NONLINEAR ARRAY ALGEBRA IN DIGITAL PHOTOGRAMMETRY

Urho A. Rauhala
General Dynamics Electronics
P.O.Box 85468 MZ 6108-A
San Diego, CA 92186

## ABSTRACT

A highly automated image mensuration technique of Global Least Squares Matching (GLSM) and the subsequent estimation processes of digital photogrammetry are expanded by a unified theory of estimation, numerical analysis and signal processing of nonlinear array algebra. The classical Newton-Gauss (NG) normal equation matrix overlooks the linear component arising from the product of the 3-D array of second order partials with the constant column of observables in the associated Newton-Raphson (NR) matrix. The array algebra expansion of polynomials can handle the higher order Taylor terms of nonlinear normals. A general set of nonlinear estimation techniques using minimized residuals is shown, including the nonlinear expansion of robust estimation and least squares as special cases. An analytic theory of direct linearized systems of nonlinear estimation is introduced. A new constrained nonlinear solution integrates several iterations and initial values beyond NR and other known techniques. It often converges in one iteration, far beyond the NG convergence limit. The nonlinear GLSM application can reach a stereo mensuration speed of thousands of points per sec using today's low cost computers and softcopy systems automating the emerging digital image mapping process.

Key Words: Nonlinear, Array, Algebra, Estimation

## 1. INTRODUCTION

Computer technology advances allow application of the mapping procedures of analytical photogrammetry to digital stereo images. However, without the integration of similar array algebra advances in software of multi-ray digital photogrammetry, the resulting "digital analytical plotter" remains labor intensive, costly and slow. An advanced softcopy mapping system requires an integration of the following three developments:

1. Advances in array algebra for unified statistical estimation, numerical analysis and signal processing
2. Exploitation of multi-ray models in image mapping
3. Continuing refinements of softcopy workstations.

Advances in computer technology reduce problems in developments 2 and 3. Efficient software and systems integration of array algebra will reduce them further or may in fact eliminate them altogether. This explains why array algebra is the focus of this paper, especially its application to the nonlinear Global Least Squares Matching (GLSM) technique, (R et.al, 1992). References to "Rauhala" will be abbreviated by R herein. GLSM image mensuration and related triangulation techniques of array algebra are integrated into automated digital image mapping and object reconstruction in (R 1986, 1988, 1992). The resulting system of Global Digital Image Map (GDIMAP) expands multi-ray stereo intersections of a bundle adjustment from the few traditional tie points to very dense (1-4 pe spacing) Digital Elevation Model (DEM) and feature mensuration points of softcopy stereo models using array algebra.

Reviews of array algebra can be found in (R 1972, 1974, 1976, 1980, 1986, 1990), (R et.al., 1989). Array algebra was introduced in photogrammetry and geodesy as an expansion of adjustment calculus in surveying mathematics. Surveyors, map makers (cartographers, geodesists, photogrammetrists etc.) and other mensuration scientists integrated the main ingredients of mathematics into the adjustment calculus, such as:

- Vector, matrix, tensor and array calculus
- Applied mathematics of linear and nonlinear systems
- Linear and nonlinear statistical estimation theory
- Geometry and physics of mensuration
- Numerical analysis
- Signal processing and general fast transforms
- Computer and information sciences

Adjustment calculus relates most closely to least squares estimation and numerical analysis. These will be reviewed next, before their expansion by the general fast transforms of array algebra. The author belongs to the transitional generation that had to solve problems the hard way, before computers eased the workload. Often, the hard way required some innovative analysis and problem formulation to become practically feasible. The resulting innovative solution algorithms, even in today's powerful computers, are often the only feasible way of solving several computational least squares problems - brute-force approaches are often defeated by their own enormity. In this context, some basic concepts of linear estimation are first reviewed using loop inverse estimation and fast linear array algebra. Then, these concepts are expanded into a general nonlinear statistical estimation, numerical analysis and signal processing procedures of array algebra.

## 2. BASIC THEORY OF LEAST SQUARES ESTIMATION

Use of computers simplifies implementation of the linear least squares solution of an overdetermined system of m equations coupling m measured values L of observables and their random residual errors V with n unknown parameters X through a full-rank matrix A in

$$\underset{mn}{A}\,\underset{n1}{X} = \underset{m1}{L} + \underset{m1}{V}. \tag{1a}$$

The traditional least squares solution that we are going to generalize in this paper minimizes the sum of the squares of the residuals resulting in

$$\hat{X} = A^{\ell} L\,, \tag{1b}$$
$$\hat{V}^T\hat{V} = (L - A\,\hat{X})^T(L - A\,\hat{X}) = \min., \tag{1c}$$

where superscript T denotes the transpose of a matrix and the full-rank least squares or l-inverse of A is

$$A^{\ell} = (A^T A)^{-1} A^T. \tag{1d}$$

To simplify the notations and derivations, only the unweighted case is treated or the a priori weights are already included in $A^T = A^T P$ such that $\hat{V}^T P \hat{V} = \min$. The linear least squares (l.s.) solution is used as an (overly crude) approximation to nonlinear estimation by truncating the Taylor expansion of a nonlinear observable after the constant and linear term in $f(X^o + dX) = f(X^o) + f'(X^o)\,dX$. Matrix A makes use of the first partial derivative $f'(X^o)$. This derivative is evaluated for each of a series of m observations using the latest estimates $X^o$ of parameters X. Typically, m exceeds the rank of matrix A or the number of unkowns which constitute the elements of column matrix dX. The linearized system is solved for the small corrections dX, and the process is iterated until it converges. It may also diverge if the initial values of X are beyond the pull-in range of A. This process is known as the Newton-Gauss (NG) solution technique of nonlinear l.s. estimation.

