Peter J.G. TEUNISSEN
Faculty of Geodesy
Delft University of Technology
Delft (The Netherlands)

FIRST AND SECOND MOMENTS OF NON–LINEAR LEAST–SQUARES ESTIMATORS

Abstract

The first two moments of non-linear and least-squares estimators are studied. Approximate expressions for the moments are derived and discussed. The results are compared with those of Jeudy [*1988*].

1. Introduction

Despite the fact that almost all functional relations in our geodetic models are non-linear, we predominantly use the ideas, concepts and results from the theory of linear estimation in our geodetic adjustment. In order to be able to justify such a linearized approach, we need to have ways of assessing the amount of non-linearity and methods to prove whether a linear(ized) model is a sufficient approximation. We, therefore, need to know how non-linearity manifests itself at the various stages of an adjustment.

In this paper, we will restrict ourselves to least-squares (LS) estimators and only study the effect of non-linearity on their first two moments.

Although we emphasize a diagnostic point of view, most of the objectives of this paper are essentially the same as those of the interesting paper [Jeudy, 1988]. We will compare our results – parts of which are taken from Teunissen [1984, 1985] and Teunissen and Knickmeyer [1988] – with those of [ibid]. We will also try to clarify some points which were raised in [ibid].

2. First moment of a non-linear estimator

One of the commonest problems in estimation theory is, given a series of random variables $\underline{y}^i$, $i = 1, \ldots, m$ (the underscore in y^i typifies randomness), to find a set of functions of these, $x^\alpha(y^i)$, $\alpha = 1, \ldots, n$, which should provide an estimator, $\hat{\underline{x}}^\alpha = x^\alpha(\underline{y}^i)$, of an unknown set of parameters x^α (in this paper, Greek indices range from 1 to n, while Roman indices range from 1 to m). There exist many general considerations for the development of "good" estimators [see e.g. Mood et al., 1974]. One of these is unbiasedness (although in some cases it may be useful to sacrifice the unbiasedness condition in favour of other, more desirable, properties; e.g.

Bull. Géod. 63 (1989) pp. 253–262.

a better precision). An estimator $\underline{\hat{x}}^a$ of x^a is said to be unbiased if $E\left\{\underline{\hat{x}}^a\right\} = x^a$, where $E\left\{.\right\}$ stands for the mathematical expectation operator. The difference $E\left\{\underline{\hat{x}}^a\right\} - x^a$ is called the bias of $\underline{\hat{x}}^a$ and it will be denoted by $b^a_{\hat{x}}$.

In general, we can hardly expect non-linear estimators to be unbiased, since if $x^a(.)$ is non-linear, then $E\left\{x^a(\underline{y}^i)\right\} \neq x^a(E\left\{\underline{y}^i\right\})$, i.e. the mean of the image differs generally from the image of the mean. If the non-linear functions $x^a(.)$ admit a Taylor expansion, it is not difficult to derive an expression which approximates the bias $b^a_{\hat{x}}$. The usefulness of such an expression is threefold. Firstly, such an expression will give us valuable insight in the factors contributing to the bias $b^a_{\hat{x}}$. Secondly, the expression may show us the way to derive practical measures for testing the significance of the bias. And, finally, if the bias turns out to be significant, the expression might be used to correct the estimator $\underline{\hat{x}}^a$ for biasedness.

Let us assume that the unknown set of parameters x^a which are to be estimated, satisfy $x^a = x^a(y^i)$, and that the random variables $\underline{y}^i$ have a mean $E\left\{\underline{y}^i\right\}$ and covariance matrix $\sigma^2 g^{ij}$, where σ^2 stands for the variance factor of unit weight. We also assume that the bias $b^i_y = E\left\{\underline{y}^i\right\} - y^i$ of $\underline{y}^i$ as an estimator of y^i, is not necessarily equal to zero. A Taylor expansion of $\underline{\hat{x}}^a = x^a(\underline{y}^i)$ at y^i gives then :

