Algorithms for Isolating Worst Case Systematic Data Errors

David W. Curkendall*
Jet Propulsion Laboratory, Pasadena, Calif.

Algorithms which are computationally simple and easy to apply and interpret are derived to investigate filter sensitivity to data error sets which have the most power, for a given norm, to create estimate error. These techniques are applied to a baseline estimation problem using very long baseline interferometry (VLBI) data. The VLBI technique itself is reviewed and the worse case error finding algorithms are motivated by noting that VLBI is a technique that permits great observational flexibility in order to desensitize the estimator to at least some sequences of data error. Finding the worst case data error sets thus has the twin utilities of 1) specifying what are the otherwise very elusive performance bounds, and 2) suggesting observational strategy modification to build an overall more robust situation.

Introduction

MOST estimation algorithms, regardless of the specific application, are formulated using an additive white, or perhaps exponentially correlated, noise model for the assumed errors corrupting the data. Where appropriate, similar models are used to account for random corruption of the state variables as well. A description of the expected estimation errors, given that the real noise models are congruent to those postulated, is inherent in the filter computation itself; in the simplest of circumstances the usual $(A^TWA)^{-1}$ computation gives the covariance of these errors.[1,2] In testing practical algorithms for susceptibility to unmodeled effects, it is common practice, for example, to compute the error sensitivity between an estimated parameter and the desired state variables.[3] Techniques for investigating the effects of measurement noise that does not behave in accordance with the postulated model exist,[3] but they are rarely applied. The application rarity comes because of two basic problems: 1) existing techniques are computationally rather involved, and 2) the alternative noise model is a rather vague undertaking with which no one seems to know just how to proceed. However, in many applications, the knowledge is almost certain that systematic error in the data space is what really limits the estimation error—all in all a very unsatisfactory set of affairs.

Absolute Worst Case Analysis

In approaching the problem of isolating data error signatures to which an estimation process is most sensitive, it is natural to seek a data error set that is small in some sense yet can produce a significant solution error. One such approach would be to minimize the mean square of the data error sequence to produce, say, a unit error in a selected component of the solution.

The following formal problem is posed: Identify the n-dimensional data error sequence $\xi_1, \xi_2, \ldots, \xi_n$, having minimum Euclidean norm, which can produce, say, a unit error in a selected component of the estimate $\hat{x}_i$. In the linear model

$$z = Ax + \xi \tag{1}$$

where

z = the $n \times 1$ observation vector
x = the $m \times 1$ parameter vector, to be estimated
A = the $n \times m$ matrix of observation partials
$\xi^T = [\xi_1, \xi_2, \ldots, \xi_n]$, the observation error

Depending upon the measurement noise model being postulated, a weighting matrix W is constructed so that

$$E[\xi \ \xi^T] = W^{-1}$$

where $E[\ \ ]$ is the expected value operator. The parameter vector x is estimated via the usual normal equation approach

$$\hat{x} = (A^TWA)^{-1}A^TWz$$

The error in $\hat{x}$,

$$\hat{x} - x = F\xi$$

where

$$F = (A^TWA)^{-1}A^TW = \begin{bmatrix} f_1^T \\ \vdots \\ f_i^T \\ \vdots \\ f_n^T \end{bmatrix}$$

From the Cauchy-Schwartz inequality, which states that the dot product of two vectors of fixed length is maximized when the two vectors are colinear, it directly follows that

$$\xi^* = kf_i$$

is the data error vector which can most efficiently produce estimate error in $\hat{x}_i$. In terms of the original problem of determining the minimum Euclidean-length ξ which can produce unit error in $\hat{x}_i$,

$$\xi^* = f_i / f_i^T f_i \tag{2}$$

Comments

This is a general result for any least-squares problem of the type posed in Eq. (1). In the special case of equally weighted data ($W^{-1} = \sigma_d^2 I$ where σ_d is the standard deviation of the postulated independent noise sequence)

$$F = (A^TA)^{-1}A^T$$

and $f_i^T f_i$ is the i,i term of

$$FF^T = (A^TA)^{-1}A^TA(A^TA)^{-1} = (A^TA)^{-1}$$

Received July 25, 1977; presented at the AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyo., Sept. 7-9, 1977; revision received Oct. 3, 1977.

Index categories: Computer Communications, Information Processing and Software; Analytical and Numerical Methods.

*Deputy Manager, Tracking Systems and Applications Section.

Since the computed covariance of the ith component would be

$$\sigma_i^2 = i,i \text{ term of } [(A^TWA)^{-1} = \sigma_d^2(A^TA)^{-1}]$$

it follows that

$$f_i^T f_i = \sigma_i^2/\sigma_d^2$$

The Euclidean length of the ξ given by Eq. (2) will then be

$$\|\xi^*\| = \sigma_d/\sigma_i$$

It is often more convenient to think in terms of the rms value of the data error components themselves

$$\text{rms } \xi^* = \frac{\sigma_d\,|\hat{x}_i - x_i|}{\sqrt{n}\sigma_i} \tag{3}$$

In words (and slightly rearranged), Eq. (3) declares that if the rms data error is equal to σ_d, then the estimate error can be no larger than $\sqrt{n}$ times the computed estimate standard deviation σ_i.

