Astrometric Errors Correlated Strongly Across Multiple SIRTF Images

John Fowler
28 March, 2003

The possibility exists that after pointing transfer has been performed for each BCD (i.e., a calibrated image, also called a "frame") in an AOR (a collection of related images), the pointing uncertainties may be dominated by a systematic error due to a slowly varying misalignment between the star sensor and the telescope systems. This dominant systematic error would produce a strong correlation between the pointing reconstruction errors for any two BCDs. This would be important for pointing refinement when absolute astrometric references are not employed, e.g., in the longer-wavelength MIPS channels and probably MIPS24 as well. The current design of the pointing refinement software assumes that the BCD pointing uncertainties are uncorrelated and uses them accordingly to derive weights for part of the cost function minimized in the global solution for image frame coordinate corrections. A large systematic error component would effectively reduce the frame weights to zero, leaving the solution vulnerable to noisy point-source data. This paper examines the impact of strongly correlated errors in the frame coordinates and suggests simple enhancements to the software to accommodate them.

In the general case, the coordinate errors in right ascension and declination can be written

$$\varepsilon_\alpha = \varepsilon_{\alpha r} + \varepsilon_{\alpha s} \quad , \quad \varepsilon_\delta = \varepsilon_{\delta r} + \varepsilon_{\delta s}$$

where *r* indicates "random" and *s* indicates "systematic". We will treat the systematic error as a constant but unknown error whose size is a single random draw from a parent distribution of such errors, i.e., although its value is a constant for all frames in an AOR, our knowledge of it is a random variable characterized by a probability density function with a standard deviation that we will denote $\sigma_{\alpha s}$ for right ascension and $\sigma_{\delta s}$ for declination. Similar remarks apply to the "random" errors, except that they are uncorrelated over all frames in an AOR.

To summarize: once a measurement has been made, both "random" errors and "systematic" errors are unknown constants; the difference is that the "systematic" errors are expected to have the same unknown value on each measurement, whereas the unknown values of the "random" errors are expected to have no correlation from one measurement to another. In both cases, the only knowledge we have concerning the values that can occur is statistical in nature, characterized by probability density functions in the same way for both types. This statistical characterization is obtained from error models that are determined *a priori*. This is the source of values for the standard deviations used below.

As usual in error analysis, we are interested in the expectation values of the three pairwise products of the errors on the left-hand sides of the equations above.

$$\varepsilon_\alpha \varepsilon_\alpha = \varepsilon_{\alpha r}^2 + 2\,\varepsilon_{\alpha r}\,\varepsilon_{\alpha s} + \varepsilon_{\alpha s}^2$$
$$\varepsilon_\delta \varepsilon_\delta = \varepsilon_{\delta r}^2 + 2\,\varepsilon_{\delta r}\,\varepsilon_{\delta s} + \varepsilon_{\delta s}^2$$
$$\varepsilon_\alpha \varepsilon_\delta = \varepsilon_{\alpha r}\,\varepsilon_{\delta r} + \varepsilon_{\alpha r}\,\varepsilon_{\delta s} + \varepsilon_{\alpha s}\,\varepsilon_{\delta r} + \varepsilon_{\alpha s}\,\varepsilon_{\delta s}$$
$$v_\alpha = \left\langle \varepsilon_\alpha \varepsilon_\alpha \right\rangle$$
$$v_\delta = \left\langle \varepsilon_\delta \varepsilon_\delta \right\rangle$$
$$v_{\alpha\delta} = \left\langle \varepsilon_\alpha \varepsilon_\delta \right\rangle$$

