SOREN W. HENRIKSEN
*Army Map Service**
Washington, D. C. 20315

Mathematical Photogrammetry

ABSTRACT: *Extension of the analytic photogrammetry procedures used on small areas to the solution of photogrammetric problems involving extremely large areas is not simply a matter of scaling upward. As the size increases, the twin problems arise of properly handling the increased connectivity between variables and of keeping the problem within the solving ability of available computers. Fortunately, the increase in size allows a large variety of problems which are different* in the small *to be treated in a common manner. The basic mathematical formulas common to problems* in the large *are given in their most general form, and the treatment and methods of solution are described. Many parts of the general problem are explored inadequately even yet, and some of these are discussed.*

THE WORD *Photogrammetry* covers activities in a diversity of fields. Because all of these activities involve the measurement of photographs, a discipline exists which is common to all these fields. If the measurements are of positions on the photographs, the discipline can be further particularized by a statement in mathematical terms. In the following set of three studies, the mathematical basis common to terrestrial photogrammetry in the large, establishment of lunar control (in its present form), and the determination of star positions is set down. The basis could be extended even farther into satellite tracking, nuclear particle tracking, mapping by radar, etc; the three specific subjects here considered were selected by the writer because they have been and are being worked on.

At their present stages, all three fields are using essentially the same mathematical technique. In each, we have given a set of photographs on which are identifiable image points; these points have associated with them measured coordinate pairs (x^1, x^2). The image points are mapped into a set of points in object space; these points have coordinates (X^1, X^2, X^3), not necessarily cartesian and not necessarily in 3-space. Some of these coordinates are known, either from direct measurement or by computation from auxiliary data. In addition, there are given conditions which must be satisfied by the object space coordinates. The mapping

$$(x^1, x^2) \rightarrow\rightarrow (X^1, X^2, X^3) \tag{1}$$

is considered defined by a number of parameters

$$(p_1, p_2 \cdots p_n)$$

and these parameters, related by condition equations, must also be found.

SINCE the mapping $(x^i)\rightarrow\rightarrow(X^i)$ is nonlinear, the first step is to reduce the problem to a linear form. This "linearization" is also advantageous in that the total number of equations available exceeds the number of unknowns so that a statistical adjustment is needed; such an adjustment is easier to make if the relationships are linear. In its linear form, the problem is stated in terms of corrections to the variables involved, rather than in terms of the variables themselves. After linearizing the problem and separating the new variables into "known" and "unknown" sets, the transformations which constitute most of photogrammetry—the mapping of 2-space or 3-space into a set of 3-space—are described mathematically by the equations

$$[\Delta x] = [\overline{A}][\Delta X] + [\overline{B}][\Delta p] + [\overline{C}][k] \tag{2}$$

$$[O] = [\overline{D}][\Delta X] \tag{3}$$

$$[O] = [\overline{E}][\Delta p]. \tag{4}$$

These relate the corrections $[\Delta X]$ to the "causative" variables with the corrections $[\Delta p]$ to the mapping parameters through the observation equations in the observables $[\Delta x]$ and the condition equations (and Lagrangian multipliers k). The entire problem can be now comprehended in the equation

* Mr. Henriksen is now affiliated with Autometric-Raytheon, Alexandria, Virginia.

$$[\Delta x] = [A][\Delta X] \tag{5}$$

where the vector on the left is the discrepancy or residual vector, and the vector on the right is the correction vector containing all the unknowns, including Lagrangian multipliers, etc. (Weights are assumed already introduced into the matrix and are not noted explicitly.)

The solution is then obtained by the usual process

$$[\Delta X] = [A]^{-1}[\Delta x] \tag{6}$$

where the reciprocal of the observation matrix is to be understood in the sense of Bjerhammer[1]. The covariance matrix of the unknowns is then,

$$[\Sigma^2(\Delta X)] = [A]^{-1}[\Sigma^2(\Delta x)][A]^{-T} \tag{7}$$

where $[\Sigma^2(\Delta x)]$ is the covariance matrix of the observations.

As will be mentioned again in the discussion of star positions, a slightly different approach to the problem is possible by considering all equations as condition equations[2] rather than dividing them into two sets: observation equations and condition equations. In the linear case, the two approaches are equivalent; in the non-linear cases they are not, and the implication for non-linear cases are being studied.

