28.

A note on the bias in the Symmetric Helmert Transformation

by

Peter J.G. Teunissen

Delft University of Technology

Faculty of Geodesy

Thijsseweg 11

NL-2629 JA DELFT

The Netherlands

ABSTRACT

In this note the bias of the least-squares estimators of the parameters of the Symmetric Helmert Transformation is derived. The approach taken makes use of the concept of normal curvature of the manifold described by the non-linear observation equations.

1. INTRODUCTION

The Helmert transformation furnishes the functional model for connecting two or more pointfields. For one and two dimensions the model is linear and simple analytical solutions for the least-squares estimators can be derived. For dimensions higher than two the model becomes nonlinear. In these cases iterative methods are needed for obtaining the least-squares solution. Based on a geometrical analysis of the model of the Helmert transformation, a numerically attractive solution method was developed in (Krarup, 1985).

In the classical Helmert transformation one pointfield is held fixed and is not adjusted for. In the symmetric Helmert transformation (Teunissen, 1985) however, both pointfields are adjusted for. Contrary to the classical Helmert transformation, the model of the Symmetric Helmert transformation is nonlinear for all dimensions. But for the one and two dimensional cases still relatively simple analytical solutions are available.

It can be shown that if both the pointfields need to be adjusted for, the classical Helmert transformation systematically underestimates the scale when compared to the scale estimator of the Symmetric Helmert transformation (Teunissen, 1985). A beautiful geometric analysis of a class of problems related to the Symmetric Helmert transformation was given in (Krarup, 1987).

In this note we will derive an approximation of the bias in the Symmetric Helmert transformation. In section 2 we briefly review some of the theory of bias due to nonlinearity. In section 3 we define the normal curvature and show how it affects the bias in the least-squares residual vector. Finally in section 4 the results of the previous sections are applied to the Symmetric Helmert transformation.

2. BIAS

Consider the nonlinear model

$$E\{\underline{y}\} = A(x) \ , \ \sigma^2 Q_y \ , \qquad (1)$$

where $E\{.\}$ is the mathematical expectation, $\underline{y}$ is a random m-vector with covariance matrix $\sigma^2 Q_y$, x is a p-vector of fixed but unknown parameters and $A(.)$ is a nonlinear map from R^p into R^m with $m > p$. We will assume that the the Jacobian of $A(.)$ and Q_y are both of full rank.

In order to appreciate the bias situation in the least-squares inversion of (1), we will first discuss briefly the forward problem, i.e. the bias situation if we go from x to y.

Forward problem: Let $\underline{\hat{x}}$ be an estimator of x with covariance matrix $\sigma^2 Q_{\hat{x}}$ and a bias $b_{\hat{x}} = E\{\hat{x}-x\}$ of the order σ^2. With a Taylor expansion of A(.) follows that the bias $b_{\hat{y}} = E\{\underline{\hat{y}}-\underline{y}\}$ in $\underline{\hat{y}} = A(\underline{\hat{x}})$ can be approximated as

$$\boxed{b_{\hat{y}} = \partial_x A\ b_{\hat{x}} + \frac{1}{2}\ \sigma^2\ \text{trace}\ (\partial^2_{xx} A Q_{\hat{x}})}\ . \tag{2}$$

In this approximation terms of the order σ^3 and higher are neglected. The second term on the right-hand side of (2) is the bias contribution due to non-linearity. A useful upperbound on this term can be obtained with the Cauchy-Schwarz inequality $(x^*\partial^2_{xx}A^i Q_{\hat{x}}x)^2 \le (x^*\partial^2_{xx}A^i\partial^2_{xx}A^i x)(x^*Q_{\hat{x}}Q_{\hat{x}}x)$. From this inequality follows for the eigenvalues that $|\mu(\partial^2_{xx}A^i Q_{\hat{x}})|_{max} \le |\mu(\partial^2_{xx}A^i)|_{max}\mu_{max}(Q_{\hat{x}})$ and therefore that

$$\frac{1}{2}\ \sigma^2|\text{trace}\ (\partial^2_{xx}A^i Q_{\hat{x}})| \le \frac{1}{2}\ \sigma^2 p|\mu(\partial^2_{xx}A^i)|_{max}\ \text{trace}(Q_{\hat{x}}) \tag{3}$$

Due to the sparsity of the Hessian $\partial^2_{xx}A^i$ it is usually not too difficult to obtain a realistic estimate of its in absolute value maximum eigenvalue.

