Testing the compatibility of constraints for parameters of a geodetic adjustment model

Authors

Prof. Dr.-Ing. Rüdiger Lehmann
University of Applied Sciences Dresden
Faculty of Spatial Information
Friedrich-List-Platz 1
D-01069 Dresden
Tel +49 351 462 3146
Fax +49 351 462 2191
mailto:r.lehmann@htw-dresden.de

Prof. Dr.-Ing. Frank Neitzel
Fachgebiet Geodäsie und Ausgleichungsrechnung
Institut für Geodäsie und Geoinformationstechnik
Technische Universität Berlin, Sekr. H20
Str. des 17. Juni 135
D 10623 Berlin
Sekretariat: +49-30/314-23315
Tel: +49-30/314-22375
Fax: +49-30/314-21119
E-mail : frank.neitzel@tu-berlin.de

Abstract

Geodetic adjustment models are often set up in a way that the model parameters need to fulfil certain constraints. The normalized Lagrange multipliers have been used as a measure of the strength of constraint in such a way that if one of them exceeds in magnitude a certain threshold then the corresponding constraint is likely to be incompatible with the observations and the rest of the constraints. We show that these and similar measures can be deduced as test statistics of a likelihood ratio test of the statistical hypothesis that some constraints are incompatible in the same sense. This has been done before only for special constraints (Teunissen 1985) . We start from the simplest case, that the full set of constraints is to be tested, and arrive at the advanced case, that each constraint is to be tested individually. Every test is worked out both for a known as well as for an unknown prior variance factor. The corresponding distributions under null and alternative hypothesis are derived. The theory is illustrated by the example of a double levelled line.

Keywords

Geodetic adjustment; Gauss Markov model with constraints; Lagrange multipliers; likelihood ratio test; Constraints compatibility

1 Introduction

Geodetic adjustment models are often set up in a way that the model parameters x need to fulfil certain constraints. Such constraints are expressed in a way that the parameters satisfy a set of equations or inequations.
In the following we will restrict ourselves to linear equality constraints

$$B^T x = b \tag{1.1}$$

where B is a $u \times m$-matrix and b is a m-vector. Inequality constraints are treated e.g. in Schaffrin (1981) , Roese-Koerner et al. (2012) .

The number of equations m must be smaller than the number of model parameters u. Sometimes it is possible to solve the equations for m model parameters and to substitute these parameters in the adjustment model by the remaining $u - m$ parameters. This approach transforms the constrained adjustment problem into an unconstrained one. Even though it is possible to do so, this approach has disadvantages:

1. The observation equations become more complex . This partly complicates the solution.
2. If a constraint is erroneous in the way that the true parameters do not fulfil it then this model misspecification does not clearly protrude anymore.

Example: If a levelled loop is not truly closed then this effect may be seen in the large magnitude of the estimated loop misclosure. Assume that the model parameters are the height differences of adjacent points. If one parameter is substituted by the negative sum of the rest of the parameters in this loop then this model disturbance only shows up in large magnitudes of all residuals (estimated observation errors) .

Constrained adjustment problems are most often solved using the concept of **Lagrange multipliers** (LMs) . The genesis of the LMs is analyzed by Bussotti (2003) . Particularly, the author shows that this mathematical approach was introduced by Lagrange in the framework of statics in order to determine the general equations of equilibrium for problems with constraints.

Later the concept of LMs became a very widespread tool for the solution of constrained optimization problems, e.g. of constrained least squares problems in geodetic adjustment. The mathematical formulation of such problems is oftentimes called Gauss-Helmert-Model. Helmert treated the case of "constrained observations" in his classical textbook (Helmert 1872, p.202-215) . He refered to the LMs as "correlates" (german: "Korrelaten", see also Jäger et al. 2005, p.161-173) . Another synonym for "LMs" is "Lagrange correlates" (e.g. Krakiwsky and Biacs 1990) .

Today the use of LMs is built on a broader foundation: Schaffrin and Felus (2005) introduce linear constraints into the novel framework of total least-squares (TLS) adjustment. Koch (2012) uses LMs for a robust estimation method based on the expectation maximization (EM) algorithm. The method is demonstrated by the analysis of measurements with gross errors of a laser scanner.

It is well known that if a constraint forces the solution of an optimization problem to be very different from the unconstrained solution then this causes the corresponding LM to become large in magnitude. That is why Wziontek et al. (2001) introduce the **LMs as measures of the strength of a constraint** in geodetic adjustment problems. They argue that if a LM divided by its standard deviation is large in magnitude then the corresponding constraint is likely to be **incompatible** with the observations. A numerical example is given in form of a 2D-triangulation network. An exact formulation in terms of statistical hypothesis tests, as it will be given below, is missing.

The concept of using LMs as a measure of the strength of constraint is reconsidered by Wienholz (2002) for the calculation of GPS phase ambiguities. The application is restricted to a numerical example in form of a 3D-trilateration network.

Independently Clements and Costa (1998) as well as Lourenço et al. (2004) derive the concept of **normalized LMs** (LMs divided by their standard deviation) as a measure of the strength of a constraint for topology error identification in power system state estimation. A threshold of 3 is used. A **normalized**

LM above 3 indicates an error in the model. Also here no direct relationship to statistical hypothesis tests is established.

We will show in this paper that these heuristic measures have a profound statistical justification. Teunissen (1985) approached this point from the opposite direction: Setting up the hypothesis test for quality control in geodetic networks, where (1.1) assumes a special form, he showed that the resulting test statistic is in fact a LM divided by its standard deviation. In the language introduced in our paper this test statistic is an individual normalized LM. Besides this we will derive extreme normalized as well as individual and extreme studentized LMs.

The paper is organized as follows: After introducing hypothesis tests for geodetic adjustment we recall the adjustment solutions in the classical Gauss Markov Modell (GMM) both without and with constraints. The mainstay of the paper is presented in section 5. Here we develop the hypothesis tests for compatibility of certain constraints of a GMM with the observations and the rest of the constraints. We start from the simplest case, that the full set of constraints is to be tested, and arrive at the advanced case, that each constraint is to be tested individually. The theory is illustrated by the example of a double levelled line.

2   Hypothesis tests in geodetic adjustment

The problem of compatibility of constraints for parameters in geodetic adjustment should be rigorously treated from the viewpoint of statistical testing theory. In this paper we will design a statistical test for the null hypothesis that all constraints are compatible vs. the alternative hypothesis that some constraints are erroneous and therefore incompatible with the observations and the rest of the constraints.

In statistical testing theory there are three important types of hypothesis tests relevant for our problem:

1. the likelihood ratio (LR) tests, which is the standard type applied in geodesy (Koch 1999, Teunissen 2000 p.53-62, Kargoll 2012 p.29) ,
2. the Rao score (RS) tests, also known as Lagrange multiplier (LM) test, introduced by Silvey (1959) , for use in geodetic adjustment see also (Kargoll 2012, p.33-34) , and
3. the Wald tests (Wald 1943) .

Since we have to deal with LMs in the following, it seems as if the LM tests are better suited than the LR or Wald tests. But this is not true. Engle (1983) shows that all three types of hypothesis tests are in some senses asymptotically equivalent. Practically we desire to apply the test, which maximizes the test power. According to the Neyman–Pearson lemma this holds for an LR test as long as we deal with simple hypotheses (Teunissen 2000, Kargoll 2012) . Then the LR test is called uniformly most powerful (UMP) . Unfortunately, all hypotheses under consideration in this paper are not simple but composite, such that a strict UMP property is not guaranteed. Thus, we cannot be sure to obtain a test with much power. Nonetheless, the LR tests are still a good choice, also because they are familiar to most geodesists. In the following we will restrict ourselves to the LR tests.

