Research Article **Open Access**

R. Lehmann* and M. Lösler

Hypothesis testing in non-linear models exemplified by the planar coordinate transformations

https://doi.org/10.1515/jogs-2018-0009
Received February 13, 2018; accepted June 25, 2018

Abstract: In geodesy, hypothesis testing is applied to a wide area of applications e.g. outlier detection, deformation analysis or, more generally, model optimisation. Due to the possible far-reaching consequences of a decision, high statistical test power of such a hypothesis test is needed. The Neyman–Pearson lemma states that under strict assumptions the often-applied likelihood ratio test has highest statistical test power and may thus fulfill the requirement. The application, however, is made more difficult as most of the decision problems are non-linear and, thus, the probability density function of the parameters does not belong to the well-known set of statistical test distributions. Moreover, the statistical test power may change, if linear approximations of the likelihood ratio test are applied.
The influence of the non-linearity on hypothesis testing is investigated and exemplified by the planar coordinate transformations. Whereas several mathematical equivalent expressions are conceivable to evaluate the rotation parameter of the transformation, the decisions and, thus, the probabilities of type 1 and 2 decision errors of the related hypothesis testing are unequal to each other. Based on Monte Carlo integration, the effective decision errors are estimated and used as a basis of valuation for linear and non-linear equivalents.

Keywords: Hypothesis testing; Likelihood ratio test; Monte Carlo integration; Non-linear model; Coordinate transformation

1 Introduction

Hypothesis testing plays an important role in the framework of parameter estimation. In the context of outlier detection, hypothesis testing is used to detect and to identify implausible observations (e.g. Lehmann and Lösler 2016, Klein et al. 2017). In congruence analysis, hypothesis testing is introduced to distinguish stable points or areas from instable parts of an epochal observed network (e.g. Velsink 2015, Lehmann and Lösler 2017). To find an adequate number of model parameters, e.g. in the framework of reverse engineering, hypothesis testing indicates the benefit of a more complex model versus a simplified model (e.g. Ahn 2005).

In geodesy, the likelihood ratio (LR) test is most often applied (Koch 1999, Teunissen 2000). It bases on the Neyman–Pearson lemma, which demonstrates that under various assumptions such a test has the highest statistical test power (Neyman and Pearson 1933). In practice, most of the decision problems are non-linear and the underlying likelihood function must be maximized iteratively, e.g. by ordinary least-squares techniques with the risk of finding only a local maximum. Moreover, the often-used LR test in the linearized model deteriorates the decision due to a potential loss of statistical test power. Finally, the true probability density function of such a test does not belong to well-known class of statistical test distributions, and therefore, critical values cannot be computed with standard statistical functions. To derive the true probability density function as well as corresponding critical values, a Monte Carlo integration can be carried out (see e.g. Lehmann 2012).

Estimation in non-linear geodetic models has been widely investigated. Teunissen (1985) found that two types of non-linearity exist: The first is inherent in the problem and manifests itself in the non-linearity of the model operator. The second is perhaps introduced by a parametrization, which can even make an inherently linear problem non-linear. This is the case when the planar four parame-

***Corresponding Author: R. Lehmann:** University of Applied Sciences Dresden, Faculty of Spatial Information, E-mail:
M. Lösler: Frankfurt University of Applied Sciences, Faculty of Architecture, Civil Engineering and Geomatics

ter transformation is parameterized by rotation angle and scale.

This investigation focuses on the influence of the non-linearity on hypothesis testing exemplified by the planar coordinate transformations. Here, several mathematical equivalent expressions are conceivable to evaluate the rotation parameter of the transformation by hypothesis testing. Depending on the degree of non-linearity, the effective α can differ in comparison to its usually used χ^2 equivalent. The planar geodetic coordinate transformation is a good example to study non-linear effects in geodetic models, because under standard assumptions on the covariance matrix it admits an analytical solution (Teunissen 1985, 1986). Moreover, the planar coordinate transformations have a wide range of applications in geodesy.

The paper is organized as follows: After briefly introducing the non-linear Gauss–Markov model, we focus on the LR test as a general decision method. Then the least squares solutions of planar coordinate transformations are introduced. As an example for hypothesis testing in non-linear models, we set up a test problem for the rotation angle and solve it by various different applications of the LR test. Finally, we compare these different solutions in terms of decision errors, for which the method of Monte Carlo integration is used.

2 Hypothesis test in the non-linear Gauss–Markov model

Throughout this paper, true values of quantities will be denoted by tilde and estimates by hat.

We start from the non-linear Gauss–Markov model (GMM)

$$Y = \mathcal{A}\left(\tilde{X}\right) + e \tag{2.1}$$

where Y is a n-vector of observations and $\tilde{X}$ is a u-vector of unknown true model parameters. $\mathcal{A}$ is a known non-linear operator mapping from the u-dimensional parameter space to the n-dimensional observation space. e is an unknown random n-vector of normally distributed observation errors. The associated stochastic model reads:

$$e \sim N(0,\ \sigma^2 P^{-1}) \tag{2.2}$$

P is a known positive definite $n \times n$-matrix of weights (weight matrix). σ^2 is the a priori variance factor, which may be either known or unknown. Estimates $\hat{X}$, $\hat{Y}$ of the unknown true values of $\tilde{X}$, $\tilde{Y} = Y - e$ are desired.

In geodesy, a decision problem is generally posed as a statistical hypothesis test. Opposing the special model represented by the GMM Eqs. (2.1), (2.2) augmented by non-linear equality constraints

$$\mathcal{B}\left(\tilde{X}\right) = \mathcal{b} \tag{2.3}$$

to a general model represented by the GMM without equality constraints is equivalent to opposing the null hypothesis

$$H_0 : \mathcal{B}\left(\tilde{X}\right) = \mathcal{b} \tag{2.4}$$

to the alternative hypothesis

$$H_A : \mathcal{B}\left(\tilde{X}\right) \neq \mathcal{b}. \tag{2.5}$$

The standard solution of the testing problem in classical statistics goes as follows (e.g. Tanizaki 2004 p. 49 ff):

1. A test statistic $T(Y)$ is introduced, which is known to assume extreme values if H_0 does not hold true.
2. Under the condition that H_0 holds true, the probability distribution of $T(Y)$ is derived, represented by a cumulative distribution function (CDF) $F(T|H_0)$.
3. A probability of type 1 decision error α (significance level) is suitably defined (say 0.01 or 0.05 or 0.10), see Fig. 1.
4. For one-sided tests, a critical value c is derived by $c = F^{-1}(1 - \alpha|H_0)$, where F^{-1} denotes the inverse CDF (also known as quantile function) of $T|H_0$. (For two-sided tests two critical values are needed, but this case does not show up in this investigation.)
5. The empirical value of the test statistic $T(Y)$ is computed from the given observations Y. If $T(Y) > c$ then H_0 must be rejected, otherwise we fail to reject H_0.

In principle, we are free to choose a test statistic. Even heuristic choices like

$$T(Y) := \left\| \mathcal{B}\left(\hat{X}\right) - \mathcal{b} \right\| \tag{2.6}$$

with some suitable norm $\|\cdot\|$ are conceivable. Although the statistical test power (probability of rejection of H_0 when it is false) of such a test might be non-optimal or even poor.

3 The likelihood ratio test

In geodesy, we most often apply the likelihood ratio (LR) test (e.g. Tanizaki 2004 p. 54 ff). The test statistic of the LR test reads

$$T_{LR}(Y) := \frac{\max\left\{L\left(X, \sigma^2|Y\right) : \mathcal{B}(X) = \mathcal{b}\right\}}{\max\left\{L\left(X, \sigma^2|Y\right)\right\}} \tag{3.1}$$

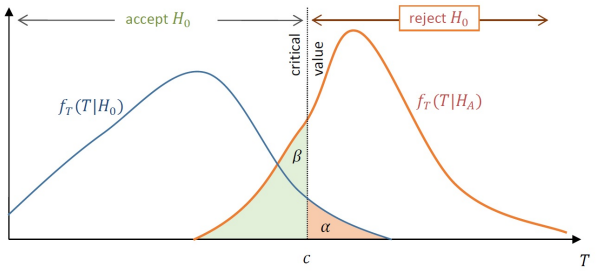

Figure 1: Probability density functions f of test statistic T under H_0 and H_A and decision errors α, β.

where $L\left(X, \sigma^2|Y\right)$ denotes the likelihood function of the GMM to be maximized with no restriction (denominator) and with the restriction $B(X) = b$ (numerator). For the GMM Eqs. (2.1), (2.2) the likelihood function reads

$$L\left(X, \sigma^2|Y\right) = \left(\det\left(2\pi\sigma^2 P^{-1}\right)\right)^{-0.5} \exp\left(-\frac{1}{2\sigma^2}(Y - \mathcal{A}(X))^T P (Y - \mathcal{A}(X))\right) \tag{3.2}$$

It is well known that maximizing $L\left(X, \sigma^2|Y\right)$ is equivalent to minimizing the least squares error functional (e.g. Koch 1999 p. 161f, Lösler et al. 2017)

$$\Omega(X) := (Y - \mathcal{A}(X))^T P (Y - \mathcal{A}(X)) \tag{3.3}$$

either with constraints $\mathcal{B}(X) = b$ or without constraints. In the first case, Ω is augmented by the Lagrange term

$$\Omega'(X, k) = \Omega(X) + 2k^T\left(\mathcal{B}(X) - b\right) \tag{3.4}$$

where k is the vector of Lagrange multipliers, in geodesy also known as correlates.

To simplify matters, we will restrict the derivation to the case of a known a priori variance factor σ^2. In this case, (3.1) can be expressed as

$$T_{LR}(Y) = \frac{\exp\left(-\frac{\min(\Omega'(X,k))}{2\sigma^2}\right)}{\exp\left(-\frac{\min(\Omega(X))}{2\sigma^2}\right)} = \exp\left(-\frac{\min(\Omega'(X, k)) - \min(\Omega(X))}{2\sigma^2}\right) \tag{3.5}$$

Moreover, we may replace $T_{LR}(Y)$ by the fully equivalent test statistic

$$T(Y) := -2 \cdot \log\left(T_{LR}(Y)\right) = \frac{\min(\Omega'(X, k)) - \min(\Omega(X))}{\sigma^2} \tag{3.6}$$

If $T(Y) > c$ with a properly chosen critical value c, then H_0 must be rejected, otherwise we fail to reject H_0.

Note that all these derivations are fully valid even if $\mathcal{A}$ or $\mathcal{B}$ are non-linear operators.

