QUALITY ANALYSIS TESTING FOR QUANTIFYING RELIABILITY PROPERTIES

Edward Nowak
Warsaw University of Technology
Institute of Geodesy and Geodetic Astronomy

ABSTRACT

Quality control generally requires data of higher quality to quantify the error in any measured quantity. The following work shows that the Prószyński Reliability Criterion (PRC) is fulfilled if the measurement variance is greater than its prediction function based on the remaining observations.
The partial derivatives with respect to the parameters of the residuals' equations are taking the role of the observations within the system. Therefore the prediction variance can be measured. If the prediction function is bounded then a reliable observational system (i.e. each observation has fulfilled the PRC) can be achieved.

1. INTRODUCTION

A versatile, accurate and practical shape approximation imposes both the observational system and model approximation to be compliant with the reliability requirements. In the previous work (Nowak E., Nowak J., 2006), 3D shapes were modelled and analysed focusing on approximating the 3D shape based on polynomial interpolation (simple, orthogonal, Legendre' and Bernstein polynomials), spline interpolation, Bezier's functions, and the NURBS technique. The approximation was made on the basis of *n* evenly-distributed points that allow for model quality control within the measured set. It was shown that $n \geq 2u^2 - 2u + 3$ can be used as a reliability criterion for approximating 3D shapes based on polynomials and spline.

2. THE RELIABILITY CRITERION IN REGARDS TO THE QUALITY OF THE ADJUSTED OBSERVATIONS

The following analysis is limited to the case of the independent observations' adjustment $C_{ll} = diag(\sigma_i^2)$. A v_i / σ_i normalisation reduces it to the Least Square Method (LSM), which is a Euclidean metrics ($\min v^T v$, where the standardized residua equations $V = Ax + l$). Whereas the reliability matrix *R* is a variance-covariance matrix

of adjusted residua (referred also as to C_{vv} or $Cov(\hat{v})$), and $R = I - Q$; where $Q = Cov(\hat{l}) = A(A^T A)^{-} A^T$.

If the R matrix (Prószyński W., 1994) is known, then both the adjustment of the observational set $\hat{v} = Rl$ and the formulation of the Observation Reliability Criterion $R_{ii} > 1/2$ (1) are feasible. Taking advantage of the duality characteristic of the Least Squares Method (Perelmuter A., Nowak E., Prószyński W., 1994., Nowak E., Nowak, J., 2005). , the Reliability Criterion of the parametric subspace is given as:

$$Q_{ii} = \sigma^2 (\hat{\ }) = A_i (A^T A)^{-1} A_i^T < 1/2 \qquad \textbf{(1)}$$

The formula (1) may be interpreted as follows: the Reliability Criterion is fulfilled if the adjustment results in the observational variance are at least two times smaller than before the adjustment.

3. THE RELIABILITY OF THE AVERAGE OF THE REPEATED OBSERVATIONS

Let us begin with a simple example; several observations of the same measure. Then:

$$A = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad A^T A = n, Q_{ii} = \frac{1}{n}, \qquad \textbf{(1a)}$$

and the Reliability Criterion is expressed as follows: $n > 2$

The method of the observation validation based on its repetitions is well-known and not complicated but it requires more than two observations for each unknown. Such conclusion can be reached by the act of deducting, as follows:

- **We make the first observation ($n = 1$) and we are unable to validate it.**
- **We make the second observations ($n = 2$) of the same measure, then we are able to check if their difference is within the given tolerance but in case of intolerable difference, we are unable to distinguish the true value.**
- **We make the third observations ($n = 3$) of the same measure, and then we are able to detect the error in any measured quantity by differences' analysis.**

Thus the triple repetition of each of the observations allows for the observation validation regardless any regression modelling.