$$\underline{\hat{x}}^a = x^a + \partial_i x^a(\underline{y}^i - y^i) + {}^1\!/_2\, \partial^2_{ij} x^a(\underline{y}^i - y^i)(\underline{y}^j - y^j) \quad \ldots, \tag{1}$$

where $\partial_i x^a$ and $\partial^2_{ij} x^a$ stand for the first and second order partial derivatives of $x^a(.)$ evaluated at y^i respectively, and Einstein's summation convention is used for repeated indices. Upon taking the expectation of (1) we get :

$$\boxed{b^a_{\hat{x}} \doteq \partial_i x^a b^i_y + {}^1\!/_2\, \partial^2_{ij} x^a(\sigma^2 g^{ij} + b^i_y b^j_y)} \tag{2}$$

Note that $\partial^2_{ij} x^a g^{ij}$ in (2) stands for a weighted trace. It reduces to the ordinary trace of $\partial^2_{ij} x^a$, if g^{ij} equals the identity matrix.

Before we make some remarks on (2), let us first see this formula at work on an example. We have chosen the following illustrative, but simple and artificial example :

Let the parameter x, which is to be estimated, be given by $x = y^i \delta_{ij} y^j$, $i, j = 1, \ldots, m$, where $y^i = y\ \forall_i$ and thus $x = my^2$ (δ_{ij} stands for the Kronecker delta : $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$). The partial derivatives evaluated at y^i read then $\partial_i x = 2y^i = 2y$, $\forall_i$ and $\partial^2_{ij} x = 2\delta_{ij}$. Let $\underline{y}^i$ be normally distributed as $\underline{y}^i \sim N(0, \sigma^2 \delta^{ij})$. Then $b^i_y = -y$, $\forall_i$. Hence, with the aid of (2) we find that $b_{\hat{x}} = -2my^2 + {}^1\!/_2 \cdot 2 \cdot \delta_{ij}(\sigma^2 + y^2)\delta^{ij} = m(\sigma^2 - y^2)$. This result is easily verified, since clearly the estimator $\underline{\hat{x}} = \underline{y}^i \delta_{ij} \underline{y}^j$ has a $\sigma^2 \chi^2$-distribution with m degrees of freedom and thus a mean of $E\left\{\underline{\hat{x}}\right\} = m\sigma^2$. Hence we find again that

$E\left\{\hat{x}\right\} - x = m\sigma^2 - my^2$.

With respect to formula (2), the following remarks are in order :

1. For formula (2) to be useful, we need to know the functions $x^a(.)$, or at least their partial derivatives of lower order. In a non-linear least-squares context this is a non-trivial matter. We will come back to this in section 3

2. If the functions $x^a(.)$ are linear, then $\partial_i x^a$ is independent of y^i and second and higher order partial derivatives vanish. Formula (2) reduces then to

$$b^a_{\hat{x}} = \partial_i x^a \, b^i_y \tag{3}$$

Here we recognize the well-known bias propagation formula, which is used extensively in linear estimation theory and in reliability theory. A useful upperbound on the absolute values of the elements of the bias vector (3) of $\underline{\hat{x}}^a$ can be obtained through the use of Cauchy-Schwarz'inequality :

$$|b^a_{\hat{x}}| \leq \sigma_{\hat{x}^a} \| b_y \| , \tag{3'}$$

where $\sigma^2_{\hat{x}^a}$ stands for the variance of $\underline{\hat{x}}^a$ and $\| b_y \|^2 = \sigma^{-2} b^i_y g_{ij} b^j_y$, with $g^{ij} g_{jk} = \delta^i_k$. A similar bound will be derived in section 3 for the bias in the non-linear least-squares estimator.

3. If the functions $x^a(.)$ are quadratic, then $\partial^2_{ij} x^a$ is independent of y^i and formula (2) holds exactly (see example above).

