Presented Paper to Commission III, 15th ISPRS Congress, Rio de Janeiro, 1984

AUTOMATIC GROSS ERROR DETECTION BY ROBUST ESTIMATORS

Hans Werner

Abstract: The theoretical foundation of robust estimators, which are unaffected by gross errors, is presented. A way is shown how to robustify the least squares method against gross errors with weight functions. Finally, some empirical results on the elimination of gross errors by robust adjustment, concerning horizontal blockadjustment with independent models, are shown.

Zusammenfassung: Die theoretischen Grundlagen von Robusten Schätzern, die unempfindlich gegen grobe Fehler sind, werden vorgestellt. Es wird gezeigt, wie die Ausgleichung nach der Methode der kleinsten Quadrate durch Einführung von Gewichtsfunktionen robust gegen grobe Fehler werden kann. Einige Ergebnisse empirischer Untersuchungen zur Elimination grober Fehler mit einer robusten Ausgleichung anhand der Blockausgleichung mit unabhängigen Modellen (Lage) werden erläutert.

Résumé: Dans l'article suivant seront présentées les bases théoriques d'estimateurs robustes insensibles à de fautes. On montrera comment la compensation suivant la méthode des moindres carrées peut être rendue robuste à de fautes par l'introduction de fonctions de poids. On expliquera quelques résultats de recherches préliminaires empiriques pour éliminer de fautes par une compensation robuste à l'aide de la compensation planimétrique de blocs avec des modèles indépendents.

1. INTRODUCTION

Since it is possible in geodesy, by application of efficient computer programs, to treat large data-sets in Least-Squares-Adjustments, the automatic gross error detection has become more and more urgent, because manual error detection in such cases is lengthy and difficult.

The recent error-searching-methods in three steps of mathematical modelling, adjustment, and test of the results presuppose a certain statistic distribution (normal distribution at most) of the data. But this assumption is not fulfilled if gross errors exist, so the statistic tests may fail sometimes.

A possible way out is the application of robust estimators, which are largely unaffected by gross errors. This insensitivity against gross errors is attainable in one step by modification of well-known methods.

We will call any method as robust, when it allows to get a result unaffected by gross errors. The iterative manual procedure of error detecting by adjustment, examination of the residuals, rejection of observations and new adjustment is in this meaning a robust method. Contrary to that, however, robust adjustment should be fully automatically, without any manual operations.

At this point we can consider as the main aim of robust estimation:

1. isolating clear outliers,
2. building in safeguards against unsuspectedly large amounts of gross errors,
3. putting a bound on the influence of hidden contamination and questionable outliers, and
4. still being nearly optimal at the strict parametric model (see Hampel, 1973, p.91).

The detection of very large gross errors and error clusters has to be done by means of plausibility controls, for example sequential methods like searching for groups of correct observations (see Weller, 1982). Medium gross errors can be eliminated by robust adjustment (Werner, 1982), while small gross errors can be detected with (statistical) tests (Klein, Förstner, 1981). Since the used robust estimator is sensitiv enough, it is under favourable circumstances no longer necessary to subsequently test the results of a robust adjustment, because a robust estimator can operate for small gross errors like a statistical test.

First of all, we will reflect upon the Least-Squares-Adjustment in the next chapter in a more general frame.

2. ESTIMATION OF UNKNOWN PARAMETERS

With a given sample (e.g. photogrammetric model coordinates), taken from a population, unknown parameters are to be estimated. A stochastic and a functional model is necessary in order to estimate these unknowns by means of an estimation function. With this estimator we get approximations of the unknowns. These approximations are in a sense optimal, depending on the actual purpose of the estimation method.

Starting with the observation equations

$$l+v = Ax \quad \text{or} \quad l_i+v_i = \sum_{j=1}^{h} a_{ij}x_j \; ; \qquad i = 1\ldots n \tag{2.1}$$

the Least-Squares-Adjustment (Gauss-Markov-Model) will minimize the weighted sum of the residual squares:

$$v^T P v = \min \quad \text{or} \quad \sum_{i=1}^{n} p_i \,(l_i - \sum_{j=1}^{h} a_{ij}x_j)^2 = \sum_{i=1}^{n} p_i v_i^2 = \min \tag{2.2}$$

which leads to the normal equations

$$A^T P A x - A^T P l = 0 \quad \text{or} \quad \sum_{i=1}^{n} p_i \,(l_i - \sum_{k=1}^{h} a_{ik}x_k)a_{ij} = 0 \; ; \; j = 1\ldots h \tag{2.3}$$

