ON SPARSE MATRIX TECHNIQUES FOR COMPUTING WEIGHT COEFFICIENT MATRICES

Tapani Sarjakoski
Technical Research Centre of Finland
Finland
Commission III

ABSTRACT

A method for computing selected parts of the rigorous inverse of a positive definite matrix is reported. The method is a generalization of those methods where only the banded or banded-bordered part of the inverse is computed. The use of the method for photogrammetric tasks is illustrated, especially for the procedures of gross error detection.

1. INTRODUCTION

The computation of weight coefficient matrix of unknown parameters of an adjustment problem has been in the interest of photogrammetrists for years. This is due to its computational difficulty: Assuming the triangular factorization of the coefficient matrix of normal equations is available, still approximately $1/3n^3$ multiplications are needed in general for complete inversion. A block adjustment of medium size already involves thousands of unknowns making the rigorous inversion impossible in practice. Additionally, in most cases only the parts of the inverse corresponding the non-zero parts of the original coefficient matrix are required.

The utilization of the banded structure of the coefficient matrix of the reduced normal equations in the solution as well as in the inversion was proposed by Gyer (1967). The inversion requires approximately twice as much work as the factorization. This is reasonable and enables an application of rigorous statistical tests for evaluation of accuracies of individual parameters also in practice.

Later the method - usually discussed under the title of recursive partitioning - has been used especially by Brown (1976) and generalized for banded-bordered matrices. In this presentation the method is further generalized for matrices with arbitrary sparsity structures.

In recent years, the methods for gross error detection have turned toward the use of rigorous statistical methods. The "data-snooping" method by Baarda (1968) and the method proposed by Stevanovic (1978) are the most famous ones. Both of them rely on the use of the weight coefficient matrix of residuals, Q_{vv} matrix. The data-snooping method requires only diagonal elements of Q_{vv} matrix. Effective methods for computing them utilize the sparsity of the design matrix as well as the techniques for computing Q_{xx} matrix which are discussed above. They have been proposed by Klein & Förstner (1981), later by Schwarz et al. (1982) and Larsson (1983).

In this paper the techniques for computing Q_{vv} matrix are further developed. Especially, a method for computing off-diagonal elements is reported. The use of the method is motivated by bringing forth the nature of photogrammetric image observations which favours the use of two-dimensional tests.

2. BLOCK FACTORIZATION

In most adjustment tasks of surveying we have as basic units of unknown parameters rather groups of parameters than single parameters. In the self calibrating bundle adjustment we naturally have a group of unknown parameters for each ground point (3 param. for each), for each photo (6 param. for each) and for each subblock of photos which are assumed to have a common deformation, each subblock having the number of parameters according to the selected model extension. It is thus natural to represent the resulting design matrix and positive definite coefficient matrix of normal equations accordingly.

Let A be a positive definite n·n matrix consisting of m hyperrows and hypercolumns and thus having a partitioning

$$A = \begin{bmatrix} A_{11} & & & \\ A_{21} & A_{22} & \text{symm.} & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix} \qquad (1)$$

each index k (k = 1,...,m) being connected with the dimension n_m corresponding to the number of parameters in each individual group. Matrix A can be factorized to the form

$$A = LDL^T \qquad (2)$$

where L is a lower triangular matrix with identity matrix as L_{kk} (k = 1,...,m),

$$L = \begin{bmatrix} L_{11} & & & \\ L_{21} & L_{22} & & \\ \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ L_{m1} & L_{m2} & \cdots & L_{mm} \end{bmatrix} \qquad (3)$$

and D is a blockdiagonal matrix with diagonal matrices D_{kk} (k = 1,...,m).

