Some empirical corrections to the IDT information matrix
by S. Söderhjelm, Lund Observatory

1. Introduction

While starting the work on an improved SEEKDBL-program, I soon realized that the naive expectation of 'flat' $\Delta \chi^2$ -histograms (cf. NDAC/LO/090) for single stars was not fulfilled. This has prompted the present (unplanned) investigation, which may be have some importance also for the main HIPPARCOS reductions. What I have done is mainly to 'correct' empirically the weighting in the IDT preprocessing, changing thereby (mostly) the derived information matrices for the $\beta$-parameters. With some simple (but ad hoc) adjustments, I finally get reasonably flat histograms for $\Delta \chi^2$, but this means only that the 44 and 45 components of the information matrices have sensible values. I have not checked the consistency with respect to the other $\beta$-parameters, nor any theoretical justifications for my expressions. This is hereby left for further consideration in the final design of the IDT preprocessing.

2. The calibration of $\beta_4$ and $\beta_5$

The definition of $\Delta \chi^2$ is simply

$$
\Delta \chi^2 = \text{Inf}(4,4)(\Delta \beta_4)^2 + 2\text{Inf}(4,5)\Delta \beta_4 \Delta \beta_5 + \text{Inf}(5,5)(\Delta \beta_5)^2
$$

where $\Delta \beta_i$ is the observed minus the calibrated value of $\beta_i$, and where $\text{Inf}(i,j)$ is the information matrix derived in the IDT preprocessing. For a single star, we expect $\Delta \chi^2$ to follow a 2 d.o.f. $\chi^2$-distribution, and as described in NDAC/LO/090, this is equivalent to equal counts in each of eight bins of an accumulated histogram of $\Delta \chi^2$-values. In my present simulations, this was simply not true, causing a too large proportion of true singles to be flagged as double.

Rather quickly, I realized that my 'calibration' values of $\beta_4(= M_2/M_1 \cos v)$ were inaccurate. I had used simply the exact input to the simulation program, but a subtle effect of the binning in the IDT preprocessing is a reduced ratio of the second to the first harmonic amplitude, as noted by Lindegrén in NDAC/LO/058. My present results show a somewhat larger effect, with a simple linear magnitude-dependence, viz.

$$(M_2/M_1)_{\text{eff}} = (M_2/M_1)_{\text{true}}[0.9870 + 0.0010(H - 3.5)]$$

The exact value is important only for the brightest stars ($H < 3$), and at $H > 8$ eq. (2) is rather arbitrary. (In the real reductions, the $\beta_{4,5}$-values will be carefully calibrated using bright single stars, and no bias is expected).
3. The information matrix

Unfortunately, even after the improved calibration, the $\Delta \chi^2$ is came out too large, and the only remedy is then to diminish the $InJ$-values. This may be done in many ways, and first, I just replaced the denominator $I_k$ (expected number of counts in sample k) in the single ML-iteration by $I_k(1 + \epsilon_2 I_k)$. This has a natural interpretation as an added 'fixed' variance $\epsilon_2$ that may not be reduced by ever so favourable photon-statistics for the brightest stars. This idea was rather successful, but it is necessary to allow $\epsilon_2$ to vary slightly with the magnitude. From a lot of tests for bright stars ($H < 3.5$), I get almost flat $\Delta \chi^2$-histograms using

$$\epsilon_2 = 0.00041 \times 10^{0.04H}$$  \hspace{1cm} (3) $$

(For $H < 0$, there is an excess of both large and small $\Delta \chi^2$-values, for any value of $\epsilon_2$. This is of no practical consequence, however, because these stars are so few).

For the 'normally' faint HIPPARCOS-stars ($H > 8$), the IDT counts per sample are less than about 3, and the $\epsilon_2$-parameter has very little effect. Still, the $InJ$-values are too large, and increasingly so at fainter magnitudes. This is probably due to an increasing discrepancy between the true and estimated $I_k$ expectation values as the stars get fainter, and the cure is now to replace $I_k$ by $I_k(1 + \epsilon_1 + \epsilon_2 I_k)$. The new parameter $\epsilon_1$ is more variable with magnitude (and less easily understandable), but again I have succeeded in getting reasonably flat $\Delta \chi^2$-histograms by using simply

$$\epsilon_1 = 0.0086 \times 10^{0.12H}$$  \hspace{1cm} (4) $$

4. Sample results

Using the $\epsilon$-values given in eqs. (3) and (4), I have made 40 single star simulations at each of five magnitudes. The 40 histograms at each magnitude (each one including about 8 (frames) x 150 (FOV=8)= 1200 $\Delta \chi^2$-values) were then added to give final 'summed' histograms with some 6000 individual $\Delta \chi^2$-values in each bin. The corresponding uncalibrated ($\epsilon=0$) histograms could be obtained simultaneously, and Figs 1-5 show the two histograms at each magnitude. The left (uncalibrated) ones are obviously distorted, especially at extreme magnitudes, and the 'adjusted' ones sufficiently flat that the double-star detection will work easily. The remaining 'peaks' in the leftmost bin (corresponding to the largest $\Delta \chi^2$-values) may probably be reduced even further by 'fine-tuning' of the $\epsilon$-values, but this is hardly worthwhile at present.

The important untackled question is instead whether the modified information matrices give realistic errors in the more fundamental phase-determinations ($\beta_3$) also. This will be partly answered in my further double-star solution-tests, but it is important to make some further investigations also at the fundamental IDT preprocessing level. It should be noted, however, that the effects sought are relatively small, and that large numbers of simulation-experiments are therefore needed.
Fig. 1-3. Summed $\Delta \chi^2$-histograms for 40 single stars of magnitude $H=1.0$, $H=3.5$ and $H=6.0$. The left ones are with the original information-matrices ($\epsilon=0$), the right ones with the corrections described in the text.
Fig. 4-5. Summed $\Delta \chi^2$-histograms for 40 single stars of magnitude $H=8.5$ and $H=11.0$. The left ones are with the original information-matrices ($\epsilon=0$), the right ones with the corrections described in the text.