

Canadian Board of Examiners for Professional Surveyors
Core Syllabus Item
C 6: GEODETIC POSITIONING

Study Guide:

The material is organized in sections that reflect the recommended progression of study. Often questions are followed by some additional explanations which are by no means exhaustive and do not replace the study of the references. They are meant to help linking different pieces of information from different textbooks in order to get a better insight. It is recommended to first answer the questions before reading the explanations.

The topics covered are:

1. Coordinate systems (revision)
2. The Earth's gravity field
3. Geodetic reference systems
4. Time scales and astronomy
5. Geopotential numbers and orthometric heights
6. Gravity measurement
7. Calculation of coordinates from terrestrial measurement
8. Space Reference systems used worldwide and in Canada.
9. The Global Positioning System and other Global Navigation Satellite Systems

1. Coordinate systems (revision)

Review the details of the following coordinate types and their related transformations:

- 3D Cartesian geocentric coordinates (xyz)
- Ellipsoidal coordinates ($\varphi \lambda h$)
- Planar coordinates related to a conformal projection ($x_p y_p$)
- Transformation between 3D-Cartesian and ellipsoidal coordinates
- Transformation between ellipsoidal and map projection coordinates

References:

[Torge (2001): chapter-4.1: The Rotational Ellipsoid]

[Hofmann-Wellenhof et al. (2005): chapter-5.6.1: Coordinate transformation: Ellipsoidal and rectangular coordinates]

[Hofmann-Wellenhof et al. (2007): chapter-8.2.3: Ellipsoidal coordinates and plane coordinates]

Additional explanations:

3D Cartesian coordinates xyz are mathematically the most simple type of coordinates. Modern space techniques, like GPS for example, allow direct determination of this type of coordinates. However, they are not very appropriate for most surveying applications.

One can associate a 3D Cartesian coordinate system to an oblate spheroid by making its centre coincide with the origin of the 3D Cartesian coordinate system. An oblate spheroid is a quadric surface obtained by rotating an ellipse about its minor axis. We align the minor axis with the z -axis. In addition, 2 parameters defining the shape of the oblate spheroid, also called rotational ellipsoid, are needed. Normally the semi-major axis and the flattening are used. This allows introducing ellipsoidal coordinates, the latitude, the longitude, and the height above the ellipsoid.

Passing from ellipsoidal coordinates to map projection coordinates reduces from 3D to 2D, only $(\varphi \lambda)$ are needed to obtain $(x_p y_p)$. No height is involved. This transformation introduces distortions. You should review the concept of scale factor and the convergence of meridians related to conformal map projections.

Sample Questions:

- Q1.1. What is the order of magnitude of the maximal convergence of meridian in the UTM projection?
- Q1.2. Where does this maximum occur?
- Q1.3. How much is the convergence at the central meridian?
- Q1.4. What is the order of magnitude of the maximal scale factor of the UTM projection?
- Q1.5. Where does this maximum occur?
- Q1.6. What is the scale factor at the central meridian?

2. The Earth's gravity field

Demonstrate an understanding of the fundamental concepts of physics, related to the gravity field of the Earth.

A good overview is given in:

[Hofmann-Wellenhof et al. (2005): chapter-2.1: Gravity]

[Hofmann-Wellenhof et al. (2005): chapter-2.2: Level surfaces and plumb line]

[Torge (2001): chapter-3.1: Fundamentals of gravity field theory]

[Torge (2001): chapter-3.2: Geometry of the gravity field]

[Torge (2001): chapter-3.4: The Geoid]

Sample Questions:

- Q2.1. What causes gravity?
- Q2.2. What are the units used to express gravity?
- Q2.3. What is the relation between gravity and attracting force?
- Q2.4. How is the potential of the Earth's gravity field, also called geopotential, defined?
- Q2.5. What are the units related to the geopotential?
- Q2.6. Why is a level surface also called a equipotential surface?
- Q2.7. What is the fundamental relationship between the potential and the gravity vector?
- Q2.8. Do equipotential surfaces intersect? Why?
- Q2.9. Are equipotential surfaces equidistant? Why?

Q2.10. On an equipotential surface is the potential the same everywhere? Why?

Q2.11. On an equipotential surface is the gravity the same everywhere? Why?

Q2.12. What is the geoid?

Additional explanations:

The Newtonian attraction of masses causes gravity. However, the distribution of the masses in the interior of the Earth is inhomogeneous. This leads to equipotential surfaces with a complicated shape, the most important one being the geoid. It corresponds to the mean surface of the oceans. So at least over the oceans this surface is visible. It is clear that the geoid exists also under the continents. This definition is no longer valid today because satellite altimetry missions show the temporal variability of the sea's surface. However, it is still the starting point for developing a good comprehension of what the geoid represents, and of its predominant role as reference surface for height.

