K152

DEPARTMENT OF EARTH AND PLANETARY SCIENCES

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139

July 17, 1978

MEMORANDUM

To: Distribution

From: J. F. Chandler

Subject: Extension of PEP Normal Equation Error Statistics

References:

1. R.D. Reasenberg, on "Convergence Indicator in PEP", #74-1, 1974 Jan 24.
2. R.D. Reasenberg, on "The PEP A Priori Facility", #75-2, 1975 Jul 28.
3. R.D. Reasenberg, on "Partial Prereduction of the Normal Equations", #75-3, 1975 Jul 29.

At present, the formation or restoration of normal equations for parameter estimation in PEP produces an nxn coefficient matrix B, an n-vector right-hand side u, and the sum, sum absolute and sum-squared weighted residuals. These are sufficient for solving for adjustments to parameters and predicting the post-fit sum-squared and, thus, also the root-mean-squared (RMS) weighted residuals, but not for predicting the post-fit sum or sum absolute weighted residuals. Indeed, the sum of absolute values of the residuals after fitting (or before, for that matter) can be found only by going through the observations one by one. However, the sum of weighted residuals is a simple linear combination of observables and computed values and the post-fit sum can be predicted by applying a correction to the prefit sum.

Let us begin with a brief development of the least-squares estimation technique. Consider a set of observations represented as an m-vector z with associated errors ν. Suppose we have a set

of parameters α such that we can compute the theoretical values $H(\alpha)$ for the observations, given the nominal values α_o. A convenient approach is to form differenced variables based on the nominal parameter values as follows.

$$x = \alpha - \alpha_o$$

$$y \equiv z - H(\alpha_o)$$

$$A \equiv \left.\frac{\partial H(\alpha)}{\partial \alpha}\right|_{\alpha=\alpha_o} \qquad (1)$$

$$R \equiv \langle \nu\nu^+ \rangle$$

$$W \equiv R^{-1}$$

In general, the error cross correlation matrix R must be used for a maximum likelihood estimation of parameter adjustments x. In practice, however, the errors are assumed uncorrelated -- if necessary, they are made so by modeling observation biases in $H(\alpha)$. Thus, both R and w are diagonal, and the least-squares fit is maximum likelihood. First, we linearize the problem by expanding to first order about α_o and express the sum-squared weighted residuals in terms of x and y.

$$z - H(\alpha) = y - Ax$$

$$S(x) \equiv (y-Ax)^+ W(y-Ax)$$

$$= y^+Wy - 2x^+A^+Wy + x^+A^+WAx \qquad (2)$$

The least-squares solution $\hat{x}$ is given by

$$\left.\frac{\partial S(x)}{\partial x}\right|_{x=\hat{x}} = 0$$

$$A^{+}WA\hat{x} = A^{+}Wy$$

$$B\hat{x} = u \qquad (3)$$

$$B \equiv A^{+}WA$$

$$u \equiv A^{+}Wy$$

Thus, the best-fit parameter values are

$$\hat{\alpha} = \alpha_o + \hat{x}$$

$$= \alpha_o + B^{-1}u \qquad (4)$$

and the post-fit residuals are

$$\hat{y} = z - H(\hat{\alpha})$$

$$= y - A\hat{x} \qquad (5)$$

The pre- and post-fit sum-squared weighted residuals are $S(0)$ and $S(\hat{x})$, and we find, by substituting into (2)

$$S(0) = y^{+}Wy$$

$$S(\hat{x}) = S(0) - \hat{x}^{+}u \qquad (6)$$

Note that $\hat{x}^{+}u$ is just the norm-square (N^2) of the adjustment as derived in Reference 1.

