23. FUTURE PROSPECTS

23.1. The Merits of a Scanning Astrometric Mission

A concept for a future space astrometry mission based on an extrapolation of the principles adopted by Hipparcos has recently been formulated (Lindegren & Perryman 1996), and has been recommended for further study within ESA's long-term scientific programme. A small interferometer, with a baseline of about 2.5 m, and equipped with CCD detectors, should be capable of measuring the astrometric parameters of every object down to 15 mag or fainter (some 50 million or more), with an accuracy of some 10 microarcsec at 15 mag or some 2–3 microarcsec at about 10 mag. Instrumental optimisation could lead to the measurement of a significant proportion of objects down to 20 mag, with an improved accuracy of about 2–3 microarcsec at 15 mag.

Scientifically, the attractions of such a mission are very broad. Distances of objects throughout the Galaxy would be measured (with a 10 per cent accuracy at distances of the galactic centre), and space velocities would be acquired with an accuracy of around 1 km/sec even at 20 kpc. In addition to the detailed motions and properties of individual stars and stellar groups throughout the Galaxy, metric terms would be directly measurable (with a precision in the PPN parameter γ of the order of 1 part in 10^6), and planetary companions of a few Jupiter masses would be observable out to a few hundred parsecs. The appeal of such large-scale, high-accuracy astrometric measurements, and the technological prospects of conducting them within the next one or two decades, provokes the question of the extent to which the Hipparcos experiences can be carried forward to space astrometry in the future.

In general terms, the measurements conducted by such a continuously scanning satellite can be shown to be almost optimally efficient, with each photon acquired during a scan contributing to the precision of the resulting astrometric parameters. Although every object down to the limiting magnitude of the Hipparcos instrument could not be observed, and significant inefficiencies resulted from the sequential mode of operation of the detector, a future mission would most probably be able to observe the objects passing across the field of view simultaneously, with every star above the corresponding signal-to-noise threshold ultimately contained within the final catalogue. The small conceptual appeal of being able to devote more observing time to a particular object of high scientific interest by means of a payload which can 'stop and stare' at a given region of sky appears to be completely outweighed by the very high accuracy that is achievable in any case on such a large number of objects. One of the scientific targets of a future astrometric mission, indeed, will be the large-scale dynamical motions of stars, associations, clusters, and galactic spiral arms, that can only be tackled by access to the distances and motions of large samples of stars.

The scanning satellite concept also leads directly to the construction of a global reference frame into which each object is placed in an absolute sense. One of the great merits of Hipparcos is that it generates a reference frame within which parallaxes and proper motions are rigidly defined. A future astrometric mission, reaching to 15 mag or fainter, would circumvent one of the problems faced by the Hipparcos mission in linking the resulting reference frame to an inertial system, through the direct observation of extragalactic objects. The wide separation of two separate viewing directions would be preserved, since it leads to the determination of absolute trigonometric parallaxes, and thereby circumvents the problem which has plagued ground-based parallax determinations, namely the transformation of relative parallaxes to absolute distances. The successful implementation of these concepts has been convincingly demonstrated by the Hipparcos mission.

The continuously scanning satellite approach leads to two further important attributes of the resulting data. The first of these is the wealth of photometric information that is acquired by an instrument which continuously scans the celestial sphere in a reasonably uniform manner. The calibrated photometric results from Hipparcos surpass in quantity, quality and uniformity the corresponding ground-based results acquired over many decades. The application of the photometric data to the study of stellar variability, and the direct astrophysical value of high-accuracy magnitudes and colours, is already evident from the Hipparcos results.

The other feature of the global astrometric data which is such an important pointer for the future is the capability of determining the astrometric parameters of double and multiple systems. Although posing a considerable and continual challenge to the instrument design, the data acquisition, the data analysis, and the final catalogue production, the wealth of information contained in the Hipparcos results provides an insight into the importance of double and multiple systems within the context of a future catalogue of 50 million objects with microarcsec accuracy. At this level, the complexity of the systems already evident in the Hipparcos Double and Multiple Systems Annex will be compounded, and a powerful observational system which samples the stellar images and their photocentric motions semi-continuously will reveal much about star formation, the initial and subsequent mass functions, n-body interactions, and many other details of stellar structure and evolution. The scanning satellite concept is important in that a semi-continuous sampling of the double or multiple star geometry is possible, and is again directly placed within the overall reference frame of the global catalogue.

Finally, in all of these considerations, it should be stressed that both for Hipparcos, and for an advanced mission based on similar concepts, the number of distinct astrometric observations per star is very much larger than the number of variables characterising the stellar motion. In this sense the overall instrument is self-calibrating, and the resulting astrometric parameters are determined along with estimates of their standard errors and correlations. This provides the possibilities of an accurate and unambiguous calibration of the instrumental geometry, and standard errors of the astrometric parameters which are expected to be a realistic indication of the true errors. For the rigorous scientific exploitation of the astrometric data such confidence in the error estimates is crucial.