If one considers a perfect liquid on which only the Newtonian attraction of its own mass acts, this liquid takes the shape of a perfect sphere. Note that its surface is necessarily an equipotential surface. If one puts this liquid in slow rotation around an axis, it will deform by flattening at the poles, the regions around the rotation axes, and bulging at the equator, the regions in and around the plane perpendicular to the rotation axis going through the center of mass of the liquid. Again the surface of the liquid will correspond to an equipotential surface. Such a surface can be approximated by an ellipsoid. The Earth is not such a perfect liquid. In addition there are other celestial bodies like the Sun and the Moon acting on the Earth. Nevertheless the ellipsoidal model is a good approximation and plays an important role in geodesy. We end up with 2 different reference surfaces, the ellipsoid as reference surface for horizontal coordinates and the geoid for height. This is the classical approach of the pre-satellite era. The use of artificial satellites for coordinate determination has revolutionized this approach by making the direct determination of 3D Cartesian coordinates possible.

You should now read:

[Torge (2001): chapter-1.3: Historical development of Geodesy].

It gives good information of how the model of the Earth evolved from a spherical to an ellipsoidal and then to a geoidal one, as well as introducing the problems related to modern 3D approaches to actual coordinate determination.

3. Geodetic reference systems

This topic is covered by:

[Torge (2001): chapter-2: Reference systems]

The entire chapter, including the following three sections (time systems, ICRS and ITRS, and gravity field), should be studied carefully.

Time systems:

[Torge (2001): chapter-2.2: Time systems]

Sample Questions:

Q3.1. What does TAI stand for?

Q3.2. How is TAI realized?

- Q3.3. What is Dynamical Time?
- Q3.4. What is Sidereal Time?
- Q3.5. What is Universal Time (especially UT1)?
- Q3.6. How is Sidereal Time defined?
- Q3.7. How is Sidereal Time related to UT1?
- Q3.8. How is UTC defined?
- Q3.9. What is UTC used for?

Additional explanations:

Time plays a fundamental role in modern geodesy. All space borne techniques rely on very accurate time measurements. This is the case for VLBI, GNSS, and Satellite Laser Ranging for example. Atomic clocks are a core component of all of these methods. A network of atomic clocks assures a reference time called TAI. So TAI relies only on atomic clocks. Relativistic effects, time dilatation due to speed makes it necessary to associate an origin for the time system, which leads to the distinction of barycentric time and geocentric time. Note that all clocks are installed on the surface of the Earth, which means that BDT for example cannot be observed directly (impossible to put an atomic clock at this location). It is obtained by applying relativistic corrections. It is important to account for these very small differences in orbit determination, but all time systems discussed until now are fundamentally similar in the sense that they are derived from atomic clocks.

Sidereal time is not a time system strictly speaking. It is the angular position of the Earth with respect to space, given in units of time. To be more precise, it is the angle in the equatorial plane between the direction of the vernal equinox and the Greenwich meridian. The answer to why it is then called time has to do with the historical development: As long as you consider the angular velocity of the Earth as constant, this angle changes linearly with time, it becomes a measure for time. Until the invention of atomic clocks there was no phenomenon known that was more regular than the Earth's spin (rotation around its axis). Time keeping was done by angular observations on stars exploiting this celestial clock. Only the stability of modern atomic clocks allows for decoupling the angular position of the Earth from the time. The name remained for the units also, and the angular position of the Earth is still called sidereal time and given in hours, minutes, and seconds. As mentioned the vernal direction, or direction of the vernal equinox, is used as a reference direction. Since precession and nutation affect this direction, one has to distinguish between mean sidereal time (without nutation) and apparent sidereal time (including nutation).

Universal time is obtained from the observed sidereal time, i.e. the angular position of the Earth. As mentioned previously this is done by a combination of different space borne techniques. Currently, UT0 and UT2 are no longer needed, only UT1 remains. [Torge (2001): page 23-24] is slightly out-of-date in this regard.

ICRS and ITRS:

- [Torge (2001): chapter-2.4: Celestial Reference System]
- [Torge (2001): chapter-2.5: Terrestrial Reference System]
- [Hofmann-Wellenhof et al. (2007): chapter-2: Reference systems]
- [Torge (2001): chapter-2.3: International Earth Rotation Service]

Sample Questions:

- Q3.10. How is the celestial reference system defined?
- Q3.11. What does ICRS stand for?
- Q3.12. How is the global earth-fixed geocentric system defined?
- Q3.13. What does ITRS stand for?
- Q3.14. What are precession and nutation?
- Q3.15. What are polar motions?
- Q3.16. How does ICRS and ITRS relate to one another?

Additional explanations:

The ICRS is the international celestial reference system. It is inertial by definition, which means it does not rotate with respect to space. The ITRS is the International Terrestrial reference system. It is Earth-fixed and rotates with the Earth. The rotation axis of the Earth is not pointing in the same direction over time. It changes its direction with respect to space due to the precession and nutation caused by the Newtonian attraction of the Sun and Moon on the equatorial bulge, which is asymmetrical with respect to the ecliptical plane. The angular velocity of the Earth varies slightly. Seen from a stable Earth, the rotation axis is not constant either, due to the polar motion. In a physically more meaningful sense, one should say that, seen from space, the Earth as a whole is wobbling around the rotation axis. The transformation between the ICRS and the ITRS is therefore composed by basically 3 rotation matrices (see Torge (2001): equation 2.16). A first one depending on the precession-nutation ($N(t)P(t)$), the main rotation $R_z(\text{GAST})$, which accounts for the angular position of the Earth, and a last one accounting for the polar motion $R_2(-x_p) R_1(-y_p)$.

