Content of the HIP double star annex

L. Lindegren
Lund Observatory

1. Introduction

This note contains some general considerations on the design of the double star annex (DSA) to the output catalogue, both the printed and machine-readable versions, without providing a detailed proposal. It should be regarded as a starting point for our discussions.

The main part of HIP will contain summary data on entries that are resolved by the reductions into distinct components. As a general principle, we have tried to put as much useful data as possible into the main catalogue, in particular the relative positions and magnitudes of binary systems without detectable relative motion (this covers, after all, a majority of all cases). In the DSA, full information must be given on all kinds of systems. Consequently there is inevitably a good deal of overlap between the main catalogue and the DSA.

As a basic rule I would insist that the DSA only contains data derived from the satellite. This is difficult enough, without the added complications of having to incorporate ground-based information in a consistent way, and it is in line with the philosophy adopted for the main catalogue.

2. Data hierarchy

The essential difference between the main catalogue and the DSA has to do with the organization of the data: in the main catalogue they are arranged into entries distinguished by the HIP numbers, and all information on one entry must be given on a single line. In the DSA we have more freedom to organize the data as required by the different types of object. In particular, it is natural to adopt a hierarchic ordering with three levels:

(i) at the highest level there are the (double or multiple) 'systems', distinguished by a CCDM or IDS type identifier;

(ii) the next level consists of the 'components', i.e., single stars or very close binaries, unresolved by Hipparcos ($\rho < 100$). Within each system, the components are distinguished by additional letters or digits (see also Section 5);

(iii) the lowest level consists of 'normal points' calculated from the observations, identified by an epoch. Exactly what the normal point should contain is discussed below.

Hierarchic data structures are well adapted to the computer, e.g., in the form of a relational database, but in printed form it may lead to a lot of duplication of data and/or empty fields. This follows from the necessity to arrange the data in lines and columns, which is not by itself a hierarchic structure. (The same applies to the machine-readable form if given as a simple table rather than a database.) With a good typographical layout this may not be a problem, though.
Only components observed by Hipparcos will be included. This means that a HIP number is associated with every component in the DSA, although more than one component may have the same HIP number.

3. Definition of a normal point

I assume that the concepts of a system and a component are relatively clear. However, we need to define what constitutes a 'normal point'.

Each normal point should ideally represent an independent estimate of a fixed set of parameters. The assumption of independency is mainly a practical one: it will simply not be practical to specify the correlations between data on different lines, because there is an indefinite number of possible such correlations. (The same convention applies to the main catalogue, where we neglect correlations between different entries.) The simplest normal point would be an estimate of the (barycentric) position at a given epoch. However, since practically all components have detectable proper motion in an absolute frame, this would give a minimum of two normal points, or two lines in the DSA, for every component, which seems excessive. The second simplest normal point is an estimate of the position and proper motion at a given epoch. With this choice, one normal point will be sufficient for most components; only those with detectably non-linear motion will require two or more normal points (each consisting of an epoch, a position at the epoch, and a proper motion at the epoch). The third simplest normal point consists of the position, proper motion, and acceleration at a given epoch, but then the acceleration would mostly be insignificant (or forced to zero).

The most practical solution is therefore to define the normal point as a quasi-independent estimate of the position and proper motion at a given epoch. This also conforms best with the main catalogue format. Indeed, many columns in the DSA could have exactly the same format as in the main catalogue. For components with a quadratic term there must then be two normal points at different epochs, each with a position and a proper motion. The epoch of each normal point should be chosen to minimize the correlation between the position at that epoch and the proper motion estimate. This is not completely trivial, because the best epoch is usually different in RA and Dec (or any other direction), so some compromise must be found. Appendix A contains some additional considerations on the choice of epochs.

It should be noted that this structure logically suits astrometric binaries as well, i.e., systems with just one component, but non-linear motion and therefore two or more normal points.

4. Convenience data

According to the above principles the astrometric data for each component is given in the form of absolute quantities — position, proper motion and parallax —, rather than the standard relative data for the secondary, etc. This may look strange to some users, and it certainly makes it difficult to get a quick overview of what kind of system it is. I propose therefore that approximate relative data are added as a convenience, but to make it clear that the full information is only represented by the absolute quantities.