These comments are included to show the simple functional character of the worst case error sequences. As has already been stated, they are only strictly true for uniformly weighted data without correlations. Similar remarks would be true in general for the slightly modified problem of minimizing $\xi^TW\xi$.

Application of the Worst Case Sequence to a VLBI Problem

In very long baseline interferometry (VLBI), the essential problem is to measure the difference in the time of arrival of a radio signal at each of two antennas separated by distances as large as the diameter of the Earth. Descriptions of this technique and its applications are provided in Refs. 4-6. A prime use of the VLBI technique is the precise measurement of the baseline vector between the two antennas. Instrumentation and data analysis programs directed toward geometrical measurements of this kind at the decimeter level and below are being actively pursued at several institutions.[7-9] These programs are currently in an intense validation phase, where attempts are being made to certify that the claimed accuracies are actually being achieved. As part of this validation process, it is important to understand the maximum sensitivity of the technique to data error. To that end, the following analysis is being pursued. A specific but typical example will be analyzed here. In this example, the baseline is between antenna sites at the Jet Propulsion Laboratory and Goldstone, both located in California, and the measurement sequence analyzed was an actual one used in an ARIES[7] experiment. For the purposes at hand, it is nearly sufficient to fix one end (station) of the interferometer and focus on the other as if it were the only one measuring event times and experiencing error. Figure 1 shows the local geometry involved and states the partial derivatives for the delay sensitivities with respect to the three directions: vertical (V), west (W), and north (N). The use of azimuth (a_z) and elevation (Σ) angles as the independent variables, while not the only choices possible, are natural ones because: 1) the equations are not complex, 2) the antennas normally used have an azimuth-elevation mount, and 3) troposphere and ionosphere errors are systematic in elevation angle.

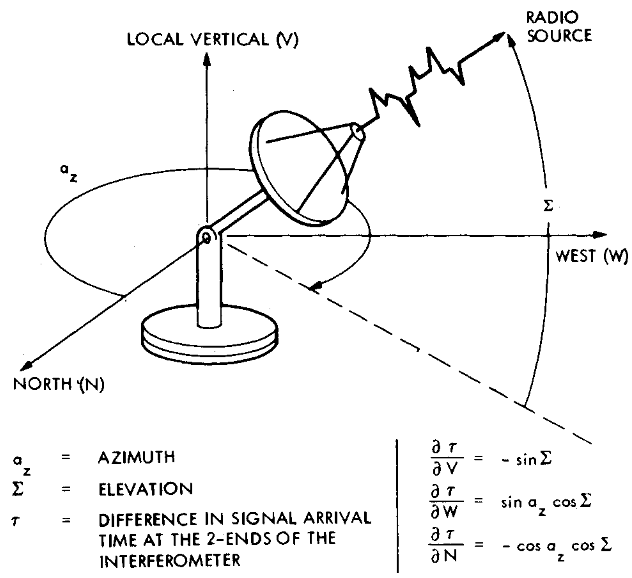


Fig. 1 Local geometry and time delay partials for one end of the interferometer.

In this typical experiment, twenty relative time delay measurements are recorded from ten natural radio sources of assumed known locations and a baseline vector solution is calculated. The solution typically includes a station delay constant and a station delay rate constant. That is, the measurements are assumed to contain a bias and a rate error, these terms accounting primarily for the two stations' differential clock effects.

Figure 2 shows the observing sequence used in terms of the a_z and Σ angles required to align the antennas toward the radio sources. Not shown well in this polar plot is the pattern of repeat measurements of a single quasar; but repeat numbers mean repeat measurements. This plot illustrates a unique property of VLBI experiments in that the three important variables—time, azimuth, and elevation angles—are not smoothly varying functions of each other. This is important both in finding data error signatures that can produce large solution errors and in interpreting residual patterns in the data subsequent to the fit.

Figure 3 displays the worst case solution obtained from direct application of Eq. (2) to the observation strategy depicted in Fig. 2. Shown is the minimum data error sequence capable of producing a decimeter error in the local vertical. This data error sequence has an rms strength of 0.059 ns, i.e.,

