

MORE REMARKS ON SELF-CALIBRATION

Y.D. Huang

Photogrammetric Laboratory, Civil Engineering Department
Northern Jiaotong University, Beijing, P.R. China

ABSTRACT

The **Problems** including the determination of systematic error models, the selection of additional parameters and the treatment of approximate correlation are studied more theoretically, which must be considered in the application of self-calibration. It is pointed out that self-calibration always enlarges the variances of the estimators of the main unknowns although it may rectify their biases, and that the mean square error reflecting both the variance and the bias should be adopted as the measure of quality of estimators and be minimized in order to attain the best effect of self-calibration. The proposition is proved that the reversible linear transformation of the systematic error models has no effect on the least squares estimators of the main unknowns. A formulated criterion is given for the selection of parameters, based on which a new proposal is made for the statistical test as a tool of selecting additional parameters. It is also pointed out that the approximate correlation among unknowns in self-calibration not only causes difficulty in numerical solution but also depresses the quality of estimators, and that the key to treating it is to improve the estimating method.

INTRODUCTION

After many years of study and test, the technique of self-calibration has been widely used in photogrammetric practice. It has brought a considerable improvement to the quality of photogrammetric results due to its function in compensating the systematic errors. Nevertheless, it should be admitted that quite a few problems with self-calibration have not yet been solved completely. Self-calibration is not always effective. Sometimes it even deteriorates the quality of results. So far we have known little about the regularity of function of self-calibration, and cannot be sure about whether it is beneficial in a particular case or not. Statements and realizations on some questions are still not identical. In the past, researches were made mainly by proposing some method or approach based on experience or intuition and then testifying it by experiments, but seldom into the theoretical basis. Now it is time for an overall theoretical study on self-calibration in order to provide a systematic and reliable theoretical basis for the effective and safe application of self-calibration. The theoretical study is also required for the perfection of the theory itself of self-calibration. In this paper, some problems including the determination of systematic error models, the selection of additional parameters and the treatment of approximate correlation are studied more theoretically, which must be considered in the application of self-calibration. Some new viewpoints are put forward.

SEEKING MINIMUM MEAN SQUARE ERROR

As is well known, direct or indirect observations are all random variables. So they have such characteristic quantities of probability distribution as the expectation and the variance from statistics. Their errors appear only when they are used to estimate unknown quantities. If \tilde{X} denotes the observation vector and X denotes the true value vector of the quantities to be estimated by \tilde{X} , then the true error vector of the estimation denoted by e is

$$e = \tilde{X} - X .$$

e is also a random vector. The sum of all the second moments of its elements, i.e. the statistical mean of $e'e$, is called the mean square error of \tilde{X} estimating X and denoted by Mse in regression analysis. That is

$$Mse(\tilde{X}) = E(e'e) = E((\tilde{X}-X)'(\tilde{X}-X)) \quad (1)$$

where $E(\cdot)$ denotes an expectation. In the case of single element vectors, the Mse is evident to be able to measure the proximity of the estimator to the estimated quantity — the accuracy, so can be used to assess the quality of the estimator, which was proposed by Gauss (Mikhail, 1976). In general, this reason also applies to vectors with more than one elements.

Evolving Eq.(1) with taking account of the statistic definitions of variance and bias, we have

$$\begin{aligned} Mse(\tilde{X}) &= E((\tilde{X}-E\tilde{X})'(\tilde{X}-E\tilde{X})) + E((E\tilde{X}-X)'(E\tilde{X}-X)) \\ &= \text{tr}(\text{Var}(\tilde{X})) + (\text{Bias}(\tilde{X}))'(\text{Bias}(\tilde{X})) \quad , \end{aligned} \quad (2)$$

which shows that the mean square error consists of the variance term and the bias term. As a rule, it is thought that the variance reflects the random error and the bias reflects the systematic error. So the mean square error can be considered as a quality measure that reflects both the random errors and the systematic errors, and accordingly the real quality of estimators.

The variance of a random variable measures the extent to which a group of values of this random variable concentrate. It is regarded as the measure of precision of estimators in a lot of literature. Now it can be seen from Eq.(2) how precision differs from accuracy. The variance is just a part of the mean square error. Only when the bias is equal to zero is it reasonable for the variance to replace the mean square error to assess the quality of result. A smaller value of either variance or bias only does not necessarily mean a better result. In estimation we should strive to minimize the combined effect of bias and variance, that is, to minimize the mean square error. Especially when decreasing the variance contradicts decreasing the bias, it is crucial to consider both the variance and the bias comprehensively in order to procure the best result.

