recommended. The method even tolerates a longitude error of up to 1°, in which case the resulting latitude error is still smaller than 1'. Latitude by Polaris is basically a variant of the ex-meridian sight. A rigorous procedure is given further below.

Noon latitude (latitude by maximum altitude)

This is a very simple method enabling the observer to determine the latitude of his position by measuring the maximum altitude of the Sun (or any other object). A very accurate time measurement is not required.

The altitude of the Sun passes through a flat maximum **approximately** (see noon longitude) at the moment of upper meridian passage (**local apparent noon, LAN**) when t equals 0 and the GP of the Sun is either north or south of the observer, depending on the declination of the Sun and the observer's geographic latitude.

The latter is easily calculated by forming the algebraic sum or difference of the declination and observed zenith distance z (90°-Ho) of the Sun, depending on whether the Sun is north or south of the observer (*Fig. 6-2*).

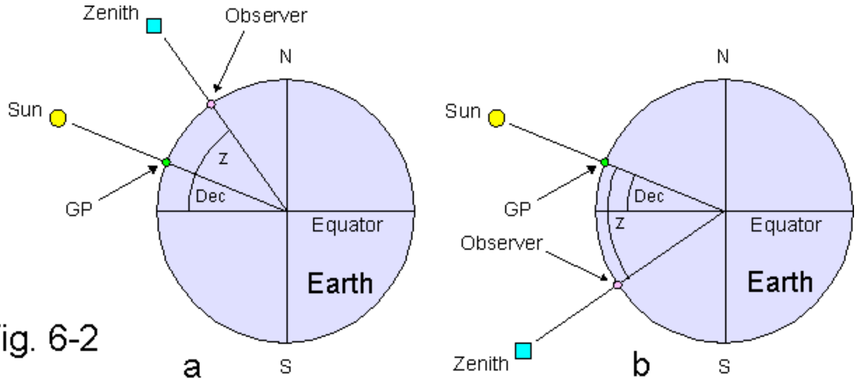


Fig. 6-2

1. Sun south of observer (*Fig. 6-2a*): $Lat = Dec + z = Dec - Ho + 90°$

2. Sun north of observer (*Fig. 6-2b*): $Lat = Dec - z = Dec + Ho - 90°$

Northern declination is positive, southern declination negative.

Before starting the observations, we need a rough estimate of our current longitude to know the time of meridian transit. We look up the time (UT) of Greenwich meridian transit of the Sun on the daily page of the Nautical Almanac and add 4 minutes for each degree of western longitude or subtract 4 minutes for each degree of eastern longitude. To determine the maximum altitude, we start observing the Sun approximately 15 minutes before the expected meridian transit. We follow the increasing altitude of the Sun with the sextant, note the maximum altitude when the Sun starts descending again, and apply the usual corrections. We look up the declination of the Sun at the approximate time (UT) of local meridian passage on the daily page of the Nautical Almanac and apply one of the above formulas.

The method may produce erratic results when the Sun culminates close to the zenith, in which case it is difficult to find if the Sun is north or south of the observer. Historically, noon latitude and latitude by Polaris are among the oldest methods of celestial navigation. With circumpolar bodies, latitude can also be found by minimum altitude.

Ex-meridian sight

Sometimes, it may be impossible to measure the maximum altitude of a body. For example, the latter may be obscured by a cloud at the instant of culmination. If we have a chance to measure the altitude some time before or after meridian transit, we are still able to find our latitude, provided we know the exact longitude of our position.

First, we use the law of sines for spherical triangles to calculate the azimuth angle from Dec, t, and Ho (see chapter 10 & 11). In contrast to the other azimuth formulas (see chapter 4), this one does not require the latitude to be known. The azimuth angle thus obtained is only used as an intermediate quantity here. It is not to be confused with the true Azimuth, Az_N!

$$\sin Az = \frac{\cos Dec \cdot \sin t}{\cos Ho} \qquad\qquad Az = \arcsin \frac{\cos Dec \cdot \sin t}{\cos Ho}$$

Since sin Az = sin (180°- Az), the above equation has two solutions, Az and 180°-Az. This corresponds to the fact that the circle of equal altitude usually intersects the local meridian at two points of different latitude. We enter the following formula which is based on Napier’s analogies (chapter 11), with Az. Afterwards, we repeat the calculation with the supplement angle, 180°-Az.

$$Lat = 90° - 2 \cdot \arctan \frac{\cos \dfrac{Az + t}{2}}{\tan \dfrac{Ho + Dec}{2} \cdot \cos \dfrac{Az - t}{2}}$$

Thus, we obtain two latitudes, Lat_{Az} and $Lat_{180°-Az}$. If either of these is greater than 90°, we replace it with its supplement angle, 180°−Lat.

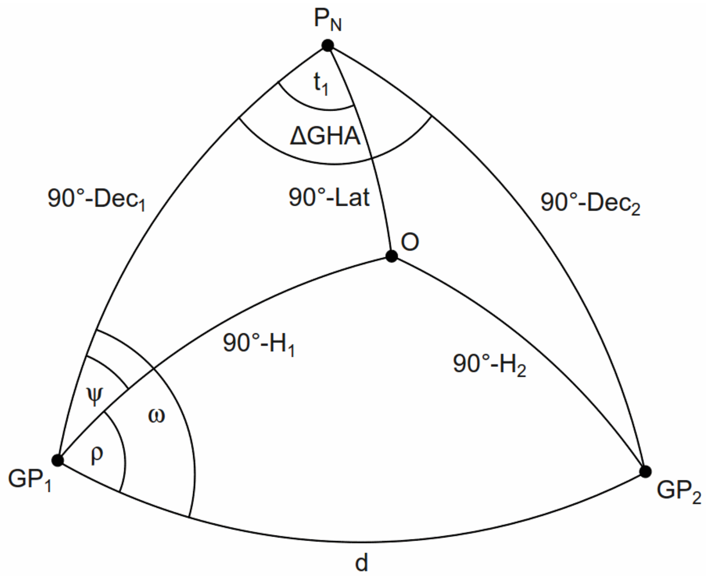
After that, we choose the angle which is nearest to our estimated latitude. Both solutions merge as Az approaches +90° or –90°. The sight has to be discarded when the difference between the two latitudes is too small for a clear distinction, depending on the reliability of our estimate. Critical judgment is required.

With the advent of position line navigation, the ex-meridian sight became more or less obsolete and is mainly of theoretical interest today. Usually, the navigator knows the latitude of his position better than the longitude. In that case, the latter can easily be found by a **time sight** (see further below).

Latitude by two altitudes

Even if the longitude is unknown, the exact latitude can still be found by observation of two celestial bodies. The required quantities are Greenwich hour angle (or sidereal hour angle), declination, and the observed altitude of each body [7]. The calculations are based upon spherical triangles (see chapter 10 & chapter 11). In *Fig. 6-3*, P_N denotes the north pole, O the observer's unknown position, GP_1 the geographic position of the first body, and GP_2 the position of the second body.

Fig. 6-3



First, we consider the spherical triangle [GP_1, P_N, GP_2]. *Fig. 6-3* shows only one of several possible configurations. O may as well be outside the triangle [GP_1, P_N, GP_2]. ΔGHA is the difference between both Greenwich hour angles which is equal to the difference between both sidereal hour angles.

$$\Delta GHA = \left| GHA_2 - GHA_1 \right| = \left| SHA_2 - SHA_1 \right|$$

Using the law of cosines for sides (chapter 10), we calculate d, the great circle distance between GP_1 and GP_2. We can use the absolute value of ΔGHA since cos (ΔGHA) = cos (-ΔGHA).

$$\cos d = \sin Dec_1 \cdot \sin Dec_2 + \cos Dec_1 \cdot \cos Dec_2 \cdot \cos(\Delta GHA)$$

$$d = \arccos\left[\sin Dec_1 \cdot \sin Dec_2 + \cos Dec_1 \cdot \cos Dec_2 \cdot \cos(\Delta GHA) \right]$$

Now we solve the same triangle for the angle ω, the horizontal distance between P_N and GP_2, measured at GP_1:

$$\cos \omega = \frac{\sin Dec_2 - \sin Dec_1 \cdot \cos d}{\cos Dec_1 \cdot \sin d}$$

$$\omega = \arccos \frac{\sin Dec_2 - \sin Dec_1 \cdot \cos d}{\cos Dec_1 \cdot \sin d}$$

For the spherical triangle [GP_1, O, GP_2], we calculate the angle ρ, the horizontal distance between O and GP_2, measured at GP_1.

$$\cos\rho = \frac{\sin H_2 - \sin H_1 \cdot \cos d}{\cos H_1 \cdot \sin d}$$

$$\rho = \arccos\frac{\sin H_2 - \sin H_1 \cdot \cos d}{\cos H_1 \cdot \sin d}$$

We calculate the angle ψ, the horizontal distance between P_N and O, measured at GP_1. There are two solutions, ψ_1 and ψ_2, since cos ρ = cos (-ρ):

$$\psi_1 = |\omega - \rho| \qquad \psi_2 = \omega + \rho$$

The circles of equal altitude intersect each other at two points. The corresponding positions are on opposite sides of the great circle going through GP_1 and GP_2 (not shown in *Fig. 6-3*). Using the law of cosines for sides again, we solve the spherical triangle [GP_1, P_N, O] for Lat. Since we have two solutions for ψ, we obtain two possible latitudes, Lat_1 and Lat_2.

$$\sin Lat_1 = \sin H_1 \cdot \sin Dec_1 + \cos H_1 \cdot \cos Dec_1 \cdot \cos\psi_1$$

$$Lat_1 = \arcsin\left(\sin H_1 \cdot \sin Dec_1 + \cos H_1 \cdot \cos Dec_1 \cdot \cos\psi_1\right)$$

$$\sin Lat_2 = \sin H_1 \cdot \sin Dec_1 + \cos H_1 \cdot \cos Dec_1 \cdot \cos\psi_2$$

$$Lat_2 = \arcsin\left(\sin H_1 \cdot \sin Dec_1 + \cos H_1 \cdot \cos Dec_1 \cdot \cos\psi_2\right)$$

We choose the value nearest to our estimated latitude and discard the other one. If both solutions are very similar and a clear distinction is not possible, one of the sights should be discarded, and a body with a more favorable position should be chosen.

Although the method requires more complicated calculations than, e. g., a latitude by Polaris, it has the advantage that measuring two altitudes usually takes less time than finding the maximum altitude of a single body. Moreover, if fixed stars are observed, even a chronometer error of several hours has no significant influence on the resulting latitude since sidereal hour angles and declinations of stars change rather slowly. If the exact time of observation is known, even the observer's longitude and, thus, his position can be calculated precisely (see end of chapter).

When the horizontal distance between the observed bodies is in the vicinity of 0° or 180°, the observer's position is close to the great circle going through $GP_1$and GP_2. In this case, the two solutions for latitude are similar, and finding which one corresponds with the actual latitude may be difficult (depending on the quality of the estimate). The resulting latitudes are also close to each other when the observed bodies have approximately the same Greenwich hour angle.

Noon longitude (longitude by equal altitudes)

The following method is based upon the equatorial coordinates of the apparent Sun. Thus, knowledge of the exact equation of time (see chapter 3) is not required here.

At the instant of local meridian transit (local apparent noon, LAN), the longitude of the geographic position of the Sun equals the observer‘s longitude. Thus, we only have to convert the Greenwich hour angle of the Sun to another format to obtain the corresponding longitude.

$$Lon = \left\{ \begin{array}{rcl} -GHA_{transit} & IF & GHA_{transit} < 180° \\ 360° - GHA_{transit} & IF & GHA_{transit} \geq 180° \end{array} \right\} = 180° - mod\left(GHA_{transit} + 180°, 360\right)$$

Unfortunately, the exact moment of LAN is very difficult to observe since the altitude of the Sun passes through a rather flat maximum. However, we can circumvent this problem by observing a chosen altitude of the ascending Sun a few hours before LAN and noting the exact moment of observation ($Date_1$, UT_1). After waiting a few hours, we note the instant at which we observe the same altitude of the descending Sun after LAN ($Date_2$, UT_2). For this purpose, we set our sextant or theodolite to a convenient altitude and note the two instants at which the chosen limb of the Sun crosses the reference line indicated by the instrument. Afterwards, we find the corresponding values for Greenwich hour angle and declination of the Sun in the almanac. We denote the values corresponding with $Date_1$ and UT_1 as GHA_1 and Dec_1 and the values corresponding with $Date_2$ and UT_2 as GHA_2 and Dec_2, respectively.

$$Date_1\ ,\ UT_1 \quad \rightarrow \quad GHA_1\ ,\ Dec_1$$
$$Date_2\ ,\ UT_2 \quad \rightarrow \quad GHA_2\ ,\ Dec_2$$

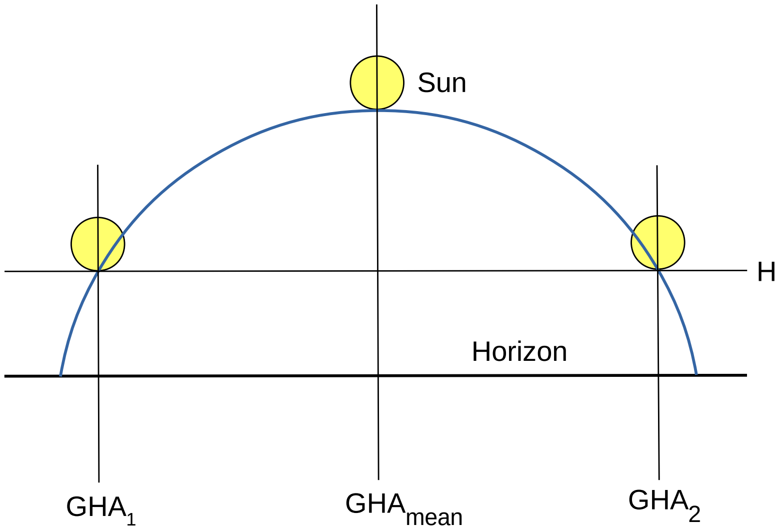
The altitude value itself is not of interest (no altitude corrections either, provided there are no extreme changes in temperature or atmospheric pressure between the observations). Only the corresponding GHA and Dec values are needed for the calculations.

Assuming a stationary observer and a constant declination of the Sun (solstices), the arc described by the apparent Sun in the sky is almost symmetrical and $GHA_{transit}$ is approximately the arithmetic mean of GHA_1 and GHA_2 (*Fig. 6-4*).*

$$GHA_{transit} \approx GHA_{mean} = \left\{ \begin{array}{lll} \dfrac{GHA_1 + GHA_2}{2} & IF & GHA_2 > GHA_1 \\ \dfrac{GHA_1 + GHA_2}{2} - 180° & IF & GHA_2 < GHA_1 \end{array} \right\}$$

* GHA_2 is smaller than GHA_1 when the calendar date changes between both observations ($Date_2 > Date_1$).

Fig. 6-4



This would be the ideal case. In reality, the observer‘s position often changes during the observation interval, particularly at sea. The change of position, defined by ΔLon and ΔLat, can be calculated from course and average speed over ground during the observation interval, UT_2-UT_1 (see further below). Further, the declination of the Sun changes measurably during the greater part of the year, the rate of change being greatest around the times of the equinoxes (up to approx. ±1′/h). Each of these influencing factors requires a correction to be applied.

1st correction (change of longitude):

If the ship moves westward (ΔLon < 0), the apparent motion of the Sun is slower than seen from a stationary point of observation. Accordingly, an observer moving eastward (ΔLon > 0) will notice a faster motion of the Sun in the sky. We correct for ΔLon as follows:

$$GHA_2^* = GHA_2 + \Delta Lon$$

2nd correction (change of latitude):

The longitude error caused by a change in latitude can be dramatic and requires the navigator's particular attention, even if the vessel travels at a moderate speed. Let us assume the ship moves along a meridian during the observation interval. Lat_1 be the observer‘s latitude at UT_1 and Lat_2 (obtained by dead reckoning) be the latitude at UT_2. If the exact value of Lat_1 is not known, an estimated value may be used. Using the **equation of equal altitudes** [5], we calculate the second correction to be applied to GHA_2, ΔGHA_{Lat}. For this purpose, we first have to find the approximate meridian angle of the Sun corresponding with GHA_2^*:

$$t \approx \frac{GHA_2^* - GHA_1}{2}$$

The equation of equal altitudes describes the change of hour angle equivalent to a small change of latitude:

$$\Delta GHA_{Lat} \;=\; \Delta t \;\approx\; \left(\frac{\tan Dec_{mean}}{\sin t} - \frac{\tan Lat_{mean}}{\tan t} \right) \cdot \Delta Lat$$

$$\Delta Lat \;=\; Lat_2 - Lat_1 \qquad Lat_{mean} \;=\; \frac{Lat_1 + Lat_2}{2} \qquad Dec_{mean} \;=\; \frac{Dec_1 + Dec_2}{2}$$

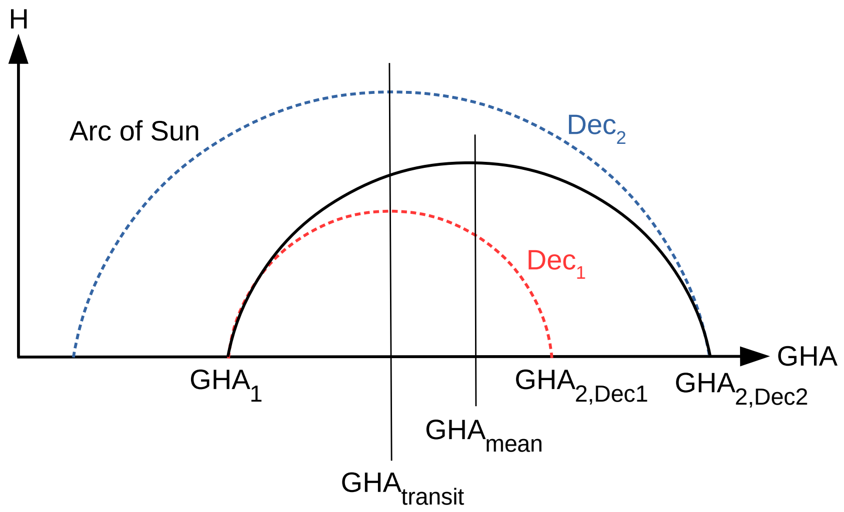
Now, we have the Greenwich hour angle of the Sun corrected for a change of position defined by ΔLon and ΔLat.

$$GHA_2^{**} \;=\; GHA_2^{*} - \Delta GHA_{Lat}$$

3rd correction (change of declination):

Fig. 6-5 is a plot of the altitude of the Sun versus the GHA of the latter. It demonstrates how the shape of the apparent arc described by the Sun is affected when the declination is not constant but changes from Dec_1 to Dec_2 between both observations. The example shows a scenario (strongly exaggerated) in which the declination of the Sun changes toward the observer‘s latitude during the interval of observation.

Fig. 6-5



The dotted red line shows the path of the Sun for a given constant declination, Dec_1. The dotted blue line shows how the path would look with a different declination, Dec_2. In both cases, the apparent path of the Sun is symmetrical with respect to $GHA_{transit}$ and the latter equals GHA_{mean}. However, if the Sun's declination varies from Dec_1 to Dec_2 during the observation interval, the path shown by the continuous black line will result. Now, there is a measurable difference between $GHA_{transit}$ and GHA_{mean}, and without further corrections there would be a significant error in the longitude thus found. A change of the observer‘s latitude (see 2nd correction) toward the declination of the Sun causes a similar effect.

The resulting error in longitude is greatest around the times of the equinoxes when the rate of change of Dec is at its maximum. Moreover, the error increases with the observer's distance from the equator and may be quite dramatic in polar regions.

The third correction to be applied to GHA_2, ΔGHA_{Dec}, is calculated with another version of the **equation of equal altitudes** [5]:

$$\Delta GHA_{Dec} \;=\; \Delta t \;\approx\; \left(\frac{\tan Lat_{mean}}{\sin t} - \frac{\tan Dec_{mean}}{\tan t} \right) \cdot \Delta Dec$$

$$\Delta Dec \;=\; Dec_2 - Dec_1$$

Applying the third correction, we get

$$GHA_2^{***} \;=\; GHA_2^{**} - \Delta GHA_{Dec}$$

The improved value for GHA**transit** , including the three corrections, is

$$GHA_{\text{transit, impr.}} \;\approx\; \frac{GHA_1 + GHA_2^{***}}{2}$$

Finally, we have

$$Lon_1 = \left\{ \begin{array}{lll} -\,GHA_{\text{transit, impr.}} & IF & GHA_{\text{transit, impr.}} < 180° \\ 360° - GHA_{\text{transit, impr.}} & IF & GHA_{\text{transit, impr.}} \geq 180° \end{array} \right\} = 180° - mod\left(GHA_{\text{transit, impr.}} + 180°,\; 360\right)$$

and

$$Lon_2 \;=\; Lon_1 + \Delta Lon$$

Although this is not a rigorous method, it is quite accurate. At 80° latitude (+ or −), the inherent error is a small fraction of an arcsecond (≈ ±0.1"). As we move towards the equator, the error becomes even smaller.

Calculating with the Greenwich hour angle of the (apparent) Sun instead of UT has the advantage that we do not need to know the current values for the equation of time.

ΔLon and ΔLat are calculated from course, C, and average speed, v, as follows (see chapter 12):

$$\Delta Lat['] \;=\; v[kn] \cdot \cos C \cdot \left(UT_2[h] - UT_1[h]\right)$$

$$Lat_2 \;=\; Lat_1 + \Delta Lat$$

$$\Delta Lon['] \;\approx\; v[kn] \cdot \frac{\sin C}{\cos Lat_{mean}} \cdot \left(UT_2[h] - UT_1[h]\right)$$

$$Lon_2 \;=\; Lon_1 + \Delta Lon$$

$$1\,kn\,(knot) \;=\; 1\; nm/h$$

C is measured clockwise from true north (0°...360°).

The above considerations clearly demonstrate that determining one's **exact** longitude by equal altitudes of the Sun is not as simple as it seems to be at first glance, particularly on a moving vessel. It is therefore understandable that with the development of position line navigation (including simple graphic solutions for a traveling vessel), longitude by equal altitudes became less important.

Under certain circumstances (high latitudes, arctic or antarctic summer), it may be useful to observe the Sun a few hours before and after **local apparent midnight** (lower local meridian transit). To obtain the correct longitude in this case, we have to increase the meridian angle, t, by 180° before calculating the 2nd and 3rd correction and subtract 180° from the longitude thus obtained.

Compared with a time sight (see farther below), a longitude by equal altitudes appears more cumbersome. However, a time sight requires knowledge of the exact latitude, otherwise a large longitude error may result. In contrast, an estimated latitude is sufficient for a longitude by equal altitudes since a latitude error will only influence the relatively small corrections for changes in latitude and/or declination.

Observation planning:

For observation planning, the approximate time of meridian transit should be known. It can be calculated using the following formula. Estimated values for equation of time and geographic longitude are sufficient for this purpose.

$$UT_{transit} \;\approx\; 12\,h \;-\; EoT[h] \;-\; \frac{Lon[°]}{15}$$

Theory of the equation of equal altitudes

The equation of equal altitudes is derived from the altitude formula (see chapter 4) using differential calculus.

