

# Transformation model selection by multiple hypotheses testing

Rüdiger Lehmann

Received: 4 November 2013 / Accepted: 1 July 2014 / Published online: 24 July 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** Transformations between different geodetic reference frames are often performed such that first the transformation parameters are determined from control points. If in the first place we do not know which of the numerous transformation models is appropriate then we can set up a multiple hypotheses test. The paper extends the common method of testing transformation parameters for significance, to the case that also constraints for such parameters are tested. This provides more flexibility when setting up such a test. One can formulate a general model with a maximum number of transformation parameters and specialize it by adding constraints to those parameters, which need to be tested. The proper test statistic in a multiple test is shown to be either the extreme normalized or the extreme studentized Lagrange multiplier. They are shown to perform superior to the more intuitive test statistics derived from misclosures. It is shown how model selection by multiple hypotheses testing relates to the use of information criteria like AICc and Mallows'  $C_p$ , which are based on an information theoretic approach. Nevertheless, whenever comparable, the results of an exemplary computation almost coincide.

**Keywords** Coordinate transformation · Hypothesis test · Affine transformation · Similarity transformation · Gauss–Markov model with constraints · Normalized Lagrange multiplier · Studentized Lagrange multiplier · Akaike information criterion · Mallows'  $C_p$

## 1 Introduction

Today, geodetic information is often expressed with respect to various geodetic reference frames. To convert such information (like coordinates of points) from one frame to another, coordinate transformations are widely needed and applied in all branches of the modern geodetic profession. Those fields of application range from satellite navigation (e.g. Zhang et al 2012) to cadastral surveying (e.g. Deakin 1998, 2007) and to photogrammetry (e.g. Goktepe and Kocaman 2010). The most important theoretical and practical problems in these fields are solved.

Nonetheless, there are many new developments regarding the computation of coordinate transformation parameters from a set of points with given coordinates in two different reference frames, known as control points (synonymously referred to as identical or homologous points):

- There are new results regarding the transformation accuracy. Lehmann (2010) analyzes why and under which conditions the accuracy of transformed points is optimal in the barycenter of the control points.
- Nowadays, the adjustment of transformation parameters is often computed by robust methods (Kampmann 1996; Kanani 2000; Carosio et al 2006; Ge et al. 2013). Here, it is shown that outliers in coordinates of control points are less influential to the transformation parameters.
- A more recent development is the total least squares approach (Schaffrin and Felus 2008; Mahboub 2012). Here, errors are not only assigned to the coordinates of control points, but also to the elements of the system matrix. However, based on the example of a planar similarity transformation Neitzel (2010) shows that the total least squares solution can be obtained easily from a rigorous evaluation of the Gauss–Helmert model.

---

R. Lehmann (✉)  
Faculty of Spatial Information, University of Applied Sciences Dresden,  
Friedrich-List-Platz 1, 01069 Dresden, Germany  
e-mail: r.lehmann@htw-dresden.de

In the following, we restrict ourselves to three-dimensional (3D) reference frames. In geodesy, we use a variety of 3D transformation models. The most important are (Andrei 2006):

- translation (three parameters)
- rotation (three parameters)
- rotation and translation (six parameters)
- rotation, one scalation and translation (similarity transformation, seven parameters)
- rotation, two scalations and translation (eight parameters)
- rotation, three scalations and translation (nine parameters)
- affine transformation (12 parameters)

Even more complex transformation models are used in geodesy, e.g., transformations using thin plate splines (Donato and Belongie 2002). Alternatively, we find piecewise approaches (Lippus 2004), where the region covered by control points is partitioned. In each partition, the transformation model is computed independently and later the pieces are glued together into a single transformation.

In this paper, we are concerned with the following problem: Given a sufficiently large number of control points, which transformation model should be selected. This problem belongs to a class of problems, which is often referred to as model selection (Burnham and Anderson 2002). Often a strong preference for a transformation model can be deduced from the relationship between the reference frames. For example, if they are strictly related by a conformal mapping then the similarity transformation is a proper choice. But often there is no such preference, and the proper model must be selected by means of the control points.

If one model contains some extra parameters with respect to the other, like rotation and translation versus similarity transformation, the standard geodetic approach is to test statistically, if the additional parameters are significant, i.e., if the estimated values of these parameters are significantly different from zero. This test is often called  $\nu$  test in geodesy (Teunissen 2000). Such an approach is, e.g., used in Andrei (2006). In the case of the rotation and translation versus the similarity transformation the task would be to test if the scale parameter is significantly different from unity, which can be done by the  $\nu$  test.

But often this is not possible because one model is set up with completely different parameters than another. For example, the parameters of the spatial affine transformation are generally not scales and rotation angles. Therefore, a comparison to the similarity transformation cannot be done by testing the significance of some extra parameters. Making the  $\nu$  test applicable would require (if possible at all) an unnatural change of parameterization. In other cases, we must select one out of more than two transformation models.

In our paper, we basically follow the hypothesis testing approach, but extend it, such that not only parameters themselves can be tested, but also constraints on parameters. It will turn out that this case arises, e.g. in the problem of transformation model selection. Moreover, it is necessary to understand that we are performing a multiple hypotheses test, which adheres to other laws than the classical hypothesis test (Miller 1981). This is not always properly understood and leads to misconceptions. For example, the significance test of the transformation model parameters performed by Ziggah et al. (2013) should have been such a multiple test.

Other approaches to model selection used in statistics are based on information criteria (Burnham and Anderson 2002). The idea is that more complex models can generally fit the data better, but this may result in overfitting, i.e., unduly complex models partly fit the observation errors. Therefore, pure goodness of fit is not a valid criterion for model selection, but a penalty term for model complexity needs to be introduced. The most important information criteria are

- Akaike information criterion (AIC), see Akaike (1974),
- its alternate version (AICc, which means AIC with a correction for small data sets),
- Bayesian information criterion (BIC) and
- Mallows'  $C_p$ , see Mallows (1973).

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC or AICc or BIC value, or the one with  $C_p$  value approaching the number of model parameters. The AICc has been used for the transformation model selection by Felus and Felus (2009) because it is recommended for small sets of observations. We will return to this approach in Sect. 8.

The paper is organized as follows: after introducing transformation equations and constraints for common spatial coordinate transformations we set up a Gauss–Markov model (GMM) with constraints and solve the model selection problem by a multiple hypotheses test. Following Lehmann and Neitzel (2013), it is shown that the proper test statistics for such a test are the extreme normalized or externally studentized Lagrange multipliers (LMs, also known as correlates in geodesy). We emphasize that the use of normalized LMs as test statistics in geodesy goes back to Teunissen (1985). In a numerical example, it is shown how these tests work and how they are superior to more intuitively defined test statistics. Finally, we comment on the relationship of this model selection strategy in relation to information criteria, exemplified for AICc.

## 2 Transformation equations and constraints

Given a number of points with coordinates in two different spatial reference frames (known as control points), the prob-

lem is to find a good model for the transformation between these two frames. The given coordinates may be affected by random observation errors with known stochastic properties.

We start from the 3D affine transformation, which obeys the following system of equations:

$$\begin{pmatrix} \xi_b \\ \eta_b \\ \zeta_b \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} + \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \begin{pmatrix} \xi_a \\ \eta_a \\ \zeta_a \end{pmatrix} =: t + T \begin{pmatrix} \xi_a \\ \eta_a \\ \zeta_a \end{pmatrix} \tag{1}$$

Here  $\xi_a, \eta_a, \zeta_a$  denote the coordinates of a point in the initial reference frame and  $\xi_b, \eta_b, \zeta_b$  denote the coordinates of the same point in the target reference frame. This transformation model has 12 transformation parameters  $t_1, t_2, t_3, t_{11}, \dots, t_{33}$ . All other relevant models can be derived therefrom by imposing constraints on these 12 parameters as follows:

A nine-parameter transformation is derived by imposing orthogonality of the rows of  $T$ . In other words, the elements of  $T$  need to fulfill the following three constraints (e.g. Andrei 2006, chapter 2):

$$t_{11}t_{21} + t_{12}t_{22} + t_{13}t_{23} = 0 \tag{2}$$

$$t_{11}t_{31} + t_{12}t_{32} + t_{13}t_{33} = 0 \tag{3}$$

$$t_{21}t_{31} + t_{22}t_{32} + t_{23}t_{33} = 0 \tag{4}$$

In this special case, we can express (1) as

$$\begin{pmatrix} \xi_b \\ \eta_b \\ \zeta_b \end{pmatrix} = t + \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix} R \begin{pmatrix} \xi_a \\ \eta_a \\ \zeta_a \end{pmatrix} \tag{5}$$

where

$$\mu_1 = \sqrt{t_{11}^2 + t_{12}^2 + t_{13}^2}$$

$$\mu_2 = \sqrt{t_{21}^2 + t_{22}^2 + t_{23}^2}$$

$$\mu_3 = \sqrt{t_{31}^2 + t_{32}^2 + t_{33}^2}$$

are scale factors and  $R$  is a rotation matrix (orthogonal matrix with determinant 1). A way of expressing this transformation by nine parameters is by  $t_1, t_2, t_3, \mu_1, \mu_2, \mu_3$  and by three Eulerian rotation angles. Another nine-parameter transformation can be defined by requiring the columns of  $T$  to be orthogonal, rather than the rows. Here the diagonal matrix of scale factors and  $R$  are interchanged. Both transformations are essentially different.

