K. KUBIK
ITC – Delft.

THE ESTIMATION OF THE WEIGHTS OF MEASURED QUANTITIES WITHIN THE METHOD OF LEAST SQUARES

1. – Introduction

This paper proposes an extension of the familiar method of least squares : for in addition to the more commonly computed unknowns, the variances of the original measurements and the relationships between these variances are estimated in this method.

The method of least squares serves to estimate the unknown magnitudes in linear models and the essential prerequisite for this is the knowledge of the "weights" of the measured quantities. Frequently, however, these weights are not known adequately and to establish them hypothetically inevitably leads to systematic deviations in the results, the limits of which are difficult to control. In these circumstances it is advisable to estimate the weights simultaneously with the other unknowns in the model. And such is the purpose of this method.

Apart from being regarded as an extended method of least squares, it may also be seen as an extension of those ideas of Helmert (1924), who proposed estimating the weights within the context of the method of least squares.

The method discussed in this paper has wide applicability although it was especially developed for investigations into geodetic problems : so it may be used for example for estimating the weights of direction and distance measurements in geodetic networks. It may also be applied to estimate the weights of relative and absolute gravity measurements in gravimetric levelling.

In the next section (2) those basic premises, that are needed for the later sections, are stated. Since this proposed method has many similarities to the usual method of least squares, the latter is shortly reviewed in section 3. The detailed discussion of the extended method is made in section 4. The estimators for the weights and the other unknowns in the model are derived by the method of maximum likelihood. Two different algorithms for estimating the weights are given. The estimators derived from the first algorithm are "biased" for a certain group of models. The second algorithm always yields unbiased estimators, however, the calculations are often more extensive. To complete the paper a numerical example is given in section 5 and serves to illustrate how this method can be used for

estimating the weights of direction and distance measurements, within a geodetic network.

2. – Assumptions and basic relations

First the assumptions which are needed for the later sections of the paper, shall be stated. The following linear model for a physical system is considered :

$$\mathbf{A} \cdot \mathbf{v} + \mathbf{B} \cdot \mathbf{x} + \mathbf{w} = \mathbf{0} . \tag{1}$$

The vector $\mathbf{v}$ is a vector of random variables $[\mathbf{v1}, \mathbf{v2}, \ldots \mathbf{vn}]$, normally distributed with mean : $\mathbf{E}(\mathbf{v}) = \mathbf{0}$ and variance matrix $\mathbf{V}(\mathbf{v}) = \mathbf{Q}$. The vector $\mathbf{x} = [\mathbf{x1} \ldots \mathbf{xm}]$ represents non–stochastic magnitudes. The matrices $\mathbf{A}$ and $\mathbf{B}$ are matrices of known coefficients. The random vector $\mathbf{w} = [\mathbf{w1} \ldots \mathbf{wk}]$ is measurable within this model, which in turn means thas a sample $\mathbf{W} = [\mathbf{W1}, \mathbf{W2}, \ldots \mathbf{Wk}]$ of measurements can be taken from $\mathbf{w}$.

In the following $\mathbf{Q}$ will be assumed to be a diagonal matrix of the type shown in figure (2). It follows from this figure, that the vector $\mathbf{v}$ can be partitioned into subvectors $\mathbf{v1}, \mathbf{v2}, \ldots \mathbf{vt}$, so that the variance of every element in the vector $\mathbf{vi}$ $(\mathbf{i} = 1 \ldots \mathbf{t})$ is $\mathbf{qi}$. Moreover, no elements in the vector $\mathbf{v}$ are correlated. The matrix $\mathbf{A}$ may be partitioned according to $\mathbf{v}$ and consequently equation (1) may also be written as follows :

$$\mathbf{A1} \cdot \mathbf{v1} + \mathbf{A2} \cdot \mathbf{v2} + \mathbf{A3} \cdot \mathbf{v3} + \ldots \mathbf{At} \cdot \mathbf{vt} + \mathbf{B} \cdot \mathbf{x} + \mathbf{w} = \mathbf{0}. \tag{3}$$

The reciprocals of the variances $\mathbf{qi}$ $(\mathbf{i} = 1 \ldots \mathbf{t})$ are usually referred to as "weights".

The model defined by (2) and (3) is valid, for example, in a geodetic network were both directions and distances have been measured. In this case $\mathbf{v1}$ represents the corrections to the direction measurements, $\mathbf{v2}$ represents the corrections to the distance measurements and $\mathbf{x}$ symbolizes the unknown point coordinates in the network.

Since the random vector $\mathbf{v}$ is normally distributed, every linear function $(\mathbf{T}.\mathbf{v} + \mathbf{C})$ of $\mathbf{v}$ will be also normally distributed with variance matrix $(\mathbf{T}^* \mathbf{QT})$ and mean $\mathbf{T}.\mathbf{E}(\mathbf{v}) + \mathbf{C}$ (Wilks, 1962, p. 158).

Consequently the vector $\mathbf{w}$ is also normally distributed with the variance matrix $\mathbf{AQA}^* = \mathbf{N}$ and mean $\mathbf{E}(\mathbf{w}) = \mathbf{B}.\mathbf{x}$, i.e.

$$\mathbf{pr}(\mathbf{w}) = \frac{1}{(2\pi)^{k/2}} \quad \frac{1}{|\mathbf{N}|^{\frac{1}{2}}} \quad \exp -\frac{1}{2} (\mathbf{w} + \mathbf{Bx})^* \ \mathbf{N}^{-1} \ (\mathbf{w} + \mathbf{Bx}) ; \tag{4}$$