In the formulation of the problem, the work of setting up Equation 5 becomes difficult if the number of different categories of residuals and unknowns is considerable. Simplification is achieved by using the multidimensional matrix (tensor) calculus of G. Kron,[3] but since Kron's calculus was set up excluding non-(hyper)-cube matrices as well as reciprocals, modification is needed. Just as the inverse of a non-square can be defined usefully,[2] so can operations on rectangular matrices in hyper-space. Transposition, addition, and multiplication are definable if attention is paid to the order in which the operations along the various axes are taken. Inversion is much more difficult to define. Fortunately, only the inversion of hyper-cube matrices need be considered and this operation always can be accomplished in practice by first mapping the matrix into a two-dimensional form and then inverting.

IT IS WORTHWHILE giving special attention to the fact that the theory here presented assumes for the stochastic variables a Gaussian distribution, so that the most probable values of the unknowns are found by minimizing the sum of the squares of the deviations. If the distribution is non-Gaussian, the least-squares criterion, in general, will not be valid, as can be shown easily. That the distribution is not Gaussian is fairly certain; a recent paper by Stearn[6] shows a non-Gaussian distribution for angles measured with a theodolite and a similar situation can be expected to hold for angles measured by a camera. Stearn also shows that in his experiments the difference between the Pearson curve actually fitted and a Gaussian curve is too small to cause trouble; this need not hold true in photogrammetric problems. If the true distribution can be identified, the solution is obvious if not necessarily simple; otherwise, real difficulties arise. A non-Gaussian distribution, if it does not differ too much from the Gaussian, can be introduced as a perturbation on the Gaussian. An alternative, in those cases where the deviation is large, is to introduce a weighting function which will let the least-squares procedures be used.

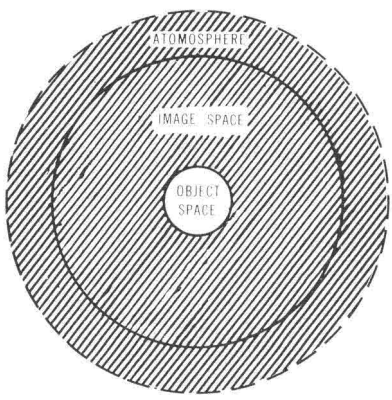

FIG. 1

In a series of papers to follow, the application of the above mathematical formalism to specific problems will be considered.

THE FIRST problem is an extended concept of terrestrial analytic photogrammetry. Here both image space and object space lie in the same atmosphere, with a global, or at least, multiply-connected, coverage of the object space by the image space (Figure 1).

The second problem is that of establishing "control" points on the moon using earth-based photography. In this problem only the image space lies within an atmosphere, and the object space coverage is only semi-global but simply connected (Figure 2).

The third and last problem is that of finding the coordinates (exclusive of distance) of all stars brighter than a certain magnitude, again using earth-based photography. As in the second problem, only image space is

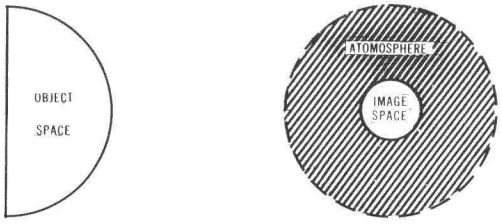

FIG. 2

within the atmosphere, but now the coverage is again global. Image space now lies completely inside object space instead of completely outside, and the two spaces are extremely far apart instead of very close together (Figure 3).

Investigation of the mathematical basis common to the three examples mentioned shows that two major and related mathematical difficulties are to be expected:

1. Organization of the observation matrix $[A]$ into such a form that $[A]^{-1}$ can be most easily computed; and
2. Creation of algorithms which will permit numerical solution of photogrammetric problems within a reasonable interval of time.

Neither of these difficulties can be considered solved as yet. In special cases,[4] very rapid computing methods can be found and even demonstrated to be optimum. Whether Equations 5 and 7 can be made to fit into these broad cases, or whether the special cases can be broadened to cover the equations above is not yet known for certain, but studies so far show that the two difficulties can be brought down to a reasonable size.