An eight-parameter transformation is practically less important. It is motivated by the fact that horizontal coordinates are sometimes determined by different technologies than vertical coordinates. Therefore, the scales in horizontal directions may be equal, i.e.,  $\mu_1 = \mu_2$ , while  $\mu_3$  is kept as an independent parameter (e.g. Andrei 2006, chapter 3):

$$\begin{pmatrix} \xi_b \\ \eta_b \\ \zeta_b \end{pmatrix} = t + \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_1 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix} R \begin{pmatrix} \xi_a \\ \eta_a \\ \zeta_a \end{pmatrix} \tag{6}$$

The specialization from (5) to (6) is equivalent to the constraint  $\mu_1 = \mu_2$ , but since we want to impose this constraint on the model (1), it must be expressed in terms of the transformation parameters of this model:

$$t_{11}^2 + t_{12}^2 + t_{13}^2 - t_{21}^2 - t_{22}^2 - t_{23}^2 = 0 \tag{7}$$

Also in (6), the diagonal matrix of scale factors and  $R$  may be interchanged, yielding a different transformation model with eight parameters.

The seven-parameter similarity transformation (in geodesy also known as spatial Helmert transformation) is obtained by requiring all scales to be equal, i.e.,  $\mu := \mu_1 = \mu_2 = \mu_3$ . The system of transformation equations now reads (e.g. Andrei 2006, section 1.2):

$$\begin{pmatrix} \xi_b \\ \eta_b \\ \zeta_b \end{pmatrix} = t + \mu R \begin{pmatrix} \xi_a \\ \eta_a \\ \zeta_a \end{pmatrix} \tag{8}$$

Here  $\mu$  and  $R$  can obviously be interchanged without changing the transformation. To restrict the eight-parameter transformation (6) to the similarity transformation (8), a further constraint must be added. There are several equivalent possibilities, how such a constraint could read. We favor the following:

$$t_{31}^2 + t_{32}^2 + t_{33}^2 - (t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2) / 2 = 0 \tag{9}$$

The reason is that this constraint makes sense even without (7):  $\mu_3$  equals the quadratic mean of  $\mu_1$  and  $\mu_2$ , or equivalently,  $\mu_3^2$  equals the mean of  $\mu_1^2$  and  $\mu_2^2$ . This is instructive even if  $\mu_1$  and  $\mu_2$  are different: a transformation using (9) without (7) deforms a sphere to an ellipsoid with one axis length being the quadratic mean of the other two axis lengths.

The spatial rotation and translation is obtained by requiring  $\mu = 1$ :

$$\begin{pmatrix} \xi_b \\ \eta_b \\ \zeta_b \end{pmatrix} = t + R \begin{pmatrix} \xi_a \\ \eta_a \\ \zeta_a \end{pmatrix} \tag{10}$$

To restrict the similarity transformation (8) to this transformation, we favor the following form of the constraint  $\mu = 1$ :

$$t_{31}^2 + t_{32}^2 + t_{33}^2 + t_{11}^2 + t_{12}^2 + t_{13}^2 + t_{21}^2 + t_{22}^2 + t_{23}^2 = 3 \tag{11}$$

The reason is that this constraint makes senses even without (7) and (9): the quadratic mean of  $\mu_1, \mu_2$  and  $\mu_3$  equals unity,

or equivalently the mean of  $\mu_1^2$ ,  $\mu_2^2$  and  $\mu_3^2$  equals unity. This is instructive even if  $\mu_1, \mu_2, \mu_3$  are different: a transformation using (11) without (7) and (9) deforms a sphere to an ellipsoid such that the space diagonal of the bounding cuboid preserves length.

The pure *spatial rotation* is derived by additionally requiring that

$$t_1 = 0, t_2 = 0, t_3 = 0$$

The pure *spatial translation* is obtained by requiring  $R$  to be the unit matrix. This could be achieved by additionally imposing the constraints

$$t_{11} = 1, t_{22} = 1, t_{33} = 1$$

(Remember that a rotation matrix with 1,1,1 on the main diagonal is uniquely determined as the unit matrix). But since such transformation models are rarely used in geodesy, the last two sets of constraints will not be used in the sequel.

If we want to decide if the general transformation like the affine transformation (1) is the correct model or a more special one like the similarity transformation (8), then we have to decide, whether the constraints restricting the generality of (1)–(8) are compatible with the given coordinates of the control points or not. Due to inevitable observation errors we can in general not expect the estimated parameters to fulfill such constraints exactly. But if the constraints show only small misclosures then we may assume that the special model is sufficient to represent the relationship between both reference frames. It remains to be shown how smallness of misclosures or a similar criterion is to be defined.

### 3 Gauss–Markov model

The general transformation problem can be formulated as a non-linear GMM

$$Y = \mathcal{A}(X) - e \tag{12}$$

where  $Y$  is a  $n$ -vector of given coordinates of control points and  $X$  is a  $u$ -vector of unknown transformation model parameters, augmented by some unknown true values of coordinates (see below).  $\mathcal{A}$  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:

$$e \sim N(0, \sigma^2 P^{-1}) \tag{13}$$

$P$  is a known  $n \times n$ -matrix of weights (weight matrix).  $\sigma^2$  is the a priori variance factor, which may be either known or unknown.

It is also customary to formulate coordinate transformations as a Gauss–Helmert model. But this model can be transformed into the GMM by the simple variable substitution given in (Koch 1999, p.212). When testing the compatibility of constraints for parameters, it is better to restrict ourselves to GMM because then we can immediately use the results given by Lehmann and Neitzel (2013).

For the 3D affine transformation (1), the non-linear observation equations associated with a control point having six observed coordinates read

$$\begin{aligned} \xi_a &= \xi_a^{\text{true}} - e_{\xi_a} \\ \eta_a &= \eta_a^{\text{true}} - e_{\eta_a} \\ \zeta_a &= \zeta_a^{\text{true}} - e_{\zeta_a} \\ \xi_b &= t_1 + t_{11} \cdot \xi_a^{\text{true}} + t_{12} \cdot \eta_a^{\text{true}} + t_{13} \cdot \zeta_a^{\text{true}} - e_{\xi_b} \\ \eta_b &= t_2 + t_{21} \cdot \xi_a^{\text{true}} + t_{22} \cdot \eta_a^{\text{true}} + t_{23} \cdot \zeta_a^{\text{true}} - e_{\eta_b} \\ \zeta_b &= t_3 + t_{31} \cdot \xi_a^{\text{true}} + t_{32} \cdot \eta_a^{\text{true}} + t_{33} \cdot \zeta_a^{\text{true}} - e_{\zeta_b} \end{aligned} \tag{14}$$

Let there be given  $p$  control points. Then  $n = 6p$  and the vector of observations reads

$$Y = (\dots, \xi_a, \eta_a, \zeta_a, \xi_b, \eta_b, \zeta_b, \dots)^T \tag{15}$$

Also  $u = 3p + 12$  and the vector of GMM parameters reads

$$X = (\dots, \xi_a^{\text{true}}, \eta_a^{\text{true}}, \zeta_a^{\text{true}}, \dots, t_1, t_2, t_3, t_{11}, t_{12}, \dots, t_{32}, t_{33})^T \tag{16}$$

Note the difference between transformation parameters and GMM parameters. The latter set also comprises the unknown true values of coordinates in the initial frame.

$\mathcal{A}$  is clearly a non-linear operator here. (The affine transformation model would be immediately linear only if all  $\xi_a, \eta_a, \zeta_a$  could be treated as error-free.) In the sequel, we exclude singular configurations of the control points (like coplanarity), such that all parameters can be uniquely determined in the unconstrained GMM.

The  $m$  constraints restricting the general transformation model to the special one can be formulated as

$$\mathcal{B}(X) = \ell \tag{17}$$

where  $\mathcal{B}$  is a generally non-linear operator mapping the unknown GMM parameter vector  $X$  to a known  $m$ -vector  $\ell$ . In fact, due to the non-linearity of the constraints (2), (3), (4), (7), (9), (11) the operator  $\mathcal{B}$  related to the affine transformation problem is non-linear.