All error distributions are assumed herein to be zero-mean. Since actual values of systematic errors are assumed to be constant, we have

$$\left\langle \varepsilon_{\alpha r}\,\varepsilon_{\alpha s} \right\rangle = 0$$
$$\left\langle \varepsilon_{\delta r}\,\varepsilon_{\delta s} \right\rangle = 0$$
$$\left\langle \varepsilon_{\alpha r}\,\varepsilon_{\delta s} \right\rangle = 0$$
$$\left\langle \varepsilon_{\alpha s}\,\varepsilon_{\delta r} \right\rangle = 0$$

so that

$$v_\alpha \equiv \left\langle \varepsilon_{\alpha r}^2 \right\rangle + \left\langle \varepsilon_{\alpha s}^2 \right\rangle = \sigma_{\alpha r}^2 + \sigma_{\alpha s}^2 \equiv v_{\alpha r} + v_{\alpha s}$$
$$v_\delta \equiv \left\langle \varepsilon_{\delta r}^2 \right\rangle + \left\langle \varepsilon_{\delta s}^2 \right\rangle = \sigma_{\delta r}^2 + \sigma_{\delta s}^2 \equiv v_{\delta r} + v_{\delta s}$$
$$v_{\alpha\delta} \equiv \left\langle \varepsilon_{\alpha r}\,\varepsilon_{\delta r} \right\rangle + \left\langle \varepsilon_{\alpha s}\,\varepsilon_{\delta s} \right\rangle \equiv v_{\alpha\delta r} + v_{\alpha\delta s}$$

These are the quantities that are assumed to be known from the *a priori* error model for each image. In general, the values would not be expected to be exactly the same for each frame in an AOR, but there is no apparent reason for them to vary greatly either, so we will consider the case in which they are the same for each frame.

The full error covariance matrix is a $2 \times 2 \times N \times N$ matrix, where N is the number of frames in the AOR. The covariance of right ascension with declination tends to be weak, however, and is ignored in the pointing refinement processing. If we ignore it here, a considerable simplification results, since the matrix dimensions become just $2 \times N \times N$, and furthermore the lack of any coupling between right ascension and declination permits this to be treated as two separate $N \times N$ covariance matrices. There is no formal difference in the way these two matrices are used, so we will confine the discussion to the right ascension error covariance matrix. For any two frames, which we will indicate with subscripts 1 and 2, the errors in right ascension are

$$\varepsilon_{\alpha 1} = \varepsilon_{\alpha r 1} + \varepsilon_{\alpha s 1}$$
$$\varepsilon_{\alpha 2} = \varepsilon_{\alpha r 2} + \varepsilon_{\alpha s 2}$$

The expectation value for the product of the two errors is

$$\left\langle \varepsilon_{\alpha 1}\ \varepsilon_{\alpha 2} \right\rangle = \left\langle \varepsilon_{\alpha r 1}\ \varepsilon_{\alpha r 2} \right\rangle + \left\langle \varepsilon_{\alpha r 1}\ \varepsilon_{\alpha s 2} \right\rangle + \left\langle \varepsilon_{\alpha s 1}\ \varepsilon_{\alpha r 2} \right\rangle + \left\langle \varepsilon_{\alpha s 1}\ \varepsilon_{\alpha r 2} \right\rangle$$

Since

$$\left\langle \varepsilon_{\alpha r 1}\ \varepsilon_{\alpha r 2} \right\rangle = \left\langle \varepsilon_{\alpha r 1}\ \varepsilon_{\alpha s 2} \right\rangle = \left\langle \varepsilon_{\alpha s 1}\ \varepsilon_{\alpha r 2} \right\rangle = 0$$

we have

$$\left\langle \varepsilon_{\alpha 1}\ \varepsilon_{\alpha 2} \right\rangle = \left\langle \varepsilon_{\alpha s 1}\ \varepsilon_{\alpha s 2} \right\rangle$$

and since we are considering the case in which the error distributions are the same for each frame, this becomes

$$\left\langle \varepsilon_{\alpha 1}\ \varepsilon_{\alpha 2} \right\rangle = \left\langle \varepsilon_{\alpha s 1}\ \varepsilon_{\alpha s 1} \right\rangle = v_{\alpha s}$$