23.2. The Space Astrometry Problem Revisited

Looking back on the many years of planning and execution of the data reductions for Hipparcos, it is easy to find instances where a somewhat different approach to the analysis of the satellite data might have been advantageous. In several cases more direct and accurate methods would certainly have been adopted, given the availability of present-day computing facilities. This experience must be taken into account in any future space astrometry project. In this context it is perhaps of some interest to reconsider the space astrometry problem in very general terms.

Stellar astrometric observations from space aim at the determination of a finite set of parameters describing the barycentric motion of each star. These parameters may be summarised in a vector of unknowns, **a**. The observations consist of instantaneous measurements of the centroids of stellar images on the detector, expressed in detector coordinates, such as slits or pixels, denoted G and H. Each observation, k, is therefore characterised by the time t_k, a measurement vector $\mathbf{g}_k = (G_k, H_k)'$, and associated statistics.

Very generally, the space astrometry problem can be formulated as the minimisation problem:

$$\min_{\mathbf{a},\mathbf{n}} \left\| \mathbf{g}^{\text{obs}} - \mathbf{g}^{\text{calc}}(\mathbf{a}, \mathbf{n}) \right\|_M \qquad [23.1]$$

where $\mathbf{g}^{\text{obs}}$ is the vector of all measurements and $\mathbf{g}^{\text{calc}}$ the vector of detector coordinates calculated from the astrometric parameters. The norm is calculated in a metric M defined by the statistics of the data, which in the general, non-linear case need not be Gaussian. In this equation **n** is a vector of parameters which are of no direct interest to the astronomical problem at hand, but which are nevertheless required for a physically realistic modelling of the data and therefore have to be estimated along with the astrometric parameters. The practical formulation of the problem is mainly related to the specification of the 'nuisance parameters' **n**, which naturally depends on the type of mission considered. Subsequently a continuously scanning satellite, such as Hipparcos or GAIA, will be assumed.

The modelling of the observables **g** is done by three successive transformations: (1) from astrometric parameters to the celestial directions of the star at the instants of observation, using an astrometric model; (2) from celestial to instrumental frame directions using an attitude model; and (3) from instrumental directions to detector coordinates using an instrument model.

Astrometric Model

In the simplest case, as applied to most of the Hipparcos stars, the modelling of the satellitocentric direction to star i at time t_k depends on just five parameters intrinsic to the star, the so-called five astrometric parameters: α_i, δ_i, π_i, $\mu_{\alpha*i}$, and $\mu_{\delta i}$, referred to a given epoch and being defined with respect to the solar system barycentre. More generally, the stellar astrometric parameters could include, for instance, the orbital parameters of binary stars.

One of the important insights gained from the Hipparcos mission concerned the impact of stellar duplicity on high-accuracy astrometry, and the sometimes astonishing complications brought about by this well-known, but easily forgotten, phenomenon of the common stars. For the present discussion it is assumed that the vector $\mathbf{a}_i$ includes whatever parameters are needed to represent the motion to the required accuracy.

Calculation of the observable (proper) direction of the star at an arbitrary instant requires a set of auxiliary data $\mathbf{e}$ which are regarded as known, i.e. not subject to improvement from the observations. Most importantly this set includes the barycentric ephemeris of the satellite. The transformation to proper direction, largely covered in Chapter 12, is written symbolically:

$$\mathbf{u}_{ik} = \mathbf{u}(\mathbf{a}_i | t_k, \mathbf{e}) \qquad [23.2]$$

Note that the auxiliary data $\mathbf{e}$ are not part of the nuisance parameters. Hence they are placed, with time, to the right of the bar in Equation 23.2, indicating that they are 'given'.

Continuous Attitude Model

The attitude specifies the instantaneous orientation of the instrument axes in the same celestial reference frame as used for the astrometric parameters. The instrument axes are defined by means of the celestial projections of certain reference points on the detector. Clearly the attitude angles enter as unknowns in the general problem. There are two rather different ways in which they can be handled: as discrete or continuous variables.

In the discrete case there is an independent set of (three) attitude angles for every instant t_k. Given that each observation provides two coordinates, a prerequisite for this model is that at least two observations are made at each instant. In principle the attitude parameters can be eliminated 'on the spot', leaving a set of equations representing the instantaneous relative measurements, e.g. in the form of the angular separations of stellar images expressed in detector coordinates. A pointing space observatory is the most obvious example where the discrete attitude model applies.

