AN EFFICIENT ALGORITHM FOR THE COMPUTATION OF THE INVERSE IN BUNDLE ADJUSTMENT SYSTEMS

Rune Larsson
Department of Photogrammetry
Royal Institute of Technology
Sweden
Commission III

ABSTRACT

A brief review of the Gentri system for bundle adjustment is given. The point by point processing utilised in Gentri lends itself to a straightforward computation of the band part of the inverse of the normal equations. The algorithm and data structures for that computation is presented. For each point the following is computed in the last iteration:

- o final ground point co-ordinates
- o co-ordinate error estimates
- o image co-ordinate residuals
- o residual error estimates (redundancy numbers).

All necessary data is then available for blunder detection, determination of reliability and estimation of variance components (a posteriori variance estimation).

SUMMARY OF THE BUNDLE ADJUSTMENT

The bundle adjustment system Gentri is designed with the aim to incorporate geodetic and other observations along with the photo observations and adjust them simultaneously (Larsson 1982a, 1982b, 1983).

In order to process data efficiently, a suitable data structure was developed. The necessary data is divided into three groups, namely photo data, observation data and point data.

The observation is always connected to one photo and one point, and serves as the primary search unit. From each observation, all connected data can be accessed. This structure could with success also be applied to other types of observations.

The choice was also made to process all observations to one point at the same time. The idea was fruitful because a point observed geodetically could easily be discovered and be processed accordingly.

This approach made it possible to access both the observations and the points in sequence, whereas the photos were accessed at random. The number of photo accesses were minimised with a suitable preparatory sorting of the input data, together with a dynamic allocation facility for photo data.

Other advantages were the immediate reduction of point unknowns and that the submatrices for the reduction of each point could be saved on an intermediate sequential file for later use.

DERIVATION OF THE SOLUTION

The basic equation for bundle adjustment is based on collinearity between the point on the ground, the projection centre and the point on the photo. If this nonlinear equation is

$$F(l, x) = 0 \qquad \ldots (1)$$

then an approximate solution $\tilde{x}$ is computed. At $\tilde{x}$, the function is linearized

$$\frac{\delta F}{\delta l} v + \frac{\delta F}{\delta x} dx = F(l, \tilde{x}) \qquad \ldots (2)$$

Now, with x a vector of unknowns and F a vector of functions, (2) is a linear system of equations which can be solved for dx. A better solution to (1) is then

$$x = \tilde{x} + dx \qquad \ldots (3)$$

The equations (1) can be detailed as follows:

$$\begin{aligned} F_p(l_p, x_p) &= 0 \\ F_i(l_i, x_p, x_i, x_a) &= 0 \\ F_a(l_a, x_a) &= 0 \end{aligned} \qquad \ldots (4)$$

where

F_p and l_p are the functions and the ficticious observations for the control points,

F_i and l_i are the collinearity functions and photo co-ordinate observations,

F_a and l_a are ficticious or actual observations for the additional parameters.

x_p are the unknown point co-ordinates,

x_i are the unknown photo orientation parameters and

x_a are the unknown additional parameters.

The corresponding linearized system of equations are

$$v + A_{dx} = w$$

or

$$\begin{pmatrix} v_p \\ v_i \\ v_a \end{pmatrix} + \begin{pmatrix} A_{pp} & 0 & 0 \\ A_{ip} & A_{ii} & +A_{ia} \\ 0 & 0 & +A_{aa} \end{pmatrix} \begin{pmatrix} dx_p \\ dx_i \\ dx_a \end{pmatrix} = \begin{pmatrix} w_p \\ w_i \\ w_a \end{pmatrix} \qquad \ldots (5)$$

where

A.. are the matrices of partial derivatives of the functions with respect to the relevant unknowns,

v. are the vectors of residuals and

w. are the vectors of discrepancies.