Now we turn to the sum of weighted residuals and define the corresponding quantities analogous to equation (2).

$$G(x) \equiv (y-Ax)^{+}\omega$$

$$k \equiv A^{+}\omega \qquad (7)$$

$$\omega \equiv \text{vector of 1/errors}$$

In terms of the generalized maximum likelihood weighting matrix W, ω could be expressed as a projection into a vector of the square-root of W, but this reduces to just the reciprocal errors for the individual observations for uncorrelated errors. The vector k is, thus, a weighted sensitivity vector, that is, a weighted projection of the partial derivative matrix. Substituting into (7) we find the pre- and post-fit sums.

$$G(0) = y^{+}\omega$$

$$G(\hat{x}) = G(0) - \hat{x}^{+}k \qquad (8)$$

The realization of the prediction of post-fit average weighted residuals, then, requires only the accumulation of the vector k, quite analogously with the right hand side u. In particular, we note that k may be considered as extended with arbitrarily many zeros for parameters that have no effect on the observations. Further, we may combine observation sets directly by just adding the k's component-wise. Consider two sets of data (a) and (b) already processed separately. Each has a saved set of quantities B, u, S(0), G(0), and k, but, of course, the original solutions involved separate y, A, w, and ω. Now,

$$y = \begin{pmatrix} y^{(a)} \\ y^{(b)} \end{pmatrix}, \quad A = \begin{pmatrix} A^{(a)} \\ A^{(b)} \end{pmatrix}, \quad w = \begin{pmatrix} w^{(a)} & 0 \\ 0 & w^{(b)} \end{pmatrix}, \quad \omega = \begin{pmatrix} \omega^{(a)} \\ \omega^{(b)} \end{pmatrix} \tag{9}$$

$$\text{So,} \quad G(0) = \begin{pmatrix} y^{(a)} \\ y^{(b)} \end{pmatrix}^{+} \begin{pmatrix} \omega^{(a)} \\ \omega^{(b)} \end{pmatrix} = y^{(a)+}\omega^{(a)} + y^{(b)+}\omega^{(b)}$$

$$= G(0)^{(a)} + G(0)^{(b)} \tag{10}$$

$$k = \begin{pmatrix} A^{(a)} \\ A^{(b)} \end{pmatrix}^{+} \begin{pmatrix} \omega^{(a)} \\ \omega^{(b)} \end{pmatrix} = k^{(a)} + k^{(b)}$$

and $G(\hat{x}) = G(0) - \hat{x}^{+}k$ in just the same way as before.

PEP parameter estimation has a number of features which depart from the straightforward solution via (1-4), notably the inclusion of a priori information in the normal equations (see Reference 2). A priori input unfortunately destroys the simplicity of predicting post-fit RMS residuals. If we let (d) represent the data portions and (a) represents the a priori, we have

$$\begin{aligned} B &= B^{(d)} + B^{(a)} \\ u &= u^{(d)} + u^{(a)} \\ u^{(a)} &= B^{(a)} x^{(a)} \\ x^{(a)} &= \alpha^{(a)} - \alpha_o \\ x &= (B^{(a)} + B^{(d)})^{-1} (u^{(a)} + u^{(d)}) \end{aligned} \tag{11}$$

$$\begin{aligned} S(\hat{x}) &= y^{+}Wy - 2\hat{x}^{+}A^{+}wy + \hat{x}^{+}A^{+}wA\hat{x} \\ &= S(0) - 2\hat{x}^{+}u^{(d)} + \hat{x}^{+}B^{(d)}\hat{x} \\ &= S(0) - \hat{x}^{+}u + \hat{x}^{+}u^{(a)} - \hat{x}^{+}u^{(d)} + \hat{x}^{+}(u - B^{(a)}\hat{x}) \\ &= S(0) - N^2 + \hat{x}^{+}(2u^{(a)} - B^{(a)}x) \end{aligned}$$

However, since k is unaffected by a priori information and G is linear in x, no such problem exists for predicting average residuals.

The partitioning of the parameter set α for partial pre-reduction of the equations (see Reference 3) likewise presents no difficulty.

