Peter Teunissen

Faculteit Civiele Techniek en Geowetenschappen
Technische Universiteit Delft
Postbus 5048
2600 GA Delft
p.j.g.teunissen@citg.tudelft.nl

Vakantiecursus 2003

GPS and integer estimation

Het Global Positioning System (GPS) is een wereldwijd plaatsbepalingssysteem op basis van satellieten. De eerste plannen en ontwerpen voor het systeem dateren uit de vroege jaren zeventig in de vorige eeuw; reeds in februari van 1978 werd de eerste GPS-satelliet gelanceerd. De nominale constellatie omvat 24 satellieten, die elk in ongeveer 12 uur om de aarde cirkelen. Daardoor kunnen overal ter wereld, doorgaans minstens vier satellieten tegelijkertijd waargenomen worden. Peter Teunissen, hoogleraar mathematische geodesie en positiebepaling aan de Technische Universiteit Delft verzorgde in de zomer van 2003 colleges over Global Positioning Systems in het kader van de, door het CWI georganiseerde, vakantiecursus voor wiskundeleraren.

The Global Positioning System (GPS) nowadays is used for a whole variety of positioning and navigation applications. These applications range from navigating your private sailboat to determining the millimeter movements of the earth's tectonic plates. For the very high-accuracy applications of GPS one needs very precise range information. These precise ranges for positioning with GPS are obtained from carrier phase measurements. These measurements of range inherently contain unknown integer ambiguities to account for the mismatch of a whole number of wavelengths or cycles. This contribution describes the problem of GPS carrier phase ambiguity resolution, discusses some relevant elements of integer estimation theory and reviews some of the high precision positioning applications that come into reach when the integer carrier phase ambiguities can be resolved quickly and correctly.

Redundant measurements

As in other physical sciences, empirical data are used in geodesy to make inferences so as to describe the physical reality. Many such problems involve the determination of a number of unknown parameters which bear a linear or linearized relationship to the set of data. In order to be able to check for errors or to reduce for the effect these errors have on the final result, the collected data often exceed the minimum necessary for a unique solution (redundant data). As a consequence of measurement uncertainty, redundant data are usually inconsistent in the sense that each sufficient subset yields different results from another subset. Hence, redundancy generally leads to an inconsistent system of linear(ized) equations, say

$$y \cong Ax \tag{1}$$

where vector y contains the m observations, vector x the n unknown parameters. The $m \times n$ matrix A relates the observations to the parameters. Redundancy of the above system is defined as $m - \text{rank}A$, which in case of a full rank matrix simplifies to $m - n$, the difference between the number of observations and the number of unknown parameters.

The above inconsistent system is without additional criteria not uniquely solvable. The problem of solving an inconsistent system of equations has attracted the attention of leading scientists in the middle of the 18th century. Historically, the first methods of combining redundant measurements originate from studies in
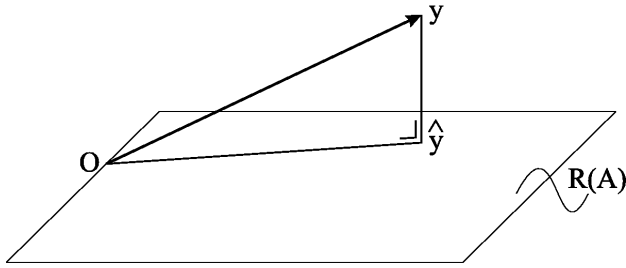
Figuur 1 Least-squares estimation implies a (n orthogonal) projection of the observation vector y onto the plane spanned by the columns of matrix A. Example with three observations and two unknown parameters.

geodesy and astronomy, namely from the problem of determining the size and shape of the earth, and the problem of finding a mathematical representation of the motions of the Moon. Since its discovery almost 200 years ago, the method of least-squares has been and still is too a large extent one of the most popular methods of solving an inconsistent system of equations. Although the method of least-squares may seem 'natural' for a student in modern times, its discovery evolved only slowly from earlier methods of combining redundant observations [1]