From (5), a system of normal equations is computed:

$$\begin{pmatrix} N_{pp} & N_{pi} & N_{pa} \\ & N_{ii} & N_{ia} \\ \text{symm} & & N_{aa} \end{pmatrix} \begin{pmatrix} dx_p \\ dx_i \\ dx_a \end{pmatrix} = \begin{pmatrix} u_p \\ u_i \\ u_a \end{pmatrix} \quad \ldots (6)$$

where

dx_p, dx_i and dx_a are unknown corrections to point co-ordinates, image orientation parameters and additional parameters, respectively, and

$N_{pp} = A'_{pp} A_{pp} + A'_{ip} A_{ip}$

$N_{pi} = A'_{ip} A_{ii}$

$N_{pa} = A'_{ip} A_{ia}$

$N_{ii} = A'_{ii} A_{ii}$

$N_{ia} = A'_{ii} A_{ia}$

$N_{aa} = A'_{ia} A_{ia} + A'_{aa} A_{aa}$

$u_p = A'_{pp} w_p + A'_{ip} w_i$

$u_i = A'_{ii} w_i$

$u_a = A'_{ia} w_i + A'_{aa} w_a$

Above was mentioned that the unknowns for point co-ordinates, dx_p, were eliminated from the system. This leads to a reduced system of normals

$$\begin{pmatrix} \bar{N}_{ii} & \bar{N}_{ia} \\ \text{symm} & \bar{N}_{aa} \end{pmatrix} \begin{pmatrix} dx_i \\ dx_a \end{pmatrix} = \begin{pmatrix} \bar{u}_i \\ \bar{u}_a \end{pmatrix} \quad \ldots (7)$$

where

$\bar{N}_{ii} = N_{ii} - N'_{pi} N_{pp}^{-1} N_{pi}$,

$\bar{N}_{ia} = N_{ia} - N'_{pi} N_{pp}^{-1} N_{pa}$,

$\bar{N}_{aa} = N_{aa} - N'_{pa} N_{pp}^{-1} N_{pa}$,

$\bar{u}_i = u_i - N'_{pi} N_{pp}^{-1} u_p$ and

$\bar{u}_a = u_a - N'_{pa} N_{pp}^{-1} u_p$.

This reduction yields the wellknown banded bordered system of normal equations, $\bar{N}_{ii}$ is banded and $\bar{N}_{ia}$ is the border. Thes matrices are large and to be processed in the computer, they are divided into submatrices, each containing a number of normal equations. Each submatrix or block is stored in a record of a random access disk file. This situation is shown in Figure 1.

During the decomposition of the normal equations, each block will modify a number of equation coefficients. The coefficients altered by the decomposition of the upper block in Figure 1 is shown shaded.

The decomposition of $\bar{N}_{ii}$ is

$$\bar{T}_{ii}\, dx_i + \bar{T}'^{-1}_{ii} \bar{N}_{ia}\, dx_a = \bar{T}'^{-1}_{ii} \bar{u}_i \quad \ldots (8)$$

