

Aust. J. Geod. Photogram. Surv.  
Nos 46 & 47 December, 1987.  
pp 57 - 68

*In 2009 I retyped this  
paper and changed  
symbols eg  $\hat{\sigma}_o^2$  to VF for  
my students' benefit*

## DEGREES OF FREEDOM - SIMPLIFIED

Bruce R. Harvey  
Division of Radiophysics, C.S.I.R.O.  
P.O. Box 76, Epping, NSW, 2121, Australia.

*Now at School of Surveying and Spatial  
Information Systems, University of NSW*

### ABSTRACT

A simple equation is given to calculate the degrees of freedom of a least squares computation which has a priori weights on the parameters and on the observations. The method can be applied easily because it requires a few simple calculations rather than multiplying several large matrices. The method also clearly indicates whether an a priori weight on any parameter contributes significantly to the least squares solution or not.

## 1. INTRODUCTION

This paper outlines the equations used in a general least squares and in Bayesian least squares. Bayesian least squares includes weighted a priori estimates of the parameters and is very useful in a number of geodetic applications. However, a rigorous method of calculating the degrees of freedom of the solution, as presented by Theil (1963) and described by Bossler (1972), is rather difficult to apply in practice.

An approximate method is presented which is easy to apply in practice and is sufficiently accurate in most cases. The method can be computed easily on a pocket calculator using results that are usually printed out by least squares computer programs without modifying the computer program. An example is given which compares the results of the rigorous and simplified methods. The example also shows that the simplified method is easy to apply and that it indicates clearly whether an a priori weight on an individual parameter is significant or not.

## 2. LEAST SQUARES THEORY REVIEWED

Least squares theory is covered in many textbooks (e.g. Mikhail, 1976; Krakiwsky, 1981; Harvey, 2006). A general method follows which takes into account the a priori estimates of the parameters and their variance covariance matrix (VCV), equations which relate observables to each other and equations which relate observables to the parameters to be estimated.

The general mathematical model relating the true parameters and the true observations is

$$F(X, L) = 0$$

The linearised model is

$$A\Delta x + Bv = b \quad (1)$$

and the results are obtained from

$$\Delta x = [A^T(BQB^T)^{-1}A + Q_{xa}^{-1}]^{-1}A^T(BQB^T)^{-1}b \quad (2)$$

$$v = QB^T(BQB^T)^{-1}(b - A\Delta x) \quad (3)$$

$$VF = \hat{\sigma}_o^2 = \frac{v^T Q^{-1}v + \Delta x^T Q_{xa}^{-1}\Delta x}{n_o - m + u} \quad (4)$$

$$Q_x = [A^T(BQB^T)^{-1}A + Q_{xa}^{-1}]^{-1} \quad (5)$$

$$Q_L = Q + QB^T(BQB^T)^{-1}AQ_xA^T(BQB^T)^{-1}BQ - QB^T(BQB^T)^{-1}BQ \quad (6)$$

where:

- $x_a$  are the a priori parameters
- $\mathbf{X}$  are the adjusted parameters
- $\ell$  are the observations
- $L$  are the adjusted observations
- $\mathbf{v}$  =  $L - \ell$  are the residuals; the order of  $\mathbf{v}$  is  $(n, 1)$
- $\Delta\mathbf{x}$  =  $\mathbf{X} - x_a$  are the corrections to parameters
- $\mathbf{A}$  =  $\partial F / \partial \mathbf{X}$  at  $(x_a, \ell)$ ; the order of  $\mathbf{A}$  is  $(n_o, m)$
- $\mathbf{B}$  =  $\partial F / \partial L$  at  $(x_a, \ell)$ ; the order of  $\mathbf{B}$  is  $(n_o, n)$
- $\mathbf{b}$  =  $-F(x_a, \ell)$ ; the order of  $\mathbf{b}$  is  $(n_o, 1)$
- $VF$  is the a posteriori estimate of the variance factor =  $\hat{\sigma}_o^2$
- $\mathbf{Q}$  is VCV matrix of observations
- $\mathbf{Q}_{x_a}$  is VCV matrix of a priori parameters
- $\mathbf{Q}_L$  is cofactor matrix of adjusted observations
- $\mathbf{Q}_X$  is cofactor matrix of estimated parameters
- $n_o$  = number of observation equations
- $n$  = number of observations
- $m$  = total number of parameters
- $u$  = number of parameters with a priori weights