Precession-nutation, variation of the angular speed, and polar motion are called the EOP Earth rotation parameters. Precession-nutation can be obtained from models. The other parameters cannot be predicted at least over long time periods. All the parameters are now monitored by space geodetic methods like VLBI, GPS, etc. using permanent global networks. This is coordinated by the IERS. IERS has a very informative web-page that is easy to find. Determining these parameters was formerly carried out using astrogeodetic angular observations.

Gravity field:

[Torge (2001): chapter-2.6: Gravity Field related reference systems]

[Hofmann-Wellenhof et al. (2005): chapter 2.4: Natural coordinates]

Sample Questions:

- Q3.17. How are astronomical latitude and longitude defined?
- Q3.18. How is a local astronomical horizontal system defined?
- Q3.19. How is it related to a global geocentric system?
- Q3.20. How is a local geodetic horizontal system defined?
- Q3.21. How is it related to a global geocentric system?
- Q3.22. What is the difference between the local astronomical and the local geodetic system?
- Q3.23. What are natural coordinates?
- Q3.24. Why are they still important today even with GPS?

Additional explanations:

The starting point is to be aware that gravity is a physical phenomenon. It allows us to distinguish between top and bottom. Through the use of a plumb bob you can visualize easily the direction of the gravity vector at your location. The majority of geodetic measurements refer to the gravity field of the Earth through this orientation, which is given by two Eulerian angles the astronomic latitude Φ and the astronomic longitude Λ . Therefore it is convenient to introduce a local astronomic horizontal system, which can be related to the global geocentric system, implying rotation matrices depending on Φ and Λ .

In the local astronomic system the vertical is defined by the direction of the gravity vector given by (Φ, Λ) . In the local geodetic system the vertical is defined by the direction of the normal to the ellipsoid, given by (φ, λ) . They differ slightly. The differences are the deflection of the vertical. Be aware of the fact that they depend on the ellipsoidal datum used. Compare [Torge (2001): Fig.2.15] and [Torge (2001): Fig.4.6].

4. Time scales and astronomy

You should study the following chapters:

[Torge (2001): chapter-2.6.2: Local astronomic Systems]

[Torge (2001): chapter-5.3: Geodetic Astronomy]

Sample Questions:

- Q4.1. Do equipotential surfaces intersect? Why?
- Q4.2. How are astronomic latitude and longitude defined?
- Q4.3. How is an astronomic azimuth defined?
- Q4.4. What are the fundamental equations in the determination of astronomic latitude and longitude?
- Q4.5. Define declination and right ascension of a star.
- Q4.6. Define the hour angle of a star.
- Q4.7. What is the difference between Greenwich sidereal time and local sidereal time?
- Q4.8. What is the difference between a mean sidereal time and an apparent sidereal time?
- Q4.9. Explain the observation and computation of astronomic latitude and longitude.
- Q4.10. Explain the observation and computation of astronomic azimuth, using Polaris or the Sun.

Additional explanations:

Theoretically, astronomic latitude and longitude do not correspond to positions, but rather to the orientation of the gravity vector. This is an important one to remember. Only if one neglects the deflection of the vertical and considers astronomical and ellipsoidal quantities to be equal, one may interpret them as a position. In this case the resulting resolution cannot be better than the error introduced by neglecting the deflections of the vertical. If one assumes an order of magnitude of several arcseconds for the deflections, the resulting resolution is in the range of several hundred metres. This was the typical procedure in navigation during the eighteenth and nineteenth centuries, at a time where the resolution of the instruments were still poor enough to neglect the deflection of the vertical. Later, the determination of latitude and longitude played an important role in the establishment of horizontal networks. They were necessary in order to reduce the measured astronomical azimuth to ellipsoidal azimuth.

and to constrain the network at a fundamental station. The orientation of the network was obtained by measuring astronomical azimuths.

Until GPS become available as surveying tool, the orientation of local networks in remote areas was often done by measuring an astronomical azimuth. Today, astronomical measurements have become obsolete at least as a common surveying tool. They still play a certain role in local geoid determination and in very highly accurate special networks. Also, much information still used today, especially azimuths or bearings, has been based on astronomical methods. Torge (2001): [chapter-5.3: Geodetic Astronomy] gives a short but good overview.

The starting point for the determination of the above mentioned astronomical quantities are the fundamental equations given in [Torge (2001): equ.2.21-2.23]. There are a lot of parameters which appear. Note which parameters are known, which parameters are unknowns, and which parameters are observations. This will help you when studying [Torge (2001): chapter-5.3: Geodetic Astronomy].

5. Geopotential numbers and orthometric heights

You should study in detail:

[Hofmann-Wellenhof et al. (2005): chapter-4 Heights]

A less explicit discussion can also be found in [Torge (2001): chapter-6.4 Height determination]. This is an important topic which needs some care to master the concepts related to the gravity field of the Earth.

Additional information on levelling can be found in:

[Torge (2001): chapter 5.5.3 Levelling]

Sample Questions:

- Q5.1. How are astronomical latitude and longitude defined?
- Q5.2. Why is there a theoretical misclosure in a conventional spirit levelling in a loop?
- Q5.3. What are geopotential numbers?
- Q5.4. What is the unit used to express geopotential numbers?
- Q5.5. How are geopotential numbers obtained?
- Q5.6. What are the types of measurements necessary to determine geopotential numbers?