1st case: latitude change with constant declination

Altitude formula:

$$\sin H = \sin Lat \cdot \sin Dec + \cos Lat \cdot \cos Dec \cdot \cos t$$

First, we need to know how a small change in latitude would affect sin H. We form the partial derivative with respect to Lat:

$$\frac{\partial(\sin H)}{\partial Lat} = \cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t$$

Thus, the change in sin H caused by an infinitesimal change in latitude, d Lat, is:

$$\frac{\partial(\sin H)}{\partial Dec} \cdot d\,Lat = (\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t) \cdot d\,Lat$$

Now, we form the partial derivative with respect to t in order to find out how a small change in the meridian angle would affect sin H:

$$\frac{\partial(\sin H)}{\partial t} = -\cos Lat \cdot \cos Dec \cdot \sin t$$

Since the combined effects of latitude and meridian angle cancel each other with respect to their influence on sin H, the total differential is zero:

$$\frac{\partial(\sin H)}{\partial Lat} \cdot d\,Lat + \frac{\partial(\sin H)}{\partial t} \cdot dt = 0$$

$$-\frac{\partial(\sin H)}{\partial t} \cdot d\,t = \frac{\partial(\sin H)}{\partial Lat} \cdot d\,Lat$$

$$\cos Lat \cdot \cos Dec \cdot \sin t \cdot d\,t = (\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t) \cdot d\,Lat$$

Solving for dt, we obtain the following formula for an infinitesimally small change in latitude:

$$d\,t = \frac{\cos Lat \cdot \sin Dec - \sin Lat \cdot \cos Dec \cdot \cos t}{\cos Lat \cdot \cos Dec \cdot \sin t} \cdot d\,Lat$$

$$dt = \left(\frac{\tan Dec}{\sin t} - \frac{\tan Lat}{\tan t}\right) \cdot d\,Lat$$

For a measurable small change in latitude, we get

$$\Delta t \approx \left(\frac{\tan Dec}{\sin t} - \frac{\tan Lat_{mean}}{\tan t}\right) \cdot \Delta Lat$$

2nd case: declination change with constant latitude

The calculations are done in the same way as with a changing latitude.

$$\sin H = \sin Lat \cdot \sin Dec + \cos Lat \cdot \cos Dec \cdot \cos t$$

Now, we need to know how a small change in declination would affect sin H.

We form the partial derivative with respect to Dec:

$$\frac{\partial(\sin H)}{\partial\, Dec} \;=\; \sin Lat \cdot \cos Dec \,-\, \cos Lat \cdot \sin Dec \cdot \cos t$$

Thus, the change in sin H caused by an infinitesimal change in declination, d Dec, is:

$$\frac{\partial(\sin H)}{\partial\, Dec} \cdot d\,Dec \;=\; \left(\sin Lat \cdot \cos Dec - \cos Lat \cdot \sin Dec \cdot \cos t\right) \cdot d\,Dec$$

As shown above, we form the partial derivative with respect to t in order to find out how a small change in the meridian angle would affect sin H:

$$\frac{\partial(\sin H)}{\partial t} \;=\; -\cos Lat \cdot \cos Dec \cdot \sin t$$

Again, the change in sin H caused by an infinitesimal change in the meridian angle, dt, is:

$$\frac{\partial(\sin H)}{\partial t} \cdot d\,t \;=\; -\cos Lat \cdot \cos Dec \cdot \sin t \cdot d\,t$$

Since both effects cancel each other, the total differential is zero:

$$\frac{\partial(\sin H)}{\partial\, Dec} \cdot d\;Dec + \frac{\partial(\sin H)}{\partial t} \cdot d\,t \;=\; 0$$

$$-\frac{\partial(\sin H)}{\partial\, t} \cdot d\,t \;=\; \frac{\partial(\sin H)}{\partial\, Dec} \cdot d\,Dec$$

$$\cos Lat \cdot \cos Dec \cdot \sin t \cdot d\,t \;=\; \left(\sin Lat \cdot \cos Dec - \cos Lat \cdot \sin Dec \cdot \cos t\right) \cdot d\,Dec$$

$$d\,t \;=\; \frac{\sin Lat \cdot \cos Dec - \cos Lat \cdot \sin Dec \cdot \cos t}{\cos Lat \cdot \cos Dec \cdot \sin\, t} \cdot d\,Dec$$

Thus, we get the formula for a small change in declination:

$$d\,t \;=\; \left(\frac{\tan Lat}{\sin\, t} - \frac{\tan\, Dec}{\tan\, t}\right) \cdot d\;Dec$$

$$\Delta t \;\approx\; \left(\frac{\tan Lat}{\sin\, t} - \frac{\tan\, Dec_{mean}}{\tan\, t}\right) \cdot \Delta\, Dec$$

The meridian angle of the Sun at maximum altitude (stationary observer)

Fig. 6-5 indicates that the maximum altitude of the Sun is slightly different from the altitude at the moment of meridian passage if the declination changes.

At maximum altitude, the rate of change of altitude caused by the changing declination cancels the rate of change of altitude caused by the changing meridian angle. The equation of equal altitude enables us to calculate the meridian angle of the Sun at this moment.

$$d\,t \;=\; \left(\frac{\tan Lat}{\sin\;t} - \frac{\tan\,Dec}{\tan\;t}\right)\cdot d\,Dec$$

Dividing by dDec, we get

$$\frac{d\,t}{d\,Dec} \;=\; \frac{\tan Lat}{\sin\;t} - \frac{\tan Dec}{\tan\;t}$$

Since in <u>this</u> case t is a very small angle, we can substitute tan t for sin t (or vice versa).

$$\frac{d\,t}{d\,Dec} \;\approx\; \frac{\tan Lat - \tan Dec}{\tan t}$$

Solving for tan t, we get

$$\tan t \;\approx\; (\tan Lat - \tan Dec)\cdot\frac{d\,Dec}{d\,t}$$

$$t \;\approx\; \arctan\left[(\tan Lat - \tan Dec)\cdot\frac{d\,Dec}{d\,t}\right]$$

Measuring the rate of change of declination in arcminutes per hour (1 h ≙ 15°), we get

$$t \;\approx\; \arctan\left[(\tan Lat - \tan Dec)\cdot\frac{1}{900}\cdot\frac{d\,Dec[']}{d\,t[h]}\right]$$

For example, at the time of the spring equinox (Dec ≈ 0, dDec/dT ≈ +1'/h) an observer being at +80° (N) latitude would observe the maximum altitude of the Sun at t ≈ +21.7', i. e., 86.6 seconds after local meridian transit (LAN). An observer at +45° latitude, however, would observe the maximum altitude at t ≈ +3.8', i. e., only 15.3 seconds after meridian transit.

The maximum altitude of the Sun

We can use the last equation to estimate the systematic error of a noon latitude. The latter is based upon the maximum altitude of the Sun which may slightly differ from the altitude at the moment of meridian transit. Following the above example, the observer at 80° latitude would observe the maximum altitude at t = 21,7' .

Between meridian transit (t = 0) and t = 21.7' , the declination of the Sun would have changed by ΔDec:

$$\Delta\,Dec \;\approx\; \frac{Dec_2 - Dec_1}{GHA_2 - GHA_1}\cdot t$$

Using the diffentiated form of the altitude formula, we get

$$\Delta\,H \;\approx\; \frac{\sin Lat\,\cdot\,\cos Dec_{mean} - \cos Lat\,\cdot\sin Dec_{mean}\cdot\cos t}{\cos H_{transit}}\cdot\Delta\,Dec$$

$$H_{transit} \approx\; 90° - |\,Lat - Dec_{mean}\,|$$

For the above example (Lat = 80°), a maximum altitude of approx. 10° 00' 00.7'' instead of exactly 10° would result.

This demonstrates that even at the times of the equinoxes, the systematic error of a noon latitude caused by the changing declination of the Sun is not significant because it is much smaller than other observational errors, e. g., the errors in dip or refraction.

A measurable error in latitude can only occur if the observer is very close to one of the poles (tan Lat!). Around the times of the solstices, the error in latitude is practically non-existent.

Time sight

The process of deriving the longitude from a single altitude of a body (as well as the observation made for this purpose) is called **time sight**.
In contrast to a noon longitude, this method requires knowledge of the exact (!) latitude, e. g., a noon latitude or a latitude by two altitudes. Solving the navigational triangle (chapter 11) for the meridian angle, t, we get:

$$t = \pm \arccos \frac{\sin Ho - \sin Lat \cdot \sin Dec}{\cos Lat \cdot \cos Dec}$$

Alternatively, we can use this formula which is based on the haversine formula (chapter 10 & 11):

$$t = \pm \arccos \left(1 - \frac{\cos(Lat - Dec) - \sin Ho}{\cos Lat \cdot \cos Dec} \right)$$

Both equations have two solutions, +t and –t, since cos t = cos (–t). Geometrically, this corresponds with the fact that the circle of equal altitude intersects the parallel of latitude at two points with a longitude difference of 2·t.

Using the following formulas and rules, we obtain the longitudes of these points of intersection, Lon_1 and Lon_2:

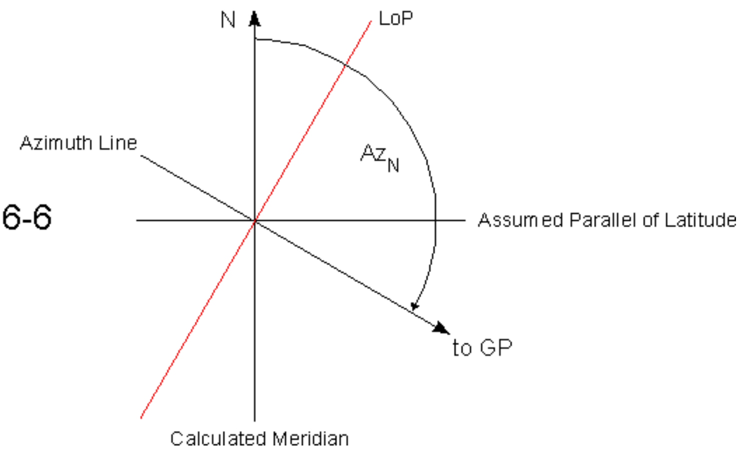
$$Lon_1 = t - GHA$$

$$Lon_2 = 360° - t - GHA$$

$$\begin{array}{llll} \text{If} & Lon_1 < -180° & \rightarrow & Lon_1 + 360° \\ \text{If} & Lon_2 < -180° & \rightarrow & Lon_2 + 360° \\ \text{If} & Lon_2 > +180° & \rightarrow & Lon_2 - 360° \end{array}$$

Even if we do not know the exact latitude, we can still use a time sight to derive a line of position from an assumed or estimated latitude. After solving the time sight, we plot the assumed parallel of latitude and the calculated meridian. Next, we calculate the azimuth of the body with respect to the position thus obtained (azimuth formulas, chapter 4) and plot the azimuth line. Our line of position is the perpendicular of the azimuth line going through the calculated position (*Fig. 6-6*).

Fig. 6-6



Finding a LOP by time sight is mainly of historical interest. Today, most navigators prefer the intercept method (chapter 4) which can be used without any restrictions regarding meridian angle (local hour angle), latitude, and declination (see below). A time sight is not reliable when the body is close to the meridian. Using differential calculus, we can demonstrate that the error of the meridian angle, dt, resulting from an altitude error, dH, varies in proportion with 1/sin t:

$$dt = -\frac{\cos Ho}{\cos Lat \cdot \cos Dec \cdot \sin t} \cdot dH$$

Moreover, dt varies inversely with cos Lat and cos Dec. Therefore, high latitudes and declinations should be avoided as well.

Last but not least, a small meridian angle increases the risk of choosing the wrong point of intersection on the parallel of latitude (ambiguity).

Sumner'*s* method of finding a position line is based on two time sights (chapter 4).

Direct computation of a position

If we know the exact time, the observations for a **latitude by two altitudes** even enable us to calculate our position directly, without graphic plot. After obtaining our latitude, Lat, from two altitudes (see above), we use the **time sight** formula to calculate the meridian angle of one of the bodies. In case of the first body, for example, we calculate $\pm t_1$ from the quantities Lat, Dec_1, and H_1 (see *Fig. 6-3*). Two possible longitudes result from the meridian angles thus obtained. We choose the one nearest to our estimated longitude. This is a **rigorous method**, not an approximation.

In a similar manner, a position can be found by combining a **noon latitude** with a **noon longitude**.

Direct computation is rarely used since the calculations are more complicated than those required for graphic solutions. Of course, in the age of computers the complexity of the method does not present a problem anymore.

Chapter 7

Finding Time and Longitude by Lunar Distances

In celestial navigation, time and longitude are interdependent. Finding one's longitude at sea or in unknown terrain is impossible without knowing the exact time and vice versa. Therefore, old-time navigators were basically restricted to latitude sailing on long voyages, i. e., they had to sail along a chosen parallel of latitude until they came in sight of the coast. Since there was no reliable estimate of the time of arrival, many ships ran ashore during periods of darkness or bad visibility. Spurred by heavy losses of men and material, scientists tried to solve the longitude problem by using astronomical events as time marks. In principle, such a method is only suitable when the observed time of the event is virtually independent of the observer's geographic position.

Measuring time by the apparent movement of the Moon with respect to the background of fixed stars was suggested in the 15th century already (*Regiomontanus*) but proved impracticable since neither reliable ephemerides for the Moon nor precise instruments for measuring angles were available at that time.

Around the middle of the 18th century, astronomy and instrument making had finally reached a stage of development that made time measurement by lunar observations possible. Particularly, deriving the time from a so-called **lunar distance**, the angular distance of the Moon from a chosen reference body, became a popular method. Although the procedure is rather cumbersome, it became an essential part of celestial navigation and was used far into the 19th century, long after the invention of the mechanical chronometer (*Harrison*, 1736). This was mainly due to the limited availability of reliable chronometers and their exorbitant price. When chronometers became affordable around the middle of the 19th century, lunar distances gradually went out of use. Until 1906, the Nautical Almanac included lunar distance tables showing predicted geocentric angular distances between the Moon and selected bodies in 3-hour intervals.* After the tables were dropped, lunar distances fell more or less into oblivion. Not much later, radio time signals became available world-wide, and the longitude problem was solved once and for all. Today, lunar distances are mainly of historical interest. The method is so ingenious, however, that a detailed study is worthwhile.

The basic idea of the lunar distance method is easy to comprehend. Since the Moon moves across the celestial sphere at a rate of about 0.5° per hour, the angular distance between the Moon, M, and a body in her path, B, varies at a similar rate and rapidly enough to be used to measure the time. The time corresponding with an observed lunar distance can be found by comparison with tabulated values.

Tabulated lunar distances are calculated from the geocentric equatorial coordinates of M and B using the cosine law:

$$\cos D = \sin Dec_M \cdot \sin Dec_B + \cos Dec_M \cdot \cos Dec_B \cdot \cos(GHA_M - GHA_B)$$

or

$$\cos D = \sin Dec_M \cdot \sin Dec_B + \cos Dec_M \cdot \cos Dec_B \cdot \cos\left[15 \cdot \left(RA_M[h] - RA_B[h]\right)\right]$$

D is the geocentric lunar distance. These formulas can be used to set up one's own table with the aid of the Nautical Almanac or any computer almanac if a lunar distance table is not available.

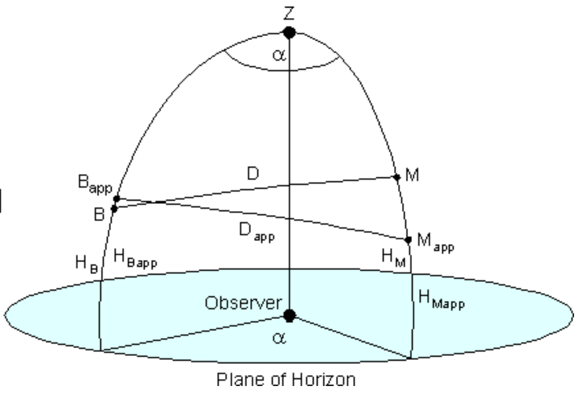
*Nowadays, it is possible to set up precomputed lunar distance tables for chosen bodies with the aid of a computer almanac and one of the above formulas.

Clearing the lunar distance

Before a lunar distance measured by the observer can be compared with tabulated values, it has to be reduced to the corresponding geocentric angle by clearing it from the effects of refraction and parallax. This essential process is called **clearing the lunar distance**. Numerous procedures have been developed, among them rigorous and "quick" methods. In the following, we will discuss the almost identical methods by *Dunthorne* (1766) and *Young* (1856). They are rigorous for a spherical model of the Earth.

Fig. 7-1 shows the positions of the Moon, M, and a reference body, B, in the coordinate system of the horizon. We denote the apparent positions of the centers of the Moon and the reference body by M_{app} and B_{app}, respectively. Z is the zenith.

Fig. 7-1



The side D_{app} of the spherical triangle B_{app}-Z-M_{app} is the apparent lunar distance. The altitudes of M_{app} and B_{app} (obtained after applying the corrections for index error, dip, and semidiameter) are H_{Mapp} and H_{Bapp}, respectively. The vertical circles of both bodies form the angle α, the difference between the azimuth of the Moon, Az_M, and the azimuth of the reference body, Az_B:

$$\alpha = Az_M - Az_B$$

The position of each body is shifted along its vertical circle by atmospheric refraction and parallax in altitude. After correcting H_{Mapp} and H_{Bapp} for both effects, we obtain the geocentric positions M and B. We denote the altitude of M by H_M and the altitude of B by H_B. H_M is always greater than H_{Mapp} because the parallax of the Moon is always greater than refraction. The angle α is neither affected by refraction nor by the parallax in altitude:

$$Az_M = Az_{Mapp} \qquad Az_B = Az_{Bapp}$$

The side D of the spherical triangle B-Z-M is the unknown geocentric lunar distance. If we knew the exact value for α, calculation of D would be very simple (cosine law). Unfortunately, the navigator has no means for measuring α precisely. It is possible, however, to calculate D solely from the five quantities D_{app}, H_{Mapp}, H_M, H_{Bapp}, and H_B.

Applying the cosine formula to the spherical triangle formed by the zenith and the apparent positions, we get:

$$\cos D_{app} = \sin H_{Mapp} \cdot \sin H_{Bapp} + \cos H_{Mapp} \cdot \cos H_{Bapp} \cdot \cos\alpha$$

$$\cos\alpha = \frac{\cos D_{app} - \sin H_{Mapp} \cdot \sin H_{Bapp}}{\cos H_{Mapp} \cdot \cos H_{Bapp}}$$

Repeating the procedure with the spherical triangle formed by the zenith and the geocentric positions, we get:

$$\cos D = \sin H_M \cdot \sin H_B + \cos H_M \cdot \cos H_B \cdot \cos\alpha$$

$$\cos\alpha = \frac{\cos D - \sin H_M \cdot \sin H_B}{\cos H_M \cdot \cos H_B}$$

Since α is constant, we can combine both azimuth formulas:

$$\frac{\cos D - \sin H_M \cdot \sin H_B}{\cos H_M \cdot \cos H_B} = \frac{\cos D_{app} - \sin H_{Mapp} \cdot \sin H_{Bapp}}{\cos H_{Mapp} \cdot \cos H_{Bapp}}$$

Thus, we have eliminated the unknown angle α. Now, we subtract unity from both sides of the equation:

$$\frac{\cos D - \sin H_M \cdot \sin H_B}{\cos H_M \cdot \cos H_B} - 1 = \frac{\cos D_{app} - \sin H_{Mapp} \cdot \sin H_{Bapp}}{\cos H_{Mapp} \cdot \cos H_{Bapp}} - 1$$

$$\frac{\cos D - \sin H_M \cdot \sin H_B}{\cos H_M \cdot \cos H_B} - \frac{\cos H_M \cdot \cos H_B}{\cos H_M \cdot \cos H_B} = \frac{\cos D_{app} - \sin H_{Mapp} \cdot \sin H_{Bapp}}{\cos H_{Mapp} \cdot \cos H_{Bapp}} - \frac{\cos H_{Mapp} \cdot \cos H_{Bapp}}{\cos H_{Mapp} \cdot \cos H_{Bapp}}$$

$$\frac{\cos D - \sin H_M \cdot \sin H_B - \cos H_M \cdot \cos H_B}{\cos H_M \cdot \cos H_B} = \frac{\cos D_{app} - \sin H_{Mapp} \cdot \sin H_{Bapp} - \cos H_{Mapp} \cdot \cos H_{Bapp}}{\cos H_{Mapp} \cdot \cos H_{Bapp}}$$

Using the addition formula for cosines, we have:

$$\frac{\cos D - \cos(H_M - H_B)}{\cos H_M \cdot \cos H_B} = \frac{\cos D_{app} - \cos(H_{Mapp} - H_{Bapp})}{\cos H_{Mapp} \cdot \cos H_{Bapp}}$$

Solving for cos D, we obtain *Dunthorne's* formula for clearing the lunar distance:

$$\cos D = \frac{\cos H_M \cdot \cos H_B}{\cos H_{Mapp} \cdot \cos H_{Bapp}} \cdot \left[\cos D_{app} - \cos(H_{Mapp} - H_{Bapp})\right] + \cos(H_M - H_B)$$

Adding unity to both sides of the equation instead of subtracting it, leads to *Young's* formula:

$$\cos D = \frac{\cos H_M \cdot \cos H_B}{\cos H_{Mapp} \cdot \cos H_{Bapp}} \cdot \left[\cos D_{app} + \cos(H_{Mapp} + H_{Bapp})\right] - \cos(H_M + H_B)$$

Procedure

Deriving UT from a lunar distance comprises the following steps:

1.

We measure the altitude of the upper or lower limb of the Moon, whichever is visible, and note the time of the observation indicated by our watch, $WT1_{LMapp}$.
We apply the corrections for index error and dip (if necessary) and get the apparent altitude of the limb, $H1_{LMapp}$. We repeat the procedure with the reference body and obtain the watch time $WT1_{Bapp}$ and the altitude $H1_{Bapp}$.

2.

We measure the angular distance between the limb of the Moon and the reference body, D_{Lapp}, and note the corresponding watch time, WT_D. The angle D_{Lapp} has to be measured with the greatest possible precision. It is recommended to measure a few D_{Lapp} values and their corresponding WT_D values in rapid succession and calculate the respective average value. When the Moon is almost full, it is not quite easy to distinguish the limb of the Moon from the terminator (shadow line). In general, the limb has a sharp appearance whereas the terminator is slightly indistinct.

3.

We measure the altitudes of both bodies again, as described above. We denote them by $H2_{LMapp}$ and $H2_{Bapp}$, and note the corresponding watch times of observation, $WT2_{LMapp}$ and $WT2_{Bapp}$.

4.

Since the observations are only a few minutes apart, we can calculate the altitude of the respective body at the moment of the lunar distance observation by linear interpolation:

$$H_{LMapp} = H1_{LMapp} + (H2_{LMapp} - H1_{LMapp}) \cdot \frac{WT_D - WT1_{LMapp}}{WT2_{LMapp} - WT1_{LMapp}}$$

$$H_{Bapp} = H1_{Bapp} + (H2_{Bapp} - H1_{Bapp}) \cdot \frac{WT_D - WT1_{Bapp}}{WT2_{Bapp} - WT1_{Bapp}}$$

5.