If the special transformation does not describe the relationship between the reference frames correctly then we get true misclosures

$$\widehat{W} := \mathcal{B}(X) - \ell \tag{18}$$

After computing an estimate  $\hat{X}$  of the parameters of the unconstrained GMM we can insert  $\hat{X}$  into the constraints and come up with estimated misclosures:

$$\widehat{W} := \mathcal{B}(\hat{X}) - \ell \quad (19)$$

Due to inevitable observation errors we get in general  $\widehat{W} \neq 0$  even if  $W = 0$  holds. For example, the misclosures related to constraints (2), (3), (4) can be interpreted as the sines of the three shear angles related to the affine transformation.

#### 4 Hypothesis test in the non-linear model

In geodesy, the decision problem on the proper transformation model is generally posed as a statistical hypothesis test. Opposing the special model represented by the GMM (12), (13) augmented by constraints (17) to a general model represented by the unconstrained GMM is equivalent to opposing the null hypothesis

$$H_0 : W = 0 \quad (20a)$$

to the alternative hypothesis

$$H_A : W \neq 0. \quad (20b)$$

If  $H_0$  is to be rejected then we decide on the general model, otherwise the special model is used for the transformation between the given reference frames.

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 I decision error  $\alpha$  (significance level) is suitably defined (say 0.01 or 0.05 or 0.10).
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

$$T(Y) := \|\widehat{W}\| \quad (21)$$

with some suitable norm  $\|\bullet\|$  are conceivable. Although the statistical power (probability of rejection of  $H_0$  when it is false) of such a test might not be optimal or even poor.

Consider for example the problem of opposing the affine transformation model (1) with the nine-parameter transformation model (6).  $\widehat{W}$  would be the vector of sines of the shear angles computed from estimated affine transformation parameters  $\hat{t}_{11}, \hat{t}_{12}, \dots, \hat{t}_{32}, \hat{t}_{33}$ . A possible test statistic would be the RMS or maximum absolute value of these estimated misclosures. In some instances, the misclosure is directly interpreted as a deviation of a parameter from a fixed value: Consider for example the problem of opposing the similarity transformation model (8) with the rotation and translation model (10). The effective constraint is  $\mu = 1$  here, and we can test if the estimated parameter  $\hat{\mu}$  in (8) is significantly different from unity. It is here equivalent to (21) because  $\hat{\mu} - 1$  is nothing but the estimated misclosure (19) of the related constraint  $\mu = 1$ .

In geodesy, we most often apply the likelihood ratio (LR) test (e.g. Tanizaki 2004, p. 54 ff). The rationale of the LR test is provided by the famous Neyman–Pearson lemma (1933), which demonstrates that under various assumptions such a test has the highest 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 (Teunissen 2000; Kargoll 2012).

Moreover, we can oppose the general model to a set of special models in parallel. This is equivalent to opposing  $H_0$  in (20a) to a set of multiple alternative hypotheses  $H_{A1}, H_{A2}, \dots, H_{Am}$ . Each of them proposes that only a subset of constraints is violated, or equivalently, a subset of elements of  $W$  is non-zero. In this way, we come up with a multiple hypotheses test. It is performed by testing  $H_0$  vs.  $H_{A1}$ ,  $H_0$  vs.  $H_{A2}$ ,  $\dots$ ,  $H_0$  vs.  $H_{Am}$ , and  $H_0$  is rejected if and only if it is rejected in any of the  $m$  tests. However, these tests are not performed with quantile probability  $\alpha$ , but  $\alpha/m$ . The rationale for this is that the producers' risk must be portioned to  $m$  alternative hypotheses. Nonetheless, this treatment is fully valid only if the  $m$  test statistics are statistically independent, which is often violated. Lehmann (2012) shows how to improve this in the case of geodetic outlier detection. A recommendable textbook on this topic is (Miller 1981).

Consider for example the case that we want to test if

- $H_0$  : the rotation and translation (10) is the correct model or
- $H_{A1}$  : the similarity transformation model (8) or
- $H_{A2}$  : the eight-parameter transformation model (6) or
- $H_{A3}$  : the nine-parameter transformation model (5) or
- $H_{A4}$  : the affine transformation model (1).

We set up the observation equations of the affine transformation and augment the resulting GMM by the constraints (2), (3), (4), (7), (9), (11).  $H_{A1}$  is then the hypothesis that (11) is in conflict with the observations and the rest of the constraints.  $H_{A2}$  is then the hypothesis that (9) and (11) are in conflict with the observations and (2), (3), (4), (7).  $H_{A3}$  is now the hypothesis that (7), (9), (11) are in conflict with the observations and (2), (3), (4) and finally  $H_{A4}$  is the hypothesis that all six constraints are in conflict with the observations (i.e., they produce true misclosures  $W \neq 0$ ).

**5 Linearization**

In general non-linear models the desired CDF  $F(T|H_0)$  cannot be analytically derived. Not even the CDF of  $\hat{X}$  can be analytically derived here. A numerical technique for deriving such distributions is the Monte Carlo method, but it is often computationally costly.

The family of normal distributions enjoys the famous property of constituting a family of stable distributions, i.e., linear combinations of normal random variables are also normally distributed. Therefore, in a linear model, where  $\mathcal{A}$ ,  $\mathcal{B}$  are linear operators, the distributions of  $\hat{X}$ ,  $\hat{W}$  etc. are known to be normal too. However, if the non-linear model is somehow close to a linear model, then the relevant distributions are still somehow close to normal. (Otherwise the representation of the solution by the estimate  $\hat{X}$  and possibly a covariance matrix associated with it would be meaningless).

The common procedure in this case is to introduce approximate parameters  $X^0$ , e.g. by solving 12 selected affine transformation equations neglecting observation errors. Then we get with  $x := X - X^0$ ,  $y := Y - \mathcal{A}(X^0)$ ,  $b := \ell - \mathcal{B}(X^0)$  the linearized GMM

$$y = Ax - e \tag{22}$$

with linearized constraints

$$B^T x = b \tag{23}$$

$A$  and  $B^T$  denote the Jacobian matrices of  $\mathcal{A}$  and  $\mathcal{B}$  at  $X^0$ . Transposition of  $B$  is introduced here to come close to the standard geodetic notation, also used by Lehmann and Neitzel (2013).

The linearized true misclosures now read

$$w := B^T x - b \tag{24}$$

In a 3D affine transformation it is customary to use

$$X^0 := (\dots, \xi_a, \eta_a, \zeta_a, \dots, t_1^0, t_2^0, t_3^0, t_{11}^0, t_{12}^0, \dots, t_{32}^0, t_{33}^0)^T \tag{25}$$

where  $t_1^0, t_2^0, t_3^0, t_{11}^0, t_{12}^0, \dots, t_{32}^0, t_{33}^0$  are computed from four selected control points (not coplanar).

The resulting linearized observation equations (14) and linearized constraints (2), (3), (4), (7), (9), (11) in the form of (22), (23) are built of:

$$y = \begin{pmatrix} \vdots \\ 0 \\ 0 \\ 0 \\ \xi_b - t_1^0 - t_{11}^0 \cdot \xi_a - t_{12}^0 \cdot \eta_a - t_{13}^0 \cdot \zeta_a \\ \xi_b - t_2^0 - t_{21}^0 \cdot \xi_a - t_{22}^0 \cdot \eta_a - t_{23}^0 \cdot \zeta_a \\ \xi_b - t_3^0 - t_{31}^0 \cdot \xi_a - t_{32}^0 \cdot \eta_a - t_{33}^0 \cdot \zeta_a \\ \vdots \end{pmatrix} \tag{26}$$

$$A = \begin{pmatrix} \vdots \\ \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & t_{11}^0 & t_{12}^0 & t_{13}^0 & \dots & 1 & 0 & 0 & \xi_a & \eta_a & \zeta_a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & t_{21}^0 & t_{22}^0 & t_{23}^0 & \dots & 0 & 1 & 0 & 0 & 0 & 0 & \xi_a & \eta_a & \zeta_a & 0 & 0 & 0 & 0 \\ \dots & t_{31}^0 & t_{32}^0 & t_{33}^0 & \dots & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_a & \eta_a & \zeta_a \\ \vdots \end{pmatrix} \tag{27}$$

