Covariance of IDT signal parameters with different weightings

by L Lindegren

Introduction

This note gives a theoretical and numerical comparison of the phase accuracy of the IDT signal, using least-squares estimation of the signal parameters with different weightings of the individual samples. In particular, we compare (i) the Cramér-Rao bound, (ii) unweighted least-squares, and (iii) unweighted least-squares followed by a single ML iteration.

Theory

The IDT counts for a certain star in a frame, $\{N_k\}$, are modelled as a Poisson process with expectation

$$E(N_k) = e_k = b_1 + b_2 \cos H_k + b_3 \sin H_k + b_4 \cos 2H_k + b_5 \sin 2H_k$$

$$= c_k' b_k$$  \hspace{1cm} (1)

where $H_k$ is the known reference phase of each sample and $c_k' = (1 \cos H_k \ldots \sin 2H_k)$. Let $n$ be the number of samples in the observation and $C$ the $(5,n)$-matrix the columns of which are the 5-vectors $c_k$. The $n$ observation equations can now be written in matrix form

$$C'b + v = n$$  \hspace{1cm} (2)

where $v$ is the noise vector and $n = (N_1 N_2 \ldots N_n)'$ the vector of counts.

Introducing a general weight matrix $W$ of dimension $(n,n)$, the least-squares solution of (2) is

$$\hat{b} = (CW'C)^{-1}CWn$$  \hspace{1cm} (3)

Now if $b_0$ is the true vector of harmonic coefficients, we have

$$E(n) = e = C'b_0$$  \hspace{1cm} (4)
and consequently

$$E(\hat{b}) = (CWC')^{-1}CW E(n) = (CWC')^{-1}CWC'b_0 = b_0$$

(5)

showing that $\hat{b}$ is an unbiased estimate of $b_0$ for any weight matrix (provided that $CWC'$ is positive definite).

By choosing different weight matrices $W$, we therefore obtain different least-squares estimates $\hat{b}$ which are all unbiased, but some of which may be "better" than others. In particular, we want to study how the covariance of the estimate depends on $W$ and whether we can minimize the variance of the phase estimate by a suitable choice of $W$.

The covariance of $\hat{b}$ for arbitrary (symmetric) weight matrix is

$$V = E[(\hat{b} - b_0)(\hat{b} - b_0)'] = (CWC')^{-1}CW E[(n - e)(n - e)'] W C' (CWC')^{-1}$$

$$= (CWC')^{-1} C W E W C'(CWC')^{-1}$$

(6)

where

$$E = \text{diag}(e) = \text{diag}(e_1 e_2 \ldots e_n)$$

(7)

Case 1. The Cramér-Rao bound

Choosing the weight matrix

$$W_1 = E^{-1}$$

(8)

in (6) we have the covariance matrix

$$V_1 = (CE^{-1}C')^{-1}$$

(9)

which is in fact the Cramér-Rao bound. That is, for arbitrary $W$ we have

$$V \geq V_1$$

(10)

(where the inequality means that $V - V_1$ is non-negative definite). To prove this inequality in the general case shall not be done here, but it is useful to make a parabolic expansion around $V_1$ to study the sensitivity of $V$ to small deviations from the optimal weighting. In terms of the differential
weight matrix

\[ D = W - W_1. \]  

we have

\[ (C W E C') = (E^{-1} + D) E (E^{-1} + D) C' = V_1^{-1} + 2 C D C' + C W E C' \]  

and to second degree in \( D \),

\[ ((C W C')^{-1}) = (I + V_1 C D C')^{-1} V_1 \]

\[ = (I - V_1 C D C' + V_1 C D C' V_1 C D C') V_1 \]

\[ = V_1 - V_1 C D C' V_1 + V_1 C D C' V_1 C D C' V_1 \]  

Introducing (12) and (13) in (6) and retaining only terms up to the second degree in \( D \) we find

\[ V \propto V_1 + V_1 C W E C' V_1 - V_1 C D C' V_1 C D C' V_1 \]

\[ = V_1 + V_1 C D W C' V_1 \]  

where

\[ Q = E - C' V_1 C \]  