SELF-CALIBRATION DECREASES BIAS BUT INCREASES VARIANCE

When it is required to use a group of observations to estimate some unknown quantities which are relevant to those observations, we always first establish based on some ideal the function model which describes their relation and the stochastic model of the errors, then estimate the unknown quantities according to the least squares principle. Suppose that L is the observation vector, X is the vector of unknowns to be estimated and the models are

$$E(L) = AX \quad , \quad \text{and} \quad \text{Var}(L) = \sigma^2 I \quad . \quad (3)$$

Then the vector of least squares (abbreviated LS) estimator of X , denoted by \hat{X} , is

$$\hat{X} = (A'A)^{-1}A'L = N^{-1}A'L \quad . \quad (4)$$

If the models (3) are true, then

$$E(\hat{X}) = N^{-1}A'E(L) = (A'A)^{-1}A'AX = X \quad (5)$$

$$\text{Var}(\hat{X}) = N^{-1}A'\text{Var}(L)AN^{-1} = \sigma^2 N^{-1} \quad . \quad (6)$$

\hat{X} is the best linear unbiased estimate (vector) of X . But in practice, there is always some difference between the actual observations and the assumed ideal function model. The reason for that can be explained from different standpoints as the poor fidelity of the model or as the systematic errors in the observations. Now the true function model may be

$$E(L) = AX + V_S \quad , \quad V_S \neq 0 \quad , \quad (7)$$

rather than the one in models(3) (the stochastic model remains as that in models(3) all the time in the context). In this case, for \hat{X} derived from Eq.(4), we have

$$E(\hat{X}) = N^{-1}A'E(L) = N^{-1}A'(AX+V_S) = X + N^{-1}A'V_S \neq X \quad (8)$$

It shows that \hat{X} is no longer unbiased at this moment. If V_S is too large, \hat{X} will not be acceptable. One way to rectify the bias of \hat{X} is to refine the function model by supplementing a term, say HS, to describe the systematic error V_S so that the refined function model fits the observations, that is, there is

$$E(L) = AX + HS \quad , \quad (9)$$

then to execute the least squares method with regarding S as well as X as unknowns. As a result, the unbiased estimator of X accompanied by that of S is acquired. For distinguishing from the main unknown parameters in X, the elements of S are called additional parameters, which are introduced only for rectifying the biases. Such an approach of compensating systematic errors is referred to as self-calibration in photogrammetry.

It will be seen from the forthcoming Eq.(15) and Eq.(14) that self-calibration can rectify the biases of estimators but always inflates their variances. Therefore, in order to gain quality from self-calibration, we must take account of both aspects of variance and bias and take the mean square error as the measure of quality. The determination of systematic error models, the selection of additional parameters, the treatment of approximate correlation and suchlike measures must all clearly aim at minimizing the mean square error.

ADDITIONAL PARAMETER REVERSIBLE LINEAR TRANSFORMATION HAS NO EFFECT ON MAIN PARAMETER ESTIMATORS

It is to compensate systematic errors and acquire unbiased estimators of unknowns that self-calibration is applied. So it is evident that the supplemented model must be able to describe the systematic errors really, in other words, to have a high fidelity.

The current methods for the determination of systematic error models can be classified into two categories. One is to decompose the systematic errors into a number of elements to each of which a physical explanation and a describing submodel can be given, then to compose all the submodels into the systematic error model, which is known as the physical model. The other method is to use some ideal mathematical formula to fit the total systematic errors without accounting for the physical factors and the explanation of the parameters. Such models are named empirical models. So far whether physical models or empirical models should be adopted has not been reconciled yet. In fact, it is difficult to give a general answer to that. Experiences in analytical aerotriangulation have shown that none of the tested models can be superior to the others in all cases. Therefore, the key seems to be making it clear in which cases a model has a good description of the systematic error sources and adopting the particular model in the particular cases.

Once a systematic error model is determined, its reversible linear transformation does not affect the least squares estimators of the main unknown parameters. In fact, if LS estimation is made according to the following model:

$$L = AX + HCS + V_r$$

in which C is a reversible matrix, then the LS estimator \hat{X} of X is derived from

$$\begin{aligned} \hat{X} &= N^{-1}A'(I-HC(C'H'(I-AN^{-1}A')HC)^{-1}C'H'(I-AN^{-1}A'))L \\ &= N^{-1}A'(I-H(H'(I-AN^{-1}A')H)^{-1}H'(I-AN^{-1}A'))L \end{aligned} \quad (10)$$

It is seen from the final expression in Eq.(10) that the LS estimator \hat{X} of the main parameter vector X is independent of the transformation matrix C. This fact suggest us that the form of a model is not significant to self-calibration, and that in the determination of systematic error models consideration should be given mainly to the space described by the model and to making a sufficient space described with as few parameters as possible. In addition, by proper

transformation a model can be changed into an orthogonal or simpler form so as to facilitate the computation without losing its fidelity.