The use of an absolute position for each component avoids the risk of approximation errors in converting from relative to absolute data. This may seem like a trivial point, but I think there would be a real danger in giving high-precision relative data (cf. Appendix B).
5. Designation of components

In HIC the components of double/multiple stars are identified by letters A, B, etc. A significant problem might be to identify the components we obtain in our solution with these a priori components. In many (most?) cases it will be easy but I can easily imagine that it will be very difficult for many systems, and impossible for some. I propose a radical solution to this problem: Use only digits (1, 2, ...) or perhaps lower-case letters (a, b, ...) to number the components obtained in the Hipparcos analysis. We can (perhaps) do the numbering in such a way that for the majority of systems there is a direct relationship $1 \equiv A$, $2 \equiv B$, etc. The important thing is that we do not claim to identify the components in the ground-based catalogues. If this principle is adopted, then the same thing must of course apply to the main catalogue.

6. Correlations between components or normal points

As mentioned before, the tabular format of the printed DSA does not easily allow to include correlations between the data on different lines. It could perhaps be achieved for simple systems, consisting of just two components with one normal point each, but it would be rather complicated and could not be extended to multiple systems etc.

Yet it may be quite useful to have inter-component correlations. They are in fact needed, e.g., to compute the standard errors of any combination of data from the different components, such as the position of the photocentre (or mass centre, if the masses are known).

For many close binaries the main catalogue will give data for the photocentre, including its standard errors, so in many cases this information is still present in the catalogue. However, systematic inclusion of correlations in the printed catalogue seems unfeasible, and should be relegated to the machine-readable version. Here we may take advantage of the greater freedom to use records with different contents. Thus, for each system one may have a numbering $(i = 1, 2, \ldots)$ of the records containing the normal points in that system, and the normal points are then followed by a number of records containing the correlations for all relevant combinations of indices $i$. For instance, a quadruple star could have four data records for the normal points, which are then followed by up to six records for the correlations.

7. An example of the printed format

The attached table is a first attempt at defining the content and layout of the printed DSA. There are probably more flags required, but the important aspect right now is whether the general principles described above will give an acceptable format.

APPENDIX A: Decorrelating the position/proper motion data

The following examples considering the motion in one coordinate ($x$) only may give some insight into the problem of finding the optimum epochs to represent a non-linear motion. It is assumed that the motion is in the DS analysis represented by a polynomial of degree $n$, and that the normal points ($x$ and $\dot{x}$) are to be evaluated at $n$ different epochs.
For linear motion the coordinate along one axis can be written \( x = a_0 + a_1 t \). The DS analysis provides an estimate of the vector \((a_0, a_1)\) and its \(2 \times 2\) covariance matrix with elements \(c_{00}, c_{11}, \text{ and } c_{11}\). At the arbitrary epoch \(t\) the covariance of \(x(t)\) and \(\dot{x}(t) = a_1\) is \(c_{01} + c_{11}t\). The correlation between the position and the proper motion therefore vanishes for \(t = -c_{01}/c_{11}\). (It is easily seen that this also minimizes the variance of \(x\).)

Now consider quadratic motion, \(x = a_0 + a_1 t + a_2 t^2\), for which we have the \(3 \times 3\) covariance matrix with elements \(c_{00}, ..., c_{22}\). The covariance of \(x(t)\) and \(\dot{x}(t) = a_1 + 2a_2 t\) is now found to be \(c_{01} + (c_{11} + 2c_{02})t + 3c_{12}t^2 + 2c_{22}t^3\). Putting this equal to zero gives a cubic equation whose extreme roots are the desired epochs. (I surmise that there are always three real roots if the covariance matrix is positive definite.) For instance, if the observations are uniformly distributed in the interval \(-T < t < +T\), then the extreme roots are \(t = \pm \sqrt{2/ST}\).

For \(n > 2\) it is not clear if any useful results are obtained with the algebraic method, but it is in any case unlikely that such polynomials will be used in the DS analysis.

In relativity there are two (in general correlated) coordinates, such as RA and Dec. For linear motions I have pointed out elsewhere that the trace of the covariance matrix is invariant to rotations of the coordinate axes, and that minimizing the trace may thus be a suitable way to define a coordinate-independent mean epoch. In terms of the above covariance terms \(c_{ij}\) in one coordinate (RA, say) and \(c'_{ij}\) in the other (Dec), the effective epoch then becomes \(t = -(c_{01} + c'_{01})/(c_{11} + c'_{11})\). I assume that this procedure can be extended to the quadratic case, so that two effective epochs can be defined in a coordinate-independent way.