Consider a linear null hypothesis, represented by the system of m linear equations

$$H_0: B^T x = b$$

(2.1)

for the u-vector of model parameters x, opposed by the alternative hypothesis

$$H_A: B^T x \neq b. \tag{2.2}$$

Let $L(x|y)$ denote the likelihood function of x, given the n-vector of observations y. The LR test defines the test statistic as (e.g. Teunissen 2000 p.53)

$$T_1^{LR} := \frac{sup\{L(x|y): B^T x = b\}}{sup\{L(x|y)\}}. \tag{2.3}$$

In the numerator the hypothesis (2.1) is introduced as a constraint. The LR test requires the solution of two optimization problems: a constrained and an unconstrained maximum likelihood estimate (MLE) for $L(x|y)$. The first is solved using the method of LMs: The least squares minimum functional Ω (weighted sum of squared residuals, see (3.3) below) is extended by the additional LM-term to the LM minimum functional

$$\Omega' = \Omega + 2k^T \left(B^T x - b \right) \tag{2.4}$$

where k is the m-vector of LMs. If the ratio T_1^{LR} of the suprema falls below a critical value c then H_0 must be rejected, otherwise we fail to reject H_0.

If we have normally distributed observation errors then the MLE coincides with the well-known least squares estimate of the parameters with and without constraints: (2.3) assumes the form

$$T_1^{LR} = \frac{\exp\left(-\frac{\Omega'}{2\sigma^2}\right)}{\exp\left(-\frac{\Omega}{2\sigma^2}\right)} = \exp\left(-\frac{\Omega' - \Omega}{2\sigma^2}\right). \tag{2.5}$$

Alternatively we may use the fully equivalent test statistic

$$T_1' := \frac{\Omega' - \Omega}{\sigma^2}. \tag{2.6}$$

If $T_1' > c_1'$ with a properly chosen critical value c_1' then H_0 must be rejected, otherwise we fail to reject H_0.

Now assume that additionally the variance factor σ^2 of the observations is unknown. Let $L(x, \sigma^2|y)$ denote the joint likelihood function of x and σ^2, given the n-vector of observations y. The LR test defines the test statistic now as

$$T_2^{LR} := \frac{sup\{L(x, \sigma^2|y): B^T x = b\}}{sup\{L(x, \sigma^2|y)\}} \tag{2.7}$$

If we have normally distributed observation errors then the MLE for σ^2 without and with constraints $B^T x = b$ read (e.g. Koch 1999)

$$\hat{\sigma}^2 = \Omega/n \tag{2.8}$$

$$\hat{\sigma}'^2 = \Omega'/n \tag{2.9}$$

respectively. These formulae are obtained from the necessary condition $\partial L(x, \sigma^2|y)/\partial \sigma^2 = 0$, both in the constrained and in the unconstrained case. Note that the LM term is free of σ^2 and disappears when taking the derivative.

Thus, (2.7) assumes the form

$$T_2^{LR} = \frac{(\hat{\sigma}'^2)^{-n/2} \exp\left(-\frac{\Omega'}{2\hat{\sigma}'^2}\right)}{(\hat{\sigma}^2)^{-n/2} \exp\left(-\frac{\Omega}{2\hat{\sigma}^2}\right)} = \left(\frac{\hat{\sigma}'^2}{\hat{\sigma}^2}\right)^{-n/2} = \left(\frac{\Omega'}{\Omega}\right)^{-n/2} = \left(1 + \frac{\Omega' - \Omega}{\Omega}\right)^{-n/2} \tag{2.10}$$

Alternatively we may use the fully equivalent test statistic

$$T_2' := \frac{n-u}{m} \cdot \frac{\Omega' - \Omega}{\Omega}. \tag{2.11}$$

The factor $(n-u)/m$ is chosen such that under standard assumptions T_2' follows a F-distribution, see (5.9), (5.10) below.

If $T_2' > c_2'$ with a properly chosen critical value c_2' then H_0 must be rejected, otherwise we fail to reject H_0.

3 GMM without constraints

We start with the linear or properly linearized GMM for the u-vector of parameters x and the n-vector of observations y. The observation equations (functional model)

$$y = Ax + e \tag{3.1}$$

contain random observation errors (stochastical model)

$$e \sim (0, \sigma^2 P^{-1}). \tag{3.2}$$

We require the $n \times u$-matrix A (design matrix) and the $n \times n$-matrix P (weight matrix) to be of full rank. The least squares minimum functional without constraints reads

$$\Omega = \hat{e}^T P \hat{e} \rightarrow min \tag{3.3}$$

The classical least squares estimate without constraints is obtained as (Koch 1999, Teunissen 2001 p.44, Jäger et al. 2005 p.160)

$$\hat{x} = (A^T P A)^{-1} A^T P y \tag{3.4}$$

$$Q_{\hat{x}\hat{x}} = (A^T P A)^{-1} \tag{3.5}$$

$$\hat{e} = A\hat{x} - y \tag{3.6}$$

It is well known that the least squares estimate $\hat{x}$ is unbiased:

$$E\{\hat{x}\} = Q_{\hat{x}\hat{x}} A^T P E\{y\} = Q_{\hat{x}\hat{x}} A^T P A x = x \tag{3.7}$$

Example: A levelling line A-C (see Fig. 1) is levelled in both directions. The observations are the observed height differences between adjacent points in forward and backward direction. The parameters are the corresponding true height differences. The $n = 2u$ observation equations read

$$\begin{aligned} y_{2i-1} &= x_i + e_{2i-1} \\ y_{2i} &= x_i + e_{2i} \end{aligned}, i = 1, \dots, u \tag{3.8}$$

Typically the weights are here identical for each pair of observations:

$$p_{2i-1} = p_{2i}, i = 1, \dots, u \tag{3.9}$$

According to (3.3) - (3.6) we get the estimates

$$\hat{x}_i = \frac{y_{2i-1} + y_{2i}}{2}, i = 1, \dots, u \tag{3.10}$$

$$Q_{\hat{x}\hat{x}} = \text{diag}\left(\frac{1}{2p_i}, i = 1, \dots, u\right) \tag{3.11}$$

$$\hat{e}_{2i} = -\hat{e}_{2i-1} = \frac{y_{2i-1} - y_{2i}}{2}, i = 1, \dots, u \tag{3.12}$$

$$\Omega = 2 \sum_{i=1}^{u} p_{2i} \left(\frac{y_{2i-1} - y_{2i}}{2}\right)^2 \tag{3.13}$$

Fig. 1. Double levelled line with $n = 8, u = 4$

4 GMM with constraints for parameters

We impose the m constraints (1.1) on the parameters. Additionally we require the $u \times m$-matrix B to be of full rank.

Even though the true parameters x fulfil the constraints, i.e. the vector of true misclosures

$$w := B^T x - b \tag{4.1}$$

equals the zero vector, we cannot expect the estimated parameters $\hat{x}$ in (3.4) to do so. This is due to the effect of inevitable observation errors e in (3.1) . There would remain a vector of estimated misclosures

$$\hat{w} = B^T \hat{x} - b \tag{4.2}$$

with the cofactor matrix obtained by variance-covariance propagation as

$$Q_{\hat{w}\hat{w}} = B^T Q_{\hat{x}\hat{x}} B. \tag{4.3}$$

We must explicitly impose the constraints (1.1) on the estimated parameters. This is done by the method of LMs (2.4) : The least squares minimum functional without constraints Ω in (3.3) is extended to the Lagrange minimum functional

$$\Omega' = \hat{e}'^T P \hat{e}' + 2\hat{k}'^T (B^T \hat{x}' - b) \rightarrow min \tag{4.4}$$

The solution of the constrained optimization problem, here distinguished from the unconstrained solution by primes, is (Koch 1999, Teunissen 2001 p.86, Jäger et al. 2005 p.166)

$$\hat{k}' = (B^T Q_{\hat{x}\hat{x}} B)^{-1}(B^T \hat{x} - b) = Q_{\hat{w}\hat{w}}^{-1} \hat{w} \tag{4.5}$$

$$Q_{\hat{k}'\hat{k}'} = Q_{\hat{w}\hat{w}}^{-1} = (B^T Q_{\hat{x}\hat{x}} B)^{-1} \tag{4.6}$$

$$\hat{x}' = \hat{x} - Q_{\hat{x}\hat{x}} B \hat{k}' \tag{4.7}$$

$$Q_{\hat{x}'\hat{x}'} = Q_{\hat{x}\hat{x}} - Q_{\hat{x}\hat{x}} B Q_{\hat{k}'\hat{k}'} B^T Q_{\hat{x}\hat{x}} \tag{4.8}$$

$$\hat{e}' = A\hat{x}' - y. \tag{4.9}$$

The new estimated parameters fulfil the constraints. This is shown as follows:

$$B^T \hat{x}' = B^T \hat{x} - B^T Q_{\hat{x}\hat{x}} B \hat{k}' = B^T \hat{x} - B^T Q_{\hat{x}\hat{x}} B Q_{\hat{k}'\hat{k}'} (B^T \hat{x} - b)$$
$$= B^T \hat{x} - (B^T \hat{x} - b) = b. \tag{4.10}$$