So the off-diagonal element of the error covariance matrix for frames 1 and 2 is

$$v_{\alpha 12} = v_{\alpha s}$$

The general form of the right-ascension error covariance matrix for *N* frames is

$$\Omega_{\alpha} = \begin{pmatrix} v_{\alpha 11} & v_{\alpha 12} & v_{\alpha 13} & \cdots & v_{\alpha 1N} \\ v_{\alpha 21} & v_{\alpha 22} & v_{\alpha 23} & \cdots & v_{\alpha 2N} \\ v_{\alpha 31} & v_{\alpha 32} & v_{\alpha 33} & \cdots & v_{\alpha 3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{\alpha N1} & v_{\alpha N2} & v_{\alpha N3} & \cdots & v_{\alpha NN} \end{pmatrix}$$

where

$$v_{\alpha ii} = v_{\alpha r} + v_{\alpha s}$$
$$v_{\alpha i \neq j} = v_{\alpha s}$$

We will make explicit use of the fact that all error covariance matrices are symmetric. Furthermore, we are considering the case in which the diagonal elements are all equal (i.e., the total error in right ascension has the same distribution for each frame) and the off-diagonal elements are all equal (i.e., the same systematic error is common to any pair of frames). Suppressing the subscript alpha and using these symmetries, the error covariance matrix becomes

$$\Omega = \begin{pmatrix} v_{11} & v_{12} & v_{12} & \cdots & v_{12} \\ v_{12} & v_{11} & v_{12} & \cdots & v_{12} \\ v_{12} & v_{12} & v_{11} & \cdots & v_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{12} & v_{12} & v_{12} & \cdots & v_{11} \end{pmatrix}$$

Typical values for the standard deviations of the random and systematic errors are expected to be 1 and 5 arcseconds, respectively. For example, in the case $N = 2$, this yields

$$\Omega = \begin{pmatrix} 26 & 25 \\ 25 & 26 \end{pmatrix}$$

The eigenvalues of this matrix are 51 and 1, and therefore the one-sigma errors in the diagonalized form are 7.14 and 1.0 arcseconds, and the principal axes are rotated 45 degrees (the eigenvectors are [1,1] and [-1,1], respectively). Note that the smaller eigenvalue is just the original random error. For the case $N = 3$ with the same error values,

$$\Omega = \begin{pmatrix} 26 & 25 & 25 \\ 25 & 26 & 25 \\ 25 & 25 & 26 \end{pmatrix}$$

Now the eigenvalues are 76, 1, and 1, so the one-sigma values corresponding to the diagonalized form are 8.72, 1.0, and 1.0 arcseconds. Investigating the cases $N = 4$, $N = 5$, etc., for various values of "random" and "systematic" error variances reveals that one always obtains one large eigenvalue and N-1 degenerate eigenvalues given respectively by

$$\lambda_1 = v_{11} + (N - 1)\, v_{12}$$
$$\lambda_2 = v_{11} - v_{12}$$

This can be shown to be a property of nonsingular square matrices with equal diagonal elements and equal off-diagonal elements (see Appendix B). The eigenvector associated with the larger eigenvalue is always an N-vector whose components are all 1, and the N-1 eigenvectors associated with the smaller eigenvalues are N-vectors whose first component is -1 and remaining components form all permutations of zeros and a single 1. In terms of the right-ascension errors,

$$\lambda_1 = v_{\alpha r} + N\, v_{\alpha s}$$
$$\lambda_2 = v_{\alpha r}$$

It may be challenging to the intuition to interpret the space in which the error covariance matrix is diagonal. When considering correlated errors in right ascension and declination, the interpretation of the diagonalized space is straightforward: the basic nature of the space is unchanged; it is just a two-dimensional position space whose coordinates are merely rotated so that their axes align with the principal axes of the error ellipse. Thus the parameters whose coordinate axes define the rotated space are linear combinations of the right ascension and declination parameters (technically, this is true only in the Cartesian approximation, i.e., the small-angle approximation, which is of interest here; nonlinear mappings are needed in general for spherical-coordinate mappings). In the case of the $N \times N$ error covariance matrix above, each parameter contributing to the definition of the "space" is a right-ascension error parameter *belonging to a specific measurement*. It may seem that there is only one right ascension error parameter, but it is common in statistics to treat a given physical parameter involved in two measurements as two separate parameters; a closely related example is the construction of joint density functions, wherein N samples of a random variable defined on the same domain are treated as N separate parameters.

