

DELFT UNIVERSITY OF TECHNOLOGY  
REPORTS OF THE FACULTY OF GEODESY  
MATHEMATICAL AND PHYSICAL GEODESY

87.1

THE 1 AND 2D SYMMETRIC HELMERT TRANSFORMATION:  
AN EXACT NON-LINEAR LEAST-SQUARES SOLUTION

by

P.J.G. Teunissen

Delft, The Netherlands

June 1987

Contents.

	Page.
Abstract	2
1. Introduction	2
2. A particular class of non-linear least-squares problems	3
3. The 1D Helmert transformation	9
4. The 2D Helmert transformation with a rotational invariant covariance structure	15
5. The 2D Symmetric Helmert transformation	17
6. Concluding remarks	22
7. References	24
Appendix	25

Peter J.G. Teunissen  
Faculty of Geodesy,  
Delft University of Technology,  
Thijsseweg 11, NL - 2629 JA Delft,  
The Netherlands

THE 1 AND 2D SYMMETRIC HELMERT TRANSFORMATION:  
AN EXACT NON-LINEAR LEAST-SQUARES SOLUTION.

Abstract

In this paper a particular class of non-linear least-squares problems for which it is possible to take advantage of the special structure of the non-linear model, is discussed. The non-linear models are of the ruled-type (Teunissen, 1985a). The proposed solution strategy for this class of problems is applied to the 1 and 2D non-linear Symmetric Helmert transformation. Exact non-linear least-squares solutions are derived.

1. Introduction

The aim of the present paper is to derive an exact non-linear least-squares solution for the 1 and 2D Symmetric Helmert transformation. In section two we discuss a particular class of non-linear least-squares problems for which it is possible to take advantage of the special structure of the non-linear model. The non-linear models are manifolds of the ruled-type (see Teunissen, 1985a). We show that for this class of non-linear least-squares problems a two-step procedure can be devised. The first step consists of a linear least-squares problem, while the second step consists of a non-linear least-squares problem of a reduced dimension. In general the second step has to be solved through the use of linearization and iteration techniques, such as Gauss' method or variations thereof. A theorem is given which justifies the proposed two-step procedure. In section three we first consider the linear 1D Helmert transformation. We show that the product of the scale estimator  $\hat{\lambda}_H$  and the scale estimator  $\hat{\lambda}'_H$  of the inverse 1D Helmert transformation does not satisfy  $\hat{\lambda}_H \cdot \hat{\lambda}'_H = 1$ . This is unsatisfactory and a consequence of the fact that the linear Helmert trans-

formation does not treat the two coordinate sets on an equal basis. We therefore introduce our 1D Symmetric Helmert transformation. The corresponding model is non-linear, but a member of the class of models considered in section 1. The proposed two-step procedure is therefore applied and it is shown that the second step reduces to an eigenvalue problem which can be solved in an analytical way.

In section four we generalize the stochastic model of the classical linear 2D Helmert transformation to rotational-invariant covariance matrices. The linear least-squares solution is given.

In section five we introduce our new non-linear 2D Symmetric Helmert transformation. A rotational-invariant covariance structure is assumed. The non-linear least-squares solution is derived with the proposed two-step procedure. We show that the product of the scale estimators  $\hat{\lambda}_{SH}$  and  $\hat{\lambda}'_{SH}$  of the Symmetric Helmert transformation and its inverse satisfies  $\hat{\lambda}_{SH} \cdot \hat{\lambda}'_{SH} = 1$ . We also show that in general one systematically underestimates the scale when using the classical Helmert transformation.

The appendix contains a proof of an expression for the derivative of an orthogonal projector. This result is useful in itself for perturbation analysis and is needed when one wants to apply Gauss' iteration method to the second step of the proposed two-step procedure.