$$b = \begin{pmatrix} -t_{11}^0 t_{21}^0 - t_{12}^0 t_{22}^0 - t_{13}^0 t_{23}^0 \\ -t_{11}^0 t_{31}^0 - t_{12}^0 t_{32}^0 - t_{13}^0 t_{33}^0 \\ -t_{21}^0 t_{31}^0 - t_{22}^0 t_{32}^0 - t_{23}^0 t_{33}^0 \\ (t_{21}^0)^2 + (t_{22}^0)^2 + (t_{23}^0)^2 - (t_{11}^0)^2 - (t_{12}^0)^2 - (t_{13}^0)^2 \\ ((t_{11}^0)^2 + (t_{12}^0)^2 + (t_{13}^0)^2 + (t_{21}^0)^2 + (t_{22}^0)^2 + (t_{23}^0)^2)/2 - (t_{31}^0)^2 - (t_{32}^0)^2 - (t_{33}^0)^2 \\ 3 - (t_{11}^0)^2 - (t_{12}^0)^2 - (t_{13}^0)^2 - (t_{21}^0)^2 - (t_{22}^0)^2 - (t_{23}^0)^2 - (t_{31}^0)^2 - (t_{32}^0)^2 - (t_{33}^0)^2 \end{pmatrix} \tag{28}$$

$$B^T = \begin{pmatrix} 0 & \cdots & 0 & t_{21}^0 & t_{22}^0 & t_{23}^0 & t_{11}^0 & t_{12}^0 & t_{13}^0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & t_{31}^0 & t_{32}^0 & t_{33}^0 & 0 & 0 & 0 & t_{11}^0 & t_{12}^0 & t_{13}^0 \\ 0 & \cdots & 0 & 0 & 0 & 0 & t_{31}^0 & t_{32}^0 & t_{33}^0 & t_{21}^0 & t_{22}^0 & t_{23}^0 \\ 0 & \cdots & 0 & 2t_{11}^0 & 2t_{12}^0 & 2t_{13}^0 & -2t_{21}^0 & -2t_{22}^0 & -2t_{23}^0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & -t_{11}^0 & -t_{12}^0 & -t_{13}^0 & -t_{21}^0 & -t_{22}^0 & -t_{23}^0 & 2t_{31}^0 & 2t_{32}^0 & 2t_{33}^0 \\ 0 & \cdots & 0 & 2t_{11}^0 & 2t_{12}^0 & 2t_{13}^0 & 2t_{21}^0 & 2t_{22}^0 & 2t_{23}^0 & 2t_{31}^0 & 2t_{32}^0 & 2t_{33}^0 \end{pmatrix} \tag{29}$$

**6 Hypothesis test in the linearized model**

The problem is now to identify constraints which in the linearized GMM with constraints are in conflict with the observations and the rest of the constraints, indicating that the transformation model is too special. The standard approach is to test, if all constraints are in conflict with the observations. The hypotheses to be tested here, read

$$H_0 : w = 0 \text{ versus } H_A : w \neq 0 \tag{30}$$

In the case developed in the last section this would mean to discriminate between the two models of the rotation and translation ( $H_0$  is true) and of the affine transformation ( $H_A$  is true).

As Lehmann and Neitzel (2013) have shown, it is possible to identify also conflicting subsets of constraints. If e.g. only the last three constraints (7), (9), (11) are conflicting then the nine-parameter transformation would be the transformation model of choice. This is equivalent to eliminating the conflicting constraints. Since we have more than two options how to build subsets of constraints, this requires a multiple test.

The most appealing layout of such a multiple test would be to test if an individual constraint is in conflict with the observations and the rest of the constraints. The drawback is that eliminating one of the first five constraints, the resulting transformation does not have a name in geodesy. However, this does not mean that it is excluded to apply such a transformation practically.

Since a priori we do not know which of the considered six constraints is the best candidate for elimination, we need to test the compatibility of all individual constraints in parallel. Following Lehmann and Neitzel (2013) the test statistic of such a test derived from the classical likelihood ratio is

- either the extreme normalized Lagrange multiplier (LM)

$$T_5'' = \max_{i=1, \dots, m} \left| \frac{\hat{k}'_i}{\sigma_{\hat{k}'_i \hat{k}'_i}} \right| \tag{31}$$

- or the extreme externally studentized Lagrange multiplier (LM)

$$T_6'' = \max_{i=1, \dots, m} \left| \frac{\hat{k}'_i}{\hat{\sigma}''_{\hat{k}'_i \hat{k}'_i}} \right| \tag{32}$$

$T_5''$  should be used if the a priori variance factor  $\sigma^2$  is known and  $T_6''$  should be used otherwise. For convenience, the notation of (Lehmann and Neitzel 2013, eqs. 5.13, 5.14, 5.75, 5.76) is adopted here as follows:  $\hat{k}'_i$  denotes the estimate of the LM related to the  $i$ th constraint, when solving the fully constrained GMM,  $\sigma_{\hat{k}'_i \hat{k}'_i}$  denotes the standard deviation of this value using the a priori variance factor  $\sigma^2$  and  $\hat{\sigma}''_{\hat{k}'_i \hat{k}'_i}$  denotes the standard deviation of  $\hat{k}'_i$  using an estimate  $\hat{\sigma}''^2$  of  $\sigma^2$ . This estimate denotes the common best quadratic unbiased estimate, but in the semi-constrained GMM. This means that for computing the estimate  $\hat{\sigma}''^2$  of  $\sigma^2$  the  $i$ th constraint must be dropped.

From (Lehmann and Neitzel 2013, eqs. 5.53-5.56) the following distributional results can be adopted:

$$\frac{\hat{k}'_i}{\sigma_{\hat{k}'_i \hat{k}'_i}} | H_0 \sim N(0, 1) \tag{33a}$$

$$\frac{\hat{k}'_i}{\sigma_{\hat{k}'_i \hat{k}'_i}} | H_A \sim N(\lambda_i, 1) \tag{33b}$$

$$\frac{\hat{k}'_i}{\hat{\sigma}''_{\hat{k}'_i \hat{k}'_i}} | H_0 \sim t(n - u + m - 1) \tag{34a}$$

$$\frac{\hat{k}'_i}{\hat{\sigma}''_{\hat{k}'_i \hat{k}'_i}} | H_A \sim t'(n - u + m - 1, \lambda_i) \tag{34b}$$

Here  $t(f)$  and  $t'(f, \lambda)$  denote the central and non-central Student's  $t$  distribution with  $f$  degrees of freedom and non-centrality parameter  $\lambda$ . Both in (33b) and (34b) the non-centrality parameter  $\lambda_i$  reads

$$\lambda_i = \frac{w_i \sqrt{q_{\hat{k}'_i \hat{k}'_i}}}{\sigma} = \frac{w_i \sigma_{\hat{k}'_i \hat{k}'_i}}{\sigma^2} \tag{35}$$

Practically, the true misclosure  $w_i$  in the  $i$ th constraint is unknown, and perhaps also the a priori variance factor  $\sigma^2$ . The cofactor  $q_{\hat{k}'_i \hat{k}'_i}$  of  $\hat{k}'_i$  is always known.

After choosing a probability of type I error  $\alpha$ , a critical value must be taken from either of the distributions (33a) or (34a), but with quantile probability  $\alpha/m$  in both cases, because the actual test statistics in the multiple test are the extreme LMs  $T_5''$  or  $T_6''$  (see Sect. 4). If the critical value is exceeded by the related test statistic  $T_5''$  in (31) or  $T_6''$  in (32) then we are inclined to reject  $H_0 : w = 0$ . This means here that the rotation and translation transformation model is not adequate. We should now drop the constraint, at which the maximum in (31) or (32) is attained.

Not unlike the common practice in geodesy when dealing with extreme normalized or studentized residuals in outlier detection, it is possible to iterate the procedure, until no further conflicting constraint can be identified.

### 7 Example: 3D transformation based on six control points

#### 7.1 Null hypothesis is true

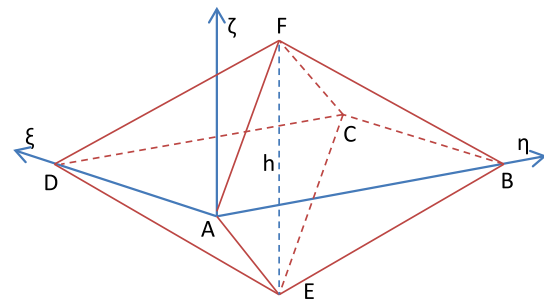
We illustrate and investigate the procedure in a setup of six control points forming a flattened octahedron, see Table 1 and Fig. 1. The height  $h$  of the octahedron will be varied.