Some of the existing systematic error models in photogrammetry can be transformed linearly from one to another. For example, the spherical harmonic function model containing ten parameters can be transformed linearly to the complete three degree polynomial model (Schut,1979). From above discussion we know those models have the same effect in self-calibration . The major difference is that the models of good orthogonality, such as the spherical harmonic function model, can make the numerical solution stable and the statistical test of the parameters convenient. Many empirical models were proposed for facilitating the computation and fitting the anomalous systematic errors which cannot be explained. Hence they can bring less difficulty in numerical solution. But the empirical models could not fit the several major systematic errors as well as the physical models. Recently, S.Murai derived the conclusion that the physical models were superior to the empirical models in fitting the image systematic errors from his study on comparing nine models including three physical ones in the close-range photogrammetric cases with nonmetric cameras (Murai,1984). This result might prove that the physical models possess higher fidelities than the empirical models. The chief defect of the physical models is likely to be the possible approximate correlation among their parameters that can make the LS solution unstable or even fail. But this defect can be overcome by other means, such as, applying linear transformation to the additional parameters, using the Marquardt method(Marquardt, 1963) or other improved LS methods to realize the LS solution, and for the convenience of statistical tests, implementing the posteriori orthogonalization of the additional parameters (A. Grün,1978), and etc.. For the above reasons, the empirical models should not be complemented too highly, especially when their fidelities have not been ensured.

Fidelity mentioned here has not been defined strictly. In the author's opinion, the fidelity of a model should be measured by the residuals of image coordinate observations rather than the residuals (or true errors) of the photogrammetrically determined coordinates of object points. If a systematic error model, when used in self-calibration, leads to a big drop in the residuals of image observations, it is thought to be of high fidelity. It is then evident that adding parameters to a model always gains fidelity and decreases the fitting errors. But it will be seen in the section right below that a model of high fidelity only dose not necessarily bring a good final result and that some parameters need to be rejected. Therefore, we should choose such "candidates" for systematic error models that can most probably describe the systematic errors and also contain the fewest additional parameters.

FORMULATED CRITERION FOR ADDITIONAL PARAMETER SELECTION

Since systematic errors are caused by a variety of factors, a mathematical formula describing the space in which the systematic errors vary always contains a large number of parameters. However, for a certain realization of photogrammetry, not all those factors have a great effect, always some have, but some not. Its reflection on the mathematical model is that some parameters differ significantly from zero and some approach zero. Then including all the parameters in self-calibration will increase the amount of computational work not necessarily, and worse, will inflate the effect of random errors on unknowns' estimators. Therefore, it is always hoped to reject those additional parameters which have little effect of compensating systematic errors or even deteriorate the result.

The fact has been testified by a great deal of practice that the self-calibration including inappropriate additional parameters can deteriorate the result of estimation. The reason for that has been explained in a popular and intuitive manner before. But those explanations are not sufficient to give instructions for preventing that phenomenon. Now we shall analyse the phenomenon in more mathematical terms.

Suppose that Eq.(9) is the true model and $H=(H_1 H_2)$, $S'=(S_1' S_2')$. Then the model for estimation corresponding to Eq.(9) is

$$L - V = AX + HS = AX + H_1S_1 + H_2S_2 \quad , \quad (11)$$

which is named complete model or true model. And the model with a part of all the (additional) parameters, say S_2 , rejected is

$$L - V = AX + H_1S_1 \quad , \quad (12)$$

which is named submodel. Now examine the least squares estimators of the main unknowns (vector X) obtained respectively by model(11) and model(12).