The task of a problem analyst employing l.s. estimation is to define a linear parametric model between the true values of observables and parameters, or to derive the row elements $f'(X^{o})$ as the first partial derivatives of a nonlinear model. The l.s. solution often has to be completed with reliability and accuracy estimates of $\hat{X}$ and adjusted observations $A\ \hat{X}$ using the variance - covariance terms of $\hat{V}^T\hat{V}$, $(A^TA)^{-1}$and $AA^{\ell}$. The special l.s. estimate $\hat{X}$ of (1b) coincides with the Best Linear Unbiased Estimator (BLUE) of estimable X under the Gauss - Markov model E(L) = A X where E denotes the expectation operator. Term "estimable" refers to linear functions that can be unbiasedly estimated.

All elements of X are estimable, $E(\hat{X}) = X$, in this special case when matrices A and $A^TA$ are nonsingular (have the full rank). The general rule for a linear function $A_o$ X (one element at a time) to be estimable as H L under the Gauss - Markov model is $A_o = HA = A_oG\ A$. A general matrix inverse G was generalized in the theory of loop inverses beyond the restriction of the g-inverse of A, defined by the condition A G A = A for any matrix A. The BLUE or 'Best' estimator H L among the estimable $A_oX$ (satisfying $A_o = H\ A$) minimizes its variance $H\ H^T$, (Rauhala, 1976, 1980, 1981).

The general l.s. inverse $G = A_{\ell}^{g}$ of a singular matrix A provides nonunique estimates $\hat{X} = A_{\ell}^{g}\ L$ such that their projections into the estimable domain by $\hat{L} = A\ \hat{X}$ remain unique and are BLUE of L. It satisfies the conditions $(AG)^T = AG$ and A G A = A that are equivalent to the single condition $A^TAG = A^T$, (Rauhala, 1976 p.93). There is typically an infinite number of math models and choices of modeling parameters X that can produce the unique BLUE elements of the estimable space but this does not affect the estimability of parameters X themselves. It does not help to make the l.s. estimate $\hat{X}$ artificially BLUE of X by forcing matrix A to have a full rank, e.g., by a gradual reduction of such parameters among X that are highly correlated. A concrete example illustrates this philosophy.

Point variant errors of interior orientation (due to unflatness of film, refraction, scanning etc) under the assumption of image invariance (ie., the same systematic error repeating itself from one image to the next at a given image cooordinate) would require an empirical finite element model with post-spacings of only a few mm (Brown, 1984). This implies over 1,000 parameters to cover a 9"x9" image area. Their effect on the measured image coordinates are not separable in the traditional bundle adjustment from that of exterior or other interior orientation elements and from that of the object space parameters. Without their explicit prior correction (by better sensor design or directly measuring and correcting the effect of e.g the 3-D micro topography of the film) and leaving them out from the bundle adjustment does not make the remaining parameters of the l.s. solution BLUE of X although matrix A achieves full rank. The unmodeled systematic errors are compensated by invisible deformations of the object space coordinates among parameters X while their projections or the unique (BLUE) estimators of the observables and their residuals remain the same.

The invisibility of the bias of X in the observable (always estimable) space prompted the name of "fool's paradise" by Brown (1984). The differences or shear error of such model deformations of a block become visible at the overlap of the traditional 2-ray stereo models after the bundle adjustment, especially at the DEM and feature points between the few discrete tie points of triangulation. Their removal by some ad hoc techniques from the extracted map data base may actually enlarge the bias. The biases can be reduced in a rigorous bundle adjustment of a multi-ray model (where all DEM and feature points are measured and used in the tie point mode) with a pre-calibration of the point variant and self-calibration of the image invariant errors, (Brown, 1984). Integration of GPS and inertial systems to such a bundle adjustment is today providing a viable control hierarchy for the subsequent softcopy mapping procedures. Loop inverse estimation can include some image variant errors into self-calibration, (R 1972, 1974), ultimately at the 1-4 pe node spacing of GLSM tie points of the GDIMAP system concept, (R 1986, 1988, 1992).

Numerical analysis addresses the computational aspects of l.s. estimation. The general brute-force l.s. solution of n parameters takes the magnitude of n**3 operations (additions and multiplications) and n**2 high speed memory words. Consequently, n must be restricted to the order of only few thousand parameters, even in super computers. Many problems can split, directly or by some careful analysis of the problem and its math model, matrix A, into sparse (banded, null, diagonal etc) and non-sparse partitions. The resulting l.s. solution can be programmed to avoid unnecessary operations on zero elements with significant savings in RAM and computing time. The traditional "fast transforms" of signal processing are seldom applicable to general problems of numerical analysis. Before an introduction of the generalized fast transforms of linear array algebra, loop inverse estimation will provide the foundations of choosing the modeling parameters from the estimable space, such as image vs. object space in photogrammetry. The general fast transforms of array algebra get applicable in this domain by a careful analysis and systems design, indirectly solving or simplifying also the estimation process of the biased parameters X.

## 3. GENERAL LINEAR ESTIMATION OF LOOP INVERSES

Theory of loop inverses changes the physically explainable "Copernicus-Gauss" modeling parameters X of the Gauss-Markov model E(L) = A X into an estimable set of independent linear functions of X, such as any base combination spanning the domain of the observables L = A X. An illustrative example is the estimation problem of image matching.

Image-to-image shifts of conjugate points of a visible object surface and their illumination change are unbiasedly estimable in the image space. This holds true even when the interior and exterior orientation parameters are unknown or are not estimable. The measured image coordinates (a regular grid of integer line and sample locations of the reference image + the matched shifts to get their conjugate image coordinates in the slave images) are nonlinear functions of an infinite number of physically and empirically explainable parameters X to connect them to their object space coordinates. These parameters interconnecting the image and object space coordinates are seldom, if ever, fully estimable depending on how well the sensor is designed, built and operated in the metric sense.

The bias caused by local effects of atmospheric and other geometric disturbances is the limiting factor in estimating the image-to-object and its inverse transform parameters. The opposite is true for image-to-image transform parameters. The differences of image coordinates and point illumination (at visible surface points) are typically estimable from the observed gray values. Points of small correlation or matching signature, not estimable in the single point mode, get estimable under the globally constrained finite element Gauss - Markov model of GLSM or by a similar global stereo fusion process of an operator pointing to the conjugate image details. This example is analogous to so called station adjustment in surveying and closely related to the idea of loop inverses.