$|\cdot|$ symbol for determinant

3. – The least squares method

In this familiar method the unknown magnitudes – the vectors $\mathbf{v}$ and $\mathbf{x}$ – are estimated from the measured sample $\mathbf{W}$. The estimators for $\mathbf{v}$ and $\mathbf{x}$ are consequently functions of $\mathbf{W}$ and are denoted by $\tilde{\mathbf{v}} = \tilde{\mathbf{v}}(\mathbf{W})$ and $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}(\mathbf{W})$. These least squares estimators are obtained by minimizing the sum of the squares :

$$\tilde{\mathbf{v}}^* \mathbf{G}^{-1} \tilde{\mathbf{v}},$$

where the matrix $\mathbf{G}$ differs from $\mathbf{Q}$ by an arbitrary multiple factor σ_0^2. As a constraint, the estimators have to satisfy equation (1) :

$$\mathbf{A}\tilde{\mathbf{v}} + \mathbf{B}\tilde{\mathbf{x}} + \mathbf{W} = \mathbf{0}. \tag{5}$$

This problem of minimization can be readily reduced to the problem of solving the following linear equation system (normal equation system) :

$$\boxed{\begin{array}{l} \bar{\mathbf{N}}.\mathbf{k} + \mathbf{B}.\tilde{\mathbf{x}} + \mathbf{W} = \mathbf{0} \\ \mathbf{B}^*.\mathbf{k} \qquad\qquad = \mathbf{0} \\ \tilde{\mathbf{v}} = \mathbf{G}\mathbf{A}^*\mathbf{k}, \quad \bar{\mathbf{N}} = \mathbf{A}\mathbf{G}\mathbf{A}^* \end{array}} \tag{6}$$

There the vector $\mathbf{k}$ represents Lagrange multipliers which are caused by the condition equations (5). The least squares method applied in model (1) is referred to as standard problem IV (according to Tienstra, 1956). Another commonly used model is : $\mathbf{Av} + \mathbf{w} = \mathbf{0}$ (where $\mathbf{B}$ is not defined). The least square adjustment within this model is referred to as standard problem I ; and the least squares method applied in the model $\mathbf{Bx} + \mathbf{w} = \mathbf{v}$ (where $\mathbf{A}$ is a unit matrix) is called standard problem II.

After computing the magnitudes $\tilde{\mathbf{v}}$, a factor

$$\tilde{\sigma}_0^2 = \frac{\tilde{\mathbf{v}}^* \mathbf{G}^{-1} \tilde{\mathbf{v}}}{\mathbf{f}}$$

($\mathbf{f}$: degrees of freedom in the problem)

may be estimated, which, when used with the matrix $\mathbf{G}$, gives an estimated value for the variance matrix of $\mathbf{v}$, $\tilde{\mathbf{V}}(\mathbf{v}) = \tilde{\sigma}_0^2 . \mathbf{G}$.

However, when using the method of least squares, the weight relations between random variables $\mathbf{v}$, or groups of random variables cannot be estimated ; for this method the a priori knowledge of the weight relations is required. Frequently

they are, however, not known with sufficient accuracy and so it is desirable to estimate them too, from the available information. In the following section methods for doing this will be presented.

4. — The extended least squares method

4.1. The objective of this extended method is to deduce estimators $\tilde{\mathbf{x}}$, $\tilde{\mathbf{v}}$, and $\tilde{\mathbf{q}}$ for the magnitudes $\mathbf{x}$, $\mathbf{v}$, and $\mathbf{q} = [\, \mathbf{q1}, \mathbf{q2}, \ldots \mathbf{qt} \,]$, which are assumed to be unknown in the model (1), (2). The estimators are functions of the measurable random variables $\mathbf{w}$ and they are deduced using the principle of maximum likelihood. This principles states that those numerical values are selected for $\tilde{\mathbf{x}}$, $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{q}}$, that maximize the "likelihood function"

$$\mathbf{L} = \mathbf{pr}\,(\mathbf{w} = \mathbf{W}\,;\, \tilde{\mathbf{x}}\,,\, \tilde{\mathbf{q}})\,. \tag{7}$$

In other words, we choose those values $\tilde{\mathbf{x}}$, $\tilde{\mathbf{v}}$, and $\tilde{\mathbf{q}}$ which will make it most probable that $\mathbf{w} = \mathbf{W}$. For numerical reasons, however, the natural logarithm of $\mathbf{L}$ is frequently made to a maximum instead of $\mathbf{L}$ itself, i.e.

$$\ln \mathbf{pr}\,(\mathbf{W}\,;\, \tilde{\mathbf{q}}, \tilde{\mathbf{x}},) = \text{const}, -\tfrac{1}{2} \ln |\,\mathbf{N}\,| \; -\tfrac{1}{2}\,(\mathbf{W} + \mathbf{B}\tilde{\mathbf{x}})^{*}\,\mathbf{N}^{-1}\,(\mathbf{W} + \mathbf{B}\tilde{\mathbf{x}}) \rightarrow \max. \tag{8}$$

This problem of maximization is identical to problem (7). It is solved by equating the first partial derivatives of $\ln \mathbf{pr}\,(\mathbf{W})$ to zero :

$$\frac{\partial \ln \mathbf{pr}\,(\mathbf{W})}{\partial \tilde{\mathbf{q}}} = \mathbf{0}\,; \quad \frac{\partial \ln \mathbf{pr}\,(\mathbf{W})}{\partial \tilde{\mathbf{x}}} = \mathbf{0}\,. \tag{9}$$

For computing these derivatives the following theorem must be used :

Theorem 1 : Let the elements $\mathbf{n}^{\alpha}_{\beta}$ of a matrix $\mathbf{N}$ be functions of elements $\mathbf{qj}$ of the vector $\mathbf{q}$. Further, let $\mathbf{N}$ be bounded and assume that both $\dfrac{\partial \mathbf{N}}{\partial \mathbf{q}}$ and $\mathbf{N}^{-1}$ exist. Then the following relations are valid :

$$\frac{\partial \mathbf{N}^{-1}}{\partial \mathbf{q}} = -\mathbf{N}^{-1} \quad \frac{\partial \mathbf{N}}{\partial \mathbf{q}}\, \mathbf{N}^{-1}\,, \tag{10a}$$

$$\frac{\partial \ln |\mathbf{N}|}{\partial \mathbf{q}} = \mathbf{tr}\left(\mathbf{N}^{-1} \; \frac{\partial \mathbf{N}}{\partial \mathbf{q}}\right); \tag{10b}$$

tr trace of a matrix ; sum of the diagonal elements of the matrix.