(The inequality (10) can now be proved, locally at least, by showing that \( Q \) is non-negative definite. It is seen that \( Q E^{-1} = I - C' V_1 C E^{-1} \) is idempotent, i.e. \( (Q E^{-1})^m = Q E^{-1} \) for arbitrary \( m > 0 \); therefore, its eigenvalues are all either 0 or 1. Since \( E \) is moreover positive definite, it follows that \( Q \) is non-negative definite, Q.E.D. In fact, \( Q \) is singular (as can be seen from \( Q E^{-1} C' = C' - C' V_1 C E^{-1} C' = 0 \)), so that the optimum weight matrix is non-unique; this just means that the estimate is invariant to scaling of \( W \).)

It is easy to see now why the Cramér-Rao bound in general cannot be attained in a practical computation. The optimum weight function \( W_1 \) depends on the true expected counts \( e \) [through (7)-(8)], which are not known. The best we can do is to estimate the optimum weight matrix by using (e.g.) the
estimate \( \hat{b} \) to calculate estimated counts \( \hat{\alpha} = C'\hat{b} \) in an iterative solution. This is equivalent to a Maximum Likelihood solution. Even after full convergence, however, the weight matrix will be slightly sub-optimal, and the covariance of the estimate slightly greater than \( V_1 \). It is also easy to see that this effect becomes relatively more important, the fainter the star is (and the more uncertain \( \hat{W}_1 \) becomes).

Case 2. Unweighted least-squares

With the weight matrix

\[
W_2 = I
\]  

(16)

we find the covariance of the unweighted least-squares solution, \( \hat{b}_2 \),

\[
V_2 = (CC')^{-1} CEC' (CC')^{-1}
\]  

(17)

In contrast to Case 1, this solution can be obtained exactly and without iteration. (It should be remembered, however, that the calculation of \( V_2 \) still requires knowledge of the true expected counts, \( E \), and we can therefore only produce an estimate of \( V_2 \) in practical computation.)

Case 3. Unweighted least-squares followed by a single ML iteration

The unweighted least-squares solution \( \hat{b}_2 \) may be considered good enough for calculating an approximation to the weight matrix, viz.

\[
W_3 = \text{diag}(C'\hat{b}_2)^{-1}
\]  

(18)

The resulting estimate is called \( \hat{b}_3 \) and its variance can be estimated by means of the parabolic expansion (14)-(15). The elements of \( D \) (on the diagonal) are

\[
D_{\lambda\lambda} = 1/(C'\hat{b}_2) - 1/(C'\hat{b}_0) = -C'\hat{b}_2 - b_0/e_{\lambda}^2
\]  

(19)

so that the expectation of the elements of \( DOD \) is

\[
\mathbb{E}(DOD)_{\lambda m} = 0_{\lambda m} + \frac{1}{e_{\lambda}^2 - 2 e_{\lambda} e_{m}} - e_{\lambda} e_{m}
\]  

(20)

and finally
\[ V_3 = V_1 + V_1 \left[ \Sigma_\kappa \Sigma_m c_\kappa s_m c'_m \right] V_1 \]

with

\[ s_{km} = (\delta_{km} e_k - c_k V_2 c_m)(c_k V_2 c_m)(e_k e_m)^{-2} \]

(22)

It can be noted that the number of operations required to calculate \( V_3 \) is proportional to \( n^2 \), whereas it goes linearly with \( n \) for \( V_1 \) and \( V_2 \). It is totally out of question to perform this calculation of \( V_3 \) in the real data processing.

Accuracy of phase estimates

In order to compare the theoretical performances of the least-squares estimates, it is convenient to calculate a single figure of merit rather than comparing the full covariance matrices. A natural figure of merit is the expected standard deviation of the phase estimate. Now, we can obtain at least three different phase estimates for any single least-squares solution, namely those of the individual harmonics, e.g.

\[ \hat{\phi}_1 = \text{ATAN2}(\hat{b}_3, \hat{b}_2), \quad \hat{\phi}_2 = \text{iATAN2}(\hat{b}_5, \hat{b}_4) \]

(23)

and a weighted mean of these,

\[ \hat{\phi} = w\hat{\phi}_1 + (1-w)\hat{\phi}_2 \]

(24)

where \( w \) may be chosen to minimize the variance of \( \hat{\phi} \). We call the corresponding standard deviations \( \sigma_1 \), \( \sigma_2 \), and \( \sigma \). Normally, we shall take \( \sigma \) as the relevant figure of merit for the least-squares estimation.