Some related quantities are:

$$\mathbf{P}_{x_a} = \mathbf{Q}_{x_a}^{-1} \quad \text{and} \quad \mathbf{P} = \mathbf{Q}^{-1}$$

$$\Sigma_L = VF \mathbf{Q}_L \quad \text{is VCV matrix of adjusted observations}$$

$$\Sigma_X = VF \mathbf{Q}_X \quad \text{is VCV matrix of estimated parameters}$$

Now consider the special case where  $\mathbf{B} = -\mathbf{I}$  and  $n_o = n$ ; that is, in the mathematical model each equation contains contributions from only one observation (there are no conditions relating observations). The above equations then reduce to:

Harvey: Degrees of freedom

$$A\Delta x = b + v \quad (7)$$

$$\Delta x = [A^T P A + P_{xa}]^{-1} A^T P b \quad (8)$$

$$Q_x = [A^T P A + P_{xa}]^{-1} \quad (9)$$

$$v = A \Delta x - b \quad (10)$$

$$Q_L = A Q_x A^T \quad (11)$$

$$VF = \hat{\sigma}_o^2 = \frac{v^T P v + \Delta x^T P_{xa} \Delta x}{n - m + u} \quad (12)$$

The set of Equations (7) to (12) is commonly used in geodetic adjustments, such as VLBI.

## 2.1 Applications of a priori Weights on Parameters

There are many applications of least squares analysis where it is convenient to use prior knowledge of the parameters. These cases arise where model variables have been measured or estimated prior to the current data set being analysed but are not known well enough to hold them fixed. Some examples are as follows:

- i. Observations are often made in VLBI of radio sources whose positions have been determined in previous experiments and are recorded in catalogues. The accuracy of these positions is usually unknown. Additional observations may also be made of 'new' sources. The observations are used to determine parameters such as baseline vectors, positions of the new sources, changes in polar motion, etc. Obviously a better solution is obtained if the a priori accuracies of the catalogue sources are used in the solution rather than holding these source positions fixed. An example of this application is given later in this paper.
- ii. In satellite positioning (e.g. GPS or SLR) it may be feasible in some cases to include an a priori 'ephemeris' and an estimate of the VCV of its terms, obtained from an independent tracking network. This may then lead to an improvement in the determination of the satellite orbit. Depending on the circumstances, it may also be better than either completely solving for the orbit with no a priori information, or holding the given orbit parameters fixed.
- iii. If a survey team measures a geodetic network and connects on to points measured by another survey team, it may be preferable to include the given coordinates of these points and an estimate of their accuracies than to hold these points fixed. An example of this application is given by Bossler and Hanson (1980).

### 3. DEGREES OF FREEDOM

#### 3.1 Introduction

The degrees of freedom of a solution are required for several calculations and statistical tests. As shown in (4) and (12), they are required in the calculation of VF. A statistical test may be applied to determine whether VF is significantly different from the a priori variance factor ( $=\sigma_o^2$ ). This is useful for determining whether the models are reasonable and whether the data is likely to be severely contaminated by gross errors.

Moreover VF is also often used to scale the estimated cofactor matrices of estimated parameters and adjusted observations (see equations for the VCV of adjusted and observed parameters). This most often occurs when  $\sigma_o^2$  is poorly known, as is the case with new measurement techniques. Note that an error in the degrees of freedom will then directly affect the estimated precisions, error ellipses and confidence intervals of the least squares results.

Many techniques for the detection of gross errors by statistically analysing the observation residuals require reliable knowledge of the degrees of freedom of the least squares solution.

In a standard least squares adjustment (sometimes called weighted least squares or parametric adjustment), with no a priori weights on the parameters, the a posteriori variance factor (VF) is found from:

$$VF = \frac{v^T P v}{r}$$

where  $r$  is the degree of freedom in the adjustment and equals the number of observations minus the number of (free) parameters.