A series of theodolite or levelling readings are often measured and reduced in a local field or "station adjustment" to provide more reliable (blunder free) and locally calibrated l.s. estimates of fictitious observations. The reduction may also model and remove some systematic errors that cannot be included into parameters X without making matrix A ill-conditioned or without destroying the special structure of the matrix that allows an efficient solution. Estimation of all parameters X of the complete net is not feasible at this sequential stage of the mensuration process. Even after the completion of all blunder free and locally calibrated estimates of the fictitious observations or linear functions L = A X, actually used in the l.s. adjustment process in place of the raw field measurements, all parameters in X may not be estimable.

Loop inverses expand the technique of fictitious l.s. adjustment into the problem of a general matrix inverse such that the redundancy is removed by the l.s. = BLUE solution of an independent set of observable parameters spanning the always estimable space of observables. In analogy to the technique of l.s. condition adjustment of surveying, the chosen set of adjusted observables is then transformed (by consistent vs. redundant equations) into estimates of biased or unbiased parameters X. Instead of using some intuitive rules of condition adjustment, the loop inverse technique formalized the process by generalizing the theories of matrix inverses and linear estimation. This invention process has continued from the late 1960's to the present findings of the more general nonlinear estimation theory of array algebra.

A starting point (lm-inverse) of loop inverse estimation reverses the consistent linear transform $L_o = A_o X$ by selecting $L_o$ to be a set of p independent observables, often among L = A X themselves such that $A_o$ is a horizontal partition of A. The full-rank minimum norm m-inverse $A_o^m = A_o^T(A_o A_o^T)^{-1}$ exists because the rows of $A_o$ are independent. The parameter transform $X = A_o^m L_o$ substituted into (1a) makes the classical linear l.s. estimation applicable for $L_o$ by replacing the full-rank l-inverse of matrix A with that of $A A_o^m$ in (1d). The (BLUE) l.s. solution $\hat{L}_o$ is transformed or back substituted to estimates (often biased) of parameters X through the consistent transform $\hat{X} = A_o^m \hat{L}_o$. The complete chain of linear matrix operators from the original observables L to estimates of $\hat{X} = A^{lm} L$ is found through the lm-inverse $A^{lm} = A_o^m (A A_o^m)^l$.

The starting lm-operator of loop inverses (similar "loops" of parameter transformations can be applied to more general inverses than the full-rank l- and m-inverses) expanded the basic theories of estimability and general inverses, (R 1974, 1976, 1980, 1981, 1982).

If p equals to the rank of matrix A, the lm-inverse and estimates of X become invariant on the chosen $L_o$ and $A_o$ yielding the unique pseudo-inverse $A^+$ of singular A as a special case. The constraints p = rank(A) = n yield the l-inverse of (1d) as a special case of $A^{lm}$ and $A^+$.

## 4. LINEAR ARRAY ALGEBRA

Array algebra generalizes special fast transforms to become applicable to the computational solution of linear and nonlinear estimation. This work started in search of practical systems and algorithmic solutions to multi-ray analytical and digital photogrammetry. New techniques and algorithms found in this search could not be readily expressed in terms of traditional vector, matrix and tensor concepts. This prompted the introduction of array calculus in (R, 1972) which coupled to the general estimation and matrix operators of loop inverses resulted in linear array algebra (R, 1974, 1976).

The basic idea of array algebra makes fast transforms and matrix calculus applicable in general estimation problems of linear systems. An example is the expansion of matrix equation $AXB^T = L + V$ to 3-D arrays of X,L,V and a third matrix $C^T$ in a regression analysis of three variables. The starting 2-D system of matrix notations prompted this idea in 1970-71 experiments of fitting separable empirical functions of the estimable loop inverse domain to gridded reseau observations of the Hasselblad moon camera used in Apollo missions.

The operations count of forward array multiplications is reduced from MN of the equivalent long hand matrix multiplication to the order of log(N) N where N is the total number of parameters in array X or $N = n_1 n_2 n_3 ..$ and $M = m_1 m_2 m_3 ..$ is the number of elements in L. Unlike the very restricted fast transforms of signal processing, array algebra can handle general rectangular and singular sub matrices A,B,C.. and their general inverses. The operations count of inverse array multiplications in estimating N array parameters X from M observed values of array L is reduced from N**3 of the long hand solution (1b) to the order of log(N) N operations.

The fast array algebra l.s. solution can be mapped into terms of (1a) by rearranging the array elements X into a long column vector in the same fashion as a computer treats an array. The resulting large M,N matrix becomes a Kronecker or tensor product of the small sub matrices A,B,C... The operations count to get the l-inverses (1d) of the small sub matrices A,B,C,.. is often negligible in comparison to their array (general matrix) multiplications with L such as in $A^l L B^{lT}$ of the 2-D l.s. matrix solution of array $\hat{X}$. Similar savings take place in the high speed (RAM) computer storage requirements, especially when A,B,C.. become sparse matrices in loop inverse estimation, making many problem formulations involving millions of parameters practically feasible for the very first time, (R 1976 p.73).

Array algebra often allows the use of inexpensive computers to attack problems of the otherwise overwhelming dimensions. Tailored hardware to implement these fast algorithms (such as finite element and the related inverse cosine transforms) results in Array Algebra Computers (AAC) where solution speed is measured in MHz rates. A single such AAC chip can therefore have the equivalent computing power of tera FLOPS, (R 1977, 1986, 1992). The array algebra software solution speed for GLSM normals, in a general purpose computer of 1 MIPS and 0.25 MFLOPS (VAX 780), is already over 7,000 nodes or posts per sec. This translates into over

100,000 parameters per sec in today's workstations. A practical example illustrates this speed.