Because all three problems have a similar topological structure, it can be shown that the matrices $[A]$ will have the same form for all three problems, and this form is one which allows the non-zero elements to be grouped in a symmetric fashion about the main diagonal.

THE FORM of the $[A]$ matrix in Equation 2 depends on the form and arrangement of the $[\Delta X]$ matrix. If the Kron formalism is used, the $[A]$ matrix has small, completely filled blocks of matrices along the main diagonal. If this matrix is peeled apart and rearranged to give the familiar two-dimensional matrix, the blocks along the main diagonal are separated and re-distributed to give a much more open matrix (in the sense of having zero matrices interspersed among the basic 3×3 matrices). A useful viewpoint is to think of the image space as a covering of a projection of the object space. The mapping matrix then is a good picture of the projection in its hyper-dimensional form, and this can be transformed into the two or one dimensional form, as the special problem may dictate.

IN A GLOBAL PROBLEM where the mapping of object space into image space is only partial for a particular set of images (i.e., only a part of object space is mapped onto a particular photograph), the form of the expanded $[A]$ matrix is

$$\begin{bmatrix} x & & & & & & & & & 0 \\ & x & x & & & & & & & \\ & & x & x & x & & & & & \\ & & & x & x & x & x & & & \\ & & & & x & x & x & x & x & \\ & & & & \cdot & \cdot & & \cdot & \cdot & \\ & & & & & \cdot & \cdot & & \cdot & \cdot \\ & & & & & & x & x & x & \\ & & & & & & & & x & x \\ 0 & & & & & & & & & x \end{bmatrix}. \tag{8}$$

Each row represents a mapping of a particular group of object space points onto a particular image-set (photograph); each x represents a rectangular matrix. Each rectangular sub-matrix x is itself made up of 3×3 (or smaller size) sub-matrices scattered irregularly along the length of the larger sub-matrix. The degree of irregularity and amount of scattering depends on organization of the object-space matrix; in most cases the form will be somewhat as

$$[yy \cdots y0 \cdots 0y] \tag{9}$$

where the y's denote sets of 3×3 matrices. Absence of points in particular sections of image space does not effect the general form (8); it merely thins out the sub-matrices of form (9). If carried too far, however, it may

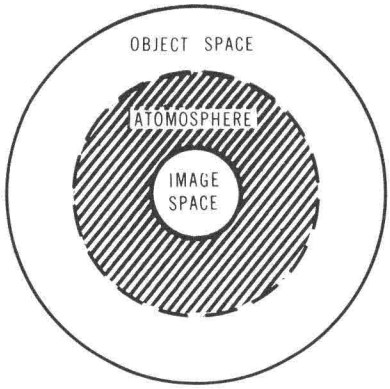

FIG. 3

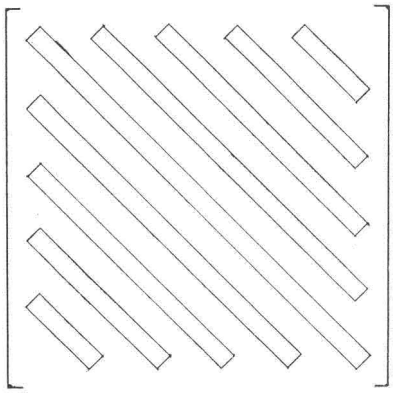

FIG. 4

result in a lack of overlap between rows, in which case the matrix would separate into independent sub-matrices.

The parameter matrix $[B]$ in Equation 2, has the form

$$\begin{bmatrix} x & & & & & & \\ & x & & & & & \\ & & x & & & & \\ & & & x & & & \\ & & & & x & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & & & & & & \cdot \\ & & & & & & & & x \end{bmatrix}$$

where the individual sub-matrices $[x]$ have the dimensions "number of images in photograph" by "number of parameters per photograph."

It has been pointed out[2] that the configurations of matrices $[\overline{A}]$ and $[\overline{B}]$ are not independent, and that the arrangement adopted for the $[\Delta x]$ matrix should perhaps be such as to optimize the computation of the quantities sought.