GPS positioning basically is determining the location of a (user) receiver with respect to satellites of which the locations (orbits) are known. This determination takes place by measuring distances, and from a geometric point of view three measurements would suffice to determine the three coordinates of the user (fortunately we know on which side of the satellite configuration the earth is located). The simplest example of (1) in case of GPS is therefore when distances are measured from an unknown GPS receiver position to more than three GPS satellites of which the positions are known. Since the distance from the unknown receiver position r to the known position of satellite s is a nonlinear function of the unknown position coordinates,

$$ l_r^s = \sqrt{(x^s - x_r)^2 + (y^s - y_r)^2 + (z^s - z_r)^2} \qquad (2) $$

the common approach is to approximate this relation by a linearized version, that is, developing the nonlinear relation in a Taylor series with zeroth and first order terms only, using good approximate values for the unknown parameters. As a result the (increments of the) observed distances are collected in vector y, the (increments of the) three unknown coordinates in vector x and the partial derivatives in matrix A. In reality the equations are far more complicated than (2) due to the fact that one also has to account for clock errors, atmospheric delays and orbital errors.

Least-squares
Around 1800 Legendre and Gauss at the same time (most likely independently), invented the method of least-squares for solving an inconsistent system of equations. The least-squares solution to (1) reads

$$ \hat{x} = (A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} y \qquad (3) $$

with Q_y^{-1} being the weight matrix. This solution is obtained by first adding an unknown error vector e to (1), giving the consistent but undetermined system $y = Ax + e$, and then minimizing the weighted norm of e, $\| e \|_{Q_y}$, as function of x. The least-squares estimator has various desirable properties. When the positive definite matrix Q_y is chosen as the variance-covariance matrix of the observations, the least-squares estimator has the smallest variance (best possible precision) of all linear unbiased estimators.

The geometric interpretation of what least-squares does to the observations is shown in figure 1. The inconsistency between observations on one hand and model (with unknown parameters) on the other is removed by orthogonal projection. Vector $\hat{y} = A\hat{x}$ eventually lies in the plane or linear manifold spanned by the columns of matrix A (indicated by $R(A)$). The orthogonal projection realizes shortest distance between the original observation values y and the adjusted ones $\hat{y}$; the observation values are modified as little as possible, though satisfying the assumed model afterwards. This follows from interpreting the least-squares estimation principle as the principle of least distance

$$ \min_x \| y - Ax \|_{Q_y}^2. \qquad (4) $$

The (squared and weighted) distance between y and $\hat{y} = A\hat{x}$ is minimized.

In order to evaluate the quality of the least-squares solution in a probabilistic sense, we need the probability density function (PDF) of $\hat{x}$. Since $\hat{x}$ of (3) is a linear function of y, the least-squares estimator has a Gaussian distribution whenever y is Gaussian distributed. The PDF of the unbiased least-squares estimator $\hat{x}$ can therefore be uniquely characterized by means of the variance-covariance matrix of $\hat{x}$. With Q_y being the variance-covariance matrix of the observations, application of the error propagation law to (3) gives the variance-covariance matrix of the estimated parameters as

$$ Q_{\hat{x}} = (A^T Q_y^{-1} A)^{-1} \qquad (5) $$

This matrix can be used to evaluate the precision of the parameter estimators, as for instance the position coordinates.

GPS carrier phase observable
GPS observations of distance or range are obtained by measuring signal travel-times (from satellite to receiver) and multiplying these by the speed of light. Two types of distance measurements are employed: pseudo range code and carrier phase. The code observation is based on the (binary) code the satellite modulates onto the signal carrier; the distance can be measured virtually unambiguously. For the carrier phase, the receiver measures the difference in phase between the carrier wave received from the satellite and the reference carrier wave it generated itself. The (physical) phase difference reads

$$ \psi_r^s = \phi_r - \phi^s. $$

With some simplifying assumptions, the phase of a carrier wave at some epoch t equals frequency f multiplied by time t: $\phi = ft$. The receiver compares the reference carrier at time of observation t_r with the carrier received from the satellite, which was generated a little earlier in order to be 'in time' at the receiver, namely at $t_r - \tau_r^s$, where τ_r^s is the signal travel time from satellite to re-

Figuur 2   A GPS receiver and antenna permanently installed for precisely monitoring motions of the earth's crust. Site Ranchita in California in the US. Photo taken from album at http://www.scign.org/

ceiver.