where

$\bar{T}'_{ii} \bar{T}_{ii} = \bar{N}_{ii}$, $\bar{T}_{ii}$ upper triangular

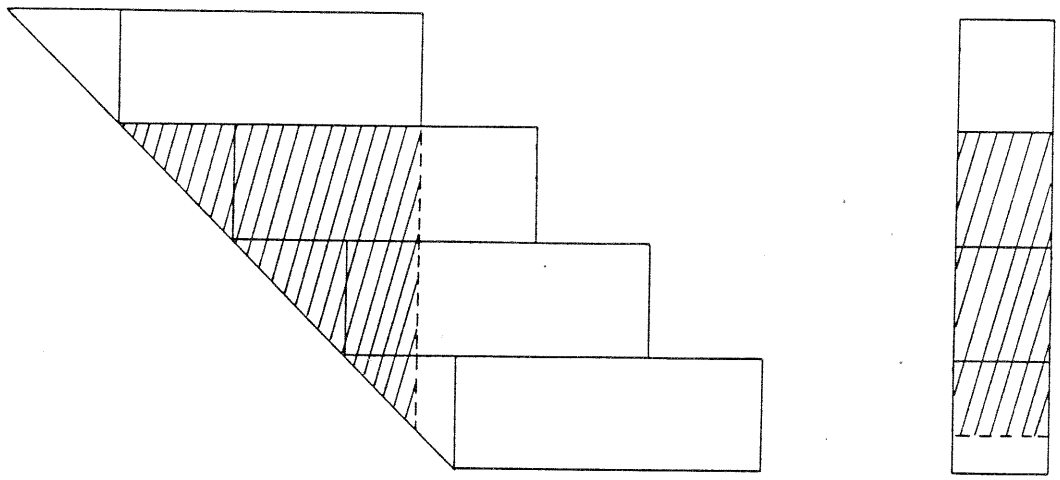

Figure 1 Blocking of band and border

The next step is then to compute

$$\left[\bar{N}_{aa} - (\bar{T}_{ii}'^{-1}\bar{N}_{ia})' (\bar{T}_{ii}'^{-1} \bar{N}_{ia})\right] dx_a = \bar{u}_a - (\bar{T}_{ii}'^{-1} \bar{N}_{ia}) (\bar{T}_{ii}'^{-1} \bar{u}_i)$$

or

$$\left[\bar{N}_{aa} - \bar{N}_{ia}' \bar{N}_{ii}^{-1} \bar{N}_{ia}\right] dx_a = \bar{u}_a - \bar{N}_{ia}' \bar{N}_{ii}^{-1} \bar{u}_i \qquad \ldots (9)$$

which is the same. The small remaining system of normal equations is solved conventionally to yield the corrections to the added parameters, dx_a.

ESTIMATES OF STANDARD ERRORS

When computing the inverse of a nonlinear system of equations such as this, it is essential to iterate the solution until the final corrections become zero, otherwise the derived quantities will be biased
Let

$$v + Ax = w \qquad \ldots (10)$$

be the linearized total set of error equations, where

v is the vector of residuals,

A the error equation matrix,

x the last correction (dx_p, dx_i, dx_a) to the solution and

w is the vector of discrepancies.

Then

$$x = (A' P A)^{-1} A' P w \qquad \ldots (11)$$

and

$$v = (I - A(A' P A)^{-1} A'P)w$$

$$= (I - A_{op})w \qquad \ldots (12)$$

and the reference variance is

$$\sigma_o^2 = \frac{v' P v}{r} \quad ; r = n - u \qquad \ldots (13)$$

where

n is the number of linearly independent observation,

u is the number of unknowns and

r is the redundancy.

The standard error of the unknowns are

$$\hat{s}_{xi}^2 = \hat{\sigma}_o^2 Q_{xx}|_{ii} \qquad \ldots (14)$$

where

$$Q_{xx} = (A' P A)^{-1} \qquad \ldots (15)$$

and the corresponding estimates for the residuals are

$$\hat{s}_{vj} = \hat{\sigma}_o^2 Q_{vv}|_{jj} \qquad \ldots (16)$$

where

$$Q_{vv} = P^{-1} - A(A' P A)^{-1} A' = P^{-1} - A Q_{xx} A' \qquad \ldots (17)$$

Let us study (17) to see how it is computed. The weight matrix P is a diagonal matrix. From the inverse of P a product of three matrices is subtracted. It is the full inverse of the normal equations, Q_{xx}, pre- and postmultiplied by the error equations A suitably transposed (c.f. (5)). Let us now study the calculation of Q_{xx}.

INVERSION OF THE SYSTEM OF NORMAL EQUATIONS

Initially some rules must be shown. For an arbitrary partitioning of the unknowns in two groups, the following is valid for the inverse of the normal equations

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{12}' & B_{22} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \qquad \ldots (18)$$

where

A.. are parts of the normal equations,

B.. are parts of the inverse and

I is a unit matrix of suitable order.