The true transformation parameters are defined as zero except  $t_{11} = t_{22} = t_{33} = \mu^{\text{true}}$ . This means that the similarity transformation model (8) is the proper model, except when  $\mu^{\text{true}} = 1$ . In this case, the rotation and translation transformation model (10) is adequate. This setup has  $n = 2 * 3 * 6 = 36$  observations. (The frames could also have been rotated or translated with respect to each other without changing the subsequent results because the constraints for “no rotation” and “no translation” are not tested in this section).

A priori we assume not to know which transformation model is adequate. Therefore we use the affine transformation model (1) with  $u = 3 * 6 + 12 = 30$  model parameters and try to specialize it by applying  $m = 6$  constraints (2), (3), (4), (7), (9), (11) as derived above. This yields a total redundancy of  $n - u + m = 12$ .

**Table 1** True coordinates of control points used in Sect. 7 (flattened octahedron)

	$\xi_a^{\text{true}}$	$\eta_a^{\text{true}}$	$\zeta_a^{\text{true}}$
A	0	0	0
B	0	100	0
C	100	100	0
D	100	0	0
E	71	71	$-h/2$
F	71	71	$h/2$



**Fig. 1** Configuration of control points used in Sect. 7 (flattened octahedron)

Let us start with the case  $\mu^{\text{true}} = 1$ . Here  $H_0 : w = 0$  is true and should be rejected by the multiple test only with probability  $\alpha$ .

Observations (14) are generated in two different ways, by adding pseudo-random noise according to (13) with  $P = I$  (identity weight matrix), both with  $\sigma^2 = 10^{-4}$  and with  $\sigma^2 = 10^{-2}$ . From this we compute the test statistics (31), (32) and compare them with their critical values, deciding if  $H_0$  must be rejected or not. In a Monte Carlo approach, we repeat this procedure  $10^6$  times and compute the relative frequency of  $H_0$  rejected. This value is expected to be equal to  $\alpha$  because in this simulation we tacitly know that  $H_0$  is in fact true. The results are given in Table 2. It is seen that relative frequencies for normalized LM are a little bit smaller than expected, for studentized LM a little bit larger.

There are three potential causes why these relative frequencies are perhaps not exactly equal to  $\alpha$ :

1.  $10^6$  repetitions are not enough. This cause can be ruled out by re-starting the procedure with different pseudo-random numbers and comparing the results. If the relative frequencies differ from  $\alpha$  in the same way then  $10^6$  repetitions are enough, otherwise the number must be increased.
2. The original model is non-linear, and consequently the distributional results (33a)–(34b) are at best approximately valid. This cause can be ruled out by re-computing with larger  $\sigma^2$ , which makes the non-linearity worse. If the relative frequency now differs substantially more from  $\alpha$  then the linearization is to blame. This can also be done in the reverse direction.
3. When computing the critical values of (31) or (32), the simple portioning of  $\alpha$  onto the  $m$  constraints using  $\alpha/m$  as significance level is only valid if all  $\hat{k}_i'$  are statistically independent. This is at best approximately true.

Although the results in Table 2 have been computed with two different  $\sigma^2$  and at this opportunity also with different pseudo-random numbers, the deviation from  $\alpha$  is practically the same. This proves that cause three produces the observed



**Table 2** Various values of the relative frequency of  $H_0$  rejected, when it is true

Significance level $\alpha$	Critical value For $T_5''$ from (33a) with $\alpha/6$	$h = 10$		$h = 100$	
		Ratio of rejected $H_0 \sigma^2 = 10^{-4}, \sigma^2 = 10^{-2}$		Ratio of rejected $H_0 \sigma^2 = 10^{-4}, \sigma^2 = 10^{-2}$	
0.01	3.19	0.008	0.009	0.010	0.009
0.02	2.98	0.017	0.017	0.019	0.019
0.05	2.69	0.043	0.043	0.049	0.048
0.10	2.45	0.084	0.083	0.095	0.094

Significance level $\alpha$	Critical value For $T_6''$ from (34a) with $\alpha/6$	$h = 10$		$h = 100$	
		Ratio of rejected $H_0 \sigma^2 = 10^{-4}, \sigma^2 = 10^{-6}$		Ratio of rejected $H_0 \sigma^2 = 10^{-4}, \sigma^2 = 10^{-6}$	
0.01	3.80	0.012	0.013	0.013	0.013
0.02	3.48	0.023	0.024	0.027	0.026
0.05	3.06	0.055	0.057	0.064	0.063
0.10	2.73	0.107	0.108	0.123	0.122

effect. However, at least in this small-scale example, the deviations from  $\alpha$  seem to be tolerable.

### 7.2 Null hypothesis is false

Next, we consider the case that  $H_0$  is false. The test should now reject  $H_0$ . The ability to reject  $H_0$  when it is false, is known as the power of the test. The power  $\Pi$  is usually smaller, when  $H_0$  is only slightly violated (small true misclosure  $w$ ) and larger otherwise. This relationship is called the power function of the test  $\Pi(w)$ .

More specifically,  $\Pi(w)$  equals the probability that  $H_0$  is rejected by the test as a function of the true misclosure  $w$ . It is computed from the CDF of  $T|H_A$ , see (33b), (34b). Software packages with implemented quantile function of the non-central Student's  $t$  distribution in (34b) are less widely spread. In MATLAB, we find the function `ncinv`.

In our simulation study, we are in the position to implement a true misclosure into the model. As a test we implement  $\mu_1^{\text{true}} = \mu_2^{\text{true}} = \mu_3^{\text{true}} = \mu^{\text{true}} \neq 1$ , i.e., a true misclosure  $w_6 = 3(\mu^{\text{true}})^2 - 3 \neq 0$ . All other constraints remain valid. As a test statistic we again use only the extreme normalized and externally studentized LMs, i.e., we assume not to know which constraint is violated.

The results in terms of the two power functions are given in Fig. 2. We restrict ourselves to  $\sigma^2 = 10^{-4}$  here and to  $\mu^{\text{true}} > 1$ , i.e.,  $w_6 > 0$ , because  $\Pi(w)$  does not change when  $w$  changes sign. Firstly, we observe that the power increases with the significance level  $\alpha$ . This is the typical behavior because a higher  $\alpha$  means that  $H_0$  is more often rejected and therefore less often falsely accepted. Secondly, we observe that the power increases with the magnitude of the true misclosure  $|w_6|$ . This is also expected because a more distinct separation between  $H_0$  and  $H_A$  makes type II decision errors (failures to reject a false  $H_0$ ) less probable. stop The dif-

ference between the power of the test statistics (31), (32) is clearly seen. The normalized LMs require  $\sigma^2$  to be known. If  $\sigma^2$  is unknown then we must resort to studentized LMs with a significant loss of test power. The reason for this is a typical smearing effect: If constraint (11) is in effect, the inconsistent scales between both coordinate frames are partly interpreted as observation errors, increasing the estimated residuals. In this way, the variance factor  $\sigma^2$  is mostly overestimated by  $(\hat{\sigma}'')^2$ . This makes  $\hat{\sigma}''_{\hat{k}_i \hat{k}_i}$  in (32) too large. Consequently,  $T_6''$  becomes too small, such that it does not exceed its critical value as often as  $T_5''$  does.

Finally, we see nearly no effect in the power of  $T_5''$  and  $T_6''$ , when changing  $h$ , and in this way changing the conditioning of the normal system of the GMM. (Note that for  $h = 0$  the system is singular due to a coplanarity of all control points).

### 7.3 Using the extreme normalized misclosure as a test statistic

Test statistics must not be restricted to likelihood ratios, but could perhaps be defined by plausibility reasoning. See (21) and discussion below. Not uncommon in practical geodesy, such a test statistic could be the extreme normalized or studentized estimated misclosure. In this subsection, the extreme normalized estimated misclosure

$$T_w := \max_{i=1, \dots, m} \left| \frac{\hat{w}_i}{\sigma_{\hat{w}_i} \hat{w}_i} \right| \tag{36}$$

is defined as test statistics and will be considered as a substitute for  $T_5''$  in (31).  $T_w$  uses the estimated misclosures  $\hat{w}_i$  of the unconstrained solution, which is the affine transformation model (14). A large value of such a test statistic would also indicate an incompatibility of the constraint, for which the maximum in (36) is attained. The critical values of  $T_w$  are the same as of  $T_5''$  and can be taken from Table 2.