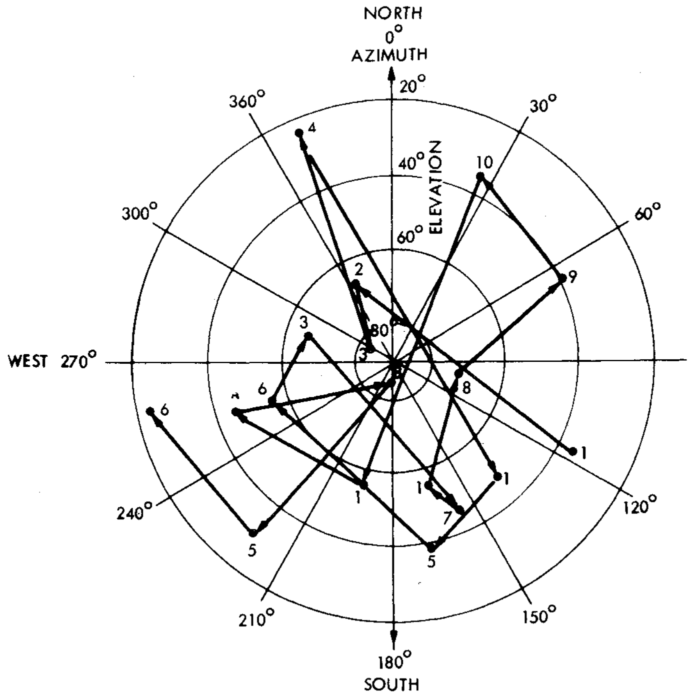
$$\text{rms}\xi^* = \sqrt{\xi \cdot \xi / n} = 0.059 \text{ ns}$$

where n is the number of data points.

The first observation to make about this result is that a 0.059 ns error signature (about 1.8 cm in light-seconds) is very small to be doing a decimeter's damage in the final result. This would mean that there can be up to about a factor of 5 multiplication from the size of the error source and the resulting baseline error. The second observation should be that the absolute worse case error looks not at all systematic, but a good deal like random noise, i.e., noise carefully specified, not to average out, but to add up to the maximum error possible. Hence, in a very real sense, all we have shown thus far is what was already known: If you do not assume some cancellation of the effects of the noise components, the resultant error can grow very large. These results are certainly useful as upper sensitivity bounds, but they do not aid directly in isolating truly systematic error patterns of reasonably likely incidence (e.g., slowly varying ones) that might produce particularly large estimate errors.

Modified Worst Case Analysis

In order to answer our own objections to the absolute worst case approach of the previous section, we pose the following modified problem: What is the sequence of data errors of minimum Euclidean norm that can be generated by a jth degree polynominal and produce a unit error in a selected component of the baseline? This point of view brings two



SOURCE NO.	SOURCE ID
1	3C 273
2	4C 55.16
3	4C 39.25
4	DW 02244
5	P 1127-14
6	OJ 287
7	3C 279
8	3C 286
9	3C 345
10	3C 371

Fig. 2 Radio source observing strategy (Aries experiment 76A, Jan. 15, 1976).

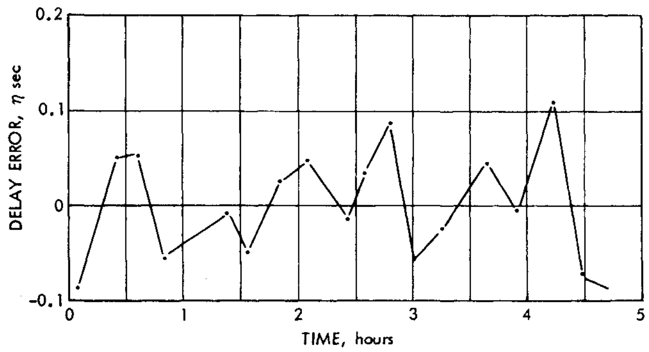


Fig. 3 Minimum norm data error sequence producing 1 dm error in the local vertical.

changes: 1) it imposes a smoothness condition whose severity is a function of j on the resulting data error sequence, and 2) the result will now depend on what variable is considered the independent variable—time, elevation angle, etc.—in the VLBI example.

In the following development, time will be considered the independent variable, but it should be clear that this selection is arbitrary. We shall work with a finite sum of Legendre polynomials.[10]

$$\xi_l = \sum_{m=0}^{j} a_m P_m(t_l) \tag{4}$$

where t_l are the observation times normalized on the -1, 1 interval to conform with the Legendre construction, and P_m represents the mth degree polynomial.

In terms of Eq. (4), the problem is to select a_m, $m=0, j$, which minimizes $\xi^T\xi$ subject to the constraint that $f_i^T\xi = 1$. Define the functional

$$J = \xi^T\xi + \lambda(f_i^T\xi - 1) = a^T S^T S a + \lambda[f_i^T S a - 1] \tag{5}$$

where

$$a = \begin{bmatrix} a_0 \\ \vdots \\ a_j \end{bmatrix}$$

λ is the undetermined Lagrange multiplier, S is an $n \times j+1$ matrix giving the numerical relationship between the values of the a_m coefficients and the data at t_l. That is, $\xi = Sa$.

Taking variations of Eq. (5)

$$\delta J = 2a^T S^T \; S\delta a \lambda f_i^T S\delta a + \delta\lambda [f_i^T Sa - 1] \tag{6}$$

For a true minimum, Eq. (6) must equal zero for arbitrary perturbations of a and λ. Thus, two conditions are obtained

$$2a^T S^T S + \lambda f_i^T S = 0$$

$$f_i^T Sa - 1 = 0 \tag{7}$$

Solving Eq. (7) for the desired coefficients yields

$$a^* = (S^T S)^{-1} S^T f_i [f_i^T S (S^T S)^{-1} S^T f_i]^{-1} \tag{8}$$

Comments

1) The solution bears a close relationship to that of a least squares fit of the first $j+1$ polynomials to the worst case data error sequence of Eq. (2). The solution to that problem is readily shown to be

$$\tilde{a} = (S^T S)^{-1} S^T f_i [f_i^T f_i]^{-1}$$

This is to within a scalar multiplicative constant of Eq. (8). The modified worst case determination can thus be described as: 1) create a least-squares fit to the absolute worst case solution with the first $j+1$ polynomials, and 2) scale the result upward until a unit error is produced.