Two of the four possible relations are used

$$A_{11} B_{12} + A_{12} B_{22} = 0$$

$$A_{12}' B_{12} + A_{22} B_{22} = I$$

From the first

$$B_{12} = - A_{11}^{-1} A_{12} B_{22}$$

which is substituted for in the second relation

$$B_{22} = (A_{22} - A_{12}' A_{11}^{-1} A_{12})^{-1} \qquad \ldots (19)$$

The matrix to be inverted can be compared to the expression for $\bar{N}_{ii}$ in (7) and to the reduced normal equations in (9). It turns out that the inverse of the reduced normal equations equals the corresponding part of the inverse

of the entire matrix i.e. submatrices of the inverse can be computed through inversion of the relevant reduced normal equations. This can be shown strictly by applying (19) recursively, first to the partitioning in (7) and then to the same for (9).

The first inverse to compute is thus

$$Q_{aa} = (\bar{N}_{aa} - \bar{N}_{ia}' \bar{N}_{ii}^{-1} \bar{N}_{ia})^{-1} \qquad \ldots (20)$$

which matrix is obtained during the process of stepwise reduction. Q_{aa} is used to compute standard errors for and correlations between the additional parameters according to (14).

Let us then look at a larger part of the normal equations and include the band and border part of the reduced system (7). The following will then hold for the inverse

$$\begin{pmatrix} \bar{N}_{ii} & \bar{N}_{ia} \\ \bar{N}_{ia}' & \bar{N}_{aa} \end{pmatrix} \begin{pmatrix} Q_{ii} & Q_{ia} \\ Q_{ia}' & Q_{aa} \end{pmatrix} = \begin{pmatrix} I & O \\ O & I \end{pmatrix} \qquad \ldots (21)$$

where

$Q..$ are submatrices of the inverse,

from which

$$Q_{ia} = - \bar{N}_{ii}^{-1} \bar{N}_{ia} Q_{aa}$$

$$\begin{aligned} Q_{ii} &= \bar{N}_{ii}^{-1} (I - \bar{N}_{ia} Q_{ia}') \qquad \ldots (22) \\ &= \bar{N}_{ii}^{-1} + \bar{N}_{ia} Q_{aa} \bar{N}_{ia}' \bar{N}_{ii}^{-1} \end{aligned}$$

The matrix Q_{aa} is already available. The heavy task is to invert $\bar{N}_{ii}$. It is the matrix corresponding to the photo unknowns and it is large and banded, which means that the inverse, Q_{ii}, is full.

The most interesting part of the inverse is however, Q_{pp}, the cofactor matrix for point co-ordinate unknowns. To compute Q_{pp}, the formula (22) is used with another partitioning. We have

$$\begin{pmatrix} Q_{pi} & Q_{pa} \end{pmatrix} = - N_{pp}^{-1} \begin{pmatrix} N_{pi} & N_{pa} \end{pmatrix} \begin{pmatrix} Q_{ii} & Q_{ia} \\ Q_{ia}' & Q_{aa} \end{pmatrix}$$

$$Q_{pp} = N_{pp}^{-1} (I - \begin{pmatrix} N_{pi} & N_{pa} \end{pmatrix} \begin{pmatrix} Q_{pi}' \\ Q_{pa}' \end{pmatrix})$$

where the matrices within brackets are obtained by joining the indicated matrices. The expression for Q_{pp} will be

$$Q_{pp} = N_{pp}^{-1} + N_{pp}^{-1} \begin{pmatrix} N_{pi} & N_{pa} \end{pmatrix} \begin{pmatrix} Q_{ii} & Q_{ia} \\ Q_{ia} & Q_{aa} \end{pmatrix} \begin{pmatrix} N_{pi}' \\ N_{pa}' \end{pmatrix} N_{pp}^{-1} \qquad \ldots (23)$$

The three matrices N_{pp}, N_{pi} and N_{pa} are produced when reducing the point co-ordinate unknowns from the adjustment. They were processed sequentially,

point by point, and were saved in a sequential file. The structure of N_{pi} and N_{pa} is important. They are sparse and the non-zero elements correspond only to those photos where the point is measured and to the additional parameters. Thus to compute Q_{pp} for a point only the parts of Q_{ii} that are within the border are needed, along with Q_{ia} and Q_{aa}, i.e. the band and border part of the inverse of the normal equations.