The continuous attitude model is only applicable to a scanning satellite. It describes the attitude in the form of continuous functions of time, using a reduced set of parameters $\mathbf{c}$. These could be, for instance, the spline coefficients for the three attitude angles with respect to an analytical reference model. Provided that the actual attitude motion is sufficiently smooth, this model has a significant advantage over the discrete model, owing to the smaller number of parameters, or degrees of freedom, that have to be estimated. The optimum dimension of $\mathbf{c}$ is a compromise between the measurement-induced error and the modelling error. Considering the relatively short dynamical memory of the satellite it is reasonable to use an independent set of attitude parameters, $\mathbf{c}_j$, for each time interval $\mathcal{T}_j$ of several hours.

Angular coordinates on the sky, measured with respect to the projected axes of the instrument, are called 'field angles' and denoted (η, ζ). Given the proper direction to a star and the attitude parameters, the field angles of the object at the time $t_k \in \mathcal{T}_j$ can be written:

$$\mathbf{f}_{ik} = \mathbf{f}(\mathbf{u}_{ik}, \mathbf{c}_j | t_k) \qquad [23.3]$$

where $\mathbf{f}_{ik}$ is the vector of field angles.

Instrument Model

The final transformation is from field angles $\mathbf{f}$ to detector coordinates $\mathbf{g}$, i.e. G, H; this is the field-to-detector transformation:

$$\mathbf{g}_{ik} = \mathbf{g}(\mathbf{f}_{ik}, \mathbf{d}_j | t_k) \tag{23.4}$$

It depends on the instrument parameter vector $\mathbf{d}_j$ describing the scale, detector orientation, optical and mechanical distortions, etc. The set of parameters $\mathbf{d}_j$ is also assumed to be defined on the interval $\mathcal{T}_j$ but may contain a subset which is constant over much longer times, e.g. for the medium- or small-scale distortion.

The practical formulation of the field-to-detector transformation is rather dependent on the hardware of the optics and detector system. The scan field mosaic of the Hipparcos main grid naturally led to a model with two components (Chapter 10): one fixed, medium-scale distortion pattern representing the physical deformations of the scan fields, and a variable, large-scale polynomial component capable of absorbing all kinds of optical distortion, chromaticity, etc.

Model Synthesis

The overall transformation can be written:

$$\begin{aligned} \mathbf{g}_{ik} &= \mathbf{g}(\mathbf{f}(\mathbf{u}(\mathbf{a}_i | t_k, \mathbf{e}), \mathbf{c}_j | t_k), \mathbf{d}_j | t_k) \\ &\equiv \mathbf{h}(\mathbf{a}_i, \mathbf{c}_j, \mathbf{d}_j | t_k, \mathbf{e}) \end{aligned} \tag{23.5}$$

The general minimisation problem thus becomes:

$$\min_{\mathbf{a},\mathbf{c},\mathbf{d}} \| \mathbf{g}^{\text{obs}} - \mathbf{h}(\mathbf{a}, \mathbf{c}, \mathbf{d} | t, \mathbf{e}) \|_M \tag{23.6}$$

where the indices i, j and k have been dropped since the norm is to be computed over the whole range of the indices.

The observations are invariant with respect to a uniform, rigid rotation $\mathcal{S}$ of the celestial coordinate system. The rigorous formulation must therefore be such that:

$$\mathbf{h}(\mathcal{S}\mathbf{a}, \mathcal{S}\mathbf{c}, \mathcal{S}\mathbf{d} | t, \mathcal{S}\mathbf{e}) = \mathbf{h}(\mathbf{a}, \mathbf{c}, \mathbf{d} | t, \mathbf{e}) \tag{23.7}$$

for any such transformation $\mathcal{S}$ of the parameter vectors. Only the instrument description, which does not involve celestial coordinates, can be assumed to be independent of this transformation: $\mathcal{S}\mathbf{d} = \mathbf{d}$.

In the Hipparcos reductions this invariance was most strikingly demonstrated by the different choices of celestial reference frame—ecliptic versus equatorial—by the two consortia. On a more subtle scale it was manifested in the small global orientation and spin differences found after transformation to equatorial coordinates (Chapter 16). The discussion of the rank-deficiency problem in Chapter 11 showed that this invariance was not an obvious property of the data reduction problem in its usual formulation based on the so-called ‘three-step’ method (Chapter 4). One conclusion for the future is that the invariance with respect to uniform rotations should be carefully considered and built into the equations from the very start, resulting in minimally constrained solutions for the reference frame.