By means of the partial derivative vectors

\[ \hat{f}_1 = \frac{\partial \hat{\phi}_1}{\partial \hat{b}} = \begin{bmatrix} 0 \\ -b_3/(b_2^2 + b_3^2) \\ b_2/(b_2^2 + b_3^2) \\ 0 \\ 0 \end{bmatrix}, \quad \hat{f}_2 = \frac{\partial \hat{\phi}_2}{\partial \hat{b}} = \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{2}b_5/(b_4^2 + b_5^2) \\ \frac{1}{2}b_4/(b_4^2 + b_5^2) \end{bmatrix} \]

(25)

we calculate, for any covariance matrix \( V \), the three scalars
\[ u_{ij} = \frac{f_i V F_j}{\lambda_{ij}}, \quad 1 \leq i \leq j \leq 2 \]  

and obtain

\[ \sigma_1 = \sqrt{u_{11}}, \quad \sigma_2 = \sqrt{u_{22}}, \quad \rho = \frac{u_{12}}{\sqrt{(u_{11}u_{22})}} \]  

(\( \rho \) is the correlation between the two estimates \( \hat{\phi}_1 \) and \( \hat{\phi}_2 \)). For arbitrary \( w \) we have

\[ \sigma^2 = [w \xi_1^2 + (1-w)\xi_2^2] \sqrt{[w \xi_1^2 + (1-w)\xi_2^2]} \]

\[ = w^2 u_{11} + 2w(1-w)u_{12} + (1-w)^2 u_{22} \]  

which is minimized for

\[ w = \frac{(u_{22} - u_{12})}{(u_{11} - 2u_{12} + u_{22})} \]  

yielding

\[ \sigma^2 = \frac{(u_{11}u_{22} - u_{12}^2)}{(u_{11} - 2u_{12} + u_{22})^2} \]  

In general we shall find \( \sigma < \sigma_1 < \sigma_2 \). It should be remarked that the Location Estimator may produce an even better phase estimate than \( \hat{\phi} \), viz., by introducing the a priori knowledge of the ratio \( M_2/M_1 = \sqrt{(b_4^2 + b_5^2)/(b_2^2 + b_3^2)} \) from the OTF calibration. This is not taken into consideration here.

**Numerical results**

The covariances \( V_1 \), \( V_2 \), and \( V_3 \) and corresponding standard deviations were calculated with the following parameters (cf. MAT-HIP-06076):

- slit period (s) = 1.208 as
- scan rate = 168.75 as/s
- sample rate = 1200 Hz
- flux for \( B=9 \) = 2139 Hz
- background flux = 49.3 Hz

The reference phase of the first sample was \( H_1 = 1.7314344 \) rad and increasing with \( \lambda \). The samples in a frame were divided in 16 equal groups, according to the 7.5 Hz IFOV cycling.
Table 1 shows the resulting $\sigma$ (in mas) for simulations with magnitude $B = 6, 9, \text{ and } 12$ and with up to 512 samples per frame.

Table 1. The mean error of the weighted mean of $\hat{\phi}_1$ and $\hat{\phi}_2$.