The solution is called Least-Squares-Estimation of the parameters

$$\hat{x} = (A^T P A)^{-1} A^T P l \tag{2.4}$$

and of the adjusted observations

$$\hat{l} = A(A^T P A)^{-1} A^T P l \tag{2.5}$$

This method may be understood as a special case of a more general estimation method, defined as follows:

"Each estimation function $\hat{\theta}_n$, defined by a minimum problem of the form

$$\Sigma\, \rho(l,\hat{\theta}_n) = \min \tag{2.6}$$

or by an implicite equation

$$\Sigma\, \psi(l,\hat{\theta}_n) = 0 \tag{2.7}$$

l: observation vector
$\hat{\theta}_n$: estimation function

with

$$\psi(1,\hat{\theta}_i) = \partial\rho(1,\theta_i)/\partial\theta_i \tag{2.8}$$

is called M-estimate (or Maximum likelihood type estimator)"(see Huber, 1981, p. 43 or Rey, 1983, p.90).

If we choose for the estimation function ρ in eq. (2.6) a logarithmic density function

$$\rho(x,\hat{\theta}) = \ln f(x) \tag{2.9}$$

we obtain the ordinary Maximum-Likelihood-Estimator:

$$\Sigma \ln f(x) = \max \tag{2.10}$$

or

$$\Sigma \frac{\partial}{\partial x_i} \ln f(x_i) = 0 \tag{2.11}$$

In the group of M-Estimates exist robust ones which are less sensitive against outliers than the Least-Squares-Adjustment. In the next chapter we will modify the Least-Squares-Adjustment in a simple way and robustify it against gross errors.

3. ROBUST ESTIMATORS

In order to robustify the M-Estimators against gross errors and to gain the aims of robust estimation (see above), for $\rho(1, \hat{\theta}_n)$ or $\psi(1, \hat{\theta}_n)$, respectively, in eq. (2.6) and (2.7) functions are used, which are less sensitive against gross errors. The sensitivity against gross errors of functions ρ or ψ respectively, can be shown by means of the influence function. The influence function indicates the effects of outliers on the results and is proportional to the function ψ (see Hampel, 1973, p.98; Huber, 1981, p. 45 or Werner, 1982, p. 14). In other words: with the selection of a function ψ we can design robust M-Estimators with certain properties.

An M-Estimator can be robustified by the following arrangements (see Fig. 3.1):

① by putting a bound on $|\psi|$: gross-error-sensitivity,

② by putting a bound on the absolute value of the slope: local-shift-sensitivity, the sensitivity against small shifts between model and estimation,

③ by setting $\Psi = 0$ outside a certain point: outliers, located beyond this point will be eliminated,

④ by going down to zero (in the location case) on a certain hyperbolic tangent, the change-of-variance-sensitivity: small changes of observation errors shall not cause large changes of the influence on the results and

⑤ by putting (in the same case) a stricter bound from below on the slope in order to minimize the number of iteration steps (see Hampel, 1973, p. 98).

The following diagram serves to illustrate these conditions: the function ψ should be designed so, that it will not leave the dotted area.

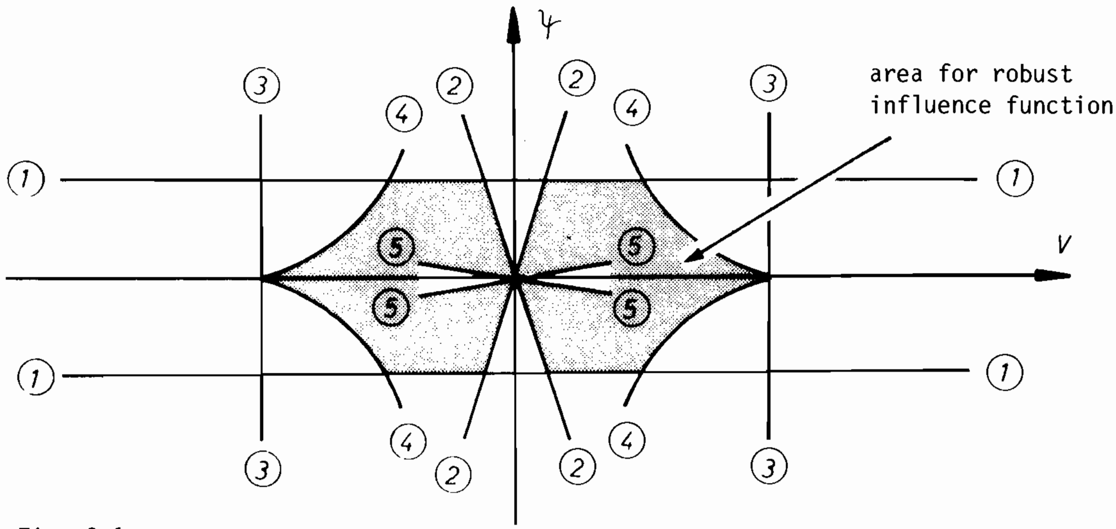

Fig. 3.1

To robustify the Least-Squares-Method for practical application we use variable observation weights. In the case of ordinary Least-Squares-Adjustment we minimize:

$$\sum_{i=1}^{n} \rho(v_i) = \min \qquad \text{(see eq.(2.6))} \tag{3.1}$$

with

$$\rho(v_i) = p_i \; v_i^2 \tag{3.2}$$

where the weights p_i are constant during the whole adjustment. For Robust-Least-Squares-Adjustment we choose weights as a function of the residuals v_i :

$$\rho(v_i) = p(v_i) \; v_i^2 \tag{3.3}$$

The first step of iteration is a conventional Least-Squares-Adjustment with constant weights. New observation weights are computed from the residuals of the previous adjustment. With these new weights a next adjustment is started and this procedure is repeated 5...20 times. At last observations contaminated with gross errors obtain the weight $p_i = 0$ and their residuals are a measure for the magnitude of the gross errors (see Krarup et al, 1980, p. 373 and Werner, 1982, p. 16) .

In order to gain such weight functions $p(v_i)$ we have two alternatives:

1. choice of a function $\rho(v)$ or $\psi(v) = \partial\rho(v)/\partial v$ and derivation of a weight function by transforming with eq. (3.3):

$$p(v) = \frac{\rho(v)}{v^2+c} \quad ; \quad c<<1 \tag{3.4}$$

   (c is a small constant to avoid division by zero in case of exactly $v_i = 0$; see Krarup, 1980, p. 378).

2. choice of a weight function $p(v)$ and verification of it's robustness with eq. (3.3) and eq. (2.8).

For example the Minimum-Norm-Method uses the following functions:

$$\rho(v) = |v|^q \quad ; \quad 0 \le q \le 2 \tag{3.5}$$

$$\psi(v) = \text{sign}(v)\; q\; |v|^{(q-1)} \tag{3.6}$$

$$p(v) = \frac{1}{|v|^{(2-q)}+c} \quad ; \quad c<<1 \tag{3.7}$$

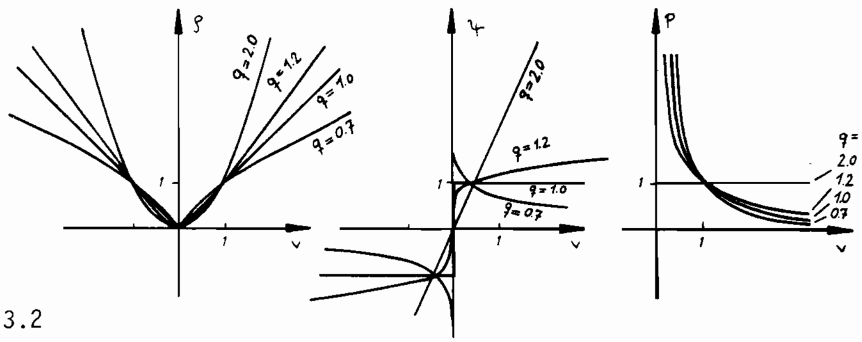

Fig. 3.2

For $q \approx 1.2$ the conditions ① and ② are nearly fulfilled, but not the other ones (only poor robust characteristics). For q=1 all errors cause the same influence, independent from their size. Finally for q=2 we get a non-robust Least-Squares-Adjustment with constant weights, wherein the influence of an error on the adjustment results grows proportional to the size of the error.

While under the Minimum-Norm-Method the weight functions are independent of the number of iteration steps and of precision of observations, under the Danish Method (see Krarup et al, 1980, p. 374 or Krarup/Juhl, 1983, p. 132) the weight functions are controlled by the global precision σ and the number of iteration steps:

first iteration: $p = 1$ (3.8)

2^{nd} and 3^{rd} iteration: $p = \exp(-0.05(\frac{v}{\sigma})^{4.4})$ (3.9)

following iterations: $p = \exp(-0.05(\frac{v}{\sigma})^{3.0})$ (3.10)