Thus the space implicit in the error covariance matrix is more abstract than ordinary position-coordinate spaces, but it still obeys the same mathematical rules. Specifically, there exists a single rotation (whose axis is generally not one of the coordinate axes) that yields a new space whose axes are aligned with the principal axes of the N-dimensional error ellipsoid. The parameters corresponding to these rotated axes are linear combinations of the original axes; that is, a "measurement" on any one such axis is a linear combination of measurements on *all* the original axes, arranged in such a way that the "errors" in these new "measurements" are all *independent*.

The above remarks show that computations requiring independent frame errors can be made by transforming coordinates to the space in which all errors are independent, making whatever computations are desired, and then transforming the results back to the original space. The fact that the intermediate space has a somewhat obscure interpretation plays no role in the procedure.

But the specific symmetry properties of the error covariance matrix considered herein provide a simpler approach. All but one of the of the independent-error variances are the same as the "random" error variance in the original space. Thus the N measurements in the rotated space are comprised of one with a rather large uncertainty (the variance is the random-error variance plus N times the systematic-error variance) and N-1 measurements with relatively small uncertainties (the random-error variance alone). This is a manifestation of the fact that the N original measurements have rather small uncertainties in their positions relative to each other, but a rather large uncertainty in how the entire assemblage is positioned on the sky. Since pointing refinement in the absence of absolute astrometric references is concerned only with uncertainties in the frame positions relative to each other, this suggests using only the random-error variances in the inverse-variance weighting employed in the cost function that is globally minimized. Such an approach is intuitively appealing, and the fact that the random-error variances turn out to survive the diagonalization to become all but one of the eigenvalues lends support to this idea.

Another way of stating this is as follows. The rigorous weighting of the frame position measurements would involve the full use of the error covariance matrix; but this is equivalent to transforming coordinates to the diagonalized system and evaluating this portion of the cost function therein. Without astrometric references, one of the SIRTF frames is chosen as the fiducial frame to which the others are adjusted, and this can be chosen to correspond to the measurement in the rotated system that has the single large error variance; all other frame-to-frame uncertainties depend only on the relatively small random-error uncertainty, which is numerically the same in the original system. So the corresponding weighting might as well be done in the original system without either using the entire error covariance matrix (which requires inverting it, although this is easy for the special form we are considering, as shown in Appendix A) or diagonalizing it (which is also easy, as shown in Appendix B).

The following suggestion is therefore made: when pointing refinement is performed without absolute astrometric references, the weights assigned to the frame positions themselves in the cost function should be inverse variances computed only from the random-error portion of the total uncertainty.

The remaining problem is that currently the error covariance matrix computed for the position of a frame is not broken down into random and systematic components. The only significant systematic error stems from the telescope boresight uncertainty, and this is not provided in such terms. The only workaround is to estimate the random portion of the uncertainty from the observed dispersion in the celestial coordinates, SIGRA and SIGDEC in the FITS headers. These could be used *with suitable lower limits* to provide a much better approximation to the weighting than simply using the stated uncertainty as though it were uncorrelated frame-to-frame, which would in effect delete the frame position information from the cost function and leave the solution vulnerable to sparse and/or low S/N point source data. The lower limits should be enforced to avoid unrealistically low estimates due to statistical fluctuations. No upper limit is needed, since if the observed dispersion is high, the uncertainty really is high. The use of SIGRA and SIGDEC is also preferable to any attempt to back out the random component of error by "un-RSSing" a standard estimate of the systematic component, since the total error is strongly dominated by the component being removed, and therefore the residual must be considered unacceptably inaccurate.

It should be noted that SIGRA and SIGDEC are generally not the same in every frame of an AOR, although there is no reason to expect them to fluctuate greatly. Nevertheless, the great algebraic simplifications due to the symmetries involved in having only two distinct numerical values in the covariance matrix are lost with even the slightest variation among the diagonal elements or the off-diagonal elements. Numerical studies show that for small fluctuations, however, the results are only marginally affected, as one would expect. Conclusions drawn from the simplified algebra, therefore, are not significantly altered, and there is no need to employ, e.g., AOR-averaged values of SIGRA and SIGDEC. The small fluctuations that do exist should be permitted to influence the solution, i.e., slightly better measurements should get a slightly greater vote in the outcome.