By model(11), the LS estimator \hat{X} of X is

$$\hat{X} = N^{-1}A'(I - HQ_{SS}^{\wedge}H'Q_{VV}^{\wedge})L \quad (13)$$

in which $Q_{SS}^{\wedge} = (H'Q_{VV}H)^{-1} \triangleq \begin{pmatrix} Q_{S_1S_1}^{\wedge} & Q_{S_1S_2}^{\wedge} \\ Q_{S_2S_1}^{\wedge} & Q_{S_2S_2}^{\wedge} \end{pmatrix}$, $Q_{VV}^{\wedge} = I - AN^{-1}A'$

The cofactor matrix of \hat{X} is

$$Q_{\hat{X}\hat{X}}^{\wedge} = N^{-1} + N^{-1}A'HQ_{SS}^{\wedge}H'AN^{-1} \quad , \quad (14)$$

and the bias of \hat{X} is

$$\text{Bias}(\hat{X}) = E(\hat{X}) - X = N^{-1}A'(I - HQ_{SS}^{\wedge}H'Q_{VV}^{\wedge})(AX + HS) - X = 0 \quad . \quad (15)$$

The mean square error of \hat{X} is then derived from Eq.(2),Eq.(14)and Eq.(15) as follows:

$$\text{Mse}(\hat{X}) = \sigma^2 \text{tr}(Q_{\hat{X}\hat{X}}^{\wedge}) \quad . \quad (16)$$

And by model(12), the LS estimator of X , denoted by $\hat{\hat{X}}$, is

$$\hat{\hat{X}} = N^{-1}A'(I - H_1Q_{S_1S_1}^{\wedge}H_1'Q_{VV}^{\wedge})L \quad , \quad (17)$$

where $Q_{S_1S_1}^{\wedge} = (H_1'Q_{VV}H_1)^{-1}$.

The cofactor matrix and the bias of $\hat{\hat{X}}$ are respectively

$$Q_{\hat{\hat{X}}\hat{\hat{X}}}^{\wedge} = N^{-1} + N^{-1}A'H_1Q_{S_1S_1}^{\wedge}H_1'AN^{-1} \quad (18)$$

and $\text{Bias}(\hat{\hat{X}}) = E(\hat{\hat{X}}) - X = N^{-1}A'(I - H_1Q_{S_1S_1}^{\wedge}H_1'Q_{VV}^{\wedge})(AX + H_1S_1 + H_2S_2) - X$
 $= N^{-1}A'(I - H_1Q_{S_1S_1}^{\wedge}H_1'Q_{VV}^{\wedge})H_2S_2 \quad . \quad (19)$

The mean square error of $\hat{\hat{X}}$ is then as follows:

$$\text{Mse}(\hat{\hat{X}}) = \sigma^2 \text{tr}(Q_{\hat{\hat{X}}\hat{\hat{X}}}^{\wedge}) + \text{tr}((E(\hat{\hat{X}}) - X)(E(\hat{\hat{X}}) - X)') \quad . \quad (20)$$

Now compare the mean square errors of \hat{X} and $\hat{\hat{X}}$ by subtracting one from the other. Then

$$\begin{aligned} \text{Mse}(\hat{X}) - \text{Mse}(\hat{\hat{X}}) &= \text{tr}(N^{-1}A'(I - H_1Q_{S_1S_1}^{\wedge}H_1'Q_{VV}^{\wedge})H_2' \\ &\quad \cdot (\sigma^2 Q_{S_2S_2}^{\wedge} - S_2S_2') \cdot \\ &\quad \cdot H_2'(I - H_1Q_{S_1S_1}^{\wedge}H_1'Q_{VV}^{\wedge})'AN^{-1}) \quad , \end{aligned} \quad (21)$$

in which $Q_{S_2S_2}^{\wedge} = (H_2'Q_{VV}H_2 - H_2'Q_{VV}H_1Q_{S_1S_1}^{\wedge}H_1'Q_{VV}H_2)^{-1}$.

It follows that if

$$\sigma^2 Q_{\hat{S}_2 \hat{S}_2} - S_2 S_2' > 0 \quad (22a)$$

that is, the left side is a positive definite matrix, then there will be

$$Mse(\hat{X}) > Mse(\hat{\hat{X}}) .$$

The inequality (22a) can be proved to be equivalent to the following one:

$$S_2' Q_{\hat{S}_2 \hat{S}_2}^{-1} S_2 / \sigma^2 < 1 . \quad (22b)$$

The above reasoning shows that in the case of the inequality (22) being satisfied the estimator of the main unknowns X by the true model(11) is inferior to that by the submodel(12) in which the parameters S_2 have been rejected. That is just the phenomenon which is referred to as the overparameterization in photogrammetric self-calibration. When it is present, rejecting parameters S_2 will be beneficial. That is to say, the inequality (22) provides us a formulated explanation of the overparameterization and a formulated criterion for the selection of additional parameters.