We correct the altitude of the Moon and the angular distance D_{Lapp} for the augmented semidiameter of the Moon, SD_{aug}. The latter can be calculated directly from the altitude of the upper or lower limb of the Moon:

$$\tan SD_{aug} = \frac{k}{\left[\sqrt{\frac{1}{\sin^2 HP_M} - (\cos H_{LMapp} \pm k)^2}\right] - \sin H_{LMapp}} \qquad k = 0.2725$$

$$\text{upper limb:} \quad \cos H_{LMapp} - k \qquad\qquad \text{lower limb:} \quad \cos H_{LMapp} + k$$

The altitude correction is:

$$\text{Lower limb:} \quad H_{Mapp} = H_{LMapp} + SD_{aug}$$
$$\text{Upper limb:} \quad H_{Mapp} = H_{LMapp} - SD_{aug}$$

The rules for the lunar distance correction are:

$$\text{Limb of moon towards reference body:} \quad D_{app} = D_{Lapp} + SD_{aug}$$
$$\text{Limb of moon away from reference body:} \quad D_{app} = D_{Lapp} - SD_{aug}$$

The above procedure is an approximation since the augmented semidiameter is a function of the altitude corrected for refraction. Since refraction is a small quantity and since the total augmentation between 0° and 90° altitude is only approx. 0.3’, the resulting error is very small and may be ignored.

The Sun, when chosen as reference body, requires the same corrections for semidiameter. Since the Sun does not show a measurable augmentation, we can use the geocentric semidiameter tabulated in the Nautical Almanac or calculated with a computer program.

6.

We correct both altitudes, H_{Mapp} and H_{Bapp}, for atmospheric refraction, R.

$$R_i['] = \frac{p[mbar]}{1010} \cdot \frac{283}{T[°C] + 273} \cdot \left(\frac{0.96474}{\tan H_i} - \frac{0.00113}{\tan^3 H_i}\right) \qquad i = Mapp, Bapp \qquad H_i > 10°$$

R_i is subtracted from the respective altitude. The refraction formula is only accurate for altitudes above approx. 10°. Lower altitudes should be avoided anyway since refraction may become erratic and since the apparent disk of the Moon (and Sun) assumes an oval shape caused by an increasing difference in refraction for upper and lower limb. This distortion would affect the semidiameter with respect to the reference body in a complicated way.

7.

We correct the altitudes for the parallax in altitude:

$$\sin P_M = \sin HP_M \cdot \cos(H_{Mapp} - R_{Mapp}) \qquad \sin P_B = \sin HP_B \cdot \cos(H_{Bapp} - R_{Bapp})$$

We apply the altitude corrections as follows:

$$H_M = H_{Mapp} - R_{Mapp} + P_M \qquad H_B = H_{Bapp} - R_{Bapp} + P_B$$

The correction for parallax is not applied to the altitude of a fixed star (HPB = 0).

8.

With D_{app}, H_{Mapp}, H_M, H_{Bapp}, and H_B, we calculate D using *Dunthorne's* or *Young's* formula.

9.

The time corresponding with the geocentric distance D is found by interpolation. Lunar distance tables show D as a function of time, T (UT). If the rate of change of D does not vary too much (less than approx. 0.3' in 3 hours), we can use linear interpolation. However, in order to find T, we have to consider T as a function of D (inverse interpolation).

$$T_D = T_1 + (T_2 - T_1) \cdot \frac{D - D_1}{D_2 - D_1}$$

T_D is the unknown time corresponding with D. D_1 and D_2 are tabulated lunar distances. T_1 and T_2 are the corresponding time (UT) values ($T_2 = T_1 + 3h$). D is the geocentric lunar distance calculated from D_{app}. D has to be between D_1 and D_2.

If the rate of change of D varies significantly, more accurate results are obtained with methods for non-linear interpolation, for example, with 3-point *Lagrange* interpolation. Choosing three pairs of tabulated values, (T_1, D_1), (T_2, D_2), and (T_3, D_3), T_D is calculated as follows:

$$T_D = T_1 \cdot \frac{(D - D_2) \cdot (D - D_3)}{(D_1 - D_2) \cdot (D_1 - D_3)} + T_2 \cdot \frac{(D - D_1) \cdot (D - D_3)}{(D_2 - D_1) \cdot (D_2 - D_3)} + T_3 \cdot \frac{(D - D_1) \cdot (D - D_2)}{(D_3 - D_1) \cdot (D_3 - D_2)}$$

$$T_2 = T_1 + 3\text{h} \qquad T_3 = T_2 + 3\text{h} \qquad D_1 < D_2 < D_3 \quad \text{or} \quad D_1 > D_2 > D_3$$

D may have any value between D_1 and D_3.

There must not be a minimum or maximum of D in the time interval [T_1, T_3]. This problem does not occur with a properly chosen body having a suitable rate of change of D. Near a minimum or maximum of D, ΔD/ΔT would be very small, and the observation would be erratic anyway.

After finding T_D, we can calculate the watch error, ΔT.

$$\Delta T = WT_D - T_D$$

ΔT is the difference between our watch time at the moment of observation, WT_D, and the time found by interpolation, T_D.

Subtracting the watch error from the watch time, WT, results in UT.

$$UT = WT - \Delta T$$

Improvements

The procedures described so far refer to a spherical Earth. In reality, however, the Earth has approximately the shape of an ellipsoid flattened at the poles. This leads to small but measurable effects when observing the Moon, the body nearest to the Earth. First, the parallax in altitude differs slightly from the value calculated for a spherical Earth. Second, there is a small parallax in azimuth which would not exist if the Earth were a sphere (see chapter 9). If no correction is applied, D may contain an error of up to approx. 0.2'. The following formulas refer to an observer on the surface of the reference ellipsoid (approximately at sea level).

The corrections require knowledge of the observer's latitude, Lat, the true azimuth of the Moon, Az_M, and the true azimuth of the reference body, Az_B.

Since the corrections are small, the three values do not need to be very accurate. Errors of a few degrees are tolerable. Instead of the azimuth, the compass bearing of each body, corrected for magnetic declination, may be used.

Parallax in altitude:

This correction is applied to the parallax in altitude and is used to calculate H_M with higher precision before clearing the lunar distance.

$$\Delta P_M \approx f \cdot HP_M \cdot \left[\sin(2 \cdot Lat) \cdot \cos Az_M \cdot \sin H_{Mapp} - \sin^2 Lat \cdot \cos H_{Mapp}\right]$$

f is the flattening of the Earth: $f = \dfrac{1}{298.257}$

$$P_{M,\,improved} = P_M + \Delta P_M$$

$$H_M = H_{mapp} - R_{Mapp} + P_{M,\,improved}$$

Parallax in azimuth:

The correction for the parallax in azimuth is applied after calculating H_M and D. The following formula is a fairly accurate approximation of the parallax in azimuth, ΔAz_M:

$$\Delta Az_M \approx f \cdot HP_M \cdot \frac{\sin(2 \cdot Lat) \cdot \sin Az_M}{\cos H_M}$$

In order to find how ΔAz_M affects D, we go back to the cosine formula:

$$\cos D = \sin H_M \cdot \sin H_B + \cos H_M \cdot \cos H_B \cdot \cos \alpha$$

We differentiate the equation with respect to α:

$$\frac{d(\cos D)}{d\alpha} = -\cos H_M \cdot \cos H_B \cdot \sin \alpha$$

$$d(\cos D) = -\sin D \cdot dD$$

$$-\sin D \cdot dD = -\cos H_M \cdot \cos H_B \cdot \sin \alpha \cdot d\alpha$$

$$dD = \frac{\cos H_M \cdot \cos H_B \cdot \sin \alpha}{\sin D} \cdot d\alpha$$

Since $d\alpha = d\,Az_M$, the change in D caused by an infinitesimal change in Az_M is:

$$dD = \frac{\cos H_M \cdot \cos H_B \cdot \sin \alpha}{\sin D} \cdot d\,Az_M$$

With a small but measurable change in Az_M, we have:

$$\Delta D \approx \frac{\cos H_M \cdot \cos H_B \cdot \sin \alpha}{\sin D} \cdot \Delta Az_M$$

$$D_{improved} \approx D + \Delta D$$

Combining the formulas for ΔAz_M and ΔD, we get:

$$D_{improved} \approx D + f \cdot HP_M \cdot \frac{\cos H_B \cdot \sin(2 \cdot Lat) \cdot \sin Az_M \cdot \sin(Az_M - Az_B)}{\sin D}$$

In most cases, $D_{improved}$ will be accurate to 0.1" (provided the measurements are error-free). The correction formula is less accurate when the topocentric (~ apparent) positions of Moon and reference body are close (< 5°) together. The formula should not be applied when the reference body is less than about four semidiameters (~1°) away from the center of the Moon.

Accuracy

According to modern requirements, the lunar distance method is rather inaccurate. In the 18th and early 19th century, however, this was generally accepted because a longitude with an error of 0.5°-1° was still better than no longitude measurement at all. Said error is the approximate result of an error of only 1' in the measurement of D_{Lapp}, not uncommon for a sextant reading under practical conditions. Therefore, D_{Lapp} should be measured with greatest care.

The altitudes of both bodies do not quite require the same degree of precision because a small error in the apparent altitude leads to about the same error in the geocentric altitude. Since both errors cancel each other to a large extent, the resulting error in D is comparatively small. An altitude error of a few arcminutes is tolerable in most cases. Therefore, measuring two altitudes of each body and finding the altitude at the moment of the lunar distance observation by interpolation is not absolutely necessary. Measuring a single altitude of each body shortly before or after the lunar distance measurement is sufficient if a small loss in accuracy is accepted.

The position of the reference body with respect to the Moon is crucial. The rate of change of D should not be too low. It becomes zero when D passes through a minimum or maximum, making an observation useless. This can be checked with lunar distance tables. Since the plane of the lunar orbit forms a relatively small angle (approx. 5°) with the ecliptic, bright bodies in the vicinity of the ecliptic are most suitable (Sun, planets, selected stars).

The stars generally recommended for the lunar distance method are Aldebaran, Altair, Antares, Fomalhaut, Hamal, Markab, Pollux, Regulus, and Spica, but other stars close to the ecliptic may be used as well, e. g., Nunki. The historic lunar distance tables of the *Nautical Almanac* contained only D values for those bodies having a favorable position with respect to the Moon on the day of observation.

Chapter 8

Rise, Set, Twilight

General Conditions for Visibility

For the planning of observations, it is useful to know the times during which a certain body is above the horizon as well as the times of sunrise, sunset, and twilight.

A body can be always above the horizon, always below the horizon, or above the horizon during a part of the day, depending on the observer's latitude and the declination of the body.

A body is above the celestial horizon at any time of the day when the zenith distance is smaller than 90° at the moment of **lower** meridian passage, i. e., when the body is on the lower branch of the local meridian (*Fig 8-1a*). This is the case if

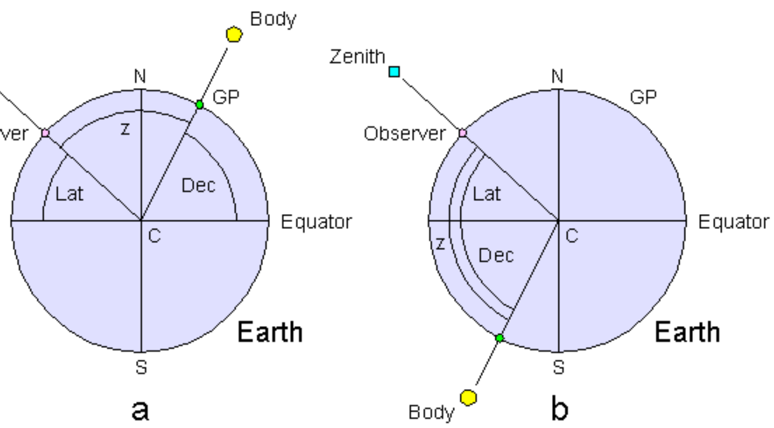
$$|Lat + Dec| > 90°$$

If this condition is fulfilled all year round, the body is **circumpolar**, i. e., it never sets.

A body never rises above the celestial horizon when the zenith distance is greater than 90° at the instant of **upper** meridian passage (*Fig 8-1b*). This is the case if

$$|Lat - Dec| > 90°$$

Fig. 8-1



A celestial body being on the same hemisphere as the observer is either sometimes or permanently above the celestial horizon. A body being on the opposite hemisphere is either sometimes above the horizon or permanently invisible.

The Sun provides a good example of how the visibility of a body is affected by latitude and declination. At the time of the summer solstice (Dec = +23.44°), the Sun (center) remains above the celestial horizon all day if the observer is north of the **arctic circle** (Lat > +66.56°). At the same time, the Sun remains below the celestial horizon all day if the observer is south of the **antarctic circle** (Lat < -66.56°). At the times of the equinoxes (Dec ≈ 0°), an observer at either of the poles will observe the Sun moving along the horizon during the course of the day. At the time of the winter solstice (Dec = −23.44°), the Sun is permanently visible south of the antarctic circle and invisible north of the arctic circle. If the observer is between arctic and antarctic circle, the Sun is visible during a part of the day all year round.

Rise and Set

The events of rise and set can be used to determine latitude, longitude, or time. One should not expect very accurate results, however, since atmospheric refraction may be erratic if the body is on or near the horizon.

The geometric rise or set of a body occurs when the center of the body passes through the celestial horizon (H = 0°). Due to the influence of atmospheric refraction, all bodies except the Moon appear above the sensible horizon at this instant.

The Moon is not visible at the moment of her geometric rise or set since the depressing effect of the horizontal parallax (~1°) is greater than the elevating effect of atmospheric refraction.

We begin with the well-known altitude formula (chapter 4) and set H at 0°:

$$\sin H \;=\; \sin Lat \cdot \sin Dec + \cos Lat \cdot \cos Dec \cdot \cos\, t \;=\; 0$$

$$\cos\, t \;=\; -\,\frac{\sin Lat \cdot \sin Dec}{\cos Lat \cdot \cos Dec}$$

Solving the equation for the meridian angle, t, we get :

$$t \;=\; \pm \arccos\left(-\,\tan Lat \cdot \tan Dec\right)$$

The equation has no solution when the argument of the inverse cosine is smaller than -1 or greater than 1. In the first case, the body is permanently above the horizon, in the latter case, the body remains below the horizon at any time. Otherwise, the arccos function returns values in the range from 0° through 180°. Due to the ambiguity of the arccos function, the equation has two solutions, one for rise and one for set. For the calculations below, we have to observe the following rules:

If the body is **rising** (body eastward from the observer), t is treated as a **negative** quantity.

If the body is **setting** (body westward from the observer), t is treated as a **positive** quantity.

If we know our latitude and the time of rise or set, we can calculate our longitude:

$$Lon \;=\; -GHA \pm t$$

GHA is the Greenwich hour angle of the body at the moment of rise or set. The sign of t has to be observed carefully (see above). If the resulting longitude is smaller than -180°, we add 360°.

Knowing our position, we can calculate the times of sunrise and sunset:

$$UT_{Sunrise\,,Sunset} \;\approx\; 12 - \,EoT - \frac{Lon[\,°\,]}{15} \pm \frac{t\,[\,°\,]}{15}$$

The UT value thus obtained may be negative or greater than 24h. Either case indicates a change of date occurring between local meridian transit and the instant of rise or set. Add or subtract 24h if necessary.

The moments of sunrise and sunset obtained with the above formula are not quite accurate since Dec and EoT are variable. Since we do not know the exact time of rise or set at the beginning, we have to use approximate values for Dec and EoT initially. The time of rise or set is improved by iteration. For this purpose, we find the values for Dec and EoT at the time of rise or set thus calculated. Then we re-calculate t and UT. If necessary, we repeat this procedure until the results converge sufficiently.

The times thus calculated are influenced by the irregularities of atmospheric refraction near the horizon. Therefore, a time error of ±2 minutes is not unusual.

Accordingly, we can calculate our longitude from the time of sunrise or sunset if we know our latitude:

$$Lon[\,°\,] \;\approx\; 15 \cdot \left(12 - UT_{Sunrise\,,\,Sunset} - \,EoT\right) \pm\, t\,[\,°\,]$$

Again, this is not a very precise method, and an error of several arcminutes in longitude is not unlikely.

Knowing our longitude, we are able to determine our approximate latitude from the time of sunrise or sunset:

$$t\,[\,°\,] \;\approx\; Lon[\,°\,] - 15 \cdot \left(12 - UT_{Sunrise\,,\,Sunset} - \,EoT\,\right)$$

$$Lat \;=\; \arctan\left(-\,\frac{\cos t}{\tan Dec}\right)$$

Sunrise and sunset are defined as the instants at which the upper limb of the Sun appears on the astronomical (= sensible) horizon (see glossary in [10] and [17]).

Taking into account the effects of refraction, horizontal parallax, and semidiameter, the geocentric altitude of the Sun's center is negative at the moment of rise or set.

$$H_{Sun} = HP - SD - R_H < 0$$

By definition, the standard refraction for a body being on the sensible horizon at sea level, R_H, is 34'. In reality, R_H is subject to random variations, mainly if temperature and atmospheric pressure differ from standard conditions.

Horizontal parallax (~0.15') and the small variations in semidiameter (~16') are customarily ignored [17]. Accordingly, the corresponding altitude of the center of the Sun with respect to the celestial horizon is -50'.

$$\mathrm{H} \approx -16' - 34' = -50' = -0.8333°$$

(When observing a rise or set at sea, we further have to subtract the dip of horizon from H.)

Thus, referring to the upper limb of the Sun and the sensible horizon (Dip=0), the meridian angle at the time of sunrise or sunset is:

$$t = \pm \arccos \frac{\sin(-0.8333°) - \sin Lat \cdot \sin Dec}{\cos Lat \cdot \cos Dec}$$

The time of rise or set is calculated in the same way as shown above. Again, we have to re-iterate to improve the result.

Azimuth and Amplitude

The azimuth angle of a rising or setting body is calculated with the azimuth formula (see chapter 4):

$$Az = \arccos \frac{\sin Dec - \sin H \cdot \sin Lat}{\cos H \cdot \cos Lat}$$

With H=0, we get:

$$Az = \arccos \frac{\sin Dec}{\cos Lat}$$

Az is +90° (rise) and -90° (set) if the declination of the body is zero, regardless of the observer's latitude. Accordingly, the Sun rises almost exactly in the east and sets in the west at the times of the equinoxes (geometric rise and set).

Observing the upper limb of the Sun crossing the sensible horizon (H = −0.8333°), we have:

$$Az \approx \arccos \frac{\sin Dec - \sin(-0.8333°) \cdot \sin Lat}{\cos Lat}$$

The true azimuth of the rising or setting body is:

$$Az_N = \begin{cases} Az & \text{if} \quad t < 0 \\ 360° - Az & \text{if} \quad t > 0 \end{cases}$$

The azimuth of a body at the moment of rise or set can be used to find the magnetic declination at the observer's position (compare with chapter 13).

The horizontal angular distance of a rising or setting body from the east or west point on the horizon is called **amplitude** and can be calculated from the azimuth. An amplitude of E45°N, for instance, means that the body rises 45° north of the east point on the horizon.

Twilight

At sea, twilight is important for the observation of stars and planets since it is the only time when these bodies **and** the horizon are visible. By definition, there are three kinds of twilight. The altitude, H, refers to the center of the Sun and the celestial horizon.

Civil twilight: $-0.8333° > H \geq -6°$

Nautical twilight: $-6° > H \geq -12°$

Astronomical twilight: $-12° > H \geq -18°$

In general, the nautical twilight is considered the best time window for observations of stars and planets at sea (visibility of brighter stars coinciding with visibility of sea horizon). However, exceptions to this rule are possible, depending on the actual weather conditions and the brightness of the observed body.

The meridian angle for the Sun (center) at -9° altitude (middle of nautical twilight) is:

$$t = \pm \arccos \frac{\sin(-9°) - \sin Lat \cdot \sin Dec}{\cos Lat \cdot \cos Dec}$$

Using this formula, we can find the approximate time for our observations (in analogy to sunrise and sunset). As mentioned above, the simultaneous observation of stars or planets **and** the horizon is possible during a limited time interval only.

To calculate the length of this interval, ΔT, we use the altitude formula and differentiate sin H with respect to the meridian angle, t:

$$\frac{d(\sin H)}{dt} = -\cos Lat \cdot \cos Dec \cdot \sin t$$

$$d(\sin H) = -\cos Lat \cdot \cos Dec \cdot \sin t \cdot dt$$

Substituting cosH·dH for d(sinH) and solving for dt, we get the change in the meridian angle, dt, as a function of a change in altitude, dH:

$$dt = -\frac{\cos H}{\cos Lat \cdot \cos Dec \cdot \sin t} \cdot dH$$

With H = -9° and dH ≈ ΔH = 6° (H = -6°...-12°), we get:

$$\Delta t[°] \approx -\frac{5.93}{\cos Lat \cdot \cos Dec \cdot \sin t}$$

Converting the change in the meridian angle to a time span (measured in minutes) and ignoring the sign, the equation is stated as:

$$\Delta T[m] \approx \frac{24}{\cos Lat \cdot \cos Dec \cdot \sin t}$$

The shortest possible time interval for our observations (Lat = 0, Dec = 0, H = -9°, t = 99°) lasts approx. 24 minutes. As the observer moves northward or southward from the equator, cos Lat decreases. Accordingly, the duration of twilight increases. When t is 180°, ΔT is infinite. This is confirmed by the well-known fact that the duration of twilight is shortest in equatorial regions and longest in polar regions.

We would obtain the same result when calculating t for H = -6° and H = -12°, respectively:

$$\Delta T[m] = 4 \cdot \left(t_{-12°}[°] - t_{-6°}[°]\right)$$

The Nautical Almanac provides tabulated values for the times of sunrise, sunset, civil twilight and nautical twilight for latitudes between -60° and +72° (referring to an observer being at the Greenwich meridian). In addition, times of moonrise and moonset are given.

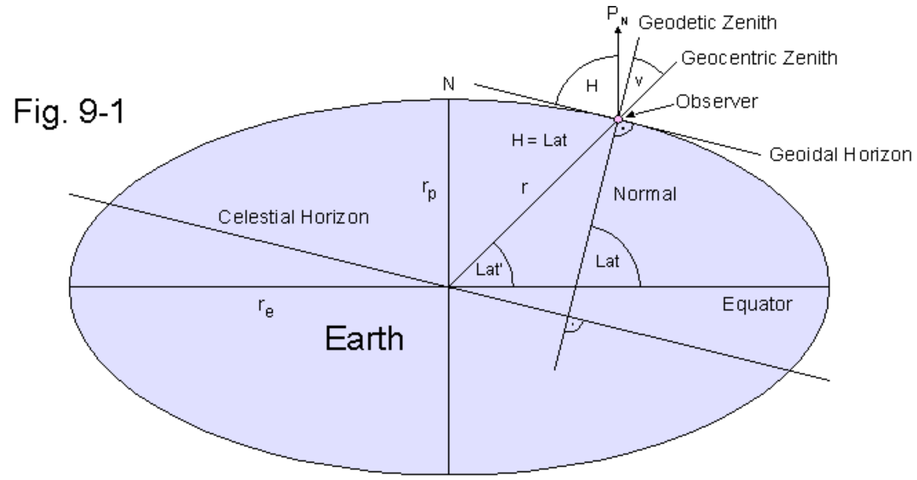
Chapter 9

Geodetic Aspects of Celestial Navigation

The Ellipsoid

Celestial navigation is based upon the assumption that the Earth is a sphere. Accordingly, calculations are based on the laws of spherical trigonometry. In reality, the shape of the Earth rather resembles an **oblate spheroid (ellipsoid)** resulting from two forces, **gravitation** and **centrifugal force**, acting on the viscous body of the Earth. While gravitation alone would force the Earth to assume the shape of a sphere, the state of lowest potential energy, the centrifugal force caused by the Earth's rotation contracts the Earth along the axis of rotation (polar axis) and stretches it along the plane of the equator. The local vector sum of both forces is called **gravity**.

