Statistical methods in geodesy

15

> [. . .] There has not been a single date in the history of the theory of gravitation when a modern significance test would not have rejected all laws and left us with no law. Nevertheless the law did lead to improvement for centuries, [. . .]
>
> Harold Jeffreys, 1939 (Jeffreys, 1998, page 391)

IN THIS CHAPTER we present two interrelated subjects:

- Statistical testing, in the context of the validation of and outlier detection in geodetic network measurements. poikkeava havaintoarvo
- The reliability of geodetic networks.

The framework of hypothesis testing with null and alternative hypotheses is adopted. We also show how this framework may be used, for example for geodetic deformation analysis.

We would be amiss in not pointing out that hypothesis testing is not the appropriate framework for settling all scientific disputes. There are other, often more appropriate, techniques, like the Akaike information criterion (Burnham and Anderson, 2013) and Bayesian approaches. The above quote from Harold Jeffreys[1] is apposite. 1

15.1 The method of least-squares

Explaining the method of least-squares is simplest if one assumes that all observables are stochastic quantities that are *normally distributed* (figure 2.5), both individually and together: they form a *multivariate* pienimmän neliösumman menetelmä

[1] Sir Harold Jeffreys FRS (1891–1989) was a British mathematician, statistician, geophysicist, and astronomer, an influential advocate of Bayesian statistics.

normal distribution. If normally distributed observables are statistically independent of each other — for example, if they were produced by independent measurement processes — this is automatically the case.

The method of least-squares as a way to minimise the random errors in estimated quantities such as point co-ordinates has been presented above. In chapter 14, the *parametric adjustment method* was presented, in which observations are expressed as functions of the unknowns. The alternative, adjustment by conditions, is suitable, for example, for computing traverses.

parametrinen tasoitus

ehtotasoitus

monikulmiojono

Here, the parametric method will be discussed in more detail.

Let the *observations*, as a vector $\underline{\ell}$, be linear functions of the *unknowns* x:

$$\overbrace{\begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ \vdots \\ \underline{\ell}_n \end{bmatrix}}^{\underline{\ell}}_{[n]} = \overbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}}^{A}_{[n\times m]} \overbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}}^{\mathrm{x}}_{[m]} + \overbrace{\begin{bmatrix} \underline{n}_1 \\ \underline{n}_2 \\ \vdots \\ \underline{n}_n \end{bmatrix}}^{\underline{\mathrm{n}}}_{[n]} .$$

Here, n is the number of observations, m the number of unknowns. The observations, elements of the vector $\underline{\ell}$, are stochastic quantities. Assume that they are normally distributed around the "true" value of the observed quantity. Then, the elements of the vector of *observation errors* $\underline{\mathrm{n}}$ are also normally distributed.[2]

[2]

pienimmän neliösumman ratkaisu

In this, rather general, case we may compute the *least-squares solution* in the following way:

$$\underset{[m]}{\widehat{\mathrm{x}}} = \underbrace{\left(A^{\mathsf{T}} Q_{\ell\ell}^{-1} A\right)^{-1}}_{[m\times m]^{-1}} \underbrace{\left(A^{\mathsf{T}} Q_{\ell\ell}^{-1}\right)}_{[m\times n]} \underset{[n]}{\underline{\ell}}, \tag{15.1}$$

in which $Q_{\ell\ell}$ is the *weight-coefficient matrix* of the observations, size $[n \times n]$:

$$Q_{\ell\ell} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix} .$$

[2] Often, they are also assumed to be statistically independent from each other, meaning that their random variations happen independently of each other. However, neither the elements of solution vector $\widehat{\mathrm{x}}$ nor those of the vector of residuals $\underline{\mathrm{v}}$ will be statistically independent of each other.

This matrix is related to the variance matrix of the observations as follows:

$$\mathrm{Var}\{\underline{\ell}\} \stackrel{\text{def}}{=} \Sigma_{\ell\ell} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} = \sigma^2 Q_{\ell\ell},$$

in which

$$\sigma_i^2 = \mathrm{Var}\{\underline{\ell}_i\} = E\left\{\left(\underline{\ell}_i - E\{\underline{\ell}_i\}\right)^2\right\} = \sigma^2 q_{ii},$$

$$\sigma_{ij} = \mathrm{Cov}\{\underline{\ell}_i, \underline{\ell}_j\} = E\left\{\left(\underline{\ell}_i - E\{\underline{\ell}_i\}\right)\left(\underline{\ell}_j - E\{\underline{\ell}_j\}\right)\right\} = \sigma^2 q_{ij}.$$

Here, σ is the mean error of unit weight.