In Bayesian least squares, where a priori weights are assigned to the parameters, VF is found from (Krakiwsky, 1981):

$$VF = \frac{v^T P v + \Delta x^T P_{xa} \Delta x}{r'}$$

where  $r'$  is approximately equal to the number of observations minus the number of parameters without a priori weights.

This is an approximate formula that works best when  $r'$  is large. When  $r'$  is small a slight error in it will cause significant errors in VF. Another problem is the magnitude of the a priori weight of a parameter. If the weight is large - i.e. with small variance - then the parameter estimate is obviously affected by this weight.

## Harvey: Degrees of freedom

If the weight is small then it may not have much effect on the solution, so counting it as a weighted parameter would give a misleading value of  $r'$ . In such a case the analyst may choose to regard parameters with small weights as not weighted for the purposes of calculating  $r'$  and VF. Then the problem of determining whether a weight is large or small arises. Again this problem is not so critical when there are many more observations than parameters, i.e.  $r'$  is large. Theil (1963) describes a procedure to overcome this problem and to obtain more accurate estimates of  $r'$  and VF. The procedure to be used with weighted parameter solutions is outlined below.

### 3.2 Rigorous Calculation of Degrees of Freedom

Step i. Compute a value of VF ( $VF_f$ ) from a solution with no weights on the parameters. Using

$$v = A \Delta x - b$$

$$\Delta x = (A^T P A)^{-1} A^T P b$$

$$VF_f = \frac{v^T P v}{n - m}$$

Step ii. Multiply the VCV of the observations by  $VF_f$

Step iii. Compute  $\Delta x$  and  $v$  from a solution with weighted parameters (use Equations 8 and 10).

Step iv. Compute  $u'$ , the number of unweighted parameters, from (Theil, 1963)

$$u' = \text{tr} \left\{ \frac{A^T P A}{VF_f} \left( \frac{A^T P A}{VF_f} + P_{xa} \right)^{-1} \right\} \quad (13)$$

$$u' = \text{tr} \left\{ A^T P_t A \left( A^T P_t A + P_{xa} \right)^{-1} \right\} \quad (14)$$

where  $P_t = \left( \frac{1}{VF_f} \right) P$  and  $\text{tr}\{ \}$  is the trace i.e. sum of the diagonals, of a matrix

$P$  being the inverse of the original VCV of the observations, and

$P_t$  the inverse of the new (by step ii) VCV of the observations. Note that  $u'$  is not necessarily an integer. Therefore the degrees of freedom in the adjustment, which is the number of observations minus  $u'$ , is not necessarily an integer.

Step v. Compute the final VF from

$$VF = \frac{v^T P_t v + \Delta x^T P_{xa} \Delta x}{n - u'}$$

Step vi. Compute  $\Omega_s$ , the share of VF due to the VCV of the observations, and  $\Omega_p$ , the share of VF due to the a priori weights of the parameters, where

$$\Omega_s = u'/m \quad \text{and} \quad \Omega_p = 1 - \Omega_s$$

### 3.3 A Disadvantage of Theil's Method

Many computer programs do not write out the A matrix, so in order to apply Equation (14) the program has to be modified, either to write out the A matrix or to do the complete calculation. However, a number of least squares programs, especially commercially available programs, do not supply a listing of the program source code. In this case it is not possible for the user to modify the program.

### 3.4 Simplified Equations for Degrees of Freedom

If good estimates of the VCV of the observations are available then  $VF_f$  is usually close to 1 (assuming the a priori  $\sigma_o^2$  is 1, as in common practice). If VF is close to 1 then it can be ignored. In any case it is usually simple enough to compute two solutions, one with  $P_{xa} = 0$  and one with  $P_{xa} \neq 0$ . That is, the second solution is computed with a new, scaled, VCV matrix - i.e. a new P matrix,  $P_t$ . In the following sections we deal with the results of the second solution. From (9) we have