2) In the computation of a, it is required both that $(S^T S)$ be invertible and that $f_i^T S$ be other than the null vector. In theory this first condition will be satisfied for any discrete set of t_m, although in practice we did experience a numerically singular condition when $J+1=n$. Fortunately, this is also the case when S is square and Eq. (8) reduces to

$$a^* = S_\square^{-1} f_i / f_i^T f_i \tag{8a}$$

$$j+1 = n$$

$S_\square$ itself has always proved numerically invertible. In the example case, the second condition is not satisfied for $j=0,1$.

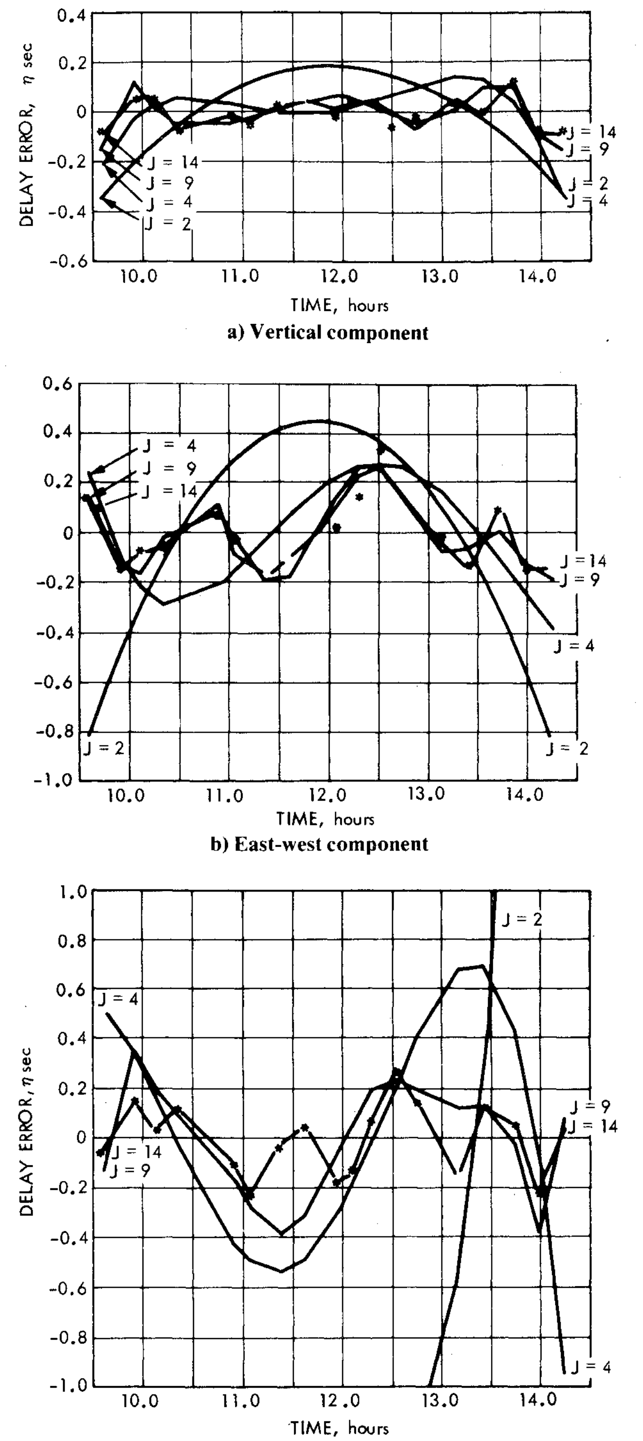


Fig. 4 Minimum norm error signatures generated from restricted degree polynomials producing 1 dm error. Independent variable: time.