The algorithm for constructing L and D by recursive reduction of one hyperrow at a time is as follows:

$$D_{aa}^{(k)} = \bar{A}_{aa}^{(k)} \qquad (4a)$$

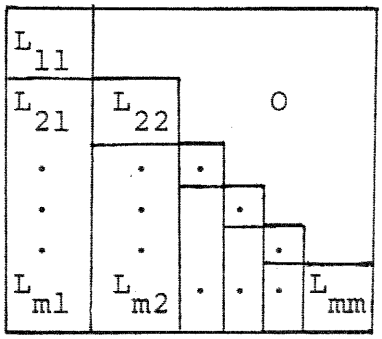
$$L_{ba}^{(k)} = \bar{A}_{ba}^{(k)} \, D_{aa}^{(k)^{-1}} \qquad (k = 1,\ldots,m) \qquad (4b)$$

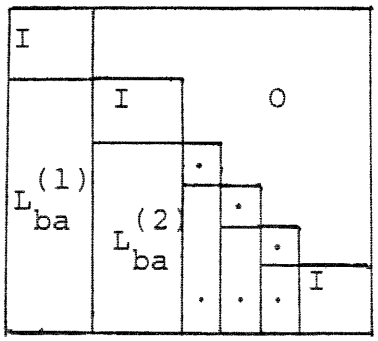
$$\bar{A}^{(k+1)} = \bar{A}_{bb}^{(k)} - L_{ba}^{(k)} \, \bar{A}_{ba}^{(k)^{T}} \qquad (4c)$$

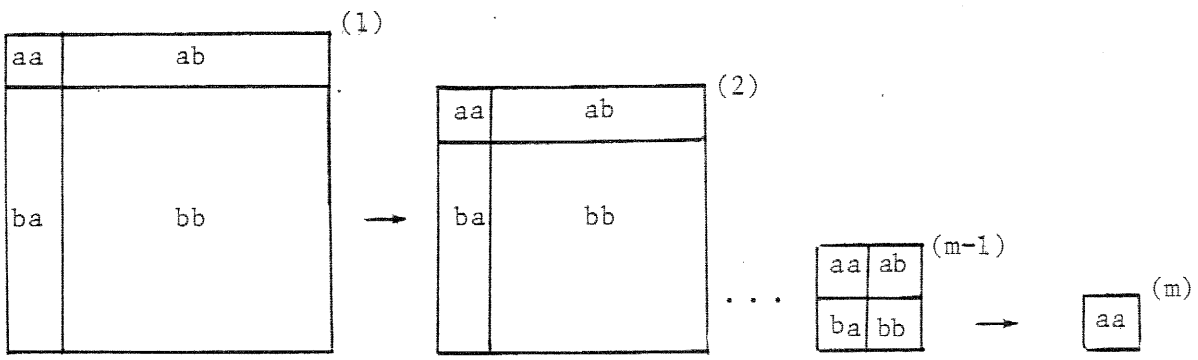
where $\bar{A}^{(k)}$ notes the reduced coefficient matrix at step k and $\bar{A}^{(1)} = A$.

The diagrams below show the correspondence of the partitionings:

$$L = \qquad = \qquad (5)$$







Matrices A and D are partitioned respectively.

The analogy between the conventional LDL^T-factorization and the above given block-LDL^T-factorization is obvious: Instead of single elements we have submatrices as working units. Zero matrices can be easily skipped in computations because the partitioning used in the decomposition fits with the natural partitioning of the problem.

The algorithm above follows the idea of the outer product technique (see e.g. George & Liu 1981). However, also the inner product technique and the bordering technique (see ibid.) can be applied analogically. They render double precision computations without usage of double precision storage space (contrary to the outer product technique) being thus favourable in some circumstances.

The hyper-Cholesky method (Ackermann 1972) as well as the recursive partitioning used by Gyer and Brown can be derived of the block-LDL^T-factorization by treating the submatrices D_{kk} (k = 1,...,m) in different ways.