For a comparison, some super computer CRAY-XMP results of (Fulton and Fausett, 1991) are shown in the elementary case of separable 2-D array solution where non-sparse matrices were used in the pre- and post-multiplications of a matrix solution discussed in (R et.al., 1989). The RAM user allocation limit of the conventional reference solution was reached in CRAY at 40x40=1600 parameters, taking 69.38 sec CPU time, while the array algebra solution of the same 40x40 parameters required 0.02 sec. That is, array algebra attained a speed of 80,000 parameters per sec. The CRAY RAM allocation limit was exhausted with the non-sparse array algebra solution at 600x600=360,000 parameters, taking 26.48 sec. This is about the same speed per parameter as with the VAX 780 sparse array algebra GLSM solution (2-3 parameters per node) which is invariant on the number of nodes. In other words, the efficient software of sparse array algebra makes VAX 780 and the comparable micro computers to compete with the speed of a super computer solution using the non-sparse array algebra software. The brute-force solution of a practical size of the problem is defeated by its enormity in any computer.

Some recent VAX 780 GLSM algorithms can handle close to 490,000 (700x700) nodes or over one million parameters in one batch. They can enforce seamless equality or mosaicking constraints among the neighboring frames in real-time such that the mosaicked solution is the same as if the whole area of unlimited size had been processed simultaneously. A typical mapping area involves billions of GLSM modeling parameters.

We will discuss GLSM in more detail in the application section of the nonlinear array algebra.

## 5. NONLINEAR ARRAY ALGEBRA

Nonlinear expansion of array algebra unifies the theories of nonlinear estimation, numerical analysis and signal processing. It can efficiently apply general Taylor expansions of functions to nonlinear estimation, (R 1990). New solutions of nonlinear systems are then found which surpass the capabilities of the Newton-Gauss (NG) least squares and linear robust estimation. They also expand Newton-Raphson (NR) and some general techniques of mathematical geodesy, (Pope, 1974), (Meissl, 1979), (Blaha, 1987). The fast numerical analysis and signal processing of linear array algebra is generalized to become applicable to the resulting nonlinear solutions. The basic ideas of the nonlinear array algebra estimation theory are:

- Expansion of Taylor series by array multiplications
- Derivation of nonlinear normals
- Direct linear solution of nonlinear problems
- Linearization of general nonlinear normals
- Expansion of "super" iteration
- Nonlinear estimation beyond least squares
- "Fast" GLSM type application of the theory

### 5.1 Array Algebra Expansion of Taylor Series

The concepts of array algebra serve to apply Taylor series:

$$f(x+dx) = f(x) + f'(x)\,dx + 1/2\,f''(x)\,dx^{**}2 + 1/6\,f'''(x)\,dx^{**}3 \ldots \qquad (2)$$

to the general case of multiple parameters when these are expressed as column vectors X and dX. Use of these general operators will show a shortcoming of the NG technique, explaining its narrow pull-in range and slow convergence. This shortcoming is corrected in NR but is usually ignored because of the resulting computational burden and increased analytic work of deriving the higher order partial derivatives f″ and f‴. Array algebra, however, can often eliminate the computational problem justifying the increased analysis work.

Matrix algebra of the traditional NG iterations of (1b) can handle the constant and linear terms of Taylor series when the scalar dx is expanded to a column matrix dX of n parameters. The constant term $f(X^o)$ is the "predicted" or evaluated value of the nonlinear function and $f'(X^o)$ is a row matrix of its first partial derivatives with respect to each parameter or element of dX. Contraction or inner product of the row by column matrix multiplication f′(X ) dX in (2) produces an element as the correction $df(X^o)$ to an observable due to the unknown parameters dX. But what about the product $f''(X^o)\,dX^{**}2$ in (2) and its derivative $d/dX = df'(X^o) = 2f''(X^o)\,dX$ that will cause the missing linear component of NG to be shown in the more rigorous NR solution of (11)?

Operator $f''(X^o)$ is a 3-D (1,n,n) array of one row, n columns and n "depth slices" of the second partial derivatives of nonlinear function f using the initial values $X^o$. It grows into an (m,n,n) array F″ of m observables. Array algebra expands matrix multiplication and other rules beyond the traditional limitation of 2-D matrices. This facilitates estimation theory and numerical analysis better than the indical notations of tensor calculus. The product df(i) = f″(i) dX**2 for an observable i consists of the double summation or contraction

$$\underset{m1}{dF} = df(i) = \underset{mnn}{F''}\,\underset{n1}{dX}{}^{**}2 = \sum_{j}\sum_{k} f''(i,j,k)\,dx(j)\,dx(k) \qquad (3)$$

where i,j,k are the indices of array F″ and vector dX resulting in vector dF of m elements. There are several short hand notations to express the double summation in (3). It can be expressed in traditional matrix notations as

$$df(i) = \underset{1n}{dX^T}\,\underset{nn}{F''(i)}\,\underset{n1}{dX}, \quad i=1,2,..\, m, \qquad (3a)$$

by transposing the m horizontal slices of 3-D array F″, or exchanging indices i and k, into (n,n) matrices F″(i), i=1,2..m. Each matrix is then multiplied by the row matrix $dX^T$ and post-multiplied by column matrix dX.

The expanded rule of array multiplication in (3) for contraction of the second and third index (j and k) of F″ with the post and "depth" vectors dX has two alternative notations. There is no need of transposing the 2-D matrices or sub-arrays of F″ into a position where the old matrix rules apply. The first alternative emphasizes the analogy of array multiplication to expanded matrix multiplications and polynomials by

$$\underset{m1}{dF} = \overset{\underset{n1}{dX}}{\underset{mnn}{F''}}\,\underset{n1}{dX} = F''\,dX^{**}2\,. \qquad (3b)$$

Post matrix multiplication of 3-D array F″ by dX in (3b) contracts the second index of array F″ (thereby allowing array F″ to have any number of indices) with the first index j of matrix dX written after it. Following this analogy, a "depth" matrix written above the array contracts the third array index with the first matrix index k. A matrix written in front (unit matrix in (3b)) of an array contracts the first index i of the array with the second matrix index in analogy to the rule of matrix pre-multiplication.