The above phase difference becomes

$$\psi_r^s = f\tau_r^s,$$

and when multiplied by wavelength $\lambda = \frac{c}{f}$, $\lambda\psi_r^s = c\tau_r^s = l_r^s$, the distance in meters is obtained; it equals the travel time premultiplied by the speed of light c, exactly as with the code observation.
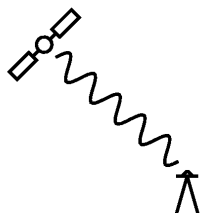
Figuur 3   Measurement of phase on the continuous carrier wave transmitted by the satellite. The satellite keeps on transmitting the carrier wave, cycle after identical cycle.

As a consequence of carrying out measurements on a (monotone) continuous carrier wave, the receiver can not distinguish one cycle from another. The satellite keeps on transmitting the carrier wave, in principle cycle after identical cycle, see figure 3.

At some epoch in time the receiver simply starts outputting the measured *fractional* difference in phase: $\text{frac}(\psi_r^s) \in [0, 1\rangle$ cycle. The full (physical) phase difference is then decomposed into

$$\psi_r^s = \underbrace{\text{int}(\psi_r^s)}_{N_r^s} + \underbrace{\text{frac}(\psi_r^s)}_{\phi_r^s}.$$

The observed (fractional) phase difference ϕ_r^s (times the wavelength) does thereby not equal the distance from satellite to receiver l_r^s, but equals this distance apart from an integer number of wavelengths

$$\lambda\phi_r^s = l_r^s - \lambda N_r^s.$$

As a consequence the vector x in (1) will, next to the unknown receiver coordinates, now also contain unknown *integer* cycle ambiguities N_r^s.
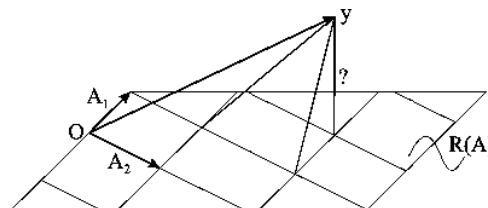
Figuur 4   Least-squares with integer parameters: possible solutions for the vector of observations form a grid in the column-space of matrix A (A_1 and A_2 are two columns of matrix A); the solution is no longer allowed to lie anywhere in the plane $R(A)$.

Integer least-squares

The least-squares solution (3) is obtained from solving (4), where x is allowed to vary over the whole n- dimensional space of *real* numbers. In case of GPS however, when use is made of the carrier phase observations, the vector of unknown parameters x consists of both real-valued and integer valued parameters (real-valued coordinates and integer-valued carrier phase ambiguities). We therefore need to modify the solution (3) so as to take the integerness of some of the parameters into account. To keep the discussion simple, it will be assumed here that *all* parameters in vector x are integer-valued. Due to the integerness of the parameters, orthogonal projection of y will now not do the job properly, see figure 4. Nevertheless one can start with 'ordinary' least-squares as a first step, see figure 5. The solution so obtained for the unknown parameters will be real-valued and is usually referred to as the 'float' solution.
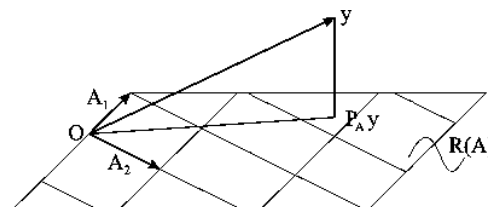
Figuur 5   Least-squares with integer parameters: the first step consists of 'ordinary' least-squares (orthogonal projection); the solution $\hat{x}$ for the parameters will consist of real-valued numbers.