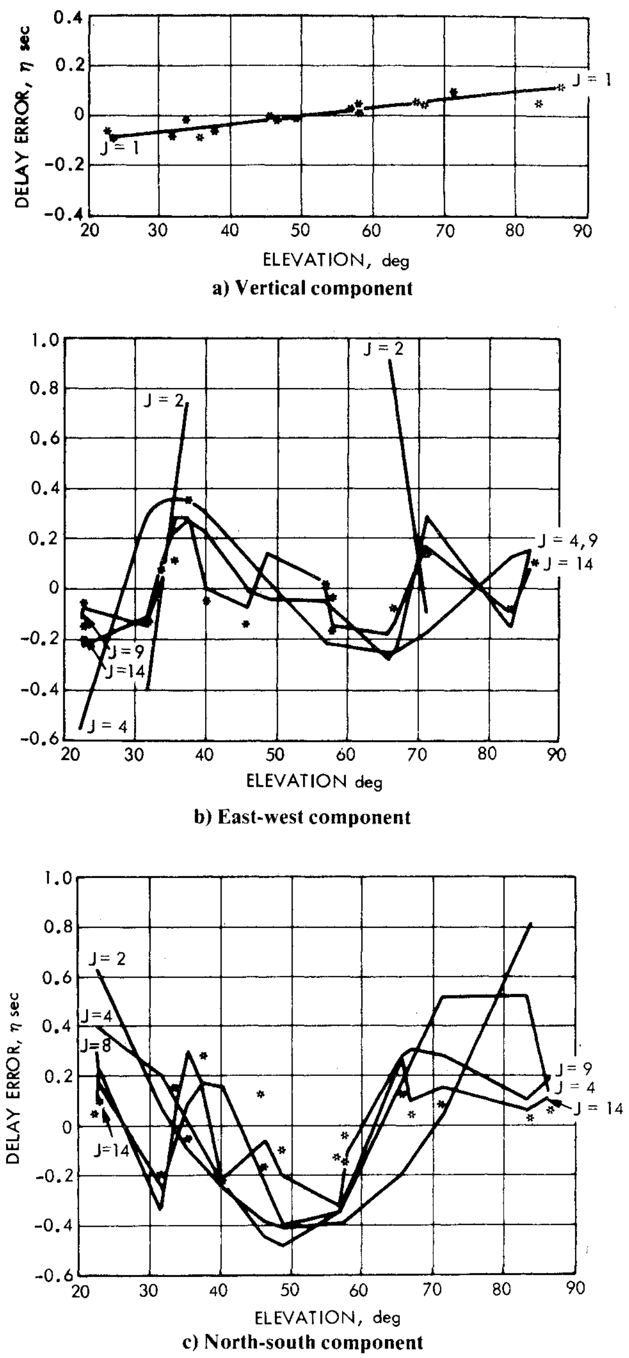


Fig. 5 Minimum norm error signatures generated from restricted degree polynomials producing 1 dm error. Independent variable: elevation angle.

This arises because the solution vector includes constant and linear terms (in time) and it is a property of least-squares estimation that f_i will therefore be orthogonal to both any constant or linear effect. In physical terms, this means nothing more than that these two effects cannot produce baseline error.

Figure 4 shows the results of applying this analysis to the example case. The * are the absolute worst case discussed in the previous section. The connected line plots result from evaluating the constrained degree worst case polynomial at the data times. Table 1 summarizes the rms value of the data error signature required to produce the canonical 1 dm baseline error for these and other cases. As already mentioned, there is no solution for $j=0,1$. Note that the curves converge to the absolute worst case point set as j approaches 17 ($n=18$, for the example). Note also from the table that the rms ξ required is about 0.18 ns (5.4 cm) for $j=2$ and gradually diminishes to the previously quoted 0.059 ns for large j.

Figures 4b and 4c display the same information except that now the signatures are optimized to produce the largest effect in the E-W and then the N-S directions. As shown in Table 1, the rms ξ required for the absolute worst case rises to 0.13 and 0.14 ns, respectively. This is in the same ratio as are the calculated σ's for these baseline components computed in the usual way via $(A^TWA)^{-1}$ (c.f. the comments discussed previously). These ratios are not well maintained for the modified worst cases shown, however. For example, the N-S component is relatively immune to data errors generated from low degree polynomials ($j\leq 4$), although it is nearly equal to

Table 1 Minimum norm time dependent error signatures required for generating 1 dm baseline component error

Baseline component	Computed component standard deviation, cm	rms ξ, ns Absolute worst case	Generating polynomial degree $J=2$	$J=4$	$J=9$	$J=14$
Vertical	12	0.059	0.179	0.119	0.075	0.063
East-west	5.2	0.131	0.430	0.215	0.146	0.138
North-south	5.0	0.140	2.34	0.455	0.221	0.140

Table 2 Minimum norm elevation angle dependent error signatures required for generating 1 dm baseline component error

Baseline component	rms ξ, ns Absolute worst case	Generating polynomial degree $J=1$	$J=2$	$J=4$	$J=9$	$J=14$
Vertical	0.059	0.063	0.062	0.062	0.060	0.059
East-west	0.131	15.58	2.07	0.284	0.175	0.145
North-south	0.140	2.65	0.444	0.344	0.250	0.191

Table 3 Minimum norm azimuth angle dependent error signatures required for generating 1 dm baseline component error

Baseline component	rms ξ, ns Absolute worst case	Generating polynomial degree $J=1$	$J=2$	$J=4$	$J=9$	$J=14$
Vertical	0.059	0.54	0.40	0.31	0.090	0.066
East-west	0.131	0.161	0.158	0.141	0.135	0.134
North-south	0.140	2.15	0.18	0.161	0.151	0.148

the E-W sensitivity for the higher degrees. Both of these components are uniformly less sensitive than was the vertical.