Additional explanations:

The theoretical misclosure is due to the fact that the equipotential surfaces are not equidistant [Fig. 4.2 page 158 in Hofmann-Wellenhof et al. (2005)]. The next step is to remedy this situation. Spirit levelling alone is not sufficient. Additional observations are necessary. One has to observe gravity along the levelling lines. Gravity is observed by using gravity meters, mentioned later. For the time being it is sufficient to know that the observed value corresponds to the amount of gravitational attraction [units = gals] in the direction of the plumb line. Combining observed height differences with gravity observations leads to geopotential differences and finally to geopotential numbers. This means that in order to obtain the geopotential difference between two points, A and B, one needs to do conventional spirit levelling plus surface gravity measurements along the same line from A to B.

By definition, the geopotential number is zero on the geoid. This is used in the establishment of vertical datums, by including a tide gauge station as a bench mark, preferably one with a long history in order to get a good mean value of the sea level.

Read also:

[Torge (2001): chapter-7.2 Vertical control networks].

It is essential to be aware of the fact that, prior to the adjustment of a levelling network, the observed raw height differences [unit = metres] can be transformed to geopotential differences [unit = gpu]. The adjustment is then done using geopotential differences instead of height differences leading to geopotential numbers for all bench marks of the vertical control network. This is the first step. We have now geopotential numbers for every bench mark of the vertical datum.

Sample Questions:

Q5.7. Do equipotential surfaces intersect? Why?

Q5.8. How are dynamic heights defined?

Q5.9. How are dynamic heights related to geopotential numbers?

Q5.10. What are orthometric heights?

Q5.11. How are they obtained from geopotential numbers?

Additional explanations:

The next step is to convert the geopotential numbers into dynamic heights, orthometric heights or normal heights. Dynamic heights differ from geopotential numbers only in the scale or the unit. They do not have any geometrical meaning. The most important are orthometric heights because they use the geoid as reference surface. However, the conversion from geopotential numbers to orthometric heights involves the mean value of the gravity along the plumb line between the bench mark situated on the topography and on the geoid. The only accessible point is the bench mark itself. In order to obtain a mean value one has to do some assumptions on the variation of "g" along the plumb line. The simplest approximation consists of replacing the terrain by an infinite layer of constant density (the "Bouguer plate"). The mass attraction of the Bouguer plate can be expressed by an analytical function which allows obtaining a closed expression for the mean value of "g" along the plumb line. Applying this approximation one obtains Helmert-heights.

If one goes through the way geopotential numbers and, further on, the way orthometric heights are obtained, no ellipsoidal model at all is needed. This is important to notice. This approach is the old way of how a vertical datum was established.

6. Gravity measurement

You should study:

[Torge (2001): chapter-5.4.1 Absolute Gravity measurements]

[Torge (2001): chapter-5.4.2 Relative Gravity measurements]

[Torge (2001): chapter-3.5 Temporal Gravity Variations]

All the details of the different instruments are beyond the scope of this item. However, you should master the basic functional principles on which absolute gravimeters (especially the

most frequently used FG5) and relative gravimeters rely, the operating mode of this type of instrument and the obtainable accuracy.

Additional explanations:

Gravity measurements may be used for different purposes. As mentioned above, gravity observations along levelling lines are necessary in order to obtain geopotential numbers in the establishment of a vertical datum. For these applications, relative gravimeters are commonly used. This means that the observations have to be tied to a known station. Gravity is affected by tides, which means that the measurements have to be corrected for this effect. On the other hand permanently operating gravity meters (relative or absolute instruments) are used to observe the tidal variations.

7. Calculation of coordinates from terrestrial measurement

You should study:

[Torge (2001): chapter-5.5.1: Horizontal and vertical angle measurement]

[Torge (2001): chapter-5.5.2: Distance measurements, total stations]

[Torge (2001): chapter-6.4: Horizontal positioning]

Additional explanations:

On the one hand there are the measurements as obtained by a total station which are basically horizontal directions, vertical angles and distances. If the horizontal directions are oriented, these measurements can be viewed as spherical coordinates with respect to the local astronomical system. On the other hand, there are the unknown coordinates. The measurements are normally carried out to determine coordinates. Then there is the question of which type of coordinates is used.

There are different coordinates systems which may be considered:

- 3D Cartesian coordinates with respect to the local astronomical horizontal system;
- 3D Cartesian coordinates with respect to the local ellipsoidal horizontal system;
- 3D Cartesian geocentric equatorial coordinates;
- 2D ellipsoidal coordinates;
- 2D plane coordinates related to a conformal projection.

The choice depends on the type of application. Cartesian coordinates in the local horizontal system are practically never used as a final product. However, they play an important role in establishing the relationship between the observations and the 3D Cartesian geocentric equatorial coordinates.

From the point of view of adjustment theory, there are observations and unknowns (coordinates and may be other unknowns like orientation parameters). In order to perform an adjustment, one has to find the mathematical expressions which relate the measurements to the unknown coordinates and which depend on the choice of the coordinate type. In addition, the raw measurements have to be reduced to the same coordinate surface as the equations. (A simple example: The slant distance is measured, but in the equation the horizontal distance is used making a reduction from the slant distance to a horizontal distance necessary).