If we generalize the eq. (3.9) and eq. (3.10), we have

$$p_i = \exp(-\alpha\; |v_i|)^d \tag{3.11}$$

with

$$\alpha = \frac{1}{\sigma_{v_i}\; k} = \frac{1}{\sqrt{r_i}\; \sigma_{l_i}\; k} \tag{3.12}$$

and

$$\left.\begin{array}{lll} k = k(\frac{\hat{\sigma}_o}{\sigma_o}) \;; & \min,\max(k) = 1.0,\; 6.0 & \hat{\sigma}_o\text{: estimated precision} \\ d = d(\frac{\hat{\sigma}_o}{\sigma_o}) \;; & \min,\max(d) = 1.0,\; 10.0 & \sigma_o\text{: } \sigma_o\text{-a-priori} \end{array}\right\} \tag{3.13}$$

The new weights are dependent of the residuals v_i, their precision σ_{v_i} and the ratio between estimated precision $\hat{\sigma}_o$ and σ_o-a-priori (Werner, 1982, p.20 and p.50). Since eq. (3.11) leads fast to very small weights, Werner (1982, p.20) used besides the exponential weight functions eq. (3.11) a hyperbolic function:

$$p = \frac{1}{1+(\alpha\ |v|)^d} \tag{3.14}$$

The influence functions, belonging to eq. (3.11) and (3.14), are fulfilling the conditions ① - ⑤ (see above), or differ from them only slightly. Therefore sufficient robust characteristics are reached, as shown in the next chapter.

In order to gain an unobjectionable Least-Squares-Solution, at the end of the robust adjustment, each observation with weight p_i less than a bound p_o, will be eliminated by setting the weight to zero: $p_i = 0$. The observations with weight p_i greater than the bound p_o will be used in the last Least-Squares-iterationstep with their a-priori-weight:

$$p_i^{(\nu)} < p_o \quad : \quad p_i^{(\nu+1)} = 0 \tag{3.15}$$

$$p_i^{(\nu)} \geq p_o \quad : \quad p_i^{(\nu+1)} = p_i - a - priori \tag{3.16}$$

$(\nu+1)$: last iteration step

The experimental program SNOOPY (see Werner, 1982), which enables a robust block adjustment with independent models for photogrammetric horizontal blocks, was written to demonstrate the practicability of a robust adjustment for the elimination of medium and small gross errors. A few experimental results are presented in the next chapter.

4. APPLICATION TO HORIZONTAL BLOCK ADJUSTMENT WITH INDEPENDENT MODELS

The horizontal block adjustment has been chosen, because it is significant twice for photogrammetry: at first it enables to compute the horizontal coordinates in a two-step block adjustment (horizontal/vertical); secondly the horizontal block adjustment can serve to compute approximative horizontal coordinates in a one-step block adjustment with bundles.

Three questions had to be answered to realize a robust adjustment with the program SNOOPY:

1. choice of parameters for the weight functions,
2. break up point and convergence behaviour of the iterative adjustment process, and
3. the consideration of the geometry by means of redundancy numbers r_i .

Only a few results of the investigations concerning these questions can be presented here, for further details see Werner (1982).

The parameters in eq. (3.11)-(3.14) can be thought as "sensitivity screws" of the robust adjustment. The essential problem at this point is to find a compromise between a too weak adjustment, where too few observations contaminated with gross errors are eliminated, and a too strict adjustment, where too many observations are marked as wrong. An independence of the parameters from the adjustment problem and from actual data could not be obtained.

					results				
block	points per model	$\bar{r}_i = \frac{r}{n}$: average redundancy numbers	consideration of geometry by means of r_i	number of gross errors in block	number of iterations	number of eliminated gross errors	number of <u>not</u> eliminated gross errors	number of eliminated correct observations	number of false decisions (sum of columns 8 and 9)
1	2	3	4	5	6	7	8	9	10
A1	4	o.22	yes	10	22	9	1	4	5
A2			no		23	-	-	10	10
B1	6	o.31	yes	9	14	9	-	1	1
B2			no		20	8	1	1	2
C1	~15	o.46	yes	7	8	7	-	-	-
C2			no		13	7	-	-	-

<u>Table 4.1</u> contains the results of robust adjustment concerning three different horizontal blocks with increasing number of points per model (columns 2 and 3). Each block has been adjusted with and without consideration of geometry by means of redundancy numbers r_i (see col. 4).

The horizontal block adjustment with independent models is normally treated as a linear problem. However, strictly viewed the observation equations are nonlinear, because the unknowns are connected by a product:

$$\begin{aligned} a(x^1+v_{x^1}) + b(x^2+v_{x^2}) + c - y^1 &= 0 \\ -b(x^1+v_{x^1}) + a(x^2+v_{x^2}) + d - y^2 &= 0 \end{aligned} \qquad (4.1)$$

with the observations x^1, x^2 : measured model coordinates

and the unknowns y^1, y^2 : adjusted ground coordinates

a,b,c,d: transformation parameters

Therefore we have to iterate and must use approximation values for the unknowns. In the case of ordinary Least-Squares-Adjustment it converges after three iterationsteps at most. Usual the iterations will be ended, if the maximum addition to the approximative coordinates is less than a given bound (convergence) or if the addition will grow (divergence). Concerning the robust adjustment the maximum addition may vary by reason of different weights per iteration step. In this case the addition is not able to show convergence or divergence.