In addition, it can be discerned from the equation of comparison (21) that the effectiveness of self-calibration is relative to the ratio of the magnitude of additional parameters to the reference variance, in the current case, that is the left side of the inequality (22b), which reflects the "signal to noise ratio". The larger the ratio, the more effective the self-calibration, and less effective otherwise. This property of self-calibration was once derived specially from simulated experiments (Li,1981).

Because the selection of additional parameters has much to do with the effect of self-calibration, this problem has received a widespread attention. But the problem is a rather difficult one. The inequality (22) only provides the criterion for whether a certain group of parameters should be rejected or not. While the optimum group of parameters need to be selected from the complete combination of all the additional parameters so as that the self-calibration may improve the quality of estimation to an extreme extent. Besides, such criteria as the inequality (22) all contain unknowns to be estimated such as S_2 and σ . In practice those unknowns are always replaced by their estimators which come from the observations. That thus makes the selection of parameters stochastic. The less accurate the estimators of those unknowns, the more probable a misleading selection. Frequently, the more important the selection of parameters becomes, the more difficult to procure the optimum parameter group.

At present, the common method for selecting additional parameters in photogrammetry is to make the statistical test and the check of determinability to the additional parameters with their LS estimators obtained from the initial adjustment and thereby to decide whether to accept or to reject the parameters. The statistical test is to test the null hypothesis $H_0: s=0$ (s may be any single additional parameter in S , but the subscripts are omitted here and in the following for all single elements.) using the statistic \hat{T} as follows:

$$\hat{T} = \hat{s} / \hat{\sigma} \sqrt{q}$$

in which \hat{s} and $\hat{\sigma}$ are respectively the LS estimators of s and σ obtained from the initial adjustment using a model with sufficient number of parameters that it is granted as the true model, and q is the diagonal element in $Q_{\hat{S}\hat{S}}$ corresponding to \hat{s} . If the orthogonality among the elements of S is poor, the following F-test is made in addition:

$$\hat{F} = \hat{S}' Q_{\hat{S}\hat{S}}^{-1} \hat{S} / \hat{\sigma}^2 .$$

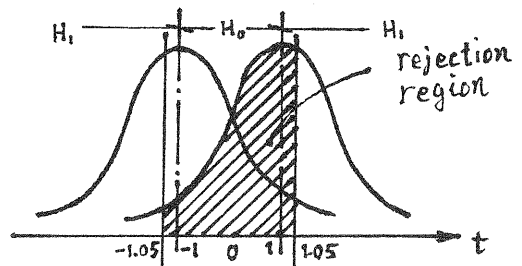
The significance level α for the tests is determined from experiences, which is usually one of 0.1, 0.05, and 0.01 as for the common significance tests. The check of determinability is to check the following measures of internal and external reliability of additional parameters:

$$v_s = \sigma \delta_s \sqrt{q}$$

and $v_f = \delta_f \sigma_f$

(Ackermann, 1981; Förstner, 1980). When the measures are large enough, the parameter s is thought to be non-determinable. Now the principle for parameter selection can be stated as being taking those additional parameters which are significantly different from zero and determinable and deleting those which are nearly zero and non-determinable. Such a method for parameter selection has been reported to be effective in many of the current papers.

However, in the current author's opinion, such a method is questionable in some places. For the sake of prominence of the contradictions, the LS estimators of additional parameters are assumed not to be correlated with each other. And only individual parameters are concerned. First, the significance level was specified to be unreasonably small and so the absolute values of the critical values were too large. From the Eq.(21) it is known that when S_2 contains only one element s , the inequality(22) becomes the sufficient and necessary condition of $Mse(\hat{X}) > Mse(\hat{X})$. That is to say, if $T^2 = s^2/\sigma^2 q < 1$, rejecting s is preferred to accepting it, otherwise, accepting s is preferred. Therefore the test for deciding whether to accept or to reject s should take $H_0: T^2 \leq 1$ as the null hypothesis and $H_1: T^2 > 1$ as the alternate hypothesis. Because mistaken acceptance and mistaken rejection of a parameter are equivalently harmful to the quality. Hence, if there is no priori information about the parameters available, the probability of accepting the false and that of rejecting the true should be specified to be the same for the just mentioned test. Now take it for granted approximately that \hat{T} follows the normal distribution $N(T, 1)$. Accordingly the critical values are ± 1.05 which correspond to $\alpha = \beta = 50\%$, which are respectively the maximum probabilities of rejecting the true and accepting the false. That is to reject s when $|\hat{T}| \leq 1.05$ with the maximum probability of making a mistake being 50%, and to accept s when $|\hat{T}| > 1.05$, again with the maximum probability of making a mistake being 50%. Selecting additional parameters as stated above is thought to be more reasonable by the author. In the past, this sort of test had a much larger absolute critical value than 1.05. That could lead to a mistaken rejection of parameters more probably. The reason for such a large absolute critical value being specified may be that the test using the statistic \hat{T} was carried out referring to the null hypothesis $H_0: s=0$ and a high credit was given to that the null hypothesis is true regardless of which parameter and what case they were, and that the result by the non-self-calibration model without any additional parameters was taken as the reference to which improvement was spoken about and comparison was not made between the two decisions of accepting or rejecting some parameter (or group). Those considerations are not reasonable.