3. INVERSION

Noting by B the inverse of A and following the partitioning of (4) and (5) we get a recursive algorithm for inversion:

$$\left.\begin{aligned} B_{ba}^{(k)} &= - B_{bb}^{(k)} L_{ba}^{(k)} && (6a)\\ B_{aa}^{(k)} &= D_{aa}^{(k)\,-1} - L_{ba}^{(k)\,T} B_{ba}^{(k)} && (6b) \end{aligned}\right\} \quad (k = m,\ldots,1)$$

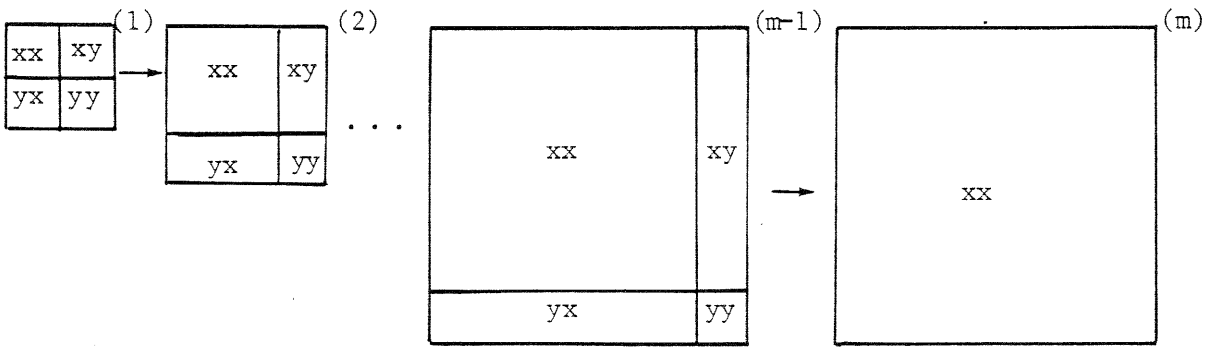
The algorithm above is the conventional method working backwards. Alternatively we can construct the inverse by an approch working forwards, which is known as the bordering method. Before its introduction the partitionings and the notations are revised:

$$\left.\begin{aligned} A^{(k)} &= \left[\begin{array}{c|c} A_{xx}^{(k)} & A_{xy}^{(k)} \\ \hline A_{yx}^{(k)} & A_{yy}^{(k)} \end{array}\right] && (7a)\\ A_{xx}^{(k)} &= A^{(k-1)} && (7b) \end{aligned}\right\} \quad (k = 1,\ldots,m)$$

$$A^{(1)} = A_{11}, \qquad A^{(m)} = A \qquad (7c),\ (7d)$$

$$C^{(k)} = A^{(k)\,-1} \qquad (k = 1,\ldots,m) \qquad (8a)$$

$$C = B = A^{-1} \qquad (8b)$$



Additionally, the partitioning of L and D is revised respectively. The bordering algorithm where the inverse is updated on each recursion cycle is as follows:

$$L_{xx}^{(k)T}\, M_{yx}^{(k)T} = L_{yx}^{(k)T} \qquad \text{solve } M_{yx}^{(k)} \qquad (9a)$$

$$C_{yy}^{(k)} = D_{yy}^{(k)-1} \qquad (9b)$$

$$C_{yx}^{(k)} = -\, C_{yy}^{(k)}\, M_{yx}^{(k)} \qquad (k = 1,\ldots,m) \qquad (9c)$$

$$C_{xx}^{(k)} = C^{(k-1)} - M_{yx}^{(k)T}\, C_{yx}^{(k)} \qquad (9d)$$

$M_{yx}^{(k)}$ from (9a) will actually be solved by a backwards substitution process applying the a/b-partitioning for $L_{xx}^{(k)}$, $L_{yx}^{(k)}$ and $M_{yx}^{(k)}$ respectively.