The technique is primarily a computational shortcut. In brief, a sub-matrix of B for part of the parameters is inverted once, and the solution is applied to the remainder of the elements of B and u to produce a smaller set of normal equations for the remaining parameters. The vector k is simply an n-vector of sensitivities and may be partitioned in the same was as u. Thus, the saved sum of reduced residuals can be formed analogously with the reduced sum-squared residuals. Let

$$u = \begin{pmatrix} u_1 \\ \\ u_2 \end{pmatrix}, \quad B = \begin{pmatrix} C & F \\ \\ F^+ & D \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \\ x_2 \end{pmatrix} \tag{13}$$

Then

$$\begin{aligned} u_1 &= C\hat{x}_1 + F\hat{x}_2 \\ u_2 &= F^+\hat{x}_1 + D\hat{x}_2 \\ \hat{x}_2 &= D^{-1}(u_2 - F^+\hat{x}_1) \end{aligned} \tag{14}$$

and the reduced normal equations appear as

$$\bar{u}_1 = \bar{C}\hat{x}_1 \tag{15}$$

where

$$\bar{u}_1 \equiv u_1 - FD^{-1}u_2 \tag{16}$$

$$\bar{C} \equiv C - FD^{-1}F^{+}$$

Examining the sum-squared residuals, we find

$$S(\hat{x}) = S(0) - \hat{x}_1^{+}u_1 - \hat{x}_2^{+}u_2$$

$$= S(0) - \bar{x}_2^{+}u_2 - \hat{x}_1^{+}\bar{u}_1$$

$$\bar{x}_2 \equiv D^{-1}u_2 \tag{17}$$

Thus, we may define a reduced sum-squared quantity

$$\bar{S}(x_1) \equiv S(0) - \bar{x}_2^{+}u_2 \tag{18}$$

The sum of residuals yield similar results:

$$G(\hat{x}) = G(0) - \hat{x}_1^{+}k_1 - \hat{x}_2^{+}k_2$$

$$= G(0) - u_2^{+}D^{-1}k_2 + \hat{x}_1^{+}D^{-1}k_2 - \hat{x}_1^{+}k_1$$

$$= \bar{G}(0) - x_1^{+}\bar{k}_1 \tag{19}$$

where

$$\bar{k}_1 \equiv k_1 - FD^{-1}k_2$$

$$\bar{G}(x_1) \equiv G(0) - \bar{x}_2^{+}k_2 \tag{20}$$

In some applications it may be desirable to transform the normal equations by changing the nominal parameter values α_o (for example, to add two sets of equations with different nominals). The procedure bears some similarity to the inclusion of a priori parameter estimates in that the right-hand side u must be adjusted by $-B\Delta\alpha_o$. At the same time the error statistics must also be adjusted to reflect the new computed values of observations.

$$y^* = y - A\Delta\alpha_o \tag{21}$$

$$u^* = u - B\Delta\alpha_o \tag{22}$$

First, we have the sum-squared residuals

$$\begin{aligned} S^*(x) &= S(x+\Delta\alpha_o) \\ S^*(0) &= S(0) - 2\Delta\alpha_o{}^+A^+Wy + \Delta\alpha_o{}^+A^+WA\Delta\alpha_o \\ &= S(0) - 2\Delta\alpha_o{}^+u + \Delta\alpha_o{}^+B\Delta\alpha_o \\ &= S(0) - \Delta\alpha_o{}^+(u+u^*) \end{aligned} \tag{23}$$

Then, the summed residuals

$$\begin{aligned} G^*(x) &= G(x+\Delta\alpha_o) \\ G^*(0) &= G(0) - \Delta\alpha_o{}^+k \end{aligned} \tag{24}$$

We have seen that the maintenance of the sensitivity vector k can be accomplished with a relatively simple set of operations and, further, that it is necessary for the proper handling of error statistics in nominal value shifting. The implementation of this feature will make possible the prediction of post-fit average weighted residuals without predicting the residuals themselves.

Distribution:

M. E. Ash
R. J. Cappallo
P. G. Ford
A. Forni
R. B. Goldstein
S. A. Gourevitch
R. W. King
P. E. MacNeil
G. H. Pettengill
R. D. Reasenberg
I. I. Shapiro