There are several reference ellipsoids in use to describe the shape of the Earth, for example the **World Geodetic System ellipsoid of 1984 (WGS 84)**. An important characteristic of the WGS 84 ellipsoid is that its center conincides with the mass center of the Earth. There are special reference ellipsoids whose centers are not identical with the mass center. Off-center ellipsoids are constructed to obtain a better fit for a particular region. The following considerations refer to the WGS 84 ellipsoid which gives the best universal fit and is accurate enough for the purpose of navigation in most cases. *Fig. 9-1* shows a meridional section of the ellipsoid.



Earth data (WGS 84 ellipsoid) :

Equatorial radius	r_e	6378137.0 m
Polar radius	r_p	6356752.3142 m
Flattening	$f = (r_e - r_p) / r_e$	1/298.25722

Due to the flattening of the Earth, we have to distinguish between **geodetic** and **geocentric** latitude of a given position. The geodetic latitude, Lat, is the angle between the local normal (perpendicular) to the surface of the reference ellipsoid and the line of intersection formed by the plane of the equator and the plane of the local meridian. The geocentric latitude, Lat', is the angle formed by the local radius vector and said line of intersection. Geodetic and geocentric latitude are interrelated as follows:

$$\tan Lat' = (1-f)^2 \cdot \tan Lat$$

If the Earth were a sphere (f = 0), geodetic and geocentric latitude would be the same. With the spheroid, both quantities are equal only at the poles and on the equator. At all other places, the absolute value of the geocentric latitude is smaller than the absolute value of the geodetic latitude. **Due to the rotational symmetry of the ellipsoid with respect to the polar axis, geodetic and geocentric longitude are equal, provided the same reference meridian is used***. Maps are usually based upon **geodetic coordinates** which are also referred to as **geographic coordinates** [1]**. In this context it should be mentioned that the term "**geographic position**", applied to a celestial body, is misleading, since Greenwich hour angle and declination are **geocentric coordinates** (see chapter 3).

*Actually, the WGS84 (GPS) reference meridian is located 5.3 arcseconds east of the 1884 Greenwich meridian which passes through the historic Airy transit instrument. The reason is the local deflection of the vertical (see further below).

**In other publications, e. g., [10], astronomical coordinates (see below) and geographic coordinates are considered as identical.

In the following, we will discuss the effects of the oblateness (flattening) of the Earth on celestial navigation.

Any zenith distance (and corresponding altitude) measured by the navigator refers to the local **direction of gravity** (plumb line) which points to the **astronomical nadir** and thus defines the **astronomical zenith** which is exactly opposite to the nadir. Even the visible sea horizon is defined by the astronomical zenith since the plane tangent to the water surface at the observer's position is perpendicular to the local direction of gravity.

Under the hypothetical (!) assumption that the mass distribution inside the ellipsoid is in a hydrostatic equilibrium, the plumb line coincides with the local normal to the ellipsoid which passes through the **geodetic zenith**. Thus, astronomical and geodetic zenith are identical. Accordingly, the **astronomical coordinates** of a terrestrial position (obtained by astronomical observations) are equal to the geodetic (geographic) coordinates. As demonstrated in *Fig. 9-1* for example, the altitude of the celestial north pole, P_N, with respect to the geoidal horizon equals the geodetic, not the geocentric latitude. A noon latitude, calculated from the (geocentric) declination and the zenith distance with respect to the astronomical zenith would lead to the same result.

The **geocentric zenith** is defined as the point where a straight line originating from the center of the Earth and passing through the observer's position intersects the celestial sphere. The angle between this line and the local normal to the reference ellipsoid is called **angle of the vertical**, **v**. The angle of the vertical lies on the plane of the local meridian and is a function of the geodetic latitude. The following formula was proposed by *Smart* [9]:

$$v[''] \approx 692.666 \cdot \sin(2 \cdot Lat) - 1.163 \cdot \sin(4 \cdot Lat) + 0.026 \cdot \sin(6 \cdot Lat)$$

The coefficients of the above formula refer to the proportions of the WGS 84 ellipsoid.

The angle of the vertical at a given position equals the difference between geodetic and geocentric latitude (*Fig. 9-1*):

$$v = Lat - Lat'$$

The maximum value of v, occurring at 45° geodetic latitude, is approx. 11.5'. Thus, the geocentric latitude of an observer being at 45°geodetic latitude is only 44° 48.5'.

The navigator, of course, wants to know if the oblateness of the Earth causes significant errors due to the fact that calculations of celestial navigation are based on the laws of spherical trigonometry. According to the above values for polar radius and equatorial radius of the WGS 84 ellipsoid, the great circle distance of one arcminute is 1849 m at the poles and 1855 m at the equator. This small difference does not produce a significant error when plotting lines of position. It is therefore sufficient to use the adopted mean value (1 nautical mile = 1.852 km). However, when calculating the great circle distance (chapter 11) of two locations thousands of nautical miles apart, the error caused by the oblateness of the Earth can increase to several nautical miles. If high accuracy is required, the formulas for geodetic distance should be used [2]. The shortest path between two points on the surface of an ellipsoid is called **geodesic line**. It is the equivalent to the arc of a great circle on the surface of a sphere.

The Geoid

The spheroid is an idealized description of the shape of the Earth. In reality, the Earth has a non-uniform mass distribution and is not in a state of hydrostatic equilibrium. The rather irregular shape of the Earth is more accurately described by the **geoid**, an equipotential surface of gravity.

The geoid has local anomalies in the form of elevations and depressions. Elevations occur at local accumulations of matter (mountains, ore deposits), depressions at local deficiencies of matter (large water bodies, valleys, caverns). The elevation or depression of each surface point of the geoid with respect to the reference ellipsoid is found by gravity measurement.

On the slope of an elevation or depression of the geoid, the direction of gravity (the normal to the geoid) does not coincide with the normal to the reference ellipsoid, i. e., the astronomical zenith differs from the geodetic zenith in such places. The small angle between the local direction of gravity and the local normal to the reference ellipsoid, i. e., the angular distance between astronomical and geodetic zenith, is called **deflection of the vertical**. The latter is composed of a **meridional (north-south) component** and a **zonal (west-east) component**.

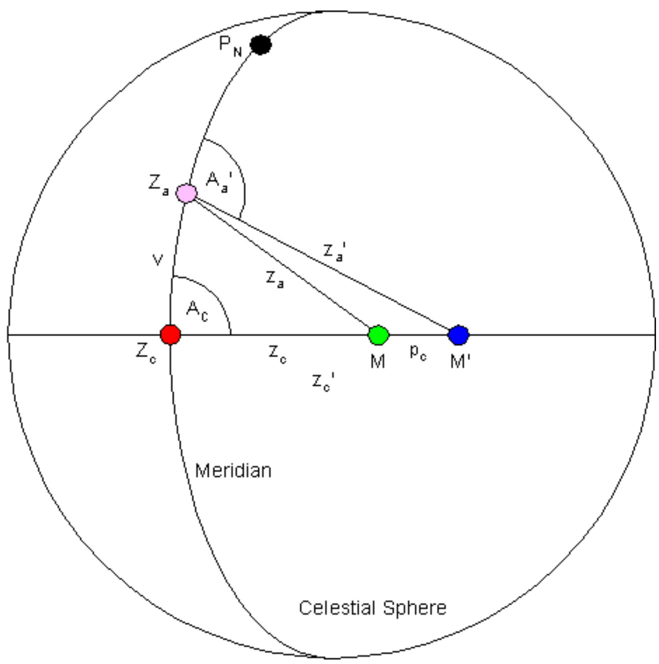
The deflection of the vertical is usually negligible at sea and thus ignored by mariners. In the vicinity of mountain ranges, however, significant deflections of the vertical (up to approx. 1 arcminute) have been reported (chapter 2). Thus, an **astronomical position** may be measurably different from the geodetic (geographic) position. This is important to surveying and map-making. Therefore, local corrections for the meridional and zonal component may have to be applied to an astronomical position, depending on the required precision.

The Parallax of the Moon

Due to the oblateness of the Earth, the distance between geoidal and celestial horizon is not constant but can assume any value between r_p and r_e, depending on the observer's latitude. This has a measurable effect on the parallax in altitude of the Moon since tabulated values for HP refer to the equatorial radius, r_e. The apparent position of the Moon is further affected by the fact that usually the local direction of gravity does not pass through the center of the ellipsoid. This displacement of the plumb line from the Earth's center causes a small (usually negligible) parallax in azimuth unless the Moon is on the local meridian. In the following, we will calculate the effects of the oblateness of the Earth on the parallax of the Moon with the exact formulas of spherical astronomy [9]. The effect of the oblateness of the Earth on the apparent position of other bodies is negligible.

Fig. 9-2 shows a projection of the astronomical zenith, Z_a, the geocentric zenith, Z_c, and the geographic position of the Moon, M, on the **celestial sphere**, an imaginary hollow sphere of infinite diameter with the Earth at its center.

Fig. 9-2



The geocentric zenith, Z_c, is the point where a straight line from the Earth's center through the observer's position intersects the celestial sphere. The astronomical zenith, Z_a, is the point at which the plumb line going through the observer's position intersects the celestial sphere. Z_a and Z_c are on the same celestial meridian. M is the projected geocentric position of the Moon defined by Greenwich hour angle and declination.

M' is the point where a straight line from the observer through the Moon's center intersects the celestial sphere. Z_c, M, and M' are on a great circle. The zenith distance measured by the observer is z_a' because the astronomical zenith is the only reference available. The quantity we want to know is z_a, the astronomical zenith distance corrected for parallax in altitude. This is the angular distance of the Moon from the astronomical zenith, measured by a fictitious observer at the Earth's center.

The known quantities are v, A_a', and z_a'. In contrast to the astronomer, the navigator is usually not able to measure A_a' precisely. For navigational purposes, the calculated azimuth (see chapter 4) may be substituted for A_a'.

We have three spherical triangles, Z_aZ_cM', Z_aZ_cM, and Z_aMM'. First, we calculate z_c' from z_a', v, and A_a' using the **law of cosines for sides** (see chapter 10):

$$\cos z_c' = \cos z_a' \cdot \cos v + \sin z_a' \cdot \sin v \cdot \cos\left(180° - A_a'\right)$$

$$z_c' = \arccos\left(\cos z_a' \cdot \cos v - \sin z_a' \cdot \sin v \cdot \cos A_a'\right)$$

To obtain z_c, we first have to calculate the relative ($r_e = 1$) local radius, ρ, and the geocentric parallax, p_c:

$$\rho = \frac{r}{r_e} = \sqrt{\frac{1-(2e^2-e^4)\cdot\sin^2 Lat}{1-e^2\cdot\sin^2 Lat}} \qquad e^2 = 1-\frac{r_p^2}{r_e^2}$$

$$p_c = \arcsin\left(\rho\cdot\sin HP\cdot\sin z_c^l\right)$$

HP is the equatorial horizontal parallax. The geocentric zenith distance corrected for parallax is:

$$z_c = z_c^l - p_c$$

Using the cosine formula again, we calculate A_c, the azimuth angle of the Moon with respect to the geocentric zenith:

$$A_c = \arccos\frac{\cos z_a^l - \cos z_c^l\cdot\cos v}{\sin z_c^l\cdot\sin v}$$

The astronomical zenith distance corrected for parallax is:

$$z_a = \arccos\left(\cos z_c\cdot\cos v + \sin z_c\cdot\sin v\cdot\cos A_c\right)$$

Thus, the **parallax in altitude** (astronomical) is:

$$PA = z_a^l - z_a$$

The small angle between M and M', measured at Z_a, is the **parallax in azimuth**, p_{az}:

$$p_{az} = \arccos\frac{\cos p_c - \cos z_a\cdot\cos z_a^l}{\sin z_a\cdot\sin z_a^l}$$

The parallax in azimuth does not exist when the Moon is on the local meridian. It is further non-existant when the observer is at one of the poles or on the equator (v = 0) but greatest when the observer is at medium latitudes. As a consequence of the parallax in azimuth, the horizontal direction of the Moon observed from the surface of the ellipsoid is always a little closer to the elevated pole (the celestial pole above the horizon) than the horizontal direction observed from the center of the ellipsoid. The parallax in azimuth does not exceed ±f·HP when the Moon is on the horizon but increases with increasing altitude. In most cases, particularly at sea, the navigator will not notice the influence of the flattening of the Earth. Traditionally, the apparent altitude of a body is reduced to the geocentric altitude through the established altitude correction procedure (including the correction for parallax in altitude). The intercept method (chapter 4) compares the observed altitude thus obtained with the geocentric altitude calculated from the assumed geodetic (geographic) coordinates of the observer and the geocentric equatorial coordinates (chapter 3) of the observed body. The difference between observed and calculated altitude is the intercept. The calculated azimuth is geocentric. A correction for the parallax in azimuth (see above) is usually omitted since such a degree of precision can not be reproduced when plotting position lines on a nautical chart. On land, however, more accurate altitude measurement is possible, and the navigator or surveyor may wish to use refined methods for the calculation of his position when observing the Moon.

Medium-precision method

During the course of altitude corrections, we calculate the parallax in altitude, P, with the formulas for spherical bodies (chapter 2). After doing this, we calculate the approximate correction for the flattening of the Earth, ΔP:

$$\Delta P \approx f\cdot HP\cdot\left[\sin(2\cdot Lat)\cdot\cos Az_N\cdot\sin H - \sin^2 Lat\cdot\cos H\right]$$

Adding ΔP to P, we get the improved parallax in altitude which we use for our further calculations instead of P:

$$P_{improved} = P + \Delta P$$

As a result, we obtain a more accurate intercept (chapter 4). The above correction formula is accurate to a fraction of an arcsecond.

The approximate parallax in azimuth is obtained through a simple formula:

$$\Delta Az_N \approx f \cdot HP \cdot \frac{\sin(2 \cdot Lat) \cdot \sin Az_{N,geocentric}}{\cos H_{geocentric}}$$

The topocentric true azimuth is

$$Az_{N,topocentric} = Az_{N,geocentric} - \Delta Az_N$$

The formula for the parallax in azimuth is also accurate to a fraction of an arcsecond. It becomes less accurate as the altitude approaches 90°. Observing bodies with such altitudes, however, is difficult and usually avoided.

Rigorous method*

For even more accurate results, we use the **topocentric equatorial coordinates** of the observed body for sight reduction. Instead of the center of the Earth, the observer's position is the origin of this coordinate system. The plane of the topocentric equator is parallel to the geocentric equator. The plane of the local meridian remains the same. The values for altitude and true azimuth calculated from the topocentric coordinates of the observed body are topocentric as well. There is neither a parallax in altitude nor a parallax in azimuth, so we have to skip the parallax correction and have to correct for the topocentric (augmented) semidiameter of the body when performing the altitude corrections.

The topocentric equatorial coordinates of a celestial body are obtained from the geocentric ones through coordinate transformation. The given quantities are:

Geographic latitude of the observer	Lat
Geocentric meridian angle	t
Geocentric declination	Dec
Equatorial horizontal parallax	HP
Polar radius of the Earth	r_p
Equatorial radius of the Earth	r_e

To be calculated:

Topocentric meridian angle	t'
Topocentric declination	Dec'

First, we calculate a number of auxiliary quantities:

Eccentricity of the ellipsoid, distance between center and focal point of a meridional section ($r_e = 1$):

$$e = \sqrt{1-\left(\frac{r_p}{r_e}\right)^2}$$

Local radius ($r_e = 1$):

$$\rho = \frac{r_p}{r_e} \cdot \frac{1}{\sqrt{1-e^2 \cdot \cos^2 Lat'}}$$

*The formulas are rigorous for an observer on the surface of a reference ellipsoid the center of which coincides with the mass center of the Earth.

Geocentric latitude of the observer:

$$Lat' = \arctan\left[\left(\frac{r_p}{r_e}\right)^2 \cdot \tan Lat\right] = \arctan\left[(1-f)^2 \cdot \tan Lat\right]$$

The topocentric coordinates of the body, t' and Dec', are calculated as follows. Δt is the parallax in hour angle:

$$\Delta t = \arctan\frac{\rho \cdot \cos Lat' \cdot \sin HP \cdot \sin t}{\cos Dec - \rho \cdot \cos Lat' \cdot \sin HP \cdot \cos t}$$

$$t' = t + \Delta t$$

$$Dec' = \arctan\frac{(\sin Dec - \rho \cdot \sin Lat' \cdot \sin HP) \cdot \cos \Delta t}{\cos Dec - \rho \cdot \cos Lat' \cdot \sin HP \cdot \cos t}$$

Chapter 10

Spherical Trigonometry

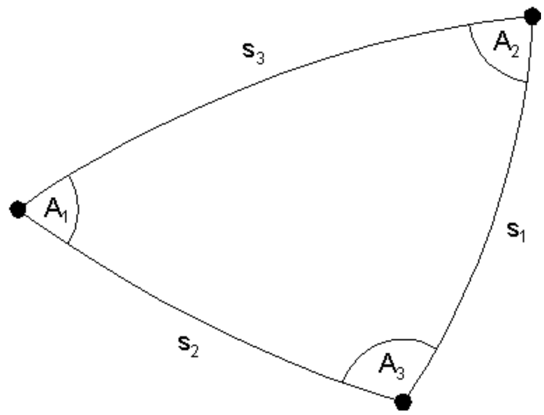
The Earth is usually regarded as a sphere in celestial navigation although an oblate spheroid would be a better approximation. Otherwise, navigational calculations would become too difficult for practical use. The position error introduced by the spherical Earth model is usually very small and stays within the "statistical noise" caused by other omnipresent errors like, e.g., abnormal refraction, rounding errors, etc.

Although it is possible to perform navigational calculations solely with the aid of tables (H.O. 229, H.O. 211, etc.) and with little mathematics, the principles of celestial navigation can not be comprehended without knowing the elements of spherical trigonometry.

The Oblique Spherical Triangle

Like any triangle, a spherical triangle is characterized by three sides and three angles. However, a spherical triangle is part of the surface of a sphere, and the sides are not straight lines but arcs of great circles (*Fig. 10-1*).

Fig. 10-1



A **great circle** is a circle on the surface of a sphere whose plane passes through the center of the sphere (see chapter 3).

Any side of a spherical triangle can be regarded as an angle - the angular distance between the adjacent vertices, measured at the center of the sphere. The interrelations between angles and sides of a spherical triangle are described by the **law of sines**, the **law of cosines for sides**, the **law of cosines for angles**, the **law of sines and cosines**, the **law of cotangents**, **Napier's analogies**, and **Gauss' formulas** (apart from other formulas).

Law of sines:

$$\frac{\sin A_1}{\sin s_1} = \frac{\sin A_2}{\sin s_2} = \frac{\sin A_3}{\sin s_3}$$

Law of cosines for sides:

$$\cos s_1 = \cos s_2 \cdot \cos s_3 + \sin s_2 \cdot \sin s_3 \cdot \cos A_1$$
$$\cos s_2 = \cos s_1 \cdot \cos s_3 + \sin s_1 \cdot \sin s_3 \cdot \cos A_2$$
$$\cos s_3 = \cos s_1 \cdot \cos s_2 + \sin s_1 \cdot \sin s_2 \cdot \cos A_3$$

Law of cosines for angles:

$$\cos A_1 = -\cos A_2 \cdot \cos A_3 + \sin A_2 \cdot \sin A_3 \cdot \cos s_1$$
$$\cos A_2 = -\cos A_1 \cdot \cos A_3 + \sin A_1 \cdot \sin A_3 \cdot \cos s_2$$
$$\cos A_3 = -\cos A_1 \cdot \cos A_2 + \sin A_1 \cdot \sin A_2 \cdot \cos s_3$$

Law of sines and cosines:

$$\sin s_1 \cdot \cos A_2 = \cos s_2 \cdot \sin s_3 - \sin s_2 \cdot \cos s_3 \cdot \cos A_1$$
$$\sin s_2 \cdot \cos A_3 = \cos s_3 \cdot \sin s_1 - \sin s_3 \cdot \cos s_1 \cdot \cos A_2$$
$$\sin s_3 \cdot \cos A_1 = \cos s_1 \cdot \sin s_2 - \sin s_1 \cdot \cos s_2 \cdot \cos A_3$$
$$\sin s_1 \cdot \cos A_3 = \cos s_3 \cdot \sin s_2 - \sin s_3 \cdot \cos s_2 \cdot \cos A_1$$
$$\sin s_2 \cdot \cos A_1 = \cos s_1 \cdot \sin s_3 - \sin s_1 \cdot \cos s_3 \cdot \cos A_2$$
$$\sin s_3 \cdot \cos A_2 = \cos s_2 \cdot \sin s_1 - \sin s_2 \cdot \cos s_1 \cdot \cos A_3$$

Law of cotangents:

$$\sin A_1 \cdot \cot A_2 = \cot s_2 \cdot \sin s_3 - \cos s_3 \cdot \cos A_1$$
$$\sin A_1 \cdot \cot A_3 = \cot s_3 \cdot \sin s_2 - \cos s_2 \cdot \cos A_1$$
$$\sin A_2 \cdot \cot A_3 = \cot s_3 \cdot \sin s_1 - \cos s_1 \cdot \cos A_2$$
$$\sin A_2 \cdot \cot A_1 = \cot s_1 \cdot \sin s_3 - \cos s_3 \cdot \cos A_2$$
$$\sin A_3 \cdot \cot A_1 = \cot s_1 \cdot \sin s_2 - \cos s_2 \cdot \cos A_3$$
$$\sin A_3 \cdot \cot A_2 = \cot s_2 \cdot \sin s_1 - \cos s_1 \cdot \cos A_3$$

Napier's analogies:

$$\tan\frac{A_1 + A_2}{2} \cdot \tan\frac{A_3}{2} = \frac{\cos\frac{s_1 - s_2}{2}}{\cos\frac{s_1 + s_2}{2}} \qquad \tan\frac{A_1 - A_2}{2} \cdot \tan\frac{A_3}{2} = \frac{\sin\frac{s_1 - s_2}{2}}{\sin\frac{s_1 + s_2}{2}}$$

$$\frac{\tan\frac{s_1 + s_2}{2}}{\tan\frac{s_3}{2}} = \frac{\cos\frac{A_1 - A_2}{2}}{\cos\frac{A_1 + A_2}{2}} \qquad \frac{\tan\frac{s_1 - s_2}{2}}{\tan\frac{s_3}{2}} = \frac{\sin\frac{A_1 - A_2}{2}}{\sin\frac{A_1 + A_2}{2}}$$

Gauss' formulas:

$$\frac{\sin\frac{A_1 + A_2}{2}}{\cos\frac{A_3}{2}} = \frac{\cos\frac{s_1 - s_2}{2}}{\cos\frac{s_3}{2}} \qquad \frac{\cos\frac{A_1 + A_2}{2}}{\sin\frac{A_3}{2}} = \frac{\cos\frac{s_1 + s_2}{2}}{\cos\frac{s_3}{2}}$$

$$\frac{\sin\frac{A_1 - A_2}{2}}{\cos\frac{A_3}{2}} = \frac{\sin\frac{s_1 - s_2}{2}}{\sin\frac{s_3}{2}} \qquad \frac{\cos\frac{A_1 - A_2}{2}}{\sin\frac{A_3}{2}} = \frac{\sin\frac{s_1 + s_2}{2}}{\sin\frac{s_3}{2}}$$

These formulas and others derived thereof enable any quantity (angle or side) of a spherical triangle to be calculated if three other quantities are known.