painoyksikön keskivirhe

The variance matrix of the solution is obtained though propagation of variances. Let

varianssien kasautumislaki

$$\widehat{\mathrm{x}} = L\underline{\ell},$$

in which

$$L \stackrel{\text{def}}{=} \left(A^{\mathsf{T}} Q_{\ell\ell}^{-1} A\right)^{-1} A^{\mathsf{T}} Q_{\ell\ell}^{-1}.$$

Then, based on equation 15.1:

$$Q_{\mathrm{xx}} = L Q_{\ell\ell} L^{\mathsf{T}} = \left(A^{\mathsf{T}} Q_{\ell\ell}^{-1} A\right)^{-1} A^{\mathsf{T}} Q_{\ell\ell}^{-1} \cdot Q_{\ell\ell} \cdot Q_{\ell\ell}^{-1} A \left(A^{\mathsf{T}} Q_{\ell\ell}^{-1} A\right)^{-1} = \\ = \left(A^{\mathsf{T}} Q_{\ell\ell}^{-1} A\right)^{-1}$$

by suitable elimination. So, the variance matrix of the solution vector $\widehat{\mathrm{x}}$, $\Sigma_{\mathrm{xx}} = \sigma^2 Q_{\mathrm{xx}}$, is obtained in any case as a side product of computing the solution, equation 15.1!

15.2 The residuals from the adjustment

The least-squares estimators of the observations $\widehat{\ell}$ and unknowns $\widehat{\mathrm{x}}$ are connected to each other through the functional model

$$\widehat{\ell} = A\widehat{\mathrm{x}},$$

and from the original observations

$$\underline{\ell} = A\mathrm{x} + \underline{\mathrm{n}}$$

one computes the *residuals*:[3]

jäännösvirhe 3

$$\underline{\mathrm{v}} \stackrel{\text{def}}{=} \widehat{\ell} - \underline{\ell} = A\widehat{\mathrm{x}} - \underline{\ell} = A\left(\widehat{\mathrm{x}} - \mathrm{x}\right) - \underline{\mathrm{n}}.$$

Residuals are central in the quality control of geodetic network solutions.

karkea virhe

- The size of the residuals tells something about the contradictions present in the network solution, possible gross errors, or even model errors.
- The size of the residual of a certain observation can indicate whether there may be a gross error hiding in this observation.
- The network must be *reliable*: there has to be redundancy, an over-determination by the observational material. For example, all kinds of closing errors offer possibilities for checking.

  With little or no redundancy, the residuals may well be small or even zero, but this means nothing!

An often-used form of the *observation equations* is

$$\underline{\ell} + \underline{v} = A\widehat{\underline{x}}.$$

pienimmän neliösumman tasoitus

The residuals $\underline{v}$ of a least-squares adjustment have four nice properties, here given without proof:

1. The quadratic form

$$\mathcal{E} \overset{\text{def}}{=} \underline{v}^{\mathsf{T}} P_{\ell\ell} \underline{v} = \underline{v}^{\mathsf{T}} Q_{\ell\ell}^{-1} \underline{v} = \sigma^2 \left(\underline{v}^{\mathsf{T}} \Sigma_{\ell\ell}^{-1} \underline{v} \right),$$

   the weighted sum of the squares of the residuals, is *minimised* — this is what the methods of least-squares got its name from. In fact, the square root of this quantity is the *norm* of the vector of residuals $\underline{v}$, or its length, in the $Q_{\ell\ell}$ metric, which is thus minimised:[4] $\mathcal{E} = \|\underline{v}\|_Q^2$.

4

[3] The vector $\underline{v}$ of residuals is not the same as the vector of observation errors, or "noise", $\underline{n}$! The residual is the difference between the original observation and the adjusted observation: in other words, a *correction*. However, not even an adjusted observation — or unknown — is the "truth". The truth is not precisely knowable; it is only approximable at best, and the values of the elements of the vector $\underline{n}$, unlike the values of the elements of the vector $\underline{v}$, cannot be computed.

[4] One could eliminate the weight matrix altogether by applying a co-ordinate transformation in the vector space of observations: do a Cholesky decomposition $P_{\ell\ell} = \Gamma\Gamma^{\mathsf{T}}$, resulting in $\mathcal{E} = \underline{v}^{\mathsf{T}} P_{\ell\ell} \underline{v} = \underline{v}^{\mathsf{T}} \Gamma\Gamma^{\mathsf{T}} \underline{v} = \widetilde{\underline{v}}^{\mathsf{T}} \widetilde{\underline{v}}$, with $\widetilde{\underline{v}} \overset{\text{def}}{=} \Gamma^{\mathsf{T}} \underline{v}$. This is automatically achieved by redefining the observables as $\widetilde{\underline{\ell}} \overset{\text{def}}{=} \Gamma^{\mathsf{T}} \underline{\ell}$. This is the straightforward way of reducing the general least-squares problem 15.1 to the simpler unweighted one 14.3. The new observables $\widetilde{\underline{\ell}}$ do not correlate with each other and have identical mean errors.