4. If $\underline{y}^i$ is an unbiased estimator of y^i, i.e. $b^i_y = 0$, then (2) reduces to :

$$b^a_{\hat{x}} \doteq \frac{1}{2} \sigma^2 \partial^2_{ij} x^a g^{ij} \tag{4}$$

This is formula (13), with (14), of [Jeudy, 1988] (the difference in sign between the two formulae is due to a different convention in the bias definition). The approximation (4) neglects terms of the order σ^3 and higher.

5. Formula (4) becomes an improved approximation of the bias in $\underline{\hat{x}}^a$ if $\underline{y}^i$ is an unbiased estimator of y^i and $\underline{y}^i$ is normally distributed. In this case terms of the order σ^4 and higher are neglected, due to the fact that the third order central moments vanish. Hence, if the functions $x^a(.)$ are cubic in this case, then (4) holds exactly.

6. If $\underline{y}^i$ is a non-linear least-squares estimator of y^i, then (see next section) b^i_y is of the order σ^2. Hence, by retaining quantities of the order σ^2 formula (2) can be simplified to :

$$b^a_{\hat{x}} \doteq \partial_i x^a b^i_y + \frac{1}{2} \sigma^2 \partial^2_{ij} x^a g^{ij} \tag{5}$$

7. Once the bias in $\underline{\hat{x}}^a$ is known, one can derive, using exactly the same Taylor

expansion technique as was used above, an approximation to the covariance matrix of $\hat{\underline{x}}^a$. For the case that $\underline{y}^i$ is an unbiased estimator of y^i, this will give formula (19) of [ibid]. Note that if $\underline{y}^i$ is normally distributed, this formula simplifies because of the terms of order σ^3.

8. In [Jeudy, 1988], under the heading "practical considerations", the interesting point is raised concerning the evaluation of the partial derivatives of $x^a(.)$. In (2), as in formulae (13), (19) and (21) of [ibid], the partial derivatives need to be evaluated at y^i, which is unknown in general. In practice one will therefore, if $\underline{y}^i$ is an unbiased estimator of y^i, evaluate the partial derivatives at the known sample value of $\underline{y}^i$. And, as is correctly pointed out in [ibid], this will introduce another approximation in the computation of the bias.

In [Jeudy, 1988] a numerical example is given with the purpose of testing the adequacy of his formulae (13) (our formula (4)) and (21) for the bias and covariance matrix respectively. Unfortunately the author has missed here the opportunity to investigate for his particular example, the effect of evaluating the partial derivatives at a sample value instead of evaluating at y^i. For the example given in [ibid], the closed form expressions for the bias and variance respectively, are namely easily derived. They read in the notation of [ibid] :

$$\hat{X}(L) - E\left\{\hat{X}(\tilde{L})\right\} = \exp(-\lambda L^2) - E\left\{\exp(-\lambda\tilde{L}^2)\right\} \quad \text{and}$$

$$E\left\{(\hat{X}(\tilde{L}) - E\left\{\hat{X}(\tilde{L})\right\})^2\right\} = E\left\{\exp(-2\lambda\tilde{L}^2)\right\} - E\left\{\exp(-\lambda\tilde{L}^2)\right\}^2,$$

$$\text{with } E\left\{\exp(-\lambda\tilde{L}^2)\right\} = (2\lambda\sigma^2+1)^{-1/2}\cdot\exp(-\lambda(2\lambda\sigma^2+1)^{-1}L^2).$$

For the bias formula (2) we belief that the problem of having to evaluate the partial derivatives at a sample value instead of at y^i, should not be of too much concern to us. First of all, one can always, from a diagnostic point of view, evaluate the partial derivatives at all likely values of y^i, in order to infer whether non-linearity is likely to produce a significant bias or not. In some practical important cases one can even derive a closed formula for the bias, and thus study the behaviour of bias under various circumstances (see e.g. [Teunissen, 1985] or [Teunissen and Knickmeyer, 1988] for the bias in the scale estimator of the Symmetric Helmert transformation). Secondly, if the partial derivatives of (2) are evaluated at a sample value of $\underline{y}^i$ and if $\underline{y}^i$ is either an unbiased estimator of y^i or a non-linear least-squares estimator of y^i, one can show with the same Taylor expansion technique of above, but now applied to the partial derivatives of $x^a(.)$, that the additional error introduced will be of the same order, on the average, as the order of the terms one is already neglecting in (2).