To apply the least-squares principle (4), but now under the condition that the parameters in x are all integers, a second step has to be carried out. Since the first step projects orthogonally to the plane $R(A)$, the second step takes place *in* the plane. From the orthogonal decomposition

$$\|y - Ax\|_{Q_y}^2 = \|y - \hat{y}\|_{Q_y}^2 + \|\hat{y} - Ax\|_{Q_y}^2 \qquad (6)$$

it follows that the second step amounts to solving the minimization problem

$$\min_x(\hat{y} - Ax)^T Q_y^{-1}(\hat{y} - Ax) =$$

$$\min_x(\hat{x} - x)^T A^T Q_y^{-1} A(\hat{x} - x) = \min_x(\hat{x} - x)^T Q_{\hat{x}}^{-1}(\hat{x} - x) \qquad (7)$$

for x being integer, where in the last equation (5) has been used. This minimization can also be visualized in the parameter space, see figure 6, instead of in the observation space as in figures 1 and 4.
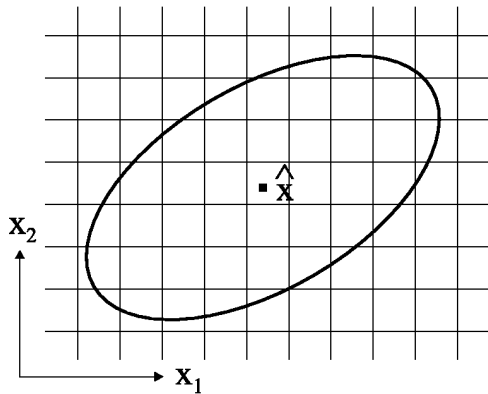
Figuur 6 Least-squares with integer parameters: in the second step the integer solution for x is sought that is *closest* to the real-valued solution $\hat{x}$ of the first step; 'closest' is to be measured in the metric of the variance-covariance matrix $Q_{\hat{x}}$; the quadratic form (7), set equal to a constant, is represented by the ellipse in this example with two ambiguities x_1 and x_2.

The integer least-squares principle has been applied very successfully to GPS ambiguity resolution. By the presence of the variance-covariance matrix $Q_{\hat{x}}$ in (7), the precision and correlation of the individual real-valued ambiguity estimates is properly and fully exploited. In contrast to the 'ordinary' least-squares solution (3), there does not exist an analytical solution to (7). In practice a search over possible integer solutions has to be carried out. The space of integer solutions is restricted by limiting the squared and weighted distance in (7) to a convenient value. As a result, the volume of the corresponding ellipse (or hyper-ellipsoid in higher dimensions) has to be searched through in order to find the integer least-squares solution of x.

When the ambiguities of the first step are of poor precision and at the same time highly correlated, the ellipse or ellipsoid gets very elongated and narrow. As a consequence the discrete search may get computationally inefficient. For computational efficiency the quadratic form (7) can be integer transformed, so that the resulting ellipsoid becomes more sphere-like and the transformed ambiguities become less correlated [2–3].

Alternative integer estimators

Instead of the integer least-squares estimator one can also think of alternative integer estimators. Starting from the 'float' solution, such an estimator $\check{x} = F(\hat{x})$ will consist of a mapping $F : R^n \mapsto Z^n$ from the n-dimensional space of real numbers to the n-dimensional space of integers. Due to the discrete nature of Z^n, the map F will not be one-to-one. This implies that different real-valued ambiguity vectors will be mapped to the same integer vector. One can therefore assign a subset $S_z \subset R^n$ to each integer vector $z \in Z^n$:

$$S_z = \{x \in R^n \mid z = F(x)\}, \quad z \in Z^n \tag{8}$$

The subset S_z contains all real-valued ambiguity vectors that will be mapped by F to the same integer vector $z \in Z^n$. This subset is referred to as the *pull-in-region* of z. It is the region in which all am-

biguity 'float' solutions are pulled to the same 'fixed' ambiguity vector z.

Since the pull-in-regions define the integer estimator completely, one can define classes of integer estimators by imposing various conditions on the pull-in-regions. One such class is given as follows [4].