Based on the above analysis, it is suggested that in the case of no priori information available about the additional parameters the test using T statistic for the selection of additional parameters should take ± 1.05 as its critical values and the critical values should not be far from ± 1.05 without sufficient reason. If there is priori information available about the distributions of the additional parameters, the selection of additional parameters should depend on the conditional probabilities of the parameters under the conditions of their estimators having been realized, rejecting the parameters when their conditional probabilities for that the inequality (22) is satisfied exceed 50%, and accepting them otherwise. For example, suppose that

$$s \sim N(s_0, \sigma_s^2) \quad (\sim \text{means following.})$$

which is equivalent to that

$$T = s/\sigma_s\sqrt{q} \sim N(T_0, \sigma_t^2) ,$$

in which $T_0 = s_0/\sigma_s\sqrt{q}$ and $\sigma_t^2 = \sigma_s^2/\sigma_s^2\sqrt{q}$. Since it is known that $\hat{s} \sim N(s, \sigma_q^2)$ for given s , the conditional probability distribution of T for given \hat{s} is then a normal distribution with the conditional expectation and variance as follows:

$$E(T|\hat{s}) = \frac{\sigma_s^2\hat{s}/\sigma_s\sqrt{q} + s_0\sigma_s\sqrt{q}}{\sigma_q^2 + \sigma_s^2} = \frac{\sigma_s^2\hat{T} + s_0\sigma_s\sqrt{q}}{\hat{\sigma}_q^2 + \sigma_s^2}$$

$$\text{Var}(T|\hat{s}) = \frac{\sigma_s^2}{\sigma_q^2 + \sigma_s^2} = \frac{\sigma_s^2}{\hat{\sigma}_q^2 + \sigma_s^2}$$

in which the final expressions can be tolerated when the redundancy is sufficiently large. Thus the conditional probability for that the inequality (22) is satisfied for given s is

$$\text{Pr}(|T| \leq 1 | \hat{s}) = \int_{-1}^1 \frac{1}{\sqrt{2\pi \text{Var}(T|\hat{s})}} \exp\left[-\frac{(t - E(T|\hat{s}))^2}{2\text{Var}(T|\hat{s})}\right] dt = p(\hat{s}) ,$$

which is a function of \hat{s} . Then reject s if \hat{s} makes $p(\hat{s}) \geq 50\%$, and accept s otherwise. When $\sigma_s = \infty$, $p(\pm 1.05\sigma_s\sqrt{q}) = 50\%$ so the same critical values are obtained as in the case of no priori information about the parameter available.

Secondly, the check of determinability had not an objective standard for how large a value when $\nabla_s s$ reaches s is determinable or non-determinable. For such a check to act on the selection of parameters will make the selection of parameters rather arbitrary. It is more questionable that in some procedures parameters were selected only according to the check of determinability. That is because the measures is only concerned with the geometry of the system not the magnitudes of the parameters. Such a case may be present that the non-determinable parameters take so large values that the errors caused by treating them as zero are larger than their errors of determination. In fact, if the statistical test as mentioned above has been carried out, the check of determinability is no longer necessary. If the check of determinability has been removed, the problem whether a non-significant parameter, after its determinability has been assured, must be deleted in all cases (Ackermann, 1981) disappears.

PARAMETER APPROXIMATE CORRELATION CAUSES NUMERICAL INSTABILITY AND POOR ESTIMATES

For a group of vectors a_1, a_2, \dots, a_m and a group of real numbers k_1, k_2, \dots, k_m which are not all equal to zero, if the following equation is true, the group of vectors is defined to have linear correlation in it, and not to have linear correlation in it otherwise.

$$k_1 a_1 + k_2 a_2 + \dots + k_m a_m = 0 \quad (23)$$

If the column vectors in the coefficient matrix A in the linear model (3) have nearly the relation as Eq.(23) describes, the vectors or the elements of X are said to have approximate correlation among them. In regression analysis, that phenomenon is specially named multicollinearity in recent years.