$$(A^T P_t A + P_{xa})^{-1} = Q_x$$

$$\text{so} \quad u' = \text{tr}\{A^T P_t A Q_x\}$$

$$\text{now} \quad A^T P_t A + P_{xa} = Q_x^{-1}$$

$$\text{so} \quad A^T P_t A = Q_x^{-1} - P_{xa}$$

$$\text{thus} \quad u' = \text{tr}\{(Q_x^{-1} - P_{xa}) Q_x\}$$

$$= \text{tr}\{I - P_{xa} Q_x\}$$

$$= \text{tr} I - \text{tr}\{P_{xa} Q_x\}$$

$$= m - \text{tr}\{P_{xa} Q_x\}$$

(15)

## Harvey: Degrees of freedom

Equation (15) is simpler than (14). The degrees of freedom,  $r'$ , equals  $n - u'$ , so:

$$r' = n - m + \text{tr}(P_x Q_x) \quad (16)$$

Since  $P_{x_a}$  is known (it is input to the solution) and  $Q_x$  is usually output, the degrees of freedom can be calculated without modifying the program. Note that some programs may not produce  $Q_x$  but merely give the standard deviations of the estimated parameters, or the standard deviations plus the correlations between the estimated parameters. In these cases  $Q_x$  can be regenerated or at least approximated. Equation (16) does not require the A matrix, which may be large and is rarely output by least squares programs.

Further simplifications may be made depending on the structure of  $P_{x_a}$  and  $Q_x$ .  $P_{x_a}$  is often a diagonal matrix. The  $i^{\text{th}}$  diagonal term is  $1/\sigma_{ai}^2$ , where  $\sigma_{ai}^2$  is the a priori variance of the  $i^{\text{th}}$  parameter.  $Q_x$  is usually not diagonal. However, if the correlations between the estimates of the parameters are small then  $Q_x$  will be close to diagonal. If  $P_{x_a}$  and  $Q_x$  are both diagonal then the calculation  $P_{x_a} Q_x$  is simple. Let the  $i^{\text{th}}$  diagonal terms of  $Q_x$  be  $\sigma_{ei}^2$  i.e. the estimated variances. Then the number of weighted parameters is

$$\text{tr}\{P_{x_a} Q_x\} = \sum_{i=1}^m \frac{\sigma_{ei}^2}{\sigma_{ai}^2} = \sum_{i=1}^m \left( \frac{\sigma_{ei}}{\sigma_{ai}} \right)^2$$

$$\text{Thus } r' = n - m + \sum_{i=1}^m \left( \frac{\sigma_{ei}}{\sigma_{ai}} \right)^2 \quad (17)$$

So dividing the estimated standard deviation of a parameter by the corresponding a priori standard deviation and then squaring gives the contribution of that parameter to the degrees of freedom. If either  $P_{x_a}$  or  $Q_x$  or both contain large correlations then it may be necessary to calculate  $\text{tr}\{P_{x_a} Q_x\}$ .

It can be shown, from (5) that

$$\sigma_{ei} \leq \sigma_{ai}$$

$$\text{so } 0 \leq \sigma_{ei}^2 / \sigma_{ai}^2 \leq 1$$

If the a priori variance is small the a priori weight will be large. This means the observations will not contribute very much to the final parameter estimate. Thus  $\sigma_{ei}$  will be approximately equal to  $\sigma_{ai}$  and so  $(\sigma_{ei}/\sigma_{ai})^2$  will be close to 1. In this case the parameter can be considered significantly weighted. (Remember, the degrees of freedom equals number of observations minus the number of unweighted parameters).

Conversely, if a parameter has a small weight then the observations will contribute substantially to the parameter estimate. Thus  $\sigma_e$  will be considerably smaller than  $\sigma_a$  and  $(\sigma_e/\sigma_a)^2$  will be very small, approaching zero as  $\sigma_a$  tends to infinity.