In the expanded polynomial notation of array multiplication in F″ dX**2 of (3) and (3b), the two (post and depth) matrices are the same vector dX. They contract the last two indices of array F″ in two repeated (exponential) array multiplications. Notice that the first power has to be noted in F dX**1 to indicate that contraction is to be done for the last index of any array F and not the second index in F dX of the expanded rule of post matrix multiplication.

The second more general notation alternative (not detailed here) of array multiplication can handle arrays of more than three indices as it is difficult to continue the analogy for pre, post and "depth" matrix multiplications for the fourth and higher order indices. This general notation convention need not transpose the second and third matrix before the post and depth multiplications. In this review paper, only a summary of nonlinear array algebra can be given. Details of the array algebra notations and their comparisons to the indical notations of tensor calculus are found in (R 1974, 1976, 1980, 1990).

In general nonlinear estimation, each element of row vector $f'(X^o)$ is a nonlinear function of all unknown elements of dX. The analysis work to derive $f''(X^o)$ is thereby multiplied by a factor of n in comparison to NG that uses only $f'(X^o)$ ignoring all $f''(X^o)$. It may appear that we need not care about the nonlinear term $f''(X^o)$ dX**2 of the Taylor expansion as we have to linearize the problem anyway. The partial derivative d/dX of this term is linear. It equals to 2 $f''(X^o)$ dX**1 (the last or third array index is contracted) = 2 $f''(X^o)$ dX (the second index is contracted but because of the symmetry of f″(i,j,k) among indices j,k the result is the same matrix after an array transpose). Its contribution on the linear term of the nonlinear normals is neglected in NG but included in NR as will be detailed in equations (9)-(12).

The use of the 4-D array f‴ is not as impractical as it may appear (increasing the work of an analyst by a factor of n**2 over NG). In most problems of global modeling philosophy, the terms of f‴ at a given nonlinear finite element reduce to constants or null elements. In other words, they are invariant on X and dX. Similarly, elements of $F''=f''(X^o)$ of all observables need to be evaluated only once with very crude initial values. They can be updated (seldom needed in practice or only after the first iteration) by 3-D array corrections dF″ = F‴ dX**1 as a contraction among the accumulated corrections of dX and the prederived scalars of F‴, (Blaha, 1987), (R 1990).

### 5.2 Nonlinear Normals and Their Taylor Expansion

We first derive the nonlinear normals for the elementary scalar case of one single unknown dx using the nonlinear observation equations

f(x + dx) = g(x) + v(x). (4)

In an example of nonlinear single element l.s. matching, x is the integer image coordinate of the observed slave gray values g(x) and unknown residuals v(x). The unknown shift dx of the nonlinear reference function f(x) is modeled by one constant within a small window. Since f(x) and g(x) are taken to be known, as discrete gray-value grids approximating continuous functions, they could be matched with an infinite number of l.s. observations using integral calculus vs. the discrete samples. The derivation and integration of such closed nonlinear normals will be discussed in section 5.3 as a limiting case of the regular discrete l.s. observations derived next.

The function of (4) to be minimized in least squares estimation is

sum(v(x)**2) = sum( ( f(x+dx) - g(x) )**2 ) (5)

where the summation is made over the index of observables or variable x in our example. This nonlinear function of dx reaches its minimum when its first derivative d/dx vanishes. The consistent system of nonlinear normal equations

n(x,dx) = sum( f′(x+dx)( f(x+dx) - g(x) ) ) = 0 (6)

is found by applying the elementary rule d/dx(f(x)**2) = 2 f′(x) f(x) to (5) and scaling both sides of equation (6) by the constant factor of 2.

The Taylor series of (6) is the product of the derivative of (2)

f′(x+dx) = f′(x) + f″(x) dx + 1/2 f‴(x) dx**2 (7)

and f(x+dx) of (2) minus the observed g(x). This product results, when f(x+dx) of (4) is truncated after term f‴ in (2), in a fifth degree polynomial approximation n(x,dx) to the closed normals of (6) by

n(x,dx)= 1/12 sum(f‴f‴) dx**5
+ sum(5/12 f″f‴) dx**4
+ sum(2/3 f′f‴ + 1/2 f″f″) dx**3
+ sum(1/2 f‴ l + 3/2 f′f″) dx**2
+ sum(f′f′ + f″l) dx + sum(f′l) = 0 (6b)

where

l(x) = f(x) - g(x).

This scalar case of a single element in dx is analogous to general array polynomials of parameter vector dX where f′ is a matrix, f″ is a 3-D array and f‴ is a 4-D array, (R 1990). The summation over products, such as sum(5/12 f″f‴) associated with dX**4 contracts the (first) index of observables from the 3-D and 4-D arrays resulting in an output array of 3+4-2=5 indices. Its array polynomial or quadruple contraction with dX**4 leaves out an array of only one index or a column vector of the same size as the constant terms $F'^T L$.

### 5.3 Direct Linear Solutions of Nonlinear Problems

A problem analyst should try to mentally derive the analytic expressions of the closed nonlinear normal (6), especially in such a simple problem as the present example. The reader can do this by specifying a nonlinear f(x) in (4), such as a quadratic or cubic polynomial of known coefficients, or by switching over to your favorite nonlinear problem. The author made some preliminary explorations of nonlinear analytic normals of the bundle strip adjustment and range observations uncovering techniques similar to those of Direct Linear Transforms (DLT) of photogrammetry or the direct linear GPS reduction technique of Meissl (1979). The direct analytic derivation of normals of nonlinear least squares or robust estimation offers challenges to today's computer generation of problem analysts. It may take some time before these analytic derivations can be made in computers. The mental derivations are complicated by the fact that the number and distribution of the (fictitious grid) observations affect the solution and that all arithmetic expressions of n nonlinear equations of n parameters have to be explicitly derived.

