The Use of a Vectorprocessor in Geodetic Applications

Karsten Engsager

Geodetic Institute
Copenhagen
1982

Abstract.

On basis of "A Priory Prediction of Roundoff Error Accumulation in the Solution of a Super-Large Geodetic Normal Equation System" by Peter Meissl it is discussed whether it is feasible to make the NAD-adjustment on the computer of Geodetic Institute, RC 8000, and a parallel discussion is made on the Danish Network.

It is further discussed the number of digits there are needed in such network computations.

At last it is shown how the use of Wilkinson's ideas concerning single/double precision will change the safe estimate of the global round off errors.

An example of the implementation is given using a dedicated vectorprocessor developed by Geodetic Institute in cooperation with the manufacturer RC Computer.

I. Introduction.

Using the same notation as in Meissl (1980) some definitions should be repeated.

The normal equations A X = B resulting from least squares adjustment are solved using Cholesky's method. Formation and solution of the system are organized according to the Helmert blocking scheme.

The elementary round off error is defined as

$$\epsilon = a \odot b - a \vartriangle b \tag{1}$$

where $\vartriangle$ is a mathematical correct operation on a and b while $\odot$ is the computer operation on a and b.

By true rounding is the expectation

$$E\{\epsilon\} = 0 .$$

A β base machine with τ digits gives the standard deviation (root mean square error)

$$\sigma\{\epsilon\} = \frac{c}{\sqrt{12}} \beta^{-\tau} \tag{2}$$

where $c = \beta^{\nu}$ and

$$\nu = \mathrm{MIN}\{\nu \mid \beta^{\nu} \geqslant |a \vartriangle b|\}.$$

The round off errors perturb the normal equations

$$(A + \epsilon)(x + \xi) = b + \eta . \tag{3}$$

giving an offset of the solution x by

$$\xi = -A^{-1}\epsilon x + A^{-1}\eta . \tag{4}$$

Let μ_{ij} be the number of nonzero products in the triangular decomposition

$$a_{ij} - \sum_{k=1}^{i-1} r_{ki} \; r_{kj}$$

then the number Π of nonzero elements ($a_{ij} \neq 0$ or places (i,j) of "fill in") is

$$\Pi = \sum_{i=1}^{n} \mu_{ii} + n \qquad (5)$$

where n is the dimension of the matrix A.
The number of nonzero products to be evaluated and added during the triangular decomposition phase is

$$\Gamma = \sum_{1 \leqslant i \leqslant j \leqslant n}^{n} \mu_{ij}. \qquad (6)$$

II. Round off estimates for the RC 8000.

The RC 8000 is a true rounding two complement machine with β = 2 and τ = 36.

On the RC 8000 computer are estimates of the global round off errors during the triangular decomposition as given in Meissl (1980) Chapter 4.1.4

	$E\{\xi_i\}$	$\sigma\{\xi_i\}$
IBM 360	167 m	0.0028 m
CDC 6600	0 m	0.025 m
RC 8000	0 m	102 m

and during backsolution

	$E\{\xi_i\}$	$\sigma\{\xi_i\}$
IBM 360	$1.9\ 10^{-8}$ m	$6.7\ 10^{-13}$ m
CDC 6600	0 m	$1.7\ 10^{-10}$ m
RC 8000	0 m	$7.7\ 10^{-5}$ m

On RC 8000 is the refined estimate for a 2^{o} by 2^{o} quad given in Meissl (1980) Chapter 8

$$\sigma\{\xi_p\} = 0.45 \text{ m}$$

It is seen that the mantissa of the RC 8000 is too short (36 bit) to make the NAD network adjustment be feasible.

III. Estimate of the global round off errors in the Danish Network.

The Danish Network has about 40.000 stations covering the whole country. It is a homogeneous area network having no measurements of extremely high precision. It has at present 3 Dopplerstations and 9 Laplace azimuths. There are about 25.000 distances and about 330.000 directions with about 90.000 orientation unknowns (which are eliminated by Schreiber's method).

In the following we consider true rounding machines only, consequently $E\{\xi_i\} = 0$.