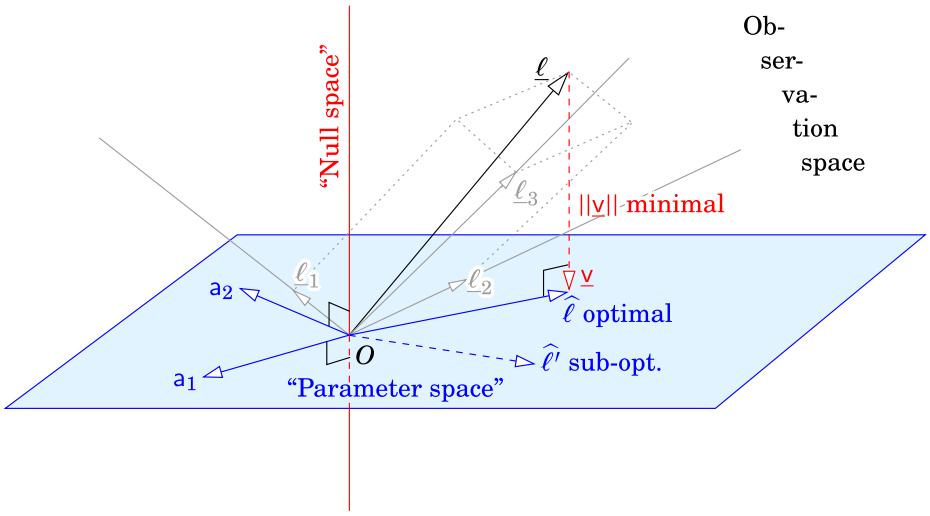


FIGURE 15.1. Least-squares adjustment as an orthogonal projection. For visualisation, it is assumed that the observation space has three dimensions, the parameter space two, and thus the number of degrees of freedom, the dimensionality of the null space, is one.

2. The variance $\Sigma_{\lambda\lambda}$ of an *arbitrary linear combination* $\widehat{\lambda} = \Lambda\widehat{\mathsf{x}}$ of the unknowns $\widehat{\mathsf{x}}$ (and its mean error $\sqrt{\Sigma_{\lambda\lambda}}$) is minimised.
3. The adjusted observables $\widehat{\ell}$ and the residuals $\underline{\mathsf{v}}$ are mutually *orthogonal* in the $Q_{\ell\ell}$ metric: if the scalar product is defined as $\langle \mathsf{a}\cdot\mathsf{b}\rangle_Q \stackrel{\text{def}}{=} \mathsf{a}^\mathsf{T} Q_{\ell\ell}^{-1}\mathsf{b}$:

   $$\langle \widehat{\ell}\cdot\underline{\mathsf{v}}\rangle_Q = \langle A\widehat{\mathsf{x}}\cdot\underline{\mathsf{v}}\rangle_Q = (A\widehat{\mathsf{x}})^\mathsf{T} Q_{\ell\ell}^{-1}\underline{\mathsf{v}} = \widehat{\mathsf{x}}^\mathsf{T} A^\mathsf{T} Q_{\ell\ell}^{-1}\underline{\mathsf{v}} = 0,$$

   because

   $$\mathsf{a}_i^\mathsf{T} Q_{\ell\ell}^{-1}\underline{\mathsf{v}} = 0, \qquad i = 1,\ldots,m,$$

   in other words, the vector of residuals is orthogonal to all columns a_i of the design matrix A.

   Figure 15.1 gives a geometrical interpretation: the unknowns are those coefficients in the linear combination of the columns of the A matrix that minimise the norm of the vector of residuals.
4. The covariance matrix between the vector of unknowns $\widehat{\mathsf{x}}$ and the vector of residuals $\underline{\mathsf{v}}$ vanishes: $\Sigma_{\mathsf{x}\mathsf{v}} = \sigma^2 Q_{\mathsf{x}\mathsf{v}} = 0$. So, they *do not correlate* with each other.

tasoitettu havaintosuure

rakennematriisi

jäännösvirheiden vektori

Because

$$\underline{\ell} = A\widehat{\mathsf{x}} - \underline{\mathsf{v}},$$

it follows, based on the law of propagation of variances and the above-mentioned property 4, that

$$Q_{\ell\ell} = AQ_{xx}A^{\mathsf{T}} + Q_{vv} \quad \Longrightarrow \quad Q_{vv} = Q_{\ell\ell} - AQ_{xx}A^{\mathsf{T}}, \tag{15.2}$$

a useful equation for computing the *weight-coefficient matrix* Q_{vv} — and the variance matrix $\Sigma_{vv} = \sigma^2 Q_{vv}$ — *of the residuals*.