## 2.A particular class of Non-Linear Least-Squares Problems

We will study a method that takes advantage of a special structure of an optimization problem, which is expressed so that the optimization with respect to some of the variables is easier than with respect to the others. Consider the unconstrained minimization problem

$$(2.1) \quad \min_{w \in R^n} f(w)$$

Suppose we can partition  $w$  into

$$(2.2) \quad w = \begin{pmatrix} u \\ v \end{pmatrix}, \quad u \in R^{n_1}, \quad v \in R^{n_2}, \quad n_1 + n_2 = n,$$

in such a way that the subproblem

$$(2.3) \quad \min_{u \in R^{n_1}} f(u,v)$$

is easy to solve for every  $v$  in the domain of consideration. Let  $u(v)$  denote one solution of (2.3) and formulate the problem

$$(2.4) \quad \min_{v \in R^{n_2}} f(u(v),v) .$$

We can now replace the original  $n$  dimensional minimization problem (2.1) by a  $n_2$  dimensional one (2.4), where each evaluation of the object function needs the solution of a  $n_1$  dimensional minimization problem (2.3).

In principle every partition of the variables is possible. But to be of advantage practically, (2.3) must be simpler to solve than the corresponding problem with respect to  $v$ .

If we restrict ourselves to least-squares problems, the above two-step procedure becomes particularly advantageous when some of the variables occur linearly. In this case (2.3) is simply a linear least-squares problem.

Example: Orthogonal projection onto a ruled surface (see Teunissen, 1985a)

A ruled surface is a surface which has the property that through every point of the surface there passes a straight line which lies entirely in the surface. Thus the surface is covered by straight lines, called rulings which form a family depending on one parameter.

In order to find a parametrization of a ruled surface choose on the surface a curve transversal to the rulings. Let this curve be given by  $c(v)$ ,  $v \in R$ . At any point of this curve take a vector  $t$  of the ruling which passes through this point. This vector obviously depends on  $v$ . Thus we have  $t(v)$ . Now we can write the equation of the surface as

$$(2.5) \quad a(u,v) = c(u) + ut(v), \quad u,v \in R, \quad a,c,t \in R^3.$$

The parameter  $v$  indicates the ruling on the surface and the parameter  $u$  shows the position on the ruling.

Now let us assume that we have to solve for the following non-linear least-squares problem:

$$(2.6) \quad \min_{u,v} \|y-a(u,v)\|^2 ,$$

with the norm  $\|.\|^2 = (.)^* Q_y^{-1} (.)$  .

Since the ruled surface is flat in the directions of the rulings, whilst curved in the directions transversal to it, it becomes advantageous to perform the adjustment in two steps. In the first step one would then solve for a linear least-squares adjustment problem, and in the second step for a non-linear adjustment problem of a reduced dimension. That is, one first solves for

$$(2.7) \quad \min_u \| (y-c(v)) - t(v)u \|^2 ,$$

which gives

$$(2.8) \quad u(v) = [t^*(v)Q_y^{-1}t(v)]^{-1}t^*(v)Q_y^{-1}(y-c(v)) .$$

Then in the second step one solves for the non-linear problem

$$(2.9) \quad \min_v \|y-(c(v) + t(v)u(v))\|^2 .$$

As a generalization of the foregoing example, we are interested in solving the non-linear model

$$(2.10) \quad E\{y\} = A(z)x , \quad \text{Cov.}\{y\} = Q_y ,$$

in a least-squares sense; where  $y$  is the  $m$  dimensional vector of observational variates,  $E\{.\}$  stands for the mathematical expectation,  $A(z)$  is a  $m \times n_1$  matrix,  $Q_y$  is the  $m \times m$  positive definite covariance matrix of  $y$ , and  $x$  and  $z$  are respectively the  $n_1$ - and  $n_2$  dimensional vectors of unknown parameters. We will assume that matrix  $A(z)$  has constant full rank for all  $z$  of interest.

We can write (2.10) in index notation as

$$(2.10') \quad E\{y^i\} = A_{\alpha}^i(z)x^{\alpha}, \quad \text{Cov.}\{y^i\} = g^{ij}.$$

We will assume that the  $m_{n_1}$  functions  $A_{\alpha}^i(z)$  are continuously differentiable.