<table>
<thead>
<tr>
<th>$B$</th>
<th>$n$</th>
<th>$N_s$</th>
<th>$N_b$</th>
<th>$\sigma_{CR}$</th>
<th>$\sigma_{ULS}$</th>
<th>$\sigma_{ML1}$</th>
<th>$\Delta_{ULS}$</th>
<th>$\Delta_{ML1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>128</td>
<td>3616</td>
<td>5.3</td>
<td>5.675</td>
<td>5.841</td>
<td>5.676</td>
<td>2.9%</td>
<td>0.02%</td>
</tr>
<tr>
<td>9.0</td>
<td>16</td>
<td>28.5</td>
<td>0.7</td>
<td>62.50</td>
<td>64.26</td>
<td>63.93</td>
<td>2.8%</td>
<td>2.3%</td>
</tr>
<tr>
<td>9.0</td>
<td>32</td>
<td>57.1</td>
<td>1.3</td>
<td>46.26</td>
<td>47.68</td>
<td>46.84</td>
<td>3.1%</td>
<td>1.3%</td>
</tr>
<tr>
<td>9.0</td>
<td>64</td>
<td>114</td>
<td>2.6</td>
<td>32.23</td>
<td>33.14</td>
<td>32.42</td>
<td>2.8%</td>
<td>0.6%</td>
</tr>
<tr>
<td>9.0</td>
<td>128</td>
<td>228</td>
<td>5.3</td>
<td>22.79</td>
<td>23.43</td>
<td>22.86</td>
<td>2.8%</td>
<td>0.3%</td>
</tr>
<tr>
<td>9.0</td>
<td>512</td>
<td>913</td>
<td>21</td>
<td>11.42</td>
<td>11.74</td>
<td>11.43</td>
<td>2.8%</td>
<td>0.08%</td>
</tr>
<tr>
<td>12.0</td>
<td>128</td>
<td>14.4</td>
<td>5.3</td>
<td>103.96</td>
<td>105.80</td>
<td>107.31</td>
<td>1.8%</td>
<td>3.2%</td>
</tr>
<tr>
<td>12.0</td>
<td>512</td>
<td>57.6</td>
<td>21</td>
<td>52.16</td>
<td>53.09</td>
<td>52.59</td>
<td>1.8%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

Notations: $B$ = $B$ magnitude ($B-V = 0.5$), $n$ = number of samples in a frame, $N_s$ = total number of stellar photons collected in the frame, $N_b$ = total number of background photons collected in frame, $\sigma_{CR}$ = mean error of $\hat{\phi}$ according to Cramér-Rao bound [mas], $\sigma_{ULS}$ = mean error of $\hat{\phi}$ for unweighted least-squares, $\sigma_{ML1}$ = mean error of $\hat{\phi}$ for ULS followed by one ML iteration, $\Delta_{ULS} = \sigma_{ULS}/\sigma_{CR} - 1$ = relative loss of ULS w.r.t. CR, $\Delta_{ML1} = \sigma_{ML1}/\sigma_{CR} - 1$ = relative loss of ML1 w.r.t. CR.

Table 2 shows a comparison of $\sigma$, $\sigma_1$, and $\sigma_2$ for three representative runs, and also the correlation between $\hat{\phi}_1$ and $\hat{\phi}_2$ and the optimum weighting of the two harmonics.

Table 2. $\sigma$, $\sigma_1$, and $\sigma_2$ together with the correlation (\rho) and weight factor (w) of the first harmonic.

<table>
<thead>
<tr>
<th>$B$</th>
<th>$n$</th>
<th>$\sigma$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\rho$</th>
<th>$w$</th>
<th>estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.0</td>
<td>128</td>
<td>5.675</td>
<td>6.085</td>
<td>8.559</td>
<td>0.348</td>
<td>0.745</td>
<td>CR</td>
</tr>
<tr>
<td>9.0</td>
<td>128</td>
<td>22.79</td>
<td>24.45</td>
<td>34.40</td>
<td>0.339</td>
<td>0.742</td>
<td>CR</td>
</tr>
<tr>
<td>12.0</td>
<td>128</td>
<td>103.96</td>
<td>114.92</td>
<td>162.63</td>
<td>0.251</td>
<td>0.719</td>
<td>CR</td>
</tr>
<tr>
<td>6.0</td>
<td>128</td>
<td>5.841</td>
<td>6.147</td>
<td>9.126</td>
<td>0.369</td>
<td>0.785</td>
<td>ULS</td>
</tr>
<tr>
<td>9.0</td>
<td>128</td>
<td>23.43</td>
<td>24.73</td>
<td>36.65</td>
<td>0.359</td>
<td>0.780</td>
<td>ULS</td>
</tr>
<tr>
<td>12.0</td>
<td>128</td>
<td>105.80</td>
<td>115.48</td>
<td>168.24</td>
<td>0.264</td>
<td>0.739</td>
<td>ULS</td>
</tr>
</tbody>
</table>
Remarks

1. \( \rho \) and \( w \) are, for CR and ULS, independent of \( n \) at a given ratio \( N_b/N_s \).