Example (cont'd) : Now we introduce a known height difference ΔH_{AC} between end point and starting point of the levelling line in Fig. 1. This imposes a constraint for the parameters:

$$\sum_{i=1}^{u} x_i = (1 \dots 1)x = \Delta H_{AC}. \tag{4.11}$$

From (4.2) , (4.3) , (4.5) , (4.6) we derive

$$\hat{w} = (1 \dots 1)\hat{x} - \Delta H_{AC} = \sum_{i=1}^{u} \frac{y_{2i-1} + y_{2i}}{2} - \Delta H_{AC} \tag{4.12}$$

$$Q_{\hat{w}\hat{w}} = (1 \dots 1)\text{diag}\left(\frac{1}{2p_i}, i = 1, \dots, u\right)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{u} \frac{1}{2p_i} \tag{4.13}$$

$$\hat{k}' = \left(\sum_{i=1}^{u} \frac{1}{2p_i}\right)^{-1} \hat{w} \qquad (4.14)$$

$$Q_{\hat{k}'\hat{k}'} = \left(\sum_{i=1}^{u} \frac{1}{2p_i}\right)^{-1}. \qquad (4.15)$$

Taking account of the constraint, the estimated parameters $\hat{x}'_i$ are obtained from (4.7) as

$$\hat{x}'_i = \frac{y_{2i-1} + y_{2i}}{2} - \frac{\hat{w}}{p_i}\left(\sum_{j=1}^{u} \frac{1}{p_j}\right)^{-1}, i = 1, \dots, u \qquad (4.16)$$

As the result we obtain the well-known rule that the misclosure must be divided inversely proportional to the weights and subtracted from the mean observations.

It is obvious that $\hat{k}' = 0$ is equivalent to $B^T\hat{x} = b$ and to $\hat{w} = 0$. In other words, the estimated vector of LM equals the zero vector if and only if the estimated parameters $\hat{x}$ in (3.4) already fulfil the constraints.

Using (4.7) the minimum of the Lagrange functional (2.4) can be rewritten as

$$\Omega' = (A\hat{x}' - y)^T P(A\hat{x}' - y) + 2\hat{k}'^T (B^T\hat{x} - b)$$
$$= \left(A\hat{x} - AQ_{\hat{x}\hat{x}}B\hat{k}' - y\right)^T P\left(A\hat{x} - AQ_{\hat{x}\hat{x}}B\hat{k}' - y\right) + 2\hat{k}'^T (b - b)$$
$$= \left(\hat{e} - AQ_{\hat{x}\hat{x}}B\hat{k}'\right)^T P\left(\hat{e} - AQ_{\hat{x}\hat{x}}B\hat{k}'\right)$$
$$= \Omega - 2\hat{e}^T PAQ_{\hat{x}\hat{x}}B\hat{k} + \left(AQ_{\hat{x}\hat{x}}B\hat{k}'\right)^T P\left(AQ_{\hat{x}\hat{x}}B\hat{k}'\right). \qquad (4.17)$$

Utilizing the well-known equivalence $\hat{e}^T PA = 0$ and (4.5) we get the increase of the minimum value of the least squares minimum functional enforced by the constraints

$$\Omega' - \Omega = \hat{k}'^T B^T Q_{\hat{x}\hat{x}} A^T PAQ_{\hat{x}\hat{x}} B\hat{k}' = \hat{k}'^T Q_{\hat{k}'\hat{k}'}^{-1} \hat{k}' = \hat{w}^T Q_{\hat{w}\hat{w}}^{-1} \hat{w}. \qquad (4.18)$$

5 Compatibility of constraints

If the minimum value of the Lagrange functional Ω' is increased considerably with respect to Ω then this indicates that the constraints (1.1) are not almost automatically fulfiled by the information provided by the observations, but a great force is needed to constrain the adjustment. This may show that the model is not correct. Possibly not even the true parameters x fulfil the constraints, such that true misclosures (4.1) are left.

Example (cont'd) : (4.14) shows that if the estimated misclosure $\hat{w}$ is large in magnitude then so is the LM $\hat{k}'$. This may indicate that the model is not correct. A possible explanation is that the constraint and the observations are incompatible. Possibly the known height difference ΔH_{AC} is not correct.

The approach presented here is to test statistically if the constraints are compatible with the information provided by the observations. The test applied here will be equivalent to a LR test, particularly of the form (2.6) or (2.11) . In the following we assume normally distributed observation errors

$$e \sim N(0, \sigma^2 P^{-1}) \qquad (5.1)$$

as a special case of (3.2) .

5.1 Testing the compatibility of the full set of constraints

Using (3.7) , (4.7) we rewrite

$$B^T x = B^T E\{\hat{x}\} = E\{B^T \hat{x}\} = E\{B^T \hat{x}' + Q_{\hat{k}'\hat{k}'}^{-1}\hat{k}'\} = b + Q_{\hat{k}'\hat{k}'}^{-1}E\{\hat{k}'\} = b + E\{\hat{w}\} \quad (5.2)$$

we observe that $B^T x = b$ is equivalent to $E\{\hat{k}'\} = E\{\hat{w}\} = 0$. In the constrained GMM we propose the hypotheses

null hypothesis $\qquad H_0: E\{\hat{k}'\} = B^T E\{\hat{x}\} - b = 0 \quad$ vs.

alternative hypothesis $\quad H_A: E\{\hat{k}'\} = B^T E\{\hat{x}\} - b \neq 0.$

which are linear hypotheses for the parameters of the unconstrained GMM.

To begin with, assume that the prior variance factor σ^2 is known, e.g. from long standing experiences with observations of this kind. The LR test statistic (2.6) can here be written with (4.18) as

$$T_1' = \frac{\Omega' - \Omega}{\sigma^2} = \frac{\hat{k}'^T Q_{\hat{k}'\hat{k}'}^{-1}\hat{k}'}{\sigma^2} \tag{5.3}$$

Thus, T_1' can be written as a quadratic form of a vector and it's corresponding covariance matrix, both with $\hat{k}'$ and $\hat{w}$. The distribution of such a quadratic form is known to be a central or non-central χ^2-distribution with m degrees of freedom (Koch 1999, Teunissen 2000 p.125) , denoted as $\chi^2(m)$ and $\chi'^2(m, \Lambda')$ respectively. The distribution is central if and only if $E\{\hat{k}'\} = E\{\hat{w}\} = w = 0$, i.e. if H_0 holds true:

$$T_1'|H_0 \sim \chi^2\,(m) \tag{5.4}$$

$$T_1'|H_A \sim \chi'^2(m, \Lambda') \tag{5.5}$$

with the non-centrality parameter

$$\Lambda' = \frac{w^T Q_{\hat{w}\hat{w}}^{-1} w}{\sigma^2}. \tag{5.6}$$

The true misclosure w is naturally unknown. Hence, in practical application the distribution of $T_1'|H_A$ cannot be computed.

Now assume that the prior variance factor σ^2 is unknown. In this case we get with (4.18) the LR test statistic (2.11)

$$T_2' = \frac{n-u}{m} \cdot \frac{\Omega' - \Omega}{\Omega} = \frac{\hat{k}'^T Q_{\hat{k}'\hat{k}'}^{-1}\hat{k}'}{m\hat{\sigma}^2}. \tag{5.7}$$

Note that $\hat{\sigma}^2$ is the estimate of σ^2 in the unconstrained GMM.