Method of Solution: Direct Approach

In Equations 23.3 and 23.4 the unknowns $\mathbf{c}_j$ and $\mathbf{d}_j$ were both taken to be defined over the interval $\mathcal{T}_j$ of several hours. While the two parameter sets represent very different physical models, they are thus equivalent from a data processing point of view and may be considered together as parts of the 'local' vector of nuisance parameters, $\mathbf{n}_j$. However, as was remarked before, $\mathbf{d}_j$ may contain a part which is common to a longer interval, or even the whole mission. These 'global' nuisance parameters may be separated out as the vector $\boldsymbol{\gamma}$, and $\mathbf{n}$ may be redefined to contain only the 'local' nuisance parameters. Equation 23.5 is then recast as:

$$\mathbf{g}_{ik} = \mathbf{h}(\mathbf{a}_i, \mathbf{n}_j, \boldsymbol{\gamma} | t_k, \mathbf{e}) \qquad [23.8]$$

and the general minimisation problem becomes:

$$\min_{\mathbf{a},\mathbf{n},\boldsymbol{\gamma}} \| \mathbf{g}^{\text{obs}} - \mathbf{h}(\mathbf{a}, \mathbf{n}, \boldsymbol{\gamma} | t, \mathbf{e}) \|_M \qquad [23.9]$$

It can be noted that this form, after linearisation, has the same general structure as the least-squares problems encountered in the great-circle reductions (Equation 9.5) and the sphere solution (Equation 11.23), and could in principle be handled by the same direct method as was used in those problems. That is, after sorting the data either chronologically (by the j index) or systematically (by the i index), the corresponding unknowns ($\mathbf{n}_j$ or $\mathbf{a}_i$) may be eliminated, resulting in a rather dense system of normal equations for the remaining parameters. For Hipparcos the dimensions of $\mathbf{a}$ and $\mathbf{n}$ were, respectively, about 370 000 (the astrometric parameters for the primary reference stars, see Table 11.1) and ~ 2 000 000 (the number of spline coefficients and free instrument parameters in the FAST great-circle reductions). If the $\mathbf{n}_j$ are successively eliminated, the direct solution of the remaining system requires of the order of $n^3/3 \sim 10^{16}$ floating-point operations, and the administration of $n^2/2 \sim 6 \times 10^{10}$ double-precision reals ($\simeq$ 500 Gigabyte): a non-trivial task even for supercomputers and parallel processing. It was such considerations that lead to the idea of the 'three-step' decomposition proposed in 1976. However, the practicality of that method was gained at the expense of approximations which should now be avoided.

Global Iterative Solution

Apart from the 'three-step' method, the only alternative to the direct solution proposed to this date seems to be an iterative solution. The basic idea dates back at least to 1977, when Prof. Pierre Lacroute advocated the use of intermittent guiding of the satellite and the use of 'dynamical smoothing' in the quiet intervals. In his introductory talk at the 'Colloquium on European Satellite Astrometry', held in Padova in June 1978, the idea was formulated the following way:

> *… it is possible to represent the attitude motion during the periods of free motion by using the coordinates of the stars and all their transit times. With the help of mechanical laws the computed attitudes should be very accurate and by using them along with the transit times we could obtain better evaluations of the coordinates.*

To iterate this procedure is an obvious possibility. The resulting method, which may be referred to as the 'global iterative solution', was subsequently proposed and studied by a group at the Istituto di Topografia, Fotogrammetria e Geofisica, Milano (Betti, Sansò *et al.*, in Perryman *et al.* 1989 Volume III, Chapter 28) and further discussed by Lattanzi *et al.* (1990). In the present framework it can be described as follows.

Let $\mathbf{i}_i$ be the vector of all observations $\mathbf{g}_{ik}^{\text{obs}}$ of a particular star i, and similarly let $\mathbf{j}_j$ be the vector of all observations made in the time interval $\mathcal{T}_j$. With I and J denoting the number of stars and time intervals, respectively, $(\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_I)$ and $(\mathbf{j}_1, \mathbf{j}_2, \dots, \mathbf{j}_J)$ are thus different partitions of the total observation vector $\mathbf{g}^{\text{obs}}$. In practice they could be obtained by sorting the observations according to star index or time, respectively, although this may not be necessary depending on the administration of the equations.

If, for a moment, the astrometric parameters $\mathbf{a}$ and the global parameters $\boldsymbol{\gamma}$ are regarded as known, or rather as 'given', it is a simple matter to solve, for each time interval j, the minimisation problem:

$$\min_{\mathbf{n}_j} \|\mathbf{g}_{ik} - \mathbf{h}(\mathbf{a}_i, \mathbf{n}_j, \boldsymbol{\gamma} | t_k, \mathbf{e})\|_M \qquad [23.10]$$

involving only the observations $\mathbf{j}_j$ and resulting in a linearised system of equations with $\dim(\mathbf{n}_j)$ unknowns, i.e. typically a few hundred. The solution to this problem may be formally written as the function $\hat{\mathbf{n}}_j(\mathbf{j}_j|\mathbf{e}, \mathbf{a}, \boldsymbol{\gamma})$. This problem is somewhat analogous to the attitude reconstruction problem discussed in Chapter 7.