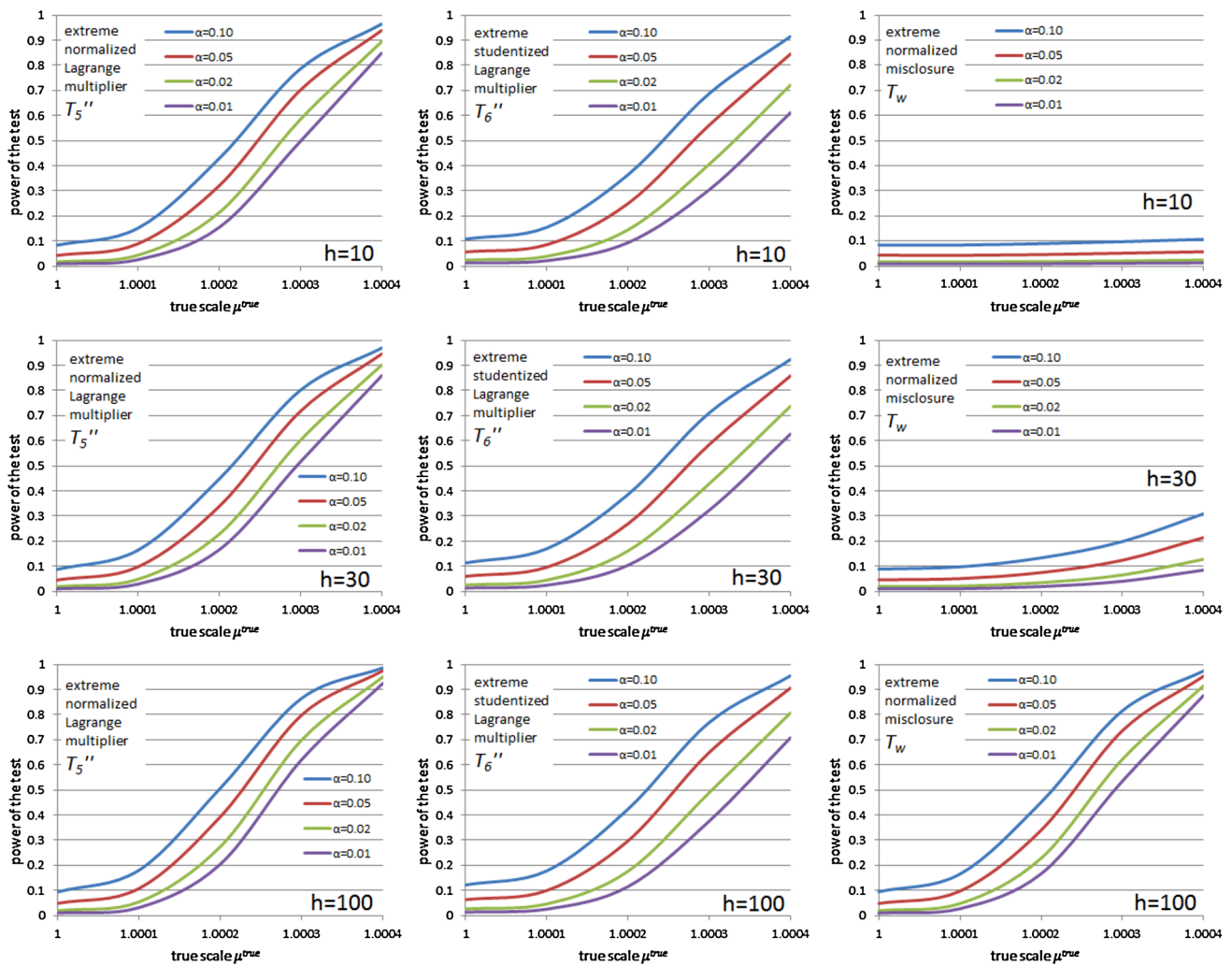


Fig. 2 Power functions related to the test statistics (31), (32), (36) for the example of Sect. 7

The resulting power functions are displayed in Fig. 2 and must be compared to that of the test statistic  $T_5''$ , where  $\sigma^2$  has also been used. We see a similar behavior, but recognize a great loss of power, when  $h$  is small, i.e., for nearly ill-conditioned normal systems. Here, the misclosure is subject to a smearing effect: incompatible scales of the coordinate frames produce large magnitudes of misclosures not primarily in  $\hat{w}_6$ , but the effect also smears over to other misclosures, such that  $T_w$  often falls below the critical value.

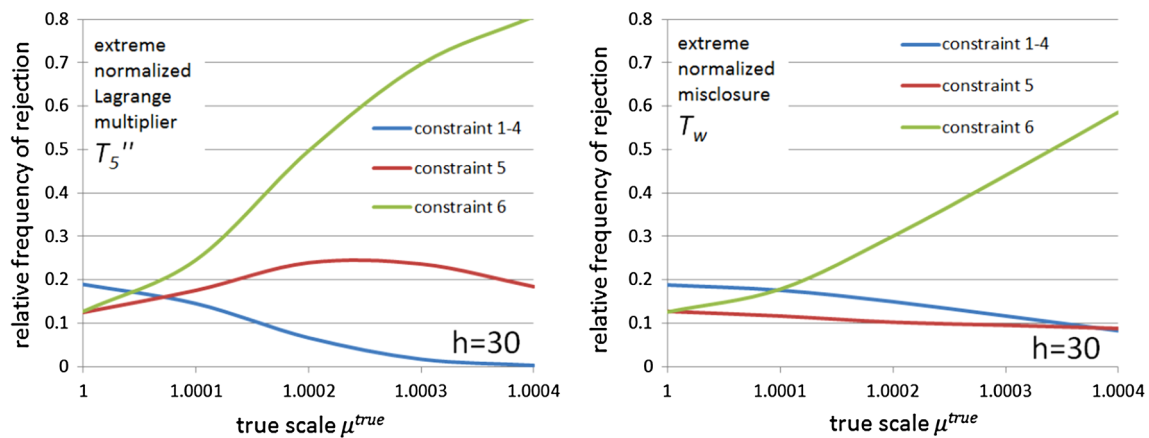
#### 7.4 Identification of the conflicting constraints

If the test statistic exceeds the critical value, then we are inclined to reject  $H_0$ . This means that the model of rotation and translation is too special for the description of the relationship between the control points. The next step would be to identify the constraint in conflict with the observations and the rest of the constraints. We can only hope that this is the

constraint, for which the maximum in (31), (32) or (36) is attained. Now, we investigate if this is true.

Firstly, note that in this respect there is no difference between normalized and studentized LMs (and also not between normalized and studentized misclosures, although the latter values have not been used here). The reason is that these values differ only in the way that  $\hat{\sigma}_{\hat{r}_i \hat{r}_i}''$  is computed with the estimated variance factor  $\hat{\sigma}^2$ , while for  $\sigma_{\hat{r}_i \hat{r}_i}$  the known value  $\sigma^2$  is used. Therefore, normalized and studentized values are proportional to each other and the maxima in (31) and (32) are attained at the same index  $i$ . Consequently, studentized values are disregarded below.

We use the computation of the last section for  $h = 30$  only. The relative frequencies, how often the maxima of (31) and (36) are attained at a certain constraint, are displayed in Fig. 3, regardless of their value, i.e., if they exceed any critical value or not. In this way, the investigation becomes independent of  $\alpha$ . We expect those maxima to be attained mostly at  $i = 6$ , which is the index of the violated constraint



**Fig. 3** Relative frequencies of rejection of a constraint, for which the maximum in (31), left subplot, and (36), right subplot, is attained. Note that the curves for constraints 1–4 are practically overlapping

(11). If this violation is only weak, i.e.,  $1 < \mu^{\text{true}} < 1.0001$  here, then these maxima are not attained primarily at  $i = 6$ . But unsurprisingly, the stronger (11) is violated, the more often the maxima are attained at  $i = 6$ . At a value of  $\mu^{\text{true}}$  at which  $H_0$  is almost certainly rejected ( $\mu^{\text{true}} > 1.0004$ ), we almost certainly accept the correct  $H_A$ .

Comparing left and right part of Fig. reffg3 we see that for LMs we get the correct result more often than for misclosures, which confirms that LMs are the superior test statistics. For  $T_5''$  the maximum in (31) is oftentimes attained at  $i = 5$ , such that the constraint (9) is sometimes wrongly identified as conflicting. For misclosures we find wrong identifications in all five other constraints.