When the coverage is not global, but is still connected, the $[\overline{A}]$ matrix, Equation 3, is truncated:

$$\begin{bmatrix} [x\ x\ x\ x\ x] & & & & & & & \\ & [x\ x\ x\ x\ x] & & & & & & \\ & & [x\ x\ x\ x] & & & & & \\ & & & [x\ x\ x] & & & & \\ & & & & [x\ x\ x\ x] & & & \\ & & & & & [x\ x\ x\ x] & & \\ & & & & & & [x\ x\ x] & \\ & & & & & & & [x\ x] \\ 0 & & & & & & & & [x] \end{bmatrix}. \qquad (10)$$

As the mapping of image space into object space approaches completeness for one or more subsets of image space, the sub-matrices become longer, so that when each photography covers the entire object, the matrix (8) fills up:

$$\begin{bmatrix} [x\ x\ x\ \cdots\ x] \\ [x\ x\ x\ \cdots\ \cdot] \\ [x\ x\ x\ x\ x] \\ [x\ x\ x\ \cdots\ x] \end{bmatrix}.$$

SOLUTION of the Equation 5 is not as easy as a cursory glance would indicate. First, the matrices are very much larger than are met within problems "in the small." The $[\Delta x]$-matrix, for instance, may be of the order of

$$1000 \times 10{,}000 \times 2 \times 10;$$

the $[\Delta X]$ matrix of the order of

$$1{,}000{,}000 \times 3 \times 2;$$

and the $[A]$ matrix correspondingly huge. A straight-forward inversion process is, of course, out of the question, and even a Gauss-Seidel iterative process applied directly would be unallowably long. Furthermore, the variance matrix $[\Sigma^2 (\Delta X)]$ must be computed, as well as the confidence limits for the elements of the variance matrix (v.i.). The very first step, therefore, must be the multiplication of Equation 5 by a permutation matrix chosen to make the solution easier. Investigation thus far[4] shows that a permutation matrix can be found which rearranges $[A]$ so that the non-zero elements occur in bands symmetrically placed about a diagonal; schematically, somewhat as shown in Figure 4.

If this matrix is manipulated, by transposition and inversion, the tendency will be for the bands to "fold over" onto each other and thus to limit the spread of the non-zero elements. For each particular problem, the permutation matrix to be used will be different, and the search for the proper matrix is a major part of the problem.

ONCE the question of the matrix topology has been settled, solution for $[\Delta X]$ becomes a computational problem involving iterative relaxation procedures. In order to preserve storage space and to keep the computing time per iteration from growing beyond bounds, significance intervals must be established for all elements of $[\Delta x]$ and $[\Delta X]$. The center of each significance interval is then shifted downward (i.e., toward zero) with

each iteration in such a manner that the number of significant figures involved in each iteration is kept approximately constant, or at any rate does not grow beyond the machine's capacity before the pre-set and of the compution.

After computation of $[\Delta X]$, the next step is the computation of confidence intervals (or their equivalent) for the $[\Delta X]$. This involves finding the variance matrix $[\Sigma^2\ (\Delta X)]$ from Equation 7. The $[\Sigma^2\ (\Delta X)]$ can either be used directly or combined with the T^2-function

$$T^2(\alpha) = \frac{(N-1)}{N-p}\, p \cdot F_{p,N-p(\alpha)}$$

where F is Snedecor's non-central F-function with p and N-p degrees of freedom, to define an ellipsoid

$$T^2(\alpha) \geq N([\Delta X]-[\overline{\Delta X}])^T[S^2]^{-1}([\Delta X]-[\overline{\Delta X}]) \quad (11)$$

This is the equation of an ellipsoid whose size is determined by the significance level parameter α and whose shape is determined by $[S^2]$, the sample variance.

If the variance matrix $[\Sigma^2\ (\Delta X)]$ is completely diagonal, then confidence limits for the elements of $[S^2]$ can be set up in terms of the χ^2-function, since

$$\frac{(N-1)S^2}{\sigma^2}$$

has a χ^2-distribution with N-1 degrees of freedom. In the multivariate case, however, the covariances will in general not be zero. If the χ^2-distribution is to be applicable, a rotation to principle axes[7] can be tried, but such a rotation will not be possible if the $[\Delta X]$ units are inhomogeneous.