Consider the case where parameters are given either very large or very small weights. This is analogous to least squares without a priori weights on parameters where such parameters are either held fixed or 'solved for' with no constraints. Then  $(\sigma_{ei}/\sigma_{ai})^2$  will equal either 0 or 1. The fixed parameters have  $(\sigma_{ei}/\sigma_{ai})^2 = 1$  and the free parameters have  $(\sigma_{ei}/\sigma_{ai})^2 = 0$ . Then  $\Sigma(\sigma_{ei}/\sigma_{ai})^2 =$  number of fixed parameters, the number of free parameters is  $m - \Sigma(\sigma_{ei}/\sigma_{ai})^2 = u'$  and the degrees of freedom is  $n - u' = r$  as obtained in a standard least squares adjustment.

Another satisfying feature confirming the illustrative nature of (17) is that as the number of weighted parameters increases the degrees of freedom increase. That is, if more independent a priori information is available the solution is more reliable, as should be expected. Examining the individual  $(\sigma_{ei}/\sigma_{ai})^2$  terms informs the analyst of the significance of the a priori constraints placed on each parameter.

#### 4. NUMERICAL EXAMPLE

Theil's (1963) procedure was applied to the analysis of the data obtained in the 1982 Australian VLBI experiment (Harvey, 1985). The Tidbinbilla-Parkes solution involved 33 source coordinates with a priori standard deviations of  $\pm 0.03''$ , 12 source coordinates with zero a priori weight, and 14 other parameters with zero a priori weight. There were a total of 59 parameters and 290 observations. Theil's procedure was followed and it was found that the number of unweighted parameters was 33.457, thus showing that a standard deviation of  $\pm 0.03''$  is a significant weight in this experiment (degrees of freedom = 256.543).

It was also found that the share of VF due to the variances of the observations was 57% and the share due to the weights of the parameters was 43%. Thus the a priori weights of the parameters do have a significant effect on the estimated variance factor.

**Table 1**

**Example of simplified calculations for degrees of freedom**

Parameter	$\sigma_{ai}$	$\sigma_{ei}$	$(\sigma_{ei} / \sigma_{ai})^2$
17	0.002	0.0020	0.99
18	0.030	0.0295	0.97
19	0.002	0.0019	0.87
20	0.030	0.0243	0.66
21	0.002	0.0019	0.91
22	0.030	0.0224	0.56
23	0.002	0.0019	0.89
24	0.030	0.0240	0.64
25	0.002	0.0019	0.88
26	0.030	0.0229	0.58
31	0.002	0.0020	0.99
32	0.030	0.0282	0.89
33	0.002	0.0019	0.89
34	0.030	0.0236	0.62
35	0.002	0.0018	0.84
36	0.030	0.0210	0.49
37	0.030	0.0222	0.55
38	0.002	0.0019	0.92
39	0.030	0.0262	0.76
40	0.002	0.0019	0.94
41	0.030	0.0260	0.75
42	0.002	0.0019	0.94
43	0.030	0.0281	0.88
46	0.002	0.0018	0.82
47	0.030	0.0206	0.47
48	0.002	0.0018	0.85
49	0.030	0.0199	0.44
52	0.002	0.0019	0.86
53	0.030	0.0221	0.54
54	0.002	0.0020	1.00
55	0.030	0.0290	0.93
58	0.002	0.0018	0.83
59	0.030	0.0193	0.41

TOTAL = 25.549

The procedure recommended in this paper (17) was also carried out, and the results are shown in Table 1. The value of  $\sigma_a$  for those parameters not listed is infinite, i.e. the corresponding weight is zero. Thus their value of  $(\sigma_{ei}/\sigma_{ai})^2$  is zero. Note that in these calculations it is not necessary to use the internal units of the program (e.g. radians or kilometres); any convenient unit (e.g. seconds of arc, seconds of time, centimetres; etc.) can be used, provided the same units are used for corresponding  $\sigma_{ei}$  and  $\sigma_{ai}$ .

In this example the number of weighted parameters is 25.549 and the degrees of freedom ( $r'$ ) of the solution is  $290 - 59 + 25.549 = 256.549$ . Considering that  $r'$  is normally rounded to the nearest integer this is not significantly different from the value obtained from the rigorous calculation.