The quadratic form $\sigma^{-2}\hat{e}^T P\hat{e}$ is known to follow a central χ^2-distribution with $n - u$ degrees of freedom (Koch 1999, Teunissen 2000 p.91) :

$$\frac{\Omega}{\sigma^2} = \frac{\hat{\sigma}^2(n-u)}{\sigma^2} \sim \chi^2(n-u). \tag{5.8}$$

Moreover, it is statistically independent from the quadratic forms in the numerator of (5.7) , (Koch 1999) . This ensures that the test statistic T_2' in (5.7) is distributed as (Koch 1999)

$$T_2'|H_0 \sim F\,(m, n-u) \tag{5.9}$$

$$T_2'|H_A \sim F'(m, n-u, \Lambda'), \tag{5.10}$$

i.e. T_2' follows a central or non-central F-distribution with m and $n - u$ degrees of freedom and with the same non-centrality parameter Λ' as in (5.6) . Again, it cannot practically be computed, because the true misclosure w is naturally unknown, this time together with the prior variance factor σ^2. Being equivalent to the LR tests (2.5) , (2.10) both tests are one-sided tests. If T_1' or T_2' exceed a critical value c_1' or c_2', respectively, then H_0 is to be rejected. The latter values are to be derived as the (1-α) -quantile of the distribution of $T_1'|H_0$ or $T_2'|H_0$, where α denotes the desired probability of a type I decision error (H_0 rejected, if it is true) .

A rejected H_0 means that there is a conflict between the observations and the constraints. Here we are inclined to assume that the true parameters do not fulfil the constraints, such that $w \neq 0$ in (4.1) .

5.2 Testing the compatibility of a single constraint

In this subsection we investigate the special case $m = 1$. Here $Q_{\widehat{w}\widehat{w}} = q_{\widehat{w}\widehat{w}}$ and $Q_{\widehat{k}'\widehat{k}'} = q_{\widehat{k}'\widehat{k}'}$ reduce to scalars. We get from (5.3) and (5.7)

$$T'_{1,m=1} = \frac{\widehat{k}'^2}{\sigma^2 q_{\widehat{k}'\widehat{k}'}} \tag{5.11}$$

$$T'_{2,m=1} = \frac{\widehat{k}'^2}{\widehat{\sigma}^2 q_{\widehat{k}'\widehat{k}'}} \tag{5.12}$$

and in the denominators we find the prior and posterior variances of $\widehat{k}'$ or $\widehat{w}$. This motivates introducing the new test statistics

$$T'_3 := \frac{\widehat{k}'}{\sigma_{\widehat{k}'\widehat{k}'}} \tag{5.13}$$

$$T'_4 := \frac{\widehat{k}'}{\widehat{\sigma}_{\widehat{k}'\widehat{k}'}}. \tag{5.14}$$

In analogy to the normalized and studentized residuals in outlier detection they can be termed **normalized and studentized LMs**. In particular, since the posterior variances are estimates in the **unconstrained** GMM, we call T'_4 **externally** studentized, in analogy to the **externally** studentized residuals in outlier detection, which are computed from the posterior variances in the GMM without the suspected outlier. (The similarity to outlier detection is more than superficial. Outliers are misspecifications of the observation equations (3.1), while true misclosures are misspecifications of the constraints (1.1).)
Since $T'_{1,m=1} = T'^2_3, T'_{2,m=1} = T'^2_4$, the new test statistics follow the distributions

$$T'_3 | H_0 \sim N(0,1) \tag{5.15}$$

$$T'_3 | H_A \sim N(\lambda', 1) \tag{5.16}$$

$$T'_4 | H_0 \sim t(n-u) \tag{5.17}$$

$$T'_4 | H_A \sim t'(n-u, \lambda') \tag{5.18}$$

where t and t' denote the central and non-central Student's t-distribution, respectively, with $n-u$ degrees of freedom. The non-centrality parameter reads

$$\lambda' = \frac{w}{\sigma q_{\widehat{w}\widehat{w}}} = \frac{w}{\sigma_{\widehat{w}\widehat{w}}}. \tag{5.19}$$

Again, it cannot be computed unless the prior variance factor σ^2 and the true misclosure w are known. At least the latter is practically out of the question.
A test using T'_3 or T'_4 is fully equivalent to the test using $T'_{1,m=1}$ or $T'_{2,m=1}$, respectively. However, both tests are two-sided tests now because T'_3 and T'_4 can be positive as well as negative. If $|T'_3|$ or $|T'_4|$ exceed critical values c'_3 or c'_4, respectively, then H_0 is to be rejected. The latter values are derived as the $(1 - \alpha/2)$-quantile of the distribution of $T'_3 | H_0$ in (5.15) and $T'_4 | H_0$ in (5.17).
Example: Let $\alpha := 0.05$. If σ^2 is known and $|T'_3| > c'_3 = \Phi^{-1}(1 - \alpha/2) = 1.96$ then H_0 is to be rejected. Here Φ denotes the cumulative distribution function (cdf) of the standard normal distribution $N(0,1)$. The critical value for $|T'_4|$ also depends on $n-u$. For example, if $n-u = 10$ then it is found from the inverse t-distribution as $c'_4 = 2.23$.
The advantages of T'_3 and T'_4 are:

- The critical values c'_3 and c'_4 can sometimes be computed more easily than c'_1 and c'_2.
- The test statistics in the form of normalized and studentized quantities can be directly interpreted as a measure of the strength of the related constraint.

Example (cont'd) : Using (4.13), (4.15), the normalized and externally studentized LMs and misclosures read

$$T_3' = \left(\sum_{i=1}^{u} \frac{1}{2p_i} \right)^{1/2} \frac{\hat{k}'}{\sigma} \tag{5.20}$$

$$T_4' = \left(\sum_{i=1}^{u} \frac{1}{2p_i} \right)^{1/2} \frac{\hat{k}}{\hat{\sigma}} \tag{5.21}$$

Using (2.8) , (3.13) we get

$$\hat{\sigma}^2 = \frac{2}{n} \sum_{i=1}^{u} p_{2i} \left(\frac{y_{2i-1} - y_{2i}}{2} \right)^2 . \tag{5.22}$$

The test statistics follow the distributions in (5.15) - (5.18) with n = 2u and

$$\lambda' = \left(\sum_{i=1}^{u} \frac{1}{2p_i} \right)^{-1/2} \frac{w}{\sigma} . \tag{5.23}$$

5.3   Testing the compatibility of a subset of constraints

The next step is to split the system of constraints into two parts:

$$B_1^T x = b_1 \tag{5.24}$$

$$B_2^T x = b_2 . \tag{5.25}$$

Suppose that the first m_1 constraints are correct. Now we test whether the second m_2 constraints are in conflict with the observations and the first part of the constraints. The LR ratio in (2.6) or (2.11) becomes a ratio of two constrained likelihood maxima: in the numerator we find the fully constrained maxima as before and in the denominator we find the semi-constrained maxima subject to the constraints $B_1^T x = b_1$ only.

