IDEAL REFERENCE FRAMES, CONCEPTS AND INTERRELATIONSHIPS

George Veis
National Technical University, Athens

ABSTRACT. Many problems of geodynamics depend on spatial relationship of points and their temporal variations. To solve these problems it is convenient, but not necessary, to use a reference frame. To use a reference frame, a scheme is needed by which the coordinates of any point expressed in this frame could be obtained. Coordinates are hardly ever measured directly. Instead, they are computed from measurements of other quantities within the framework of a theory that relates the measured quantities with the coordinates. Such a scheme will be called a reference system. Reference systems have been realized using simple theories and reducing the measurements with the best available, but not always complete, theories. This geometric (static) method has been used to a great extend to define astronomic reference systems (star catalogues) and geodetic reference systems (geodetic datums). With space techniques, a method can be used based on dynamic principles. A space object moving according to a certain theory (assumed to be known) defines in a time dependent way the representative points. A reference system of this type is the WGS 72.

Most of the problems connected to Astronomy, Geodesy, and Geodynamics are related with the spatial relationship of points, and the temporal variation of their position. In principle these problems can be formulated by expressing the position of each point by a variety of methods but it becomes much easier if the positions of all points are expressed with the same simple scheme. This is accomplished by using a reference frame in which the positions are expressed with coordinates. In what will follow we assume that euclidean space is a sufficiently good approximation and that any deviations from it can be taken care of, by appropriate small corrections.

In a one-dimensional space a point can be positioned if its distance from an arbitrarily selected point, used as origin, is given. Here we make the assumption that we have defined a scale for measuring distances. Moving in two and three dimensional spaces we need, respectively, to measure two or three distances from two or three selected points as origins.

E. M. Gaposchkin and B. Kołaczek (eds.), Reference Coordinate Systems for Earth Dynamics, 37–41.

https://doi.org/10.1017/S0252921100081082 Published online by Cambridge University Press

Defining positions with distances from selected points is not convenient because the formulation of our problems becomes rather complicated. Instead we choose to define positions by giving the distances from two lines or three planes, for a two or three dimensional space respectively. In order to make it even simpler we also choose the lines or planes to intersect at right angles. Three mutually perpendicular vectors of unit length define a triad. We will call such a triad a reference frame and the distances from the three unit vectors cartesian or rectangular coordinates. Cartesian coordinates are a commonly used form of coordinates but other forms, such as spherical, cylindrical, ellipsoid or geoidal are also used. Spherical coordinates are used in astronomy and ellipsoidal in geodesy (ellipsoidal coordinates also require to define that the parameters of the reference ellipsoid be defined). Assuming that these different forms of coordinates are referred to the same reference frame (or triad) their transformation is a trivial problem.

A three-dimensional reference frame can be arbitrarily located and oriented in space giving six degrees of freedom, three for the location and three for the orientation. There is no way by which this arbitrariness can be removed without additional information. However if the relative position and orientation of two reference frames are known (which means six parameters if they have the same scale) the conversion of coordinates from one frame to another is a simple matrix operation.

The freedom in choosing the origin and the orientation of the reference frame, as well as the form of the coordinates, could allow the selection of a coordinate system to optimize its use. As a rule this is accomplished if: a) observables are easily and simply expressed in this coordinate system, b) needed computations are easily and simply performed, and c) coordinates can readily be used for different applications. If a coordinate system is to be used only for the solution of a theoretical problem, the selection of an optimum system is not so difficult as a rule.

A coordinate system however is not used only for the solution of theoretical problems. There are many applications not only of geodesy and astronomy but also of geophysics, surveying, navigation, engineering etc. that require the use of a coordinate system. What this means is, to be able to find, in a certain reference frame, the coordinates of a physical point, and to be able to find the point in space with given coordinates. If the points are in motion, the coordinates should refer to a certain epoch.

The implementation of the selected coordinate system is a complicated operation. Except for very special cases, coordinates are not measured directly as distances from axes materialized on the ground. Instead they are computed from other measured quantities mathematically related with the coordinates.

https://doi.org/10.1017/S0252921100081082 Published online by Cambridge University Press

This mathematical relation is based on a certain theory and physical laws which may also include some parameters. In addition the measurements may have to be corrected before they are used. An example is the measurement of the distance between two points with a microwave measuring instrument. This distance is related with a simple formula with the rectangular coordinates of the two points. The measured distance however has to be corrected for refraction before being used in the formula. The use of a theory will not affect the arbitrariness of the origin and orientation of the reference frame, unless some outside positional information is introduced with the theory.

A scheme including the necessary measurements, theories, and computations, by which coordinates in a certain reference frame can be computed, defines a reference system. The internal consistancy of such a system will depend on the accuracy of the measurements and their corrections, on the completeness of the theories and the correctness of the constants, and on the precision of the computations. Coordinates derived from two different reference systems will not agree if the measurements, the theories and the computations are not consistent. In order to relate two reference systems it is necessary to find the relations between the two theories (and constants), the two sets of observations (and their corrections) and the two computations used for their definition.