The Wishert distribution function is defined by

$$w(A, \Sigma^2, N) \equiv \frac{|A|^{1/2(N-p-2)}\,|\Sigma^2|^{-1/2(N-1)} \exp\{-\frac{1}{2}\operatorname{tr}\Sigma^{-2}A\}}{2^{1/2(N-1)p}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{N-i}{2}\right)} \quad (12)$$

and is the probability density function for the elements a_{ij} of the matrix

$$[A] \equiv (N-1)[S^2(\Delta X)].$$

The matrix $[S^2]$ is the sample (maximum likelihood estimate of the population variance matrix $[\Sigma^2\ (\Delta X)]$ with dimensions $p \times p$. The cumulative distribution function (*c.d.f.*) for a_{ij} is obtained by integrating over the space of $[A]$ between 0 and a_{ij}, or between $-a_{ij}$ and $+a_{ij}$ for $i \neq j$. In the photogrammetric problem, the value of $[A]$ will be known, and an estimate of the probable range of variation of Σ^2 is wanted.

Properly, the *c.d.f.* $W(\Sigma^2, N)$ of $w(S^2, \Sigma^2, N)$ should be differentiated with respect to $[\Sigma^2(\Delta X)]$ to get the probability function associated with that matrix, and to derive from that the fiducial limits,

$$[\Sigma^2]_1 \quad \text{and} \quad [\Sigma^2]_2$$

such that

$$1-\alpha = Pr\{[\Sigma^2]_1 \leq [\Sigma^2] \leq [\Sigma^2]_2\}.$$

For all practical purposes, a set of fiducial limits on $[\Sigma^2(\Delta X)]$ can be gotten from the *c.d.f.* by solving the equations

$$(1-\tfrac{1}{2}\alpha) = W([\Sigma^2]_1, N)$$
$$(\tfrac{1}{2}\alpha) = W([\Sigma^2]_2, N)$$

for $[\Sigma^2]_1$ and $[\Sigma^2]_2$, and then assigning the probability $(1-\alpha)$ to the bracket

$$\{[\Sigma^2]_2 \leq [\Sigma^2] \leq [\Sigma^2]_1\}.$$

This does not mean that there is a probability $(1-\alpha)$ of $[\Sigma^2]$ lying between the limits shown, but the difference between this interpetation and the true interpretation is not important for the problems here discussed.

A CAREFUL note should be made that these procedures and criteria are not necessarily correct if the observation population is non-Gaussian. The non-Gaussian character of the observation is almost certain, but it is also reasonable to assume that the original departure from the Gaussian is small enough that no serious effect occurs. Emphasis is put on *original*, because weighting of observations almost certainly will change the nature of the distribution, and such change can be quite severe without being detectable. The effects usually will be to bring $[\Sigma^2(\Delta X)]$ closer to [0] and to bring $[\Sigma^2]_1$ and $[\Sigma^2]_2$ closer together. A sense of euphoria is induced without a corresponding improvement in the basic health of the data.

It is, of course, impossible to solve the equation

$$[\Delta x] = [A][\Delta X]$$

in one step on any computing machine in existence, and an iterative procedure must be used. Of the various iterative procedures that have been investigated, the over-relaxation process[4] appears to be the most useful. The size of the step from one iteration to the next is controlled by a relaxation factor ω which is related to the maximum eigenvalue of $[A]^T[A]$. For $\omega = 1$, the familiar Gauss-Seidel process results; for $\omega > 1$, the over-relaxation process. It can be shown that the over-relaxation process converges asymptotically more rapidly than does the Gauss-Seidel. A process known as the semi-iterative process achieves faster convergence by using an estimation $[\Delta Y]m$ for $[\Delta X]$ at the m-th step which is based on the two preceding estimations $[\Delta Y]_{m-1}$ and $[\Delta Y]_{m-2}$. Of course, the more rapid convergence must be paid for by the need for storage of $[\Delta Y]_{m-2}$ as well as $[\Delta Y]_{m-1}$ at each step. Furthermore, it is easy to see that for $m >> 1$, the rate of convergence approaches that of the over-relaxation method, so that the initial advantage is lost.