Proof : From the definition of the inverse of a matrix : $\mathbf{N}\,\mathbf{N}^{-1} = \mathbf{E}$ ($\mathbf{E}$ unit matrix) it follows :

$$\frac{\partial \mathbf{N}}{\partial q} \cdot \mathbf{N}^{-1} + \mathbf{N} \frac{\partial \mathbf{N}^{-1}}{\partial q} = \mathbf{0}. \tag{11}$$

The expression (10a) can now be readily verified by pre–multiplying (11) with $\mathbf{N}^{-1}$.

The second relation (10b) derives from the definition of the determinant :

$$|\mathbf{N}| = \sum_{a_1 \ldots a_m} n^1_{a_1} \; n^2_{a_2} \ldots\ldots n^m_{a_m} .$$

Under the above assumptions both $|\mathbf{N}|$ and $\ln|\mathbf{N}|$ are differentiable and the first derivative of the latter becomes

$$\frac{\partial \ln|\mathbf{N}|}{\partial q} = \frac{1}{|\mathbf{N}|} \cdot \frac{\partial |\mathbf{N}|}{\partial q} = \frac{1}{|\mathbf{N}|} \left\{ \sum_{a_1 \ldots a_m} \frac{\partial n^1_{a_1}}{\partial q} \cdot n^2_{a_2} \ldots n^m_{a_m} + \ldots \right.$$

$$\left. \sum_{a_1 \ldots a_m} n^1_{a_1} \ldots\ldots n^{m-1}_{a_{m-1}} \frac{\partial n^m_{a_m}}{\partial q} \right\} . \tag{12}$$

The algebraic sums

$$\sum_{a_2 \ldots a_m} n^2_{a_2} \ldots n^m_{a_m} \; ; \qquad \sum_{a_1 \ldots a_{m-1}} n^1_{a_1} \ldots n^{m-1}_{a_{m-1}}$$

just form the minor $\left|\mathbf{N}^1_{a_1}\right| \ldots \left|\mathbf{N}^m_{a_m}\right|$ of $\mathbf{N}$ and furthermore the expressions $\frac{\left|\mathbf{N}^{\alpha}_{\beta}\right|}{|\mathbf{N}|}$ are identical to the elements $\mathbf{I}^{\alpha}_{\beta}$ of the inverse $\mathbf{N}^{-1}$. Therefore this relation (12) may be written as follows :

$$\frac{\partial \ln|\mathbf{N}|}{\partial q} = \sum_{a_1} \frac{\partial n^1_{a_1}}{\partial q} I^1_{a_1} + \sum_{a_2} \frac{\partial n^2_{a_2}}{\partial q} I^2_{a_2} + \ldots + \sum_{a_m} \frac{\partial n^m_{a_m}}{\partial q} I^m_{a_m}$$

$$= \mathrm{tr}\left(\frac{\partial \mathbf{N}}{\partial q} \cdot \mathbf{N}^{-1} \right) , \text{Q.E.D.} \tag{13}$$

Let us now come back to the original task of evaluating the derivatives (9).

They follow with (10) to

$$\frac{\partial \ln pr(W)}{\partial \tilde{x}} = -B^* \tilde{N}^{-1} \quad B - B^* \tilde{N}^{-1} W = 0, \tag{14a}$$

$$\frac{\partial \ln pr(W)}{\partial \tilde{q}i} = \frac{1}{\tilde{q}i^2} \tilde{v}i^* \tilde{v}i - tr(Mi\tilde{N}^{-1}) = 0, \tag{14b}$$

$$\tilde{v}i = + \tilde{q}i \cdot Ai^* \tilde{N}^{-1} (B\tilde{x} + W),$$

$$\tilde{N} = (AQA^*) qi = \tilde{q}i, \quad Mi = Ai Ai^*.$$

The second term in equation (14b) resembles strongly the expectation of $\tilde{v}i^* \tilde{v}i$, which, when evaluated for the theoretical values x and q, is

$$E(\tilde{v}i^* \tilde{v}i/x, q) = qi^2 \, tr(MiN^{-1}).$$

Replacing in this expression qi by $\tilde{q}i$ and $E(\tilde{v}i^* \tilde{v}i/x, q)$ by the symbol $E_1(\tilde{v}i^* \tilde{v}i/x, q)$, it follows

$$tr(Mi\tilde{N}^{-1}) = \frac{1}{\tilde{q}i^2} \cdot E_1 (\tilde{v}i^* \tilde{v}i/x, q)$$

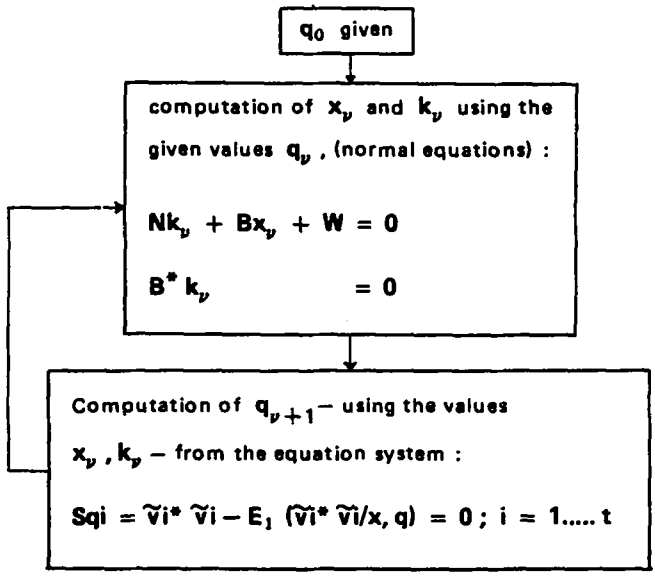
and equation (14b) becomes

$$\tilde{v}i^* \tilde{v}i - E_1 (\tilde{v}i^* \tilde{v}i/x, q) = 0 \tag{14c}$$

The system (14) is equivalent to the following larger system of equations :