There are three major models using:

1. 3D Cartesian geocentric equatorial coordinates: These coordinates are used in conjunction with space borne techniques e.g. GPS positioning, not as frequently with terrestrial measurements. There are, however, special applications like high precision networks in mountainous areas, or base networks for tunnels, which are applications where this type of coordinate plays a role. One should mention that this model is the most powerful and flexible of all three. It allows for combining terrestrial observations and GPS determined coordinates in a simple and rigorous way. The drawback: this model is also the most demanding and difficult.
2. 2D ellipsoidal coordinates: These coordinates were used exclusively in the establishment of classical 2D datum. It implies that all the terrestrial and astronomical measurements are first reduced to the ellipsoid and the adjustment is then performed using ellipsoidal latitude and longitude of the sites as unknowns. Since the networks related to such datum normally spread over at least some hundreds of kilometres, often thousand kilometres and more, the use of projection coordinates would be inappropriate. A national datum is no longer established this way.
3. 2D plane coordinates related to a conformal projection: These coordinates are used in the overwhelming majority of surveying applications, because of their simplicity.

Additional explanations:

3D Cartesian geocentric equatorial coordinates: The mathematical models are given in [Torge (2001): pp.42-44, equ.6.39]. The raw measurements have only to be reduced geometrically for the effect of the height of the instrument and the reflector/target above the ground points. Note that the astronomical latitude and longitude have to be used in this model. There are different possibilities to deal with this problem: 1) The astronomical latitudes and longitudes are measured; 2) the astronomical latitudes and longitudes are obtained from ellipsoidal latitudes and longitudes (which are always implicitly known), corrected for the deflection of the vertical taken from a model; 3) The ellipsoidal latitude and longitude are used instead, which is equivalent to neglecting the deflection of the vertical, which introduces an error.

2D ellipsoidal coordinates: The mathematical models are given in [Torge (2001): chapter-6.3.3 Computation on the ellipsoid]. The complexity of the models comes from the fact that the observation equations, the equations which relate the observations to the unknown coordinates are not closed formulas, in the solution of the inverse problem on the ellipsoid. There are several different approaches, some based on iterations and others on series expansions. The details of this computation, are beyond this item. However, you should be able to numerically calculate spherical approximations. If this model is used, all measurements have to be reduced first to the ellipsoid. [Torge (2001): chapter-6.3.2 Reductions to the ellipsoid; Hofmann-Wellenhof et al. (2005): chapter 5.12 Reduction of the astronomical measurements to the ellipsoid].

2D plane coordinates in a conformal projection: The mathematical model related to this coordinate type is likely the familiar. Here as a reminder:

$$\text{distance: } d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

$$\text{grid azimuth: } a = \arctan((Y_2 - Y_1)/(X_2 - X_1))$$

This model is the most simple, but one should be aware that even if the equations apparently involve a Euclidian metric, the metric is NOT Euclidian and these equations are only approximations which work fine for small distances. In a first step, the measurements are reduced to the ellipsoid. In the second step they are reduced to the projection. This second

step consists of applying a scale factor to the distance, and correcting for the convergence of meridians for passing from a geodetic azimuth to a grid azimuth.

A general remark: Independent of the model, if the horizontal directions are not oriented, additional unknown parameters for the orientations have to be considered.

8. Space Reference systems used worldwide and in Canada.

You should study:

[Torge (2001): chapter-5.2 Satellite Observations]

[Héroux (2006): all-articles]

The second reference [Héroux (2006)], a collection of five articles, is an excellent source of information on the state of art of Canadian reference systems condensed in one volume. It needs a lot of attention, because it describes the historical evolution and the state of the art of the fundamental reference systems that most surveyors in Canada are using in their daily work. To get some more information on the latest developments you may consult the website of Natural Resources Canada: http://www.geod.nrcan.gc.ca/index_e.php

[Héroux (2006): A1: Space Geodetic Techniques and the Canadian Spatial Reference System]

[Héroux (2006): A2: The Evolution of NAD83 in Canada]

Sample Questions:

- Q8.1. What was the horizontal reference system prior to NAD83? How had it been established? What was its accuracy?
- Q8.2. How are modern 3D datums (also called TRS = Terrestrial reference systems) defined? On which type of coordinates do they rely? Why?
- Q8.3. How did the reference systems evolve in Canada: NAD27 – NAD83–NAD83(CSRS)?
- Q8.4. What is the difference between NAD83(CSRS) and the ITRS?
- Q8.5. Why is a 7-parameter transformation not sufficient to transform coordinates from NAD83(CSRS) to ITRF?
- Q8.6. What is a 14 parameter transformation? Identify the different parameters. How does this transformation work?
- Q8.7. What is the accuracy of NAD83(CSRS)?
- Q8.8. Which control networks are related to NAD83(CSRS) and how are they monitored?

Additional explanations:

Modern geodetic Terrestrial Reference Systems (TRS) like NAD83(CSRS) are all linked to the ITRS, the International Terrestrial Reference System. They have replaced world-wide old fashioned classical reference systems like NAD27. The NAD83(CSRS) has not appeared from nowhere. There has been an evolution from NAD27 to NAD83(CSRS) with a certain number of intermediate realizations which have been used for a certain time span and often abandoned after.