From formula (4.48)

$$\sigma\{\xi_i\} \leqslant \left[\frac{4c^2}{12}\|f\|^2\|x\|^2\Gamma + \frac{d^2}{12}\|f\|^2 2\Pi\right]^{\frac{1}{2}}\beta^{-\tau} \qquad (7)$$

where $c = \beta^\nu$ bounds the quantities $c_{ijk}^{(\cdot)}$, $c_{ijk}^{(+)}$, $c_{ij}^{(-)}$, $c_{ij}^{(d)}$ involved during the triangular decomposition of the elements a_{ij} to $a_{ij}^{(i-1)}$ (see (4.10) to (4.13) and (4.17)),

$d = \beta^\nu$ bounds the right hand quantities $d_{ik}^{(\cdot)}$, $d_{ik}^{(+)}$, $d_i^{(-)}$, $d_i^{(d)}$ involved during the triangular decomposition of the right hand side (see (4.23) and (4.25)),

$\|f\|$ bounds the elements f_{ij} of A^{-1}

and $\|x\| = \underset{i}{MAX}\{|x_i|\}$.

Considering the d bound we see that d might be chosen as

$$d \geqslant \underset{\nu}{MIN}\,\beta^\nu \geqslant 2\,\underset{i}{MAX}\{a_{ii}\}\,\|x\|. \qquad (8)$$

Putting $e = 2\,\underset{i}{MAX}\{a_{ii}\}$ the formula (4.48) may be written

$$\frac{\sigma\{\xi_i\}}{\|x\|} < \left[\frac{c^2}{3}\Gamma + \frac{e^2}{6}\Pi\right]^{\frac{1}{2}}\|f\|\,\beta^{-\tau} \qquad (9)$$

which shows how many digits of x are determined by the adjustment. From this it follows that a safe estimate of the length t of the mantissa required to determine x by D decimal digits is

$$\tau = \frac{\frac{1}{2}\log[\,(c^2\Gamma/3 + e^2\Pi/6)\,\|f\|^2\,] - D}{\log \beta} \tag{10}$$

The ordering and dissection algorithm described by Mark (1981) will give

$$\Pi = n\,w = 80.000 \cdot 500 = 40\;10^6$$

and

$$\Gamma = n\,w^2 = 80.000 \cdot 550^2 = 24\;10^9$$

It is assumed that a_{ii} is of the order of $10^2 - 10^4$ then $c = 5\;10^4$ and $e = 4\;10^4$. Expecting $\|f\| = 0.25$ which corresponds to a rms accuracy of 0.5 m, we get

$$\frac{\sigma\{\xi_i\}}{\|x\|} \leq 0.016$$

of the Danish Network on RC 8000. The convergence of the nonlinear problem will be weak and too many iterations of the adjustment are claimed to reach to a stopping condition

"computational noise less than one thousandth of the observational noise".

IV. Wilkinsons recommendations concerning the reduction of normal equations.

Wilkinson has considered several reduction and backsolution algorithm used on symmetric positive definite matrices and shows that the Cholesky-reduction is the most efficient method counting the number of operations. No futhermore pivoting is needed by the Cholesky reduction.

Wilkinson (1963) recommends that the observation equations are formed in standard mantissa length τ, the normal equations are formed by adding products of observation equations of double length 2τ. The final step of the reduction (division by the diagonal element or squareroot formation) will reduce the mantissa length to the standard length τ and during the triangular decomposition the product sums of the reduced elements are formed in double length 2τ (see figure 1). The backsolution can without difficulties be made in the standard mantissa length τ.

The method proposed by Wilkinson will reduce the effect of the round off errors.

Assuming a rounding operation after each arithmetic operation we have (d means double length operations)

for addition and subtraction

$$\sigma\{\varepsilon_d\} = \frac{c}{\sqrt{12}} \beta^{-2\tau} \qquad (11a)$$

$$\sigma\{\varepsilon\} = \frac{c}{\sqrt{12}} \beta^{-\tau} \qquad (11b)$$

where $c = \beta^{\gamma} \geqslant \mathrm{MAX}\{ |a|, |b|, |a+b|, |a-b|\}$,

for multiplication to double length from two single length operands

$$\sigma\{\varepsilon_d\} = 0 , \qquad (11c)$$

for division

$$\sigma\{\varepsilon\} = \frac{c}{\sqrt{12}} \beta^{-\tau} \qquad (11d)$$

where $c = \beta^{\gamma} \geqslant a / b$

and for squareroot operation

$$\sigma\{\varepsilon\} = \frac{c}{\sqrt{12}} \beta^{-\tau} \qquad (11e)$$

where $c = \beta^{\gamma} \geqslant \mathrm{sqrt}(a)$.