$$\begin{array}{|l|l|}
\hline
\tilde{N}.k + B.\tilde{x} + W = 0 & \\
B^*.k = 0 & \\
\hline
Sqi = \tilde{v}i^* \tilde{v}i - E_1 (\tilde{v}i^* \tilde{v}i/x, q) = 0 & i = 1 \ldots\ldots t \\
\hline
\end{array} \tag{15}$$

which contains an additional unknown vector $\mathbf{k}$. The latter system can be readily transformed into (14) by merely eliminating the vector $\mathbf{k}$. It resembles strongly the normal equations (6) of the method of least squares ; the present system is, however, non–linear and contains also the unknown $\tilde{\mathbf{q}}$. A non–linear equation system can in most cases only be solved iteratively. Starting from suitable approximate values $\mathbf{x}_0$, $\mathbf{q}_0$ a sequence of better approximations $\mathbf{x}_\nu$, $\mathbf{q}_\nu$, ($\nu = 1, 2, 3 \ldots$) is computed, which finally converges to $\tilde{\mathbf{x}}$, $\tilde{\mathbf{q}}$. In this paper the sequence $\mathbf{x}_\nu$, $\mathbf{q}_\nu$, will be computed by the ***Block Gauss–Seidel Method*** for non–linear equation systems (see e.g. Varga, 1962 ; Martin and Tee, 1962), which possesses the following algorithm :



The algorithm (16) may be interpreted as a repeated least squares adjustment, followed by the improvement of the variances $\mathbf{q}$. Details about the computation of $\mathbf{q}_{\nu+1}$ are given in appendix A. The iterative computation is stopped as soon as two successive approximations $\mathbf{q}_\nu$ and $\mathbf{q}_{\nu+1}$ no longer differ significantly. The approximations computed last serve as final result.

The stochastical properties of the deduced estimators are derived within the theory of maximum likelihood (see e.g. Wilks 1962) and are studied in detail by Kubik (1967_2). There, it is shown, that the estimators $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{x}}$ form minimum variance, unbiased estimators assuming the theoretical value of $\mathbf{q}$ is known. The estimators $\tilde{\mathbf{q}}$ will converge "in probability" to the theoretical values $\mathbf{q}$ for infinite sample size and in this limiting case they will have higher accuracy than any other estimator and will also be normally distributed. The estimators $\tilde{\mathbf{q}}$ will be unbiased if the mathematical model (1) does not contain any unknown $\mathbf{x}$, i.e. $\mathbf{E}(\tilde{\mathbf{q}}) = \mathbf{q}$.

For standard problems II and IV, however, $\tilde{q}$ will be biased, the bias being

$$b(\tilde{q}i) = E(\tilde{q}i/q) - qi = \sum_{j=1}^{t} V(i,j) \cdot tr(MjN^{-1} B (B^* N^{-1} B)^{-1} B^* N^{-1}) \quad (17)$$

where $V(i,j)$ are the elements of the matrix $V = I^{-1}$ with I a matrix with the elements $I(i,j) = \frac{1}{2} tr(N^{-1} MiN^{-1} Mj)$. The bias is caused by unknowns x and q being simultaneously estimated and depends mainly on the degrees of freedom (redundancy) f in the adjustment. The larger f is, the smaller the bias will be. The variance matrix of the estimators $\tilde{q}$ may always be approximated by $V(\tilde{q}) = V$, (Wilks 1962, p. 380).

4.2. For some applications it may be advisable to find unbiased estimators for q also for standard problems II and IV. For this purpose more constraints have to be added to the optimization problem (8), in order to ensure that the estimators will be unbiased ; this leads then to the following problem

$$\ln L(w = W ; \hat{q}, \hat{x}) \rightarrow max$$

$$E(\hat{q} / q, x) = q$$

$$E(\hat{x} / q, x) = x \quad (18)$$

In the previous method of estimation (8), the bias of $\tilde{q}$ was caused by the simultaneous estimation of x and q ; suppose we now estimate x independently by

$$\hat{x} = -(B^* N^{-1} B)^{-1} B^* N^{-1} W, \quad (19)$$

using an approximation q_0 to q. We may then ask which estimator of q maximizes the likelihood for all samples giving the ascertained value of x, namely $\hat{x}$. The variations of L from sample to sample are now considered in a certain subpopulation in which x has a fixed value. The problem is then solved as follows : the likelihood function is put into the following form

$$\ln L = \ln pr(W) = \ln pr(\hat{x}) + \ln pr(\hat{v}), \quad (20)$$

where $pr(\hat{x})$ represents the distribution function of the least squares estimator $\hat{x}$, (19)

$$pr(\hat{x}) = \frac{1}{(2\pi)^{m/2}} \frac{1}{|(B^* N^{-1} B)^{-1}|^{\frac{1}{2}}} exp -\frac{1}{2} (\hat{x} - x)^* B^* N^{-1} B (\hat{x} - x) \quad (20a)$$

$\mathbf{pr}\,(\hat{\mathbf{v}})$ represents the distribution function of the least squares estimator $\hat{\mathbf{v}} = +\mathbf{q}.\mathbf{A}^*.\mathbf{N}^{-1}\,(\mathbf{B}\hat{\mathbf{x}} + \mathbf{W})$,

$$\mathbf{pr}\,(\hat{\mathbf{v}}) = \frac{1}{(2\pi)^{\frac{k-m}{2}}} \frac{|(\mathbf{B}\mathbf{N}^{-1}\mathbf{B})^{-1}|^{\frac{1}{2}}}{|\mathbf{N}|^{\frac{1}{2}}} \exp -\frac{1}{2}\mathbf{W}^*(\mathbf{N}^{-1} - \mathbf{N}^{-1}\mathbf{B}(\mathbf{B}^*\mathbf{N}^{-1}\mathbf{B})^{-1}$$

$$\mathbf{B}^*\mathbf{N}^{-1})\,\mathbf{W}. \qquad (20b)$$

This relation (19) may be readily verified, using the generalized inverse for the singular variance matrix $\mathbf{Qvv}$, (Bjierhammer 1958). If we maximize the likelihood in this form, for simultaneous variation of $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$, we arrive back at (9) and (15), as of course we must. But if $\hat{\mathbf{x}}$ has a fixed value, the logarithmic likelihood is then proportional to the second factor in (19), viz.