Perhaps the most elegant solutions in photogrammetry are still to come from the persistent analysts who endure the search of the direct analytic solutions of nonlinear least squares and robust estimation. There are some signs in the recent literature of photogrammetry and computer vision toward this type of work in nonlinear problems (perhaps reinventing a "lost wheel" of the founding days of photogrammetry). GLSM can handle every single pixel of a multi-ray stereo model as a tie point, allowing some simplifying assumptions to the nonlinear bundle adjustment such as regular and complete grid observations of quite uniform weights. These simplifications cancel many nonlinear and linear terms in the analytically derived normals.

The nonlinear problems of GLSM, bundle and range (SAR or range image) adjustments are not very far from the idea of direct linear solutions impacting their pull-in range and convergence rate. The main computational work is already done by the ingenious and hard analysis of mental derivations of the normals and their analytic solution (no rounding errors) or other reduction before the computer is employed in the final stages evaluating the numerical values. The author has routinely employed this philosophy of special "inner loop" solutions of Array Relaxation such as in inertial net or continental datum adjustment of crossing traverses and Global least squares reseau reductions, (R 1982, 1986, 1992), (Craig, 1992).

The reader can independently confirm the existence of a direct linear solution in the simple matching example of (4). A quadratic polynomial f(x) results in a closed nonlinear normal equation in dx as a cubic polynomial. An added linear illumination difference parameter in (4) results in two nonlinear equations. After its elimination from one of them, the resulting reduced single equation in dx becomes linear reading

$$f''\, dx = g' - f' \tag{8}$$

where $g'-f'$ is the difference of the central first derivatives of f(x), g(x) and $f''$ is the constant $f''(x)$.

### 5.4 Linearization of Nonlinear Normals

The iterative Newton solution of the nonlinear normal equations can be derived in several ways. The direct way resulting in the NR technique assumes an initial value of $dx^\circ$, and computes the discrepancy or residual vector $n(x,dx^\circ)$ of normals (6). An unknown correction ddx of each iteration is found by solving a linear system involving the partial derivative matrix $n'(x,dx^\circ)$ in

$$n'(x,dx^\circ)\, ddx = -\, n(x,dx^\circ) \tag{9}$$

where

$$\begin{aligned}
n'(x,dx^\circ) &= d/dx\ \mathrm{sum}(\ f'(x+dx^\circ)(\ f(x+dx^\circ) - g(x)\ )\\
&= \mathrm{sum}(\ f'(x+dx^\circ)^{**}2\ +\ f''(x+dx^\circ)(\ f(x+dx^\circ) - g(x))\ )\\
&= \mathrm{sum}(f'f' + f''l),\\
n(x,dx^\circ) &= \mathrm{sum}(\ f'(x+dx^\circ)(\ f(x+dx^\circ) - g(x))\ )\\
&= \mathrm{sum}(f'l),\\
l(x,dx^\circ) &= f(x+dx^\circ) - g(x).
\end{aligned}$$

The summations are done over the index of observables, the first index of l, $f'$ and $f''$. The same solution is found by the Taylor expansion (6b) containing only the constant and linear terms. Under the implicit assumption that $dx^\circ = 0$ or evaluation of l, $f'$, $f''$ with the latest parameter estimate, the first two Taylor terms of nonlinear normals in (6b) are

$$\mathrm{sum}(f'f' + f''l)\, ddx = -\mathrm{sum}(f'l). \tag{10}$$

The scalar normal equations of (9), (10) are analogous to the NR matrix equation of multi element ddX

$$(\underset{nn}{F'^{T}F'} + \underset{nn}{(L^{T}F'')_{T}})\ \underset{n1}{ddX} = -\ \underset{n1}{F'^{T}L}\ . \tag{11}$$

Observables $l(x+dx^\circ)$ have one index of column matrix L. First partial derivatives $f'(x+dx^\circ)$ have two indices resulting in matrix $F'$ such that sum $(f'f') = F'^{T}F'$ and $\mathrm{sum}(f'l) = F'^{T}L$ in matrix notations of NG. The second derivatives $f''(x+dx^\circ)$ have three indices. Its first index is contracted with L in $\mathrm{sum}(f''l)$. According to the expanded matrix notation of array algebra under equation (3b), this is achieved by $L^{T}F''$ resulting in an (1,n,n) array. It is then transposed in (11), by subscript T to exchange the first and third indices vs. the first and second index of superscript T, to a symmetric (n,n,1) matrix before addition to $F'^{T}F'$ of the NG matrix.

In theory, the resulting symmetrix NR matrix $N'(X,dX^\circ)$ in (11) is non-negative definite in "clean and smooth" problems but the empirical and noisy data of image matching often yields negative definite local normals in GLSM, perhaps because the "bad" areas (occlusions, etc.) do not obey the Gauss-Markov model. A derivation of the Taylor array series of matrices $N'(X^\circ,dX)$ as partial derivatives of (6b) reveals the 'decentered' NR normals of (R 1990)

$$\begin{aligned}
N'(X^\circ,dX) &= F'^{T}F' + (L^{T}(F'' + F'''dX\ ^{**}1))_{T} +\\
&\quad F'^{T}(3\, F''dX\ ^{**}1\ +\ 2\, F'''dX\ ^{**}2)\ \ldots
\end{aligned} \tag{11b}$$

All arrays $F'$, $F''$ and $F'''$ are evaluated only once with very crude initial values $X^\circ$. Updating of $F'^{T}F'$ is approximated by $F'^{T}(3dF_1' + 2dF_2')$ where matrix corrections $dF_1'$, $dF_2'$ are found by the array polynomials $F''dX^{**}1$, $F'''\, dX^{**}2$ using the accumulated dX estimates. Updating of $F''$ is done by $F'''\, dX\ ^{**}1$.

### 5.5 Array Algebra Expansion of Super Iteration

The basic 1976-77 idea of nonlinear l.s. matching in (4) introduced the local "scale space constraint" by using two closest integer locations to the latest $dx^\circ$ estimate as the initial values of two analytic NG adjustments. The combined solution from both initial values is required to result in the same (fractional) solution of dx. The application of this technique to NR in (11) combines several iterations into one "super iteration" of a much more robust solution than the use of a single initial value. Computing time is not much increased from a single iteration of a single initial value by the application of the analytic techniques in section 5.3.