In the case that $\mathcal{A}$ and $\mathcal{B}$ are both linear operators, we obtain the expression (e.g. Lehmann and Neitzel 2013)

$$T(Y) = \hat{w}^T \Sigma_{\hat{w}}^{-1} \hat{w} \tag{3.7}$$

where

$$\hat{w} := \mathcal{B}\left(\hat{X}\right) - b \tag{3.8}$$

is the vector of estimated misclosures and $\Sigma_{\hat{w}}$ is the related covariance matrix. $\hat{X}$ denotes the minimizer of $\Omega(X)$ in Eq. (3.3), known as least squares estimate of X. Equation (3.7) can be seen as a special case of Eq. (2.6). If σ^2 is known and Eq. (2.2) holds true, the test statistic Eq. (3.7) follows the distributions:

$$T(Y|H_0) \sim \chi^2(m) \tag{3.9a}$$

$$T(Y|H_A) \sim \chi'^2\left(m, \Lambda'\right) \tag{3.9b}$$

with the non-centrality parameter

$$\Lambda' = \tilde{w}^T \Sigma_{\hat{w}}^{-1} \tilde{w} \tag{3.10}$$

m denotes the number of independent constraints. The vector of true misclosures $\tilde{w} = \mathcal{B}\left(\tilde{X}\right) - b$ and hence also Λ' are naturally unknown.

All established tests in geodesy belong to the class of LR tests. The rationale of these tests is provided by the famous Neyman–Pearson lemma (Neyman and Pearson 1933), which demonstrates that under various assumptions such a test has the highest statistical test power among all competitors. It is often applied even if we cannot exactly or only approximately make these assumptions in practice, because we know that the power is still larger than for rival tests (Teunisssen 2000, Kargoll 2012, Lehmann and Voß-Böhme 2017).

In truly non-linear models, we generally encounter three special problems:

1. The likelihood function Eq. (3.2) can only be maximized iteratively with the danger of finding only a local maximum. (Global optimization methods, which promise to find also the global maximum, are not yet widely applied practically because of the considerable computational workload for multidimensional problems.)
2. Test statistic Eq. (3.7) is only an approximation of the true LR test statistic Eq. (3.6) because the likelihood ratio is taken in the linearized GMM.
3. The probability density function (PDF) of $T(Y)$ or some equivalent of it does not belong to the well-known set of statistical test distributions (t, χ^2, F etc.)

such that the critical values must be computed numerically.

In the next sections, we will illustrate some consequences of these problems. In the conclusions, we will return to these points.

4 The least squares solution of the three-parameter transformation

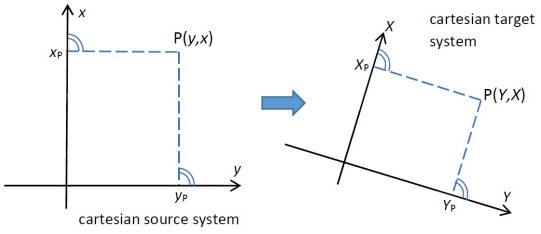

Figure 2: Planar parameter transformation

In a plane consider two Cartesian reference frames x, y and X, Y, which are related by translation and rotation, such that an arbitrary point P has coordinates x_P, y_P, X_P, Y_P satisfying the non-linear transformation equations

$$\begin{aligned} X_P &= X_0 + x_P \cdot \cos\epsilon - y_P \cdot \sin\epsilon \\ Y_P &= Y_0 + x_P \cdot \sin\epsilon + y_P \cdot \cos\epsilon \end{aligned} \tag{4.1}$$

with transformation parameters X_0, Y_0, ϵ, see Fig. 2. Related equations can be formulated for the opposite transformation direction.

We start from a set of N points having observed coordinates

$$x_1, y_1, \ldots, x_N, y_N, X_1, Y_1, \ldots, X_N, Y_N \tag{4.2}$$

in both frames. The problem is to find the best estimates for X_0, Y_0, ϵ in the least squares sense, also known as the least squares solution of the three-parameter transformation.

In the following, we restrict ourselves to the case that the coordinates of one system are non-stochastic (error-free) fixed quantities. Without restriction of generality, the error-free coordinates are denoted as $x_1, y_1, \ldots, x_N, y_N$. Moreover, we assume that for each pair of observations X_i, Y_i, both X_i and Y_i have the same weight p_i. This GMM reads

$$Y = \begin{pmatrix} X_1 \\ Y_1 \\ \vdots \\ X_N \\ Y_N \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} \tilde{\epsilon} \\ \tilde{X}_0 \\ \tilde{Y}_0 \end{pmatrix},$$

$$\mathcal{A}\left(\tilde{X}\right) = \begin{pmatrix} \tilde{X}_0 + x_1 \cdot \cos\tilde{\epsilon} - y_1 \cdot \sin\tilde{\epsilon} \\ \tilde{Y}_0 + x_1 \cdot \sin\tilde{\epsilon} + y_1 \cdot cos\tilde{\epsilon} \\ \vdots \\ \tilde{X}_0 + x_N \cdot \cos\tilde{\epsilon} - y_N \cdot \sin\tilde{\epsilon} \\ \tilde{Y}_0 + x_N \cdot \sin\tilde{\epsilon} + y_N \cdot cos\tilde{\epsilon} \end{pmatrix},$$

$$P = \begin{pmatrix} p_1 & 0 & \cdots & 0 & 0 \\ 0 & p_1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & p_N & 0 \\ 0 & 0 & \cdots & 0 & p_N \end{pmatrix} \tag{4.3}$$

This setting is one of the rare cases, where an analytical solution exists. For the sake of simplicity, we assume that σ^2 is chosen such that the weights fulfill $\Sigma p_i = 1$.

For the sake of compact notation, we introduce in either coordinate system the following abbreviations:

1. the weighted barycentres are given by

$$\begin{aligned} X_\star &:= \sum_{i=1}^{N} p_i X_i, \quad Y_\star := \sum_{i=1}^{N} p_i Y_i, \\ x_\star &:= \sum_{i=1}^{N} p_i x_i, \quad y_\star := \sum_{i=1}^{N} p_i y_i \end{aligned} \tag{4.4}$$

2. the coordinates related to the barycentres as origins

$$\begin{aligned} \Delta X_i &:= X_i - X_\star, \quad \Delta Y_i := Y_i - Y_\star, \\ \Delta x_i &:= x_i - x_\star, \quad \Delta y_i := y_i - y_\star \end{aligned} \tag{4.5}$$

3. the moments of inertia related to the barycentres

$$h := \sum_{i=1}^{N} p_i \left(\Delta x_i^2 + \Delta y_i^2\right) = -x_\star^2 - y_\star^2 + \sum_{i=1}^{N} p_i \left(x_i^2 + y_i^2\right) \tag{4.6a}$$

$$H := \sum_{i=1}^{N} p_i \left(\Delta X_i^2 + \Delta Y_i^2\right) = -X_\star^2 - Y_\star^2 + \sum_{i=1}^{N} p_i \left(X_i^2 + Y_i^2\right) \tag{4.6b}$$

4. the auxiliary terms

$$c := \sum_{i=1}^{N} p_i \left(\Delta X_i \Delta x_i + \Delta Y_i \Delta y_i\right) = \sum_{i=1}^{N} p_i \left(X_i \Delta x_i + Y_i \Delta y_i\right) \tag{4.7a}$$

$$s := \sum_{i=1}^{N} p_i \left(\Delta Y_i \Delta x_i - \Delta X_i \Delta y_i\right) = \sum_{i=1}^{N} p_i \left(Y_i \Delta x_i - X_i \Delta y_i\right) \tag{4.7b}$$

In GMM Eq. (4.3) the non-linear least squares solution for ϵ, X_0, Y_0 reads (see appendix 1)

$$\hat{\epsilon} = \arctan \frac{s}{c} \tag{4.8a}$$

$$\hat{X}_0 = X_\star - x_\star \cdot \cos\hat{\epsilon} + y_\star \cdot \sin\hat{\epsilon} = X_\star - \frac{x_\star \cdot c - y_\star \cdot s}{\sqrt{c^2 + s^2}} \tag{4.8b}$$

$$\hat{Y}_0 = Y_\star - x_\star \cdot \sin\hat{\epsilon} - y_\star \cdot \cos\hat{\epsilon} = Y_\star - \frac{x_\star \cdot s + y_\star \cdot c}{\sqrt{c^2 + s^2}} \tag{4.8c}$$

$$\Omega\left(\hat{X}\right) = h + H - 2\sqrt{c^2 + s^2} \tag{4.8d}$$

These formulas do not directly contain the observations Y, but only the statistics $c, s, X_\star, Y_\star$. All other quantities are fixed. Therefore, the vector $(c, s, X_\star, Y_\star)^T$ is a sufficient statistic of the problem. Moreover, it is normally distributed because of the linear relationship

$$Z := \begin{pmatrix} c \\ s \\ X_\star \\ Y_\star \end{pmatrix} = \begin{pmatrix} \Delta x_1 & \Delta y_1 & \cdots & \Delta y_N \\ -\Delta y_1 & \Delta x_1 & \cdots & \Delta x_N \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \end{pmatrix} PY \tag{4.9}$$

The covariance matrix of Z can be derived by covariance propagation

$$\begin{aligned} \Sigma_Z &= \begin{pmatrix} \Delta x_1 & \Delta y_1 & \cdots & \Delta y_N \\ -\Delta y_1 & \Delta x_1 & \cdots & \Delta x_N \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \end{pmatrix} \\ &\quad P\sigma^2 P^{-1} P \begin{pmatrix} \Delta x_1 & \Delta y_1 & \cdots & \Delta y_N \\ -\Delta y_1 & \Delta x_1 & \cdots & \Delta x_N \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \end{pmatrix}^T \\ &= \sigma^2 \begin{pmatrix} h & 0 & 0 & 0 \\ 0 & h & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned} \tag{4.10}$$