Introducing the notation

$$\hat{w} = \begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix} = \begin{pmatrix} B_1^T \\ B_2^T \end{pmatrix} \hat{x} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{5.26}$$

the least squares solution in the semi-constrained GMM can be derived from (4.5) - (4.9) , (4.18) with B, b, w replaced by B_1, b_1, w_1. It is here distinguished from the previous solutions by double primes:

$$\hat{k}'' = (B_1^T Q_{\hat{x}\hat{x}} B_1)^{-1} (B_1^T \hat{x} - b_1) = Q_{\hat{w}_1 \hat{w}_1}^{-1} \hat{w}_1 \tag{5.27}$$

$$Q_{\hat{k}''\hat{k}''} = Q_{\hat{w}_1 \hat{w}_1}^{-1} = (B_1^T Q_{\hat{x}\hat{x}} B_1)^{-1} \tag{5.28}$$

$$\hat{x}'' = \hat{x} - Q_{\hat{x}\hat{x}} B_1 \hat{k}'' \tag{5.29}$$

$$Q_{\hat{x}''\hat{x}''} = Q_{\hat{x}\hat{x}} - Q_{\hat{x}\hat{x}} B_1 Q_{\hat{k}''\hat{k}''} B_1^T Q_{\hat{x}\hat{x}} \tag{5.30}$$

$$\hat{e}'' = A\hat{x}'' - y \tag{5.31}$$

$$\Omega'' - \Omega = \hat{k}''^T Q_{\hat{k}''\hat{k}''}^{-1} \hat{k}'' = \hat{w}_1^T Q_{\hat{w}_1 \hat{w}_1}^{-1} \hat{w}_1 \tag{5.32}$$

The misclosure in the first set of constraints is removed by this solution, while in the second set the misclosure is estimated to

$$\hat{w}_2'' = B_2^T \hat{x}'' - b_2 \tag{5.33}$$

with cofactor matrix

$$Q_{\hat{w}_2''\hat{w}_2''} = B_2^T Q_{\hat{x}''\hat{x}''} B_2 = Q_{\hat{w}_2 \hat{w}_2} - B_2^T Q_{\hat{x}\hat{x}} B_1 Q_{\hat{k}''\hat{k}''} B_1^T Q_{\hat{x}\hat{x}} B_2 \tag{5.34}$$

The hypotheses to be tested can now be written in various equivalent forms, e.g. as

null hypothesis $\qquad H_0: B_2^T E\{\hat{x}''\} = b_2 \qquad$ vs.

alternative hypothesis $\quad H_A: B_2^T E\{\hat{x}''\} \neq b_2$

and therefrom we can derive the LR test statistics as before:

$$T_1'' = \frac{\Omega' - \Omega''}{\sigma^2} \tag{5.35}$$

$$T_2'' = \frac{n - u + m_1}{m_2} \cdot \frac{\Omega' - \Omega''}{\Omega''} = \frac{\Omega' - \Omega''}{m_2 \hat{\sigma}''^2} \tag{5.36}$$

with

$$\Omega' - \Omega'' = (\Omega' - \Omega) - (\Omega'' - \Omega) = \hat{w}^T Q_{\hat{w}\hat{w}}^{-1} \hat{w} - \hat{w}_1^T Q_{\hat{w}_1 \hat{w}_1}^{-1} \hat{w}_1$$

$$= \begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix}^T \left[\begin{pmatrix} Q_{\hat{w}_1 \hat{w}_1} & Q_{\hat{w}_1 \hat{w}_2} \\ Q_{\hat{w}_1 \hat{w}_2}^T & Q_{\hat{w}_2 \hat{w}_2} \end{pmatrix}^{-1} - \begin{pmatrix} Q_{\hat{w}_1 \hat{w}_1}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right] \begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix}. \tag{5.37}$$

$\Omega' - \Omega''$ and Ω'' enjoy the same nice properties as $\Omega' - \Omega$ and Ω. This can be understood by considering the following procedure: $B_1^T x = b_1$ could be solved for m_1 suitable parameters, and these parameters could be eliminated also from the observation equations (3.1) . Then the semi-constrained GMM is transformed into an unconstrained GMM for the remaining $u - m_1$ parameters having the same solution. This shows that (5.8) can be modified to

$$\frac{\Omega''}{\sigma^2} = \frac{\hat{\sigma}''^2 (n - u + m_1)}{\sigma^2} \sim \chi^2 (n - u + m_1). \tag{5.38}$$

Now the m_1 parameters are also eliminated from $B_2^T x = b_2$. Then the fully constrained GMM is transformed into a constrained GMM for the remaining $u - m_1$ parameters and m_2 constraints with the same solution. Consequently, we get the increase of the minimum value of the least squares minimum functional enforced by the constraints $B_2^T x = b_2$ in analogy to (4.18) as

$$\Omega' - \Omega'' = \hat{w}_2''^T Q_{\hat{w}_2'' \hat{w}_2''}^{-1} \hat{w}_2''. \tag{5.39}$$

This shows that (5.4) , (5.5) can be modified to

$$T_1'' | H_0 \sim \chi^2 (m_2) \tag{5.40}$$

$$T_1'' | H_A \sim \chi'^2 (m_2, \Lambda'') \tag{5.41}$$

with the non-centrality parameter

$$\Lambda'' = \frac{w_2^T Q_{\hat{w}_2'' \hat{w}_2''}^{-1} w_2}{\sigma^2}. \tag{5.42}$$

By the same line of reasoning as for T_2' we obtain from (5.9) , (5.10) for T_2'' the distributions

$$T_2'' | H_0 \sim F (m_2, n - u + m_1) \tag{5.43}$$

$$T_2'' | H_A \sim F'(m_2, n - u + m_1, \Lambda''). \tag{5.44}$$

5.4   Testing the compatibility of an individual constraint

In the following we want to specify the test elaborated in the preceding subsection for the null hypothesis that in a set of m constraints for parameters an individual constraint is compatible with the rest of the constraints and the observations vs. the alternative hypothesis that there is a true misclosure in that particular constraint. This is a special case of the hypotheses tested in the preceding subsection for $m_1 = m - 1, m_2 = 1$. Assume that the constraint under consideration is moved to the end of the set. Some submatrices of $Q_{\hat{k}'\hat{k}'}$ and $Q_{\hat{w}\hat{w}}$ become vectors and scalars. This will be made more clear by changing the notation from Q to q:

$$Q_{\hat{k}'\hat{k}'} = \begin{pmatrix} Q_{\hat{k}_1'\hat{k}_1'} & q_{\hat{k}_1'\hat{k}_2'} \\ q_{\hat{k}_1'\hat{k}_2'}^T & q_{\hat{k}_2'\hat{k}_2'} \end{pmatrix}, Q_{\hat{w}\hat{w}} = \begin{pmatrix} Q_{\hat{w}_1\hat{w}_1} & q_{\hat{w}_1\hat{w}_2} \\ q_{\hat{w}_1\hat{w}_2}^T & q_{\hat{w}_2\hat{w}_2} \end{pmatrix}. \tag{5.45}$$

By block matrix inversion (Koch 1999, Jäger et al. p.35) we get from (4.6)

$$Q_{\hat{k}'\hat{k}'} = Q_{\hat{w}\hat{w}}^{-1} = \begin{pmatrix} Q_{\hat{w}_1\hat{w}_1}^{-1} + q_{\hat{k}_2'\hat{k}_2'} Q_{\hat{w}_1\hat{w}_1}^{-1} q_{\hat{w}_1\hat{w}_2} q_{\hat{w}_1\hat{w}_2}^{T} Q_{\hat{w}_1\hat{w}_1}^{-1} & -q_{\hat{k}_2'\hat{k}_2'} Q_{\hat{w}_1\hat{w}_1}^{-1} q_{\hat{w}_1\hat{w}_2} \\ -q_{\hat{k}_2'\hat{k}_2'} q_{\hat{w}_1\hat{w}_2}^{T} Q_{\hat{w}_1\hat{w}_1}^{-1} & q_{\hat{k}_2'\hat{k}_2'} \end{pmatrix} \quad (5.46)$$

with

$$q_{\hat{k}_2'\hat{k}_2'} = \left(q_{\hat{w}_2\hat{w}_2} - q_{\hat{w}_1\hat{w}_2}^{T} Q_{\hat{w}_1\hat{w}_1}^{-1} q_{\hat{w}_1\hat{w}_2} \right)^{-1} \quad (5.47)$$

being the last scalar element of matrix $Q_{\hat{k}'\hat{k}'}$. Inserting this into (5.37) and rearranging yields

$$\Omega' - \Omega'' = q_{\hat{k}_2'\hat{k}_2'} \begin{pmatrix} \widehat{w}_1 \\ \widehat{w}_2 \end{pmatrix}^{T} \begin{pmatrix} Q_{\hat{w}_1\hat{w}_1}^{-1} q_{\hat{w}_1\hat{w}_2} q_{\hat{w}_1\hat{w}_2}^{T} Q_{\hat{w}_1\hat{w}_1}^{-1} & -Q_{\hat{w}_1\hat{w}_1}^{-1} q_{\hat{w}_1\hat{w}_2} \\ -q_{\hat{w}_1\hat{w}_2}^{T} Q_{\hat{w}_1\hat{w}_1}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \widehat{w}_1 \\ \widehat{w}_2 \end{pmatrix}$$

$$= q_{\hat{k}_2'\hat{k}_2'} \left(\widehat{w}_2 - q_{\hat{w}_1\hat{w}_2}^{T} Q_{\hat{w}_1\hat{w}_1}^{-1} \widehat{w}_1 \right)^2 = \frac{\left(q_{\hat{k}_1'\hat{k}_2'}^{T} \widehat{w}_1 + q_{\hat{k}_2'\hat{k}_2'} \widehat{w}_2 \right)^2}{q_{\hat{k}_2'\hat{k}_2'}} = \frac{\hat{k}_2'^2}{q_{\hat{k}_2'\hat{k}_2'}}. \quad (5.48)$$