As mentioned earlier, it is necessary to consider something other than time as the independent variable for the modified analyses. Figure 5a displays the absolute worst case points shown for the vertical component in Fig. 3, but plotted as a function of the elevation angle of the observable. Note that the points, when arranged this way, have a decided trend that looks linear except for small departures. It is easy to see that these departures have little significance, however, by calculating the $j=1$ curve which produces the decimeter vertical error. This is also shown plotted in Fig. 5a; its rms value is 0.063 ns, only some 7% higher than the absolute minimum of 0.059 ns already calculated. The results for higher degree cases are not included in the figure since they depart little from the displayed linear solution.

Thus, elevation angle dependent systematic error will couple into estimates of the local vertical very efficiently. A multiplication factor of more than 5 exists between the rms data error and its induction into the vertical. Lest this ratio seem too large, it is noted that throughout the discussion the comparison is being made between the rms data error and the total component movement. Perhaps a better basis for comparison would be the peak-to-peak data error. Viewed thus, the vertical component error would be 1.6 times the peak-to-peak data error for the linear function. Figures 5b and 5c show the same information, but for the N-S and E-W baseline components. These components are both quite insensitive to all low degree errors systematic in elevation angle. The behavior of all 3 components with systematic elevation angle error is summarized in Table 2.

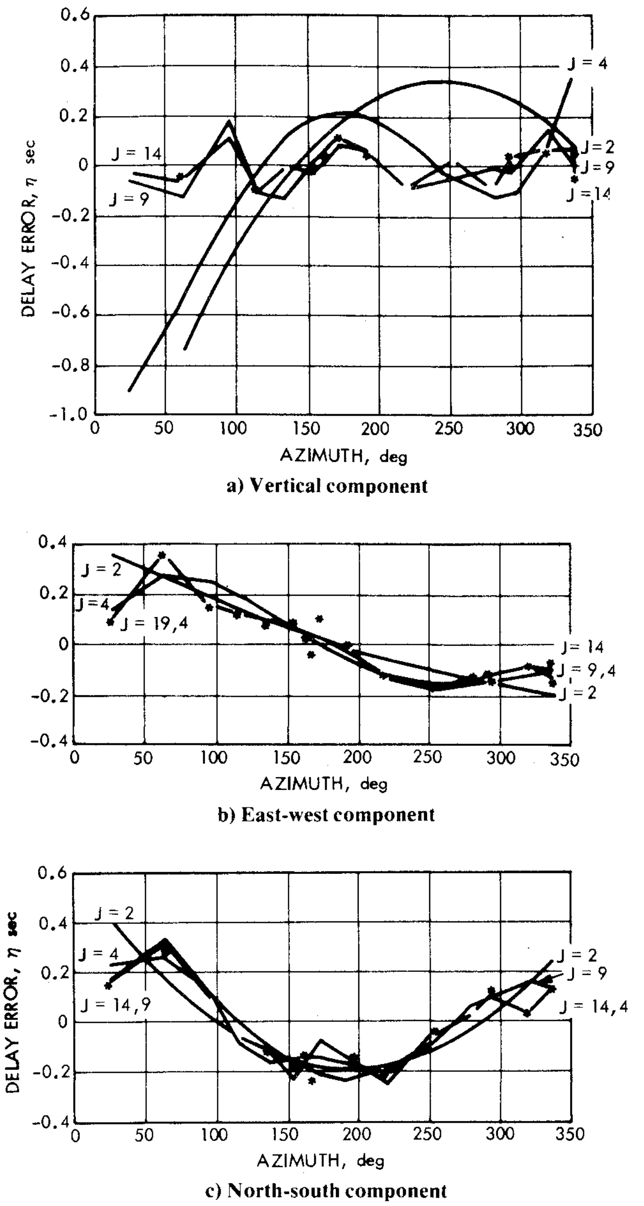
Finally, Figs. 6a-c and Table 3 display the same results, but this time the independent variable is azimuth angle. The E-W and N-S directions are intermediate in sensitivity, but this sensitivity can be very nearly reached with rather low order polynomials in azimuth.

The results shown are clearly dependent on the observation strategy and only a single strategy has been used to generate the numerical results. In particular, it should be possible to partly immunize the E-W and N-S components to systematic error using the preceding algorithms as a guide. The vertical component sensitivity appears nearly strategy-free; however, the simple $\sin\Sigma$ structure of that component's delay partial essentially prevents strategy ingenuity from helping very much.

In sumary, local vertical is by far the most sensitive direction to systematic error and this maximum sensitivity is achieved for very simply structured errors systematic in elevation angle. The E-W and N-S components are relatively more immune; they achieve their maximum error sensitivities earliest when subjected to moderately simple functions systematic in azimuth angle.

Detecting Systematic Error in the Residuals: Troposphere Error, A Special Case

A subject companion to that of charting error sensitivities is the investigation of detecting systematic error via residual analysis. Unfortunately, the worst case analysis just presented is not a very good vehicle to be used directly for this purpose. The absolute worst case error sequence produces no residuals at all, and the polynomials proximate to the worst case produce very small ones. However, we shall use a tropospheric error as an example case, representative of a physical source for what has been found to be the most



a) Vertical component

b) East-west component

c) North-south component

Fig. 6 Minimum norm error signatures generated from restricted degree polynomials producing 1 dm error. Independent variable: azimuth angle.

sensitive error coupling pair: elevation dependent error into the local vertical.