This is shown in Figure 2, where the relevant submatrices for one point is shown. When forming the matrix product of (23) only the shaded areas of the matrices are needed. Since the N_{pi}'s once was used to reduce N_{ii}, making it banded, only the parts of Q_{ii} which correspond to the band of N_{ii} are utilized.

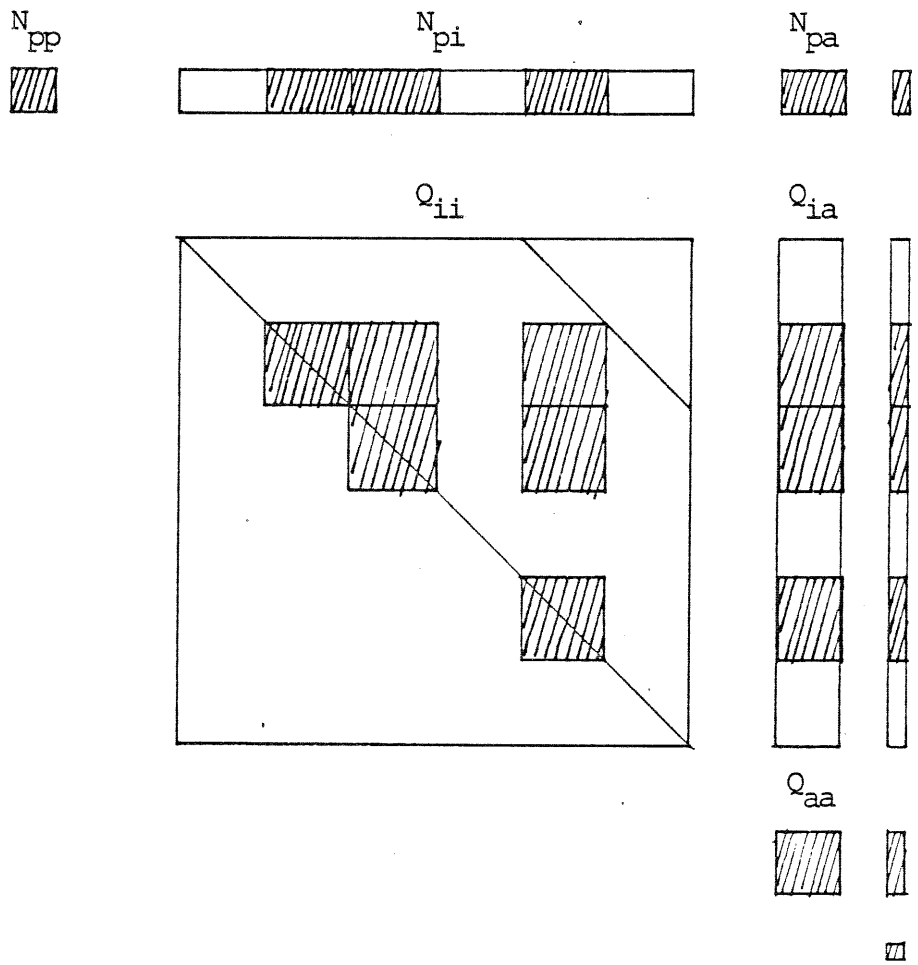

Figure 2 Computation of Q_{pp}

BAND AND BORDER OF THE INVERSE

The next step is to compute the inverses Q_{ii} and Q_{ia}. For a start the matrix Q_{aa} is already known, see formula (20). Then the following is valid