### 8 The use of information criteria

#### 8.1 Comparison to Akaike information criterion

As an alternative to transformation model selection by hypothesis testing we employ the AICc information criterion as proposed by Felus and Felus (2009). It is defined as (Burnham and Anderson 2002)

$$AICc = n \cdot \log(\text{WSSR}) + \frac{2(u - m)n}{n - u + m - 1} \tag{37}$$

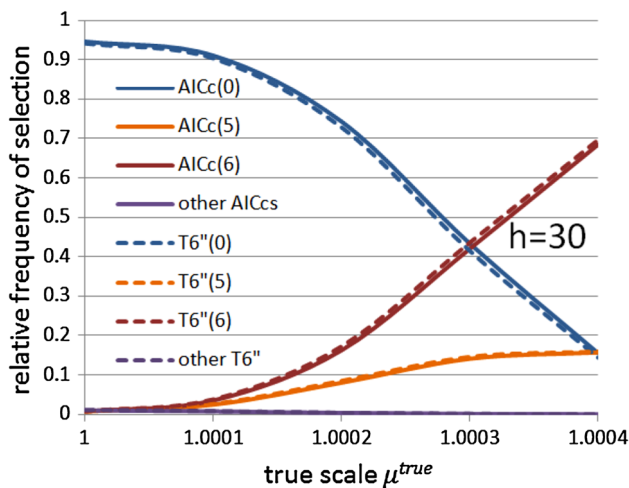
where as before  $n$ ,  $m$  and  $u$  are the number of observations, constraints and GMM parameters, respectively, and WSSR is the weighted sum of squared residuals in the GMM. The first summand measures goodness of fit and the second is a penalty term for model complexity. The model with the smallest AICc must be selected. Note that the standard formulae for AIC and AICc do not involve constraints. But it is clear that the number of constraints  $m$  must be subtracted

from the number of parameters  $u$  to get the number of effective parameters  $u - m$  as a measure of model complexity.

We apply this criterion to the example of Sect. 7. Note that AICc and similar criteria do not use a possibly known variance factor  $\sigma^2$ . Therefore, we can fairly compare AICc only to  $T_6''$  in (32). In the following, we restrict ourselves to AICc computed for

- the rotation and translation model, i.e.,  $m = 6$ , denoted as  $AICc(0)$
- all six transformation models with  $m = 5$ , denoted as  $AICc(1) \dots AICc(6)$ , and
- the eight-parameter transformation model, i.e.,  $m = 4$ , denoted as  $AICc(7)$ , as an example of an, in any case, overfitting model.

First of all, if  $\mu^{\text{true}} = 1$  then the rotation and translation is the proper model, and we expect  $AICc(0) < AICc(i), i > 0$  to show this. In the range of  $10 < h < 100$  this happens with a relative frequency of 0.94, . . . , 0.95. This can be seen in Fig. 4 for  $h = 30$ . It means that the AICc effectively relates to a significance level  $\alpha \approx 0.05$  (although it is of course not a hypothesis test). If  $\mu^{\text{true}} \neq 1$  then we expect that  $AICc(6)$  falls below  $AICc(0)$ . For each model we display the relative frequency of selection according to AICc in Fig. 4 together with the relative frequency of selection according to the hypothesis test based on  $T_6''$  in (32). By  $T_6''(0), T_6''(1) \dots T_6''(6)$  we mean the relative frequencies of selection of the rotation and translation model and the six models with  $m = 5$ , respectively, analogous to the notation of AICc. First of all, it is striking how close the corresponding curves in Fig. 4 are. Thus, AICc almost coincides with  $T_6''$  at a significance level  $\alpha = 0.05$ . The small difference is in favor of  $T_6''$ . (Remember that here  $T_6''(6)$  selects the appropriate model.) In compliance with the results of Sect. 7.3, the



**Fig. 4** Relative frequencies of selection of a model according to the AICc (37) [solid curves] and the hypothesis test based on extreme studentized LMs  $T_6''$  (32) with significance level  $\alpha = 0.05$  [dotted curves]

almost only other model with  $m = 5$  selected here is the model without constraint (9). It is wrongly selected here, but this happens only in rare cases.

It is interesting to note that not only the model selections show equal frequencies in Fig. 4, but also coincide very well. For example, for  $\mu^{\text{true}} = 1.0004$  the relative frequency of coincident model selections, may be right or wrong, is 0.985. Here, we have restricted ourselves to  $h = 30$  because the same figure for  $h = 10$  or  $h = 100$  would be nearly indistinguishable from Fig. 4.

Summarizing, we can say that model selection by AICc is very much the same as by extreme studentized LMs  $T_6''$  with significance level  $\alpha = 0.05$  here. This is appreciated by users uncertain about the choice of any  $\alpha$ , because AICc does not require such a choice. Also no critical value must be computed. This is welcome, if the distribution of the test statistic is analytically intractable or unknown. On the other hand, if for example a user wants to have the model with less parameters supported more than AICc does, then she/he cannot simply choose a smaller  $\alpha$ . Possibly, another information criterion like BIC can be employed.

## 8.2 Comparison to Mallows' $C_p$

Another common information criterion is Mallows'  $C_p$  (Mallows 1973). So far it is less popular in geodesy, at least under this name, but has been employed e.g. by Mahboub (2014). It is derived from an advanced statistic as

$$C_p = \frac{\text{WSSR}}{\sigma^2} - (n - 2(u - m)) = \frac{\text{WSSR}}{\sigma^2} - (n - 2p) \quad (38)$$

with notation as in (37). In the same way as for AIC and AICc, the standard formula for  $C_p$  does not involve con-

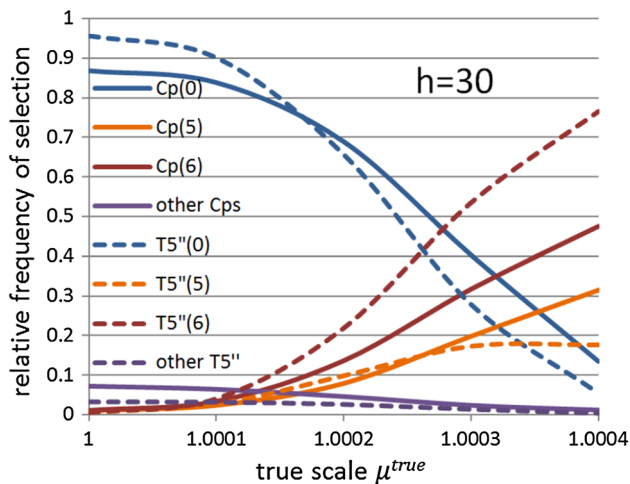
straints. But it is clear that the number of constraints  $m$  must be subtracted from the number of parameters  $u$  to get the number of effective parameters  $u - m$ , for which often the symbol  $p = u - m$  is used. The expectation of  $C_p$  is known to be equal to  $p$ , possibly plus a positive bias term due to lack of fit. Therefore, the model with the smallest value of  $C_p$ , which is somehow close to  $p$ , is selected.

Unlike AIC and AICc, Mallows'  $C_p$  is also able to use  $\sigma^2$ , should it be known. Therefore it is interesting to compare it to  $T_5''$  in (31). We repeat the computations of the last subsection, coming up with  $C_p(0), \dots, C_p(7)$  instead of  $\text{AICc}(0), \dots, \text{AICc}(7)$  and  $T_5''(0), \dots, T_5''(6)$  instead of  $T_6''(0), \dots, T_6''(6)$ .

To start with, we naively base the model selection on Mallows'  $C_p$  in such a way that each time the model with  $C_p$  closest to  $p$  is selected. The results are disappointing: For  $\mu^{\text{true}} = 1$  Mallows'  $C_p$  selects the correct model only with a relative frequency of 0.06. Instead the overfitting eight-parameter transformation model is wrongly selected most often. The reason is that in any case the values of  $C_p$  scatter perfectly around  $p$ , but the overfitting models show smaller scattering, such that there is a bigger chance to get a value of  $C_p$  closer to  $p$ . For  $\mu^{\text{true}} = 1.0004$  Mallows'  $C_p$  most often selects the correct model, but by far not as often as the other methods.

This shows that it is crucial not to select the model "with  $C_p$  closest to  $p$ ", as done before, but "with the smallest value of  $C_p$ , which is somehow close to  $p$ ". We must therefore define a permissible interval of values close to  $p$ . The disadvantage is that it may happen that none of the values of  $C_p$  falls into this interval. Some investigation using the numerical example at hand shows that the interval  $[p - 7, p + 7]$  is optimal. It is the smallest interval, which ensures that with a relative frequency of 0.95 at least one  $C_p$  is permissible. It is therefore used below.

The results are displayed in Fig. 5. While model selection based on  $T_5''$  shows almost the same, but slightly superior results as on  $T_6''$  (compare with Fig. 4), the results with Mallows'  $C_p$  are worse. For  $\mu^{\text{true}} = 1$  Mallows'  $C_p$  selects the correct model only with a relative frequency of 0.87 (curve  $C_p(0)$  in Fig. 5). This is much better than in the naïve approach, but not as good as for the other methods. The latter also applies to the cases with  $\mu^{\text{true}} > 1$ : For example, for  $\mu^{\text{true}} = 1.0004$  Mallows'  $C_p$  selects the correct model only in one out of two cases (curve  $C_p(6)$  in Fig. 5). In one out of three cases constraint (9) is wrongly identified as conflicting (curve  $C_p(5)$  in Fig. 5). This shows that Mallows'  $C_p$  is not a good alternative to multiple hypotheses testing. One reason could be that here  $p = u - m$  is not varying much, only from 24 to 26. Therefore, Mallows'  $C_p$  may be tried again in cases where in one system all coordinates of control points are considered as error-free quantities, such that  $p = u - m$  shows a stronger relative variation, here from 6 to 8.