Particularly the law of cosines for sides is of interest since it can solve the majority of navigational problems.

Haversine formula

The haversine formula was very popular in the pre-electronic age because it is particularly suitable for logarithmic calculations by means of tabulated log hav x values.

The haversine function is defined as follows:

$$\operatorname{hav}\alpha = \frac{1-\cos\alpha}{2} = \sin^2\left(\frac{\alpha}{2}\right)$$

The haversine formula for the oblique spherical triangle is stated as

$$\begin{aligned}
\operatorname{hav} s_1 &= \operatorname{hav}(s_2 - s_3) + \sin s_2 \cdot \sin s_3 \cdot \operatorname{hav} A_1 \\
\operatorname{hav} s_2 &= \operatorname{hav}(s_1 - s_3) + \sin s_1 \cdot \sin s_3 \cdot \operatorname{hav} A_2 \\
\operatorname{hav} s_3 &= \operatorname{hav}(s_1 - s_2) + \sin s_1 \cdot \sin s_2 \cdot \operatorname{hav} A_3
\end{aligned}$$

With hav x = (1 - cos x)/2, the haversine formula is stated as

$$\begin{aligned}
\cos s_1 &= \cos(s_2 - s_3) - \sin s_2 \cdot \sin s_3 \cdot (1 - \cos A_1) \\
\cos s_2 &= \cos(s_1 - s_3) - \sin s_1 \cdot \sin s_3 \cdot (1 - \cos A_2) \\
\cos s_3 &= \cos(s_1 - s_2) - \sin s_1 \cdot \sin s_2 \cdot (1 - \cos A_3)
\end{aligned}$$

The haversine formula is another form of the law of cosines for sides.

Proof:

$$\cos s_1 = \cos(s_2 - s_3) - \sin s_2 \cdot \sin s_3 \cdot (1 - \cos A_1)$$

Addition identity:

$$\cos(s_2 - s_3) = \cos s_2 \cdot \cos s_3 + \sin s_2 \cdot \sin s_3$$

Replacing $\cos(s_2 - s_3)$ with $\cos s_2 \cdot \cos s_3 + \sin s_2 \cdot \sin s_3$, we get

$$\cos s_1 = \cos s_2 \cdot \cos s_3 + \sin s_2 \cdot \sin s_3 - \sin s_2 \cdot \sin s_3 \cdot (1 - \cos A_1)$$

$$\cos s_1 = \cos s_2 \cdot \cos s_3 + \sin s_2 \cdot \sin s_3 \cdot \left(1 - (1 - \cos A_1)\right)$$

$$\cos s_1 = \cos s_2 \cdot \cos s_3 + \sin s_2 \cdot \sin s_3 \cdot \cos A_1$$

Quod erat demonstrandum.

This is how the haversine formula looks when replacing hav x with the equivalent $\sin^2(x/2)$:

$$\begin{aligned}
\sin^2\frac{s_1}{2} &= \sin^2\frac{(s_2 - s_3)}{2} + \sin s_2 \cdot \sin s_3 \cdot \sin^2\frac{A_1}{2} \\
\sin^2\frac{s_2}{2} &= \sin^2\frac{(s_1 - s_3)}{2} + \sin s_1 \cdot \sin s_3 \cdot \sin^2\frac{A_2}{2} \\
\sin^2\frac{s_3}{2} &= \sin^2\frac{(s_1 - s_2)}{2} + \sin s_1 \cdot \sin s_2 \cdot \sin^2\frac{A_3}{2}
\end{aligned}$$

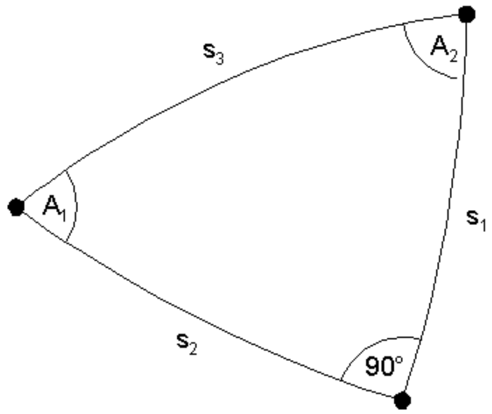
In this form, the haversine formula is often used for the calculation of great circle distances (chapter 12):

$$s_1 = 2 \cdot \arcsin\sqrt{\sin^2\frac{(s_2 - s_3)}{2} + \sin s_2 \cdot \sin s_3 \cdot \sin^2\frac{A_1}{2}}$$

The Right Spherical Triangle

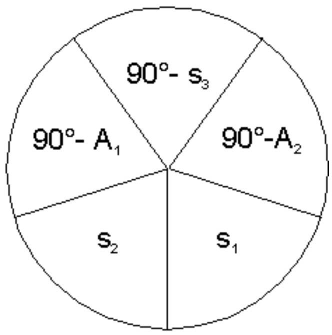
Solving a spherical triangle is less complicated when it contains a right angle (*Fig. 10-2*). Using ***Napier's* rules of circular parts**, any quantity can be calculated if only two other quantities (apart from the right angle) are known.

Fig. 10-2



We arrange the sides forming the right angle (s_1, s_2) and the **complements** of the remaining angles (A_1, A_2) and opposite side (s_3) in the form of a circular diagram consisting of five sectors, called "parts" (in the same order as they appear in the triangle). The right angle itself is omitted (*Fig. 10-3*):

Fig. 10-3



According to *Napier's* rules, the sine of any part of the diagram equals the product of the tangents of the adjacent parts and the product of the cosines of the opposite parts:

$$\sin s_1 = \tan s_2 \cdot \tan(90° - A_2) = \cos(90° - A_1) \cdot \cos(90° - s_3)$$
$$\sin s_2 = \tan(90° - A_1) \cdot \tan s_1 = \cos(90° - s_3) \cdot \cos(90° - A_2)$$
$$\sin(90° - A_1) = \tan(90° - s_3) \cdot \tan s_2 = \cos(90° - A_2) \cdot \cos s_1$$
$$\sin(90° - s_3) = \tan(90° - A_2) \cdot \tan(90° - A_1) = \cos s_1 \cdot \cos s_2$$
$$\sin(90° - A_2) = \tan s_1 \cdot \tan(90° - s_3) = \cos s_2 \cdot \cos(90° - A_1)$$

In a simpler form, these equations are stated as:

$$\sin s_1 = \frac{\tan s_2}{\tan A_2} = \sin A_1 \cdot \sin s_3$$
$$\sin s_2 = \frac{\tan s_1}{\tan A_1} = \sin s_3 \cdot \sin A_2$$
$$\cos A_1 = \frac{\tan s_2}{\tan s_3} = \sin A_2 \cdot \cos s_1$$
$$\cos s_3 = \frac{1}{\tan A_2 \cdot \tan A_1} = \cos s_1 \cdot \cos s_2$$
$$\cos A_2 = \frac{\tan s_1}{\tan s_3} = \cos s_2 \cdot \sin A_1$$

There are several applications for the right spherical triangle in navigation, for example **Ageton's sight reduction tables** (chapter 11) and **great circle navigation** (chapter 13).

Chapter 11

The Navigational Triangle

The **navigational (nautical) triangle** is the (usually) oblique spherical triangle formed by the north pole, P_N, the observer's assumed position, AP, and the geographic position of the celestial object, GP (*Fig. 11-1*). **All common sight reduction procedures are based upon the navigational triangle.**

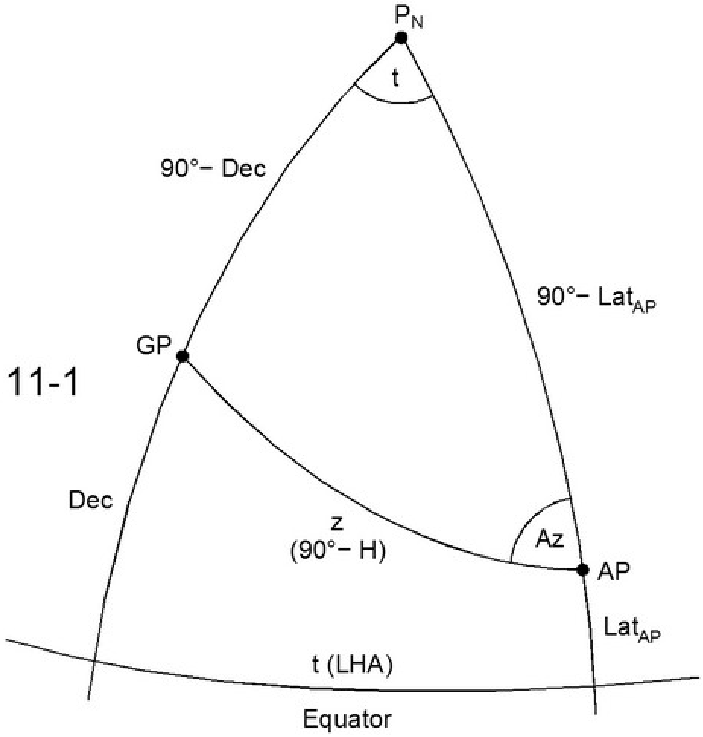


Fig. 11-1

Intercept method

When using the intercept method (chapter 4), the latitude of the observer's (assumed or real) position, Lat_{AP}, the declination of the observed celestial body, Dec, and the meridian angle, t, or the local hour angle, LHA, (calculated from the longitude of AP and the GHA of the object), are the known quantities.

The first step is calculating the side z of the navigational triangle by using the law of cosines for sides:

$$\cos z \;=\; \cos(90°-Lat_{AP})\cdot\cos(90°-Dec)+\sin(90°-Lat_{AP})\cdot\sin(90°-Dec)\cdot\cos t$$

Since cos (90°-x) equals sin x and vice versa, the equation can be written in a simpler form:

$$\cos z \;=\; \sin Lat_{AP}\cdot\sin Dec+\cos Lat_{AP}\cdot\cos Dec\cdot\cos t$$

The side z is not only the **great circle distance** between AP and GP but also the **zenith distance** of the celestial object and the great circle radius of the **circle of equal altitude** (see chapter 1).

Substituting the altitude H for z, we get

$$\sin H \;=\; \sin Lat_{AP}\cdot\sin Dec+\cos Lat_{AP}\cdot\cos Dec\cdot\cos t$$

Solving the equation for H leads to the altitude formula known from chapter 4:

$$H \;=\; \arcsin\left(\sin Lat_{AP}\cdot\sin Dec+\cos Lat_{AP}\cdot\cos Dec\cdot\cos t\right)$$

Instead of the law of cosines for sides, we can use the **haversine formula** (see chapter 10) to calculate H:

$$\text{hav}\, z \;=\; \text{hav}\left(Lat - Dec\right) + \cos Lat \cdot \cos Dec \cdot \text{hav}\, t$$

Substituting the equivalent (1-cos x)/2 for hav x and multiplying both sides with 2, we get

$$1 - \cos z \;=\; 1 - \cos\left(Lat - Dec\right) + \cos Lat \cdot \cos Dec \cdot \left(1 - \cos t\right)$$

$$\sin H \;=\; \cos\left(Lat - Dec\right) - \cos Lat \cdot \cos Dec \cdot \left(1 - \cos t\right)$$

$$H \;=\; \arcsin\left[\cos\left(Lat - Dec\right) - \cos Lat \cdot \cos Dec \cdot \left(1 - \cos t\right)\right]$$

The altitude calculated for a given position (assumed or real) is called computed altitude, Hc.

The **azimuth angle** of the observed body is also calculated by means of the law of cosines for sides:

$$\cos\left(90° - Dec\right) \;=\; \cos\left(90° - Lat_{AP}\right) \cdot \cos z + \sin\left(90° - Lat_{AP}\right) \cdot \sin z \cdot \cos Az$$

$$\sin Dec \;=\; \sin Lat_{AP} \cdot \cos z + \cos Lat_{AP} \cdot \sin z \cdot \cos Az$$

Using the computed altitude instead of the zenith distance results in the following equation:

$$\sin Dec \;=\; \sin Lat_{AP} \cdot \sin Hc + \cos Lat_{AP} \cdot \cos Hc \cdot \cos Az$$

Solving the equation for Az finally yields the altitude-azimuth formula from chapter 4:

$$Az \;=\; \arccos \frac{\sin Dec - \sin Lat_{AP} \cdot \sin Hc}{\cos Lat_{AP} \cdot \cos Hc}$$

Solving the haversine formula for Az, we get

$$Az \;=\; \arccos\left(1 - \frac{\cos\left(Lat_{AP} - Hc\right) - \sin Dec}{\cos Lat_{AP} \cdot \cos Hc}\right)$$

The arccos function returns angles between 0° and 180°. Therefore, the resulting azimuth angle is not necessarily identical with the true azimuth, Az_N (0°... 360°, measured clockwise from true north) commonly used in navigation. In all cases where t is negative (GP east of AP) , Az_N equals Az. Otherwise (t positive, GP westward from AP as shown in *Fig. 11-1*), Az_N is obtained by subtracting Az from 360°.

Time sight

When the meridian angle, t, (or the local hour angle, LHA) is the quantity to be calculated (time sight, *Sumner's* method), Dec, Lat_{AP} (the assumed latitude), and the observed zenith distance or altitude (z or Ho) are the known quantities. Again, the law of cosines for sides is applied:

$$\cos z \;=\; \cos\left(90° - Lat_{AP}\right) \cdot \cos\left(90° - Dec\right) + \sin\left(90° - Lat_{AP}\right) \cdot \sin\left(90° - Dec\right) \cdot \cos t$$

$$\sin Ho \;=\; \sin Lat_{AP} \cdot \sin Dec + \cos Lat_{AP} \cdot \cos Dec \cdot \cos t$$

$$\cos t \;=\; \frac{\sin Ho - \sin Lat_{AP} \cdot \sin Dec}{\cos Lat_{AP} \cdot \cos Dec}$$

$$t \;=\; \pm \arccos \frac{\sin Ho - \sin Lat_{AP} \cdot \sin Dec}{\cos Lat_{AP} \cdot \cos Dec}$$

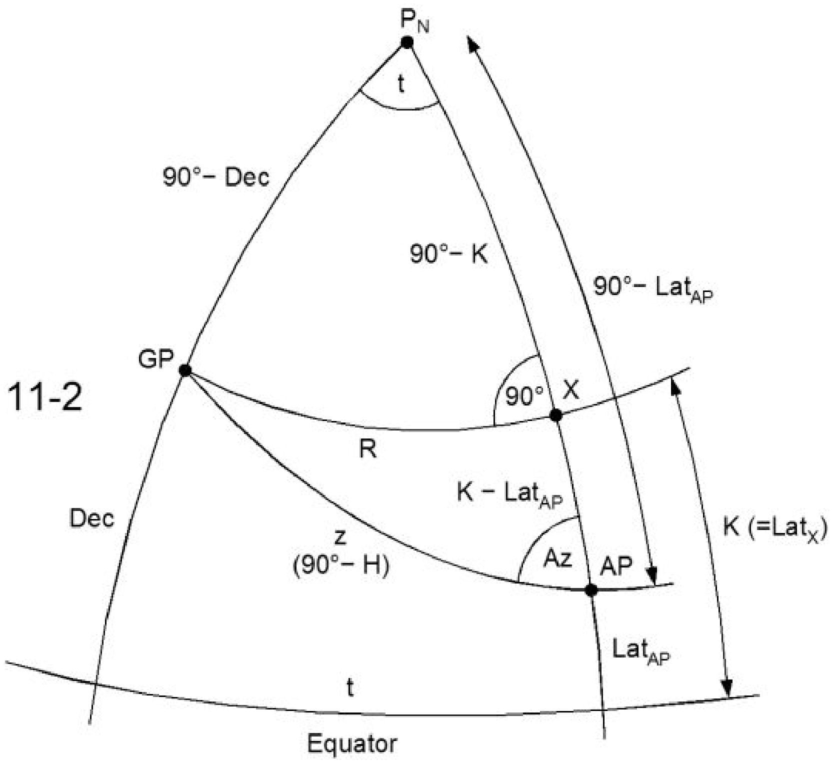
The obtained meridian angle, t, is then used as described in chapter 4 and chapter 6.

When observing a celestial body at the time of meridian passage (e. g., for determining one's latitude), the local hour angle is zero, and the navigational triangle becomes infinitesimally narrow. In this case, the sides of the spherical triangle can be calculated by simple addition or subtraction.

The Divided Navigational Triangle

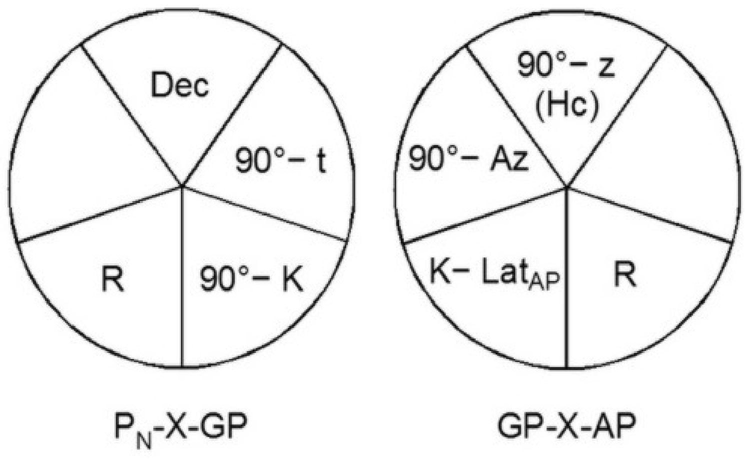
An alternative method for solving the navigational triangle is based upon two right spherical triangles obtained by constructing a great circle passing through GP and intersecting the local meridian perpendicularly at X (*Fig. 11-2*).

Fig. 11-2



The first right triangle is formed by P_N, X, and GP, the second one by GP, X, and AP. The **auxiliary parts** R and K are intermediate quantities used to calculate z (or Hc) and Az. K is the geographic latitude of X. Both triangles are solved using **Napier's rules of circular parts** (see chapter 9). *Fig. 11-3* illustrates the corresponding circular diagrams:

Fig. 11-3



According to Napier's rules, Hc and Az are calculated by means of the following formulas:

$$\sin R = \sin t \cdot \cos Dec \qquad \Rightarrow \qquad R = \arcsin(\sin t \cdot \cos Dec)$$

$$\sin Dec = \cos R \cdot \sin K \quad \Rightarrow \quad \sin K = \frac{\sin Dec}{\cos R} \quad \Rightarrow \quad K = \arcsin \frac{\sin Dec}{\cos R}$$

Substitute 180°-K for K in the following equation if |t| > 90° (or 90° < LHA < 270°):

$$\sin Hc = \cos R \cdot \cos\left(K - Lat_{AP}\right) \quad \Rightarrow \quad Hc = \arcsin\left[\cos R \cdot \cos\left(K - Lat_{AP}\right)\right]$$

$$\sin R = \cos Hc \cdot \sin Az \quad \Rightarrow \quad \sin Az = \frac{\sin R}{\cos Hc} \quad \Rightarrow \quad Az = \arcsin \frac{\sin R}{\cos Hc}$$

For further calculations, substitute 180°-Az for Az if K and Lat have opposite signs or if |K| **<** |Lat|. To obtain the true azimuth, Az_N (0°... 360°), the following rules have to be applied:

$$Az_N = \begin{Bmatrix} -Az & \text{if} & Lat_{AP} > 0\ (N) & \text{AND} & t < 0\ \ (180° < LHA < 360°) \\ 360° - Az & \text{if} & Lat_{AP} > 0\ (N) & \text{AND} & t > 0\ \ (0° < LHA < 180°) \\ 180° + Az & \text{if} & Lat_{AP} < 0\ (S) & & \end{Bmatrix}$$

The divided navigational triangle is of considerable importance since it forms the theoretical background for a number of **sight reduction tables**, e.g., the Ageton Tables (see below). It is also used for **great circle navigation** (chapter 12).

Using the secant and cosecant functions (sec x = 1/cos x, csc x = 1/sin x), we can write the equations for the divided navigational triangle in the following form:

$$\csc R = \csc t \cdot \sec Dec$$

$$\csc K = \frac{\csc Dec}{\sec R}$$

Substitute 180°-K for K in the following equation if |t| > 90°:

$$\csc Hc = \sec R \cdot \sec\left(K - Lat\right)$$

$$\csc Az = \frac{\csc R}{\sec Hc}$$

Substitute 180°-Az for Az if K and Lat have opposite signs or if |K|**<**|Lat|.

In logarithmic form, these equations are stated as:

$$\log\ \csc R = \log\ \csc t + \log \sec Dec$$

$$\log\ \csc K = \log\ \csc Dec - \log\ \sec R$$

$$\log\ \csc Hc = \log \sec R + \log \sec\left(K - Lat\right)$$

$$\log\ \csc Az = \log\ \csc R - \log\ \sec Hc$$

With the logarithms of the secants and cosecants of angles arranged in the form of a suitable table, we can solve a sight by a sequence of simple additions and subtractions. Apart from the table itself, the only tools required are a sheet of paper and a pencil.

The Ageton Tables (H.O. 211), first published in 1931, are based upon the above formulas and provide a very efficient arrangement of angles and their log secants and log cosecants on 36 pages. Since all calculations are based on absolute values, certain rules included in the instructions have to be observed.

Sight reduction tables were developed many years before electronic calculators became available in order to simplify calculations necessary to reduce a sight. Still today, sight reduction tables are preferred by people who do not want to deal with the formulas of spherical trigonometry. Moreover, they provide a valuable backup method if electronic devices fail.

Two modified versions of the Ageton Tables are available (2019) at: https://celnav.de/page3.htm

Chapter 12

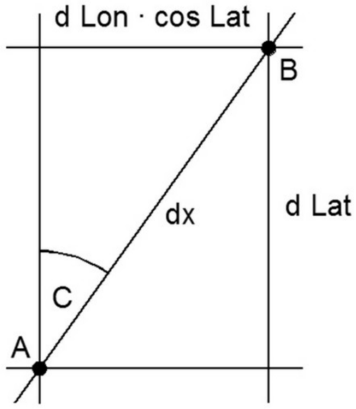
General Formulas for Navigation

Although the following formulas are not part of celestial navigation, they are indispensible because they are necessary to calculate distance and direction (course) from the **point of departure**, A, to the **point of arrival**, B, as well as to calculate the position of B from the position of A if course and distance are known. The **true course**, C, is the angle made by the vector of motion and the local meridian. It is measured from true north (clockwise through 360°). Knowing the coordinates of A, Lat_A and Lon_A, and the coordinates of B, Lat_B and Lon_B, the navigator has the principal choice between **rhumb line navigation** (simple procedure but longer distance) and **great circle navigation** (shortest possible distance on a sphere). Combinations of both methods are possible.