When considering this example it must be noted that  $P_{xa}$  was diagonal, thus making (17) more accurate in this case. It is also necessary to consider the off diagonal terms in  $Q_x$ . In this example the correlations between those parameters where  $1/\sigma_{ai} \neq 0$  were considered. Note that if  $1/\sigma_{ai} = 0$  then the correlations between the  $i^{\text{th}}$  parameter and any other parameter will not affect the result because all the terms in the  $i^{\text{th}}$  row and the  $i^{\text{th}}$  column of  $P_{xa}$  will equal zero.

The largest correlations between parameters with  $1/\sigma_{ai} \neq 0$  was 0.18. This is because they were weighted, and therefore the observations did not have much effect on the estimates of these parameters, and thus did not introduce large correlations.

#### 4.1 A Similar Method

Another simple way to calculate the degrees of freedom is to calculate the redundancy numbers of the observations (e.g. Caspary, 1987). In this case:

$$r \approx \sum_{i=1}^n \left( \frac{q_{vi}}{\sigma_{li}^2} \right) \quad (18)$$

where  $qv$  is the diagonal term of the cofactor matrix of residuals ( $Q_v$ ) and  $\sigma_l^2$  is the corresponding a priori variance of the observations. Caspary (1987) applies this equation to the case  $P_{xa} = 0$ . Equation (18) is surprisingly similar to (17) proposed in this paper, but it uses observation variances instead of parameter variances. However, (18) has two disadvantages. Firstly, there are usually many more observations than parameters, thus leading to many more terms in (18) than in (17). Secondly, and most important, many least squares computer programs do not write out the necessary cofactor matrix of the residuals. Even if the program can be modified it is usually found that considerable extra computer time and space are required to compute  $Q_v$ .

## 5. CONCLUSIONS

Analysts may have been reluctant to implement the equations and procedure recommended by Theil (1963) and Bossler (1972) because of their complexity or the need to modify programs. However, this problem is overcome with the simpler equations presented here. Moreover, (17) intuitively makes sense, is easy to understand, and may help some analysts understand what is happening in what they view as the “black box” least squares program package. An examination of the individual  $(\sigma_{ei}/\sigma_{ai})^2$  terms reveals the significance of the a priori constraints placed on each parameter.

For most applications slight errors in  $r'$  are not important, especially if  $r'$  is large. However, if correlations between those parameters with significant a priori weights are considerable, then (16) should be used. It will give the correct answer for  $r'$  and is computationally better than (14).

Moreover (17), and often also (16), can be applied even when it is not possible to modify the least squares program.

## 6. ACKNOWLEDGEMENTS

The research reported in this paper was carried out while the author was the holder of a CSIRO postdoctoral award.

## 7. REFERENCES

- Bossler, J.D., 1972. Bayesian inference in geodesy. Ph.D. Thesis, Dept. Geod.Sc., The Ohio State University (revised and reprinted, 1976)
- Bossler, J.D. and Hanson, R.H., 1980. Application of special variance estimators to geodesy. NOAA Tech. Rep. NOS 84 NGS15, U.S. Dept. of Commerce/NOAA.
- Caspary, W.F., 1987. Concepts of network and deformation analysis. Monograph 11, School of Surveying, University of N.S. W.
- Harvey, B.R., 1985. The combination of VLBI and ground data for geodesy and geophysics. UNISURV S-27, School of Surveying, University of N.S. W.
- Harvey, B.R., 2006. Practical Least Squares and Statistics for Surveyors, Monograph 13, Third Edition, School of Surveying and SIS, UNSW. 332 + x pp. ISBN 0-7334-2339-6
- Krakiwsky, E.J., 1981. A synthesis of recent advances in the method of least squares. Dept.Surv.Eng., Uni. of Calgary, Canada, Publ. 10003.
- Mikhail, E.M., 1976. Observations and Least Squares. Harper and Row, 497 pp.
- Theil, H., 1963. On the use of incomplete prior information in regression analysis. J.Am.Stat.Assoc., 58, 401-414.

Received: 1 June, 1987. Reviewed: 13 July, 1987. Accepted: 28 July, 1987.  
*Retyped, symbols changed, reference added: Sep, 2009*