$$\mathbf{lnL}\ \alpha\ \frac{1}{2}\ln\left|\,(\mathbf{B}^*\mathbf{N}^{-1}\mathbf{B})^{-1}\,\right| - \frac{1}{2}\ln|\mathbf{N}| - \frac{1}{2}\,\mathbf{W}^*(\mathbf{N}^{-1} - \mathbf{N}^{-1}\mathbf{B}(\mathbf{B}^*\mathbf{N}^{-1}\mathbf{B})^{-1}$$

$$\mathbf{B}^*\mathbf{N}^{-1})\,\mathbf{W}$$

This factor is maximized by the solution $\hat{\mathbf{q}}$ of the following equation system :

$$\mathbf{Sqi} = \hat{\mathbf{v}}\mathbf{i}^*\,\hat{\mathbf{v}}\mathbf{i} - \mathbf{E}_2\,(\hat{\mathbf{v}}\mathbf{i}^*\,\hat{\mathbf{v}}\mathbf{i}\,/\,\hat{\mathbf{x}}, \mathbf{q}) = \mathbf{0}, \qquad (21)$$

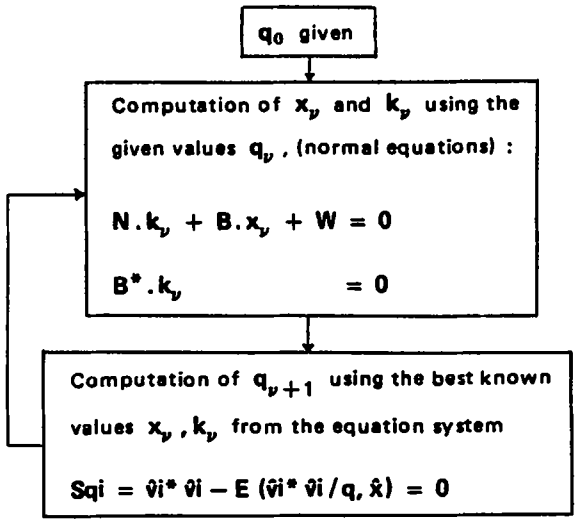
with

$$\hat{\mathbf{v}}\mathbf{i} = +\,\hat{\mathbf{q}}\mathbf{i}.\mathbf{A}\mathbf{i}^*.\hat{\mathbf{N}}^{-1}\,(\mathbf{W} + \mathbf{B}\hat{\mathbf{x}}\,,\ \hat{\mathbf{N}} = (\mathbf{N})_{\mathbf{q}=\hat{\mathbf{q}}}\,,$$

$$\mathbf{E}_2\,(\hat{\mathbf{v}}\mathbf{i}^*\,\hat{\mathbf{v}}\mathbf{i}\,/\,\hat{\mathbf{x}}, \hat{\mathbf{q}}) = \{\mathbf{E}\,(\hat{\mathbf{v}}\mathbf{i}^*\,\hat{\mathbf{v}}\mathbf{i}\,/\,\hat{\mathbf{x}}, \mathbf{q})\}_{\mathbf{q}=\hat{\mathbf{q}}} = \{\mathbf{tr}\,(\mathbf{Mi}\mathbf{N}^{-1}) - \mathbf{tr}\,(\mathbf{Mi}\mathbf{N}^{-1}\,\mathbf{B}\,(\mathbf{B}^*\mathbf{N}^{-1}\mathbf{B})^{-1}$$

$$\mathbf{B}^*\mathbf{N}^{-1})\}_{\mathbf{q}=\hat{\mathbf{q}}}$$

This $\hat{\mathbf{q}}$ is – as may be verified – an unbiased estimator for $\mathbf{q}$, if theoretical values of $\mathbf{q}$ were used for computing $\hat{\mathbf{x}}$. Since this was not the case, the computation of $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$ has to be repeated with the new approximations substituted for $\mathbf{q}$ Consequently the computation is iterative, and a sequence of values $\mathbf{x}_\nu$ and $\mathbf{q}_\nu$ ($\nu = 1, 2, 3$ .....) is computed which finally will converge towards the desired estimations $\hat{\mathbf{x}}$ and $\hat{\mathbf{q}}$. The resulting computational algorithm is very similar to (16), namely



The present algorithm differs from (6) only in the computation of q_ν which is now more elaborate. Details on the computation of q_ν are given in appendix **B**. The iterative computations (22) are stopped again as soon as two successive approximations q_ν and $q_{\nu+1}$ no longer differ significantly.

The estimators $\hat{q}$ and $\hat{x}$, which result from this computation, are unbiased and, for sufficiently large sample size, normally distributed. In the limiting case (sample size tending towards infinity), the estimator $\hat{q}$ will have higher accuracy than any other estimator. For standard problem I the estimators $\tilde{q}$ and $\hat{q}$ are identical as may be verified by neglecting in the algorithms (16) and (22) all terms containing the matrix **B**. The variance matrix of $\hat{q}$ may be approximated by the expression $V(\hat{q}) \doteq V$ with **V** defined in (17).

4.3. In both cases considered so far the extended least squares method consists of repeatedly carrying out the usual method of least squares, followed by improving the values **qi**. The algorithms, which were derived for standard problem IV, may be readily specialized for standard problem I (by neglecting all terms containing **B**), and for standard problem II (by setting **Ai** equal to a unit matrix $E_{(i)}$). This is partly elaborated in appendices **A** and **B**. For all algorithms published in this paper computer programmes are available in ALGOL 60 language.