We define

$$(2.11) \quad \begin{cases} \text{a)} & f(x,z) \triangleq \|y - A(z)x\|^2 \\ \text{b)} & f_1(z) \triangleq \|P_{A(z)}^{\perp} y\|^2 \\ \text{c)} & x(z) \triangleq A^{-}(z)y \end{cases},$$

where  $\|\cdot\|^2 = (\cdot)^* Q_y^{-1}(\cdot)$ ,  $P_{A(z)}$  is the orthogonal projector projecting onto the rangespace of  $A(z)$ ,  $P_{A(z)}^{\perp} = I - P_{A(z)}$  is the orthogonal projector projecting onto the orthogonal complement of the rangespace of  $A(z)$  and  $A^{-}(z)$  is the least-squares inverse of  $A(z)$ .

Since  $x(z)$  is the solution of  $\min_x f(x,z)$  we have that

$$(2.12) \quad \begin{cases} \text{a)} & f_1(z) = f(x(z), z) = \min_x f(x,z) \quad \forall z \\ \text{b)} & f_1(z) \leq f(x,z) \quad \forall x,z. \end{cases}$$

From (2.11) also follows that

$$(2.13) \quad \begin{cases} \text{a)} & \partial_x f(x,z) = -2(y - A(z)x)^* Q_y^{-1} A(z) \\ \text{b)} & \partial_z f(x,z) = -2(y - A(z)x)^* Q_y^{-1} \partial_z A(z)x \\ \text{c)} & \partial_z f_1(z) = -2(P_{A(z)}^{\perp} y)^* Q_y^{-1} \partial_z P_{A(z)} y. \end{cases}$$

In the appendix it is proved that

$$(2.14) \quad \partial_z P_{A(z)} = (I - P_{A(z)}) \partial_z A(z) A^{-}(z) + [Q_y^{-1} (I - P_{A(z)}) \partial_z A(z) A^{-}(z) Q_y]^*.$$

Since

$$(2.15) \quad \begin{cases} \text{a)} & P_{A(z)}^{\perp*} Q_y^{-1} P_{A(z)}^{\perp} = Q_y^{-1} P_{A(z)}^{\perp} \\ \text{b)} & A^{-}(z) P_{A(z)}^{\perp} = 0, \end{cases}$$

substitution of (2.14) into (2.13c) gives

$$(2.16) \quad \partial_z f_1(z) = -2y^* Q_y^{-1} P_{A(z)}^\perp \partial_z A(z) A^-(z) y .$$

We are now ready to proof the following theorem, which gives a justification for the discussed two-step procedure.

Theorem

(i). If  $\hat{x}$  and  $\hat{z}$  are such that

$$(2.17) \quad \text{a) } \partial_z f_1(\hat{z}) = 0, \quad \text{b) } \hat{x} = A^-(\hat{z})y ,$$

then

$$(2.18) \quad \text{a) } f_1(\hat{z}) = f(\hat{x}, \hat{z}), \quad \text{b) } \partial_x f(\hat{x}, \hat{z}) = 0, \quad \text{c) } \partial_z f(\hat{x}, \hat{z}) = 0 .$$

(ii). If  $\hat{x}$  and  $\hat{z}$  are such that

$$(2.19) \quad \text{a) } f_1(\hat{z}) \leq f_1(z) \quad \forall z, \quad \text{b) } \hat{x} = A^-(\hat{z})y ,$$

then

$$(2.20) \quad f(\hat{x}, \hat{z}) \leq f(x, z) \quad \forall x, z .$$

(iii). If  $\hat{x}$  and  $\hat{z}$  are such that

$$(2.21) \quad f(\hat{x}, \hat{z}) \leq f(x, z) \quad \forall x, z ,$$

then

$$(2.22) \quad \text{a) } f_1(\hat{z}) = f(\hat{x}, \hat{z}), \quad \text{b) } f_1(\hat{z}) \leq f_1(z) \quad \forall z .$$

proof of (i):

(2.18a) follows from (2.11c), (2.17b) and (2.12a).

(2.18b) follows from (2.17b), (2.13a), (2.15a) and the fact that  $P_{A(z)}^\perp A(z) = 0$ .

(2.18c) follows from (2.17a), (2.17b), (2.13b), (2.15a) and (2.16).