Using

$$\varepsilon_{ij}^{(p)} = \sum_{k=1}^{i-1} (-\varepsilon_{ijk}^{(\cdot)} - \varepsilon_{ijk}^{(+)}) + \varepsilon_{ij}^{(-)} \qquad (12)$$

we get

$$\sigma\{\varepsilon_{ij}^{(p)}\} \leqslant \frac{c_{ij}^{(p)}}{\sqrt{12}} \sqrt{\mu_{ij}} \beta^{-2\tau} \qquad (13a)$$

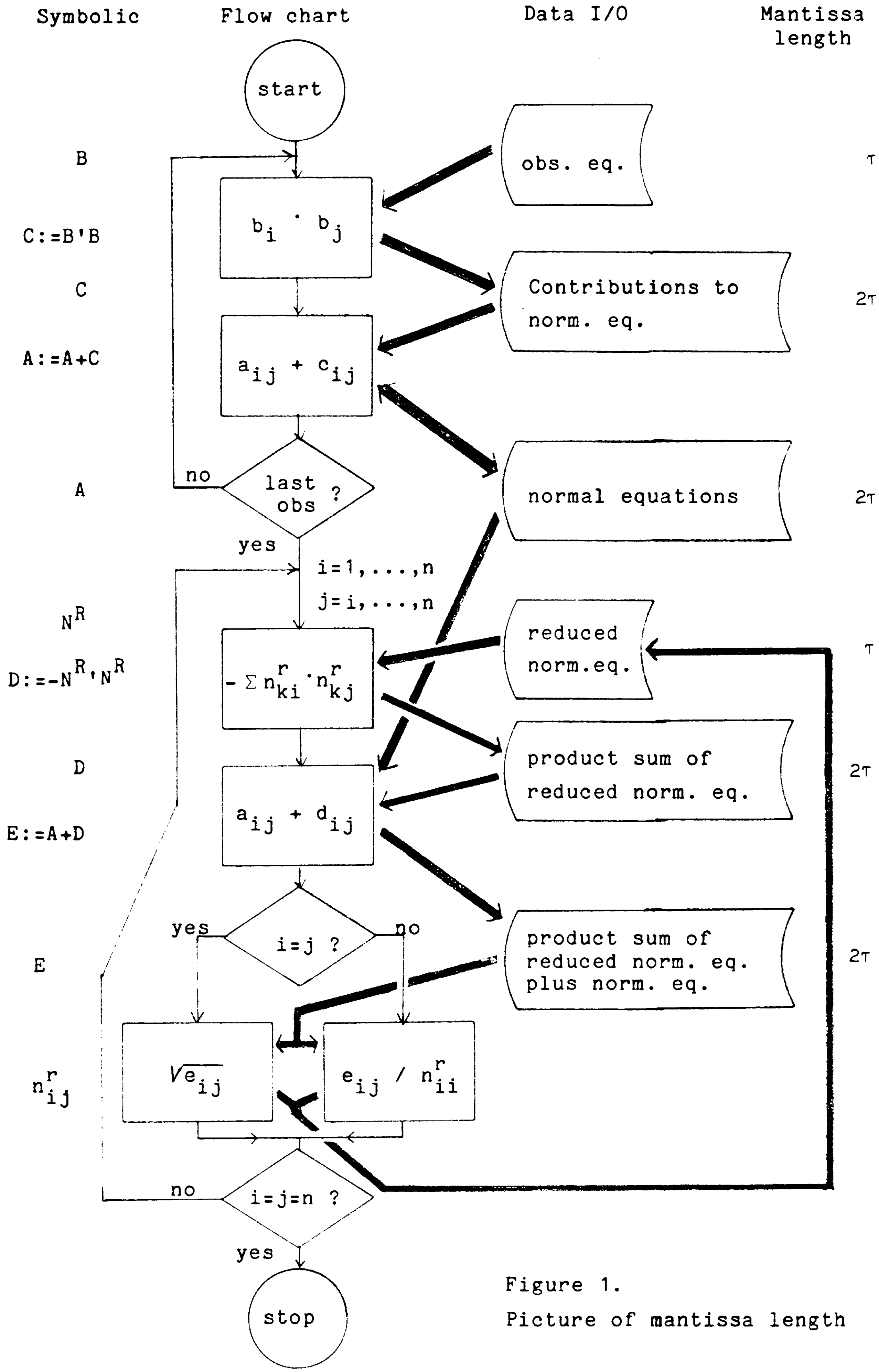

Figure 1.