Conversely, by regarding $\boldsymbol{\gamma}$ and the local parameters $\mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_J)$ as given, the astrometric parameters of each star are obtained by solving the problem:

$$\min_{\mathbf{a}_i} \|\mathbf{g}_{ik} - \mathbf{h}(\mathbf{a}_i, \mathbf{n}_j, \boldsymbol{\gamma} | t_k, \mathbf{e})\|_M \qquad [23.11]$$

involving only the observations $\mathbf{i}_i$ and resulting in a linearised system of equations with $\dim(\mathbf{a}_i)$ unknowns (typically 5). The solution to this problem, analogous to the astrometric parameter determination discussed in Chapter 11, may be written as the function $\hat{\mathbf{a}}_i(\mathbf{i}_i|\mathbf{e}, \mathbf{n}, \boldsymbol{\gamma})$.

Finally, if both the local and astrometric parameters are regarded as given, the global parameters may be obtained as the solution to the problem:

$$\min_{\boldsymbol{\gamma}} \|\mathbf{g}_{ik} - \mathbf{h}(\mathbf{a}_i, \mathbf{n}_j, \boldsymbol{\gamma} | t_k, \mathbf{e})\|_M \qquad [23.12]$$

and denoted $\hat{\boldsymbol{\gamma}}(\mathbf{g}^{\text{obs}}|\mathbf{e}, \mathbf{a}, \mathbf{n})$. This problem involves all the observations, but still results in a relatively small system of equations with $\dim(\boldsymbol{\gamma})$ unknowns.

The global iterative solution is a straightforward sequential application of the above (partial) solutions. The optimal sequence of the three estimators $\hat{\mathbf{n}}_j$, $\hat{\mathbf{a}}_i$, $\hat{\boldsymbol{\gamma}}$ is not obvious, but the following order seems intuitively natural:

$$\begin{aligned}
\mathbf{a}^{(0)} &= \text{initial catalogue} \\
\boldsymbol{\gamma}^{(0)} &= \mathbf{0} \\
\left.\begin{aligned}
\mathbf{n}_j^{(m)} &= \hat{\mathbf{n}}_j(\mathbf{j}_j|\mathbf{e}, \mathbf{a}^{(m-1)}, \boldsymbol{\gamma}^{(m-1)}), \quad j = 1, 2, \dots, J \\
\mathbf{a}_i^{(m)} &= \hat{\mathbf{a}}_i(\mathbf{i}_i|\mathbf{e}, \mathbf{n}^{(m)}, \boldsymbol{\gamma}^{(m-1)}), \quad i = 1, 2, \dots, I \\
\boldsymbol{\gamma}^{(m)} &= \hat{\boldsymbol{\gamma}}(\mathbf{g}^{\text{obs}}|\mathbf{e}, \mathbf{a}^{(m)}, \mathbf{n}^{(m)})
\end{aligned}\right\} & \quad m = 1, 2, \dots
\end{aligned} \qquad [23.13]$$

If the iterations converge, the end result is evidently equivalent to a direct solution of the global minimisation problem, Equation 23.9.

Concerning the convergence properties, it can be noted that the linearised form of the procedure, written in the form of normal equations, is equivalent to the Gauss-Seidel iteration method for the solution of the linear system of equations. It is well known that this method converges for any symmetric and positive definite matrix. Due to the (theoretical) rank deficiency of the problem, this condition is in principle not satisfied.

However, it can be argued that the particular degeneracy due to the undefined reference frame is of no practical consequence for the iterative solution, since each of the partial minimisation problems (Equations 23.10–23.12) do not suffer from this degeneracy. The tentative conclusion is therefore that the method does converge, namely to the particular solution closest, in some sense, to the initial estimate $\mathbf{a}^{(0)}$, $\mathbf{n}^{(0)}$, $\boldsymbol{\gamma}^{(0)}$.

Intuitively, the global iterative solution is expected to converge as a consequence of the geometrical structure of the problem, namely that in a given interval $\mathcal{T}_j$ many different stars contribute to the determination of $\mathbf{n}_j$, while, conversely, many different intervals contribute to the determination of a given star. Thus, an initial error in the coordinate of one star gives only a much smaller error in the attitude parameters of the affected intervals, and these errors in turn are diffused, in the next iteration, to a large number of stars, and, in rather few iterations, to the whole set of stars. It can be noted that this diffusion is strengthened by the superposition of the two fields of view in Hipparcos, by the incommensurability of the basic angle to 360°, and by the diversity of scan directions across any point on the sky; i.e. by the very properties that make the Hipparcos reference frame internally 'stiff'. It is a very likely hypothesis that the convergence properties are closely linked with the stiffness of the resulting reference frame: a well-designed space astrometry project should ensure good convergence of the global iterations.