Note the distinction between system and frame. A reference system like ITRS is a theoretical concept, a definition. A reference frame like ITRF corresponds to its realisation. Loosely speaking, the ITRS (singular) is what one wants to have and the ITRFs (plural) are

what one has. For example, it is easy to state that the origin of the system has to coincide with the centre of mass of the Earth (part of the definition of the ITRS), but it is not as easy to make sure that it really does. This led to different realisations (ITRFs) over time, which are denoted by a year, the latest realization being ITRF2005.

The ITRFs represent realisations of the most fundamental terrestrial reference system, the ITRS. There has been, e.g., ITRF88, ITRF93, ITRF97, ITRF2000, ITRF2005. The list is not exhaustive. The important fact is that the realisations are converging. The differences are getting smaller since the accuracy is getting better over time.

Plate tectonic movements are affecting the coordinates. One can no longer consider the coordinates of a site as invariant over time. One has to associate a velocity vector to every point. The way the ITRS is dealing with this problem differs substantially from the way the NAD83(CSRS) does. The motion of the individual plates can be explained mainly by a rotation. The ITRS is a global reference frame. The resulting rotation of all the plates is minimized. This leads to a noticeable rotation of the North American plate with respect to ITRS (or to the different ITRFs, if you prefer). The resulting velocities of the points are typically on the order of 1-2 cm per year. The NAD83 tries to minimize this by letting the reference system itself rotate with the plate. Result: The coordinates with respect to NAD83 are no longer affected by plate tectonics. This is not completely true: there remain intra-plate distortions, not all the points are situated on the North American plate. In addition, the rotation vector accounting for the rotation of the North American plate is taken from the Nuvel-A model which is slightly overestimating it.

Since NAD83(CSRS) is rotating with the North American plate the transformation between an ITRF_{xx} and NAD83(CSRS) can no longer be a simple static 7-parameter transformation (3 translations, 3 rotations, 1 scale). It is a dynamical one, consisting of 14 parameters. For every one of the 7 parameters, its behaviour over time is considered to be linear. This means that the changes per unit of time (one may speak of the speed of the parameters) are added for every parameter, thus doubling their number. In order to apply such a transformation, a 2 step approach is used. First, the 7 parameters for the considered epoch are calculated, then a standard 7 parameter transformation is applied. [Héroux (2006): Table-1 page 156] lists the parameters. This table implicitly contains the values of the parameters related to a 14-parameter transformation between 2 ITRFs. The parameters are stabilizing over time, converging to values which are different from zero. There is a offset between ITRS and NAD83(CSRS) as can be deduced from the values of the translation vector at the metre level, and a misalignment of the axes as shown by the rotations in the order of magnitude (30 milliarcseconds [mas] correspond to about 1 metre at the Earth's surface). These discrepancies stem from the uncertainties in tying the NAD83. Since NAD83(CSRS) was held as close as possible to NAD83, they transferred to NAD83(CSRS) (historical heritage). The rotation of the North American Plate can be found in the change rate of the rotation parameters which is a bit less than 1 mas/yr.

[Héroux (2006): A3: A Gravimetric Geoid Model as a Vertical Datum in Canada]

[Héroux (2006): A4: Crustal motion and deformation monitoring]

[Héroux (2006): A5: Global Geodetic Observation System]

Sample Questions:

Q8.9. What is the reference surface used in most national vertical datums? Why?

Q8.10. How was the vertical datum established in Canada?

Q8.11. How will it be realized in the near future? Why?

Q8.12. What are the advantages of the new approach?

Q8.13. What is postglacial rebound and how does it affect coordinates?

Q8.14. What is plate tectonics and how does it affect coordinates?

Q8.15. How are plate tectonic movements monitored today?

Additional explanations:

The new NAD83(CSRS) is a 3D reference system, thus containing a vertical component, which corresponds to the ellipsoidal height. One may now be tempted to stay with the ellipsoid as reference surface for the height. However this surface does not have any physical meaning in the sense that water may flow uphill with respect to ellipsoidal heights. Therefore the geoid as equipotential surface of the earth gravity field remains the privileged reference surface.

The existing vertical datum in Canada (CGVD28) is based on spirit levelling, which was tied to tide gauges. The mean sea level observed at a tide gauge over several decades does not coincide with the geoid due to the local systematic behaviour of water temperature, currents, salinity etc. This is called the permanent sea surface topography (SST) which may reach 1-2 metres. Integrating more than one tide gauge as zero level references introduces a systematic distortion in the datum. If one uses only one reference station (Rimouski) one gets rid of this effect.

The new approach consists in abandoning spirit levelling for establishing and monitoring the vertical datum and to use instead the fundamental relationship between ellipsoidal height (h) and orthometric height (H): $H = h - N$, where N is the geoid undulation. Basically everything is tied to the ellipsoid, through NAD83(CSRS), and a high performance geoid model allows to convert to orthometric height. The challenge is to determine this model with an appropriate accuracy. This has become feasible nowadays. The advantages are explained in detail in [Héroux (2006): A3].