2. The optimum weight factor \( (w) \) is slightly larger than the one adopted by MATRA, \( M_1^2/(M_1^2 + 4M_2^2) = 0.659 \). The increase in \( \sigma \) when using MATRA's value instead of \( w \) may be about 1% in typical cases.

3. E26-HIP-06412 (issue 1, p. 135) gives \( \sigma_{CR}^2 = 85.56/(0\tau) + 3.06/(0^2\tau) \text{ mas}^2 \) for the photon noise + background effects. This is within 1% in agreement with Table 1.

4. From Table 2 we have empirically

\[
\Delta_{ULS} \approx 0.029(1 + 1.8 \frac{N_b}{N_s})^{-1} \quad (31a)
\]

\[
\Delta_{ML1} \approx 0.74 \frac{N_s}{N_s}^{-1}(1 + 1.6 \frac{N_b}{N_s})^{-1} \quad (31b)
\]

This would mean that ULS actually performs better than ML1 for \( N_s \lesssim 27 \).

Conclusions

Using the unweighted least-squares estimate (ULS), the mean error of the phase determination is always within 2.9% of the Cramér-Rao (CR) bound. At first sight it would perhaps seem that we can do without the likelihood iteration and accept this loss of accuracy. However, we still need to calculate the covariance of the estimate, i.e. (17) for the ULS. The amount of computation involved in that process is almost the same as for a single likelihood iteration, which almost automatically gives the covariance also (at least the CR covariance). On the other hand, it appears that a single likelihood iteration is sufficient.

In summary, I recommend the following procedure:

A. Make an initial ULS estimate by accumulating \( \mathbf{CC}' \) and \( \mathbf{Cn} \) and solving \( \hat{\mathbf{b}}_2 \):

\[
[ \mathbf{CC}' \quad \mathbf{Cn} ] \rightarrow [ \ldots \hat{\mathbf{b}}_2 ] \quad (32)
\]

[(32) can be performed with the subroutine CHOL (Appendix), using redl = redr = sol = .TRUE..]

B. Calculate the estimated counts, \( \hat{c} = \mathbf{C}'\hat{\mathbf{b}}_2 \), while accumulating the weighted normals \( [\mathbf{C}\mathbf{E}^{-1}\mathbf{C}' \quad \mathbf{C}\mathbf{E}^{-1}\mathbf{n}] \) and sum-squared right-hand side \( (\mathbf{n}'\mathbf{E}^{-1}\mathbf{n}) \); then solve
for the final (ML1) estimate $\hat{\beta}_3$, invert the Cholesky factor ($R$), and compute the chi-square according to

$$
\begin{align*}
[ R'R = CE^{-1}C' \quad d = CE^{-1}n ] & \rightarrow [ R \quad h = R'^{-1}d ] & \rightarrow [ R^{-1} \quad \hat{\beta}_3 = R^{-1}h ]
\end{align*}
$$

(33)

$$
\chi^2 = n' E^{-1} n - h'h
$$

(34)

[ (33) can be performed by the subsequent calls

CALL CHOL(5, 5, 6, .TRUE., .TRUE., .FALSE., .FALSE., a, ierr)

CALL CHOL(5, 5, 6, .FALSE., .FALSE., .TRUE., .FALSE., a, ierr)

with the calculation of $\chi^2$ in between.]