A simplified version of the global iterated solution, using 2000 stars, was in fact implemented by Sansò *et al.* (in Perryman *et al.* 1989 Volume III, Chapter 28), and was found to converge in only two iterations. The block iteration method used for the FAST sphere solution (Equations 11.27–11.29) follows the same general numerical principle (although the detailed equations are different), demonstrating its feasibility for a similar problem with ~ 370 000 unknowns.

The global iterative solution thus appears to be a both practically feasible and intuitively natural method for solving the general space astrometry problem. One possible disadvantage of the method is that it seems to be difficult to estimate reliably the uncertainties of the astrometric parameters. The curvature matrix associated with the restricted problem in Equation 23.11 gives only a lower bound to the covariance matrix of $\mathbf{a}$, by neglecting the uncertainties in $\mathbf{n}$ and $\boldsymbol{\gamma}$. This aspect of the global iterative solution requires additional study.

23.3. An Attempted Global Iterative Solution

The 'three-step method' on which both the FAST and NDAC data reductions were based introduced the star abscissae as an intermediate quantity in order to allow a direct, but approximate, solution of the general space astrometry problem. The nature of this approximation was discussed in Sections 11.3 and 11.7. A particular concern was that it might introduce a distortion of the resulting system of positions and proper motions. As shown in the previous section, the approximation could be eliminated by adopting instead the 'global iterative solution'. It was also remarked that some of the key procedures necessary for the global iteration were in fact very similar to procedures already implemented in the data reductions: for Equation 23.10, the attitude reconstruction or, more precisely, the attitude smoothing included in the great-circle reductions; for Equation 23.11, the determination of astrometric parameters.

In 1993, when the end of the main astrometric reductions in NDAC appeared to be within sight, it was therefore natural to start thinking of a possible alternative treatment of the grid coordinates, eliminating the artificial division into the great-circle reductions and sphere solution. A first plan was drafted by L. Lindegren in July 1993, and most of the software was written by C.S. Petersen at Copenhagen University Observatory between March and September 1994. However, because of other commitments and the more urgent requirements of the final iteration of the nominal reductions in NDAC, it was not until April 1995 that a first successful solution was made.

The input to the Copenhagen global iterative solution consisted of two major data sets:

- the attitude files (~ 1.6 Gigabyte), containing the results of the last iteration of the NDAC attitude determination;
- the grid coordinate files (~ 2.8 Gigabyte), containing the phase determinations of all the programme stars observed in each observational frame.

These data sets were essentially the output from the first stage of the data processing (Part A in Section 4.1) performed at the Royal Greenwich Observatory, but with the along-scan attitude component updated from the great-circle reductions. Additionally, three data bases were used:

- the star catalogue from one of the last NDAC sphere solutions (N37.1);
- the instrument parameters determined in the last run of great-circle reductions;
- the mean residual maps (Section 10.3).

The output consisted of the updated star catalogue including the 5×6 normal equations system for each of the $\simeq 118\,000$ programme stars.

In order to make the best use of existing procedures and minimise the need for additional software development, the following simplifications were introduced, in comparison with Equation 23.13:

- no global parameters ($\boldsymbol{\gamma}$) were included;
- the local parameters $\mathbf{n}_j$ included only the spline coefficients for the corrections to the along-scan attitude angle (Ω), with the knot sequences taken without changes from the last great-circle reduction;
- the instrument parameters were not updated.

Each iteration consisted of three main procedures run in sequence:

(1) initialisation of the normal equations for all the stars;

(2) a loop through the attitude intervals $\mathcal{T}_j$ to determine the spline coefficients $\mathbf{n}_j$ and, using the residuals of each such fit, update the normal equations for the corresponding stars;

(3) solution of the normal equations for one star at a time.

The initial catalogue, $\mathbf{a}^{(0)}$, was taken from the NDAC sphere solution N37.1. Only a single iteration was made ($m = 1$), and took about 18 hours on a Sparc-10 workstation. Nearly all the time was spent on procedure (2) above, the other two procedures being a matter of few minutes only.

Table 23.1. Standard deviations of the differences in astrometric parameters between the global solution NG1 and four other catalogues, after elimination of orientation and spin differences: the Hipparcos Catalogue (HIP), the final FAST and NDAC sphere solutions (F37.3 and N37.5), and the NDAC sphere solution N37.1 used as starting point for the global solution. The positions were compared at the epoch J1991.25. The standard deviations were computed by the robust method of Equation 16.22. The second column gives the number of stars used in each comparison. The last column gives the geometrical mean, D, of the five standard deviations in each comparison, as a somewhat arbitrary measure of the global 'distance' between the catalogues. Differences among the comparison catalogues are given in the lower part of the table (see Tables 16.9–16.10).