Rhumb Line Navigation

A **rhumb line**, also called **loxodrome**, is a line on the surface of the Earth intersecting all meridians at a constant angle, C. Thus, a rhumb line is represented by a straight line on a Mercartor chart (see chapter 13) which makes **voyage planning** quite simple. On a globe, a rhumb line forms a spherical spiral extending from pole to pole unless it is identical with a meridian (C = 0° or 180°) or a parallel of latitude (C = 90° or 270°). A vessel steering a constant course travels along a rhumb line, provided there is no drift. Rhumb line course, C, and distance, d, are calculated as shown below. First, we imagine traveling the infinitesimal distance dx from the point of departure, A, to the point of arrival, B. Our course is C (*Fig. 12-1*):

Fig. 12-1



The distance, dx, is the vector sum of a north-south component, dLat, and a west-east component, dLon · cos Lat. The factor cos Lat is the relative circumference of the respective parallel of latitude (equator = 1):

$$\tan C = \frac{d\,Lon \cdot \cos Lat}{d\,Lat}$$

$$\frac{d\,Lat}{\cos Lat} = \frac{1}{\tan C} \cdot d\,Lon$$

If the distance between A (defined by Lat_A and Lon_A) and B (defined by Lat_B and Lon_B) is a measurable quantity, we have to integrate:

$$\int_{Lat_A}^{Lat_B} \frac{d\,Lat}{\cos Lat} = \frac{1}{\tan C} \cdot \int_{Lon_A}^{Lon_B} d\,Lon$$

$$\ln\left[\tan\left(\frac{Lat_B}{2} + \frac{\pi}{4}\right)\right] - \ln\left[\tan\left(\frac{Lat_A}{2} + \frac{\pi}{4}\right)\right] = \frac{Lon_B - Lon_A}{\tan C}$$

$$\tan C = \frac{Lon_B - Lon_A}{\ln \frac{\tan\left(\frac{Lat_B}{2} + \frac{\pi}{4}\right)}{\tan\left(\frac{Lat_A}{2} + \frac{\pi}{4}\right)}}$$

Measuring angles in degrees and solving for C, we get:

$$C = \arctan \frac{Lon_B - Lon_A}{\ln \frac{\tan\left(\frac{Lat_B}{2} + 45°\right)}{\tan\left(\frac{Lat_A}{2} + 45°\right)}}$$

The term Lon_B-Lon_A has to be in the range between –180° tand +180°. If it is outside this range, we have to add or subtract 360° before entering the rhumb line course formula.

The arctan function returns values between -90° and +90°. To obtain the true course (0°...360°), we apply the following rules:

$$C \rightarrow \left\{\begin{array}{llll} C & \text{if} & Lat_B > Lat_A & \text{AND} \quad Lon_B > Lon_A \\ 360° + C & \text{if} & Lat_B > Lat_A & \text{AND} \quad Lon_B < Lon_A \\ 180° + C & \text{if} & Lat_B < Lat_A & \end{array}\right\}$$

To find the total length of the rhumb line track, we calculate the infinitesimal distance dx:

$$dx = \frac{d\,Lat}{\cos C}$$

The total length d is found through integration:

$$d = \frac{1}{\cos C} \cdot \int_{Lat_A}^{Lat_B} d\,Lat = \frac{Lat_B - Lat_A}{\cos C}$$

Finally, we get:

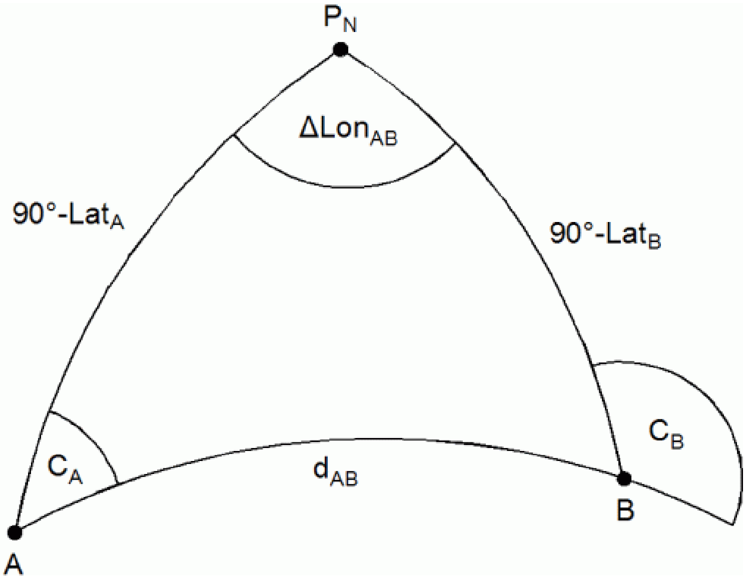
$$d[km] = \frac{40031.6}{360} \cdot \frac{Lat_B - Lat_A}{\cos C} \qquad d[nm] = 60 \cdot \frac{Lat_B - Lat_A}{\cos C}$$

If both positions have the same latitude, the distance can **not** be calculated using the above formulas. In this case, the following formulas apply (C is either 90° or 270°):

$$d[km] = \frac{40031.6}{360} \cdot (Lon_B - Lon_A) \cdot \cos Lat \qquad d[nm] = 60 \cdot (Lon_B - Lon_A) \cdot \cos Lat$$

Great circle distance, d_{AB}, and course, C_A, are calculated on the analogy of zenith distance and azimuth. For this purpose, we consider the navigational triangle (see chapter 11) and substitute A for GP, B for AP, d_{AB} for z, and ΔLon_{AB} (difference of longitude) for LHA (*Fig. 12-2*):

Fig. 12-2



$$d_{AB} = \arccos\left(\sin Lat_A \cdot \sin Lat_B + \cos Lat_A \cdot \cos Lat_B \cdot \cos \Delta Lon_{AB}\right)$$

$$\Delta Lon_{AB} = Lon_B - Lon_A$$

Quite often, the haversine formula (chapter 10 & 11) is used to calculate great circle distances:

$$\operatorname{hav} d_{AB} = \operatorname{hav}\left(\Delta Lat_{AB}\right) + \cos Lat_A \cdot \cos Lat_B \cdot \operatorname{hav}\left(\Delta Lon_{AB}\right)$$

$$\Delta Lat_{AB} = Lat_B - Lat_A$$

(Remember that northern latitude and eastern longitude are positive, southern latitude and western longitude negative.)

Replacing hav(x) with the equivalent $\sin^2(x/2)$, and solving for d_{AB}, we get

$$d_{AB} = 2 \cdot \arcsin \sqrt{\left(\sin^2\left(\frac{\Delta Lat_{AB}}{2}\right) + \cos Lat_A \cdot \cos Lat_B \cdot \sin^2\left(\frac{\Delta Lon_{AB}}{2}\right)\right)}$$

A great circle distance has the dimension of an angle (measured at the center of the Earth). To measure d_{AB} in distance units, we multiply it by 40031.6/360 (distance measured in km) or by 60 (distance in nm).

Here is another variant of the haversine formula. This one is (slightly) more accurate because the arctan function is less affected by the limited number of decimals available in electronic calculators than the arcsin function. This is due to the fact that the curve of the arctan function is rather flat near 90° (+ or -) whereas the curve of the arcsin function is very steep in this range.

$$x = \sin^2\left(\frac{\Delta Lat_{AB}}{2}\right) + \cos Lat_A \cdot \cos Lat_B \cdot \sin^2\left(\frac{\Delta Lon_{AB}}{2}\right)$$

$$d_{AB} = 2 \cdot \arctan \frac{\sqrt{x}}{\sqrt{(1-x)}} = 2 \cdot \operatorname{atan2}\left(\sqrt{1-x}, \sqrt{x}\right)$$

(Most spreadsheets use the format atan2(denominator,numerator), either with a comma or a semicolon as a separator.)

Initial Course:

$$C_A = \arccos \frac{\sin Lat_B - \sin Lat_A \cdot \cos d_{AB}}{\cos Lat_A \cdot \sin d_{AB}}$$

If the term $\sin(Lon_B - Lon_A)$ is negative, we replace C_A with 360°-C_A in order to obtain the true course (0°... 360° clockwise from true north).

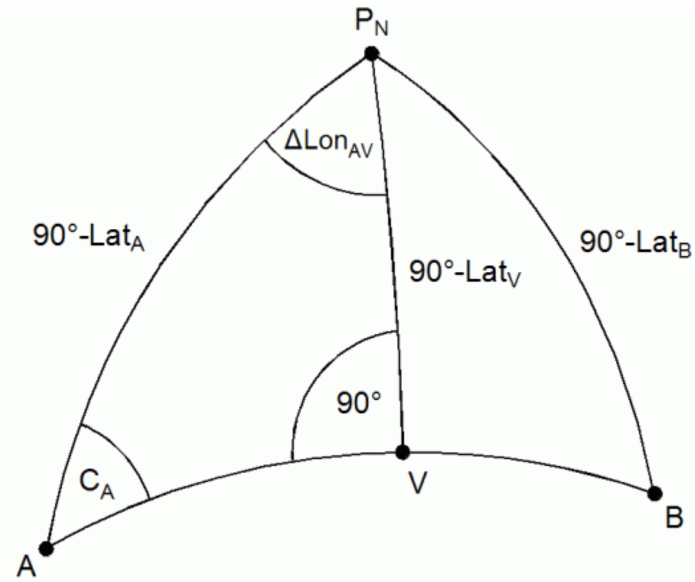
In *Fig. 12-2*, C_A is the **initial great circle course**, C_B the **final great circle course**. Since the angle between the great circle and the respective local meridian varies as we progress along the great circle (unless the great circle coincides with the equator or a meridian), we can **not** steer a constant course as we would when following a rhumb line.

Theoretically, we have to adjust the course continually. This is possible with the aid of navigation computers and autopilots. If such means are not available, we have to calculate an updated course at certain intervals (see below).

Great circle navigation requires more careful voyage planning than rhumb line navigation. On a Mercator chart (see chapter 13), a great circle track appears as a line bent towards the equator. As a result, the navigator may need more information about the intended great circle track in order to verify if it leads through navigable areas.

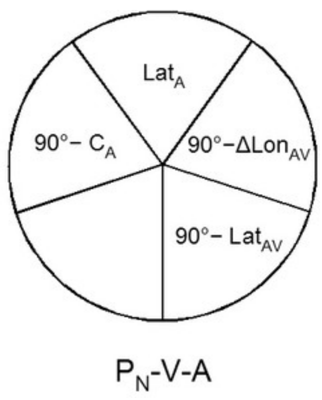
With the exception of the equator, every great circle has two **vertices**, the points farthest from the equator. The vertices have the same absolute value of latitude (with opposite sign) but are 180° apart in longitude. At each **vertex** (also called **apex**), the great circle is tangent to a parallel of latitude, and C is either 90° or 270° (cos C = 0). Thus, we have a right spherical triangle formed by the north pole, P_N, the vertex,V, and the point of departure, A (*Fig. 12-3*).

Fig. 12-3



To derive the formulas needed for the following calculations, we use Napier's rules of circular parts (*Fig. 12-4*). The right angle is at the bottom of the circular diagram. The five parts are arranged clockwise.

Fig. 12-4



First, we need the latitude of the vertex, Lat_V:

$$\cos Lat_V = \sin C_A \cdot \cos Lat_A$$

Solving for Lat_V, we get:

$$Lat_V = \pm \arccos \left(|\sin C_A| \cdot \cos Lat_A \right)$$

The absolute value of sin C_A is used to make sure that Lat_V does not exceed ±90° (the arccos function returns values between 90° and 180° for negative arguments). The equation has two solutions, according to the number of vertices. Only the vertex lying ahead of us is relevant to voyage planning. It is found using the following modified formula:

$$Lat_V = \operatorname{sgn} \left(\cos C_A \right) \cdot \arccos \left(|\sin C_A| \cdot \cos Lat_A \right)$$

sng(x) is the **signum** function:

$$\operatorname{sgn}(x) = \begin{Bmatrix} -1 & \text{if} & x<0 \\ 0 & \text{if} & x=0 \\ +1 & \text{if} & x>0 \end{Bmatrix}$$

If V is located between A and B (like shown in Fig. 12-3), our latitude passes through an extremum at the instant we reach V. This does not happen if B is between A and V.

Knowing Lat_V, we are able to calculate the longitude of V. Again, we apply Napier's rules:

$$\cos \Delta Lon_{AV} = \frac{\tan Lat_A}{\tan Lat_V} \qquad \Delta Lon_{AV} = \Delta Lon_V - \Delta Lon_A$$

Solving for ΔLon_{AV}, we get:

$$\Delta Lon_{AV} = \arccos \frac{\tan Lat_A}{\tan Lat_V}$$

The longitude of V is

$$Lon_V = Lon_A + \operatorname{sgn} \left(\sin C_A \right) \cdot \arccos \frac{\tan Lat_A}{\tan Lat_V}$$

(Add or subtract 360° if necessary.)

The term sng (sin C_A) in the above formula provides an automatic correction for the sign of ΔLon_{AV}.

Knowing the position of V (defined by Lat_V and Lon_V), we are now able to calculate the position of any chosen point, X, on the intended great circle track (substituting X for A in the right spherical triangle). Using Napier's rules once more, we get:

$$\tan Lat_X = \cos \Delta Lon_{XV} \cdot \tan Lat_V \qquad \Delta Lon_{XV} = Lon_V - Lon_X$$

$$Lat_X = \arctan \left(\cos \Delta Lon_{XV} \cdot \tan Lat_V \right)$$

Further, we can calculate the course at the point X:

$$\cos C_X = \sin \Delta Lon_{XV} \cdot \sin Lat_V$$

$$C_X = \begin{Bmatrix} \arccos \left(\sin \Delta Lon_{XV} \cdot \sin Lat_V \right) & \text{if} & \sin C_A > 0 \\ \arccos \left(\sin \Delta Lon_{XV} \cdot \sin Lat_V \right) + 180° & \text{if} & \sin C_A < 0 \end{Bmatrix}$$

Alternatively, C_X can be calculated from the oblique spherical triangle formed by X, P_N, and B.

The above formulas enable us to establish suitably spaced **waypoints** on the great circle and connect them by straight lines on the Mercator chart. The series of **legs** thus obtained, each one being a rhumb line track, is a practical approximation of the intended great circle track. Further, we are now able to see beforehand if there are obstacles in our way.

Mean latitude

Because of their simplicity, the **mean latitude** formulas are often used in everyday navigation. Mean latitude is a good approximation for rhumb line navigation for short and medium distances between A and B. The method is less suitable for polar regions (convergence of meridians).

Course:

$$C = \arctan\left(\cos Lat_M \cdot \frac{Lon_B - Lon_A}{Lat_B - Lat_A}\right) \qquad Lat_M = \frac{Lat_A + Lat_B}{2}$$

The true course is obtained by applying the same rules to C as to the rhumb line course (see above).

Distance:

$$d[km] = \frac{40031.6}{360} \cdot \frac{Lat_B - Lat_A}{\cos C} \qquad d[nm] = 60 \cdot \frac{Lat_B - Lat_A}{\cos C}$$

If C = 90° or C = 270°, we have to use the following formulas:

$$d[km] = \frac{40031.6}{360} \cdot \left(Lon_B - Lon_A\right) \cdot \cos Lat \qquad d[nm] = 60 \cdot \left(Lon_B - Lon_A\right) \cdot \cos Lat$$

Dead Reckoning

Dead reckoning is the navigational term for extrapolating one's new position, B, from the previous position, A, the **course made good, CMD**, and the distance, d. The latter is calculated from the vessel's **speed made good, SMG**, and the time elapsed:

$$d[nm] = \left(T_2[h] - T_1[h]\right) \cdot SMG[kn]$$

1 kn (knot) = 1 nm/h

The position thus obtained is called a **dead reckoning position, DRP**.

Since a DRP is only an approximate position (due to the influence of drift, etc.), the mean latitude method (see above) provides sufficient accuracy. On land, dead reckoning is of limited use since it is usually not possible to steer a constant course (apart from driving in large, entirely flat desert areas).

At sea, the DRP is needed to choose a suitable AP for the intercept method. If celestial observations are not possible and electronic navigation aids not available, dead reckoning may be the only way of keeping track of one's position. Apart from the very simple graphic solution, there are two formulas for the calculation of the DRP.

Calculation of new latitude:

$$Lat_B[°] = Lat_A[°] + \frac{360}{40031.6} \cdot d[km] \cdot \cos CMD \qquad Lat_B[°] = Lat_A[°] + \frac{d[nm] \cdot \cos CMD}{60}$$

Calculation of new longitude:

$$Lon_B[°] = Lon_A[°] + \frac{360}{40031.6} \cdot \frac{d[km] \cdot \sin CMD}{\cos Lat_M} \qquad Lon_B[°] = Lon_A[°] + \frac{d[nm] \cdot \sin CMD}{60 \cdot \cos Lat_M}$$

If the resulting longitude is greater than +180°, we subtract 360°. If it is smaller than -180°, we add 360°.

If our movement is composed of several components (including drift, etc.), we have to replace the terms d·cos C and d·sin C with the following terms:

$$\sum d_i \cdot \cos C_i \qquad \text{and} \qquad \sum d_i \cdot \sin C_i$$

Chapter 13

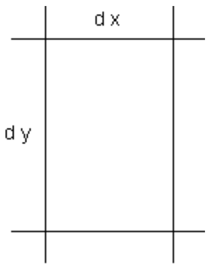
Charts and Plotting Sheets

Mercator Charts

Sophisticated navigation is not possible without the use of a map (chart), a projection of a certain area of the Earth's surface with its geographic features on a plane. Among the numerous types of map projection, the **Mercator projection,** named after the Flemish-German cartographer Gerhard Kramer (Latin: Gerardus Mercator), is mostly used in navigation because it produces charts with an orthogonal grid which is most convenient for measuring directions and plotting lines of position. Further, rhumb lines appear as straight lines on a Mercator chart. Great circles do not, apart from meridians and the equator which are also rhumb lines.

In order to construct a Mercator chart, we have to remember how the grid printed on a globe looks. At the equator, an area of, e. g., 2 by 2 degrees looks almost like a square, but it appears as a narrow trapezoid when we place it near one of the poles. While the distance between two adjacent parallels of latitude is constant, the distance between two meridians becomes progressively smaller as the latitude increases because the meridians converge to the poles. An area with the infinitesimal dimensions dLat and dLon would appear as an oblong with the dimensions dx and dy on our globe (*Fig. 13-1*):

Fig. 13-1



$$dx \; = \; c' \cdot d\,Lon \cdot \cos Lat$$

$$dy \; = \; c' \cdot d\,Lat$$

dx contains the factor cos Lat since the circumference of a parallel of latitude is proportional to cos Lat. The constant c' is the scale factor of the globe (measured in, e. g., mm/°).

Since we require **any** rhumb line to appear as a **straight** line intersecting all meridians at a constant angle, meridians have to be equally spaced vertical lines on our chart, and any infinitesimal oblong defined by dLat and dLon must have the same aspect ratio as on the globe (dy/dx = const.) at a given latitude (**conformality**). Therefore, if we transfer the oblong defined by dLat and dLon from the globe to our chart, we get the dimensions:

$$dx \; = \; c \cdot d\,Lon$$

$$dy \; = \; c \cdot \frac{d\,Lat}{\cos Lat}$$

The new constant c is the scale factor of the chart. Now, dx remains constant (parallel meridians), but dy is a function of the latitude at which our small oblong is located. To obtain the smallest distance from any point with the latitude Lat_P to the equator, we integrate:

$$Y \; = \; \int_0^Y dy \; = \; c \cdot \int_0^{Lat_P} \frac{d\,Lat}{\cos Lat} \; = \; c \cdot \ln \tan \left(\frac{Lat_P}{2} + \frac{\pi}{4} \right)$$

Y is the distance of the respective parallel of latitude from the equator. In the above equation, angles are given in circular measure (radians). If we measure angles in degrees, the equation is stated as:

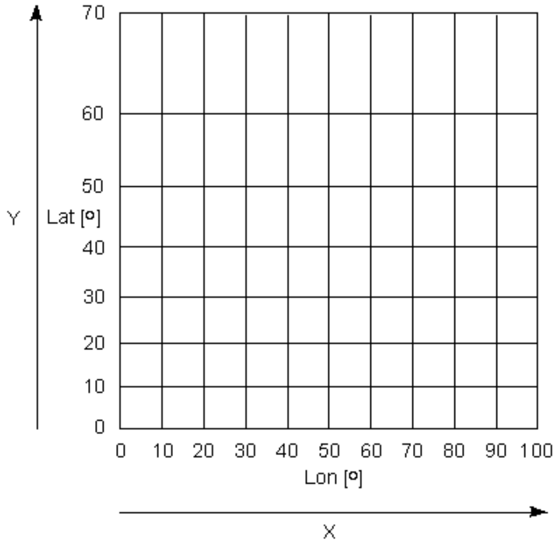
$$Y \; = \; c \cdot \ln \tan \left(\frac{Lat_P[°]}{2} + 45° \right)$$

The distance of any point from the Greenwich meridian (Lon = 0°) varies proportionally with the longitude of the point, Lon_P. X is the distance of the respective meridian from the Greenwich meridian:

$$X = \int_0^{Lon_P} dx = c \cdot Lon_P$$

Fig. 13-2 shows an example of the resulting **graticule** (10° spacing). While meridians of longitude appear as equally spaced vertical lines, parallels of latitude are horizontal lines drawn farther apart as the latitude increases. Y would be infinite at 90° latitude.

Fig. 13-2



Mercator charts have the disadvantage that geometric distortions increase as the distance from the equator increases. The Mercator projection is therefore not suitable for polar regions. A circle of equal altitude, for example, would appear as a distorted ellipse at higher latitudes. Areas near the poles, e. g., Greenland, appear much greater on a Mercator map than on a globe.

It is often said that a Mercator chart is obtained by projecting each point of the surface of a globe from the center of the globe to the inner surface of a hollow cylinder tangent to the globe at the equator. This is only a rough approximation. As a result of such a (purely geometrical) projection, Y would be proportional to tan Lat, and conformality would not be achieved.

Plotting Sheets

If we magnify a small part of a Mercator chart, e. g., an area of 30' latitude by 40' longitude, we will notice that the spacing between the parallels of latitude now seems to be almost constant. An approximated Mercator grid of such a small area can be constructed by drawing equally spaced horizontal lines, representing the parallels of latitude, and equally spaced vertical lines, representing the meridians.

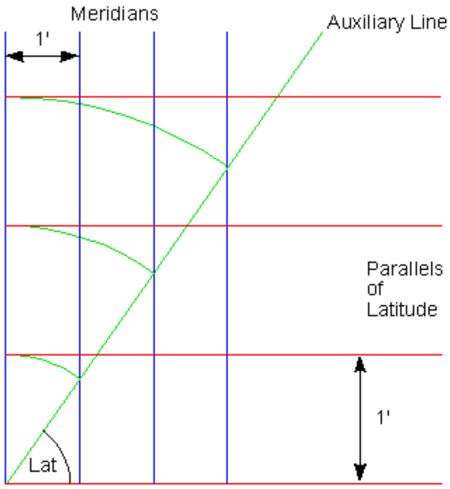
The spacing of the parallels of latitude, Δy, defines the scale of our chart, e. g., 5mm/nm. The spacing of the meridians, Δx, is a function of the mean latitude, Lat_M :

$$\Delta x = \Delta y \cdot \cos Lat_M \qquad\qquad Lat_M = \frac{Lat_{min} + Lat_{max}}{2}$$

A sheet of paper with such a simplified Mercator grid is called a **small area plotting sheet**. It is the commonly used tool for plotting lines of position and finding their intersection point.