An integer estimator is said to be *admissible* if

$$
\begin{aligned}
&(i) \bigcup_{z \in Z^n} S_z = R^n \\
&(ii) S_{z_1} \bigcap S_{z_2} = \{0\}, \quad \forall z_1, z_2 \in Z^n, z_1 \neq z_2 \\
&(iii) S_z = z + S_0, \quad \forall z \in Z^n
\end{aligned}
\tag{9}
$$

This definition is motivated as follows. Each one of the above three conditions describe a property of which it seems reasonable that it is possessed by an arbitrary integer ambiguity estimator. The first condition states that the pull-in-regions should not leave any gaps and the second that they should not overlap. The absence of gaps is needed in order to be able to map any 'float' solution $\hat{x} \in R^n$ to Z^n, while the absence of overlaps is needed to guarantee that the 'float' solution is mapped to just one integer vector. Note that the pull-in-regions are allowed to have common boundaries. This is permitted if we assume to have zero probability that $\hat{x}$ lies on one of the boundaries. This will be the case when the probability density function (PDF) of $\hat{x}$ is continuous.
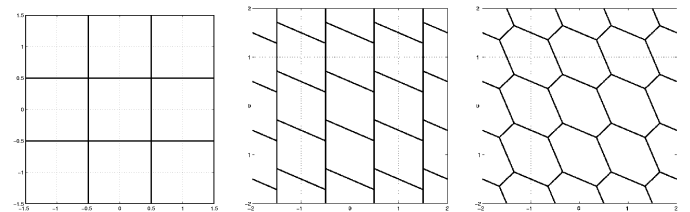
Figuur 7 Two-dimensional pull-in regions of rounding, bootstrapping and integer least-squares.

The third and last condition follows from the requirement that $F(x + z) = F(x) + z, \forall x \in R^n, z \in Z^n$. Also this condition is a reasonable one to ask for. It states that when the 'float' solution is perturbed by $z \in Z^n$, the corresponding integer solution is perturbed by the same amount. This property allows one to apply the *integer remove-restore* technique: $F(\hat{x} - z) + z = F(\hat{x})$. It therefore allows one to work with the fractional parts of the entries of $\hat{x}$, instead of with its complete entries.

There exist various admissible integer estimators. The simplest integer map is the one that corresponds to integer rounding. In this case the integer vector is obtained from a rounding of each of the entries of $\hat{x}$ to its nearest integer. Since componentwise rounding implies that each real-valued ambiguity estimate $\hat{x}_i, i = 1, \ldots, n$, is mapped to its nearest integer, the absolute value of the difference between the two is at most $\frac{1}{2}$. The subsets $S_{R,z}$ that belong to this integer estimator are therefore given as

$$S_{R,z} = \bigcap_{i=1}^{n} \left\{ \hat{x} \in R^n \mid \mid \hat{x}_i - z_i \mid \leq \frac{1}{2} \right\}, \ \forall z \in Z^n \tag{10}$$

The subset $S_{R,z}$ is an n-dimensional cube, with sides of length 1 and centered at the grid point z.

Another relatively simple integer ambiguity estimator is the integer bootstrapped estimator. This estimator can be seen as a generalization of the previous one. It still makes use of integer rounding, but it also takes some of the correlation between the ambiguities into account. The bootstrapped estimator results from a sequential conditional least-squares adjustment and is computed as follows. If n ambiguities are available, one starts with the first ambiguity $\hat{x}_1$, and rounds its value to the nearest integer. Having obtained the integer value of this first ambiguity, the real-valued estimates of all remaining ambiguities are then corrected on the basis of their correlation with the first ambiguity. Subsequently the second, but now corrected, real-valued ambiguity estimate is rounded to its nearest integer. Having obtained the integer value of the second ambiguity, the real-valued estimates of all remaining $n-2$ ambiguities are again corrected, but now on the basis of their correlation with the second ambiguity. This process of rounding and correcting is continued until all ambiguities are taken care of.