Appendix A. Inversion of the Special-Case Error Covariance Matrix

For reasons discussed in the main text, it is not necessary to invert the error covariance matrix discussed therein, but deriving the formulas needed is instructive and useful in computing the eigenvalues (see Appendix B). The special-case form of the $N \times N$ matrix is

$$\Omega = \begin{pmatrix} v_{11} & v_{12} & v_{12} & \cdots & v_{12} \\ v_{12} & v_{11} & v_{12} & \cdots & v_{12} \\ v_{12} & v_{12} & v_{11} & \cdots & v_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{12} & v_{12} & v_{12} & \cdots & v_{11} \end{pmatrix}$$

Symmetric matrices have symmetric inverses, and in this case, since the symmetries extend along the diagonal and across all off-diagonal elements, the inverse matrix must have the form

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{12} & \cdots & w_{12} \\ w_{12} & w_{11} & w_{12} & \cdots & w_{12} \\ w_{12} & w_{12} & w_{11} & \cdots & w_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{12} & w_{12} & w_{12} & \cdots & w_{11} \end{pmatrix}$$

where

$$W\Omega = I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

The matrix multiplication involves the following summations for the first two elements on the first row.

$$I_{11} = \sum_{j=1}^{N} w_{1j} v_{j1} = w_{11} v_{11} + (N-1) w_{12} v_{12} = 1$$

$$I_{12} = \sum_{j=1}^{N} w_{2j} v_{j2} = w_{11} v_{12} + w_{12} v_{11} + (N-2) w_{12} v_{12} = 0$$

where we have used

$$v_{ii} = v_{11}\ ,\ v_{ij} = v_{ji}\ ,\ v_{i \neq j} = v_{12}$$

$$w_{ii} = w_{11}\ ,\ w_{ij} = w_{ji}\ ,\ w_{i \neq j} = w_{12}$$

With the definitions

$$C_1 \equiv (N - 1)\, v_{12}$$
$$C_2 \equiv v_{11} + (N - 2)\, v_{12}$$

We can use the expressions for I_{11} and I_{12} to form the 2×2 simultaneous linear system

$$v_{11}\, w_{11} + C_1\, w_{12} = 1$$
$$v_{12}\, w_{11} + C_2\, w_{12} = 0$$

which has the solutions

$$w_{12} = \frac{v_{12}}{C_1\, v_{12} - C_2\, v_{11}}$$
$$w_{11} = \frac{-C_2\, w_{12}}{v_{12}}$$

Substituting for C_1 and C_2 and regrouping,

$$w_{12} = \frac{v_{12}}{(v_{12} - v_{11})(v_{12}\,(N - 1) + v_{11})}$$
$$w_{11} = \frac{v_{11} + (N - 2)\, v_{12}}{(v_{11} - v_{12})(v_{12}\,(N - 1) + v_{11})}$$

This supplies all elements of the inverse matrix.

Appendix B. Eigenvalues of the Special-Case Error Covariance Matrix

The eigenvalues of the special-case error covariance matrix are the solutions for λ in the equation

$$\begin{pmatrix} v_{11}-\lambda & v_{12} & v_{12} & \cdots & v_{12} \\ v_{12} & v_{11}-\lambda & v_{12} & \cdots & v_{12} \\ v_{12} & v_{12} & v_{11}-\lambda & \cdots & v_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{12} & v_{12} & v_{12} & \cdots & v_{11}-\lambda \end{pmatrix} = 0$$

which can also be written

$$\det(\Omega - \lambda I) = 0$$

where "det" refers to the determinant of the argument. Evaluation of the determinant for an $N \times N$ matrix yields an N^{th}-order polynomial (called the characteristic polynomial) whose zeros are the desired eigenvalues. For $N = 2$ and $N = 3$, the algebraic representations of the determinants are relatively straightforward, and the characteristic equations are

$$N = 2: \; v_{11}^2 - v_{12}^2 - 2\,v_{11}\,\lambda + \lambda^2 = 0$$
$$N = 3: \; v_{11}^3 - 3\,v_{11}\,v_{12}^2 + 2\,v_{12}^3 - 3\,(v_{11}^2 - v_{12}^2)\,\lambda + 3\,v_{11}\,\lambda^2 - \lambda^3 = 0$$