This method may still be extended to adjustment problems, in which **Q** possesses a more general structure than in figure (2) and where, for example, some of the non–diagonal elements of **Q** are unknown. The formulas for this case were given by Kubik (1967_2) ; here, however, the number of unknown coefficients should always be limited in order to obtain reasonably accurate estimators. A large sample for **w** is necessary in order to obtain significant estimators of the co–variances. The extension of the presented method to the matrix **Q**, (2), where the

unit matrices $\mathbf{E}_{(i)}$ are replaced by diagonal matrices with known coefficients, may be readily derived from what has been already discussed and will not be treated here.

It is possible that one or more of the estimated variances $\hat{\mathbf{q}}$ become smaller than zero. This may be the case for a small sample size or for models which do not properly describe the actual situation. In these cases, the inequality constraints $\mathbf{qi} \geq \mathbf{0}$, (23) have to be added to the optimization problems (8) and (18) in order to obtain results which do not contradict the theory of probability. This necessitates a small modification of the iterative algorithms (16) and (22). The necessary conditions of a local optimum under the inequality constraints (23) have been given by Kuhn and Tucker (1950). For the present problem they may be written as follows :

$$\mathbf{qi} \geq \mathbf{0}$$

$$\frac{\partial \ln L}{\partial \mathbf{qi}} \leq \mathbf{0}$$

$$\mathbf{qi} \cdot \frac{\partial \ln L}{\partial \mathbf{qi}} = \mathbf{0} .$$

If now after one iteration step a value $\mathbf{qi}$ smaller than zero is found, it is made equal to, and kept at, zero until $\frac{\partial \ln L}{\partial \mathbf{qi}}$ becomes $> \mathbf{0}$; then $\mathbf{qi}$ may vary again. This ensures that a solution of $\mathbf{q}$ will be found which is within the feasible region $\mathbf{qi} \geq \mathbf{0}$.

5. — A numerical example :

The theory presented will now be illustrated by a numerical example. A geodetic net, in which both directions and distances have been measured, has to be adjusted. The measured quantities and the configuration of the points in the net are shown in figure 24. The standard deviations for distance measurements and direction measurements are assumed to be $\sigma_s = \mathbf{1.5\ cm}$ and $\sigma_r = \mathbf{5\ dmg}$ respectively. Due to refraction influences, however, the relationship between these two quantities is not accurately known and this relationship still has to be improved during the adjustment.

Algorithm (22) of the extended least squares method, standard problem II, is applied to yield unbiased estimators for both the point coordinates and the variances σ_s^2 and σ_r^2. The results of the computations are summarized in table 25 and 26. Table 25 gives a list of the successive approximations to σ_s^2 prop. σ and σ_r^2 prop. σ. They are listed together with the sums of squares **v1v1, v2v2** and

the estimated variances factors $\hat{\sigma}_0^2 = \frac{\mathbf{v}^* \mathbf{Q}^{-1} \hat{\mathbf{v}}}{f}$. In table 26 the estimated point coordinates obtained from the first and last iterative steps of the computation and their estimated variance matrices are compared. The first iteration step corresponds to the usual least squares method. It is seen clearly that the use of the extended least squares method yields more reliable results than the common least squares method. This improvement will be even more noticeable in those cases where the weights are less accurately known.

6. – Final remarks

The use of this method interferes with the usual statistical tests which may be done after the least squares adjustment for detecting gross errors. It is therefore advisable to perform these tests after the first or second stage of the iterative computation of the extended method. In this method the decision whether a particular measurement is affected by a gross error or not is more critical ; because, if a doubtful measurement is accepted, then a larger variance will be estimated for it or for the corresponding group of measurements. This problem of the interaction of statistical tests with this extended estimation method still needs a more detailed study.

There are a few practical restrictions to this method. Only a limited number of parameters can be estimated from one sample of **w**. The more free parameters which are estimated from a given sample, then the larger their variance and their correlation will be. Consequently only a small number of parameters **qi** in the **Q** matrix should be regarded as unknown, in order to yield small variances for their estimators.

So the theory presented will be especially valuable in those cases where little is known about the weights in an adjustment problem and where not too many – say 2 to 4 – parameters of the **Q** matrix are regarded as unknown.

Appendix A :

Numerical solution of the equation system (6b). Since the equation system (6b) for $\mathbf{q}_{\nu+1}$ is still non–linear, it is linearized by a Taylor expansion before solving for $\mathbf{q}_{\nu+1}$ (which corresponds to the execution of one step of Newton's iterative method). The linearized equation system for the approximation $\mathbf{q}_{\nu+1}$ is written as

$$\boxed{\mathbf{H}\,(\mathbf{q}_{\nu+1} - \mathbf{q}_{\nu}) = \mathbf{Sq}} \qquad \text{(A1)}$$

with $\mathbf{Sq}^* = [\,\mathbf{Sq1}, \mathbf{Sq2}, \ldots . \mathbf{Sqt}\,]$, evaluated for the value $\tilde{\mathbf{q}} = \mathbf{q}_{\nu}$,

$\mathbf{H}$ a matrix with the elements $\mathbf{H(i,\,j)} = \mathbf{tr}\,(\mathbf{N^{-1}MiN^{-1}Mj}) - \mathbf{2D^*N^{-1}MiN^{-1}MjN^{-1}D}$

$$\mathbf{D} = \mathbf{W} + \mathbf{B}.\tilde{\mathbf{x}} = \mathbf{W} - \mathbf{B}\,(\mathbf{B^*N^{-1}\,B})^{-1}\,\mathbf{B^*N^{-1}W}$$

To simplify the computation of $\mathbf{q}_{\nu+1}$, the coefficient matrix $\mathbf{H}$ may be replaced by its expectation $\mathbf{E\,(H/x)} = \bar{\mathbf{I}}$. This expectation is computed using the following theorem :

Theorem :