Thus, $c, s, X_\star, Y_\star$ are even independent random variables. For the expectations we obtain

$$\begin{aligned} E\{c\} &= \sum_{i=1}^{N} p_i \left(E\{X_i\}\Delta x_i + E\{Y_i\}\Delta y_i\right) \\ &= \sum_{i=1}^{N} p_i \left(\left(\tilde{X}_0 + x_i \cdot \cos\tilde{\epsilon} - y_i \cdot \sin\tilde{\epsilon}\right)\Delta x_i \right. \\ &\quad \left. + \left(\tilde{Y}_0 + x_i \cdot \sin\tilde{\epsilon} + y_i \cdot \cos\tilde{\epsilon}\right)\Delta y_i\right) \\ &= \sum_{i=1}^{N} p_i \left(\left(\Delta x_i \cdot \cos\tilde{\epsilon} - \Delta y_i \cdot \sin\tilde{\epsilon}\right)\Delta x_i \right. \\ &\quad \left. + \left(\Delta x_i \cdot \sin\tilde{\epsilon} + \Delta y_i \cdot \cos\tilde{\epsilon}\right)\Delta y_i\right) = h \cdot \cos\tilde{\epsilon} \end{aligned} \tag{4.11a}$$

$$E\{s\} = \sum_{i=1}^{N} p_i \left(E\{Y_i\}\Delta x_i - E\{X_i\}\Delta y_i\right) = h \cdot \sin\tilde{\epsilon} \tag{4.11b}$$

$$\begin{aligned} E\{X_\star\} &= \sum_{i=1}^{N} p_i E\{X_i\} = \sum_{i=1}^{N} p_i \left(\tilde{X}_0 + x_i \cdot \cos\tilde{\epsilon} - y_i \cdot \sin\tilde{\epsilon}\right) \\ &= \tilde{X}_0 + x_\star \cdot \cos\tilde{\epsilon} - y_\star \cdot \sin\tilde{\epsilon} \end{aligned} \tag{4.11c}$$

$$\begin{aligned} E\{Y_\star\} &= \sum_{i=1}^{N} p_i E\{Y_i\} = \sum_{i=1}^{N} p_i \left(\tilde{Y}_0 + x_i \cdot \sin\tilde{\epsilon} + y_i \cdot \cos\tilde{\epsilon}\right) \\ &= \tilde{Y}_0 + x_\star \cdot \sin\tilde{\epsilon} + y_\star \cdot \cos\tilde{\epsilon} \end{aligned} \tag{4.11d}$$

Starting from an initial guess for ϵ, X_0, Y_0, the solution Eq. (4.8) can also be obtained as the limit of a sequence of linearized GMM. Despite of the non-linearity of the GMM, we obtain a unique solution for the parameter estimation problem. Thus, there is no danger of finding only a local minimum here.

5 The least squares solution of the four-parameter transformation

In extension of Eq. (4.1) we introduce a scale parameter μ such that the new observation equations read

$$\begin{aligned} X_P &= X_0 + \mu \cdot (x_P \cdot \cos\epsilon - y_P \cdot \sin\epsilon) \\ Y_P &= Y_0 + \mu \cdot (x_P \cdot \sin\epsilon + y_P \cdot \cos\epsilon) \end{aligned} \tag{5.1}$$

By the substitution $a := \mu \cdot \cos\epsilon,\ o := \mu \cdot \sin\epsilon$ we obtain the linear representation

$$\begin{aligned} X_P &= X_0 + x_P \cdot a - y_P \cdot o \\ Y_P &= Y_0 + x_P \cdot o + y_P \cdot a \end{aligned} \tag{5.2}$$

with the parameter vector

$$\tilde{X} = \begin{pmatrix} \tilde{a} \\ \tilde{o} \\ \tilde{X}_0 \\ \tilde{Y}_0 \end{pmatrix} \tag{5.3}$$

The least squares solution of this linear GMM is simple and well known:

$$\hat{a} = \frac{c}{h}, \quad \hat{o} = \frac{s}{h} \tag{5.4a}$$

$$\hat{X}_0 = X_\star - x_\star \cdot \hat{a} + y_\star \cdot \hat{o} \tag{5.4b}$$

$$\hat{Y}_0 = Y_\star - x_\star \cdot \hat{o} - y_\star \cdot \hat{a} \tag{5.4c}$$

$$\Omega\left(\hat{X}\right) = H - \frac{c^2 + s^2}{h} \tag{5.4d}$$

where h, H, c, s are as defined in Eqs. (4.6a,b), (4.7a,b). This solution permits an estimate of the rotation angle and scale parameter:

$$\hat{\epsilon} = \arctan\frac{\hat{o}}{\hat{a}} = \arctan\frac{s}{c} \tag{5.5a}$$

$$\hat{\mu} = \sqrt{\hat{a}^2 + \hat{o}^2} = \frac{\sqrt{c^2 + s^2}}{h} \tag{5.5b}$$

See also appendix 3. Note that Eq. (5.5a) coincides with Eq. (4.8a).

6 LR hypothesis testing in the three-parameter transformation

As an example of a hypothesis test in a planar transformation model, we want to test a hypothesis for the rotation angle $\tilde{\epsilon}$ of the form

$$H_0 : \tilde{\epsilon} = \epsilon_0 \quad \text{vs.} \quad H_A : \tilde{\epsilon} \neq \epsilon_0 \tag{6.1}$$

which can be identified as a special case of Eqs. (2.4),(2.5) by

$$\mathcal{B}\left(\tilde{X}\right) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}^T \tilde{X}, \quad b = \epsilon_0, \quad \hat{w} = \hat{\epsilon} - \epsilon_0 \tag{6.2}$$

with $m = 1$. Obviously, $\mathcal{B}$ is a linear operator, but $\mathcal{A}$ is not. To apply test statistic Eq. (3.7), $\mathcal{A}(X)$ must be linearized by Taylor expansion:

$$\mathcal{A}(X) = \mathcal{A}\left(\hat{X}\right) + A \cdot \left(X - \hat{X}\right) + o\left(\left\|X - \hat{X}\right\|\right) \tag{6.3}$$

In the following, we investigate four different derivations of a test statistic for problem Eq. (6.1).
① Starting from an initial guess for ϵ, X_0, Y_0, a sequence of linear GMM is computed, until the iteration converges. In the final step, the Jacobian matrix A assumes the form

$$A = \begin{pmatrix} -x_1 \cdot \sin\hat{\epsilon} - y_1 \cdot \cos\hat{\epsilon} & 1 & 0 \\ x_1 \cdot \cos\hat{\epsilon} - y_1 \cdot \sin\hat{\epsilon} & 0 & 1 \\ \vdots & \vdots & \vdots \\ -x_N \cdot \sin\hat{\epsilon} - y_N \cdot \cos\hat{\epsilon} & 1 & 0 \\ x_N \cdot \cos\hat{\epsilon} - y_N \cdot \sin\hat{\epsilon} & 0 & 1 \end{pmatrix} \tag{6.4}$$

This gives an approximation of the covariance matrix of the estimated parameters (see appendix 4)

$$\Sigma_{\hat{X}} = \sigma^2 \left(A^T P A\right)^{-1} = \sigma^2 \begin{pmatrix} \Sigma p_i \left(x_i^2 + y_i^2\right) & & symm. \\ -x_\star \cdot \sin\hat{\epsilon} - y_\star \cdot \cos\hat{\epsilon} & 1 & \\ x_\star \cdot \cos\hat{\epsilon} - y_\star \cdot \sin\hat{\epsilon} & 0 & 1 \end{pmatrix}^{-1}$$

$$= \frac{\sigma^2}{h} \begin{pmatrix} 1 & & symm. \\ x_\star \cdot \sin\hat{\epsilon} + y_\star \cdot \cos\hat{\epsilon} & h + (x_\star \cdot \sin\hat{\epsilon} + y_\star \cdot \cos\hat{\epsilon})^2 & \\ -x_\star \cdot \cos\hat{\epsilon} + y_\star \cdot \sin\hat{\epsilon} & \frac{y_\star^2 - x_\star^2}{2} \sin 2\hat{\epsilon} - x_\star y_\star \cos 2\hat{\epsilon} & h + (x_\star \cdot \cos\hat{\epsilon} - y_\star \cdot \sin\hat{\epsilon})^2 \end{pmatrix} \tag{6.5a}$$

$$\sigma_{\hat{\epsilon}}^2 = \frac{\sigma^2}{h} \tag{6.5b}$$

When we perform the LR test of Eq. (4.3) using the linear approximation of $\mathcal{A}$, we come up with Eq. (3.7), which reads here

$$T_{3.1}(Z) = \left(\hat{\epsilon} - \epsilon_0\right)^T \Sigma_{\hat{\epsilon}}^{-1} \left(\hat{\epsilon} - \epsilon_0\right) = \left(\frac{\hat{\epsilon} - \epsilon_0}{\sigma_{\hat{\epsilon}}}\right)^2 = \frac{h}{\sigma^2} \left(\arctan\frac{s}{c} - \epsilon_0\right)^2 \tag{6.6}$$

(“3.1” denotes here the 1st version of the three-parameter test statistic.)

② A practically equivalent formulation of Eq. (6.1) is

$$H_0 : \tan\tilde{\epsilon} = \tan\epsilon_0 \quad \text{vs.} \quad \text{HA} \quad : \tan\tilde{\epsilon} \neq \tan\epsilon_0 \tag{6.7}$$

(disregarding the impractical non-issue that $\tan\epsilon = \tan(\epsilon + \pi)$).

Here, $\mathcal{B}(X)$ is also non-linear and must be linearized by Taylor expansion:

$$\mathcal{B}(X) = \mathcal{B}\left(\hat{X}\right) + B^T \cdot \left(X - \hat{X}\right) + o\left(\left\|X - \hat{X}\right\|\right) \tag{6.8}$$

In the final step of the iteration, the Jacobian matrix B assumes the form

$$B = \begin{pmatrix} \cos^{-2}\hat{\epsilon} \\ 0 \\ 0 \end{pmatrix} \tag{6.9}$$

In this case, Eq. (3.7) reads

$$T_{3.2}(Z) = \left(\tan\hat{\epsilon} - \tan\epsilon_0\right)^T \left(B^T \Sigma_{\hat{X}} B\right)^{-1} \left(\tan\hat{\epsilon} - \tan\epsilon_0\right)$$

$$= \frac{\left(\tan\hat{\epsilon} - \tan\epsilon_0\right)^2}{\frac{\sigma^2}{h}\cos^{-4}\hat{\epsilon}} = \frac{h}{\sigma^2}\left(\frac{sc - c^2\tan\epsilon_0}{c^2 + s^2}\right)^2 \tag{6.10}$$

This result is obviously different from Eq. (6.6). One could argue that Eq. (6.10) should be less reliable than Eq. (6.6) because also $\mathcal{B}$ must now be linearized too. But this argument is not conclusive, because we could have obtained this result also by substituting $t := \tan\epsilon$ in the transformation equations and solving and testing for the new parameter t instead of ϵ. In this case, $\mathcal{B}$ would be the same as in Eq. (6.2). The same line of reasoning would apply for other trigonometric functions in Eq. (6.7).