Hence we have to change the end-of-iteration-question:

1. In the case of robust adjustment $\hat{\sigma}_o$ will converge, because after each iteration step more gross errors will be eliminated. As long as the estimated weight unity $\hat{\sigma}_o$ is larger than the σ_o-a-priori, the iteration shall be continued.
2. As convergence criterion the additions must be standardized by their precision:

$$\overline{\Delta x}_{max} = \max_j \left(\frac{|\Delta x_j|}{\sigma_{\Delta x_j}}\right) \qquad (4.2)$$

3. The minimum and maximum number of iteration steps should be ordered by the operator.

The number of false decisions increases with increasing number of iterations, as shown in table 4.1, column 6 and 10. By reason of the poorly defined boundary between small gross errors and large random errors the end of iterations is problematical (see Klein, 1984).

As table 4.1 shows, the density of points per model is essential for the gross error elimination: the larger the (average) redundancy numbers, the better are gross errors eliminated (table 4.1, col. 3,6,10). At the same time it is more and more important to use redundancy numbers, the fewer points a block has and the less homogeneous the observations are distributed. This was verified for photogrammetric measurements only, but it is certainly valid also for geodetic net adjustments.

The computation of redundancy numbers causes organizational and economic problems (e.g. computer storage is limited, costs for computing time...), and so in practice we have to choose between two alternatives: to invest time and money in the program to compute the redundancy numbers or to optimize the measurements and observation design.

5. CONCLUSION

As shown above the ordinary Least-Squares-Adjustment can be made insensitive (robust) against gross errors by means of weight functions and can thus be used to identify and eliminate outliers. After such robust adjustment automatically the weight of the remaining observations will be set to their a-priori-weight, whilst the weight for outliers is set to zero. With such weights the final Least-Squares-Adjustment will give the strict Least-Squares-Solutions. However, the choice of parameters for the weight functions and the number of iteration steps in the robust adjustment is still problematic.

From a computer program for robust adjustment we get only preliminary decisions, for the time being. The final decision, to accept the results of the automatic gross error detection, is still to be made by the operator.

The method proposed above describes the stage reached in the middle of 1983 and has been integrated and further developed in the block adjustment program PAT-M43 without explicit computation of redundancy numbers by H. Klein at the Photogrammetric Institute in Stuttgart (Klein H., Förstner W., 1984). First results have been demonstrated in fall 1983 at the 39th Photogrammetric Week in Stuttgart (Klein H., 1983), moreover Klein H. and Förstner W. submitted new results as a Presented Paper at the XVth International Congress for Photogrammetry and Remote Sensing in Rio de Janeiro.

REFERENCES

Hampel F.R. (1973): Robust Estimation: A Condensed Partial Survey. Zur Wahrscheinlichkeitstheorie, 1973

Huber P.J. (1981): Robust Statistics. New York, 1981

Klein H. (1983): Automatische Elimination grober Datenfehler im erweiterten Blockausgleichungsprogramm PAT-M. 39th Photogrammetric Week, Stuttgart 1983

Klein H., Förstner W. (1981): Strategien zur Suche grober Fehler. Photogrammetric Institute, Stuttgart University, Schriftenreihe Heft 7, 1981

Klein H., Förstner W. (1984): Realization of Automatic Error Detection in the Block-Adjustment Program PAT-M Using Robust Estimators. XVth ISPRS Congress, Rio de Janeiro 1984

Krarup, Juhl, Kubik (1980): Götterdämmerung over Least-Squares-Adjustment. XVth ISP Congress, Hamburg 1980

Krarup, Juhl (1983): The Danish Method; Experience and Philosophy. Deutsche Geodätische Kommission, Reihe A, Heft 98, München 1983

Rey W.J.J. (1983): Introduction to Robust and Quasi-Robust Statistical Methods. Berlin 1983

Weller M. (1982): Automatische Suche grober Fehler in Lageblöcken. Diplomarbeit, Photogrammetric Institute, Stuttgart University, 1982, not published

Werner H. (1982): Robuste Schätzer zur automatischen Suche grober Fehler bei der Lageblockausgleichung mit unabhängigen Modellen. Diplomarbeit, Photogrammetric Institute, Stuttgart University, 1982, not published.