BESIDES the use of automatic, iterative, relaxation procedures, there are several other techniques which must be used if the extended photogrammetric problem is to be solvable. One already mentioned is that of estimating at each step m the number $\alpha(m)$ of significant figures which must be carried along in $[A]$ and the associated matrices, and of arranging the mechanical computation so as to make use of the space that the dropped significant figures make available.

It is characteristic of and a major advantage of the iterative method that the number of significant figures that must be carried along in the computations can usually be kept constant from one iteration to the next. The amount of use one can make of this characteristic depends among other things on the eigenvalue spectra encountered, and particularly (e.g., ref. 6) on the ratio $\lambda\, min/\lambda max$ where λmin and λmax are the least and the largest eigenvalues. The process may be thought of as setting up a mathematical comb filter whose width is a function of the number of operations to be performed at each step and whose center is a function of this step. From step to step, the filter moves along the real axis to cut out more significant figures and include less significant figures. Elimination of steps in the iteration procedure is also possible by including second order terms or higher in the Taylor series expansion of $[\Delta X]$, i.e., writing

$$[\Delta x] = [A]_1[\Delta x] + [A]_2[Ax]^2 + \cdots .$$

Very little has been done in this field; the computation of $[A]_2$, which is the Jacobian

$$\left[\frac{\partial^2 \Delta x}{\partial \Delta X^2}\right]$$

is complicated, and the additional storage space required is a serious drawback. The semi-iterative procedure mentioned earlier is equivalent to the inclusion of the second-order terms, in the same sense that the method of steepest descent is equivalent to the method of least-squares, and may be used instead.

A third technique which looks very promising is that of first solving for a selected subset $[\Delta X]s \subset [\Delta X]$, the subset chosen being the maximum that can be handled by the computer available. Successive subsets $[\Delta X]_{s+n}$,

$$\{\{[\Delta X]\} - \{[\Delta X]_s\}\}$$

are then added to $\{[\Delta x]_s\}$ and the sequence of equations

$$[\Delta x]_s = [A]_s[\Delta X]_s$$
$$[\Delta x]_{s+1} = [A]_{s+1}[\Delta X]_{s+1}$$
$$\cdots\cdots\cdots$$
$$[\Delta x]_N = [A]_N[\Delta X]_N$$

solved. Since

$$[X] = [X]_{s+n} + [\Delta X]_{s+n}$$

and since $[\Delta X]_{s+n} \to\to\to O$ as $s+n \to\to\to N$ at the same time that $[\Delta^2 X]_{s+n,m}$ approaches zero with $m \to \infty$, the matrices $[A]_{s+n}$ can be kept simple.

When all data are already at hand, a partitioning technique may be found which better fits the given distribution of data. When only a part of the eventually complete set of data is available (and this is particularly true of terrestrial photogrammetry and within a few years will be true of extra-terrestrial photogrammetry), the sequential approach will be the most efficient.

IN SUMMARY one can say that a very great variety of photogrammetric problems is solvable by essentially the same set of mathematical equations which are of fairly simple form. The only real difficulties are those imposed by the limited means of computation available, and these may be diminished by looking for a mathematical basis which ex-

tends into other fields where similar difficulties have been found and surmounted.

References

1. Bjerhammer, Arre; 1951, *Applications of Calculus of Matrices to Methods of Least Squares*, Translation, Royal Institute of Technology, Stockholm, Sweden.
2. Cooge, William; 1964, Unpublished Report, Army Map Service, Washington, D. C.
3. Kron, G., 1939; *Tensor Analysis of Networks*, John Wiley & Sons, New York.
4. Varga, Richard; 1962, *Matrix Iterative Analysis*, Prentice-Hall, Inc., New Jersey, pp. 97–129.
5. Lanczos, C., 1961, *Linear Differential Operators*, Van Nostrand & Company, New York, pp. 129-ff.
6. Stearn, J. L., 1964, "Tests of Departure from the Normal Distribution for Theodolite Errors of Observation," *The Canadian Surveyor*, Volume 18, Number 1, pp. 50–53.
7. Anderson, T. W., 1958, *Introduction to Multivariate Statistical Analysis*, J. Wiley & Sons, New York, pp. 272-ff.