The main reason why (6.6) and (6.10) are different is not the “non-issue” discussed above, but the fact that the linearization errors by truncating the corresponding Taylor expansions are different. Proof: Use “cot” instead of “tan” in (6.7). Although the same $\tilde{\epsilon} = \epsilon_0 + k\pi,\ k \in \mathbb{Z}$ holds, we arrive at a different test statistic than in (6.10).

③ Applying covariance propagation to Eq. (4.8a) and using the quotient rule and the chain rule, we obtain the expression:

$$\sigma_{\hat{\epsilon}}^2 = \begin{pmatrix} \frac{-s}{\left[1 + \frac{s^2}{c^2}\right] c^2} & \frac{1}{\left[1 + \frac{s^2}{c^2}\right] c} & 0 & 0 \end{pmatrix} \Sigma_Z \begin{pmatrix} \frac{-s}{\left[1 + \frac{s^2}{c^2}\right] c^2} & \frac{1}{\left[1 + \frac{s^2}{c^2}\right] c} & 0 & 0 \end{pmatrix}^T = \frac{\left(c^2 + s^2\right)\sigma^2 h}{\left[1 + \frac{s^2}{c^2}\right]^2 c^4} = \frac{\sigma^2 h}{c^2 + s^2} \tag{6.11}$$

This is different from Eq. (6.5b), because the linearization is applied at a later stage. Therefore, we can assume that this is a better approximation than Eq. (6.5b). Using this expression in Eq. (3.7) yields

$$T_{3.3}(Z) = \left(\frac{\hat{\epsilon} - \epsilon_0}{\sigma_{\hat{\epsilon}}}\right)^2 = \frac{c^2 + s^2}{\sigma^2 h}\left(\arctan\frac{s}{c} - \epsilon_0\right)^2 \tag{6.12}$$

But still this test statistic is a linear approximation via Eq. (3.7).

④ To obtain a fully non-linear LR test statistic, we revert to Eq. (3.6):

$$\begin{aligned} T_{3.4}(Z) &= \frac{\min(\Omega') - \min(\Omega)}{\sigma^2} \\ &= \frac{\Omega(\hat{X}_0, \hat{Y}_0, \epsilon_0) - \Omega(\hat{X}_0, \hat{Y}_0, \hat{\epsilon})}{\sigma^2} \\ &= \frac{1}{\sigma^2}\left[(h + H - 2(c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0)) \right. \\ &\quad \left. - \left(h + H - 2\sqrt{c^2 + s^2}\right)\right] \\ &= \frac{2}{\sigma^2}\left(\sqrt{c^2 + s^2} - c \cdot \cos\epsilon_0 - s \cdot \sin\epsilon_0\right) \end{aligned} \tag{6.13}$$

where appendix 1 and Eq. (4.8d) have been used.

Note that this test statistic is as simple to compute as the three previous versions.

7 LR hypothesis testing in the four-parameter transformation

We want to test the same hypothesis Eq. (6.1), but now for the four-parameter transformation. In terms of the substitution model parameters Eq. (5.5a), it can be formulated as

$$H_0 : \arctan\frac{\tilde{o}}{\tilde{a}} = \epsilon_0 \quad \text{vs.} \quad H_A : \arctan\frac{\tilde{o}}{\tilde{a}} \neq \epsilon_0 \tag{7.1}$$

This can be identified as a special case of Eqs. (2.4), (2.5) by

$$\mathcal{B}\left(\tilde{X}\right) = \arctan\frac{\tilde{o}}{\tilde{a}}, \quad b = \epsilon_0, \quad \hat{w} = \arctan\frac{\hat{o}}{\hat{a}} - \epsilon_0 \tag{7.2}$$

with $m = 1$.

In the following, we investigate four different derivations of a test statistic for problem Eq. (7.1).

① Acting on a, o, operator $\mathcal{B}$ is non-linear, but $\mathcal{A}$ is linear here. Consequently, $\mathcal{B}(X)$ must be linearized as in Eq. (6.8):

$$B = \begin{pmatrix} \frac{-o}{\left[1 + \frac{o^2}{a^2}\right] a^2} & \frac{1}{\left[1 + \frac{o^2}{a^2}\right] a} & 0 & 0 \end{pmatrix}^T = \frac{1}{o^2 + a^2}\begin{pmatrix} -o \\ a \\ 0 \\ 0 \end{pmatrix} \tag{7.3}$$

In this case, Eq. (3.7) reads

$$T_{4.1}(Z) = \left(\arctan\frac{\hat{o}}{\hat{a}} - \epsilon_0\right)^T \left(B^T \Sigma_{\hat{X}} B\right)^{-1} \left(\arctan\frac{\hat{o}}{\hat{a}} - \epsilon_0\right) = \frac{c^2 + s^2}{\sigma^2 h}\left(\arctan\frac{s}{c} - \epsilon_0\right)^2 \tag{7.4}$$

where $\Sigma_{\hat{X}}$ is the well-known covariance matrix (e.g. *Wolf 1966*, *Somogyi and Kalmár 1988*)

$$\Sigma_{\hat{X}} = \sigma^2 \begin{pmatrix} 1/h & 0 & 0 & 0 \\ 0 & 1/h & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.5}$$

It turns out that $T_{4.1}(Z) \equiv T_{3.3}(Z)$. However, the corresponding models are different.

② Alternatively, we can solve the non-linear four-parameter transformation with parameters μ, ϵ instead of a, o by iteration. In the final step, the Jacobian matrix A assumes the form

$$A = \begin{pmatrix} -x_1 \cdot \hat{\mu} \cdot \sin\hat{\epsilon} - y_1 \cdot \hat{\mu} \cdot \cos\hat{\epsilon} & x_1 \cdot \cos\hat{\epsilon} - y_1 \cdot \sin\hat{\epsilon} & 1 & 0 \\ x_1 \cdot \hat{\mu} \cdot \cos\hat{\epsilon} - y_1 \cdot \hat{\mu} \cdot \sin\hat{\epsilon} & x_1 \cdot \sin\hat{\epsilon} + y_1 \cdot cos\hat{\epsilon} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ -x_N \cdot \hat{\mu} \cdot \sin\hat{\epsilon} - y_N \cdot \hat{\mu} \cdot \cos\hat{\epsilon} & x_N \cdot \cos\hat{\epsilon} - y_N \cdot \sin\hat{\epsilon} & 1 & 0 \\ x_N \cdot \hat{\mu} \cdot \cos\hat{\epsilon} - y_N \cdot \hat{\mu} \cdot \sin\hat{\epsilon} & x_N \cdot \sin\hat{\epsilon} + y_N \cdot cos\hat{\epsilon} & 0 & 1 \end{pmatrix} \tag{7.6}$$

This gives the covariance matrix of the estimated parameters (see appendix 5)

$$\Sigma_{\hat{X}} = \sigma^2 \left(A^T P A\right)^{-1} = \sigma^2 \begin{pmatrix} \hat{\mu}^2 \Sigma p_i \left(x_i^2 + y_i^2\right) & & symm. & \\ 0 & \Sigma p_i \left(x_i^2 + y_i^2\right) & & \\ \hat{\mu} \cdot \left(-x_\star \cdot \sin\hat{\epsilon} - y_\star \cdot \cos\hat{\epsilon}\right) & x_\star \cdot \cos\hat{\epsilon} - y_\star \cdot \sin\hat{\epsilon} & 1 & \\ \hat{\mu} \cdot \left(x_\star \cdot \cos\hat{\epsilon} - y_\star \cdot \sin\hat{\epsilon}\right) & x_\star \cdot \sin\hat{\epsilon} + y_\star \cdot \cos\hat{\epsilon} & 0 & 1 \end{pmatrix}^{-1}$$

$$= \frac{\sigma^2}{h} \begin{pmatrix} \hat{\mu}^{-2} & & symm. & \\ 0 & 1 & & \\ \hat{\mu}^{-1} \cdot \left(x_\star \cdot \sin\hat{\epsilon} + y_\star \cdot \cos\hat{\epsilon}\right) & y_\star \cdot \sin\hat{\epsilon} - x_\star \cdot \cos\hat{\epsilon} & \Sigma p_i \left(x_i^2 + y_i^2\right) & \\ \hat{\mu}^{-1} \cdot \left(y_\star \cdot \sin\hat{\epsilon} - x_\star \cdot \cos\hat{\epsilon}\right) & -x_\star \cdot \sin\hat{\epsilon} - y_\star \cdot \cos\hat{\epsilon} & 0 & \Sigma p_i \left(x_i^2 + y_i^2\right) \end{pmatrix} \tag{7.7a}$$

$$\sigma_{\hat{\epsilon}}^2 = \frac{\sigma^2}{h \cdot \hat{\mu}^2} \tag{7.7b}$$

The hypotheses are now formulated as in Eq. (6.1). Acting on μ, ϵ, operator $\mathcal{A}$ is non-linear, but $\mathcal{B}$ is linear here and corresponds to Eq. (6.2).

When we perform the LR test using the linear approximation of $\mathcal{A}$, we come up with Eq. (3.7), which reads here

$$T_{4.2}(Z) = \left(\hat{\epsilon} - \epsilon_0\right)^T \Sigma_{\hat{\epsilon}}^{-1} \left(\hat{\epsilon} - \epsilon_0\right) = \left(\frac{\hat{\epsilon} - \epsilon_0}{\sigma_{\hat{\epsilon}}}\right)^2 = \frac{c^2 + s^2}{h^2} \cdot \frac{h}{\sigma^2} \left(\arctan\frac{s}{c} - \epsilon_0\right)^2 \tag{7.8}$$

It turns out that $T_{4.2}(Z) \equiv T_{4.1}(Z) \equiv T_{3.3}(Z)$.

③ Let us now study the special case $\epsilon_0 = 0$. Here, the hypotheses can be written as

$$H_0 : \tilde{o} = 0 \text{ vs. } H_A : \tilde{o} \neq 0 \tag{7.9}$$

In the four-parameter transformation, this can be identified as a special case of Eqs. (2.4), (2.5) by

$$\mathcal{B}\left(\tilde{X}\right) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}^T \tilde{X}, \quad \mathfrak{b} = 0, \quad \hat{w} = \hat{o} \tag{7.10}$$

In this case, both $\mathcal{A}$ and $\mathcal{B}$ are linear operators and Eq. (3.7) reads

$$T_{4.3}(Z) = \frac{\hat{o}^2}{\sigma_{\hat{o}}^2} = \frac{h\hat{o}^2}{\sigma^2} = \frac{s^2}{\sigma^2 h} \tag{7.11}$$

where Eq. (5.4) and Eq. (7.5) have been used.