4. THE RELATIVE RELIABILITY CRITERION OF THE OBSERVATION

Let us begin with the analysis of the variances of the adjusted observations. Let us divide the set of observations into two subsets: the first one consists solely the observation being analysed (here denoted as b), the second subset consists of the all the remaining observations. Assuming that the b observation is the last one (the equivalent of the A_n), we can obtain

$$A = \begin{bmatrix} a \\ b \end{bmatrix} \qquad \textbf{(2)}$$

The establishment of the variance of the adjusted observation requires the $(A^T A)^{-1}$ matrix. In order to obtain it, we make use of the formula for the inverse of the given matrices sum (Rao C. Radhakrishna. 1982)

$$(A + BDB^T)^{-1} = A^{-1} - A^{-1}B(B^T A^{-1}B + D^{-1})^{-1} B^T A^{-1} \quad \textbf{(3)}$$

In our case the above formula is expressed as follows:

$$(A^T A)^{-1} = (a^t a + b^T b)^{-1} = (a^T a)^{-1} - (a^T a)^{-1} b^T (a^t a)^{-1} b + 1)^{-1} b (a^T a)^{-1} \quad \textbf{(4)}$$

Whereas the wanted variance (Nowak E., 1985) is formulated as:

$$Q_{nn} = \sigma^2 (l_n) = b(A^T A)^{-1} b^T = s^2 - s^2 (s^2 + 1)^{-1} s^2 = s^2 - \frac{s^4}{s^2 + 1} = \frac{s^2}{s^2 + 1}$$

where $s^2 = b(a^T a)^{-1} b^T$ **.** **(5)**

The s^2 parameter, mentioned in formula (5), is the prediction variance of the $f = l_n$ function, being determined using all the observations except the last (l_n), and it shows how these observations influence the adjustment of the l_n .

Applying (5) to the formula (1) we obtain

$$\frac{s^2}{s^2 + 1} < \frac{1}{2} \quad \textbf{(7)}$$

And finally, the Reliability Criterion can be described by the following inequality

$$s^2 < 1 \quad \textbf{(8)}$$

Not taking into account the normalisation, our analysis results in the (9) inequality

$$\frac{s^2}{s^2 + \sigma_n^2} < \frac{1}{2} \quad \textbf{(9)}$$

and subsequently $s^2 < \sigma_n^2$ **(10)**

Therefore, the Reliability Criterion is fulfilled if the observation variance is greater than its prediction function based on the remaining observations. Such prediction is used for the (new) observation validation since the new observation has to be less accurate from the standard given indirectly by the rest of the observations. By this conclusion, the relation with the classical control methods is shown there (Humienny Z., 2004).

Example 1.
Let take the mean of the observations that differ in accuracy as follows: $\sigma_1^2 \geq \sigma_2^2 \geq \sigma_3^2$**.**

Then the Reliability Criterion is given by the following inequality $\frac{\sigma_1^2 + \sigma_2^2}{4} < \sigma_3^2$**.**

5. THE RELIABILITY OF THE LINEAR APPROXIMATION (REGRESSION MODEL)

The linear regression model for the x_i, y_i **set can be written as:**

$$y = ax + b \tag{11}$$

The partial derivatives with respect to the parameters of the residuals' equations are taking the role of the observations within the system.

$$A_{i,1} = x_i, A_{i,2} = 1 \tag{12}$$

Therefore the linear regression model can be written in vector-matrix notation (as regards to the Least Squares Analysis developed by Carl Friedrich Gauss) as:

$$A^T A = \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \tag{13}$$

Coordinate system translation to the centre of the observational set ($\bar{x} = \sum x_i / n$ **) results in the elements out of the main diagonal set to 0. Thus the prediction variance for the new observation of the** x, y **point can be formulated as:**

$$s^2 = \frac{x^2}{\sum x_i^2} + \frac{1}{n}$$

The further the observational set center is, the greater the prediction varaiance is. One of its application is the determination of the unpredicted or unexplained variation in the response variable $\hat{y} \pm sg$ **, where the term *g* represents the limit of the error. The prediction variance leads to the following criterion**

$$\frac{x^2}{\sum x_i^2} + \frac{1}{n} < 1 \tag{14}$$

And finally $x^2 < \left(1 - \frac{1}{n}\right) \sum x_i^2$ **(14a)**

It means that the new observation cannot be very different from the remaining observations (that are regarded as its validation standard). Such conclusion was also shown in our prievious work (Nowak E., Nowak J., 2006).