Now the test statistics (5.35) , (5.36) simplify to

$$T_{1,m_2=1}'' = \frac{\hat{k}_2'^2}{\sigma^2 q_{\hat{k}_2'\hat{k}_2'}} = \frac{\hat{k}_2'^2}{\sigma_{\hat{k}_2'\hat{k}_2'}^2} \quad (5.49)$$

$$T_{2,m_2=1}'' = \frac{\hat{k}_2'^2}{\hat{\sigma}''^2 q_{\hat{k}_2'\hat{k}_2'}} = \frac{\hat{k}_2'^2}{\hat{\sigma}_{\hat{k}_2'\hat{k}_2'}''^2}. \quad (5.50)$$

In analogy to (5.13) , (5.14) this motivates introducing the new test statistics

$$T_3'' := \frac{\hat{k}_2'}{\sigma_{\hat{k}_2'\hat{k}_2'}} \quad (5.51)$$

$$T_4'' := \frac{\hat{k}_2'}{\hat{\sigma}_{\hat{k}_2'\hat{k}_2'}''}. \quad (5.52)$$

They can be termed **individual normalized and externally studentized** LMs. Note that the studentization must be made using the variance estimate $\hat{\sigma}_{\hat{k}_2'\hat{k}_2'}''$ of the **semi-constrained** GMM. This is indicated by the adjective "external". Since $T_{1,m_2=1}'' = T_3''^2, T_{2,m_2=1}'' = T_4''^2$, a test using T_3'' or T_4'' is fully equivalent to a test using $T_{1,m_2=1}''$ or $T_{2,m_2=1}''$. The advantages of T_3'' or T_4'' are the same as of T_3' or T_4', see subsection 5.2.

The distribution of T_3'' and T_4'' can be derived from (5.40) , (5.41) , (5.43) , (5.44) with $m_2 = 1$ as follows:

$$T_3''|H_0 \sim N(0,1) \quad (5.53)$$

$$T_3''|H_A \sim N(\lambda'', 1) \quad (5.54)$$

$$T_4''|H_0 \sim t(n - u + m - 1) \quad (5.55)$$

$$T_4''|H_A \sim t'(n - u + m - 1, \lambda'') \quad (5.56)$$

with the non-centrality parameter

$$\lambda'' = \frac{w_2}{\sigma \sqrt{q_{\hat{w}_2''\hat{w}_2''}}} = \frac{w_2}{\sigma_{\hat{w}_2''\hat{w}_2''}}. \quad (5.57)$$

An alternate version of this formula can be found from $\lambda'' = E\{T_3''|H_A\}$ and (4.5) as

$$\lambda'' = \frac{E\{\hat{k}_2'|H_A\}}{\sigma_{\hat{k}_2'\hat{k}_2'}} = \frac{q_{\hat{k}_1'\hat{k}_2'}^{T} E\{\widehat{w}_1|H_A\} + q_{\hat{k}_2'\hat{k}_2'} E\{\widehat{w}_2|H_A\}}{\sigma_{\hat{k}_2'\hat{k}_2'}} = \frac{q_{\hat{k}_2'\hat{k}_2'} w_2}{\sigma_{\hat{k}_2'\hat{k}_2'}} = \frac{w_2 \sqrt{q_{\hat{k}_2'\hat{k}_2'}}}{\sigma}. \quad (5.58)$$

Example (cont'd) : Assume that on the double levelled line from Fig. 1 there is another known point B, see Fig. 2, such that we get besides (4.11) a second constraint

$$\sum_{i=1}^{v} x = (1 \dots 1\, 0 \dots 0)x = \Delta H_{AB}. \tag{5.59}$$

The latter constraint is possibly not correct due to a suspected mistake in ΔH_{AB}. The constraint matrix and vector read

$$B^T = \begin{pmatrix} 1 & \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}, b = \begin{pmatrix} \Delta H_{AC} \\ \Delta H_{AB} \end{pmatrix}, \tag{5.60}$$

where the second row of B^T has v ones and $u - v$ zeros. We get new estimates of the misclosures

$$\widehat{w}_1 = \sum_{i=1}^{u} \frac{y_{2i-1} + y_{2i}}{2} - \Delta H_{AC} \tag{5.61}$$

$$\widehat{w}_2 = \sum_{i=1}^{v} \frac{y_{2i-1} + y_{2i}}{2} - \Delta H_{AB}. \tag{5.62}$$

In order to simplify this example as much as possible, assume that we have overall equal weights $P = I$. From (4.6), (3.11), (5.60) we get

$$Q_{\widehat{k}'\widehat{k}'} = Q_{\widehat{w}\widehat{w}}^{-1} = 2 \begin{pmatrix} u & v \\ v & v \end{pmatrix}^{-1} = \frac{2}{v(u-v)} \begin{pmatrix} v & -v \\ -v & u \end{pmatrix} \tag{5.63}$$

$$\sigma_{\widehat{k}_2'\widehat{k}_2'} = \sigma\sqrt{q_{\widehat{k}_2'\widehat{k}_2'}} = \sigma\sqrt{\frac{2u}{v(u-v)}} \tag{5.64}$$

$$\widehat{k}_2' = \frac{2}{v(u-v)}(u\widehat{w}_2 - v\widehat{w}_1) \tag{5.65}$$

$$T_3'' = \frac{\sqrt{2}(u\widehat{w}_2 - v\widehat{w}_1)}{\sigma\sqrt{uv(u-v)}}. \tag{5.66}$$

The solution of the semi-constrained GMM only taking account of the constraint (4.11) reads

$$q_{\widehat{k}''\widehat{k}''} = q_{\widehat{w}_1\widehat{w}_1}^{-1} = \frac{2}{u} \tag{5.67}$$

$$\widehat{k}'' = \frac{2}{u}\widehat{w}_1 \tag{5.68}$$

$$\Omega'' = \Omega' + \frac{\widehat{w}_1^2}{q_{\widehat{w}_1\widehat{w}_1}} = 2\sum_{i=1}^{u} \left(\frac{y_{2i} - y_{2i-1}}{2}\right)^2 + \frac{2}{u}\widehat{w}_1^2$$

$$= \frac{1}{2}\sum_{i=1}^{u}(y_{2i} - y_{2i-1})^2 + \frac{2}{u}\left(\sum_{i=1}^{2u} \frac{y_i}{2}\right)^2 \tag{5.69}$$

$$\frac{\Omega''}{\sigma^2} = \frac{1}{2\sigma^2}\sum_{i=1}^{u}(e_{2i} - e_{2i-1})^2 + \frac{2}{u\sigma^2}\left(w_1 + \sum_{i=1}^{2u} \frac{e_i}{2}\right)^2 \tag{5.70}$$

$$T_4'' = \frac{\sqrt{2}(u\widehat{w}_2 - v\widehat{w}_1)}{\sqrt{uv(u-v)}}\sqrt{\frac{u+1}{\Omega''}} = \frac{\sqrt{2u+2}(u\widehat{w}_2 - v\widehat{w}_1)}{\sqrt{\Omega''uv(u-v)}} \tag{5.71}$$

The test statistics follow the distributions (5.53) - (5.56) with $n = 2u$, $m = 2$. From (5.58) and (5.63) we get the non-centrality parameter

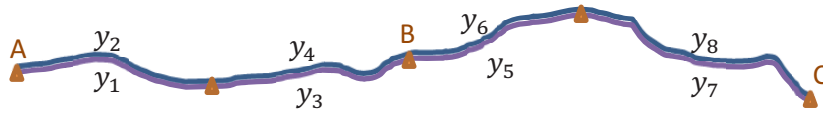$$\lambda'' = \frac{w_2}{\sigma}\sqrt{\frac{2u}{v(u-v)}}. \tag{5.72}$$

Fig. 2. Double levelled line with $n = 8, u = 4, v = 2$

5.5 Testing the compatibility of multiple individual constraints

Assume we have in a GMM a set of m constraints (1.1) . If we do not know if any individual constraint is in conflict with the observations and the rest of the constraints then we can naturally propose the hypotheses:

H_0: No individual constraint is in conflict with the observations and the rest of the constraints.

vs.