APPENDIX B: Transformation from absolute to relative position

The following expressions are sometimes used to define (explicitly or implicitly) the relative positions of two components \((\alpha_1, \delta_1)\) and \((\alpha_2, \delta_2)\):

\[
\begin{align*}
\rho \sin \theta &= \Delta \alpha \cos \delta \\
\rho \cos \theta &= \Delta \delta
\end{align*}
\]  

(B1)

where \(\Delta \alpha = \alpha_2 - \alpha_1\) and \(\Delta \delta = \delta_2 - \delta_1\). This is OK for most situations when the separation is small and/or the accuracy requirements are not too high. However, this formulation should not be used for high-precision work. There is, to begin with, an ambiguity as to which \(\delta\) should be used in the \(\cos \delta\) factor. At high declinations and for separations of a few tens of arcsec, this becomes important at the 10-50 mas level. Moreover, the \(\rho\) calculated by this formula is not the true arc length between the components, nor is the \(\theta\) the actual position angle of the system. Clearly one must use the rigorous equations, which read:

\[
\begin{align*}
\sin \rho_{12} \sin \theta_{12} &= \cos \delta_2 \sin (\alpha_2 - \alpha_1) \\
\sin \rho_{12} \cos \theta_{12} &= \cos \delta_1 \sin \delta_2 - \sin \delta_1 \cos \delta_2 \cos (\alpha_2 - \alpha_1)
\end{align*}
\]  

(B2)

Note that \(\theta_{21} \neq \theta_{12} + \pi\).

Incidentally, in the NDAC double star analysis we use local rectangular coordinates in a tangent plane at a fixed point (chosen for each system); this is similar to the use of standard coordinates in photographic astrometry. The conversion from such plane coordinates to absolute positions, and conversely, is straightforward, and from there one may compute the relative data according to (B2). This again illustrates some of the pitfalls of the \(\rho, \theta\) formalism in high-precision work. The consistent use of absolute coordinates eliminates a lot of potential problems.
Example of data format for the printed double star annex of HIP (Lindgren, 1994 March 8)