④ To obtain a fully non-linear LR test statistic, we revert to Eq. (3.6):

$$\begin{aligned} T_{4.4}(Z) &= \frac{\min(\Omega') - \min(\Omega)}{\sigma^2} \\ &= \frac{\Omega(\hat{X}_0, \hat{Y}_0, \epsilon_0, \hat{\mu}) - \Omega(\hat{X}_0, \hat{Y}_0, \hat{e}, \hat{\mu})}{\sigma^2} \\ &= \frac{1}{\sigma^2}\left(H - \frac{(c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0)^2}{h} - H + \frac{c^2+s^2}{h}\right) \\ &= \frac{c^2+s^2-(c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0)^2}{h\sigma^2} \\ &= \frac{(c \cdot \sin\epsilon_0 - s \cdot \cos\epsilon_0)^2}{h\sigma^2} \end{aligned} \tag{7.12}$$

where appendix 2 and Eq. (5.4d) have been used.

8 Distributions

Due to the coincidence with $T_{3.3}$, the test statistics $T_{4.1}, T_{4.2}$ will not be further discussed.

Note that all derived test statistics $T_i(Z)$ depend on only two of the four elements of Z, i.e. c and s. This will be highlighted by the notation $T_i(c,s)$ used below:

$$T_{3.1}(c,s) = \frac{h}{\sigma^2}\left(\arctan\frac{s}{c} - \epsilon_0\right)^2 \tag{8.1a}$$

$$T_{3.2}(c,s) = \frac{h}{\sigma^2}\left(\frac{sc - c^2\tan\epsilon_0}{c^2+s^2}\right)^2 \tag{8.1b}$$

$$T_{3.3}(c,s) = \frac{c^2+s^2}{\sigma^2 h}\left(\arctan\frac{s}{c} - \epsilon_0\right)^2 \tag{8.1c}$$

$$T_{3.4}(c,s) = \frac{2}{\sigma^2}\left(\sqrt{c^2+s^2} - c \cdot \cos\epsilon_0 - s \cdot \sin\epsilon_0\right) \tag{8.1d}$$

$$T_{4.3}(c,s) = \frac{s^2}{\sigma^2 h} \tag{8.1e}$$

$$T_{4.4}(c,s) = \frac{(c \cdot \sin\epsilon_0 - s \cdot \cos\epsilon_0)^2}{\sigma^2 h} \tag{8.1f}$$

In the linear or linearized GMM, we obtain from Eq. (3.9) the following distributions of the LR test statistics:

$$T_i(c,s)\,|H_0 \sim \chi^2(1),\ i = 3.1,\ 3.2,\ 3.3,\ 4.3, 4.4 \tag{8.2a}$$

$$T_i(c,s)\,|H_A \sim \chi'^2\left(1, \frac{h}{\sigma^2}(\tilde{\epsilon} - \epsilon_0)^2\right),\ i = 3.1,\ 3.3 \tag{8.2b}$$

$$T_{3.2}(c,s)\,|H_A \sim \chi'^2\left(1, \frac{h}{\sigma^2}(\tan\tilde{\epsilon} - \tan\epsilon_0)^2\cos^4\hat{\epsilon}\right) \tag{8.2c}$$

$$T_{4.3}(c,s)\,|H_A \sim \chi'^2\left(1, \frac{\tilde{s}^2}{\sigma^2 h}\right) \tag{8.2d}$$

$$T_{4.4}(c,s)\,|H_A \sim \chi'^2\left(1, \frac{(\tilde{c} \cdot \sin\epsilon_0 - \tilde{s} \cdot \cos\epsilon_0)^2}{\sigma^2 h}\right) \tag{8.2e}$$

However, observing that test statistics $T_i(c,s)$, $i = 3.1,\ 3.2,\ 3.3, 4.4$ in Eq. (8.2a) are obtained by linearization of $\mathcal{A}$ or $\mathcal{B}$ or both, these distributions can be no more than approximations of the true distributions of $T_i(c,s)$ in the vicinity of $\hat{e}$. But oftentimes $T_i(c,s)$ is evaluated far away from $\hat{e}$, especially if α is small. Test statistic $T_{3.4}$, $T_{4.4}$ is defined in the fully non-linear model and test statistic $T_{4.3}$ is defined in the fully linear model. Therefore, no such approximation is made here.

Remark: In Eq. (8.2b) it would not be correct to apply Eq. (6.11) instead of Eq. (6.5b). Equation (6.5b) must be used even in case $i = 3.3$, because Eq. (6.5b) is derived from $\Sigma_{\hat{X}}$ in Eq. (6.5a), as it is required by Eq. (3.9).

Simplifications:

① If we rotate the source system (x,y) by ϵ_0 about the barycentre $(x_\star, y_\star)$ and solve the same transformation problem with the rotated coordinates, c, s are replaced by

$$\begin{aligned} c' &= c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0 \\ s' &= -c \cdot \sin\epsilon_0 + s \cdot \cos\epsilon_0 \end{aligned} \tag{8.3}$$

Note that the vector $(c', s')^T$ is the result of the rotation of $(c,s)^T$ by angle $-\epsilon_0$ about the origin (0,0) and has therefore the same covariance matrix

$$\Sigma_{c',s'} = \sigma^2\begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix} \tag{8.4}$$

The transformation problems with the rotated coordinates have the solution

$$\begin{aligned} \cos\hat{\epsilon}' &= \frac{c'}{\sqrt{c'^2+s'^2}} = \frac{c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0}{\sqrt{c^2+s^2}} \\ &= \cos\hat{\epsilon} \cdot \cos\epsilon_0 + \sin\hat{\epsilon} \cdot \sin\epsilon_0 = \cos(\hat{\epsilon} - \epsilon_0) \end{aligned} \tag{8.5}$$

Now, testing Eq. (6.1) is obviously identical to testing

$$H_0 : \tilde{\epsilon}' = 0 \ \text{ vs. } \ H_A : \tilde{\epsilon}' \neq 0 \tag{8.6}$$

with the rotated coordinate system, i.e. with c', s'. This obviously results in the same test statistics $T_{3.1}, T_{3.3}$. Less obvious, the same applies to $T_{3.4}$ and $T_{4.4}$ by virtue of

$$T_{3.4}(c', s') = \frac{2}{\sigma^2}\left(\sqrt{c'^2+s'^2} - c'\right)$$

$$= \frac{2}{\sigma^2}\left(\sqrt{c^2+s^2} - c\cdot\cos\epsilon_0 - s\cdot\sin\epsilon_0\right)$$
$$= T_{3.4}(c,s) \tag{8.7a}$$

$$T_{4.4}\left(c',s'\right) = \frac{s'^2}{h\sigma^2} = \frac{(-c\cdot\sin\epsilon_0 + s\cdot\cos\epsilon_0)^2}{h\sigma^2}$$
$$= T_{4.4}(c,s) \tag{8.7b}$$

For $T_{4.3}$ no rotation is necessary, because it only applies to the special case $\epsilon_0 = 0$. Moreover, note that for $\epsilon_0 = 0$ we find a coincidence of $T_{4.3}$ and $T_{4.4}$. This shows that $T_{4.4}$ follows the χ^2-distribution Eq. (8.2a,d,e). Hence, we will further discuss only $T_{4.4}$.

However, the situation for $T_{3.2}$ is different. Here, a different result is obtained. Note that the solution in terms of the parameter $t := \tan\epsilon$ is not rotational invariant. This becomes obvious in the case of $\tilde{\epsilon} = \pm\pi/2$, where $\tilde{t}$ not even exists.

Disregarding this non-issue, we will continue with $\epsilon_0 = 0$, noting that almost no restriction of generality is made.

② If we scale both coordinate systems by the factor σ^{-1} and solve the transformation problem with the scaled coordinates, c', s', h, are replaced by

$$c'' = \frac{c'}{\sigma^2},\quad s'' = \frac{s'}{\sigma^2},\quad h'' = \frac{h}{\sigma^2},\quad \sigma''^2 = \frac{\sigma^2}{\sigma^2} = 1 \tag{8.8}$$

The weights do not change, such that $\Sigma p_i = 1$ is retained. Note that the new vector $(c'', s'')^T$ has the covariance matrix

$$\Sigma_{c'',s''} = \sigma^{-2}\begin{pmatrix} h & 0 \\ 0 & h \end{pmatrix} = \begin{pmatrix} h'' & 0 \\ 0 & h'' \end{pmatrix} \tag{8.9}$$

The new solution is $\hat{\epsilon}'' = \hat{\epsilon}'$.

Now testing Eq. (8.6) with the scaled coordinates, i.e. with c'', s'', obviously results in

$$T_{3.1}\left(c'',s''\right) = h''\left(\arctan\frac{s''}{c''}\right)^2 = T_{3.1}\left(c',s'\right) \tag{8.10a}$$

$$T_{3.2}\left(c'',s''\right) = h''\left(\frac{s''c''}{c''^2+s''^2}\right)^2 = T_{3.2}\left(c',s'\right) \tag{8.10b}$$

$$T_{3.3}\left(c'',s''\right) = \frac{c''^2+s''^2}{h''}\left(\arctan\frac{s''}{c''}\right)^2 = T_{3.3}\left(c',s'\right) \tag{8.10c}$$

$$T_{3.4}\left(c'',s''\right) = 2\left(\sqrt{c''^2+s''^2} - c''\right) = T_{3.4}\left(c',s'\right) \tag{8.10d}$$

$$T_{4.4}\left(c'',s''\right) = \frac{s''^2}{h''} = T_{4.4}\left(c',s'\right) \tag{8.10e}$$

Thus, all test statistics are scale invariant, too.

A special problem exists for $T_{3,2}$, which can be written as

$$T_{3.2}\left(c'',s''\right) = \frac{h''}{4}\sin^2 2\hat{\epsilon}'' \le \frac{h''}{4} \tag{8.11}$$

The fact that the χ^2 density function is non-zero on the whole positive real line again proves that $T_{3.2}$ has not the χ^2 distribution.

Henceforth, we drop double-primes, such that

$$\epsilon_0 := 0,\ \sigma^2 := 1 \tag{8.12}$$

is assumed with almost no loss of generality. (Remember that "almost" here concerns only $T_{3.2}$, which is not rotation invariant.)

The question, which test statistic is best, must be answered by the resulting probabilities of decision error.

1. The probability of type 1 decision error α is usually selected by the user. But if the $1-\alpha$-quantile of $\chi^2(1)$ is used for $T_{3.1}$, $T_{3.2}$, $T_{3.3}$, the effective α can be different.
2. The probability of type 2 decision error β should be small.