The expectation of an operator $\mathbf{D^*.T.D}$, where $\mathbf{D}$ is a random vector and $\mathbf{T}$ is a linear operator, is equal to

$$\mathbf{E\,(D^*\,T\,D)} = \mathbf{tr\,(E(D^*\,D)\,.\,T)}\,.$$

Proof : The operator $\mathbf{D^*\,T\,D}$ is written in index notation as

$$\sum_{\alpha,\,\beta\,=\,1}^{n} \mathbf{D}_{\alpha}\,\mathbf{T}_{\beta}^{\alpha}\,\mathbf{D}^{\beta}$$

with $\mathbf{D}_{\alpha}$, $\mathbf{T}_{\beta}^{\alpha}$, $\mathbf{D}^{\beta}$ denoting the individual elements in $\mathbf{D^*}$, $\mathbf{T}$ and $\mathbf{D}$. Its expectation may be written

$$\mathbf{E}\,\Big(\sum_{\alpha,\,\beta=1}^{n} \mathbf{D}_{\alpha}\,\mathbf{T}_{\beta}^{\alpha}\,\mathbf{D}^{\beta}\Big) = \mathbf{E}\,\Big(\sum_{\alpha,\beta=1}^{n} \mathbf{D}_{\alpha}\,\mathbf{D}^{\beta}\,\mathbf{T}_{\beta}^{\alpha}\Big)\,,$$

and since this is an algebraic sum, the symbols for sum and expectation may be interchanged :

$$\sum_{\alpha,\beta=1}^{n} E(D_\alpha D^\beta) T_\beta^\alpha$$

This expression is identical to $\mathrm{tr}(E(D^* D) T)$, Q.E.D.

From this it follows, that the matrix $E(H)$ is identical to a matrix $\bar{I}$ with the elements $\bar{I}(i,j) = -\mathrm{tr}(N^{-1} M i N^{-1} M j)$. The simplified equation system for $q_{\nu+1}$ becomes :

$$\boxed{\bar{I}(q_{\nu+1} - q_\nu) = Sq.} \qquad (A2)$$

This method was proposed by Rao (1952). However, it can only be recommended for relatively precise approximate values of q_0, since the convergence behaviour of the computation using (A2) is inferior to using the method of Newton (A1). When using imprecise values q_0, this method will often diverge. A very simple form of the equation system (6b) results in the case of standard problem II. The elements $qi_{\nu+1}$ of the vector $q_{\nu+1}$ can there be computed one by one from the equations

$$qi_{\nu+1} = \left\{ \frac{\tilde{v}i^* \tilde{v}i}{f} \right\}_{q_\nu} ,$$

with f the degree of freedom in the adjustment problem.

Appendix B :

Numerical solution of the equation system (22b). The solution of (22b) for $q_{\nu+1}$ by linearising (22b) cannot be recommended since the expressions involved are very complex. The writer is experimenting at present with other methods of solution (Powells conjugate directions method, see Powel 1964). Only in standard problem II can $q_{\nu+1}$ be computed with reasonable effort, by the algorithm

$$\boxed{\begin{aligned} H.dp &= Sq \\ qi_{\nu+1} &= \frac{qi_\nu}{1 + dp(i).qi_\nu} \end{aligned}} \qquad \begin{matrix} (B1) \\ i = 1 \ldots t \end{matrix}$$

with dp a vector with the elements $dp(i)$

$Sq^* = [\, Sqi, Sq2, \ldots . Sqt \,]$, evaluated for the value $\hat{q} = q_\nu$.

$H = \dfrac{\partial Sq}{\partial (1/qi)}$, a matrix with elements :

$$H(i,j) = -2\,\hat{v}i^*\, Bi\, N^{-1}\, Bj^*\, \hat{v}j + m.qi^2 \,.\, \delta ij - tr\,(Bi^*\, BiN^{-1}\, Bj^*\, BjN^{-1}) ;$$

δij ... Kronecker symbol ;

m .... number of unknowns ;

Bi .... submatrices of B (standard problem II), partitioned according to vi.

$$N = B^*\, Q^{-1}\, B.$$

This algorithm is simplified when H is replaced by $E(H)$, namely

$$\boxed{\begin{aligned} &E(H).\,dp = Sq \\ &qi_{\nu+1} = \frac{qi_\nu}{1 + dp\,(i).\,qi_\nu} \end{aligned}} \qquad \text{(B2)}$$

with $E(H)$ a matrix with the elements :

$$E(i,j) = tr\,(Bi^*\, BiN^{-1}\, Bj^*\, BjN^{-1}) - \delta ij.qi.tr.(Bi^*\, BiN^{-1}) + \delta ij\,.\,n_i\,.\,qi^2 ;$$

n_i number of rows in Bi.

Again this method can only be recommended for fairly precise approximate values q_0 .

REFERENCES

F.R. HELMERT (1924) : Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate, 1924, p. 362.

A. BJERHAMMER (1958) : A generalized matrix algebra. Göteberg 1958.

K. KUBIK (1967_1) : Schätzung der Gewichte der Fehlergleichungen beim Ausgleichsproblem nach vermittelnden Beobachtungen, Zeitschrift für Vermessungswesen, Vol, 92 (1967) 5, p. 173.

K. KUBIK (1967_2) : Schätzfunktionen für Varianzen, Kovarianzen und für andere Parameter in Ausgleichaufgaben. Doctoral thesis ; Technical University of Vienna, 1967.

H.W. KUHN and A.W. TUCKER (1950) : Nonlinear Programming. Proc. Second Berkeley Symposium Univ. of Calif. Press 1950 p. 481–492.

D.W. MARTIN and G.J. TEE (1962) : Iterative Methods for Linear Equations with Symmetric Positive Definite Matrix. The Computer Journal, Vol. 4 (1961/62), p. 242.

R. RAO (1952) : Advanced statistical methods in biometric research. John Wiley, 1952.

J.M. TIENSTRA (1956) : Theory of adjustment of normally distributed observations. Amsterdam 1956.

R.S. VARGA (1962) : Matrix Iterative Analysis. Prentice Hall, London 1962.

S.S. WILKS : Mathematical Statistics. John Wiley, 1962.