Thus if  $\hat{z}$  is a stationary point of  $f_1(z)$  and  $\hat{x}$  is defined by (2.17b) then  $(\hat{x}, \hat{z})$  forms a stationary point of  $f(x, z)$ .



proof of (ii):

We will give the proof by contradiction. Assume that a  $\bar{x}$  and  $\bar{z}$  exist such that  $f(\bar{x}, \bar{z}) < f(\hat{x}, \hat{z})$ . With (2.12b) this gives:  $f_1(\bar{z}) \leq f(\bar{x}, \bar{z}) < f(\hat{x}, \hat{z})$ . With (2.19b), (2.11c) and (2.12a) this gives:  $f_1(\bar{z}) \leq f(\bar{x}, \bar{z}) < f(\hat{x}, \hat{z}) = f(x(\hat{z}), \hat{z}) = f_1(\hat{z})$ . But this contradicts our assumption that  $f_1(\hat{z}) \leq f(z) \forall z$ . Hence no  $\bar{x}$  and  $\bar{z}$  exist such that  $f(\bar{x}, \bar{z}) < f(\hat{x}, \hat{z})$ .

Thus if  $\hat{z}$  is a global minimum of  $f_1(z)$  and  $\hat{x}$  is defined by (2.19b) then  $(\hat{x}, \hat{z})$  is a global minimum of  $f(x, z)$ .

proof of (iii):

First we will proof (2.22a).

From (2.12b) follows that  $f_1(\hat{z}) \leq f(\hat{x}, \hat{z})$ . Now let  $\bar{x} = A^-(\hat{z})y$ . With (2.11c) and (2.12a) follows then that  $f_1(\hat{z}) = f(\bar{x}, \hat{z}) \leq f(\hat{x}, \hat{z})$ . Since  $(\hat{x}, \hat{z})$  is a global minimum of  $f(x, z)$  we must have equality, i.e.  $f_1(\hat{z}) = f(\bar{x}, \hat{z}) = f(\hat{x}, \hat{z})$ .

We will proof (2.22b) by contradiction.

Assume that a  $\bar{z}$  exists such that  $f_1(\bar{z}) < f_1(\hat{z})$ . Now let  $\bar{x} = A^-(\bar{z})y$ . With (2.11c) and (2.12a) this gives:  $f_1(\bar{z}) = f(x(\bar{z}), \bar{z}) = f(\bar{x}, \bar{z}) < f_1(\hat{z})$ . According to (2.22a) we have  $f_1(\hat{z}) = f(\hat{x}, \hat{z})$  and thus  $f_1(\bar{z}) = f(x(\bar{z}), \bar{z}) = f(\bar{x}, \bar{z}) < f_1(\hat{z}) = f(\hat{x}, \hat{z})$ . But this contradicts our assumption that  $f(\hat{x}, \hat{z}) \leq f(x, z) \forall x, z$ .

Hence no  $\bar{z}$  exists such that  $f_1(\bar{z}) < f_1(\hat{z})$ .

Thus if  $(\hat{x}, \hat{z})$  is a global minimum of  $f(x, z)$  then  $\hat{z}$  is a global minimum of  $f_1(z)$ .

This concludes the proof of the theorem.

From (ii) and (iii) of the Theorem follows that if the global minimum  $(\hat{x}, \hat{z})$  of  $f(x, z)$  is unique, then also the global minimum of  $f_1(z)$  is unique and is given by  $\hat{z}$ . Conversely, if the global minimum  $\hat{z}$  of  $f_1(z)$  unique then the global minimum of  $f(x, z)$  is unique and is given by  $(\hat{x}, \hat{z})$ . The uniqueness of the x-component follows from the uniqueness of the least-squares inverse  $A^-(z)$ , since  $A(z)$  is assumed to be of constant full rank.

When one applies the above described two-step procedure one still has to solve for the non-linear problem  $\min_z \|P_{A(z)}^\perp y\|^2$ . This can be done by Gauss' iteration method or variations thereof. In this paper we will not discuss the application of the Gauss' iteration method to the above problem, but see e.g. (Teunissen, 1984, 1985a and b) for more details. Instead we will use the described two-step procedure to solve for the 1 and 2 dimensional non-linear Symmetric Helmert transformation in an analytical way.

First we consider the 1D Helmert transformation and its non-linear symmetrical generalization.