In self-calibration approximate correlation among unknown parameters is unavoidable. Although the approximate correlation among additional parameters can be avoided by proper linear transformation, the relation of the additional parameters as a whole with the main parameters offers no choice, depending on the systematic errors the additional parameters expect to describe. It is easy to see that when the systematic errors are orthogonal to the unknowns, they add

no bias to the LS estimators of the unknowns, self-calibration is then unnecessary, but when the systematic errors become near to correlation with the unknowns, the biases in the non-self-calibration result increase, self-calibration is now required. However, at this moment, the additional parameters that describe the systematic errors correlated approximately with the main unknowns also have approximate correlation with the main unknowns.

The approximate correlation among unknown parameters has a harmful effect on their LS estimators in two aspects. On one hand, amplifies the influence of various small disturbances on the solution and makes the LS estimation difficult to realize. When there is approximate correlation among the column vectors of A, the coefficient matrix of the normal equations $N=A'A$ will have very small eigenvalues and so always cause a large condition number. From the relation for estimating errors in computation

$$\frac{\|\delta X\|}{\|X\|} \leq \frac{K(N)}{1 - K(N)\|\delta N\|/\|N\|} \left(\frac{\|\delta N\|}{\|N\|} + \frac{\|\delta B\|}{\|B\|} \right)$$

in which B is the constant vector of the normal equations, it is seen that a large condition number $K(N)$ can make the solution of the normal equations very sensitive to the small disturbances N and B in the coefficient matrix and the constant vector. In the case of a nonlinear model, a large condition number of the coefficient matrix of the linearized normal equations can make the iterative process of solution diverge or converge to a misleading solution, so that the nonlinear LS estimation cannot be realized. On the other hand, the approximate correlation involving main unknowns can inflate the variances of the LS estimators of the main unknowns, so depress the theoretical quality of the result. As is known from the preceding sections, the sum of variances of the LS estimators of unknowns is

$$\sum_i \text{Var}(\hat{X}_i) = \sigma^2 \text{tr}(Q_{\hat{X}\hat{X}}) = \sigma^2 \text{tr}(N^{-1}) = \sigma^2 \sum_i 1/\lambda_i$$

in which λ is an eigenvalue of N. Hence, when approximate correlation is present, the very small eigenvalues of the coefficient matrix of normal equations will enlarge the sum of variances of the estimators greatly. That is to say, the variances of the observations are transmitted to the variances of the estimators in a very large amplification in this case. So the result of estimation is very poor in quality even if the process of solution has been completed smoothly and has introduced no computational errors.

The treatments of the two different aspects of harmful effect caused by approximate correlation have different requirements and methods. The methods of treating the computational problems caused by approximate correlation are to increase the word length in computation and to adopt those algorithms of high numerical stability for solving the normal equations. If a nonlinear model is dealt with, various improved Gauss-Newton methods, like the Marquardt method or damped least squares method in nonlinear programming can be adopted to weaken the constraint to convergence and ensure the realization of LS estimation. The Marquardt method has been proved by simulated experiments in close-range photogrammetry to be effective in overcoming the difficulty in numerical solution of LS estimation. The table below shows the numbers of iterations and the LS estimators of the reference variances by the Marquardt method in the cases in which the Gauss-Newton method breaks down.

case	1	2	3	4	5	6	7	8	9	10	11	12
σ	22.0	22.0	3.0	22.0	3.0	22.0	12.0	22.0	22.0	3.0	9.0	22.0
$\hat{\sigma}$	24.3	25.5	3.3	23.8	3.4	24.5	12.6	25.4	25.7	3.2	10.4	24.0
iter.numb.	6	7	8	10	11	12	13	14	15	18	24	35

As to the degradation of the quality of estimation due to approximate correlation, improving the estimating method itself is needed to overcome it. That is not easy

as a rule. The errors of observations are always much more considerable than the disturbances in computation, hence the requirement of the condition of normal equations for ensuring a good theoretical quality of estimation is generally harder than that for ensuring a stable computation. In general, if computational instability is encountered, then even though the LS estimators have been procured by some way, the estimators are still very poor in quality. On the contrary, when the quality of estimators has been improved by refining the estimating method, the instability of computation will vanish accordingly. Therefore, in the presence of approximate correlation, improving the estimating method is most important. Of course, if approximate correlation appears only among the additional parameters, not involving the main unknowns, then only considering the computational problems may be enough.

