Memo about reweighting

L. Petrov
pet@leo.gsfc.nasa.gov

February 2, 1998

Abstract:

Discussion about why it is useful to have χ^2 per degree of freedom to be near unity and how it is possible to reach it.

The sum of weighted squares of postfit residuals $R = (A\hat{x} - y)^\top \alpha^{-1} (A\hat{x} - y)$ can be reduced to the following form:

$$\begin{aligned} R &= (\varepsilon - A V A^\top \alpha^{-1} \varepsilon)^\top \alpha^{-1} (\varepsilon - A V A^\top \alpha^{-1} \varepsilon) = \\ & \varepsilon^\top \alpha^{-1} \varepsilon \; - \; \varepsilon^\top \alpha^{-1} A V A^\top \alpha^{-1} \varepsilon \end{aligned} \tag{1}$$

where ε — (unknown) noise of measurements plus contribution of the unmodeled effects, V — covariance matrix of the estimates, A – matrix of equations of conditions, α^{-1} – weight (diagonal) matrix, y – vector of right parts “o-c”, $\hat{x}$ – vector of estimates.

Let us find mathematical expectation of R, taking into consideration that for the either vector x and the square matrix M of corresponding dimension $x^\top M x = \mathrm{Sp}(x x^\top M)$ is true and operation $\mathrm{Sp}(A B) = \mathrm{Sp}(B^\top A)$:

$$\mathcal{E}(R) = \mathrm{Sp}(\alpha^{-1} \operatorname{Cov} \varepsilon) - 2\,\mathrm{Sp}(\operatorname{Cov} \varepsilon\, \alpha^{-1} A V A^\top \alpha^{-1}) + \mathrm{Sp}(\operatorname{Cov} \varepsilon\, \alpha^{-1} A V A^\top \alpha^{-1} A V A^\top \alpha^{-1}) \tag{2}$$

Where Sp is the trace of the matrix. If no constraints have been applied the term $\mathrm{Sp}(\operatorname{Cov} \varepsilon\, \alpha^{-1} A V A^\top \alpha^{-1} A V A^\top \alpha^{-1})$ is equal to $\mathrm{Sp}(\operatorname{Cov} \varepsilon\, \alpha^{-1} A V A^\top \alpha^{-1})$ and the expression (2) is simplified to the form:

$$\mathcal{E}(R) = \mathrm{Sp}(\alpha^{-1} \operatorname{Cov} \varepsilon) - \mathrm{Sp}(\operatorname{Cov} \varepsilon\, \alpha^{-1} A V A^\top \alpha^{-1}) \tag{3}$$

It is not true for the case when constraints are applied but we will ignore the difference and use simplified expression (3). We postpone discussion whether it is legal or not and notice only that we can hope that since usually the contribution of the constraints in normal equations is not large the error introduced by this substitution should not be substantial.

While calculating we should keep in mind that $\mathrm{Sp}(AB^\top) = \sum\limits_{i,j} a_{ij} b_{ij}$ is true for arbitrary matrices of corresponding dimensions, and therefore it is not necessary to calculate the matrix multiplication directly.

Having calculated *the value* of sum of squares of weighted residuals R and their *mathematical expectation* $\mathcal{E}(R)$ we can judge how good our deterministic model (confined in matrix A), describing the influence of the effects modeled by linear equations on observations corresponds to our stochastic model (confined in weight matrix α), describing influence of the unmodeled effects and measurement noise on observations. If deterministic model of the observations is complete, we didn't lose the precision during calculations, covariance matrix of the right parts is actually diagonal, and their diagonal terms are known precisely, in that case the ratio $\frac{R}{\mathcal{E}(R)}$ will be 1. Incompleteness of the model leads to increasing "o-c" due to the contribution of unmodeled effects. "Noise" will be a sum of the errors of the measurements and the contributions of the unmodeled effects. As a result of presence systematic errors diagonal terms of the covariance matrix will be increasing and off-diagonal terms will be appearing. Therefore the ratio $\frac{R}{\mathcal{E}(R)}$ (called also "χ^2 per degree of freedom") in practice is rare equal to 1.

Now assume that the choice of our a priori covariance matrix $\mathrm{Cov}\,\varepsilon$ was not perfect and it was known up to additive (quadratically) term q: $\mathrm{Cov}\,\varepsilon = \alpha + q^2 I$, where I is the unity matrix of dimension $n \times n$. Let's calculate mathematical expectation of R for such a case.

$$\begin{aligned} \mathcal{E}(R) &= \mathrm{Sp}\left((\alpha + q^2 I)\,\alpha^{-1}\right) - \mathrm{Sp}\left((\alpha + q^2 I)\,\alpha^{-1}\left(V\,A^{\top}\alpha^{-2}A\right)\right) = \\ &\quad \left[n - \mathrm{Sp}\left(V\,A^{\top}\alpha^{-2}A\right)\right] + q^2\left[\mathrm{Sp}\left(\alpha^{-1}\right) - \mathrm{Sp}\left(V\,A^{\top}\alpha^{-2}A\right)\right] \end{aligned} \tag{4}$$

Using (4) we can directly express q:

$$q = \sqrt{\frac{\mathcal{E}(R) - \left(n_i - \mathrm{Sp}(V A^{\top}\alpha_i^{-1}A)\right)}{\mathrm{Sp}(\alpha^{-1}) \;-\; \mathrm{Sp}(V\,A^{\top}\alpha^{-2}A)}} \tag{5}$$