This leads to a situation where the tide gauge at Rimouski is no longer at a zero height, but at its corresponding SST value. The new datum will be shifted with respect to the old one. An alternative is to keep Rimouski at a zero height. In this case the resulting reference surface is no longer the geoid but the nearby equipotential surface which runs through the Rimouski zero level. This choice is also viable.

9. The Global Positioning System and other Global Navigation Satellite Systems

The use of Global Navigation Satellite Systems, mainly GPS, has become an important tool for surveying. The last part of this study guide addresses the issues of this important topic.

Introductions to this topic can be found in:

[Hofmann-Wellenhof et al. (2005): chapter-5 Part I: Global reference systems after GPS]

[Torge (2001): chapter-5.2.5: Global Positioning system (GPS)]

They do not give all the details needed, but summarize in a few pages the basic essentials. For a more detailed discussion, the main reference will be Hofmann-Wellenhof et al. (2007). This reference is a textbook of about 500 pages. In the following are comments on the importance of the different chapters with respect to the topics of this syllabus item:

Chapter-1: Introduction: One should always browse through the introduction.

Chapter-2: Reference systems: This chapter was already suggested when studying the ICRS and ITRS.

Chapter-3: Satellite orbits: You should read through chapter-3.2: Orbit description and study chapter-3.4: Orbit determination.

Chapter-4: Satellite signals: You may leave this chapter out. The information contained in [Torge (2001): chapter-5.2.5] is sufficient.

Sample Questions:

Q9.1. Explain the space segment, the control segment and the user segment?

Q9.2. What is the solution of an undisturbed orbit (2 body problem)?

Q9.3. What is the Keplerian representation of an orbit?

Q9.4. What are broadcast orbits and how are they obtained?

Q9.5. What are precise orbits and how are they obtained?

Additional explanations:

The space segment of a GNSS corresponds to the satellites emitting signals. The control segment is a network of ground based reference stations operated by the owner of the system. It is responsible for the operational aspects, like the uplink of orbit information to the satellites. The user segments are the receivers observing the signals and calculating positions. The satellites travel in orbits which can be represented by Keplerian elements. If they were only submitted to the attraction of a homogenous Earth, the resulting orbits would be perfect ellipses and the Keplerian elements would be constant over time. Since this is not the case, a Keplerian representation can still be used but the elements are changing over time. In the broadcast ephemeris, which are determined by the control segment and uploaded to satellites such an approach is used. This information is added to the signal and broadcast as a navigation message, which is extracted by the receivers. It allows for computing the position of the satellites as a function of time. The navigation message is updated every 2 hours typically. Besides the orbit information, satellite clock biases are also included. The precise ephemeris and precise satellite clock biases are computed by the IGS, the International GNSS service. There are different products and their accuracy is better than that of broadcast ones. See: <http://igs.cb.jpl.nasa.gov/> for more information.

Chapter-5: Observables: Chapter-5.1 presents the different observables and chapter-5.2, the data combinations. They are part of your study material. The remainder of this chapter presents all different error sources and effects which have to be taken into account and which affects the quality of the solution. The details are not necessary here but these effects, their order of magnitude and how to deal with them are.

Chapter-6: Mathematical models for positioning: This is an important chapter because it deals with the different type of solutions. The only subchapters that may be left out are: chapter-6.1.3 Point positioning with Doppler data, chapter-6.3.3 Correlations of the phase combinations.

Sample Questions:

Q9.6. What are the different types of observations which a GPS receiver tracks?

Q9.7. What are single frequency measurements and what are dual frequency measurements?

- Q9.8. What is the advantage of dual frequency pseudorange measurements versus single frequency?
- Q9.9. What is the advantage of dual frequency phase measurements versus single frequency?
- Q9.10. Enumerate the different error sources by decreasing importance and how to account for them. Give the order of magnitude [in metres] of the remaining effect after correction.
- Q9.11. What is the basic positioning mode for which GPS was designed?
- Q9.12. What is the basic principle of DGPS using code observations and what is the obtainable accuracy?
- Q9.13. Which type of observation is used for centimetre-level geodetic positioning?
- Q9.14. What is the basic principle of relative positioning using phase measurements and what is the obtainable accuracy?
- Q9.15. What does "RTK" mean?
- Q9.16. What are the characteristics of RTK?
- Q9.17. What is the difference between a relative static approach using postprocessing and RTK?
- Q9.18. What does "PPP" mean?
- Q9.19. What are the characteristics of PPP?