$$\begin{pmatrix} N_{nn} & N_{na} \\ N'_{na} & N_{aa} \end{pmatrix} \begin{pmatrix} Q_{nn} & Q_{na} \\ Q'_{na} & Q_{aa} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \qquad \ldots (24)$$

given that N_{nn} is the lowest partition of the band of Q_{ii} (corresponding

to photo n, say) and that all N-matrices are reduced with regard to all unknowns processed earlier (i.e. all point co-ordinate unknowns and the unknowns for photo 1, ..., n-1). This has been shown earlier and this is also the normal case when the forward decomposition of the normal equations has been performed. N_{na} is the corresponding lowest part of the band. Selecting two equations from (24) and rearranging yields

$$\begin{aligned} Q_{na} &= - N_{nn}^{-1} N_{na} Q_{aa} \\ Q_{nn} &= N_{nn}^{-1} (I - N_{na} Q_{na}') \end{aligned} \qquad \ldots (25)$$

In this step, only known matrices and inverses are utilized. The next step involves the augmenting of more submatrices to (24).

Let m = n - 1. Then

$$\begin{pmatrix} N_{mm} & N_{mn} & N_{ma} \\ N_{mn}' & N_{nn} & N_{na} \\ N_{ma}' & N_{na}' & N_{aa} \end{pmatrix} \begin{pmatrix} Q_{mm} & Q_{mn} & Q_{ma} \\ Q_{mn}' & Q_{nn} & Q_{na} \\ Q_{ma}' & Q_{na}' & Q_{aa} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \ldots (26)$$

The lower right hand parts of the inverse is already known. If we expand the matrix products from the first row of the normal equations in (26) with the three matrix columns of the inverse we get

$$\begin{aligned} Q_{ma} &= - N_{mm}^{-1} (N_{mn} Q_{na} + N_{ma} Q_{aa}) \\ Q_{mn} &= - N_{mm}^{-1} (N_{mn} Q_{nn} + N_{ma} Q_{na}') \\ Q_{mm} &= N_{mm}^{-1} (I - N_{mn} Q_{mn}' - N_{ma} Q_{ma}') \end{aligned} \qquad \ldots (27)$$