Least-squares inverse problem: If $\underline{\hat{x}}$ is the least-squares estimator of x under model (1), $-b_{\hat{y}} = E\{\underline{y}-\underline{\hat{y}}\}$ becomes the bias in the least-squares residual vector, $b_{\hat{e}}$, and (2) can be written as

$$b_y = \partial_x A\ b_{\hat{x}} + b_{\hat{e}}\ , \tag{4}$$

where we have used the abbreviation

$$b_y = -\ \frac{1}{2}\ \sigma^2\ \text{trace}(\partial^2_{xx}A\ Q_{\hat{x}}). \tag{5}$$

The structure of (4) indicates that the biases in $\underline{\hat{x}}$ and $\underline{\hat{e}}$ are given by respectively

$$\boxed{b_{\hat{x}} = (\partial_x A)^- P_{\partial_x A}\ b_y} \tag{6}$$

and

$$\boxed{b_{\hat{e}} = P^{\perp}_{\partial_x A}\ b_y}\ , \tag{7}$$

where $(\partial_x A)^-$ is a left-inverse of $\partial_x A$, $P_{\partial_x A}$ is the orthogonal projector onto the range space of $\partial_x A$ and $P^{\perp}_{\partial_x A} = I - P_{\partial_x A}$. From (4), (6) and (7) follows that

$$\| b_y \|^2_{Q_y} = \| b_{\hat{x}} \|^2_{Q_{\hat{x}}} + \| b_{\hat{e}} \|^2_{Q_y} . \tag{8}$$

Hence this relatively easy computable scalar can be used as a first indicator whether the bias due to nonlinearity in $\underline{\hat{x}}$ and $\underline{\hat{e}}$ is significant or not.

3. CURVATURE

In Gaussian surface theory the normal curvature is defined as the ratio of the second fundamental form and the first fundamental form. The second fundamental form for the map A(.) is given by $x^*(n^* Q_y^{-1} \partial^2_{xx} A) x$ where n is a unit normal vector, i.e. $n^* Q_y^{-1} \partial_x A = 0$ and $n^* Q_y^{-1} n = 1$. This form is a generalization of the classical second fundamental form. There are m-p of such forms, one for each normal direction. In analogy with the classical Gaussian surface theory we define the normal curvature for model (1) as

$$K_n(x) \triangleq \frac{x^*(n^* Q_y^{-1} \partial^2_{xx} A) x}{x^* Q_{\hat{x}}^{-1} x} \tag{9}$$

The extreme values of this ratio are the principal curvatures. They follow from the generalized eigenvalue problem

$$| n^* Q_y^{-1} \partial^2_{xx} A - \mu Q_{\hat{x}}^{-1} | = 0 . \tag{10}$$

If we denote the p principal curvatures for the normal direction n by $k^1_n, k^2_n, \ldots, k^p_n$ and let $n_1, n_2, \ldots, n_{m-p}$ be an orthonormal basis of the orthogonal complement of the range space of $\partial_x A$, then the projector $P^{\perp}_{\partial_x A}$ can be written as $P^{\perp}_{\partial_x A} = \sum_{i=1}^{m-p} n_i n_i^* Q_y^{-1}$ and the bias $b_{\hat{e}} = P^{\perp}_{\partial_x A} b_y$ can be expressed in terms of the principal curvatures as

$$\boxed{b_{\hat{e}} = -\frac{1}{2} \sigma^2 \sum_{i=1}^{m-p} n_i \sum_{\alpha=1}^{p} K^{\alpha}_{n_i}} \tag{11}$$

This result shows how the local geometry of the manifold A(.) determines the bias in the least-squares residual vector $\underline{\hat{e}}$.