H_A: At least one individual constraint is in conflict with the observations and the rest of the constraints.

This H_A can be opposed to a set of m hypotheses $H_A^{(1)}, \ldots, H_A^{(m)}$ with

$H_A^{(i)}$: The ith constraint is in conflict with the observations and the rest of the constraints.

If any of the $H_A^{(i)}$ is true, so is H_A. If all $H_A^{(i)}$ are false, so is H_A. If H_0 is tested against $H_A^{(i)}$ then a proper test statistic is either $T_3''^{(i)}$ or $T_4''^{(i)}$ where (i) indicates that the ith constraint is to be tested rather than the last one, as done in the preceding subsection. If any of the test statistics exceeds its critical value then we are inclined to reject H_0 in this particular test H_0 vs. $H_A^{(i)}$ and consequently reject H_0 also in the test H_0 vs. H_A:

$$\left|T_3''^{(i)}\right| > c_3''(\alpha') \text{ for any } i = 1, \ldots, m \ \to \text{reject } H_0 \tag{5.73}$$

or

$$\left|T_4''^{(i)}\right| > c_4''(\alpha') \text{ for any } i = 1, \ldots, m \ \to \text{reject } H_0 \tag{5.74}$$

Here $c_3''(\alpha'), c_4''(\alpha')$ denote the related critical values for a probability of a type I decision error α'. They can be computed from the inverse cdf of $N(0,1)$ and $t(n - u + m - 1)$ at $1 - \alpha'/2$. Now we introduce **extreme normalized and externally studentized LMs** as follows:

$$T_5'' := \max_{i=1,\ldots,m}\left|T_3''^{(i)}\right| \tag{5.75}$$

$$T_6'' := \max_{i=1,\ldots,m}\left|T_4''^{(i)}\right| \tag{5.76}$$

such that (5.73) and (5.74) are equivalent to

$$T_5'' > c_3''(\alpha') \to \text{reject } H_0 \tag{5.77}$$

or

$$T_6'' > c_4''(\alpha') \to \text{reject } H_0 \tag{5.78}$$

respectively. Thus, extreme normalized and externally studentized LMs immediately act as test statistics for the test of H_0 vs. H_A. However, the related probability of a type I decision error is not α', as will be demonstrated now:

Unfortunately, the distributions of T_5'' and T_6'' are more complicated than those of $T_3''^{(i)}$ and $T_4''^{(i)}$. This is due to the complicated nonlinear structure of (5.75), (5.76). But we can try to apply the common approximations well known from normalized residuals in outlier detection (Baarda 1968, Koch 1999, Lehmann 2012) : If the m random events in (5.73), (5.74) were nearly independent then we could approximately write

$$P(T_5'' < c_3''(\alpha')) \approx \prod_{i=1}^{m} P\left(-c_3''(\alpha') < T_3''^{(i)} < c_3''(\alpha')\right) \tag{5.79}$$

$$P(T_6'' < c_4''(\alpha')) \approx \prod_{i=1}^{m} P\left(-c_4''(\alpha') < T_4''^{(i)} < c_4''(\alpha')\right) \tag{5.80}$$

All those probabilities P are probabilities of a type I decision error α and α' of the related test statistics T_5'', T_6'' and $T_3''^{(i)}, T_4''^{(i)}$, respectively. This yields in both cases

$$1 - \alpha \approx \prod_{i=1}^{m}(1 - \alpha') = (1 - \alpha')^m \tag{5.81}$$

This is to say: The test with test statistic T_5'' or T_6'' and probability of type I decision error α is approximated by a **family** of m tests with test statistic $T_3''^{(i)}$ or $T_4''^{(i)}$, $i = 1, \ldots, m$ and probability of type I decision error α', respectively. Since practically we chose $\alpha \ll 1$, we find from (5.81) with good accuracy the relationship

$$\alpha \approx m\alpha' \tag{5.82}$$

It is called **Bonferroni equation** (cf. Abdi 2007) . If these approximations can be applied then we may replace α' in (5.77), (5.78) by α/m.

However, even though H_0 is true, it may happen that a large observation error occurs, and a considerable force is required to fulfil the constraints. Typically this affects multiple constraints at a time. Thus, if one LM is large in magnitude, also other LMs tend to be large in magnitude and vice versa. This effect has been disregarded by the approximation in (5.79), (5.80). In this case and if one LM is small then we still suppose that the "neighbouring" LM can be large in magnitude. In this way using the approximations in (5.79), (5.80) we overestimate the probability of large extreme normalized LMs. This effect has been demonstrated by Lehmann (2012) for extreme normalized and studentized residuals, and it arises here as well.

In the case that the approximations in (5.79), (5.80) are too coarse or that it is not known or too cumbersome to find out if they can really be used, we recommend the Monte Carlo method, which directly computes the critical values $c_5''(\alpha), c_6''(\alpha)$ of the test statistic T_5'', T_6'' for a desired α. The method was demonstrated and applied to extreme normalized and externally studentized residuals for outlier detection by Lehmann (2012). (5.77), (5.78) are then replaced by

$$T_5'' > c_5''(\alpha) \rightarrow \text{reject } H_0 \tag{5.83}$$

and

$$T_6'' > c_6''(\alpha) \rightarrow \text{reject } H_0 \tag{5.84}$$

respectively.

Example (cont'd) : If also the constraint (4.11) is to be tested then we derive the related test statistic $T_3''^{(1)}$ in the same way as for the constraint (5.59) by

$$\hat{k}'_1 = \frac{2}{u-v}(\hat{w}_1 - \hat{w}_2) \tag{5.85}$$

$$\sigma_{\hat{k}'_1 \hat{k}'_1} = \sigma\sqrt{q_{\hat{k}'_1 \hat{k}'_1}} = \sigma\sqrt{\frac{2}{u-v}} \tag{5.86}$$

$$T_3''^{(1)} = \frac{\sqrt{2}(\hat{w}_1 - \hat{w}_2)}{\sigma\sqrt{u-v}} \tag{5.87}$$