For the covariance formulae (19) and (21) of [Jeudy, 1988] the situation is more problematic. In this case the additional error introduced is of the same order, on the average, as the terms which are included in (19) and (21). If the need arises, i.e. if higher order terms turn out to be significant, it is therefore doubtful whether (19) and (21) of [ibid] can be used as improved formulae for the covariance matrix of $\hat{\underline{x}}^a$ in case evaluation is done at a sample value of $\underline{y}^i$. Nevertheless, if the first, second and third order partial derivatives are not too difficult to compute, these formulae can still

be used as diagnostic tools for inferring the effect of non-linearity. We will come back to this point in section 4 .

3. First Moment of a Non-linear LS–Estimator

Consider the non-linear model

$$\underline{y}^i = A^i(x^a) + \underline{e}^i \,,\ \sigma^2 g^{ij} \ ,\ i,j = 1,\ldots,m\,;\ a = 1,\ldots,n \ , \tag{6}$$

where $\underline{e}^i$ is random with zero mean and covariance matrix $\sigma^2 g^{ij}$. Then $\underline{y}^i$, the observables, are random with mean $A^i(x^a)$ and covariance matrix $\sigma^2 g^{ij}$. Since we will assume that $\underline{y}^i$ is normally distributed, we have, through the method of maximum likelihood (ML), a probabilistic justification for using the LS-criterium in finding the LS-estimator $\hat{\underline{x}}^a = x^a(\underline{y}^i)$ of the unknown but fixed set of parameters x^a. Again the estimator will be biased in general. From the invariance property of ML follows namely that if $\hat{\underline{x}}^a$ is the ML-estimator of x^a than for any arbitrary regular enough function $F(.)$, $F(\hat{\underline{x}}^a)$ is the ML-estimator of $F(x^a)$, and since $E\left\{F(\hat{\underline{x}}^a)\right\} \neq F(E\left\{\hat{\underline{x}}^a\right\})$ it will be the exception rather than the rule for a ML-estimator to be unbiased.

In order to be able to compute the bias in the LS-estimator $\hat{\underline{x}}^a$ of x^a, we need to know, as before the functions $x^a(.)$ or at least their second order partial derivatives evaluated at $E\left\{\underline{y}^i\right\}$. Since we do not have, in general, closed expressions available which express the estimators as known functions of the observables, we again have to take recourse to Taylor expansions. The method we will use was also used in [Teunissen, 1985], and can be described as follows :

First assume that the LS-estimator $\hat{\underline{x}}^a = x^a(\underline{y}^i)$ admits a Taylor expansion at $E\left\{\underline{y}^i\right\}$. Taylor expansion at $E\left\{\underline{y}^i\right\}$ gives an expansion in $\underline{e}^i$. The problem is now to find the coefficients of this expansion. Once these coefficients are known, we can apply formula (4) of the previous section.

The coefficients are found in the following way. We start from the orthogonality conditions $0 = \partial_a A^i(\hat{\underline{x}})\, g_{ij}\,(\underline{y}^j - A^j(\hat{\underline{x}}))$. Expansion of the right hand side of the orthogonality conditions at x^a gives an expansion in $\hat{\underline{x}}^a - x^a$. We now substitute our first expansion in $\underline{e}^i$, in the above expansion in $\hat{\underline{x}}^a - x^a$. The result is a new expansion in $\underline{e}^i$, which is identical to zero for all $\underline{e}^i$. Hence we may collect terms of the same order and set them to zero. After taking the expectations of these terms and making use of the central moments of $\underline{e}^i$, e.g.