The Reference Systems that have been implemented up to now in classical astronomy and geodesy (star catalogues and geodetic datums) were based on simple (euclidian) geometry. The measurements used were primarily angles with very few distances. Only during the last 20 years the number of distance measurements has been increased, with the introduction of new instruments such as EDM, laser etc.

In astronomy the coordinates of stars were determined using mainly Meridian instruments. These observations were reduced using a certain theory for precession and nutation to a common epoch and corrected for polar motion. The astronomic coordinates of the instruments were also unknown. With this method only a few thousand stars (fundamental stars) were determined. This fundamental network was densified by relative measurements on photographic plates and extended to some hundred thousands of stars. Fundamental stars are expected to have an uncertainty of ± 0".02 but the remaining stars may be more than 10 times worse.

In geodesy the coordinates of geodetic points (triangulation points) were determined by establishing a geodetic network of which angles and distances were measured. By adjusting this network (assuming it to have been normally projected on a reference ellipsoid) the coordinates of all the points of the network were computed. The origin and the orientation of the network was not arbitrarily set, but adopted from astronomic observations as a good approximation. So geodetic datums are, practically speaking, derived from free networks.

Within a geodetic datum the internal relative accuracy is anywhere between 10^{-6} and 10^{-5}, depending on the quality of the observations, the accuracy of the applied corrections (mainly refraction) and the completeness of the reductions (mainly from geoid to ellipsoid). Datums computed before 1950, without the use of computers, were adjusted with approximate methods and this is another source of error. A characteristic of a geodetic datum is that the positional accuracy depends on the location of the triangulation point in the network and that these errors propagate with distance. However, for most practical applications, such as cartography and engineering, the relative accuracy is sufficient.

If we exclude distortions in the networks, two geodetic reference systems (datums) should be related by translation and rotation (six parameters). The relative rotation should be rather small (of the order of 1" or less), since both datums were oriented with astronomic observations. On the other hand, since old networks have used standards of length of low accuracy for their scale, two datums may differ in scale by several parts per million. As a result, for the relation of two geodetic systems a scale factor also has to be introduced, bringing the number of parameters to seven.

The realization of these static reference systems is materialized by assigning nominal coordinates, using the above mentioned techniques, to a number of selected stars or triangulation points. It is the catalogue of these selected points (stars or triangulation points) and their coordinates that really defines the reference system. In order to calculate the coordinates of a new point in the same reference system, similar measurements are made between some of the selected points and the new one and coordinates are obtained using the same theory "de proche en proche".

With simultaneous or quasi-simultaneous observations to artificial satellites, which are used simply as moving beacons, three-dimensional geodetic networks (free nets) have been established, adjusted and calculated. The relative positions for stations separated by intercontinental distances have been determined. These nets have given positional accuracies of about ± 5 m.

Instead of using a static approach to determine nominal coordinates for selected points, we could use a moving object whose motion, expressed in some reference frame, is precisely known. Assuming we can time the observations to the needed accuracy and freeze the moving object, the problem becomes equivalent with the static geometric one. Artificial Satellites, (and the Moon) provide, at least in principle, such a moving object, since their moon is governed by the laws of celestial mechanics. The theory is there but in practice the problem is much more complicated to account for the great number of parameters whose values must be known (mainly in order to express the gravity field of the earth). These parameters can be calculated as auxiliary unknowns by successive approximations. Results obtained with this

https://doi.org/10.1017/S0252921100081082 Published online by Cambridge University Press

Dynamic Satellite Geodesy method are rather promising, especially with some special geodetic satellites like Lageos. This method has two advantages. First, it gives positions expressed in a dynamically defined inertial reference system (for periods where the secular perturbations may be ignored), and second it can give the best approximation to a geocentric system. Such a dynamic reference system becomes much more complicated and complex to define, since it includes a very complicated theory (and calculations) as well as a great number of parameters. Such a reference system is the WGS 72, especially when associated with the Transit System, which gives a very high degree of internal consistency. We should expect the GPS to be even better.

When comparing or combining dynamic systems we have to be careful, since the theories and constants used for the determination of the orbits, even the computer programs, may be different. To a first approximation, however, the dynamic systems are also related to each other by a translation, rotation and scale, and thus also with the static ones.

If we consider that the points on the earth's crust undergo secular, periodic and abrupt displacements in a pseudosystematic and random way, the implementation of a static geodetic reference system providing fixed nominal coordinates to a number of points becomes utopia. In this case we either use the minimum number of arbitrary points (6 parameters) or we use very many points assuming that the motion of the mean will be zero. The use of a dynamic reference system may be in principle a better solution, since it will eventially relate the positions to an almost inertial system. Such a system will be improved continuously be successive approximations. For better results and faster conversion, it is essential that an integrated large scale satellite system be used, and as such the GPS looks to be the most promising system.