Additional explanations:

GPS is a navigation system and the basic task is to provide the facility to determine instantaneously a 3D-position, worldwide and 24 hours a day independent of weather conditions. This is achieved through the pseudorange measurements carried out by the receiver. The obtainable accuracy lies in the metre range. The integration of the phase measurements makes GPS interesting for geodetic applications. The obtainable accuracy is at the millimetre-level. Real Time Kinematic is a mode where a nearby reference receiver transmits its measurements to the rover. The measurements are dual frequency code and phase measurements. The roving receiver is computing a "ambiguity fixed solution". This means that a major task is to solve for the integer ambiguities as rapidly as possible, quasi-instantaneously. This is necessary in order to obtain a 1 centimetre-level accuracy or even better. The advantage compared to a post-processing approach is the fact that one is able to check the quality of the solution online. This allows a better efficiency. The disadvantage is that one needs a link for transmitting the data of the reference station to rover. RTK correspond to a relative positioning mode. The coordinates of the reference station have to be known. All positions determined by the roving receivers are tied to them. The precise point position, PPP, is an absolute mode. Only data from a single receiver are needed in order to compute the coordinates with a centimetre-resolution. This approach is of great help in regions where no geodetic reference stations are available. One needs, however, static dual frequency of some few hours to reach this level of accuracy. For more detailed information see the web-site of CSRS-PPP service of Natural resources Canada: http://www.geod.nrcan.gc.ca/products-produits/ppp_e.php

The different error sources which may affect the quality of the solution are mainly the ionospheric delay, the tropospheric delay, multipath, satellite orbit uncertainties, and antenna phase centre variations. How they affect the solution depends strongly on the type of solution. The ionospheric delay can be eliminated by using dual frequency observations. In

the presence of single frequency observations only, one has to rely on models, leading to residual errors. In the relative mode, this effect is strongly attenuated. Over small distances of some few kilometres it can normally be neglected. The tropospheric delay is often computed from a model (Hopfield, Saastamoinen etc.). The remaining error is typically less than a few centimetres. This is not critical for most applications relying on code observations only, since the accuracy is not as high. In a relative mode using phase observations, this error becomes more disturbing. For small distances without larger height difference between the stations, part of the error is absorbed by forming differences. For large baselines the meteorological conditions at the stations are decorrelated and the error remains. Techniques for estimating tropospheric biases are used in such a case. The most demanding mode is PPP, since one wants to obtain in the absolute mode an accuracy similar to the relative mode. No error cancels out.

Chapter-7: Data processing: This is beyond the needs of this syllabus item, so browse first through the entire chapter and then regard the following.

Chapter-7.1: Ambiguity resolution: Study carefully the chapter-7.2.1-General aspects of ambiguity resolution. Concentrate then on the search techniques because they are essential for RTK, which is the GPS solution most frequently used in surveying (see: chapter-7.2.3 Search techniques – A standard approach – Ambiguity resolution on the fly.) The details of other techniques may be skipped.

Chapter-7.2: Adjustment, filtering and quality measures: Chapter-7.3.3.Network adjustment discusses the principle of single-baseline solutions and multipoint solutions as well as making a comparison. It also explains the least-squares adjustment of baselines which is often used in commercial software packages. Chapter-7.3.4 Dilution of precision gives the theoretical background of the DOPs. The remaining of this chapter can be left out.

Chapter-8: Data transformation: The entire chapter.

Chapter-9: GPS: Provides some information on the historical development, the state of the art, and future modernizations of the still most important and most frequently used GNSS.

Chapter-10: Glonass: may be skipped.

Chapter-11: Galileo: may be skipped.

Chapter-12: More on GNSS: The following sections:

Chapter 12.1.1 Comparison of GPS, Galileo, and Glonass.

Chapter-12.3.1 Differential systems – Principles.

Chapter-12.4.1 Space based augmentation systems – Principles – WAAS.

For the Canadian augmentation systems you should return to:

[Héroux (2006): A1-chapter-4: Products, Tools and Services for Satellite Geodesy].

Chapter-13: Applications: Chapter 13.1.2 Position determination does not introduce new material. Its importance lies in the fact that the different modes are summarized from the point of view of the applications. The remainder of chapter-13 as well as the chapter-14: Conclusion are not essential but may broaden your general knowledge on GNSS beyond pure positioning applications.

Sample Questions:

Q9.20. What are initial phase ambiguities?

Q9.21. Why does one have to consider them?

Q9.22. Why are they important?

Q9.23. What is a PDOP and what it is use?

Additional explanations:

Initial phase ambiguities are related to the phase observations. The latter are ambiguous. This leads to a situation where additional unknowns have to be considered when using phase observations: the initial phase ambiguities, succinctly called ambiguities in the following. They are integers, at least when considering the double differences. The simplest way to estimate them consists in considering them first as additional real-valued unknowns in an adjustment. They are estimated together with the unknown coordinates. Afterwards one tests if the obtained real estimations can be assigned to integers. (The value has a small standard deviation and is very close to an integer). Once the integer values of the ambiguities are identified, the solution is recomputed. In this last step the ambiguities are no longer unknowns.

The problem with this approach is the fact that it takes a certain time, typically half an hour to one hour of data for solving the ambiguities in the case of a small baseline (several kilometres). This is not acceptable for a RTK application which needs to have the values instantaneously, or almost. To solve this problem, search algorithms are used. They rely on the fact that all ambiguities have to be integers. Different sets of possible candidates are defined and tested in order to find the correct one. This approach needs dual frequency measurements or, at least, works far better in this case. The gain of accuracy in the coordinate determination when passing from real-valued ambiguities (floating ambiguity solution) to integer ambiguities (fixed ambiguity solution) depends on the observation time span. For a static 24h solution the gain becomes marginal, whereas in an RTK-solution with 30 seconds worth of data, it is typically from metre-level down to centimetre-level. This explains why fast ambiguity resolution is the key feature in a RTK solution.