$$E\left\{\underline{e}^i\right\} = 0 \qquad E\left\{\underline{e}^i\,\underline{e}^j\,\underline{e}^k\right\} = 0 \tag{7}$$

$$E\left\{\underline{e}^i\,\underline{e}^j\right\} = \sigma^2 g^{ij} \qquad E\left\{\underline{e}^i\,\underline{e}^j\,\underline{e}^k\,\underline{e}^\ell\right\} = \sigma^2\,(g^{ij}g^{k\ell} + g^{ik}g^{j\ell} + g^{i\ell}g^{jk}),$$

we can recursively determine all the coefficients sought. The higher order central moments of $\underline{e}^i$ follow from using e.g. the characteristic functions approach.

In a similar way one can find the bias in the LS-residual

$\underline{\hat{e}}^i = \underline{\hat{y}}^i - A^i(\underline{\hat{x}}^\alpha)$ or in $\underline{\hat{y}}^i = A^i(\underline{\hat{x}}^\alpha)$.

Our results read :

$$\boxed{\begin{array}{ll} \text{a)} & b^\alpha_{\hat{x}} \doteq \sigma^2 g^{\alpha\beta} \partial_\beta A^i g_{ij} b^j \\ \text{b)} & b^i_e \doteq \sigma^2 [\delta^i_j - \partial_\alpha A^i g^{\alpha\beta} \partial_\beta A^k g_{kj}] b^j \\ \text{with :} & b^j = -\frac{1}{2} \partial^2_{\alpha\beta} A^j g^{\alpha\beta} \end{array}} \qquad (8)$$

All partial derivatives in (8) are evaluated at x^α . The matrix $\sigma^2 g^{\alpha\beta}$ is to a first order the covariance matrix of $\underline{\hat{x}}^\alpha$ (see section 4) and $g^{\alpha\beta} \partial_\beta A^i g_{ij} \partial_\gamma A^j = \delta^\alpha_\gamma$.

The following remarks about (8) are in order :

1. In (8) terms of the order σ^4 and higher are neglected. If $\underline{y}^i$ is not normally distributed, then in general only terms of the order σ^3 and higher are neglected.

2. Using results from differential geometry, the right hand sides of (8) can be given simple geometric interpretations (see [Teunissen, 1985] for more details).

3. The above approximations of the biases were derived using the set-up of observation equations. For the dual set-up of adjustment by condition equations similar results can be given (see [Teunissen and Knickmeyer, 1988]) .

4. From (8) follows that the bias in the parameters is zero if either the model is linear, i.e. $b^j = 0$, or if b^j is orthogonal to the rangespace of $\partial_\alpha A^i$. The bias in the residuals is zero if either the model is linear, b^j is orthogonal to the orthogonal complement of the rangespace of $\partial_\alpha A^i$, or if $m = n$, i.e. if there is no redundancy.

5. Expression (8) shows that, roughly speaking, the following factors effect the bias : a priori precision $(\sigma^2 g^{ij})$, the design $(\partial_\alpha A^i)$ and amount of non-linearity (b^j) . This indicates that in most geodetic models it is not so much the absence of severe non-linearity which makes bias negligible, but the relatively high precision of the measurements (e.g. $\sigma = 10^{-5}$ or better). For the same reason (see section 4) the inverse of the normal matrix of linearized least-squares suffices for most applications as an approximation to the covariance matrix of the LS-estimator.

6. It is interesting to observe that the biases in the LS-estimators $\underline{\hat{x}}^\alpha$ and $\underline{\hat{e}}^i$ are computed from b^j , just like in the linear LS-case the estimators $\underline{\hat{x}}^\alpha$ and $\underline{\hat{e}}^i$ themselves are computed from $\underline{y}^i$. Hence with an available LS-software package the evaluation of the biases becomes rather simple.