Both probabilities α and β are linked via the critical value c, see Fig. 1.

9 Probability of type 1 decision error

The idea is to compare the $1-\alpha$-quantiles of $\chi^2(1)$ with the quantiles of the true distribution of $T_i|H_0$ obtained by Monte Carlo integration. This method has been successfully used e.g. by Lehmann (2012) for the computation of critical values of normalized and studentized residuals employed in geodetic outlier detection. In principle, it replaces

- random variates by computer generated pseudo random numbers,
- probability distributions by histograms and
- statistical expectations by arithmetic means

computed from a large number of Monte Carlo experiments, i.e. computations with pseudo random numbers instead of noisy observations.

In the case that H_0 is true, we have $\tilde{\epsilon} = 0$, such that from Eq. (4.11) follows

$$E\{c|H_0\} = h,\ E\{s|H_0\} = 0. \tag{9.1}$$

According to Eqs. (4.10), (8.12) we need to generate the following pseudo random numbers:

$$c|H_0 \sim N(h, h), \quad s|H_0 \sim N(0, h) \tag{9.2}$$

We use $M = 10^8$ Monte Carlo samples, which turns out to be sufficiently high, because the results only insignificantly change, when the computations are repeated with different pseudo random numbers.

We use three stages of non-linearity, expressed by the signal/noise ratio: $h = 1000$ means that the signal is 1000 times larger than the noise ($\sigma = 1$), which causes only weak non-linear effects. Analogous, $h = 100$ and $h = 10$ cause medium and strong non-linear effects, respectively.

In Table 1, the $1-\alpha$-quantiles of $\chi^2(1)$ and the quantiles of the true distribution of $T_i|H_0$, $i = 3.1, 3.2, 3.3$ are compared. For $T_{4.3} \equiv T_{4.4}$ we can directly use the quantiles of $\chi^2(1)$. As expected, the largest differences occur for $h = 10$ and $T_{3.2}$. Using the χ^2-quantile as a critical value can be both, an advantage and a disadvantage in terms of α. Consider $h = 10$ and a desired $\alpha = 0.01$ in $T_{3.1}$, we erroneously select 6.63 as a critical value, instead of 8.92. The true α for $T_{3.1}$ is not 0.01, but even larger than 0.02. By interpolation of the derived quantiles in Table 1, we obtain an effective $\alpha = 0.021$. In contrast to that, we find from Eq. (8.11) that $|T_{3.2}| < 2.5$ always holds, such that 6.63 is never exceeded, which corresponds to an effective $\alpha = 0$.

The true quantiles of $T_{3.4}$ are given in Table 2, but should not be compared to the χ^2-quantiles, because they are obtained in the non-linear model. It is perhaps unexpected that $T_{3.4}$ follows the χ^2 distribution even better than other test statistics, as can be seen from a comparison of Table 1 and 2.

10 Probability of type 2 decision error

The aim of this investigation is to find out, which test statistic has the highest statistical test power, i.e. the best ability to reject a false H_0. For comparison, we plot the power function of $T_i|H_A$, $i = 3.1, \ldots, 4.4$, denoted as

$$1 - \beta_i(|\tilde{\epsilon}|) \tag{10.1}$$

Due to the symmetry of β_i, all plots are produced only for positive $\tilde{\epsilon}$. Whenever Eq. (8.2b,c) hold only approximately, we again use Monte Carlo integration to compute the true distribution of $T_i|H_A$. According to Eqs. (4.10), (4.11), (8.12) we need to generate the following pseudo random numbers:

$$c|H_A \sim N(h \cdot \cos\tilde{\epsilon}, h), \quad s|H_A \sim N(h \cdot \sin\tilde{\epsilon}, h) \tag{10.2}$$

We find

$$1 - \beta_i(|\tilde{\epsilon}|) = \Pr(T_i > c_i|H_A),\ i = 3.1, \ldots, 4.4 \tag{10.3}$$

where c_i is the critical value, which equals the $1-\alpha$-quantile of either the $\chi^2(1)$ distribution or the true distributions obtained in the preceding section, whenever this is different. The first case is practically applied. Below we restrict ourselves to the choice of $\alpha = 0.05$.

In Fig. 3, the power function Eq. (10.3) is plotted for $T_{4.4}$, which requires no Monte Carlo integration because Eq. (8.2e) holds exactly. We see that the power is increasing with $|\tilde{\epsilon}|$, which is expected, because H_0 and H_A are getting more and more different, cf. Fig. 1. Furthermore, the statistical test power is worse when h is small, which is also expected. Remember that $h = 10$ means that the moment of inertia of the points are only 10 times larger than the standard deviations $\sigma = 1$ of the target coordinates, which makes testing hypotheses nearly hopeless.

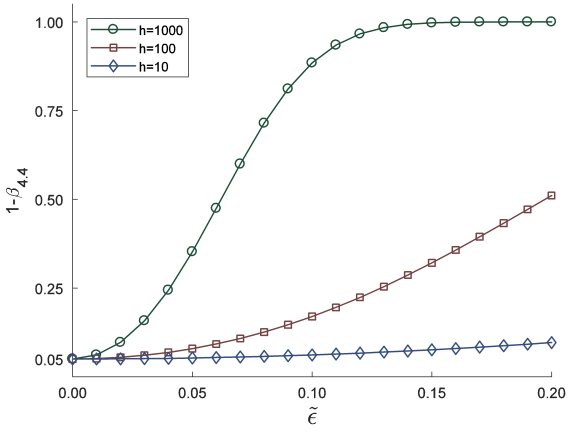

Figure 3: Power functions for $T_{4.4}$ and various values of h.

In Fig. 4-6 the other power functions Eq. (10.3) are plotted relative to that of $T_{4.4}$. A ratio > 1 means that T_i outperforms $T_{4.4}$ and vice versa. Test results with $\chi^2(1)$-quantiles are displayed by dotted curves and are denoted by $T(\chi^2)$, while those using true distributions computed by the Monte Carlo method are displayed by solid curves and are denoted by $T(\alpha)$. For $T_{3.4}$ only the solid curve makes sense.

In case of weak non-linearity, i.e. $h = 1000$, see Fig. 4, practically no difference is visible. All seven power

Table 1: Quantiles of $\chi^2(1)$(column 2) vs. quantiles of the true distribution of $T_i|H_0,\ i = 3.1,\ 3.2,\ 3.3$ (following columns)

	$T_{4.3} \equiv T_{4.4}$	$T_{3.1}$			$T_{3.2}$			$T_{3.3} \equiv T_{4.1} \equiv T_{4.2}$		
	$\chi^2(1-\alpha, 1)$	h=10	h=100	h=1000	h=10	h=100	h=1000	h=10	h=100	h=1000
α=0.10	2.71	2.99	2.73	2.71	1.93	2.63	2.70	2.99	2.73	2.71
α=0.05	3.84	4.46	3.89	3.85	2.28	3.69	3.83	4.38	3.89	3.85
α=0.02	5.41	6.78	5.51	5.42	2.46	5.12	5.38	6.40	5.51	5.42
α=0.01	6.63	8.92	6.78	6.64	2.49	6.19	6.58	8.07	6.77	6.64

Table 2: Quantiles of the true distribution of $T_{3.4}|H_0$

$T_{3.4}$	h=10	h=100	h=1000
α=0.10	2.79	2.71	2.71
α=0.05	3.96	3.85	3.84
α=0.02	5.59	5.43	5.41
α=0.01	6.86	6.65	6.63

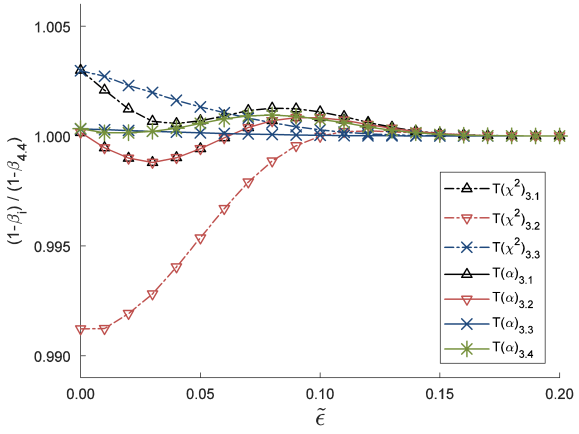

Figure 4: Power function ratios for h=1000 (weak non-linearity). Dotted curves: using $\chi^2(1)$-quantiles and are denoted by $T\ (\chi^2)$, solid curves: using true quantiles for critical values and are denoted by $T\ (\alpha)$. Black and red solid curves visually overlap.

functions behave equally well. In case of medium non-linearity, i.e. $h = 100$, there is also no great difference between the test statistics, except for $T_{3.2}$, when the $\chi^2(1)$-quantile is used (red dotted curve), see Fig. 5. The reason is that this approximate quantile ($c = 3.69$) differs much from the true value ($c = 3.84$). Otherwise, $\chi^2(1)$-quantiles are outperforming the true quantiles.

The strong non-linear case, i.e. $h = 10$, is depicted in Fig. 6. The differences between the tests are even amplified. Note, the different vertical scales in Fig. 4-6. When the $\chi^2(1)$-quantile $c = 3.84$ is used, $T_{3.2}$ is unable to reject a false H_0, no matter how large $|\tilde{\epsilon}|$ is (red dotted curve). This is a consequence of Eq. (8.11) and the price we have to pay that $\alpha = 0$ has been obtained in the preceding section.

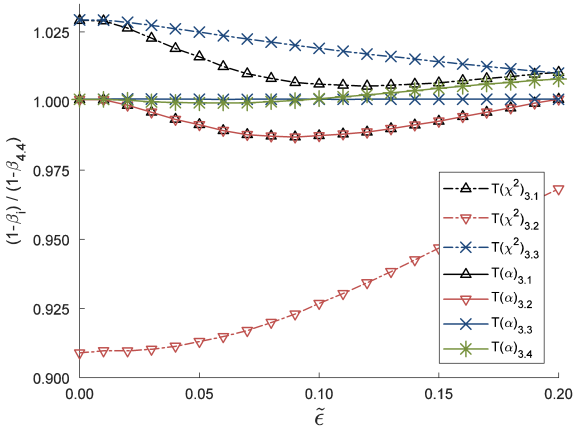

Figure 5: Power function ratios for h=100 (medium non-linearity). Dotted curves: using $\chi^2(1)$-quantiles and are denoted by $T\ (\chi^2)$, solid curves: using true quantiles for critical values and are denoted by $T\ (\alpha)$. Black and red solid curves visually overlap.