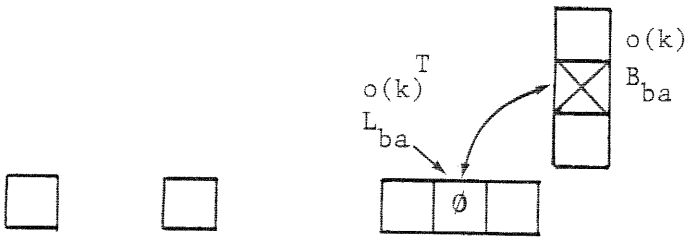
4. SPARSITY IN INVERSION

4.1 Formulas (6a) - (6b): Sparse matrix techniques used in factorization are well known. In inversion their usage is more complex because the inverse is usually a full matrix. However, the computation of inverse can be restricted to certain submatrices:

For computing diagonal submatrices of inverse B according to (6a) - (6b), it is necessary to compute only those submatrices corresponding to the non-zero submatrices of L.

The following reasoning is given to prove that this holds (see also Erisman & Tinney 1975):

Let us investigate the computation of inverse on step (k) of (6a) - (6b):



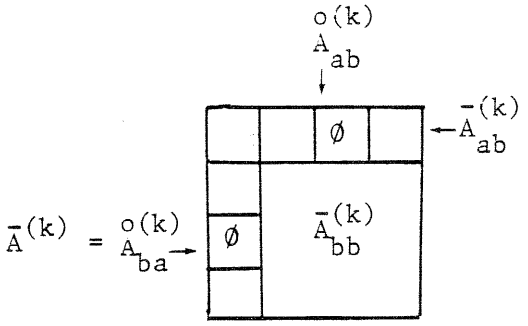
(10)

$$B_{aa}^{(k)} = D_{aa}^{(k)-1} - L_{ba}^{(k)T}\, B_{ba}^{(k)} \qquad \text{(cf. (6b))}$$

We see directly that it is not necessary to have the value of submatrix $B_{ba}^{o(k)}$ corresponding to zero submatrix of $L_{ba}^{o(k)}$.

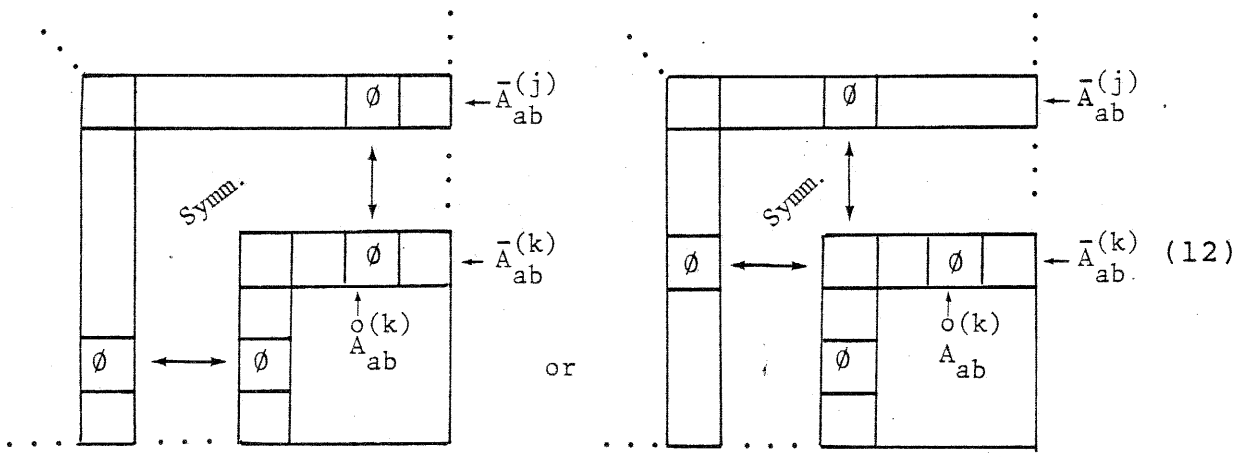
We still have to show that the value of $B_{ba}^{o(k)}$ is not needed on the subsequent steps (j = k-1,...,1) of (6a) - (6b).