NR technique of a single initial value is known to have a superior convergence region and rate over NG. These NR convergence properties are enhanced in the new Multiple Initial Value Constrained (MIVC) NR technique of array algebra because, like in GLSM, the contribution of the poor region of the chosen initial value "search range" can be overruled by proper weighting. As the search range is refined close to zero or the standard error of the adjusted parameters, the MIVC technique approaches the limiting case of NR in the final iterations. NR, in turn, approaches that of NG as elements of L get small. The more sophisticated and therefore slower nonlinear least squares techniques are only required in the initial value refinement or "pull-in" process of coarse-to-fine grid method.

Blaha (1987) showed a simulated example where the expanded NR of single initial values converged in 1-2 iterations with orders of magnitude wider pull-in range than the limit of NG convergence region of over 10 iterations. Our preliminary experiments since 1988-89 with nonlinear GLSM and MIVC techniques confirm his findings, although image matching inherently contradicts theories requiring smooth and continuous math models. The MIVC technique has a realistic potential of converging in one iteration to the nearest solution of nonlinear local l.s. minimum. It bridges the gap to the nonlinear problems that can be solved with an unlimited pull-in range of direct linear solutions.

A new finding of "nonlinear loop inverse and array algebra solution to neural nets" expands matrices and their general inverses of linear algebra into nonlinear operators, (R 1990). It continues the expansion of adjustment calculus in the fashion of linear loop inverses and array algebra.

#### 5.6 Nonlinear Estimation Beyond Least Squares

All of the above derivations, starting from the nonlinear observation equations (4), can be generalized by choosing more general minimization functions in (5) in place of the least squares or the second power of the absolute values of residuals. Laplace's estimation uses the first power @=1 and the related robust estimation (often with more general functions of residuals than a polynomial of their absolute values) employs 0<@<2. The resulting estimates $\hat{V}$ = L - A $\hat{X}$ are more robust than their BLUE of least squares in revealing the outliers, (Krarup et.al., 1980). The integer powers @ of the exponent are of a special interest. The nonlinear normals and their Taylor series show that, structurally, @=1 is the "central power" among the complete class of this unified estimation theory of minimized residuals.

The linearized normals or the basic system of equations of nonlinear estimation by minimum residuals with arbitrary @, vs. (11) of @=2, are

$$((@-1)F'^{T}PF' + (L^{T}PF'')_{T})\, ddX = -F'^{T}PL. \tag{12}$$

The reader is urged to work out the details of this derivation in analogy to equations (2)-(11). The diagonal "power weight matrix" P has elements

$$p(i,i) = (abs(l(i)) + e)^{**}(@-2)$$

where a suitable small positive constant e prevents the infinite weights of l(i)=0 for @<2.

A major expansion to the NG and linear least squares estimation (1b) is achieved in (11) and (12). The central power @=1 reduces into

$$(sign(L)^{T}F'')_{T}\, ddX = -F'^{T}sign(L). \tag{13}$$

All cases of arbitrary @ are handled by the same general least squares routine by regulating the power weight matrix P (as a part of the traditional weighting process) and the "power relaxation" term 1/(@-1). This term equals 2.0 for @=1.5 and -2.0 for its symmetric @=0.5 in the final iterations of small or no contributions from $L^{T}PF''$. The symmetric case @=0 to least squares has the reasonable "power weights" (inverse of l(i)**2). In the final iterations of small L (or small F"), it walks an equal amount with the least squares solution - but in the opposite direction! (@-1=-1 for @=0 vs. +1 for @=2).

All terms in (12) are scaled by @ such that minimizing v**0=1 of @=0 makes more sense and we should keep in mind the philosophy of loop inverse estimation. Parameters X themselves are seldom fully unbiasedly estimable and we are mainly interested in their projections and other functions A X in the observable and estimable space, including a robust rejection and automated fill-in of the outliers. Only then the second stage of estimation is started for parameters X by minimizing the bias, variance, norm or some other problem dependent estimation criterion. This widens the freedom of the problem analyst beyond the traditional one-step process of least squares estimation. While we take the new steps on this new territory, we may find the classical special case of least squares as a safe guide toward new explorations.

#### 5.7 Application in Nonlinear GLSM

The 1976-77 invention of Global Least Squares Matching (GLSM) and the associated automated technique of on-line DEM validation has prompted many inventions of array algebra and advanced system designs of multi-ray softcopy workstations. Their fast array algebra solutions detailed in (R et.al., 1989) can handle 1kx1k 2-ray stereo frames of GLSM with 1-4 pixel node spacing of over one million modeling parameters in VAX 780 and SUN work stations, in a few minutes of CPU time. The resulting 0.05-0.2 pe average standard error of x-shifts and high DEM resolution of dense posts is found (by running same test cases using both techniques) to be in the order of 4-16 times better than that of cross correlation over a rugged terrain. The automation reliability is not far from 100%, even on urban, forested and steeply mountainous areas of the poor 2-ray models, confirming the 1986-88 GLSM results of the international (ISPRS) image matching test, (R 1988), (R et.al., 1992). GLSM can handle cases, including those of close range photogrammetry, at about the same or better reliability for all (infinite number) points as the point wise (vs. global l.s.) cross correlation over few good candinate points selected by an interest operator.