7. Apart from (8), it is useful to have global scalar bias-measures available which summarize the bias present in the non-linear model. In order to discern the significance of the biases it was proposed in [Teunissen and Knickmeyer, 1988] to weight the biases

in the parameters and residuals with the inverse of $\sigma^2 g^{\alpha\beta}$ and $\sigma^2 g^{ij}$ respectively. With $\| \cdot \|_1^2 = \sigma^{-2} (.)^\alpha g_{\alpha\beta} (.)^\beta$, $\| \cdot \|_2^2 = \sigma^{-2} (.)^i g_{ij} (.)^j$, it follows from (8) that

$$\| b_{\hat{x}} \|_1 \doteq \sigma \| P_A b \|_2 , \quad \| b_{\hat{e}} \|_2 \doteq \sigma \| P_A^\perp b \|_2 , \tag{9}$$

Where P_A and $P_A^\perp$ are orthogonal projectors, projecting onto the rangespace of $\partial_\alpha A^i$ and its orthogonal complement respectively. Using "Pythagoras" the following approximate upperbounds are obtained from (9) :

$$\| b_{\hat{x}} \|_1 \leqq \| \sigma b \|_2 , \quad \| b_{\hat{e}} \|_2 \leqq \| \sigma b \|_2 . \tag{9'}$$

8. Alternative local upperbounds for the biases can be obtained from (8) using Cauchy-Schwarz' inequality :

$$| b^\alpha_{\hat{x}} | \leqq \sigma_{\hat{x}} \alpha \| \sigma b \|_2 , \quad | b^i_{\hat{e}} | \leqq \sigma_{\hat{e}} i \| \sigma b \|_2 , \tag{10}$$

where $\sigma^2_{\hat{x}} \alpha$ and $\sigma^2_{\hat{e}} i$ are, to a first order, the variances of $\hat{x}^\alpha$ and $\hat{e}^i$ respectively. Compare (10) with (3') of section 2.

9. In [Jeudy, 1988] the steps to be taken for deriving the second order partial derivatives, which are needed for the LS-bias, are outlined. Together with his formula (13) (our formula (4)) this then gives his approximation of the bias in the LS-estimator $\hat{x}^\alpha$. In order to be able to compare our result (8a) with the one of [ibid], we have to compare the different ways the partial derivatives of the functions $x^\alpha (.)$ are derived and evaluated. Above and in section 2, we have worked with the second order partial derivatives evaluated at $E \{ \underline{y}^i \}$. Also in the first part of [ibid] the bias formula is given with the second order partial derivatives (his H-array) evaluated at $E \{ \underline{y}^i \}$. In the second part of [ibid] however, the author unfortunately does not, in his sequence of steps for obtaining the H-array, evaluate at $E \{ \underline{y}^i \}$. Thus in contrast to the first part of [ibid], the derivations in the second part of [ibid] are all done with an evaluation at the sample point. Note that there is a difference whether one starts of by deriving the partial derivatives all evaluated at $E \{ \underline{y}^i \}$ and then in the final result evaluates at the sample point, or whether one starts evaluating from the beginning at the sample point.

The consequences of the approach taken in [ibid] are in our opinion somewhat unfortunate. First of all this method gives a rather complicated structure for the H-array. This may be the reason why the author does not explicitly state the final form of his H-array. In any case, when compared to (8a), the result seems to be too complicated to be practical useful. Secondly, the H-array so obtained is very difficult to interpret. One is almost certainly not able to give a simple geometric interpretation, as was given to (8) in [Teunissen, 1985]. Also the LS-interpretation, which was given above to (8), is then not possible, and the simple bounds (9') and (10) will be hard to get by. Finally, it turns out that with the approach of [ibid] one needs for the evaluation of H up to third order partial derivatives of the observation equations. In (8) we only need the first and second order partial derivatives.

4. On the Second Moment of a Non-linear LS-Estimator

In section 2 we already expressed our doubt whether in general the covariance formula (19) of [Jeudy, 1988] can be considered an improved formula when evaluated at the sample point. The sample point evaluation introduces namely additional errors that are on the average of the same order as the terms already included in the formula. Nevertheless, from a diagnostic point of view, such formulae can still be useful, e.g. to infer and study the effects of non-linearity.