Let us first look at the generation of fill-in on step (k) of (4a) - (4b): From (4a) we can conclude that $L_{ba}^{(k)}$ of (10) has exactly the same sparsity structure as $\bar{A}_{ba}^{(k)}$:



(11)

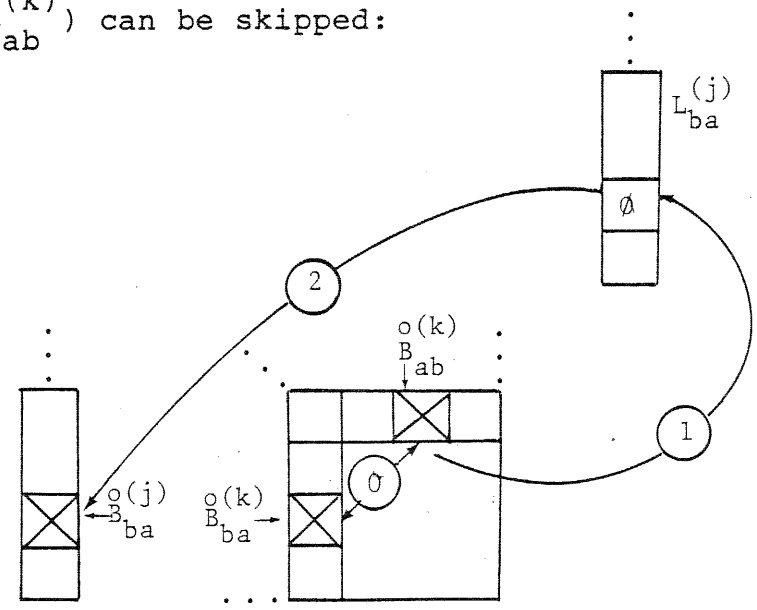
On previous recursion cycles (j = k-1,...,1) of (4a) - (4b) $\bar{A}^{(j)}$ must have structure



(12)

otherwise $\overset{o}{A}{}_{ab}^{(k)}$ would be filled.

Now we are ready to prove that the computation of $\overset{o}{B}{}_{ba}^{(k)}$ (and $\overset{o}{B}{}_{ab}^{(k)}$) can be skipped:



(13)

(cf.(6a))

$$B_{ba}^{(j)} = - B_{bb}^{(jj)} L_{ba}^{(j)} \qquad (j = k-1,\ldots,1)$$

0) these submatrices are not computed on step (k) according to the reasoning of (10)

1) the necessary condition for avoiding fill-in, case I, case II is symmetrical with case I

2) the reasoning of (10) is applied for the subsequent steps (j = k-1,...,1)

From the matrix product of (13) can be seen that matrices $\overset{o}{B}{}^{(k)}_{ba}$ and $\overset{o}{B}{}^{(k)}_{ab}$ are not needed on any step (j = k-1,...,1) of (6a) - (6b). Due to recursion this holds for any submatrix of $B^{(i)}$ corresponding to a zero submatrix of $L^{(i)}$ on any subsequent step (i = k-1,...,1).

4.2 Formulas (9a) - (9d): Because $M^{(k)}_{yx}$ in (9a) - (9d) is the solution of the blocktriangular system (9a), it is a full matrix even if the right hand side, $L^{(k)}_{yx}$ is sparse. Thus we can not save computations in (9a). However, sparsity can be fully exploited in the matrix product of (9d) by updating only desired submatrices of $C^{(k)}_{xx}$. The method is applicable especially to tasks related with regression analysis and self calibrating block adjustment (Sarjakoski 1984).

5. Q_{vv} MATRIX

5.1 Basics: In the adjustment model of indirect observations the weight coefficient matrix of residuals, Q_{vv} matrix, is given by

$$Q_{vv} = Q_{ll} - Q_{\hat{l}\hat{l}} \tag{14}$$

$$Q_{ll} = A\, Q_{xx}\, A^T \tag{15}$$

Q_{vv} weight coefficient matrix of residuals

Q_{ll} weight coefficient matrix of given observations

$Q_{\hat{l}\hat{l}}$ weight coefficient matrix of estimated values of observations

Q_{xx} weight coefficient matrix of unknown parameters, the inverse of the coefficient matrix of normal equations

A design matrix

5.2 Sparsity: In the data-snooping method of Baarda only diagonal elements of Q_{vv} are needed. By splitting (15) into two parts

$$F = A\, Q_{xx} \tag{16}$$

$$Q\hat{\hat{l}}\hat{\hat{l}} = F\, A^T \tag{17}$$

we see that for computing them

1) only those parts of F are needed which correspond to the non-zero parts of A

2) for computing these parts of F, only the parts of Q_{xx} are needed which correspond to the non-zero parts of coefficient matrix of normal equations.