Multi-temporal, SAR-to-SAR and other image registration applications where the old techniques have failed, have been reliably experimented with GLSM while taking in the order of 100-1,000 times less arithmetic operations per point than a cross correlator. GLSM sample speed covering tiny 3x3 to 5x5 pe windows in VAX 780 is over a thousand posts per sec in one iteration including the reshaping computations. This sample speed is exceeded by the earlier discussed global DEM solution of array algebra for automated edit and fill-in of the x-shifts. The GLSM speed advantage per point over the single element special case of digital LSM point transfers of about 32x32 pe (vs. 1k x 1k pe) windows is larger than over the cross correlation. A general purpose computer of 1 MIPS with the fast GLSM software has beaten many published image matching throughput rates of super computers, array processors and transputer nets by a factor of 10-100. The demonstrated 95-100 % reliability of GLSM vs. a more typical range of 60-90% of the old automated

DEM techniques removes the bottleneck of manual edit work improving the total production throughput from the sustained manual edit rate of about a second per point to 100-1,000 points per sec in the weak 2-ray mode.

The global simultaneous DEM solution of x-shifts is only one part of the overall GLSM process of each iteration. These iterations of nonlinear estimation have to be repeated several times in NG at each minification level of the hierarchical pull-in. The 2-ray throughput speed of NG varies in 200-600 nodes/sec in VAX 780 class of computers depending on the sophistication and operations mode of the algorithm to optimize the quality. The nonlinear expansion of GLSM is straight forward increasing only slightly the global NG timeline of one iteration. Because of its wider pull-in range and faster converge rate it is expected to improve the overall speed and quality (economy) after it is fully developed and tested. The processes of hierarchical multi-resolution, multiple initial values, multi-grid, multi-layer, multi-ray, multi-spectral and continuity relaxation are gradually integrated to the (inherently multi-temporal) GLSM sofware in parallel with the work on the GD/HAI sofcopy workstations, (Miller et.al., 1992).

## 6. SUMMARY

A general theory of linear estimation by loop inverse technique and fast numerical analysis of linear array algebra was reviewed with some applications in digital photogrammetry. The resulting theory was then expanded to general nonlinear estimation. The closed nonlinear normals opened a way for direct linear solution of nonlinear estimation including continuous functions of observables vs. discrete samples. Newton solution of the nonlinear normals revealed the NR technique of linearized normals where its partial derivative matrix had a contribution from the second order derivatives of the nonlinear function overlooked by the classical NG technique. NR was expanded by a general Taylor series expansion of array algebra resulting in decentered NR where the normal equations are updated without an explicit re-evaluation of the first and second partial derivatives. The generalized NR was expanded into a super iteration combining several initial values and iterations. Finally, the new nonlinear estimation theory of array algebra was generalized beyond least squares using any power of the absolute values of residuals in the sum of minimized residual functions, yielding nonlinear least squares, Laplace and robust estimation as special cases. An application in global image matching of automated softcopy mapping systems was used as an example of the generalized fast transform technology and numerical analysis of nonlinear array algebra.

ACKNOWLEDGMENT

Reviews of this paper by my colleagues H. Henning, V. Kluth, G. Kunkel and A. Whiteside are appreciated together with the internal R&D support of GD/HAI team that made the presented work possible.

## REFERENCES

Blaha, G., 1987. Nonlinear Parametric Least Squares Adjustment. Scientific Report No. 1, Air Force Geophysics Laboratory, Hanscom AFB, Mass. 01731.

Brown, D.C., 1984. A Large Format, Microprocessor Controlled Film Camera Optimized for Industrial Photogrammetry. Paper presented at ISPRS Congress, Comm. V, Rio de Janeiro.

Craig, R.D., 1992. Smoothing of Automatically-Measured Internal Reseau Nets. ASPRS, Washington D.C., August.

Fulton, C.T. and D.W. Fausett, 1990. Solving Large Least Square Problems Using Array Algebra. Dept. of Appl. Math, Florida Inst. of Technology, Fla 32901. SIAM Meeting, San Francisco, December.

Krarup, T., K.Kubic and J. Juhl, 1980. Goetterdammerung over Least Squares Adjustment. ISPRS Arch. Vol. 23 B3, Hamburg.

Meissl, P., 1979. Direct Solution of Overdetermined Algebraic Problems with Examples from Geometric Geodesy. Manuscript Geod.,Vol 4.

Miller, S., U.V. Helava and K. DeVenecia, 1992. Softcopy Photogrammetric Workstations. PERS No 1.

Pope, A., 1974. Modern Trends in Adjustment Calculus. Paper presented to Int. Symposium on NA Geodetic Networks, New Brunswick.

Rauhala, U.A., 1972. Calculus of Matrix Arrays and General Polynomial and Harmonic Interpolation by Least Squares with New Solutions in Photogrammetry and Geodesy. Fot. Medd. VI:4, Stockkolm.

---, 1974. Array Algebra with Applications in Photogrammetry and Geodesy. Fot. Medd. VI:6

---, 1976. A Review of Array Algebra. Comm. III, ISP Congress, Helsinki and Fot. Medd. 2:38

---, 1977. Array Algebra as General Base of Fast Transforms. ISP Image Processing Symposium, Graz.

---, 1980. Introduction to Array Algebra. PERS No2.

---, 1981. Note on General Linear Estimators and Matrix Inverses. Manuscript Geodaetica, Vol 6.

---, 1982. Array Algebra Estimation in Signal Processing. PERS No3.

---, 1986. Compiler Positioning of Array Algebra Technology. Int. Arch. of ISPRS, Vol. 26-3/3, Comm. III Symposium, Rovaniemi.

---, 1988. Compiler Positioning System: An Array Algebra Formulation of Digital Photogrammetry. Int. Arch. of ISPRS Vol 27 B3, Kyoto, and PERS 1989 No 3.

---, 1990. General Theory of Array Algebra in Nonlinear Least Squares and Robust Estimation. ASPRS Spring Convention, Denver.

---, 1992. Automation of Global Digital Image Mapping by Array Algebra. ASPRS, Washington D.C., August.

---, D. Davis and K. Baker, 1989. Automated DTM Validation and Progressive Sampling Algorithm of Finite Element Array Relaxation. PERS No 4. Presented at the 1988 ASPRS Spring Convention, St. Louis.

---, G. Kunkel, V. Kluth and S. Wong, 1992. Global Least Squares Matching Automates the Image Registration Process. ASPRS, Washington D.C., August.