With c_i denoting the i-th canonical unit vector having a 1 as its i-th entry, the pull-in-regions $S_{B,z}$ that belong to the bootstrapped estimator can be shown to be given as

$$S_{B,z} = \bigcap_{i=1}^{n} \left\{ \hat{x} \in R^n \mid \mid c_i^T L^{-1}(\hat{x} - z) \mid \leq \frac{1}{2} \right\} , \ \forall z \in Z^n \quad (11)$$

with matrix L being the unit lower triangular matrix of the triangular decomposition of $Q_{\hat{x}}$. Note that these pull-in-regions reduce to the ones of (10) when L becomes diagonal. This is the case when the ambiguity variance-covariance matrix is diagonal. In that case the two integer estimators $\check{x}_R$ and $\check{x}_B$ are identical.

The third admissible estimator of which the pull-in-region will be given is the integer least-squares estimator. By again using the LDL^T-decomposition of $Q_{\hat{x}}$ the least-squares' pull-in-region reads

$$S_{LS,z} = \bigcap_{c_i \in L^{-1}(Z^n)} \left\{ \hat{x} \in R^n \mid \mid c_i^T D^{-1} L^{-1}(\hat{x} - z) \mid \leq \frac{1}{2} c_i^T D^{-1} c_i \right\} \quad (12)$$

Note that (12) and (11) become identical when the matrix entries of L^{-1} are all integer. This is the case when L is an admissible ambiguity transformation.

As an example of the three types of pull-in regions consider figure 7. These three types of pull-in region correspond with the 2-by-2 variance-covariance matrix

$$Q_{\hat{x}} = \begin{pmatrix} 0.0847 & -0.0364 \\ -0.0364 & 0.0865 \end{pmatrix}$$

The ambiguity success rate

The quality of the integer ambiguity estimator is particularly of interest in case of GPS. One therefore needs the probability mass function (S) of $\check{x}$. It can be obtained as follows. Using the concept of the pull-in-region, the integer estimator is defined as $\check{x} = z \ \Leftrightarrow \ \hat{x} \in S_z$. Hence, for the probability masses one has $P(\check{x} = z) = P(\hat{x} \in S_z)$. With the PDF of $\hat{x}$ given as $p_{\hat{x}}(x)$, the PMF of $\check{x}$ follows as
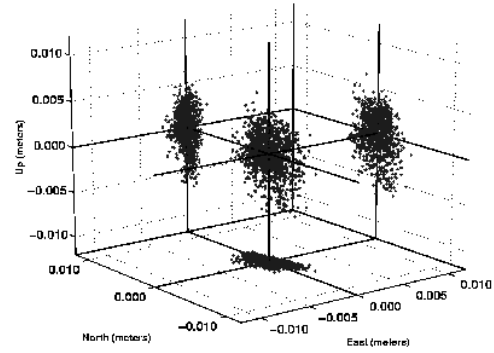

Figuur 8 Example of the repeatability of GPS positions after resolving the ambiguities by means of integer least-squares. The three-dimensional position is obtained from a single epoch of observations (so-called instantaneous positioning). The experiment has been carried out 1200 times, and shown are all 1200 ambiguities-fixed position solutions. The measurement noise in the carrier phase observation is at the few millimeter level and the consequent spread in position is clearly below 1 centimeter.

$$P(\check{x} = z) = \int_{S_z} p_{\hat{x}}(x)dx , \ \forall z \in Z^n \quad (13)$$

The *ambiguity success rate* is defined as the probability of correct integer estimation $P(\check{x} = x)$. Note that the PMF (13) as well as the success rate still depend on the type of pull-in-region and thus on the type of integer estimator chosen. Changing the geometry of the pull-in-region will change both the PMF and the ambiguity success rate. It is therefore of interest to know which integer estimator maximizes the ambiguity success rate. The answer is given by the following theorem [4]:

Theorem. *Let the PDF of $\hat{x}$ be elliptically contoured and the integer least-squares estimator be given as*

$$\check{x}_{ILS} = \arg\min_{z \in Z^n} \| \hat{x} - z \|_{Q_{\hat{x}}}^2$$

Then

$$P(\check{x}_{ILS} = x) \geq P(\check{x} = x) \quad (14)$$

for any admissible estimator $\check{x}$.