Picture of mantissa length

The round off errors from the squareroot operation and division are not affected

$$\sigma\{\varepsilon_{ii}^{(s)}\} \leqslant \frac{c_{ii}^{(s)}}{\sqrt{12}} \beta^{-\tau} \qquad (13b)$$

$$\sigma\{\varepsilon_{ij}^{(d)}\} \leqslant \frac{c_{ij}^{(d)}}{\sqrt{12}} \beta^{-\tau} , \qquad i \neq j \qquad (13c)$$

where

$$c_{ij}^{(p)} = \beta^{\gamma} \geqslant$$

$$\mathrm{MAX}\{ |a_{ij}|, |a_{ij}^{(i-1)}| ,$$

$$|r_{ki} \; r_{kj}| , \; k = 1,\ldots,i-1 ,$$

$$|\sum_{k=1}^{l} r_{ki} \; r_{kj}| , \; l = 1,\ldots,i-1 \} ,$$

$$c_{ii}^{(s)} = 2 \; r_{ii} \; \beta^{\gamma} \geqslant 2 \; r_{ii} \; r_{ii}$$

and $c_{ii}^{(d)} = r_{ii} \; \beta^{\gamma} \geqslant r_{ii} \; |r_{ij}|$.

Futhermore for the right hand side

$$\sigma\{\eta_i^{(p)}\} \leqslant \frac{d_i^{(p)}}{\sqrt{12}} \sqrt{\nu_i} \; \beta^{-2\tau} \qquad (14a)$$

$$\sigma\{\eta_i^{(d)}\} \leqslant \frac{d_i^{(d)}}{\sqrt{12}} \beta^{-\tau} \qquad (14b)$$

where $d_i^{(p)} = \beta^{\gamma} \geqslant \mathrm{MAX}\{ |b_i| , |b_i^{(i-1)}| ,$

$$|b_k^{(k-1)} \; r_{ki} / r_{kk}| , \quad k=1,\ldots,i-1 ,$$

$$|\sum_{k=1}^{l} b_k^{(k-1)} \; r_{ki} / r_{kk}| , \; l=1,\ldots,i-1 \}$$

and $d_i^{(d)} = r_{ii} \; \beta^{\gamma} \geqslant |b_i^{(i-1)}|$.

Recall the offset of the solution x in equation (4)

or

$$\xi_i = \sum_{j=1}^{n} f_{ij} x_j \; \varepsilon_{jj} - \sum_{j=1}^{n} \sum_{k=j+1}^{n} (f_{ij} \; x_k + f_{ik} \; x_j) \; \varepsilon_{jk} + \sum_{j=1}^{n} f_{ij} \; \eta_j \qquad (15)$$

Then

$$\sigma^2\{\xi_i\} = \sum_{j=1}^{n} f_{ij}^2 \; x_j^2 \; \sigma^2\{\epsilon_{ij}\}$$

$$+ \sum_{j=1}^{n} \sum_{k=j+1}^{n} (f_{ij}x_k + f_{ik}x_j)^2 \; \sigma^2\{\epsilon_{jk}\} \qquad (16)$$

$$+ \sum_{j=1}^{n} f_{ij}^2 \; \sigma^2\{\eta_j\} ,$$

where $\sigma^2\{\epsilon_{ij}\} = \sigma^2\{\epsilon_{ij}^{(p)}\} + \sigma^2\{\epsilon_{ij}^{(s)}\} + \sigma^2\{\epsilon_{ij}^{(d)}\}$

and $\sigma^2\{\eta_i\} = \sigma^2\{\eta_i^{(p)}\} + \sigma^2\{\eta_i^{(d)}\}$.

Suppose a bound $\|f\|$ is available on the elements f_{ij} of A^{-1}

$$\|f\| = \underset{i,j}{MAX} \{|f_{ij}|\} = \underset{i}{MAX} \{|f_{ii}|\},$$

and $\|x\| = \underset{i}{MAX} \{|x_i|\}$

and bounds cs and cm

$$cs = \underset{1 \leqslant i \leqslant j \leqslant n}{MAX} \{c_{ijk}^{(\cdot)} , c_{ijk}^{(+)} , k=1,\ldots,i-1 , c_{ij}^{(-)}\} ,$$

$$cm = \underset{1 \leqslant i < j \leqslant n}{MAX} \{c_{ii}^{(s)}, c_{ij}^{(d)}\}$$