Solutions	No. of	Standard deviations (mas, mas/yr)					
compared	stars	$\Delta\alpha*$	$\Delta\delta$	$\Delta\pi$	$\Delta\mu_{\alpha*}$	$\Delta\mu_\delta$	D
NG1–HIP	101 093	0.77	0.66	0.90	1.08	0.94	0.858
NG1–F37.3	101 036	1.03	0.87	1.18	1.37	1.17	1.111
NG1–N37.5	100 919	0.72	0.61	0.84	1.01	0.88	0.800
NG1–N37.1	100 713	0.73	0.62	0.85	1.05	0.93	0.822
F37.3–HIP	101 189	0.51	0.43	0.62	0.64	0.51	0.536
N37.5–HIP	101 071	0.59	0.49	0.73	0.72	0.60	0.619
N37.5–F37.3	100 894	0.97	0.81	1.17	1.19	0.98	1.014
N37.5–N37.1	100 589	0.51	0.43	0.62	0.71	0.60	0.566

Results

The first iteration of the NDAC global iterative solution, here denoted NG1, resulted in a star catalogue with 117 616 entries. However, in the following only the 101 093 entries in common with the basic subset defined in Section 16.2 will be considered, thus avoiding the major complications due to stellar duplicity.

The results of NG1 were compared with the final Hipparcos Catalogue (HIP) and the last FAST and NDAC sphere solutions (F37.3 and N37.5) according to the general principles described in Section 16.6. Additional comparisons were made with N37.1, the catalogue used as a starting approximation, in order to obtain the mean updates produced by the global solution, and between the various comparison catalogues in order to see the typical differences arising in the nominal Hipparcos processing.

First, the global orientation and spin differences of NG1 with respect to HIP were determined. The results were:

$$\boldsymbol{\varepsilon}_0 = \begin{pmatrix} -39.910 \\ -41.592 \\ +67.666 \end{pmatrix} \text{ mas [J1991.25]}, \qquad \boldsymbol{\omega} = \begin{pmatrix} -1.389 \\ +0.832 \\ +1.069 \end{pmatrix} \text{ mas/yr} \qquad [23.14]$$

These values are extremely close to the corresponding values for the sphere solution used as starting point for the global iteration, N37.1 (see Table 16.8), showing that the iteration did not introduce any significant change in the global reference frame. After elimination of the orientation and spin differences, the differences in each of the five astrometric parameters were calculated with respect to the comparison catalogues. The standard deviations of the differences, estimated according to Equation 16.22, are shown in Table 23.1.

In each comparison, the standard deviations of the differences in the five astrometric parameters vary in much the same way, and a geometrical mean, denoted D in Table 23.1, can be taken as a global measure of the 'distance' between any two solutions. In this sense, the global solution is 'nearest' to the final NDAC sphere solution ($D = 0.800$), which is not surprising, as they used the same basic input data. In this connection it is worth noting that the distance to the initial catalogue, N37.1, is slightly greater ($D = 0.822$). Clearly the global solution is rather different from both N37.1 and N37.5, although less different from these than the FAST and NDAC sphere solutions from each other. Interestingly, the global solution, while moving away from N37.1 and N37.5, does not seem to approach the FAST solution (or HIP), but rather behaves to some extent as independent of the FAST and NDAC sphere solutions. This was confirmed by the properties of the parallax distribution (see below).

The median offset in parallax was $\langle \pi_{\rm NG1} - \pi_{\rm HIP} \rangle = -0.012 \pm 0.003$ mas. The offset was found to be slightly dependent on colour, with a mean coefficient of $+0.06 \pm 0.01$ mas per magnitude of $V - I$. The hemisphere asymmetry, defined in analogy with Equation 16.24, was $\Delta\pi_0 = -0.028 \pm 0.006$ mas. The width of the parallax distribution indicated that NG1 was slightly more precise than N37.5, while the fraction of negative parallaxes lead to the contrary conclusion. Both criteria showed that a weighted mean of NG1 and N37.5, with about equal weight to the two solutions, would provide a significant improvement of the parallaxes (by $\simeq 8$ per cent in the median standard error). Even with respect to the final Hipparcos Catalogue, the global solution would contribute significant information, reducing the median standard error in parallax by a few per cent. This supports the previous conclusion that the modelling errors in NG1 are rather different from those in the sphere solutions.

Large-scale differences between NG1 and HIP, apart from the global offset in orientation and spin, were investigated by computing the rotational offsets in eight different areas of the sky (see Table 16.12). In position ($\boldsymbol{\varepsilon}_0$) the absolutely largest difference was 0.080 mas, while in proper motion ($\boldsymbol{\omega}$) it was 0.175 mas/yr. These values are somewhat larger than the FAST–NDAC differences reported in Table 16.12, but not alarmingly large and probably related to the chromatic effects described below.