At present, the common method of treating approximate correlation in photogrammetry is to treat the additional parameters which cause approximate correlation as fictitious observations with certain weights rather than as free unknowns. As a rule, the observations are designated to be zero, and the weights to be small values or such values that their corresponding standard deviations are equal to the possible magnitudes the parameters themselves might have, or to be determined iteratively. Such a method has been proved to be effective by some practical data reductions. In fact, the methods as such no longer lead to the LS estimators as originally. It has been shown in the other paper of the author's (Huang, 1986) that such methods essentially conform to the principle of biased estimation in regression analysis. Therefore they can not only overcome the numerical instability caused by approximate correlation, but also improve the quality of estimators, producing rather satisfactory results. However, it is seen in some literature that not all people doing so could realize those points.

Several years ago the methods named "fixed segmentation" and "sequential segmentation" were proposed for treating approximate correlation in place of those methods with fictitious weighted observations of additional parameters so as to avoid the time-wasting computation of matrix inversion. In those segmentation methods, all unknown parameters are segmented into several groups so that parameters in the same group have no approximate correlation with each other. Only one group of parameters are carried as unknowns in each solution of an iterative process while the others are kept as their current values. The group of parameters as unknowns is alternated for each successive iteration step. This iterative process is conducted until proper convergence is reached. Such segmentation methods can save a large amount of computation indeed. They have been incorporated into some computation programs and proved to be effective as reported in some papers. However, in the paper in which the segmentation methods were proposed, there was no theoretical basis for the methods accounted for and no comparison given between the methods with those treating parameters as fictitious weighted observations. So some questions may be put to the segmentation methods.

The current author believes that the segmentation methods cannot completely play the role of the methods treating the unknowns as fictitious weighted observations. The segmentation methods are similar to the alternating coordinates method in nonlinear programming except that the former alternates coordinates group by group while the latter one by one. They can only surmount the numerical instability caused by approximate correlation and prevent the LS solution from breaking down as various improved Gauss-Newton methods, but cannot improve the quality of estimators as the method treating unknowns as fictitious weighted observations. Because approximate correlation not only causes numerical instability but also lowers the quality of estimators. Therefore, it seems impossible that utilizing the segmentation methods in photo-variant self-calibration (Moriwa, 1980) can treat the approximate correlation due to the increase of the number of parameters satisfactorily and make the result of the photo-variant self-calibration superior to that of the block-invariant one in all cases. So far, no method with less computation has been found to possess the complete function of the method treating unknowns as fictitious weighted observations. For this reason, it is necessary to research on simplifying the

computation in the method.

REFERENCES

- Ackermann, F. (1981), "Block Adjustment with Additional Parameters", *Photogrammetria*, 36, 217-227.
- Förstner, W. (1980), "Zur Prüfung Zusätzlicher Parameter in Ausgleichungen", *ZFV*, 11, 510-519.
- Grün, A. (1978), "Progress in Photogrammetric Point Determination by Compensation of Systematic Errors and Detection of Gross Errors", *Nachr.ad.Karten-u. Verm.wesen*, Reihe II, Heft 36, Frankfurt a.M. 1978.
- Huang, Y.D. (1986), "Application of Biased Estimation to Close-Range Photogrammetric Data Reduction", *ISPRS, Comm.V, Symposium Ottawa*.
- Li, D. (1981), "On the Signal-to-noise Ratio in Self-calibrating Block Adjustment", *Acta Geodetica et Cartographica Sinica*, 11, 170-184.
- Marquardt, D.W. (1963), "An Algorithm for Least-squares Estimation of Nonlinear Parameters" *J.Soc.Indust.Appl.Math.*, 11, 431-441.
- Mikhail, E.M. (1976), "Observations and Least Squares", IEP-A Dun-donnelley Publisher, New York.
- Moniwa, H. (1980), "The Concept of 'Photo-variant' Self-calibration and Its Application in Block Adjustment with Bundles", *ISP, Comm.V/1*, Hamburg.
- Murai, Shunji, (1984), "A Study on Analytical Calibration for Non-metric Camera and Accuracy of Three Dimensional Measurement", *ISP, Comm.V*, Rio.
- Schut, G.H. (1979), "Selection of Additional Parameters for the Bundle Adjustment", *Photogrammetric Engineering and Remote Sensing*. 45,9.