and similiarly

$$ds = \underset{1 \leqslant i \leqslant n}{MAX} \{ d_{ik}^{(\cdot)}, d_{ik}^{(+)}, k=1,\ldots,n , d_i^{(-)}\} ,$$

$$dm = \underset{1 \leqslant i \leqslant n}{MAX} \{ d_i^{(d)}\} .$$

When the elementary round off errors $\epsilon_{jk}^{(el)}$ affects ξ_i by

$$\xi_i(\epsilon_{jk}^{(el)}) = \begin{cases} f_{ij} \; \epsilon_{jj}^{(el)} & , j = k \\ (f_{ij}x + f_{ik}x_j) \; \epsilon_{jk}^{(el)}, & j \neq k \end{cases} \qquad (17)$$

then the double precision operations are bounded by

$$\sigma\{ \xi_i(\epsilon_{jk}^{(double)})\} \leqslant \frac{2cs}{\sqrt{12}} \|f\| \; \|x\| \; \beta^{-2\tau} \qquad (18a)$$

and the single precision operations are bounded by

$$\sigma\{ \xi_i(\epsilon_{jk}^{(single)})\} \leqslant \frac{2cm}{\sqrt{12}} \|f\| \; \|x\| \; \beta^{-\tau} \qquad (18b)$$

and for the right hand side

$$\sigma\{\xi_i(\eta_i^{(double)})\} \leqslant \frac{ds}{2} \|f\| \beta^{-2\tau}, \qquad (19a)$$

$$\sigma\{\xi_i(\eta_i^{(single)})\} \leqslant \frac{dm}{2} \|f\| \beta^{-\tau} . \qquad (19b)$$

The number of double precision operations used during the triangular decomposition of A is

$$\Gamma_d = \sum_{i=1}^{n} \sum_{j=i+1}^{n} (\mu_{ij}+1) + \mu_{ii}, \qquad (20)$$

and the number of single precision operations is

$$\Gamma_s = \sum_{i=1}^{n} (\mu_{ii}+1) = \Pi. \qquad (21)$$

Similiarly for the right hand side

$$\Pi_d = \sum_{i=1}^{n} \mu_{ii}, \qquad (22)$$

$$\Pi_s = n . \qquad (23)$$

By summation of elementary operations we arrive at the following estimate

$$\frac{\sigma\{\xi_i\}}{\|x\|} \leqslant [(cs^2 \Gamma_d \beta^{-2\tau} + cm^2 \Pi)/3 + (\frac{ds^2}{\|x\|^2} \Pi_d \beta^{-2\tau} + \frac{dm^2}{\|x\|^2} n)]^{\frac{1}{2}} \|f\| \beta^{-\tau} \qquad (24)$$

Due to the overestimation and the factors b^{-2t} this may be written

$$\frac{\sigma\{\xi_i\}}{\|x\|} \leqslant \sqrt{\Pi/3}\ cm\ \|f\|\ \beta^{-\tau} \qquad (25)$$

where $cm = \max_{1 \leqslant i < j \leqslant n} \{c_{ii}^{(s)}, c_{ij}^{(d)}\}$

$\leqslant 2\sqrt{\beta}\ \beta^{\gamma}$

where β^{γ} is the smallest number bounding the elements a_{ij} of A.

Using the earlier stated values of p, c = cm and $\|f\|$ we get

$$\frac{\sigma\{\xi_i\}}{\|x\|}_{NAD} \leqslant 0.79$$

$$\frac{\sigma\{\xi_i\}}{\|x\|}_{DEN} \leqslant 0.6\ 10^{-3}$$

which shows that it is safe to make the Danish network adjustment with a standard floating point arithmetic of a 36 bits mantissa however accumulating product sums with a floating point arithmetic of a 72 bits mantissa, - but the NAD network adjustment is still doubtful.

Looking on the refined estimate of the NAD network adjustment we arrive at an estimate of any 2^{o} by 2^{o} quad

$$\frac{\sigma\{\xi_p\}}{\|x\|} \leqslant 0.5\ 10^{-3}$$

which shows that it from point of view of arithmetic is safe to make an adjustment of a network of the size of the NAD network, however has the size fo the datamanagement problem not been estimated and may not be able to be done by a computer of this size.

V. The vectorprocessor at Geodetic Institute.

When the new RC 8000 computer at Geodetic Institute was installed then it was decided that the configuration should include a vectorprocessor operating in accordance with the principles given by Wilkinson (1963). The vectorprocessor should be connected on the CPU-unibus of the RC 8000 connecting all the fast units of the system.