(It is instructive to try to understand why this test statistic assumes such a form. $T_3''^{(1)}$ belongs to the alternative hypothesis that ΔH_{AB} is correct, but ΔH_{AC} is not. Thus, $\Delta H_{BC} = \Delta H_{AC} - \Delta H_{AB}$ is incorrect, and this is best tested using the related estimated misclosure $\hat{w}_1 - \hat{w}_2$ in (5.87).)
From (5.58) and (5.63) we get the non-centrality parameter

$$\lambda''^{(1)} = \frac{w_2}{\sigma}\sqrt{\frac{2}{u-v}} \tag{5.88}$$

T_3'' in (5.66), T_4'' in (5.71) and λ'' in (5.72) are now renamed as $T_3''^{(2)}$, $T_4''^{(2)}$ and $\lambda''^{(2)}$, respectively. From (5.79) we conclude that the joint vector of individual test statistics

$$\begin{pmatrix} T_3''^{(1)} \\ T_3''^{(2)} \end{pmatrix} = \frac{1}{\sigma}\sqrt{\frac{2}{uv(u-v)}}\begin{pmatrix} \sqrt{uv} & -\sqrt{uv} \\ -v & u \end{pmatrix}\begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix} \tag{5.89}$$

is bivariate normally distributed

$$\begin{pmatrix} T_3''^{(1)} \\ T_3''^{(2)} \end{pmatrix}\bigg|H_0 \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma\right) \tag{5.90}$$

with covariance matrix derived by covariance propagation

$$\Sigma := \frac{1}{uv(u-v)}\begin{pmatrix} \sqrt{uv} & -\sqrt{uv} \\ -v & u \end{pmatrix}\begin{pmatrix} u & v \\ v & v \end{pmatrix}\begin{pmatrix} \sqrt{uv} & -\sqrt{uv} \\ -v & u \end{pmatrix}^T$$

$$= \begin{pmatrix} 1 & -\sqrt{v/u} \\ -\sqrt{v/u} & 1 \end{pmatrix} \tag{5.91}$$

and thus $T_3''^{(1)}$ and $T_3''^{(2)}$ are correlated with correlation coefficient $-\sqrt{v/u}$. This shows that (5.79) can be applied only when v is small. If $v > u/2$ then the second constraint is better replaced by the equivalent constraint "along the shorter way". In the worst case $v = u/2$ and for example $\alpha := 0.05$ we get $c = c_3''(\alpha/2) = \Phi^{-1}(1-\alpha/4) = 2.24$. In such a small scale problem we can perform an exact calculation without the Monte Carlo method as follows:

$$P(T_5'' > c|H_0) = 1 - P(T_5'' \le c|H_0) = 1 - P\left(\left|T_3''^{(1)}\right| \le c \wedge \left|T_3''^{(2)}\right| \le c|H_0\right)$$
$$= 1 - \Phi_{N(0,\Sigma)}(c,c) + \Phi_{N(0,\Sigma)}(c,-c) + \Phi_{N(0,\Sigma)}(-c,c) \tag{5.92}$$
$$- \Phi_{N(0,\Sigma)}(-c,-c)$$

where $\Phi_{N(0,\Sigma)}$ denotes the cdf of (5.90). This has been evaluated by the MATLAB function mvncdf. The result is $P(T_5'' > 2.24) = 0.043 \ne \alpha = 0.05$. This shows that 2.24 is not the correct critical value for T_5''. A more detailed computation would reveal that the desired critical value is $c_5''(\alpha) = 2.18$. The deviation of the approximate value 2.24 is oftentimes tolerable.
Such a computation would be infeasible when m is large due to excessive numerical expenditure, but the Monte Carlo method would still be an option.

6 Conclusions

We have demonstrated how measures of the strength of a constraint can be derived as test statistics of likelihood ratio tests. This gives those measures a profound justification. We have derived test statistics for compatibility

- of the full set of constraints (5.3) and (5.7)
- of a subset of constraints (5.35) and (5.36)
- of a specific individual constraint in the form of
  - individual normalized LMs (5.51) and
  - individual externally studentized LMs (5.52)
- of multiple individual constraints in the form of
  - extreme normalized LMs (5.75) and
  - extreme externally studentized LMs (5.76) .

All those test statistics have known distributions under the null hypothesis and also under the alternative hypothesis, except for the unknown non-centrality parameter. For the extreme LMs we derived the distributions only with the approximations (5.79) , (5.80) . Using other plausible test statistics like those employing the estimated misclosures instead of the LMs gives less powerful test results.

A number of extensions of the work presented here are required:

- We should try to use internally instead of externally studentized LMs. There is a strict relationship between both quantities. Internally studentized quantities are practically more appealing, but have more complicated distributions (so-called τ-distributions) .
- We should extend the derivation to matrices A and B not of full rank, as the appear in free network adjustments.
- We should use the Monte Carlo method for the computation of critical values of extreme normalized and studentized LMs and of power functions.
- We should try to derive the results also in such a way that the constraints are given as observation equations by letting the weights of these observations go to infinity.
- Finally, we mention that here we have exclusively considered non-compatibilities of constraints caused by non-random biases w. It is worth considering the case that they are caused by random effects. For example, a levelling line has a misclosure w because a levelling rod was by chance not erected on firm ground. Then w is a random variable and the distributions of all test statistics under the alternative hypothesis change.

Potential applications of normalized and studentized LMs are all adjustment problems with equality constraints for parameters: geodetic network adjustment, coordinate transformations, photogrammetric bundle adjustment etc.

References

Abdi H   (2007) The Bonferonni and Šidák Corrections for Multiple Comparisons. In: Neil Salkind (Ed.) Encyclopedia of Measurement and Statistics. Sage Thousand Oaks (CA)

Baarda W (1968) A testing procedure for use in geodetic networks. Netherlands Geodetic Commission, Publication on geodesy, 2 (5) , Delft, Netherlands

Bussotti P (2003) On the Genesis of the Lagrange multipliers. Journal of Optimization Theory and Applications: Vol. 117, No. 3, pp. 453–459

Clements KA, Costa AS (1998) Topology Error Identification Using Normalized LMs. IEEE Transactions on Power Systems, Vol. 13, No. 2

Engle RF (1983) . Wald, Likelihood Ratio, and LM Tests in Econometrics. In: MD Intriligator and Z Griliches: Handbook of Econometrics. II. Elsevier. pp. 796–801. ISBN 978-0-444-86185-6

Helmert FR (1872) Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate: mit Anwendungen auf die Geodäsie und die Theorie der Messinstrumente. Verlag Teubner Leipzig.

Jäger R, Müller T, Saler H, Schwäble R (2005) Klassische und robuste Ausgleichungsverfahren: ein Leitfaden für die Ausbildung und Praxis von Geodäten und Geoinformatikern. Herbert Wichmann Verlag, Heidelberg

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

Krakiwsky EJ, Biacs ZF (1990) Least squares collocation and statistical testing. Bull Geod 64:73-87; DOI: 10.1007/BF02530616

Koch KR (1999) Parameter Estimation and Hypothesis Testing in Linear Models. Springer Verlag Berlin Heidelberg New York

Koch KR (2012) Robust estimation by expectation maximization algorithm. J Geod. DOI: 10.1007/s00190-012-0582-3

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

Lourenço E, Clements K, Costa AS (2004) Bayesian-Based Hypothesis Testing for Topology Error Identification in Generalized State Estimation. IEEE Transactions on Power Systems, Vol. 19, No. 2, pp 1206-1215

Roese-Koerner L, Devaraju B, Sneeuw N, Schuh WD (2012) A stochastic framework for inequality constrained estimation. J Geod. DOI: 10.1007/s00190-012-0560-9

Schaffrin B (1981) Ausgleichung mit Bedingungsungleichungen. Allgemeine Vermessungsnachrichten 88, 227–238

Schaffrin B, Felus Y (2005) On Total Least-Squares Adjustment with Constraints. International Association of Geodesy Symposia, 2005, Volume 128, Symposium G04, 417-421, DOI: 10.1007/3-540-27432-4_71

Silvey SD (1959) The Lagrangian Multiplier Test. Annals of Mathematical Statistics 30, 389-407

Teunissen PJG (1985) Quality control in geodetic networks. In: E.W. Grafarend and F. Sanso (eds.) Optimization and Design of Geodetic Networks, pp. 526–547, Springer, Berlin Heidelberg New York

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

Teunissen PJG (2001) Adjustment theory; an introduction. 2nd edition. Series on Mathematical Geodesy and Positioning, Delft University of Technology, The Netherlands. ISBN 90-407-1974-8

Wald A (1943) Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large, Transactions of the American Mathematical Society 54, 426-482

Wziontek H, Wienholz K, Lelgemann D (2001) On the interpretation of Lagrange's multipliers as a measure of constraint in adjustment problems. In: OM Altan and L Gründig (eds.) : Fourth International Symposium "Turkish-German Joint Geodetic Days", pp. 777–783, Berlin. Technische Universität Berlin

Wienholz K (2002) Zur Bestimmung der GPS-Phasenmehrdeutigkeiten in großräumigen Netzen. Deutsche Geodätsche Kommission Reihe C, Nr. 566, München