4. THE SYMMETRIC HELMERT TRANSFORMATION

The nonlinear model of the two dimensional Symmetric Helmert transformation reads

$$E\{\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}\} = \begin{bmatrix} I_{2n} & 0 \\ \lambda R \otimes I_n & I_2 \otimes c \end{bmatrix} \begin{bmatrix} u \\ t \end{bmatrix}, \quad \sigma^2 \begin{bmatrix} I_{2n} & 0 \\ 0 & I_{2n} \end{bmatrix}, \tag{12}$$

where λ is scale, R is a two dimensional rotation matrix, "$\otimes$" stands for the Kronecker product and c is a vector of ones. If R=I and t=0, we obtain the one dimensional Symmetric Helmert transformation without translation:

$$E\{\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}\} = \begin{bmatrix} I_q \\ \lambda I_q \end{bmatrix} u \ , \quad \sigma^2 \begin{bmatrix} I_q & 0 \\ 0 & I_q \end{bmatrix}, \ q = 2n. \tag{13}$$

This model describes a (q+1) dimensional curved manifold embedded in a 2q dimensional flat space.

With

$$\partial_x A = \begin{bmatrix} I_q & 0 \\ \lambda I_q & u \end{bmatrix} \quad \text{and} \quad n^* Q_y^{-1} \partial_{xx}^2 A = \begin{bmatrix} 0 & n_2 \\ n_2^* & 0 \end{bmatrix},$$

where n is partitioned as $n = (n_1^* n_2^*)^*$, the generalized eigenvalue problem (10) for model (13) reads

$$\left| \begin{bmatrix} 0 & n_2 \\ n_2^* & 0 \end{bmatrix} - \mu \begin{bmatrix} (1+\lambda^2) I_q & \lambda u \\ \lambda u^* & \| u \|^2 \end{bmatrix} \right| = 0. \tag{14}$$

This shows that there are q-1 zero principal curvatures. The corresponding principal directions are given by the columns of the (q+1)x(q-1) matrix $(n_2^{\perp *} \ \ 0)^*$. Thus there are only two non-zero principal curvatures for each of the q-1 mutually orthogonal normal directions. Moreover, since $n_2^* u = 0$, it follows that they are independent of the chosen normal direction and that they are given by

$$K_n = \pm \| u \|^{-1} . \tag{15}$$

Hence the manifold of (13) exhibits a saddle-type geometry with two principal directions of non-zero but in absolute value equal curvatures for each of the q-1 normal directions.

If translation is included in (13) the non-zero principal curvatures become

$$\boxed{K_n = \pm \| \bar{u} \|^{-1}, \quad \text{where} \quad \bar{u} = P_c^{\perp} u} \tag{16}$$

The same result is obtained for model (12). Since the non-zero principal curvatures only differ in their sign, it follows from equation (15) that $b_{\hat{e}}=0$. Hence we have obtained the important result that the least-squares estimators of the coordinates in (12) are free from bias despite the non-linearity of the observation equations. In order to obtain the bias in the transformation parameters, we first apply formula (5). For model (12) this gives

$$b_y = \frac{1}{2} \left(\frac{\sigma_{\hat{\lambda}}}{\lambda} \right)^2 \begin{bmatrix} 0 \\ \lambda R \otimes I_n \end{bmatrix} u , \tag{17}$$

where to first order in σ^2 the variance of $\hat{\lambda}$ is given by

$$\sigma^2_{\hat{\lambda}} = (1+\lambda^2) \frac{\sigma^2}{\| \bar{u} \|^2} . \tag{18}$$