If a calculator or trigonometric table is not available, the meridian lines can be constructed with the graphic method shown in *Fig. 13-3*:

Fig. 13-3

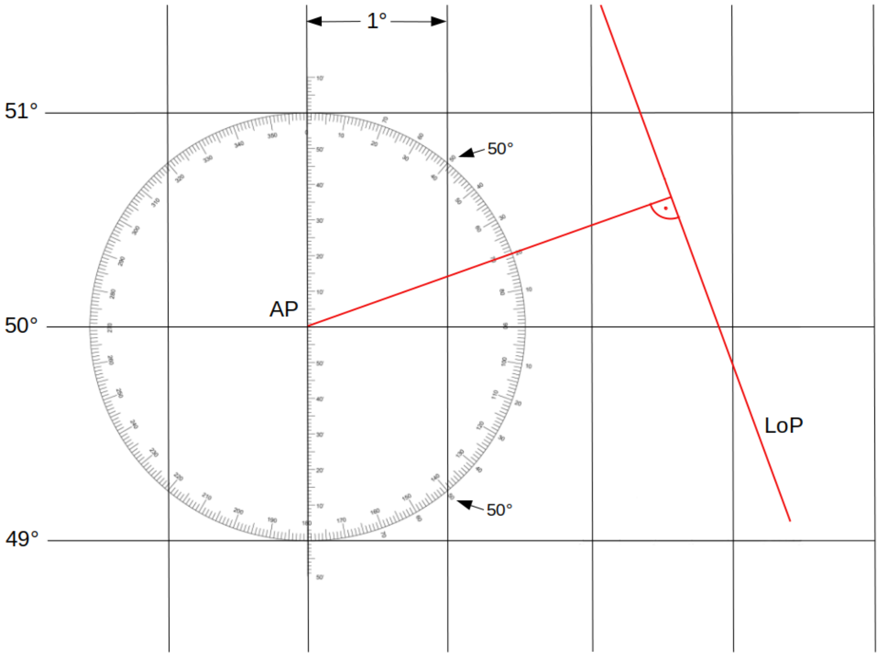


We take a sheet of blank paper and draw the required number of equally spaced horizontal lines (parallels of latitude). A spacing of 1 - 10 mm per nautical mile is recommended for most applications.

Using a protractor, we draw an auxiliary line intersecting the parallels of latitude at an angle numerically equal to the mean latitude. Then we mark the map scale (defined by the spacing of the parallels) periodically on this line, and draw the meridian lines through the points thus located. A compass can be used to transfer the map scale from the chosen meridian to the auxiliary line as demonstrated above.

Universal plotting sheets (for almost all latitudes) with an imprinted graduated circle are available at nautical book stores. These work in the same way as demonstrated in *Fig. 13-3* but do not require a separate protractor to construct the meridian lines. *Fig. 13-4* shows an example where meridian lines for a mid latitude of N 50° have been added. The outer (latitude) scale of the circle measures angles from 0° through 90°.

Fig. 13-4



If we put AP at the center of the circle, we can also plot the azimuth line without a protractor since the inner scale of the circle measures angles clockwise from 0° (N) through 360°. The azimuth displayed in *Fig. 13-4*, for example, is 70°.

Gnomonic Charts

For great circle navigation, the **gnomonic projection** offers the advantage that **any** great circle appears as a straight line. Rhumb lines, however, are curved. A gnomonic chart is obtained by projecting each point on the Earth's surface from the Earth's center to a plane tangent to the surface. Since the distance of a projected point from the point of tangency varies in proportion with the tangent of the angular distance of the original point from the point of tangency, a gnomonic chart covers less than a hemisphere, and distortions increase rapidly with increasing distance from the point of tangency. In contrast to the Mercator projection, the gnomonic projection is non-conformal (not angle-preserving).

There are three types of gnomonic projection:

If the plane of projection is tangent to the Earth at one of the poles (**polar gnomonic chart**), the meridians appear as straight lines radiating from the pole. The parallels of latitude appear as concentric circles. The spacing of the latter increases rapidly as the polar distance increases.

If the point of tangency is on the equator (**equatorial gnomonic chart**), the meridians appear as straight lines parallel to each other. Their spacing increases rapidly as their distance from the point of tangency increases. The equator appears as a straight line perpendicular to the meridians. All other parallels of latitude (small circles) are lines curved toward the respective pole. Their curvature increases with increasing latitude.

In all other cases (**oblique gnomonic chart**), the meridians appear as straight lines converging at the **elevated pole**.

The equator appears as a straight line perpendicular to the central meridian (the meridian going through the point of tangency). Parallels of latitude are lines curved toward the poles.

Fig. 13-5 shows an example of an oblique gnomonic chart.

Fig. 13-5

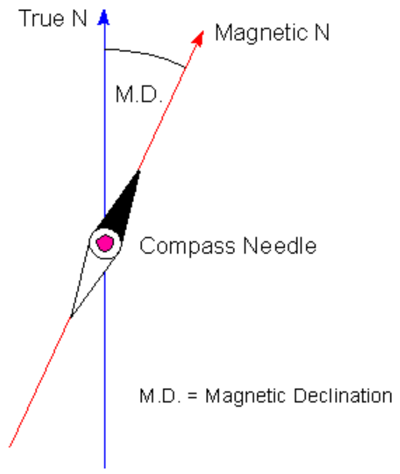
A gnomonic chart is a useful graphic tool for long-distance voyage planning. The intended great circle track is plotted as a straight line from A to B. Obstacles, if existing, become visible at once. The coordinates of the chosen waypoints (preferably those lying on meridian lines) are then read from the graticule and transferred to a Mercator chart, where the waypoints are connected by rhumb line tracks.

Chapter 14

Magnetic Declination

Since the magnetic poles of the Earth do not coincide with the geographic poles and due to other irregularities of the Earth's magnetic field, the horizontal component of the magnetic field at a given position, called **magnetic meridian**, usually forms an angle with the local geographic meridian. This angle is called **magnetic declination** or, in mariner's language, **magnetic variation**. Accordingly, the needle of a magnetic compass, aligning itself with the local magnetic meridian, does not exactly indicate the direction of true north (*Fig. 14-1*).

Fig. 14-1



Magnetic declination varies with the observer's geographic position and can exceed ±30° or even more in some areas. Knowledge of the local magnetic declination is therefore necessary to avoid dangerous navigation errors. Although magnetic declination is often given in the legend of topographic maps, the information may be outdated because magnetic declination varies with time (up to several degrees per decade). In some places, it may even differ from official statements due to local distortions of the magnetic field caused by deposits of lava, ferromagnetic ores, etc.

The time azimuth formula described in chapter 4 is a very useful tool to determine the magnetic declination at a given position. If the observer does not know his exact position, an estimate will suffice in most cases. A sextant is not required for the simple procedure:

1. We choose a celestial body being low in the sky or on the visible horizon, preferably sun or moon. We measure the magnetic compass bearing, B, of the center of the body and note the observation time. The vicinity of cars, steel objects, magnets, DC power cables, etc. has to be avoided since they distort the magnetic field locally.

2. We extract GHA and Dec of the body from the Nautical Almanac or calculate these quantities with a computer almanac.

3. We calculate the meridian angle, t (or the local hour angle, LHA), from GHA and our longitude (see chapter 4).

4. We calculate the true azimuth, Az_N, of the body from Lat, Dec, and t. The time azimuth formula (chapter 4) with its accompanying rules is particularly suitable for this purpose since it does not require an observed or computed altitude.

5. Magnetic declination, MD, is obtained by subtracting Az_N from the compass bearing, B.

$$MD = B - Az_N$$

(Add 360° if the angle thus obtained is smaller than –180°. Subtract 360° if the angle is greater than +180°.)

Eastern magnetic declination (as demonstrated in *Fig. 14-1*) is positive (0°...+180°), western declination negative (0°...–180°).

Having a direction indicated by a magnetic compass (bearing of a landmark, etc.), we have to subtract MG to obtain the true azimuth, Az_N.

Chapter 15

Ephemerides of the Sun

The Sun is certainly the most frequently observed celestial body. Greenwich hour angle and declination of the Sun as well as GHA_{Aries} and EoT can be calculated using the algorithms listed below [8,17]. These simplified, low-precision formulas are useful for navigational calculations with spreadsheets or programmable calculators.

First, the time variable, T, has to be calculated from year, month, day, and UT. T is the time of observation, measured in days and fractions of a day, before or after Jan 1, 2000, 12:00:00 UT:

$$T = 367 \cdot y - \mathrm{INT}\left\{1.75 \cdot \left[y + \mathrm{INT}\left(\frac{m+9}{12}\right)\right]\right\} + \mathrm{INT}\left(275 \cdot \frac{m}{9}\right) + d + \frac{UT[h]}{24} - 730531.5$$

y is the number of the year (4 digits), m is the number of the month, and d the number of the day in the respective month. UT is Universal Time in decimal format (e.g., 12h 30m 45s = 12 + 30/60 +45/3600 = 12.5125). For May 17, 1999, 12:30:45 UT, for example, T is -228.978646. The equation is valid from March 1, 1900 through February 28, 2100.

INT(x) is the greatest integer smaller than or equal to x. For example, INT(3.8) = 3, INT(-2.2) = -3. The INT function is part of many programming languages and spreadsheet programs. It should be checked if the INT function implemented in the chosen software handles negative values in the same way as shown above. Otherwise, the FLOOR function might be tried.

It should further be noted that in many programming languages and spreadsheets, the trigonometric and inverse trigonometric functions calculate in radians (1 radian = π/180°).

Mean anomaly of the Sun*:

$$g[°] = 0.9856003 \cdot T - 2.472$$

Mean longitude of the Sun*:

$$L_M[°] = 0.9856474 \cdot T - 79.540$$

Ecliptic longitude of the Sun*:

$$L_E[°] = L_M[°] + 1.915 \cdot \sin g + 0.02 \cdot \sin(2 \cdot g)$$

Obliquity of the ecliptic:

$$\epsilon[°] = 23.439 - 4 \cdot T \cdot 10^{-7}$$

Declination of the Sun:

$$Dec[°] = \arcsin\left(\sin L_E \cdot \sin \epsilon\right)$$

Right ascension of the Sun*:

$$RA[°] = 2 \cdot \arctan \frac{\cos\epsilon \cdot \sin L_E}{\cos Dec + \cos L_E} = 2 \cdot \mathrm{atan2}\left(\cos Dec + \cos L_E\,,\ \cos\epsilon \cdot \sin L_E\right)$$

Greenwich Mean Sidereal Time (in degrees)*:

$$GMST\,[°] = 100.46062 + 0.98564737 \cdot T + 15 \cdot UT[h]$$

Greenwich hour angle of the Sun*:

$$GHA[°] = GMST[°] - RA[°]$$

Greenwich Apparent Time:

$$GAT[h] = \frac{GHA[°]}{15} + 12\,h$$

(If GAT > 24h, subtract 24h.)

Equation of time:

$$EoT[h] = GAT[h] - UT[h]$$

(If EoT > +0.3h, subtract 24h. If EoT < −0.3h, add 24h.)

Alternative method

We calculate EoT (in degrees) directly from g and L_E. In this case, we do not need RA and GMST:

$$EoT[°] = -1.915 \cdot \sin g - 0.02 \cdot \sin(2 \cdot g) + 2.466 \cdot \sin(2 \cdot L_E) - 0.053 \cdot \sin(4 \cdot L_E)$$

Greenwich hour angle of the Sun*:

$$GHA[°] = 15 \cdot UT[h] + EoT[°] - 180°$$

*All quantities marked by an asterisk (*) have to be within the range between 0° and 360°. If necessary, add or subtract 360° or multiples thereof. This can be achieved using the following algorithm which is particularly useful for programmable calculators:

$$y = 360 \cdot \left[\frac{x}{360} - \mathrm{INT}\left(\frac{x}{360}\right)\right] = \mathrm{mod}(x, 360)$$

Semidiameter and Horizontal Parallax:

Due to the excentricity of the earth's orbit, semidiameter and horizontal parallax of the Sun change periodically during the course of a year. The SD of the Sun varies inversely with the distance earth-Sun, R:

$$R[AU] = 1.00014 - 0.01671 \cdot \cos g - 0.00014 \cdot \cos(2 \cdot g)$$

(1 AU = $149.6 \cdot 10^6$ km)

$$SD['] = \frac{16.02}{R[AU]} \qquad \text{alternatively:} \quad SD['] = \frac{16.02}{1 - 0.017 \cdot \cos g}$$

The mean horizontal parallax of the Sun is approx. 0.15'. The periodic variation of HP is too small to be of practical significance.

Accuracy:

The maximum error of GHA and Dec is about ±0.6' which is accurate enough for marine navigation. Results have been cross-checked with *Interactive Computer Ephemeris 0.51* (accurate to approx. 0.1'). Between the years 1900 and 2049, the error was smaller than ±0.3' in the majority of cases (100 dates chosen at random). EoT was accurate to approx. ±2s. The error of SD is smaller than ±0.1'. By comparison, the maximum error of GHA and Dec of the Sun extracted from the Nautical Almanac is approx. ±0.25' when using the interpolation tables.

Chapter 16

Navigational Errors

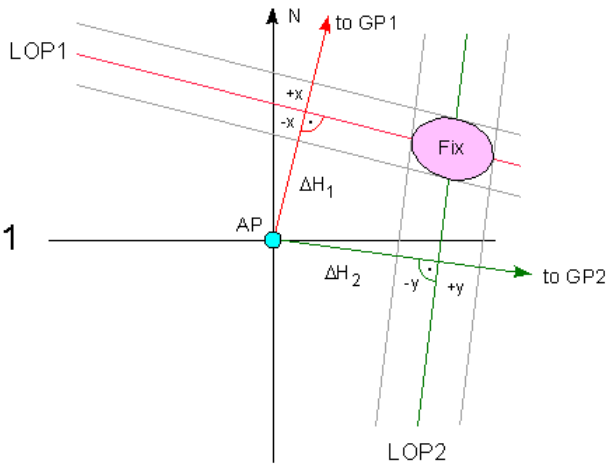
Altitude errors

Apart from systematic errors which can be corrected for the most part (see chapter 2), observed altitudes always contain random errors caused by, e.g., heavy seas, a blurred horizon, abnormal refraction, and the limited optical resolution of the human eye. Although metal sextants are usually manufactured to a mechanical and optical precision of ca. 0.1'- 0.3', the **standard deviation** of altitudes measured with a marine sextant is rather in the magnitude of 1' under fair working conditions. The standard deviation may increase to several arcminutes due to disturbing factors or if a bubble sextant or a plastic sextant is used. Altitudes measured with a theodolite are considerably more accurate (0.1'- 0.2'). For scientific applications, there are precision theodolites which can resolve angles to 1 arcsecond or even less.

Due to the influence of random observation errors, lines of position are more or less indistinct and are better considered as **bands of position**.

Two intersecting bands of position define an **area of position** (ellipse of uncertainty). *Fig. 16-1* illustrates the approximate size and shape of the ellipse of uncertainty for a given pair of position lines. The standard deviations of the intercepts (±x and ±y, respectively) are indicated by grey lines.

Fig. 16-1



The area of position is smallest if the angle between the bands is 90°. The most probable position is at the center of the area, provided the error distribution is symmetrical. Since position lines are perpendicular to their corresponding azimuth lines, objects should be chosen whose azimuths differ by approx. 90° for best accuracy. An angle between 30° and 150°, however, is tolerable in most cases.

When observing more than two bodies, the azimuths should have a roughly symmetrical distribution (**bearing spread**). With multiple observations, the optimum horizontal angle between two adjacent bodies is obtained by dividing 360° by the number of observed bodies (3 bodies: 120°, 4 bodies: 90°, 5 bodies: 72°, 6 bodies: 60°, etc.).

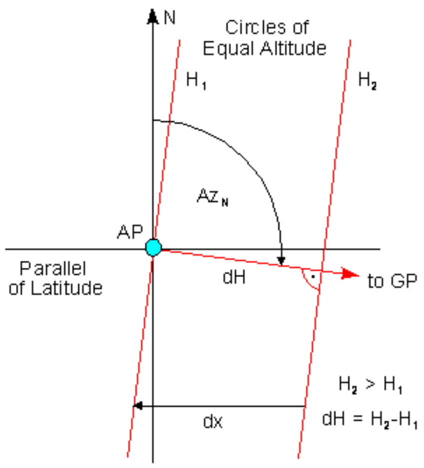
A symmetrical bearing spread not only improves geometry but also compensates for systematic errors like, e.g., the index error.

Moreover, there is an optimum range of altitudes the navigator should choose to obtain reliable results. Low altitudes increase the influence of abnormal refraction (random error), whereas high altitudes, corresponding to circles of equal altitude with small diameters, increase geometric errors due to the curvature of position lines. The generally recommended range to be used is 20° - 70° but exceptions are possible.

Time errors

The time error is as important as the altitude error since the navigator usually presets the instrument to a chosen altitude and stops the time at the instant the image of the body or its rim passes the reference line visible in the telescope. The accuracy of time measurement is usually in the range between a fraction of a second and several seconds, depending on the rate of change of altitude and other factors. Time error and altitude error are closely interrelated and can be converted to each other, as shown below (*Fig. 16-2*).

Fig. 16-2



Due to the Earth‘s rotation, the meridian angle of any celestial body, t, increases by approx. 0.25' per second. The letter d indicates a small (infinitesimal) change of a quantity (see mathematical literature).

$$dt\,[\,'\,] \;=\; 0.25 \cdot dT\,[\,s\,]$$

As the meridian angle changes by dt, the GP of the observed body travels the distance dx along a chosen parallel of latitude:

$$dx\,[\,nm\,] \;=\; \cos Lat \cdot dt\,[\,'\,]$$

dx is also the east-west shift of a circle of equal altitude during the time interval dT (tangents shown in Fig. 16-2). The cosine of Lat is the ratio of the circumference of the parallel of latitude to the circumference of the equator.

The corresponding altitude change is:

$$dH\,[\,'\,] \;=\; \sin Az_N \cdot dx\,[\,nm\,]$$

Thus, the rate of change of altitude is:

$$\frac{dH\,[\,'\,]}{dT\,[\,s\,]} \;=\; 0.25 \cdot \sin Az_N \cdot \cos Lat$$

The altitude changes most rapidly when the observer is on the equator and when the azimuth of the body is 90° (dH/dt positive) or 270° (dH/dt negative). The altitude is constant when the observer is at one of the poles, provided Dec is constant. dH/dt is zero when the azimuth is 0° or 180° which is the case at meridian transit.

An alternative formula is obtained by differentiation of the altitude with respect to the meridian angle, t (see altitude formula, chapter 4):

$$\frac{dH\,[\,'\,]}{dT\,[\,s\,]} \;=\; -\,0.25 \cdot \frac{\cos Lat \cdot \cos Dec \cdot \sin t}{\cos H}$$

In view of the above, it should be clear that is would not make sense to buy an expensive 1-arcsecond theodolite when we are not able to determine the instant of observation with the corresponding precision.

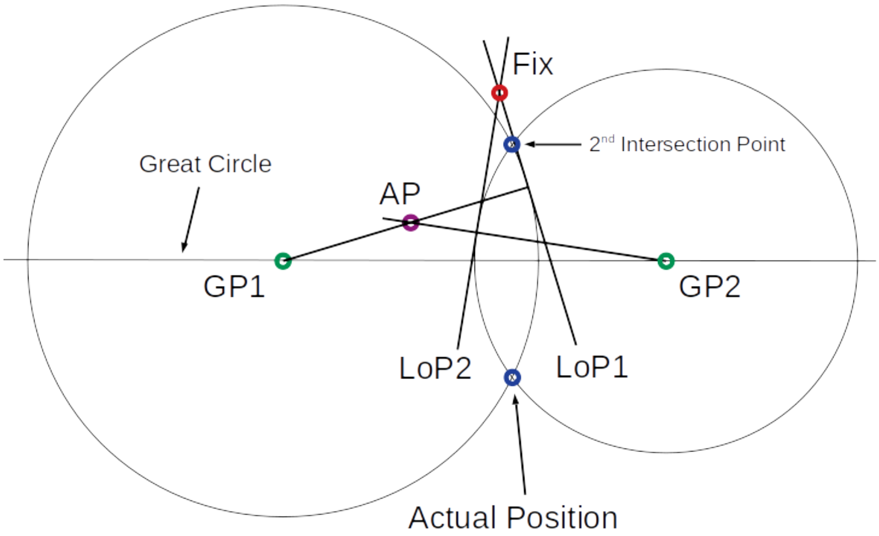
A **chronometer error** is a systematic time error (chapter 17). It influences each line of position so that only the longitude of a fix is affected whereas the obtained latitude remains unchanged, provided the declinations of the observed bodies do not change. A chronometer being 1 s slow, for example, displaces a fix by 0.25' to the west, a chronometer being 1 s fast displaces it by 0.25' to the east. If we know our exact position, we can calculate the chronometer error from the difference between our true longitude and the longitude found by our observations. If we do not know our longitude, the approximate chronometer error can be found by lunar observations (chapter 7).

Ambiguity

Poor geometry may not only decrease accuracy but may even result in an entirely wrong fix. As the observed horizontal angle (difference in azimuth) between two objects approaches 180°, the distance between the points of intersection of the corresponding circles of equal altitude becomes very small (at exactly 180°, both circles are tangent to each other). Circles of equal altitude with small diameters resulting from high altitudes also contribute to a short distance between the two intersection points.

A small distance between both points of intersection, however, increases the risk of ambiguity (*Fig. 16-3*).

Fig. 16-3



In a scenario as described above, there is an increased chance that the assumed position is too far from our actual position. As shown in Fig. 16-3, this may result in a grossly incorrect fix.

If AP is close enough to the actual position, the fix obtained by plotting the LoP's (tangents) will be almost identical with the actual position and is easily improved by iteration if necessary. The accuracy of the fix decreases as the distance of AP from the actual position becomes greater, particularly if AP approaches the great circle going through GP1 and GP2.

If AP is exactly on the great circle going through GP1 and GP2, i.e., equidistant from the actual position and the second point of intersection, the horizontal angle between GP1 and GP2, as seen from AP, will be 180°. In this case, the two LoP's obtained with the intercept method are parallel to each other or even merge with each other, and no fix can be found.

If AP is beyond the great circle going through GP1 and GP2, a fix more or less close to the second point of intersection is obtained. Fig. 16-3 shows such an extremely inconspicuous (but rather unlikely) constellation. The intercept method can not detect which of both theoretically possible positions is the true one.

Iterative application of the intercept method can only improve the fix if the initial AP is closer to the actual position than to the second point of intersection. Otherwise, an "improved" wrong position will be obtained.

Each navigational scenario should be evaluated critically before deciding if a fix is plausible or not. The distance from AP to the observer's actual position has to be considerably smaller than the distance between actual position and second point of intersection. This is usually the case when the above recommendations regarding altitude and horizontal angle are observed.

Evaluating and improving observation results with robust statistics

As described earlier, the precision of a fix can be improved through multiple observations. In this case, the method chosen for averaging is critical since an asymmetric error distribution and/or outliers may distort the result. Using **robust statistics** [23], we can reduce the influence of these error sources in a rather simple way. To apply robust statistics to a series of n measured values, we first have to find their **median**. For this purpose, we order the values from smallest to greatest:

$$x_1 \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n$$

If n is an **odd** number, the median of the data set is identical with the value at the center of the ordered data series:

$$M = x_{\frac{n+1}{2}}$$

If n is an **even** number, there are two values at the center of the data series, and the median is defined as the arithmetic mean of these:

$$M = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

Note that the median is not part of the original data series in the latter case.