Example 2.
Let take the line segment $-1 \le x \le 1$ and its measurements: one for each egde and one in the middle. And then we strength the set of the measurements by two additional observations at the $\pm x$ points. In such case, we have: $u = 2$, $n = 5$ and

$$A = \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 0 & 1 \\ -x & 1 \\ x & 1 \end{bmatrix} \qquad A^T A = \begin{bmatrix} 2+2x^2 & 0 \\ 0 & 5 \end{bmatrix}$$

Applying the Reliability Criterion (1) to the observations at the edges requires:

$$\frac{1}{2+2x^2} + \frac{1}{5} < \frac{1}{2}$$

Therefore $1 \ge x^2 > 2/3$ is the condition for reliable distribution of minimal observations' number ($n = 2u+1$).

Considering the more general case of the non-linear approximation, the equivalent conclusion can be drawn: the analogical limitation applies for the derivatives existent in the residual equations. It is an equivalent of the Lipszyc Condition for the differential equations (Demidowicz B.P., 1972).

6. THE RELIABILITY OF THE PERIODIC APPROXIMATION

We analyse the approximation of the $\{x_i, y_i\}$ set by periodic function given by the following formula:

$$y = a \sin(x + b) \tag{15}$$

The residuals equations variables are given by:

$$A_{i,1} = \sin(x_i + b), \qquad A_{i,2} = a\cos(x_i + b) \tag{16}$$

The condition for the elements out of the main diagonal being set to 0 (as regards to the Normal Gauss Equations) can be formulated as follows:

$$\sum \sin(x_i + b)\cos(x_i + b) = 0.$$

Then, expending the equations, we obtain

$$\sum (\sin x_i \cos b + \cos x_i \sin b)(\cos x_i \cos b - \sin x_i \sin b) = 0$$

And after the multiplication: $\cos 2b \sum \sin 2x_i + \sin 2b \sum \cos 2x_i = 0$. Thus the X-axis shift is defined us $\tan 2b = -\dfrac{\sum \sin 2x_i}{\sum \cos 2x_i}$ Therefore the Reliability Criterion for the new observation can be formulated as $\dfrac{\sin^2 x}{\sum \sin^2 x_i} + \dfrac{\cos^2 x}{\sum \cos^2 x_i} < 1$. And after the necessary expending:

$$\sin^2 x < \frac{1 - \frac{1}{\sum \cos^2 x_i}}{\frac{1}{\sum \sin^2 x_i} - \frac{1}{\sum \cos^2 x_i}} \qquad \textbf{(17)}$$

And finally, considering the fact that $\sin^2 x < 1$, we can express the Reliability Criterion as:

$$\sum \sin^2 x_i > 1 \qquad \textbf{(18)}$$

Drawing from specific cases for more general cases, one can state that if the observational equation derivatives are limited (bounded) then a reliable observational system (i.e. each observation has fulfilled the PRC) can be achieved.

Example 3.
Let take the approximation of the following periodic function
$y = a + b\sin x_i + c\cos x_i + d\sin 2x_i + e\cos 2x_i$

Table 1. A Matrix and the variances of the adjusted observations.

	x [grad]	A	b	C	d	e	Qii
1	0	1	1	0	1	0	5/12
2	30	1	$\sqrt{3}/2$	½	1/2	$\sqrt{3}/2$	5/12
3	60	1	1/2	$\sqrt{3}/2$	-1/2	$\sqrt{3}/2$	5/12
4	90	1	0	1	-1	0	5/12
5	120	1	-1/2	$\sqrt{3}/2$	-1/2	$-\sqrt{3}/2$	5/12
6	150	1	$-\sqrt{3}/2$	½	1/2	$-\sqrt{3}/2$	5/12
7	180	1	-1	0	1	0	5/12
8	210	1	$-\sqrt{3}/2$	-1/2	1/2	$\sqrt{3}/2$	5/12
9	240	1	-1/2	$-\sqrt{3}/2$	-1/2	$\sqrt{3}/2$	5/12
10	270	1	0	-1	-1	0	5/12
11	300	1	1/2	$-\sqrt{3}/2$	-1/2	$-\sqrt{3}/2$	5/12
12	330	1	$\sqrt{3}/2$	-1/2	1/2	$-\sqrt{3}/2$	5/12

Table 2. $A^T A$ Matrix.