In computation of Q_{xx} for 2) we can utilize the sparse matrix techniques explained above. Of special importance is that they hold for any sparsity structure.

5.3 Two-dimensional tests: The data-snooping method handles each observation as an individual thus using one-dimensional statistical tests. The photogrammetric image observations however consist of pairs of x- and y-observations, which should always be treated as a single "joint event".

The motive for treating a pair of observations as a joint event lies in the measuring process already: We are actually measuring the geometrical location of each image point, the coordinate system is used as a uniform tool for registration. Additionally, we are incapable of measuring only one of the coordinates because, to be precise, the pointing must be made in two perpendicular directions.

The F test is the proper statistical test, if the reference variance is estimated from the adjustment model. Its usage requires the 2·2 diagonal submatrix of Q_{vv} matrix for each pointing, with the off-diagonal elements being computed, too. The computation is effective because the rows of design matrix A for x- and y-observation of each pointing have the same sparsity structure; thus no extra elements of F matrix are needed.

The joint treatment gives some practical advantages: It makes the test invariant with respect to the camera coordinate system, thus avoiding the problems pointed out by Förstner (1981). It cancels completely the question about what to do if, according to a one-dimensional test, only one of the observations is erroneous.

REFERENCES

Ackermann, F., 1972: Experience with Block Triangulation by Independent Models. Proceedings of 38th Annual meeting, ASP.

Baarda, W., 1968: A Testing Procedure for Use in Geodetic Networks. Netherlands Geodetic Commission, Publications on Geodesy, Volume 2, No. 5, 1968.

Brown, D.C., 1976: The Bundle Adjustment - Progress and Prospects. International Archives of Photogrammetry, Volume 21, invited papers, Part 3, Helsinki.

Erisman, E.M., Tinney, W.F. 1975: On Computing Certain Elements of a Sparse Matrix. Communications of the ACM, Volume 18, Number 3, March 1975.

Förstner, W., 1981: Statistische Grundlagen fuer die Zuverlässigkeit von Ausgleichungsergebnissen. Vorträge des Lehrgangs Numerische Photogrammetrie (IV) an der Universität Stuttgart, Heft 7, Stuttgart.

George, A., Liu, J.W-H., 1981: Computer Solution of Large Sparse Positive Definite Systems. Prentice-Hall, Englewood Cliffs 1981.

Gyer, M.S., 1967: The Inversion of the Normal Equations of Analytical Aerotriangulation by the Method of Recursive Partitioning. RADC-TR-67-69, Rome Air Development Center, Rome, New York.

Klein, W., Förstner, W., 1981: Strategien fuer die Fehlersuche in der Aerotrianglulation. Vorträge des Lehrgangs Numerische Photogrammetrie (IV), Universität Stuttgart Heft 7, Stuttgart.

Larsson, R., 1983: Simultaneous Photogrammetric and Geodetic Adjustment - Algorithms and Data Structures, thesis, Stockholm.

Sarjakoski, T., 1984: Efficient Methods for Selecting Additional Parameters of Block Adjustment. A presented paper at the XV Congress of ISPRS, Rio de Janeiro.

Schwarz, P.G., Joosse, M., Melissen, G.M.W.J., 1982: FOTEF - Data Snooping in Independent Model Triangulation. International Archives of Photogrammetry, Volume 24-III, Helsinki.

Stevanovic, P., 1978: Blunders and Least Squares. ITC Journal 1978-1.