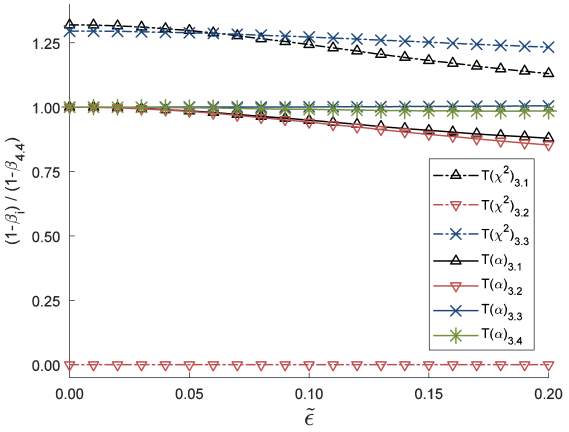

Figure 6: Power function ratios for h=10 (strong non-linearity). Dotted curves: using $\chi^2(1)$-quantiles and are denoted by $T\ (\chi^2)$, solid curves: using true quantiles for critical values and are denoted by $T\ (\alpha)$.

Due to the strong non-linearity, the power is again worst, if the true quantiles are applied. This behavior is expected, because a shift of the critical value c changes α and β in opposite directions, see Fig. 1. It follows that the increase

of probability of type 2 decision error corresponds to the loss of probability of type 1 decision error observed in the preceding section.

All solid curves are free of this effect, because they truly refer to $\alpha = 0.05$. This can easily be validated because for $\tilde{\epsilon} = 0$ the power is always equal. The only significant differences between the powers of T_i occur for strong non-linearity, so we will only focus on the case $h = 10$, see Fig. 6.

In the interval $0 < \tilde{\epsilon} < 0.2$ the best power is obtained for $T_{3.3}$, where the covariance propagation has been applied to Eq. (4.8a). This is even better than for the full non-linear test $T_{3.4}$ (green curve). But this advantage is very small and could be accidental. Remember that there is no mathematical proof that Eq. (3.6) outperforms Eq. (3.7). This has been demonstrated here. However, for values of $\tilde{\epsilon} > 0.4$ the situation changes, as is displayed in Fig. 7.

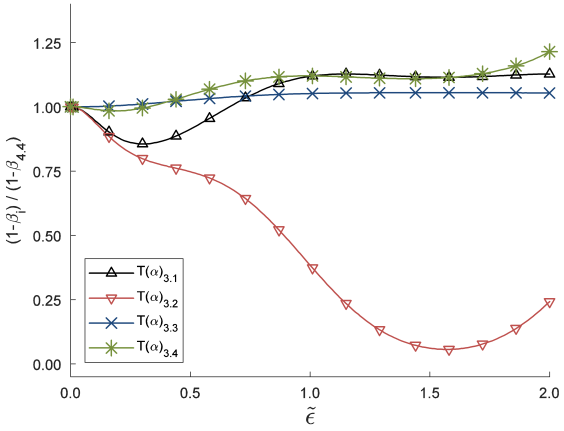

Figure 7: Same as Fig. 6, solid curves only, but larger range of $\tilde{\epsilon}$

Note that a comparison of $T_{3.i}$ vs. $T_{4.j}$ is less instructive, because if the scale is unknown, one should always use the four-parameter transformation, even though a test in a three-parameter transformation model may be more powerful.

Finally, note that the results in this section are not obtained from a "numerical experiment", but are strictly valid for all planar coordinate transformations with error-free coordinates in one coordinate system and the conventional assumption on the weights Eq. (4.3).

11 Conclusions

We have presented an analysis of the decision errors, when performing LR tests in planar coordinate transformation models. Several mathematical equivalent expressions are conceivable to apply the LR test to one specific hypothesis test Eq. (6.1), but different results are obtained.

At the end of section 3, we named three problems, which arise, if we apply the LR test to non-linear models in the usual way, which we now want to further comment on.

① The likelihood function Eq. (3.2) can only be maximized iteratively with the danger of finding only a local maximum. For problems like many transformations, which permit a unique analytical non-linear least squares solution like Eq. (4.8), this problem does not exist. The likelihood function has a unique maximum.

② Test statistic Eq. (3.7) gives an LR-Test only in the linearized GMM, i.e. not in the truly non-linear GMM. While Eq. (3.6) requires the minimization of Ω and Ω', Eq. (3.7) only relies on the minimization of Ω. $\min(\Omega') - \min(\Omega)$ is computed only by linear approximation. The consequence could be a small loss of statistical power of the test, depending on the degree of non-linearity. For the planar coordinate transformations with $\alpha = 0.05$ this has not always been found, not even for strong non-linearity. However, if α is chosen smaller, the differences between the power functions amplify.

③ The PDF of Eq. (3.6) or Eq. (3.7) does not belong to the well-known set of test distributions (t, χ^2, F etc.) such that the critical values must be computed numerically. This is usually not done, because it requires numerical effort. But using Monte Carlo integration it is simple, as has been demonstrated in section 9. The advantage would be that we effectively obtain the desired value of α. Otherwise, we found a shift of some probability from type 1 to type 2 decision error or back, which is undesired.

The same analytical computation can be done for other problems, for which explicit non-linear analytical least squares solutions exist. This encloses

- many other transformation problems, also 3D transformations (e.g. Grafarend and Awange 2003), also transformation where coordinates in both systems are error-affected (e.g. Chang 2015)
- many curve and surface fitting problems (e.g. Ahn 2005)

The four parameter transformation is an exceptional case, because it is intrinsically linear, but can be made non-linear by parameterization Eq. (5.1). The resulting non-

linear effects can be investigated easily by comparison with the linear model Eq. (5.2).
Also, more complex hypothesis tests can be studied in this way, e.g. in the framework of multiple outlier detection. The same approach can be applied to study other decision methods like model selection by information criteria, which has also been applied to transformations and other geodetic models (Lehmann 2014, 2015, Lehmann and Lösler 2016, 2017).

A Appendix 1: Analytical solution for the transformation with fixed scale parameter

The least squares error functional Eq. (3.3) to be minimized reads with Eq. (4.3)

$$\Omega(\hat{X}) = \sum_{i=1}^{N} p_i[(X_i - \hat{X}_0 - x_i \cdot \hat{\epsilon} + y_i \cdot \hat{\epsilon})^2 + (Y_i - \hat{Y}_0 - x_i \cdot \sin\hat{\epsilon} - y_i \cdot \hat{\epsilon})^2] = \min \tag{A.1}$$

Two necessary conditions for a minimum read

$$0 = \frac{\partial\Omega}{\partial\hat{X}_0} = -2\sum_{i=1}^{N} p_i \left(X_i - \hat{X}_0 - x_i \cdot \cos\hat{\epsilon} + y_i \cdot \sin\hat{\epsilon}\right)$$

$$0 = \frac{\partial\Omega}{\partial\hat{Y}_0} = -2\sum_{i=1}^{N} p_i \left(Y_i - \hat{Y}_0 - x_i \cdot \sin\hat{\epsilon} - y_i \cdot \cos\hat{\epsilon}\right)$$

Using $\Sigma p_i = 1$ gives estimates for the translation parameters:

$$\begin{aligned}\hat{X}_0 &= \sum_{i=1}^{N} p_i \left(X_i - x_i \cdot \cos\hat{\epsilon} + y_i \cdot \sin\hat{\epsilon}\right)\\ &= X_\star - x_\star \cdot \cos\hat{\epsilon} + y_\star \cdot \sin\hat{\epsilon}\\ \hat{Y}_0 &= \sum_{i=1}^{N} p_i \left(Y_i - x_i \cdot \sin\hat{\epsilon} - y_i \cdot \cos\hat{\epsilon}\right)\\ &= Y_\star - x_\star \cdot \sin\hat{\epsilon} - y_\star \cdot \cos\hat{\epsilon}\end{aligned}$$

Substitution $\hat{X}_0$, $\hat{Y}_0$ into the least squares error functional yields

$$\begin{aligned}\min = \Omega\left(\hat{X}\right) &= \sum_{i=1}^{N} p_i \Big[\left(\Delta X_i - \Delta x_i \cdot \cos\hat{\epsilon} + \Delta y_i \cdot \sin\hat{\epsilon}\right)^2\\ &+ \left(\Delta Y_i - \Delta x_i \cdot \sin\hat{\epsilon} - \Delta y_i \cdot \cos\hat{\epsilon}\right)^2\Big]\\ &= \sum_{i=1}^{N} p_i \Big[\Delta X_i^2 + \Delta Y_i^2 + \Delta x_i^2 + \Delta y_i^2\\ &-2\Delta X_i \left(\Delta x_i \cdot \cos\hat{\epsilon} - \Delta y_i \cdot \sin\hat{\epsilon}\right)\\ &-2\Delta Y_i \left(\Delta x_i \cdot \sin\hat{\epsilon} + \Delta y_i \cdot \cos\hat{\epsilon}\right)\Big]\\ &= h + H - 2\left(c \cdot \cos\hat{\epsilon} + s \cdot \sin\hat{\epsilon}\right)\end{aligned}$$

The third necessary condition for a minimum reads

$$0 = \frac{1}{2}\frac{\partial\Omega}{\partial\hat{\epsilon}} = c \cdot \sin\hat{\epsilon} - s \cdot \cos\hat{\epsilon}$$

This gives the estimate for the rotation parameter

$$\hat{\epsilon} = \arctan\frac{s}{c} = \arcsin\frac{s}{\sqrt{c^2+s^2}} = \arccos\frac{c}{\sqrt{c^2+s^2}}$$

This unique stationary point must be a minimum because Ω is bounded from below. The minimum is obtained at

$$\Omega\left(\hat{X}\right) = h + H - 2\frac{c^2+s^2}{\sqrt{c^2+s^2}} = h + H - 2\sqrt{c^2+s^2}$$

B Appendix 2: Analytical solution for the transformation with fixed rotation parameter