As to the formulae of [ibid] for approximating the covariance matrix of the LS-estimator, our comments of section 3 concerning complexity and interpretation also apply here.

Our derivation of an approximation to the covariance matrix of the LS-estimator $\hat{\underline{x}}^{\alpha}$ of x^{α}, follows along the same lines as our derivation in section 3. We will denote the covariance matrix by $\sigma^2 G^{\alpha\beta}$.
With the definitions $\Gamma^{\alpha}_{\gamma\delta} \triangleq \partial^2_{\gamma\delta} A^i g_{ij} \partial_{\beta} A^j g^{\beta\alpha}$, $B^i_{\gamma\delta} \triangleq \partial^2_{\gamma\delta} A^i - \Gamma^{\alpha}_{\gamma\delta} \partial_{\alpha} A^i$, and by taking only terms up to the order σ^4 into account, our result reads :

$$\boxed{\begin{aligned} &\sigma^2 G^{\alpha\beta} \doteq \sigma^2 g^{\alpha\beta} + \Delta^{\alpha\beta} \text{ , with} \\ &\Delta^{\alpha\beta} = \sigma^2 g^{\alpha\gamma} \partial_{\gamma} b^{\beta}_{\hat{x}} + \sigma^2 \partial_{\gamma} b^{\alpha}_{\hat{x}} g^{\gamma\beta} + \frac{1}{2}\sigma^4 \Gamma^{\alpha}_{\gamma\delta} g^{\gamma\epsilon} g^{\delta\rho} \Gamma^{\beta}_{\epsilon\rho} \\ &\qquad + \sigma^4 g^{\alpha\gamma} B^i_{\gamma\delta} g_{ij} g^{\delta\rho} B^j_{\rho\tau} g^{\tau\beta} \quad . \end{aligned}} \tag{11}$$

All the partial derivatives in the above formula are evaluated at x^{α}.

1. Although expression (11) is of a rather complicated form, some structure can be discerned. First note that the first order term $\sigma^2 g^{\alpha\beta}$ is of the form of the inverse of the normal matrix of linearized least-squares. This was to be expected. The additional term $\Delta^{\alpha\beta}$ is of the order σ^4. The first two terms of $\Delta^{\alpha\beta}$ depend on the rate of change of the bias; the third term can be shown to describe the non-linear behaviour of the x^{α}-parameter curves in the manifold described by $E\{\underline{y}^i\} = A^i(x^{\alpha})$ (the $\Gamma^{\alpha}_{\gamma\delta}$ – array is the Christoffel symbol of the second kind, see [Teunissen, 1985]); and the fourth term can be shown to describe the curvature of this manifold ($B^i_{\alpha\delta}$ is the multi-dimensional generalization of the second fundamental tensor known from classical surface theory). Hence, relative to $g^{\alpha\beta}$, this fourth term of $\Delta^{\alpha\beta}$ is invariant against reparametrizations and therefore gives a limit in the reduction one can achieve through reparametrizations.

2. Although one could use (11) to study the effect of non-linearity, a somewhat simpler expression is obtained if we study the average of $\Delta^{\alpha\beta}$ relative to $\sigma^2 g^{\alpha\beta}$. We therefore take the trace of $\sigma^{-2} g_{\gamma\alpha} \Delta^{\alpha\beta}$ and get

$$\sigma^{-2} g_{\alpha\beta} \Delta^{\alpha\beta} \doteq 2 \partial_\gamma b_{\hat{x}}^\gamma + \sigma^2 (\partial^2_{\alpha\beta} A^i) g_{ij} g^{\alpha\gamma} g^{\beta\delta} (\partial^2_{\gamma\delta} A^j)$$

$$- \frac{1}{2} \sigma^2 (\Gamma^\epsilon_{\alpha\beta} \partial_\epsilon A^i) g_{ij} g^{\alpha\gamma} g^{\beta\delta} (\Gamma^\rho_{\gamma\delta} \partial_\rho A^j) .$$