**Fig. 5** Relative frequencies of selection of a model according to Mallows'  $C_p$  (38) [solid curves] and the hypothesis test based on extreme normalized LMs  $T_5''$  (31) with significance level  $\alpha = 0.05$  [dotted curves]

## 9 Conclusions

In geodesy, we oftentimes need to transform points between two different coordinate frames. We are routinely faced with the situation that we do not know which transformation model should be selected. It is good tradition in geodesy to base such a decision on a statistical hypothesis test. If we consider more than two models then the test must be a multiple hypotheses test.

It is often not possible to simply test the significance of parameters because all considered models may have a different set of parameters. For example, there is no natural implementation of a scale parameter into the spatial affine transformation. Therefore, it is better to start with a general transformation model and try to specialize it by adding constraints. The compatibility of those constraints needs to be tested. Such a test can be intuitively based on the estimated misclosures of the constraints in the unconstrained model. From (Lehmann and Neitzel 2013) we know that better test statistics should be based on the Lagrange multipliers (LMs, also known as correlates in geodesy) of the constrained solution. They should assume the form of either the extreme normalized or the extreme externally studentized LM. The second test statistic comes into effect if the a priori variance factor is unknown.

We worked out an example of a 3D coordinate transformation based on six control points. Here it is shown that the LM-based test statistics have more statistical power than those based on the estimated misclosures. This advantage is most drastic if the configuration of the control points is poor. Moreover, in a multiple test the test statistics based on LMs more often identify the correct alternative model. They can be recommended, not only to problems with transfor-

mations, but to all geodetic adjustment problems posed as a GMM with constraints.

For transformation model selection, the AICc and Mallows'  $C_p$  are considered as an alternative to the multiple hypotheses test. It turns out that in the exemplary case the AICc almost always select the same model as the extreme externally studentized LM does. This is remarkable because the theoretical background is different. Mallows'  $C_p$  was also successfully applied, but here the results are inferior to the other methods. This may be due to the fact that there is only a small difference in the number of parameters of the models to be selected.

## References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Andrei CO (2006) 3D affine coordinate transformations. Masters of Science Thesis in Geodesy, No. 3091. School of Architecture and the Built Environment Royal Institute of Technology (KTH) Stockholm, Sweden
- Burnham KP, Anderson DR (2002) Model selection and multimodel Inference: a practical-theoretic approach, 2nd edn. Springer, New York. ISBN:0-387-95364-7
- Carosio A, Piras M, Salvini D (2006) The forward search method applied to geodetic transformations. *Studies in classification, data analysis, and knowledge organization, part III*. pp 181–188. doi:10.1007/3-540-35978-8\_21
- Deakin RE (1998) 3D coordinate transformations. In: *Surveying and land information systems*, vol 58, issue no 4. pp 223–234
- Deakin RE (2007) Coordinate transformations for Cadstral Surveying. RMIT University, School of Mathematical and Geospatial Sciences, pp 1–33
- Donato G, Belongie S (2002) Approximate thin plate spline mappings. In: *Lecture notes in computer science*, vol 2352. Springer, Berlin, pp 21–31. doi:10.1007/3-540-47977-5\_2
- Felus YA, Felus M (2009) On choosing the right coordinate transformation method. In: *Proceedings of FIG working week 2009: surveyors key role in accelerated development*. Eilat, Israel, 3–8 May 2009. [http://www.fig.net/pub/fig2009/papers/ts04c/ts04c\\_felus\\_felus\\_3313.pdf](http://www.fig.net/pub/fig2009/papers/ts04c/ts04c_felus_felus_3313.pdf)
- Ge Y, Yuan Y, Jia N (2013) More efficient methods among commonly used robust estimation methods for GPS coordinate transformation. *Surv Rev* 45(330):229–234. doi:10.1179/1752270612Y.0000000028
- Goktepe A, Kocaman E (2010) Analysis of camera calibrations using direct linear transformation and bundle adjustment methods. *Sci Res Essays* 5(9):869–872
- Kampmann G (1996) New adjustment techniques for the determination of transformation parameters for cadastral and engineering purposes. *Geomatica* 50(1):27–34
- Kanani E (2000) Robust estimators for geodetic transformations and GIS. Ph.D. thesis, Institut für Geodäsie und Photogrammetrie, ETH Zürich
- Kargoll B (2012) On the theory and application of model misspecification tests in geodesy. *Deutsche Geodätische Kommission Reihe C*, Nr. 674, München
- Koch KR (1999) Parameter estimation and hypothesis testing in linear models, 2nd edn. Springer, Berlin. ISBN:978-3-662-03976-2
- Lehmann R (2010) Im Schwerpunkt der Anschlusspunkte - Zur Genauigkeit geodätischer Koordinatentransformationen (In the barycentre of the control points—on the accuracy of geodetic coordi-

- nate transformations). *Allgemeine Vermessungsnachrichten* 4/2010. VDE-Verlag Offenbach
- Lehmann R (2012) Improved critical values for extreme normalized and studentized residuals in Gauss–Markov models. *J Geod* 86(12):1137–1146
- Lehmann R, Neitzel F (2013) Testing the compatibility of constraints for parameters of a geodetic adjustment model. *J Geod* 87:555–566. doi:[10.1007/s00190-013-0627-2](https://doi.org/10.1007/s00190-013-0627-2)
- Lippus J (2004) Transformation of coordinates using piecewise conformal mapping. *J Geod* 78(1–2):40–46. doi:[10.1007/s00190-003-0364-z](https://doi.org/10.1007/s00190-003-0364-z)
- Mahboub V (2012) On weighted total least-squares for geodetic transformations. *J Geod* 86(5):359–367. doi:[10.1007/s00190-011-0524-5](https://doi.org/10.1007/s00190-011-0524-5)
- Mahboub V (2014) Variance component estimation in errors-in-variables models and a rigorous total least squares approach. *Stud Geophys Geod* 58(2014):17–40. doi:[10.1007/s11200-013-1150-x](https://doi.org/10.1007/s11200-013-1150-x)
- Mallows CL (1973) Some comments on CP. *Technometrics* 15(4):661–675. doi:[10.2307/1267380](https://doi.org/10.2307/1267380)
- Miller RG (1981) *Simultaneous statistical inference*. Springer-Verlag New York Inc., Berlin. ISBN:0-387-90548-0.
- Neyman J, Pearson ES (1933) On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc A Math Phys Eng Sci* 231(694–706):289–337. doi:[10.1098/rsta.1933.0009](https://doi.org/10.1098/rsta.1933.0009)
- Neitzel F (2010) Generalization of total least-squares on example of unweighted and weighted 2D similarity transformation. *J Geod* 84(12):751–762. doi:[10.1007/s00190-010-0408-0](https://doi.org/10.1007/s00190-010-0408-0)
- Schaffrin B, Felus YA (2008) On the multivariate total least-squares approach to empirical coordinate transformations. Three algorithms. *J Geod* 82(6):373–383. doi:[10.1007/s00190-007-0186-5](https://doi.org/10.1007/s00190-007-0186-5)
- Tanizaki H (2004) *Computational methods in statistics and econometrics*. Marcel Dekker, New York. ISBN-13: 978–0824748043
- Teunissen PJG (1985) Quality control in geodetic networks. In: Grafarend EW, Sanso F (eds) *Optimization and design of geodetic networks*. Springer, Berlin, pp 526–547
- Teunissen PJG (2000) *Testing theory; an introduction*. In: *Series on mathematical geodesy and positioning*, 2nd edn. Delft University of Technology, The Netherlands. ISBN:90-407-1975-6
- Zhang P, Xu C, Hu C, Chen Y (2012) Coordinate transformations in satellite navigation systems. In: *Advances in electronic engineering, communication and management*, vol. 2. Springer, Berlin, pp 249–257. doi:[10.1007/978-3-642-27296-7\\_40](https://doi.org/10.1007/978-3-642-27296-7_40)
- Ziggah YY, Youjian H, Amans ChO, Fan DL (2013) Determination of GPS coordinate transformation parameters of geodetic data between reference datums—a case study of ghana geodetic reference network. *Int J Eng Sci Res Technol* 2(4):956–971