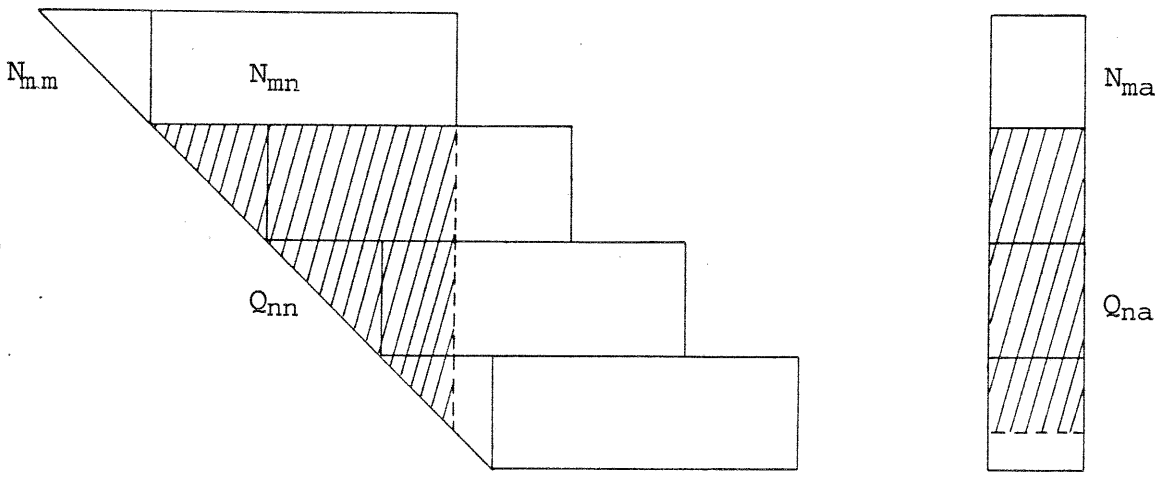

Figure 3 Computation of band and border of the inverse

This will bring us a small step up along the band. It is, however, possible to apply (27) recursively, which can be seen in Figure 3. There the horisontal partitioning of band and border show how the blocking of the matrices on a random access file is arranged. The upper block is to be inverted using (27). The shaded areas are parts of the inverse which are necessary for that computation (c.f. Wong 1973). In addition to those, also the matrix Q_{aa} is needed. By successively computing the inverse parts Q_{mm}, Q_{mn} and Q_{ma}, where

m, n are decremented, the parts of the inverse corresponding to the entire band and border are computed. The same data storage as for the normal equations are used, c.f. Figure 1.

In this process, three buffers of equation storage is needed, one for the part to be inverted (N_{mm}, N_{mn}, N_{ma}), one for the corresponding parts of the inverse and finally one for the different parts of (Q_{nn}, Q_{na}), which have to be accessed for the inversion.

Let us for one moment return to formula (23). There it was shown that the parts of the inverse that correspond to the point co-ordinate unknowns, Q_{pp}, can be computed with only partial knowledge of the inverse Q_{ii}. Now it has been shown that the necessary parts of Q_{ii} can be computed in an efficient way.

It deserves to be mentioned that this computation can be made within the same storage structure used for the solution of the normal equations with the exception that three buffers are needed instead of two.

COMPUTATION OF Q_{vv}

Let us recall the formula (17) and expand it slightly

$$Q_{vv} = P^{-1} - A\, Q_{xx}\, A^{t}$$

For a ficticious observation of a control point we have (c.f. (5))

$$Q_{vv}|_{p} = P_{p}^{-1} - A_{pp}\, Q_{pp}\, A_{pp}^{t}\,, \qquad \ldots (28)$$

for a ficticious observation of the additional parameters

$$Q_{vv}|_{a} = P_{a}^{-1} - A_{aa}\, Q_{aa}\, A_{aa}^{t} \qquad \ldots (29)$$

and finally, for the photo measurements of a point

$$Q_{vv}|_{i} = P_{i}^{-1} - \begin{pmatrix} A_{ip} & A_{ii} & A_{ia} \end{pmatrix} \begin{pmatrix} Q_{pp} & Q_{pi} & Q_{pa} \\ Q_{pi}^{t} & Q_{ii} & Q_{ia} \\ Q_{pa}^{t} & Q_{ia}^{t} & Q_{aa} \end{pmatrix} \begin{pmatrix} A_{ip}^{t} \\ A_{ii}^{t} \\ A_{ia}^{t} \end{pmatrix} \qquad \ldots (30)$$

The formula (30) is now of great interest. Since the processing of measurements is performed point by point, the necessary parts of (Q_{pp}, Q_{pi}, Q_{pa}) can be obtained easily. The computation of Q_{pp} is shown already in (23). The matrices Q_{pi} and Q_{pa} can be computed using

$$\begin{aligned} Q_{pa} &= -N_{pp}^{-1}(N_{pi}\, Q_{ia} + N_{pa}\, Q_{aa}) \\ Q_{pi} &= -N_{pp}^{-1}(N_{pi}\, Q_{ii} + N_{pa}\, Q_{ia}^{t}) \end{aligned} \qquad \ldots (31)$$

which can easily be derived using the same technique demonstrated earlier, e.g. in (26) and (27). It is important to note that the parts of Q_{ii} already computed are sufficient also for this computation.