Let us assume we have obtained n fixes through multiple observations from the same position (!), as shown in the following table (n=7):

	Latitude	Longitude
Fix 1	35° 15.7'	17° 35.2'
Fix 2	35° 31.0'	17° 36.1'
Fix 3	35° 13.2'	**17° 33.6'**
Fix 4	35° 11.9'	17° 31.9'
Fix 5	35° 19.1'	17° 32.8'
Fix 6	**35° 15.1'**	17° 32.2'
Fix 7	35° 11.6'	17° 34.7'

Table 16-1

The medians of latitude and longitude, respectively, are marked by ellipses. Thus, the coordinates of our improved position are N35°15.1' and E17°33.6'. *Fig. 16-4* shows how this configuration would look on a plotting sheet.

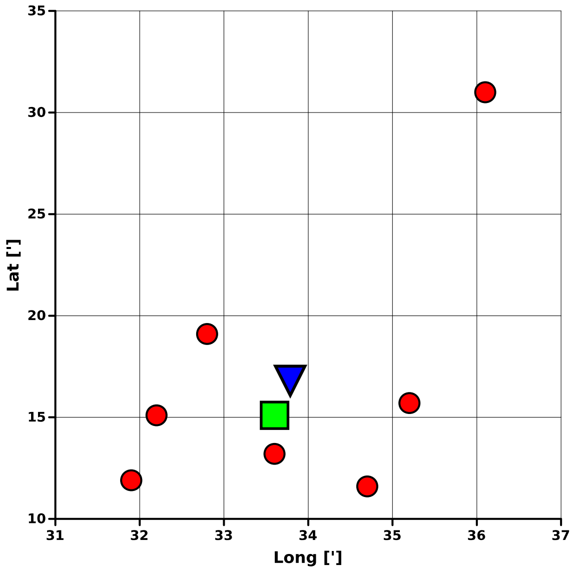


Fig. 16-4

The green square indicates the median position, the blue triangle the (arithmetic) mean position. The position in the upper right corner (Fix 2) is possibly an outlier. It is easy to see how the potential outlier influences the mean position.

With robust statistics, we can also evaluate the precision of our observations and identify outliers (if any). For this purpose, we find the **first quartile**, **Q1**, and the **third quartile**, **Q3**, of all values.

Broadly speaking, Q1 is the median of the lower data half, and Q3 is the median of the upper data half. Thus, Q1 and Q3 are easily found when we have an **even** number of values. For example the first quartile of the value series {1, 2, 5, 8, 9, 15} (n=6) is 2, the median of the series {1, 2, 5}, and the third quartile is 9, the median of the series {8, 9, 15}. Accordingly, the first quartile of the series {1, 2, 5, 8, 9, 15, 20, 25} (n=8) is 3.5, and the third quartile is 17.5.

Unfortunately, things get more complicated when we have an **odd** total number of values. In the latter case one finds various definitions of quartiles in the literature. These yield different results and obviously neither statisticians nor developers of mathematical software agree which one of them represents the value distribution in the best way.

Here are four simple methods for finding the quartiles of a series with an **odd number of values**:

The 1st method excludes the overall median, M, when finding the median of each data half. For example, the median of the ordered value series {1, 2, 5, 8, 9, 15, 17} is 8 (n = 7). Thus, $Q1_{\mathbf{exc}}$ is 2, the median of the series {1, 2, 5}. Accordingly, $Q3_{exc}$ is 15, the median of the series {9, 15, 17}.

The 2nd method includes the overall median when finding the quartiles. Accordingly, with the same data set, $Q1_{\mathbf{inc}}$ is 3.5, the median of the series {1, 2, 5, 8}, and $Q3_{\mathbf{inc}}$ is 12, the median of the series {8, 9, 15, 17}.

The 3rd method combines the two described above by calculating the arithmetic means of the quartiles obtained therewith:

$$Q1_{mean} = \frac{Q1_{exc} + Q1_{inc}}{2} \qquad Q3_{mean} = \frac{Q3_{exc} + Q3_{inc}}{2}$$

With the 3rd method, we obtain $Q1_{mean} = 2.25$ and $Q3_{mean} = 13.5$.

The 4th method ("CDF" method) is another compromise between the first two methods. First, we create two data halves which in turn have to include an **odd number** of values each. This is achieved by either including or excluding the median. M is included if n = 4k+1 whereas M is excluded if n = 4k+3 (k = 1, 2, 3, 4, 5 …). Q1 is the central value of the lower data half and Q3 the central value of the upper data half thus obtained [24]. With the above example (7 = 4k+3), the results would be Q1 = 2 and Q3 = 15. One advantage of the CDF method is that Q1 and Q3 are always elements of the original data set.

It is more or less a matter of personal preference to decide which of the above methods should be applied in case of an odd number of observations. For our purposes any of them will suffice. Methods 3 and 4 are preferred since they give a more accurate representation of a value distribution, particularly when only few values are available.

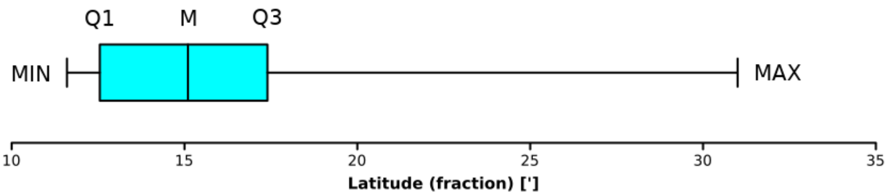
In robust statistics, the quality of our observation data is indicated by the **interquartile range**, IQR, which is the difference between first and third quartile (ignoring the method used to find both values).

$$IQR = Q3 - Q1$$

Of course, the interquartile range, too, is method-dependent since $Q1_{\mathbf{exc}}$ and $Q3_{\mathbf{exc}}$ tend to be farther apart than $Q1_{\mathbf{inc}}$ and $Q3_{\mathbf{inc}}$. As we increase the number of observations, the differences between the above methods will become smaller and approx. 50% of the observed values will be in the interval between Q1 and Q3 while approx. 25% will be smaller than Q1 and approx. 25% greater than Q3.

With a **five-number summary** consisting of the smallest value (MIN), Q1, M, Q3, and the greatest value (MAX), we are able to create a **box plot**, also known as **box-and-whisker plot**. The latter is a simplified graphic representation of the value distribution. *Fig. 16-5* shows a box plot of the latitude values from *Table 16-1* (only arcminute fraction shown). The box represents the data range between (and including) Q1 and Q3. M is marked by a line inside the box. The "whiskers" (also called antennae) attached to the box indicate the smallest and greatest value of the whole data series.

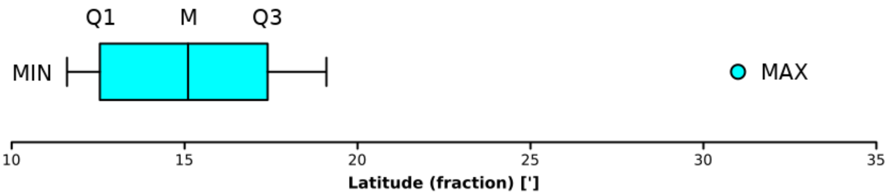
Fig. 16-5



In *Fig. 16-5*, at least one value (MAX) appears to be straying far beyond the others. **By definition, values smaller than Q1−1.5·IQR ("lower fence") or greater than Q3+1.5·IQR ("upper fence") are considered as outliers [23].** Thus, the maximum value in this example is clearly an outlier.

Fig. 16-6 shows a modified box plot in which outliers - only one in this example - are graphically distinguished from the rest of the values. The right "whisker" in *Fig. 16-6* represents the greatest value not being an outlier.

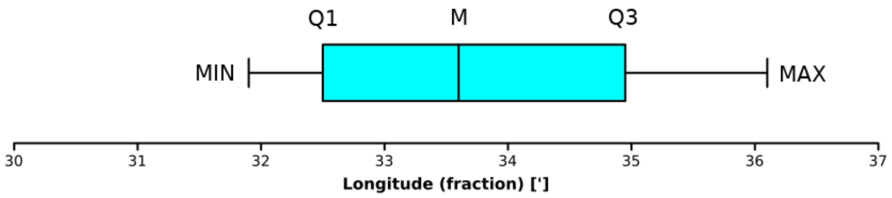
Fig. 16-6



If MAX were closer to Q3 or farther away than shown here, M would still remain the same. Even the presence of a greater number of outliers would not change the median as long as their percentage stays below 50%. However, the usefulness of a data series with such a high proportion of outliers is questionable. Outliers are not removed from the data set when finding the median!

A box plot* of the longitude values from *Table 16-1* does not indicate any outliers (*Fig. 16-7*).

Fig. 16-7



*The above box plots were created with *Gnumeric*, an open-source spreadsheet software. The method to calculate Q1 and Q3 is not one of those shown above which explaines the slightly different results.

In the above example, M is roughly in the middle between Q1 and Q3 which indicates a more or less symmetrical value distribution within this interval.

If M were closer to Q1 or closer to Q3, we would have a more **skewed** distribution. In such a case, too, the median would be much more useful than the arithmetic mean. However, in case of a strongly skewed distribution, the data set should be critically reviewed in order to find the underlying cause.

For the sake of completeness, it should be mentioned that there is another robust estimator, the **trimean**, TM, which is the weighted arithmetic mean of the median and both quartiles:

$$TM = \frac{Q1 + 2M + Q3}{4}$$

The trimean is not quite as robust as the median because it only tolerates up to 25% outliers. However, it is still much better in this respect than the arithmetic mean which tolerates no outliers at all (see *Fig 16-4*). A similar estimator is the **interquartile mean**, IQM, which is the arithmetic mean of all values included in the interval [Q1,Q3]. The interquartile mean is a special case of a **trimmed mean** which is obtained by removing a certain proportion of the highest and lowest values of a data series before calculating the arithmetic mean.

Of course, the navigator is interested to know how many observations are needed to establish a reliable position. This depends on the precision requirements which in turn depend on our situation. A skipper sailing in the vicinity of rocks, shoals, or other obstacles is advised to navigate with the utmost precision whereas he usually does not need to know his position to the arcminute when he is in the middle of an ocean and far from a frequently used traffic route.

In principle, a single value does not tell us anything about the precision of our observation. Thus, it is a matter of confidence in the general circumstances, the condition of the instrument we are using, and our own abilities to decide if we can trust our observation or not. If we deem a single fix unsafe, we may tentatively add one or two more. If the resulting group is reasonably tight, the mean or median position should be reliable enough. If we suspect any outliers to be present, we may wish to increase the number of observations and apply the methods of robust statistics. We should be aware, however, that even a rudimentary data analysis like the five-number summary cannot reasonably be done with three or five values but requires a greater number of measurements – the more, the better. Use critical judgment.

Chapter 17

The Marine Chronometer

Mechanical Chronometers

A marine chronometer is a precise timepiece kept on board as a portable time standard. In former times, the chronometer time was usually checked (compared with an optical time signal) shortly before departure. During the voyage, the chronometer had to be reliable enough to avoid dangerous longitude errors (chapter 16) even after weeks or months of service. Today, radio time signals, e. g., WWV, can be received around the world, and the chronometer can be checked as often as deemed necessary during a voyage. Therefore, a quartz watch of good quality is suitable for most navigational tasks if checked periodically, and the marine chronometer serves more or less as a back-up.

The first mechanical marine chronometer of sufficient precision was built by *Harrison* in 1736. Due to the exorbitant price of early chronometers it took decades until the marine chronometer became part of the standard navigation equipment. In the meantime, the longitude of a ship was mostly determined by lunar distance (chapter 7).

During the second half of the 20th century quartz chronometers replaced the mechanical ones almost entirely because they are much more accurate, cheaper to manufacture, and almost maintenance-free (apart from the annual battery change). Today, mechanical chronometers are valuable collector's items since there are very few manufacturers left.

Fig. 7-1 shows a POLJOT (ПОЛЁТ) 6MX, a traditional chronometer made in Russia. Note that the timepiece is suspended in gimbals to reduce the influence of ship movements (torque) on the balance wheel.

Fig 7-1

Usage

Shortly before an astronomical observation, the navigator starts a stop-watch at a chosen integer hour or minute displayed by the chronometer and makes a note of the chronometer reading. Also before the observation, the sextant or theodolite is set to a chosen altitude (unless a maximum or minimum altitude is to be observed). During the observation itself, the time is stopped at the instant the observed body makes contact with the horizontal reference line in the telescope of the instrument. This may be the sea horizon (sextant) or the cross hairs (theodolite). The sum of the previously noted chronometer time and the time measured with the stop-watch is the chronometer time of observation.

Since there is no guarantee that time signals are available at any time during a voyage, the navigator has to be familiar with the individual characteristics of the board chronometer.

The **chronometer error**, CE, is the difference between **chronometer time** (time displayed by chronometer) and **UT** (or UTC) at a given instant.

$$CE = T_{Chrono} - UT$$

The most important individual characteristic of a chronometer is the **chronometer rate**, the change of the chronometer error during a chosen time interval. The **daily rate** is measured in seconds per 24 hours. For better accuracy, the daily rate is usually obtained by measuring the change of the chronometer error within a 10-day period and deviding the result by 10. The chronometer rate can be positive (chronometer gaining) or negative (chronometer losing). Knowing the initial error and daily rate, we can extrapolate the current UT from the chronometer time and the number of days and hours elapsed since the last chronometer check.

The correction formula is:

$$UT_P = T_{Chrono} - CE_{init} - \left(d + \frac{h}{24}\right) \cdot DR$$

UT_P	Predicted UT
T_{Chrono}	Chronometer time
CE_{init}	Initial chronometer error
DR	Daily rate
d	Integer number of days elapsed since initial chronometer check. These are not calendar days but days elapsed since the exact moment of chronometer check!
h	Additional number of hours elapsed.

The daily rate of a chronometer is not always constant but may be subject to systematic changes. Therefore it should be measured periodically. At the beginning of the service life, the daily rate of a chronometer changes and finally approaches a more or less constant value („running in“). Temperature variations also affect the daily rate. The **temperature coefficient** is the change of daily rate caused by a certain temperature variation. Since Chronometers are temperature-compensated, the temperature coefficient of a mechanical marine chronometer is small, typically about ±0.1 s/day per Kelvin. Further, the daily rate exhibits random fluctuations, called **daily rate variation**, **DRV**. The latter is the difference between the daily rates measured on two consecutive days (day i and day i+1, respectively).

$$DRV_i = DR_{i+1} - DR_i$$

Variations in daily rate are usually small but should be monitored regularly. A daily rate-variation outside the range specified by the manufacturer may indicate a mechanical problem, e. g., abnormal wear. In such a case, the instrument needs overhauling or repair. The standard deviation of a series of n (usually 10) consecutive daily rate variations may be regarded as a “quality index“, ϵ [22].

$$\epsilon = \pm\sqrt{\frac{1}{n}\sum_{i=1}^{n} DRV_i^2}$$

The following table shows a series of measurements made with a mechanical chronometer (POLJOT 6MX #22787) at room temperature. Like most mechanical marine chronometers, the 6MX has a half-second beat. The chronometer time was compared with UTC (radio-controlled watch, 1-second beat). Therefore, all values have been rounded to the next half second.

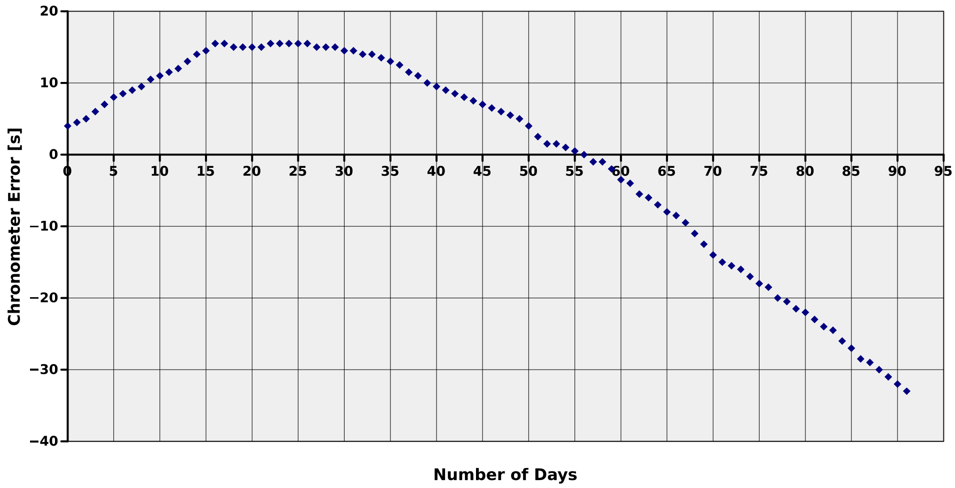
Day	CE [s]	DR [s] (1st Diff.)	DRV[s] (2nd Diff.)	DRV^2 [s^2]
0	41			
		-2		
1	39		-1	1
		-3		
2	36		0.5	0.25
		-2.5		
3	33.5		-0.5	0.25
		-3		
4	30.5		0	0
		-3		
5	27.5		0.5	0.25
		-2.5		
6	25		-0.5	0.25
		-3		
7	22		0.5	0.25
		-2.5		
8	19.5		-0.5	0.25
		-3		
9	16.5		0	0
		-3		
10	13.5		0.5	0.25
		-2.5		
11	11			Sum: 2.75

The mean daily rate of the chronometer (first 10 days) is -2.75 s (specification: ±3.5 s).

The greatest daily rate variation is -1 s (specification: ±2.3 s). The "quality index", ϵ, is ±0.52 s (not specified).

Mechanical chronometers can exhibit a strange behavior. Immediately after the above chronometer had been overhauled by a watchmaker, the author started recording the error on a daily basis. *Fig. 17-2* shows the error as a function of time.

Fig. 17-2



During the first two weeks after the overhaul, the daily rate was approx. +1 s. This was followed by a two-week plateau during which the error remained almost constant. Afterwards, the chronometer gradually became slower and after approx. 60 days, the mean daily rate finally settled at about –1s per day. A possible explanation would be that a tiny quantity of oil leaked out of the balance wheel bearing and slowly spread over the axle or even parts of the wheel, increasing the moment of inertia and thus decreasing the natural frequency of the mechanical oscillator. How this chronometer will behave in the future remains to be seen.

As a corollary, chronometers, particularly mechanical ones, should be closely monitored and checked against a reliable standard (time signal, radio-controlled watch) whenever an opportunity arises.

Quartz Chronometers

A quartz-controlled chronometer (*Fig. 7-2*) is much more reliable than a mechanical one. The specimen shown below has a certified daily rate of +0.1s. This was confirmed by own measurements (mean DR = +0.09s over a period of 200 days).

Fig. 7-2

Radio-controlled Clocks

Radio-controlled clocks and watches have become very popular and may be used to monitor a chronometer. It should be kept in mind, however, that most of these clocks are actually quartz-controlled clocks which are synchronized to UTC through an integrated time signal receiver. In order to reduce power consumption and save batteries, most clocks are not synchronized continually but once every 24 hours. If the quartz movement is of inferior quality, an error of up to several seconds can accumulate between two subsequent synchronizations. Such a clock is of limited use for measuring a current chronometer error since the exact moment of synchronization is usually not known. If it is used to measure the daily rate of a chronometer, this should be done at the same time (UTC) of day so that the daily rate of the quartz movement has no effect on the result. The occasional leap seconds introduced to UTC have to be observed.

Appendix

Literature :

[1] Bowditch, The American Practical Navigator, Pub. No. 9, Defense Mapping Agency Hydrographic/Topographic Center, Bethesda, MD, USA

[2] Jean Meeus, Astronomical Algorithms, Willmann-Bell, Inc., Richmond, VA, USA 1991

[3] Bruce A. Bauer, The Sextant Handbook, International Marine, P.O. Box 220, Camden, ME 04843, USA

[4] Charles H. Cotter, A History of Nautical Astronomy, American Elsevier Publishing Company, Inc., New York, NY, USA (This excellent book is out of print. Used examples may be available at www.amazon.com .)

[5] Charles H. Brown, Nicholl's Concise Guide to the Navigation Examinations, Vol.II, Brown, Son & Ferguson, Ltd., Glasgow, G41 2SG, UK

[6] Helmut Knopp, Astronomische Navigation, Verlag Busse + Seewald GmbH, Herford, Germany (German)

[7] Willi Kahl, Navigation für Expeditionen, Touren, Törns und Reisen, Schettler Travel Publikationen, Hattorf, Germany (German)

[8] Karl-Richard Albrand and Walter Stein, Nautische Tafeln und Formeln (German), DSV-Verlag, Germany

[9] William M. Smart, Textbook on Spherical Astronomy, 6th Edition, Cambridge University Press, 1977

[10] P. K. Seidelman (Editor), Explanatory Supplement to the Astronomical Almanac, University Science Books, Sausalito, CA 94965, USA

[11] Allan E. Bayless, Compact Sight Reduction Table (modified H. O. 211, Ageton's Table), 2nd Edition, Cornell Maritime Press, Centreville, MD 21617, USA

[12] The Nautical Almanac, US Government Printing Office, Washington, DC 20402, USA (contains not only ephemerides but also formulas and tables for sight reduction)

[13] Nautisches Jahrbuch oder Ephemeriden und Tafeln (German), Bundesamt für Seeschiffahrt und Hydrographie, Germany

[14] https://time.gov, maintained by the National Institute of Standards and Technology (NIST)

[15] IERS Rapid Service, https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html

[16] Hannu Karttunen et al., Fundamental Astronomy, 4th Ed., Springer Verlag Berlin Heidelberg New York, 2003

[17] The Astronomical Almanac for the Year 2002, US Government Printing Office, Washington, DC 20402, USA

[18] Michel Vanvaerenberg and Peter Ifland, Line of Position Navigation, Unlimited Publishing, Bloomingdale, Indiana, 2003

[19] George H. Kaplan, The Motion of the Observer in Celestial Navigation, Astronomical Applications Department, U. S. Naval Observatory, Washington DC

[20] Ed William's Aviation Formulary, http://www.edwilliams.org/avform.htm

[21] William Chauvenet, A Manual of Spherical and Practical Astronomy, Vol. I + II (reprint of the edition published in 1887 by J. B. Lippincot Company), Elibron Classics Series, Adamant Media Corporation, 2005

[22] N. Liapin, A Method of Determining the Mean Accidental Variations in Daily Rates of a Number of Chronometers, Monthly Notices of the Royal Astronomical Society 1919, Vol.80, p.64

[23] John W. Tukey (1977), Exploratory Data Analysis. Addison-Wesley, ISBN 0-201-07616-0

[24] Eric Langford (2006) Quartiles in Elementary Statistics, Journal of Statistics Education, 14:3, , DOI: 10.1080/10691898.2006.11910589

[25] https://aa.usno.navy.mil/faq/docs/RST_defs.php

GNU Free Documentation License

Version 1.3, 3 November 2008

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies.

The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a

Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  YEAR  YOUR NAME.
Permission is granted to copy, distribute and/or modify this document
under the terms of the GNU Free Documentation License, Version 1.3
or any later version published by the Free Software Foundation;
with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
A copy of the license is included in the section entitled "GNU
Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with … Texts." line with this:

```
with the Invariant Sections being LIST THEIR TITLES, with the
Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.