12	0	0	0	0
0	6	0	0	0
0	0	6	0	0
0	0	0	6	0
0	0	0	0	6

The example is characterised by the following parameters $n = 12$, $u = 5$. Unlike for the polynomial approximations ($n \geq 2u^2 - 2u + 3 = 43$), the observations number is only a bit greater than double number of unknowns.

7. TECHNOLOGICAL REMARKS/CONCLUSIONS

The theoretical work presented above has significant practical consequences. Two specific issues are described below.

A. The assumption that observations are independent may be debatable in some cases, e.g. the computation of the average from a series of observations (see section 2) by stational adjustment. In order to obtain 'true' observational independence, one must repeat the entire measurement sequence by starting from point identification and instrument centering.

B. Practical application of the linear approximation (see section 4) involves the physical representation of the line segment (and not the whole line). In this case, the derivatives (expressed in coordinates) are bounded by the size of the measured object. By analogy, the approximation in the closed subspace (e.g. based on the polynomial line segment) produces bounded derivatives (similar to the periodical approximation). The practical consequence of these findings is that the reliable observational system is achievable (e.g. for approximating the shape of a building). The unavoidable issue is the cost of measurements that increases proportionally to the square of the polynomial degree (Nowak E., Nowak J., 2006). (in the case of evenly distributed observations). Moreover, since the derivatives are susceptible to measurement errors, the measurements characterised by the greatest absolute values of their derivatives have the greatest weight with respect to the reliability characteristics of the observational system. For polynomial approximations, these maxima are placed at the edges of the line segments. Therefore, the most economical method to achieve a reliable observational system is to repeat the measurements in vulnerable areas (i.e. edges). Similarly for the surface geodetic networks, the most error prone region is the network edge.

REFERENCES

Nowak, E., Nowak, J., (2006). Projektowanie niezawodnościowe aproksymacji modelami wielomianowymi i NURBS. Mat. Sesji Naukowej z okazji 85-lecia Wydziału Geodezji i Kartografii, Politechnika Warszawska Wydział Geodezji i Kartografii, Warszawa 2006.

Prószyński, W,. (1994). Criteria for internal reliability of linear least squares models. Bull. Géodésique 68.

Perelmuter, A., Nowak, E., Prószyński, W., (1994). Practical implication of duality of the least squares method. Perelmuter Workshop on Dynamic Deformation Models, Haifa, Izrael.

Nowak, E., Nowak, J., (2005). Boundary Properties of the Reliability Matrix. Reports on Geodesy IG&GA WUT, No.3(74) 2005 PL, ISSN 0867-3179, ISBN 83-85287-73-3, pp. 65-76.

Rao C. Radhakrishna. (1982). Modele liniowe statystyki matematycznej. PWN, Warszawa 1982, pp. 51.

Nowak E., (1985). Interactive network design and analysis based on succesive modifications of covariance matrix. Zesz. Nauk. AGH Geodezja z.90, 7th International Symposium on Geodetic Computations IUGG.

Humienny, Z., (2004): Specyfikacje geometrii wyrobów. Geometrical Product Specifications – GPS. Podręcznik europejski, WNT Warszawa.

Nowak, E., Nowak, J., (2006). „Projektowanie niezawodnościowe aproksymacji modelami wielomianowymi i NURBS". Mat. Sesji Naukowej z okazji 85-lecia Wydziału Geodezji i Kartografii, Politechnika Warszawska Wydz. GiK Warszawa.

Demidowicz B.P., (1972). Matematyczna teoria stabilności. WNT Warszawa.

Acknowledgments

The presented work is a part of the Polish Ministry of Science and Higher Education research project no. 4T12E 044 28 (Doskonalenie metod projektowania systemów geodezyjnego monitorowania przemieszczeń o różnej specyfice i stopniu intensywności).