Table 25

Results of the iterative computation

of the extended method of least squares

Initial values	:	$q1_0 = 1$	or	$\sigma_s = 1.5$ cm
		$q2_0 = 0.64$		$\sigma_r = 5$ dmg
First iteration	:	vv1 = 4.7008		
		vv2 = 5.3742		
		q1 = 0.9657		
		q2 = 0.4758		$\hat{\sigma}_0 = 0.83$
Second iteration	:	vv1 = 5.5231		
		vv2 = 4.9206		
		q1: = 1.2170		
		q2: = 0.4000		$\hat{\sigma}_0 = 0.95$
Third iteration	:	vv1 = 7.1560		
		vv2 = 4.2610		
		q1 = 1.1676		
		q2 = 0.3924		$\sigma_0 = 1.00$
Fourth iteration	:	vv1 = 7.0622		
		vv2 = 4.2922		
		q1 = 1.1634		
		q2 = 0.3927		$\sigma_0 = 1.00$
Fifth iteration	:	vv1 = 7.0448		
		vv2 = 4.2981		
		q1 = 1.1634		
		q2 = 0.3927		$\hat{\sigma}_0 = 1.00$
Final values	:	q1 = 1.16	or	$\sigma_s = 1.6$ cm
		q2 = 0.39		$\sigma_r = 4$ dmg

Variance matrix of q1 , q2

$$\begin{pmatrix} 0.372 & -0.004 \\ -0.004 & 0.013 \end{pmatrix}$$

Table 26A

Results after first iteration step of extended method

Estimated Coordinates of points

point 1		point 3		point 5		point 7		point 9		point 11	
Y	X	Y	X	Y	X	Y	X	Y	X	Y	X
1000.000	0	246.337	457.407	566.164	1000.775	1000.000	2065.184	1804.130	1168.397	1525.839	686.979

Variance Matrix of Point Coordinates (elements in cm^2)

	point 1		point 3		point 5		point 7		point 9		point 11	
	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X
	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	0	0										
3	0	0	0.27									
4	0	0	0.08	0.40								
5	0	0	0.07	0.09	0.28							
6	0	0	0.00	0.22	0.06	0.21						
7	0	0	0	0	0	0	0					
8	0	0	0	0	0	0	0	0				
9	0	0	0.00	0.00	0.08	0.03	0	0	0.41			
10	0	0	−0.06	0.00	−0.02	0.04	0	0	0.08	0.32		
11	0	0	0.02	0.03	0.17	0.06	0	0	0.17	0.07	0.44	
12	0	0	−0.04	0.01	0.00	0.04	0	0	0.00	0.19	0.02	0.18

Table 26B

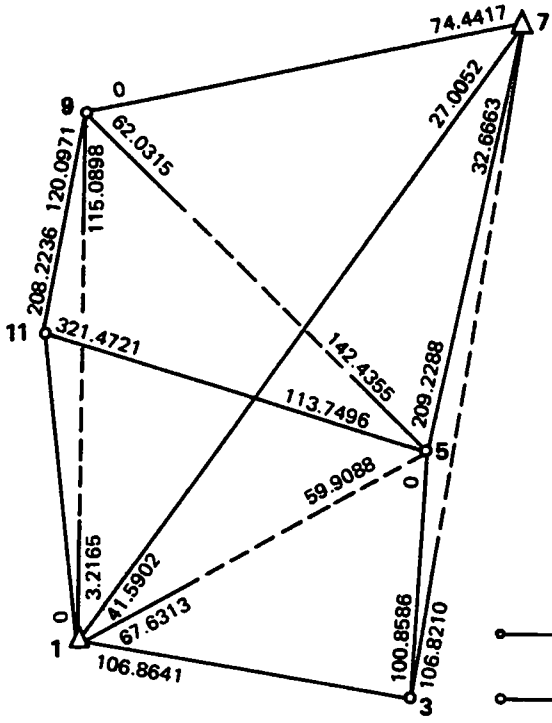
Results after last iteration step of extended method

Estimated Coordinates of Points

point 1		point 3		point 5		point 7		point 9		point 11	
Y	X	Y	X	Y	X	Y	X	Y	X	Y	X
+1000.000	0	246.335	457.408	566.163	1000.778	1000.000	2065.184	1804.130	1168.400	1525.841	686.983

Variance Matrix of Point Coordinates (elements in cm^2)

	point 1		point 3		point 5		point 7		point 9		point 11	
	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X
	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	0	0										
3	0	0	0.22									
4	0	0	0.07	0.34								
5	0	0	0.05	0.09	0.25							
6	0	0	0.00	0.19	0.06	0.18						
7	0	0	0	0	0	0	0					
8	0	0	0	0	0	0	0	0				
9	0	0	0.00	0.00	0.08	0.03	0	0	0.34			
10	0	0	0.05	0.00	0.00	0.03	0	0	0.08	0.27		
11	0	0	0.01	0.04	0.18	0.07	0	0	0.14	0.06	0.42	
12	0	0	0.03	0.01	0.00	0.03	0	0	0.01	0.16	0.02	0.15



Coordinates :

Point	Y	X
1	1000.000	0.000
7	1000.000	2065.184

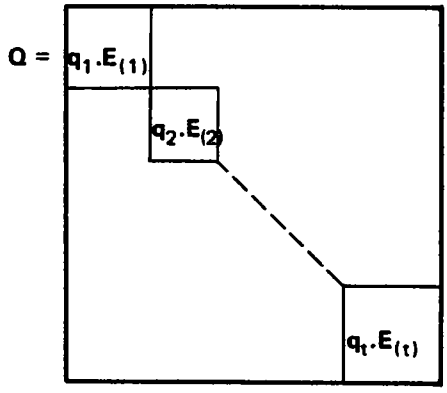
Measured Distances :

from—to	Distance (m)
1,3	881.596
1,11	865.110
3,5	630.493
5,7	1149,437
5,11	1009.659
7,9	1204.527
9,11	556,057

directions measured

both direction and distance measured

Fig. 24 : Configuration of the geodetic network



q i scalars, i = 1 . . . . . t

$E_{(i)}$ unit matrices of various dimensions

Fig. 2 : Type of matrix Q