<table>
<thead>
<tr>
<th>INDEX</th>
<th>Comp</th>
<th>Hp</th>
<th>sm</th>
<th>Ep</th>
<th>R.A.</th>
<th>Doc.</th>
<th>nua</th>
<th>nud</th>
<th>pl</th>
<th>sa</th>
<th>sd</th>
<th>cad</th>
<th>uma</th>
<th>smd</th>
<th>cmad</th>
<th>sp</th>
<th>opsa</th>
<th>opmd</th>
<th>HIP</th>
<th>thet</th>
<th>rho</th>
<th>rel</th>
<th>Rem</th>
</tr>
</thead>
<tbody>
<tr>
<td>20512+5126</td>
<td>a</td>
<td>7.301</td>
<td>.009</td>
<td>91.17</td>
<td>312.708229370</td>
<td>451.43287619</td>
<td>-50.3</td>
<td>41.7</td>
<td>38.5</td>
<td>2.7</td>
<td>3.3</td>
<td>-0.08</td>
<td>0.5</td>
<td>1.3</td>
<td>-0.22</td>
<td>2.0</td>
<td>-0.02</td>
<td>+0.00</td>
<td>102921</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>10.048</td>
<td>.067</td>
<td>91.17</td>
<td>312.708240127</td>
<td>451.43271121</td>
<td>-51.1</td>
<td>41.9</td>
<td>6.5</td>
<td>8.5</td>
<td>-0.12</td>
<td>7.6</td>
<td>2.9</td>
<td>-0.25</td>
<td>102921</td>
<td>36</td>
<td>0.91</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20514-0537</td>
<td>a</td>
<td>6.391</td>
<td>.010</td>
<td>91.02</td>
<td>312.792393370</td>
<td>-05.45967619</td>
<td>96.5</td>
<td>-7.1</td>
<td>9.1</td>
<td>2.7</td>
<td>3.3</td>
<td>-0.08</td>
<td>0.5</td>
<td>1.3</td>
<td>-0.22</td>
<td>3.1</td>
<td>+0.11</td>
<td>+0.12</td>
<td>102945</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>7.582</td>
<td>.072</td>
<td>90.47</td>
<td>312.792500647</td>
<td>-05.45951030</td>
<td>85.2</td>
<td>13.9</td>
<td>8.6</td>
<td>9.5</td>
<td>-0.10</td>
<td>3.8</td>
<td>3.9</td>
<td>+0.18</td>
<td>102945</td>
<td>219</td>
<td>0.33</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20531+2909</td>
<td>a</td>
<td>8.212</td>
<td>.006</td>
<td>90.39</td>
<td>313.025298129</td>
<td>+29.15612129</td>
<td>-30.2</td>
<td>10.5</td>
<td>105.2</td>
<td>1.9</td>
<td>2.1</td>
<td>-0.15</td>
<td>2.5</td>
<td>2.8</td>
<td>-0.12</td>
<td>1.1</td>
<td>+0.01</td>
<td>-0.12</td>
<td>103074</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>9.733</td>
<td>.050</td>
<td>91.31</td>
<td>315.01230181</td>
<td>-75.39137619</td>
<td>91.1</td>
<td>-1.5</td>
<td>6.5</td>
<td>8.5</td>
<td>-0.12</td>
<td>2.6</td>
<td>2.9</td>
<td>-0.25</td>
<td>103745</td>
<td>157</td>
<td>18.01</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21014-7526</td>
<td>a</td>
<td>8.510</td>
<td>.019</td>
<td>91.31</td>
<td>315.01230181</td>
<td>-75.39137619</td>
<td>52.3</td>
<td>-41.7</td>
<td>5.1</td>
<td>2.7</td>
<td>3.3</td>
<td>-0.08</td>
<td>0.5</td>
<td>1.3</td>
<td>-0.22</td>
<td>2.0</td>
<td>-0.02</td>
<td>+0.00</td>
<td>103745</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>9.733</td>
<td>.050</td>
<td>91.31</td>
<td>315.01211181</td>
<td>-75.39181121</td>
<td>91.1</td>
<td>-1.5</td>
<td>6.5</td>
<td>8.5</td>
<td>-0.12</td>
<td>2.6</td>
<td>2.9</td>
<td>-0.25</td>
<td>103745</td>
<td>157</td>
<td>18.01</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21114+3542</td>
<td>a</td>
<td>9.610</td>
<td>.049</td>
<td>91.11</td>
<td>315.01230181</td>
<td>+35.39127619</td>
<td>22.3</td>
<td>-11.5</td>
<td>3.1</td>
<td>3.7</td>
<td>3.3</td>
<td>-0.08</td>
<td>0.5</td>
<td>1.3</td>
<td>-0.22</td>
<td>2.0</td>
<td>-0.02</td>
<td>+0.00</td>
<td>104510</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>11.933</td>
<td>.150</td>
<td>91.11</td>
<td>315.01211181</td>
<td>+35.39181121</td>
<td>22.3</td>
<td>-11.5</td>
<td>3.1</td>
<td>3.7</td>
<td>3.3</td>
<td>-0.08</td>
<td>0.5</td>
<td>1.3</td>
<td>-0.22</td>
<td>2.0</td>
<td>-0.02</td>
<td>+0.00</td>
<td>104510</td>
<td>29</td>
<td>3.61</td>
<td>a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21514-0537</td>
<td>a</td>
<td>6.291</td>
<td>.010</td>
<td>91.22</td>
<td>312.792393370</td>
<td>-05.45967629</td>
<td>196.7</td>
<td>17.1</td>
<td>9.1</td>
<td>2.1</td>
<td>3.5</td>
<td>+0.18</td>
<td>2.5</td>
<td>2.3</td>
<td>-0.32</td>
<td>3.1</td>
<td>+0.11</td>
<td>+0.12</td>
<td>107945</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>7.592</td>
<td>.072</td>
<td>91.22</td>
<td>312.792500647</td>
<td>-05.45951030</td>
<td>196.7</td>
<td>13.9</td>
<td>8.2</td>
<td>8.2</td>
<td>+0.12</td>
<td>2.8</td>
<td>2.9</td>
<td>-0.08</td>
<td>107945</td>
<td>219</td>
<td>0.33</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>7.912</td>
<td>.110</td>
<td>91.22</td>
<td>312.79250421</td>
<td>-05.45951292</td>
<td>188.7</td>
<td>11.8</td>
<td>7.5</td>
<td>10.9</td>
<td>+0.11</td>
<td>4.7</td>
<td>4.2</td>
<td>+0.08</td>
<td>102945</td>
<td>101</td>
<td>20.93</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Col 1: INDEX = a CCM/IDS type identifier of the system
Col 2: Comp = component identifier (a, b, c, ...)
Col 3-4: Hp, sm = magnitude with standard error
Col 5: Ep = epoch for position and proper motion
Col 6-7: R.A., Doc = Right Ascension, Declination at the epoch in Col 5
Col 8-9: nua, nud = proper motions in mas/yr at the epoch in Col 5
Col 10: pl = parallax in mas
Col 11-19: ........ = standard errors and correlations as in the main catalogue
Col 20: HIP = entry number in the main catalogue
Col 21-22: thet, rho = approximate position angle (deg) and distance (arcsec) of component from another component (see Col 23)
Col 23: rel = the identifier for the component to which the data in Col 21-22 refer

Remarks: 1. A binary with linear relative motion
2. A binary with relative motion; quadratic for component b
3. Astrometric binary (quadratic proper motion)
4. Optical binary
5. Binary without relative motion
6. Triple star with relative motion

Questions: What about colours?
Flags? Order of columns; perhaps 20-23 should be somewhere in the beginning?