This theorem gives a probabilistic justification for using the *integer* least-squares estimator. It applies to GPS ambiguity resolution for which the PDF $p_{\hat{x}}(x)$ is often assumed to be a multivariate normal distribution. For GPS ambiguity resolution one is thus better off using the integer least-squares estimator than any other admissible integer estimator, such as integer rounding or integer bootstrapping.

Applications

Once the integer carrier phase cycle ambiguity has been resolved, the phase observation turns into a direct measurement of distance. These phase observations possess millimeter precision and consequently the user receiver position can be determined with a similar level of precision, see figure 8.

Already early in the history of GPS positioning, the application of surveying topography emerged. By taking the GPS receiver to

sites and features on the earth's surface, their locations can be determined and consequently be mapped. Today, GPS positioning is an important tool in producing and maintaining road-maps, town-plans and precise cadastral maps (and databases).

In the early days, precise positions got available only after considerable time spans (of one or several hours). By including the integer constraints on the ambiguities and developing efficient ways of solving the integer least-squares problem, high precision positions become available virtually immediately, see also figure 8. The ambiguities have been demonstrated to be resolved correctly using just one epoch (second) of observations, thus greatly improving surveying productivity. At present the position can be determined directly in the field, by Real-Time Kinematic (RTK) GPS, see figure 9.

Figuur 9 Real-Time Kinematic (RTK) GPS surveying: the surveyor directly 'digitizes' the points of interest in the field, by holding the antenna accurately in place for just a few seconds.

Similar equipment and algorithms can be used for high precision navigation of moving vehicles on land, vessels at sea and aircraft in the air. Challenging applications are vessel guidance through narrow straights with critical clearance and landing aircraft in conditions of poor visibility.

Precise GPS positioning anywhere on earth is of great benefit also to earth sciences. Tectonic plates may move by several centimeters a year with respect to each other. Such motions of the earth's crust can be monitored with GPS at the required level of precision. This is of particular interest in areas with considerable seismic activity. For instance in California in the United States, with emphasis on the greater Los Angeles metropolitan region, an array of GPS receivers has been installed — under the name of Southern California Integrated GPS Network (SCIGN) — to study geodynamical phenomena. Over 200 locations are covered and GPS receivers are in operation 24 hours a day, 7 days a week. Figure 2 shows an example of a station of the SCIGN.

Conclusion

In this article the problem of the integer cycle ambiguity of the GPS carrier phase observations for ranging has been addressed. The ambiguities are resolved using the integer least-squares principle thus allowing very precise and fast GPS positioning. Since various details were skipped in the above presentation, the interested reader is referred to the many textbooks available on GPS positioning [5–9].   ⬳

Acknowledgment

References

1   Teunissen, P.J.G. (2000): A brief account on the early history of adjusting geodetic and astronomical observations. *De Hollandse Cirkel*, 2(1/2), pp. 12–17.

2   Teunissen, P.J.G. (1993): *Least-squares estimation of the integer GPS ambiguities.* Invited lecture. Section IV Theory and Methodology, IAG General Meeting. Beijing, China. August. Also in LGR-series No. 6, Delft.

3   Lenstra, H.W. (1981): Integer programming with a fixed number of variables. University of Amsterdam, Dept. of Mathematics, Report 81-3.

4   Teunissen, P.J.G. (1999): An optimality property of the integer least-squares estimator. *Journal of Geodesy*, 73:587–593.

5   Leick, A. (1995): *GPS Satellite Surveying.* 2nd ed. John Wiley, New York.

6   Strang, G. and K. Borre (1997). *Linear algebra, geodesy, and GPS.* Wellesley-Cambridge Press.

7   Teunissen, P.J.G. and A. Kleusberg (1998): *GPS for Geodesy.* 2nd enlarged edition, Springer Verlag.

8   Hofmann-Wellenhof, B., H. Lichtenegger, J. Collins (2001): *Global Positioning System: Theory and Practice.* 5th ed. Springer Verlag.

9   Misra, P. and P. Enge (2001): *Global Positioning System: sSignals, Measurements and Performance*, Ganga-Jamuna Press.