Since $b_{\hat{e}}=0$, the scalar bias measure $\| b_{\hat{x}} \|^2_{Q_{\hat{x}}}$ follows form (8) and (17) as

$$\| b_{\hat{x}} \|^2_{Q_{\hat{x}}} = \frac{1}{4} \lambda^2 \left(\frac{\sigma_{\hat{\lambda}}}{\lambda} \right)^4 \frac{\| u \|^2}{\sigma^2} . \tag{19}$$

In order to obtain the individual bias components of $b_{\hat{x}}$, formula (6) has to be applied in general. For our particular case, however, the bias vector $b_{\hat{x}}$ can be obtained in a simpler way. By noting that b_y of (17) lies in the range space of the Jacobian of (12), we immediately can conclude that the biases in all parameters except scale vanishes. The bias in the least-squares estimator $\hat{\underline{\lambda}}$ follows then simply as

$$\boxed{E\{\underline{\hat{\lambda}}-\lambda\} = \frac{1}{2}\sigma_{\hat{\lambda}}\frac{\sigma_{\hat{\lambda}}}{\lambda}} \tag{20}$$

In many practical applications of model (12) it is customary to test whether $\lambda=1$ or not. The test statistic used for this test is $(\underline{\hat{\lambda}}-1)/\sigma_{\hat{\lambda}}$. It is usually assumed to have a standard normal distribution under the nulhypothesis that $\lambda=1$. The result (20) shows, however, that due to nonlinearity the mean of the test statistic $(\underline{\hat{\lambda}}-1)/\sigma_{\hat{\lambda}}$ under the hypothesis that $\lambda=1$, differs from zero by the amount $\frac{1}{2}\sigma_{\hat{\lambda}} = \sigma/\|\bar{u}\|$. This effect is practically negligible if the a priori precision is high enough (σ small enough) and the curvature small enough ($\|\bar{u}\|$ large enough). But the effect increases the smaller the network of points becomes. In order to get some indications of how the bias in scale depends on the number of network points and the distance between them, we assume that the points are distributed over a square grid of squares with side length d. With these assumptions equation (20) can be worked out to give

$$E\{\underline{\hat{\lambda}}-\lambda\} = 3\,\frac{1+\lambda^2}{\lambda n(n-1)}\left(\frac{\sigma}{d}\right)^2 \tag{21}$$

This shows that for most practical applications the bias in scale can be neglected. For $\sigma/d = 10^{-5}$, $\lambda = 1$ and $n = 4$ we have namely $b_{\hat{\lambda}} = \frac{1}{2}.10^{-10}$.

As a final remark it is interesting to note that the above results can also be used to predict the local convergence behaviour of the Gauss-Newton iteration method when applied to the Symmetric Helmert transformation. The local convergence factor (lcf.) of the Gauss-Newton method reads (Teunissen, 1985)

$$\boxed{\text{lcf.} = |\,k_{\hat{e}}\,|_{max}\|\,\hat{e}\,\|} \tag{22}$$

For model (12) this becomes

$$\text{lcf.} = \frac{\|\,\hat{e}\,\|}{\|\,\bar{u}\,\|}\,. \tag{23}$$

A useful a priori estimate of the local rate of convergence is therefore

$$\text{lcf.} \doteq \frac{\sigma(2n-4)^{\frac{1}{2}}}{\|\,\bar{u}\,\|}\,.$$

REFERENCES

Krarup, T. (1982): Non-Linear Adjustment and Curvature. In: Daar heb ik veertig jaar over nagedacht... Delft, pp. 145-159.

Krarup, T. (1985): Contribution to the Geometry of the Helmert transformation. Geodetic Institute, Denmark.

Krarup, T. (1987): Personal communication.

Teunissen, P.J.G. (1985): The Geometry of Geodetic Inverse Linear Mapping and Non-Linear Adjustment. NGC-publication, New Series, Vol. 8, No. 1, Delft.

Teunissen, P.J.G. and E.H. Knickmeyer (1988): Nonlinearity and Least-Squares. Accepted for the Canadian Surveyor.