Since the last term is non-positive we have the approximate bound :

$$g_{\alpha\beta} (G^{\alpha\beta} - g^{\alpha\beta}) \leq 2 \partial_\gamma b_{\hat{x}}^\gamma + \sigma^2 (\partial^2_{\alpha\beta} A^i) g_{ij} g^{\alpha\gamma} g^{\beta\delta} (\partial^2_{\gamma\delta} A^j) \tag{12}$$

3. A lower bound for (12) can be obtained via the celebrated Cramer-Rao bound (see e.g. [Mood et al., 1974]). In our case :

$$\sigma^2 G^{\alpha\beta} \geq (\delta^\alpha_\gamma + \partial_\gamma b^\alpha_{\hat{x}}) \sigma^2 g^{\gamma\delta} (\delta^\beta_\delta + \partial_\delta b^\beta_{\hat{x}}) - b^\alpha_{\hat{x}} b^\beta_{\hat{x}} ,$$

where the inequality should be read as a matrix inequality. Hence, neglecting the term of order σ^6 gives approximately :

$$g_{\alpha\beta} (G^{\alpha\beta} - g^{\alpha\beta}) \geq 2 \partial_\gamma b_{\hat{x}}^\gamma - \| b_{\hat{x}} \|_1^2 . \tag{13}$$

The effect of non-linearity can now be considered insignificant if one over n times the bounds are small compared to one.

5. Conclusions

In this paper we have discussed the effect of non-linearity on the first two moments of non-linear and least-squares estimators. Approximate expressions for the biases in non-linear and least-squares estimators were given. The approximations neglect terms of the order σ^3 or higher. The same approximations also hold on the average if the formulae are evaluated at the sample point instead of at the mean.

The structure of the bias formula for LS-estimators is such that it fits within the general framework of linearized LS-estimation and is therefore not difficult to compute in practice.

The factors that effect the bias are : a priori precision, the design and amount of non-linearity. Since the bias is of the order σ^2, the relatively high precision of geodetic measurements often makes bias negligible, despite strong non-linearity in the functional model.

For the same reason the inverse of the normal matrix of linearized least-squares suffices for most applications as an approximation to the covariance matrix of the LS-estimator.

An approximation to the order σ^4 of the covariance matrix of the LS-estimator was given and discussed. The result is useful as a diagnostic tool to infer the effect of non-linearity. When the intention is to use the covariance formula as an improvement over the customarily used approximation, one should keep in mind that a sample point evaluation introduces additional approximations which on the average are of the same order as the terms included in the covariance formula.

Acknowledgement

This work was funded by the Netherlands organization for scientific research (N.W.O.).

The hospitality of the Department of Surveying Engineering of the University of Calgary is greatly appreciated.

o

o o

REFERENCES

L.M.A. JEUDY (1988) : Generalized Variance-covariance Propagation Law Formulae and Application to Explicit Least-squares Adjustments. Bulletin Géodésique, Vol. 62, No. 2, pp. 113–124.

A.M. MOOD, F.A. GRAYBILL and D.C. BOES (1974) : Introduction to the theory of statistics.

P.J.G. TEUNISSEN (1984) : A note on the use of Gauss' formulas in non-linear geodetic adjustments. Reports of the Department of Geodesy, Delft, 84.4.

P.J.G. TEUNISSEN (1985) : The Geometry of Geodetic Inverse Linear Mapping and Non-linear Adjustment. Neth. Geod. Comm., Publications on Geodesy, New Series, Vol. 8, No. 1.

P.J.G. TEUNISSEN and E.H. KNICKMEYER (1988) : Non-linearity and least-squares. CISM Journal ACSGC, Vol. 42, No. 4, pp. 321–330.

Received : 30.08.1988
Accepted : 18.11.1988 after editorial review.