Similar to appendix 1, but with fixed rotation parameter ϵ_0 and with estimated scale parameter $\hat{\mu}$, we start from

$$\Omega\left(\hat{X}\right) = \sum_{i=1}^{N} p_i \Big[\left(X_i - \hat{X}_0 - x_i \cdot \hat{\mu} \cdot \cos\epsilon_0 + y_i \cdot \hat{\mu} \cdot \sin\epsilon_0\right)^2 + (Y_i - \hat{Y}_0 - x_i \cdot \hat{\mu} \cdot \sin\epsilon_0 - y_i \cdot \hat{\mu} \cdot \cos\epsilon_0)^2\Big] = \min$$

and obtain

$$\hat{X}_0 = X_\star - x_\star \cdot \hat{\mu} \cdot \cos\epsilon_0 + y_\star \cdot \hat{\mu} \cdot \sin\epsilon_0$$

$$\hat{Y}_0 = Y_\star - x_\star \cdot \hat{\mu} \cdot \sin\epsilon_0 - y_\star \cdot \hat{\mu} \cdot \cos\epsilon_0$$

Substitution $\hat{X}_0$, $\hat{Y}_0$ into the least squares error functional yields

$$\begin{aligned}\min = \Omega\left(\hat{X}\right) &= \\ &\sum_{i=1}^{N} p_i \Big[\left(\Delta X_i - \Delta x_i \cdot \hat{\mu} \cdot \cos\epsilon_0 + \Delta y_i \cdot \hat{\mu} \cdot \sin\epsilon_0\right)^2\\ &+\left(\Delta Y_i - \Delta x_i \cdot \hat{\mu} \cdot \sin\epsilon_0 - \Delta y_i \cdot \hat{\mu} \cdot \cos\epsilon_0\right)^2\Big]\\ &= \hat{\mu}^2 \cdot h + H - 2\left(c \cdot \hat{\mu} \cdot \cos\epsilon_0 + s \cdot \hat{\mu} \cdot \sin\epsilon_0\right)\end{aligned}$$

The third necessary condition for a minimum reads

$$0 = \frac{1}{2}\frac{\partial\Omega}{\partial\hat{\mu}} = \hat{\mu} \cdot h - (c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0)$$

This gives the estimate for the scale parameter

$$\hat{\mu} = \frac{c \cdot \cos\epsilon_0 + s \cdot \sin\epsilon_0}{h}$$

This unique stationary point must be a minimum because Ω is bounded from below. The minimum is obtained at

$$\begin{aligned}\Omega\left(\hat{X}\right) &= \frac{(c\cdot\cos\epsilon_0 + s\cdot\sin\epsilon_0)^2}{h}\\ &+ H - 2\frac{c\cdot\cos\epsilon_0 + s\cdot\sin\epsilon_0}{h}(c\cdot\cos\epsilon_0 + s\cdot\sin\epsilon_0)\\ &= H - \frac{(c\cdot\cos\epsilon_0 + s\cdot\sin\epsilon_0)^2}{h}\end{aligned}$$

C Appendix 3: Analytical solution for the four-parameter transformation

Following the line of appendix 2, but replacing ϵ_0 by $\hat{\epsilon}$ gives

$$\hat{\mu} = \frac{c\cdot\cos\hat{\epsilon} + s\cdot\sin\hat{\epsilon}}{h}$$

$$\Omega\left(\hat{X}\right) = H - \frac{\left(c\cdot\cos\hat{\epsilon} + s\cdot\sin\hat{\epsilon}\right)^2}{h}$$

Now also minimizing $\Omega\left(\hat{X}\right)$ for $\hat{\epsilon}$ yields a fourth necessary condition

$$0 = \frac{1}{2}\frac{\partial\Omega}{\partial\hat{\epsilon}} = \frac{1}{h}\left(c\cdot\cos\hat{\epsilon} + s\cdot\sin\hat{\epsilon}\right)\left(c\cdot\sin\hat{\epsilon} - s\cdot\cos\hat{\epsilon}\right)$$

At least one of the factors must be zero, therefore we obtain two solutions

$$\hat{\epsilon}_1 = \arctan\frac{s}{c}, \quad \hat{\epsilon}_2 = \arctan\left(-\frac{c}{s}\right)$$

but the second solution obviously belongs to a maximum of Ω and is dropped. We thus arrive at

$$\hat{\epsilon} = \arctan\frac{s}{c}$$

Inserting this for $\hat{\mu}$ and $\Omega\left(\hat{X}\right)$ gives

$$\hat{\mu} = \frac{\sqrt{c^2+s^2}}{h}$$

$$\Omega\left(\hat{X}\right) = H - \frac{\sqrt{c^2+s^2}}{h}$$

D Appendix 4: Covariance matrix of the linearized GMM of the three-parameter transformation

The normal matrix of the linearized GMM with A in Eq. (6.4) is of the form

$$A^TPA = \begin{pmatrix} u & v & w\\ v & 1 & 0\\ w & 0 & 1\end{pmatrix}$$

with

$u := \Sigma p_i\left(x_i^2 + y_i^2\right)$, $v := -x_\star\cdot\sin\hat{\epsilon} - y_\star\cdot\cos\hat{\epsilon}$, $w := x_\star\cdot\cos\hat{\epsilon} - y_\star\cdot\sin\hat{\epsilon}$.

The corresponding inverse can be readily written down:

$$\begin{aligned}\left(A^TPA\right)^{-1} &= \frac{1}{u - v^2 - w^2}\begin{pmatrix} 1 & -v & -w\\ -v & u - w^2 & vw\\ -w & vw & u - v^2\end{pmatrix}\\ &= \frac{1}{h}\begin{pmatrix} 1 & -v & -w\\ -v & h + v^2 & vw\\ -w & vw & h + w^2\end{pmatrix}\end{aligned}$$

because $u - v^2 - w^2 = -x_\star^2 - y_\star^2 + \Sigma p_i\left(x_i^2 + y_i^2\right) = h$.

E Appendix 5: Covariance matrix of the linearized GMM of the four-parameter transformation

The normal matrix of the linearized GMM with A in Eq. (7.6) is of the form

$$A^TPA = \begin{pmatrix} \hat{\mu}^2\cdot u & 0 & \hat{\mu}\cdot v & \hat{\mu}\cdot w\\ 0 & u & w & -v\\ \hat{\mu}\cdot v & w & 1 & 0\\ \hat{\mu}\cdot w & -v & 0 & 1\end{pmatrix}$$

with u, v, w as in appendix 4.

The corresponding inverse can be readily written down:

$$\left(A^TPA\right)^{-1} = \frac{1}{h}\begin{pmatrix} \hat{\mu}^{-2} & 0 & -\hat{\mu}^{-1}v & -\hat{\mu}^{-1}w\\ 0 & 1 & -w & v\\ -\hat{\mu}^{-1}v & -w & u & 0\\ -\hat{\mu}^{-1}w & v & 0 & u\end{pmatrix}$$

where $u - v^2 - w^2 = h$ has been used (see appendix 4).

References

Ahn S.J., 2005, Least squares orthogonal distance fitting of curves and surfaces in space. Lecture Notes in Computer Science (LNCS), 3151, Springer, Heidelberg, ISBN 3-540-23966-9.

Chang G., 2015, On least-squares solution to 3D similarity transformation problem under Gauss–Helmert model. J Geod., 89, 6, 573–576, DOI 10.1007/s00190-015-0799-z.

Grafarend E.W., Awange J.L., 2003, Non-linear analysis of the threedimensional datum transformation [conformal group C7(3)]. J Geod., 77, 1-2, 66–76, DOI 10.1007/s00190-002-0299-9

Kargoll B., 2012, On the Theory and Application of Model Misspecification Tests in Geodesy. Deutsche Geodätische Kommission Reihe C, Nr. 674, München.

Klein I., Matsuoka M.T., Guzatto M.P., Nievinski F.G., 2017, An approach to identify multiple outliers based on sequential likelihood ratio tests. Survey Review, 49, 357, 1-9, DOI 10.1080/00396265.2016.1212970.

Koch K.R., 1999, Parameter estimation and hypothesis testing in linear models. 2nd edn., Springer, Heidelberg, DOI 10.1007/978-3-662-03976-2.

Lehmann R., 2012, Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. J Geod., 86, 16, 1137-1146, DOI 10.1007/s00190-012-0569-0

Lehmann R., 2014, Transformation model selection by multiple hypothesis testing. J Geod., 88, 12, 1117-1130, DOI 10.1007/s00190-014-0747-3.

Lehmann R., 2015, Observation error model selection by information criteria vs. normality testing. Stud. Geophys. Geod., 59, 4, 489-504, DOI 10.1007/s11200-015-0725-0.

Lehmann R., Lösler M., 2016, Multiple outlier detection: hypothesis tests versus model selection by information criteria. J Surv. Eng., 142, 4, DOI 10.1061/(ASCE)SU.1943-5428.0000189.

Lehmann R., Lösler M., 2017, Congruence analysis of geodetic networks – hypothesis tests versus model selection by information criteria. J Appl. Geodesy, 11, 4, 271-283, DOI 10.1515/jag-2016-0049.

Lehmann R., Neitzel F., 2013, Testing the compatibility of constraints for parameters of a geodetic adjustment model. J Geod., 87, 6, 555-566, DOI 10.1007/s00190-013-0627-2.

Lehmann R., Voß-Böhme A., 2017, On the statistical power of Baarda's outlier test and some alternative. J Geod. Sci., 7, 1, 68-78, DOI:10.1515/jogs-2017-0008.

Lösler M., Lehmann R., Eschelbach C., 2017, Model selection via akaike information criterion – Application in Congruence Analysis (in German). Allgemeine Vermessungs-Nachrichten, 124, 5, 137-145.

Neyman J., Pearson E.S., 1933, On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 231, 694–706, 289–337, DOI 10.1098/rsta.1933.0009.

Somogyi J., Kalmár J., 1988, Verschiedene robuste Schätzungsverfahren für die Helmerttransformation. Allgemeine Vermessungs-Nachrichten, 95, 4, 141-146.

Tanizaki H., 2004, Computational methods in statistics and econometrics, Marcel Dekker New York, ISBN-13: 978-0824748043.

Teunissen P.J.G., 1986, Adjusting and testing with the models of the affine and similarity transformations. Manuscr. Geod., 11, 214-225.

Teunisszen P.J.G., 1985, The geometry of geodetic inverse linear mapping and non-linear adjustment. Netherlands Geodetic Commission, Publications on Geodesy, New Series, Delft, 1–186.

Teunissen P.J.G., 2000, Testing theory; an introduction. 2nd edition, Series on Mathematical Geodesy and Positioning, Delft University of Technology, The Netherlands, ISBN 90-407-1975-6.

Wolf H., 1966, Die Genauigkeit der für eine Helmert-Transformation berechneten Koordinaten. Zeitschrift für Vermessungswesen, 91, 2, 33-34.

Velsink H., 2015, On the deformation analysis of point fields. J Geod., 89, 11, 1071-1087, DOI 10.1007/s00190-015-0835-z.