It was decided to construct the GPU (Geodetic Processor Unit) in cooperation between Geodetic Institute and RC Computer, Copenhagen, using existing hardware with changed microprograms.

T.Krarup outlined the logical design and basic ideas of the unit and the present writer developed the actual instruction set of the unit including the microprogramming and the testing of the unit aided by a small working group.

The design of the GPU and the connected software is so that more than one GPU can be connected to one RC 8000 system.

The GPU has an accumulator with a 79 bits mantissa against the usual 36 bits. Products are calculated by a 71 bits mantissa from two 36 bits operands. Accumulation is done by a 79 bits mantissa except when the exponent difference is bigger than 76. The accumulator content is normalized after each operation on the accumulator content. True rounding takes place when the accumulator content is stored by adding a one bit in the position followeding the least significand bit stored.

The GPU has the RC 8000 instruction set extended with vector instructions intended for the Cholesky's algorithm. To name three:

1) a single instruction calculate the inner product af two vectors

2) a hardware squareroot function

3) the vector instruction : $A := A + c \cdot B$, where A and B are vectors and c is a constant.

The problem of 1) weight singularities and 2) the global effect of round off errors as described by Meisl1 (1980) are handled very efficiently by the double precission accumulation of products so the only not very significand round off errors takes place in

1) the computation of the squareroot of the diagonal elements followed by inversion to obtain the inverse diagonal element

2) and the subsequent multiplication of the nondiagonal elements by an inverse diagonal element.

Timestudies show that the realtime used when the GPU is used during generation, reduction and backsolution of the normal equations is reduced by 25% and the cputime decreases by 88%.

VI. Conclusion.

Meissl's findings outlined in the foregoing supplemented by the very small round off errors in the accumulation in double precission as shown in equations (11) and (24) demonstrates that the efforts laid in the development of a special double precission processor is well justified in spite of the fact that a very good ordenary arithmetic of 10-11 decimal digits with true rounding is available in the standard version of the RC 8000.

The double mantissa length floating point arithmetic circumvents the problems of weight singularities so the need of

either introducing a bigsum and a smallsum and diciding to which sum an addend should be added

or transform away the weight singularity.

The introduction of the GPU vectorprocessor parallel to the CPU of the RC 8000 computer has given Geodetic Institute a very efficient tool to handle big matrix operations.

VII. References.

Andersson, O., Engsager, K. and Weng, W. L.: "Computation and Data-management of Blocks of medium size Geodetic Networks". In Proceedings of FIG-study Group 5B "Survey Control Networks". Aalborg 1982.

Engsager, Karsten: "Reference Manual for GPU." Geodetic Institute Internal Report. Copenhagen 1982.

Meissl, Peter: "A Priori Prediction of Roundoff Error Accumulation in the Solution of a Super-Large Geodetic Normal Equation System." Rockville, Md. 1980.

Mossin, Einar: "RC 8000 Computer Family Reference Manual." Copenhagen 1979.

Poder, K. and Mark, A.: "Ordering and Dissection of Geodetic Least Squares Equations." Copenhagen 1982.

Poder, K. and Tscherning, C. C.: "Cholesky's Method on a Computer." The Danish Geodetic Institute, Internal Report No. 8. Copenhagen 1973.

Wilkinson, J. H.: "Rounding Errors in Agebraic Processes." London 1963.

Postscript on the paper

The use of a Vectorprocessor

in Geodetic Applications

All the estimates in this paper and in "A Priori Prediction of Roundoff Error Accumulation in the Solution of a Super-Large Geodetic Normal Equation System" by Peter Meissl, should be decreased by a two factor because the sign has been counted as a significand digit. This holds for a base-16 machine too!

The use of Wilkinson's ideas conserning single/double precission ('digits/2' digits) is very efficient against the occurence of weight singularities. In case of a machine having too few significand digits it will decrease the global effect of roundoff errors.

Another aproach to resolve the weight singularity is to choose one stations as main station and let the others strongly connected stations be considered as excentric station to the main station. This is of cause not so straight forward to program.

Using the transformation as proposed by Meissl this might be done simply by scaling the observation equations. But be carefull to do this by changing the exponent of the number only. I.e. the scale factors are multiples of the base of the machine.

K. Engsager.
Copenhagen 1982