POINT-BY-POINT PROCESSING

Finally, the inversion process using pointwise processing is summarized. The observations are all treated point by point, and the point co-ordinate unknowns are immediately eliminated from the normal equations. At the same time, matrices (N_{pp}, N_{pi}, N_{pa}) are stored in a temporary sequential file. The banded bordered normal equations are solved conventionally,

and the solution is iterated until convergency. In the final iteration, the band and border part of the inverse is computed simultaneously with the computation of the solution to the photo unknowns. The inversion is computed using (20), (25) and (27).

With these data, the standard errors for the solution of the additional parameters and photo orientation unknowns can be presented.

The subsequent processing is performed pointwise. All observations to one ground point are loaded, and the error equation coefficients (A_{ip}, A_{ii}, A_{ia}) are computed. From a temporary sequential file the matrices (N_{pp}, N_{pi}, N_{pa}) are read, and from these the solution of the point co-ordinate unknowns are obtained conventionally. The standard errors of the solution can now be computed using (23) and (14). For that computation, only parts of the band (Q_{ii}) and border (Q_{ia}) matrices need to be loaded. The matrix Q_{aa} finally, is permanently stored in primary storage.

With the error equation matrices, residuals for all observations to the current point are obtained.

By finally applying formulae (30) and (16), the corresponding part of Q_{vv} is computed, where the matrices Q_{pi} and Q_{pa} are computed using (31). Again, all necessary submatrices of the inverse is present, and subsequently the number of accesses to disk storage is minimized.

When all data has been presented for one point, the process continues with the next. All matrices belonging to the point (A_{ip}, A_{ii}, A_{ia}, A_{pp}, w_p, w_i, N_{pp}, N_{pi}, N_{pa}, Q_{pp}, Q_{pi}, Q_{pa}, Q_{vv}, P_i) can now be disposed of, since they are no longer necessary. That fact is essential for the efficiency of the program - the processing is adjusted to the data structure, the information is accessed easily and utilized fully when present.

REMARKS

The Gentri system for adjustment is designed to treat photogrammetric triangulation data as efficiently as possible. The sequence of points is ordered to fit as good as possible to the order of photos. This will minimize the number of accesses to photo data.

In most systems for aerial triangulation however, the program is often designed to process all measurements of one photo simultaneously. Nevertheless, these systems are at least during some phases of processing treating the data point by point. It can then be interesting to study the feasibility of implementing the inversion methods described in this paper also in these programs. The data obtained is not only useful for the computation of standard errors in the point co-ordinates, but can also be used for data snooping of all observations, using the method by Baarda(1968) modified by Pope (1976).

In this paper, some simplifications have been done which are of no importance for the theoretical derivation. One such is that when the points are reduced, the matrices saved are not (N_{pp}, N_{pi}, N_{pa}), but those obtained after the forward elimination of N_{pp}. These details are of practical importance but will not alter the principles of the data reduction presented. The same is true for the constant column and the solution of the equation systems, but they are obtained in conventional ways and need not be presented here.

Baarda, W. 1968. A Testing Procedure for Use in Geodetic Networks. Netherlands Geodetic Commission, vol 2, No 5, 1968

Larsson, R. 1982a. General triangulation - incorporating photogrammetric and other observations. International Archives of Photogrammetry, vol 24-III, pp 333-342, Otaniemi, Finland 1982

Larsson, R. 1982b. Gentri - a System for Simultaneous Adjustment of Photogrammetric and other Observations. International Archives of Photogrammetry, vol 24-V, York, England 1982

Larsson, R. 1983. Simultaneous Photogrammetric and Geodetic Adjustment - Algorithms and Data Structures. Dissertation, Royal Institute of Technology, Stockholm 1983. Also in Fotogrammetriska Meddelanden nr 2:48, Trita-Fmi-48, 1983

Pope, Allen J. 1976. The Statistics of Residuals and the Detection of Outliers. NOAA Technical Report NOS 65 NGS 1, Rockwille, Md, 1976

Wong, Kam W. 1973. Recursive Partitioning by Direct Random Access. Photogrammetric Engineering, vol 39, No 3, pp 267-274.