Figure 7 displays the observation error sequence arising from assuming a tropospheric zenith delay error of 0.1 ns (3 cm) and using the simple model that the time delay error thus induced would be

$$\xi_T = 0.1/\sin\Sigma \text{ ns}$$

If a simple baseline-only estimate (i.e., no relative clock offset and rate parameters are included) is performed using this error signature, a −5.1 cm local vertical error will arise and the rather strong and systematic residual sequence (when plotted vs elevation angle) shown in Fig. 8 is created. The sense of the error is what might be expected: a denser troposphere than that modeled will result in a depressed apparent station location. The strong residual pattern arises simply because $0.1/\sin\Sigma$ cannot be made to well match a $\sin\Sigma$ for any choice of a. The situation is quite changed, however, when the required clock parameters (bias time and rate) are included, as they must be in any solution with real data. For now $0.1/\sin\Sigma$ can be matched very well to a $\sin\Sigma + b$ by

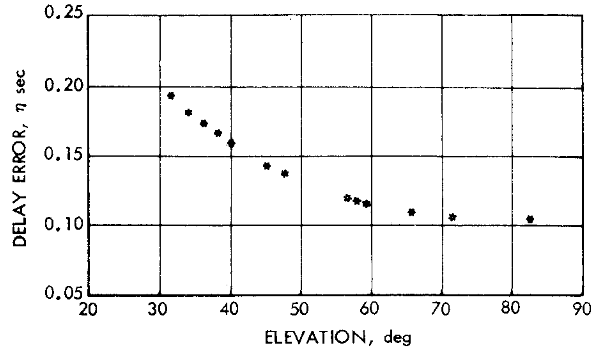


Fig. 7 Pre-fit data error sequence from a 3.0 cm tropospheric zenith delay error.

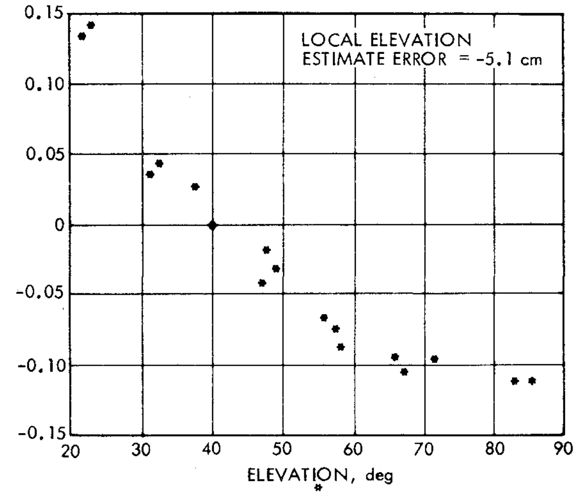


Fig. 8 Residuals from a 3.0 cm tropospheric zenith delay error subsequent to a baseline-only solution (no clock parameters).

choosing a negative a and then cancelling the bias with an appropriate selection for the clock offset estimate, b. This process is shown graphically in Fig. 10. This results in a +6.7† cm vertical error (that is the station will appear *higher* than actual in spite of the denser atmosphere) and the weak residual pattern shown in Fig. 9. (Note the scale change between Figs. 8 and 9.) The residuals are not only very small but what is left does not appear systematic even when plotted vs elevation angle.

This is a curious result since untrained intuition (the author's, at least) would note that the major effects (both the tropospheric induced data and the fit to it) were systematic in elevation; the residuals should be also. However, there is some response in both the N-S and E-W directions (~0.3 cm) so that in forming the residual

$$r = \frac{1}{\sin\Sigma} - [\sin\Sigma,\ \sin a_z \cos\Sigma,\ -\cos a_z \cos\Sigma] \begin{bmatrix} v \\ W \\ S \end{bmatrix}$$

terms dependent on a_z are injected. a_z is erratic (as shown by Fig. 2) when viewed as a function of Σ. Note that the variation from point to point in the residuals from the baseline-only solution (Fig. 8) really was present in about this amount (~.025 ns); it just is not so apparent when the basic fit is not very good.

†The rms τ for a troposphere signature that could produce a decimeter vertical estimate error is 0.233 ns, or about a factor of 4 less efficient than the absolute worst case.