The solutions can be shown to be

$$N = 2: \; \lambda_1 = v_{11} + v_{12}\,,\; \lambda_2 = v_{11} - v_{12}$$
$$N = 3: \; \lambda_1 = v_{11} + 2\,v_{12}\,,\; \lambda_2 = v_{11} - v_{12}\,,\; \lambda_3 = v_{11} - v_{12}$$

For $N > 3$, the algebraic representation of the determinant becomes progressively more nontrivial, and so we will pursue the more heuristic approach of showing that the solutions we can guess from the above continue to hold for all N. We can see from the two cases above that a pattern has begun in which a single eigenvalue exists with a value of $v_{11} + (N\text{-}1)v_{12}$ and $(N\text{-}1)$ degenerate eigenvalues equal to $v_{11}\text{-}v_{12}$. This is supported by numerical solution for higher N (cases up to 7 were checked). Furthermore, since the matrix is composed of only two distinct numerical values, it makes sense that no more than two distinct eigenvalues could ever exist.

We will use the following properties of matrix eigenvalues, where W is the inverse of Ω (see Appendix A):

$$Tr(\Omega) = \sum_{i=1}^{N} \lambda_i$$

$$Tr(W) = \sum_{i=1}^{N} \frac{1}{\lambda_i}$$

where we have used the fact that the diagonal elements of the inverse of a diagonal matrix are the inverse of the elements on the diagonal of the original matrix, and "Tr" indicates the trace of the matrix specified as an argument, i.e., the sum of the diagonal elements, so that we also have:

$$Tr(\Omega) = \sum_{i=1}^{N} v_{11} = N\, v_{11}$$

$$Tr(W) = \sum_{i=1}^{N} w_{11} = N\, w_{11}$$

Given our hypothesis that only two distinct eigenvalues exist with the values and degeneracy discussed above, these equations become

$$\lambda_1 + (N-1)\,\lambda_2 = N\, v_{11}$$

$$\frac{1}{\lambda_1} + \frac{N-1}{\lambda_2} = N\, w_{11}$$

We will work first with the second equation; Appendix A showed that

$$w_{11} = \frac{v_{11} + (N-2)\, v_{12}}{(v_{11} - v_{12})\,(v_{12}\,(N-1) + v_{11})}$$

and so

$$\frac{1}{\lambda_1} + \frac{N-1}{\lambda_2} = N\, \frac{v_{11} + (N-2)\, v_{12}}{(v_{11} - v_{12})\,(v_{12}\,(N-1) + v_{11})}$$

The right-hand side can be rearranged to obtain

$$\frac{1}{\lambda_1} + \frac{N-1}{\lambda_2} = \frac{1}{v_{11} + (N-1)\, v_{12}} + \frac{N-1}{v_{11} - v_{12}}$$

which is clearly satisfied by

$$\lambda_1 = v_{11} + (N-1)\, v_{12}$$

$$\lambda_2 = v_{11} - v_{12}$$

All that remains is to verify that these solutions satisfy the other trace equation:

$$\lambda_1 + (N - 1)\lambda_2 = N v_{11}$$

Plugging in the provisional eigenvalue expressions,

$$\begin{aligned}
\lambda_1 + (N - 1)\lambda_2 &= [v_{11} + (N - 1) v_{12}] + (N - 1)[(v_{11} - v_{12})] \\
&= v_{11} + N v_{12} - v_{12} + N v_{11} - v_{11} - N v_{12} + v_{12} \\
&= v_{11} - v_{11} + v_{12} - v_{12} + N v_{12} - N v_{12} + N v_{11} \\
&= N v_{11}
\end{aligned}$$

So both trace equations are satisfied by the eigenvalue expressions, which is what we wished to demonstrate.

The author would like to thank D. Henderson for suggesting the use of the trace equations to show that the eigenvalues have the form suggested by the numerical calculations and the formal cases for $N = 2$ and $N = 3$.