C. Convert the components of $\hat{\beta}_3$ to the required signal parameters:

$$
\begin{align*}
\hat{\beta}_1 &= b_1 \\
\hat{\beta}_2 &= (b_2^2 + b_3^2)^{1/2} \\
\hat{\beta}_3 &= \text{ATAN2}(-b_3, b_2) \\
\hat{\beta}_4 &= \frac{(b_2^2 - b_3^2)b_4 + 2b_2b_3b_5}{\beta_2^2} \\
\hat{\beta}_5 &= \frac{(b_2^2 - b_3^2)b_5 - 2b_2b_3b_4}{\beta_2^2}
\end{align*}
$$

(35)

and calculate the covariance of $\hat{\beta}$ by means of the transformation

$$
\Phi^{-1} \propto \frac{\partial \Phi}{\partial \beta} \left( \begin{array}{c}
\Phi' - \frac{\partial P}{\partial \beta}
\end{array} \right) = P(CE^{-1}C')^{-1}P' = P(R'R)^{-1}P' = PR^{-1}(PR^{-1})',
$$

(36)

where $P$ is the matrix of partial derivatives, $P_{ij} = \partial \beta_i / \partial \beta_j$, with

$$
\begin{align*}
P_{11} &= \delta_{11} \\
P_{22} &= b_2/\beta_2, \quad P_{23} = b_3/\beta_2, \quad P_{24} = P_{25} = 0 \\
P_{32} &= b_3/\beta_2, \quad P_{33} = -b_2/\beta_2, \quad P_{34} = P_{35} = 0 \\
P_{42} &= -(2b_3b_5 + b_2b_4)/\beta_2^2, \quad P_{43} = (2b_2b_5 - b_3b_4)/\beta_2^2, \quad P_{44} = (b_2^2 - b_3^2)/\beta_2^3 \\
P_{45} &= 2b_2b_3/\beta_2^3 \\
P_{52} &= (2b_3b_4 - b_2b_5)/\beta_2^2, \quad P_{53} = -(2b_2b_4 + b_3b_5)/\beta_2^2, \quad P_{54} = -P_{45}, \quad P_{55} = P_{44}
\end{align*}
$$

(37)
SUBROUTINE choi(mdim, n, redr, sol, inv, ierr)
  c Cholesky reduction etc of data array a(min) (n, LE, n)
  c Input arguments:
  c ndim = first dimension of array a(min)
  c n = number of unknowns = number of rows
  c m = number of right-hand sides (rhs's) = number of columns
  c redr = .TRUE. to reduce left-hand side (lhs = normal eqns matrix)
  c sol = .TRUE. to solve rhs's (after redr)
  c inv = .TRUE. to invert lhs (after redr)
  c a = data array
  c Output arguments:
  c a = modified data array
  c ierr = 0 for normal return (PD lhs), <0 for non-PD lhs
  c L Lindegaard 1984 May 17

IMPLICIT DOUBLE PRECISION (a-h, o-z)
DIMENSION a(min)
LOGICAL redr, sol, inv
m = n+1
ierr = 0
IF (redr) THEN
  DO 130 i = 1, n
     DO 120 j = 1, i
        s = 0.0D0
        DO 110 k = 1, j-1
           s = s + a(k,i)*a(k,j)
        CONTINUE
        110 IF (j .LT. i) THEN
          a(j,i) = (a(j,i) - s)*a(j,j)
        ELSE
          s1 = a(i,i) - s
          IF (s1 .LE. 0.0D0) THEN
            ierr = -1
            RETURN
          ENDIF
          a(i,i) = 1.0D0/dsqr(s1)
        ENDIF
     CONTINUE
  CONTINUE
ENDIF
  120 CONTINUE
  130 CONTINUE
ENDIF
IF (sol) THEN
  DO 330 i = 1, n
     DO 320 j = 1, i-1
        s = a(j,i)*a(j,j)
        a(j,i) = s
     CONTINUE
  CONTINUE
  320 CONTINUE
  330 CONTINUE
ENDIF
IF (inv) THEN
  DO 440 i = 1, 2*n-1
     s = a(i,i)
     DO 430 k = 1, i-1
        a(k,i) = -a(k,i)*s
     CONTINUE
     430 CONTINUE
     DO 440 j = i-1, 2*n-1
        s = a(j,i)*a(j,j)
        a(j,i) = s
     CONTINUE
  CONTINUE
  440 CONTINUE
ENDIF
RETURN
END