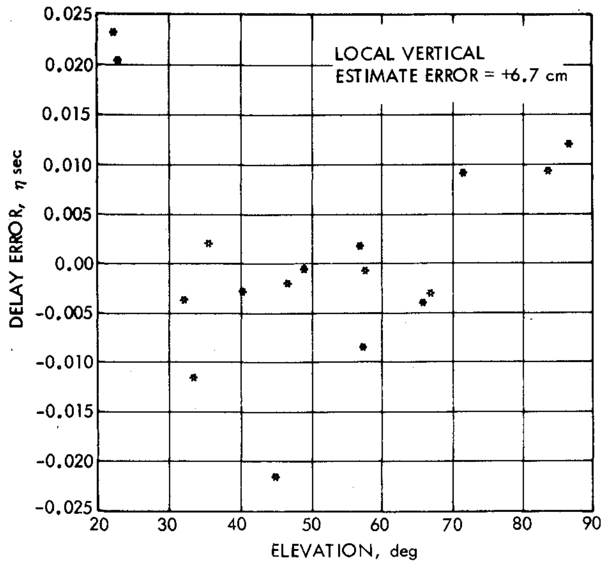


Fig. 9 Residuals from a 3.0 cm tropospheric zenith delay error subsequent to a standard fit including clock parameters.

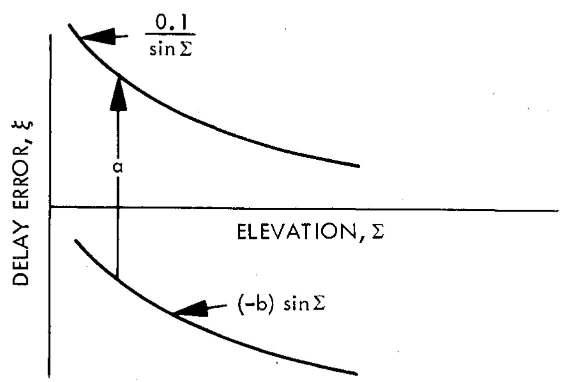


Fig. 10 Schematic showing congruence between troposphere and station vertical location effects.

This is not a matter to be taken lightly. Inspection for systematic trends in the residuals is usually a powerful means of validating the solution to an estimation problem. In fact, in the design of many instrumentation systems, the single most important reason for keeping the random noise as low as possible is so that the more subtle systematic effects (which really determine the solution accuracy) can be seen and studied. It is a troubling aspect of the VLBI technique to have the artifacts of systematic error so neatly disguised as nearly random noise.

Summary

Two separate algorithms for testing filter sensitivity to systematic data errors have been derived: one yielding the absolute minimum Euclidean norm data error for a given estimate component error and the second finding the minimum norm data error which can be generated by restricted degree Legendre polynomials. A specific VLBI baseline estimation was analyzed with these algorithms and it was determined that: 1) The local vertical is the most sensitive component to error in the data space (0.059 ns rms data error would produce 10 cm estimate error), and 2) The efficiency of a data error sequence linear in elevation angle is within 7% that of the absolute worst case sequence.

Elevation angle dependent errors were explored further and the special case of a mismodeled troposphere was treated. This data error coupled only about 25% as efficiently into component error as did the linear function. It was noted that the observational sequence flexibility of the VLBI technique, while quite useful for defeating some types of systematic error, contributed to the surprising phenomenon that the residuals from systematic error disguised themselves rather well as random noise, making presence detection more difficult.

Acknowledgment

This paper represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under NASA Contract NAS 7-100. The following people were helpful and supportive in the preparation of this paper: G. Bierman, S.G. Finley, E. Kennedy, P. MacDoran, W. Melbourne, D. Morabito, X. Newhall, G. Resch, D. Trask, and B. Thomas.

References

[1] Sorenson, H.W., "Least Squares Estimation: from Gauss to Kalman," *IEEE Spectrum,* Vol. 7, July 1970, pp. 63-68.

[2] Lawson, C.L. and Hanson, R.J., *Solving Least Squares Problems,* Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[3] Bierman, G.J., *Factorization Methods for Discrete Sequential Estimation,* Academic Press, New York, 1977, pp.178-181.

[4] Broten, N.W. et al., "Long Baseline Interferometry: A New Technique," *Science,* Vol. 156, 1967, pp. 1592-93.

[5] Williams, J.G., "Very Long Baseline Interferometry and Its Sensitivity to Geophysical and Astronomical Effects," Jet Propulsion Laboratory, Pasadena, Calif., Space Program Summary 37-62, Vol. 2, Feb. 1970.

[6] Thomas, J.B., "An Analysis of Long Baseline Radio Interferometry, Parts I-III," Deep Space Network Progress Report, Technical Report 32-1526, Vols. 7, 8, 16, Jet Propulsion Laboratory, Pasadena, Calif., 1972, 1973, pp. 37-50; 29-38; 47-64.

[7] MacDoran, P.F., "Field Demonstration Phase Project Plan for Astronomical Radio Interferometric Earth Surveying," Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1975.

[8] Thomas, J.B., et al., "Demonstration of an Independent-Station Ratio Interferometry System with 4cm Precision on a 16-km Baseline," *Journal of Geophysical Research,* Vol. 81, Feb. 1976, pp. 995-1005.

[9] Rogers, A.E., et al., "Geodesy by Radio Inteferometry," submitted for *Journal of Geophysical Research* publication, Nov. 1976.

[10] *NBS Handbook of Mathematical Functions,* Applied Mathematics Series 55, U.S. Department of Commerce, June 1964.