If we substitute to (5) the *value* of R instead of *mathematical expectation* of $\mathcal{E}(R)$ we can find the values of q[1]. Under assumption that the true covariance matrix of noise is $\mathrm{Cov}\,\varepsilon = \alpha + q^2 I$ the ratio $\frac{R}{\mathcal{E}(R)}$ will be exactly 1. But our estimates were produced under assumption that $\mathrm{Cov}\,\varepsilon = \alpha$. If the actual covariance matrix is $\alpha + q^2 I$ rather than α our estimates $\hat{x}$ will not be optimal (estimates have minimal dispersion if the matrix of weights is an inverse to the covariance matrix of noise in according with Gauss-Markov theorem). We can update weights: $\alpha_i = \alpha_{i-1} + q_i^2$ and make the next solution. It is not difficult to prove that always $|q_i| > |q_{i-1}|$ and iterations converge. Usually it is sufficiently to have 2-3 iterations to reach $\frac{R}{\mathcal{E}(R)}$ in the range [0.95, 1.05].

Notice that we can calculate partial sum of squares of weighted residuals R_i for some subset of data (observations of the certain baseline or source) and calculate partial $\mathcal{E}(R_i)$. In that case we represent $\mathrm{Cov}\,\varepsilon$ as $\mathrm{Cov}\,\varepsilon = \alpha + \sum\limits_i q_i^2 I$. Correction q_i will be expressed by the following manner:

$$q_i = \sqrt{\frac{\mathcal{E}(R_i) - \left(n_i - \mathrm{Sp}(V A^{\top}\alpha_i^{-1}A)\right)}{\mathrm{Sp}(\alpha_i^{-1}) \;-\; \mathrm{Sp}(V\,A^{\top}\alpha_i^{-2}A)}} \tag{6}$$

[1] imaginary q corresponds to the case that we should decrease diagonal terms of a priori covariance matrix. It is easy to show that in any case corrected terms of a priori covariance matrix remain real.

where n_i – number of observations in the subset and α_i has zeroes on the main diagonal for all observations not belonging to the subset.

Several remarks about the efficient way of calculation of $\mathrm{Sp}(VA^\top \alpha_i^{-1} A)$:

$$\mathrm{Sp}(VA^\top \alpha_i^{-1} A) = \sum_i^n \frac{(a_i{}^\top \cdot V\, a_i)}{\alpha_i} \tag{7}$$

where a_i is the equation of conditions for the i-th observation, and the the symbol $(\cdot)$ designates dot product. Vector a_i is usually sparse since as a rule not all observations depend on all parameters to be estimated. We don't need to calculate all elements of the vector $V\, a_i$, but only those which correspond to the non-zero elements of the equation of conditions.

In the case when B3D scheme of solving conditional equations was implemented then the expression (7) is written in the following form:

$$\begin{aligned}
&\mathrm{Sp}(VA^\top \alpha_i^{-1} A) = \\
&\sum_i^{n_i} \frac{1}{\alpha_i} \Big[\; a_i^{g\,\top} \cdot \Big(\mathrm{Cov}(\hat{x}_o, \hat{x}_o^\top)\, a_i^g + \mathrm{Cov}^\top(\hat{x}_j, \hat{x}_o^\top)\, a_i^j + \mathrm{Cov}^\top(\hat{x}_{j+1}, \hat{x}_o^\top)\, a_i^{j+1} \Big) + \\
&\qquad a_i^{j\,\top} \cdot \Big(\mathrm{Cov}(\hat{x}_j, \hat{x}_o^\top)\, a_i^g + \mathrm{Cov}(\hat{x}_j, \hat{x}_j^\top)\, a_i^j + \mathrm{Cov}^\top(\hat{x}_{j+1}, \hat{x}_j^\top)\, a_i^{j+1} \Big) + \\
&\qquad a_i^{j+1\,\top} \cdot \Big(\mathrm{Cov}(\hat{x}_{j+1}, \hat{x}_o^\top)\, a_i^g + \mathrm{Cov}(\hat{x}_{j+1}, \hat{x}_j^\top)\, a_i^j + \mathrm{Cov}(\hat{x}_{j+1}, \hat{x}_{j+1}^\top)\, a_i^{j+1} \Big) \Big]
\end{aligned} \tag{8}$$

where a_i^g, a_i^j, a_i^{j+1} — part of the equation of condition for global and segmented part, j — the number of segment to which the i-th observation is belong.

At the last I would like to emphasize that we substituted *mathematical expectation of* R by its *value* when we calculated q. This operation is correct only if we have considerably large set of data. The uncertainty $\mathcal{E}(R_i)$ is likely to have the same order as the uncertainty of estimates of dispersion: $\dfrac{2}{\sqrt{n}}$. If we have 400 observations the uncertainty will be about 10%. Thus, it is excessive to adjust $\dfrac{R_i}{\mathcal{E}(R_i)}$ to unity with all precision which computer is able to provide. To have this ratio in interval [0.95, 1.05] is more than enough.

We should also take into consideration the fact that the assumptions put on the grounds of reweighting are rather poor. We neglect off-diagonal terms and we assume that all diagonal terms of a priori covariance matrix should be corrected by *the same* value. Unfortunately, covariance matrix of unmodeled systematic errors has non-negligible off-diagonal terms. We can expect improvement of the solution when we reduce ratio $\dfrac{R_i}{\mathcal{E}(R_i)}$ from, say, 5 to 1.5 but we hardly can expect substantial improvement reducing this ratio from 1.5 to 1.001. We should not expect too much from rather poor model. Thus, mentioned above interval may be too tight.