By far the most serious systematic effects revealed by the various comparisons are related to the colours of the stars. The slight chromatic offset of the parallaxes was already noted. Chromatic effects are however much more drastic in the positions and, especially, the proper motions. They show up, for instance, as very large chromatic rotation parameters, defined as in Equation 16.27:

$$\boldsymbol{\varepsilon}_0' = \begin{pmatrix} -0.444 \\ +0.098 \\ +0.688 \end{pmatrix} \text{ mas mag}^{-1} \text{ [J1991.25]}, \qquad \boldsymbol{\omega}' = \begin{pmatrix} -1.586 \\ +0.723 \\ +0.911 \end{pmatrix} \text{ mas yr}^{-1} \text{ mag}^{-1} \qquad [23.15]$$

This is very likely caused by inadequate modelling of instrument chromaticity, in particular the 'constant chromaticity' term c_{00} not included among the instrument parameters (see Equation 10.9 and Figure 16.5). Since this would have to be included among the global parameters $\boldsymbol{\gamma}$, it was not taken into account in NG1. The effect has both the magnitude and the strong time dependence needed to explain the strong influence on the proper motion system. Some of the earlier comparisons of the NDAC and FAST sphere solutions showed similar colour-dependent differences (Table 16.13), which disappeared only in the final solutions after careful modelling of the chromaticity.

Although very robust methods were used in the comparisons, the error statistics of NG1 are somewhat degraded by the rather unclean appearance of the solution in comparison with either sphere solution or the Hipparcos Catalogue. This is manifested, for instance, in the number of parallax values below −10 mas, which is 80 in NG1, but only 9, 16, and 12 in F37.3, N37.5, and HIP (as always, only the intersections of these catalogues with NG1 and the basic subset were considered).

Finally it should be remarked that the standard errors in NG1, computed from the curvature matrix of Equation 23.11, were typically underestimated by a factor 0.5 to 0.8 compared with N37.5. This was the case even though the average unit-weight variance of the residuals was close to one. The discrepancy highlights the problem, referred to in the previous section, of finding a practical method to compute reliable covariance matrices for the global iterative solution.

Conclusions

Although the Copenhagen experiment provided only a single iteration step, the practical feasibility of the global iterative solution was clearly demonstrated. Moreover, it resulted in a solution which was not inferior to the standard sphere solution in terms of overall precision, but rather different in terms of the detailed (modelisation) errors. Given more time and work, it is probable that the major shortcomings of NG1—in particular the chromatic errors and the occurrence of outliers—could have been eliminated, resulting in a solution somewhat better than the standard NDAC sphere solution. Furthermore, rigorous inclusion of the instrument and global parameters among the unknowns, as well as fine-tuning of the attitude smoothing and more iterations, would surely result in additional improvements.

An obvious extension of the method would be to merge the FAST and NDAC data already at the grid-coordinate level and perform a global iterative solution on the merged data. However, the degree of improvement in the end results remains uncertain.

While the global iterative solution thus appeared to be a very promising alternative approach to the Hipparcos data reductions, the amount of additional work required was likely to be substantial, and it had to be abandoned as the baseline NDAC contribution to the Hipparcos Catalogue. Further study of the method should nevertheless be encouraged, especially in view of future space astrometry missions.

23.4. The Challenges for the Future

A future space astrometry mission will clearly rest very heavily on the Hipparcos experiences. Certain issues, such as the basic conceptual problems faced by the first global scanning space astrometry experiment—the derivation of absolute trigonometric parallaxes, the determination of the astrometric parameters of complex double and multiple systems, and so forth—have been convincingly demonstrated.

An experiment aiming for the cataloguing of the astrometric parameters of tens of millions of stars will certainly face numerous problems associated with the treatment of such a large quantity of data related to a very large number of stars. Not only will the Hipparcos experience help in preparing such reductions, but developments in computational power and object-oriented data bases mean that the complexities related only to the data volume will certainly not rise in proportion to the number of objects.

Apart from the instrumental challenges of designing, launching and operating a satellite with the requisite optical and geometric stability, the challenges to be faced in proceeding from milliarcsec to microarcsec astrometry will most likely be of comparable complexity as those involved in the progress to milliarcsec positional accuracy. Metric and light travel time effects will compound the complexities of the astrometric model, and its formulation and practical solution. And the conceptual definition of a reference system in which differential galactic rotation becomes a significant observable effect may demand a more complex representation of the space motion of each object observed.

What appears beyond doubt is that the principles of the Hipparcos space astrometry mission can be carried over to the realms of a microarcsec astrometry experiment, the successful completion of which would characterise to an even more significant degree the structure and evolution of stars, and our Galaxy, in a manner completely impossible using any other methods.

L. Lindegren & M.A.C. Perryman