15.3 Testing, hypotheses for testing

The observational material may contain *gross errors*. In a real-life adjustment calculus we must be able to say, based on our knowledge of the statistical distribution of the observations,

- Something about the possible occurrence of gross errors.
- How large gross errors would have to be in order to be noticed and removed.
- The propagation of gross errors of this magnitude into the unknowns of interest.

Finding gross errors belongs to the field of *statistical testing*.

Gross errors that are found can be handled in two ways:

- They are removed from the observation set, and the measurements in question are repeated. After the fact, this is laborious and costly, wherefore at least part of statistical testing is done already in the field.
- They are simply left out. This assumes that the measurement was planned redundantly from the start: so many measurements have been made that one can afford to leave a (small) fraction of them out.

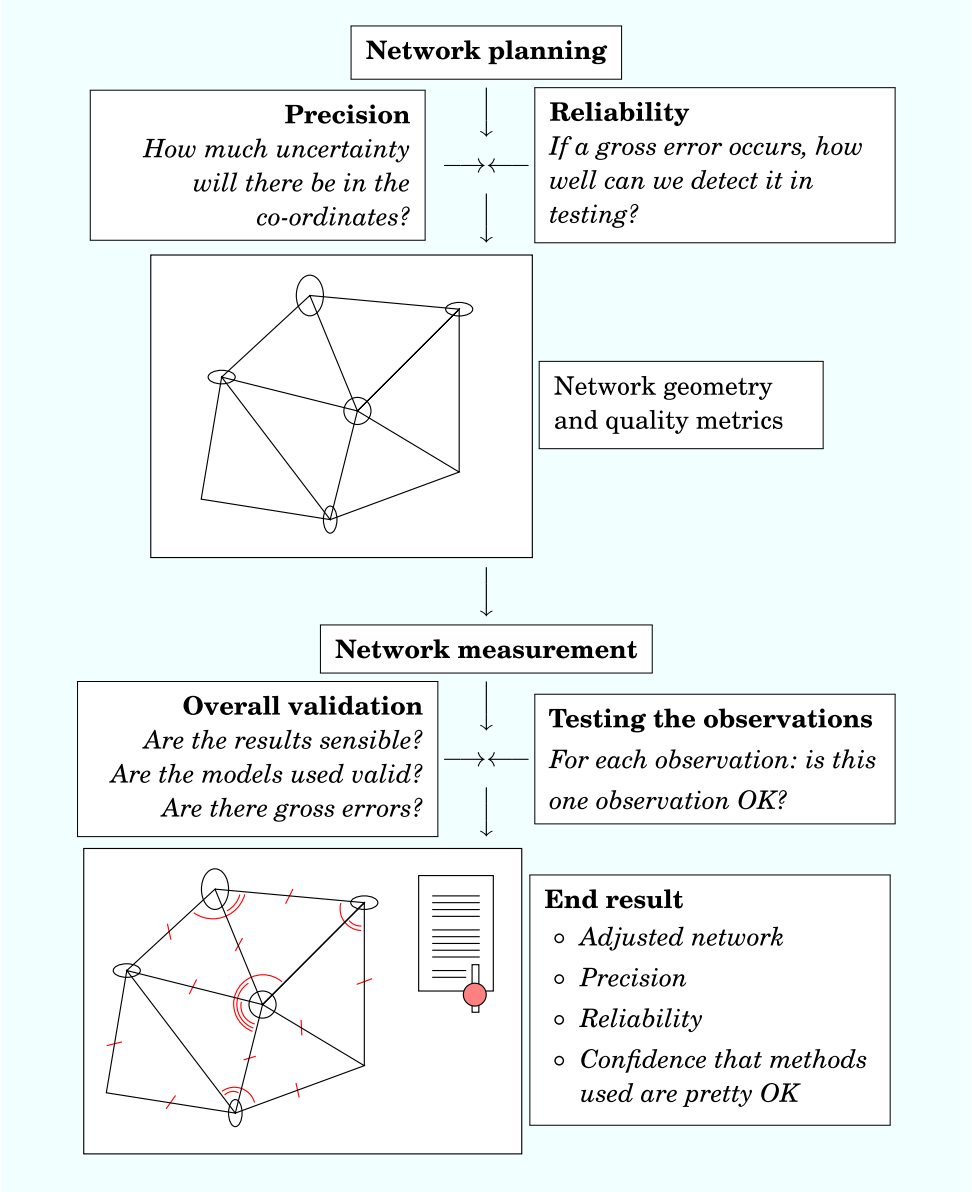
Statistical testing always requires the formulation of *hypotheses*. One of the hypotheses is always the

Null hypothesis All the measurements in the network are correct, there are no gross errors in them. This hypothesis is designated by the symbol H_0.

In addition, there must always be at least one

Alternative hypothesis The network contains some gross error, or some combination of gross errors, or a specific gross error. This hypothesis is designated by the symbol H_a.

TABLEAU 15.1. The planning and measurement process.



Generally we wish to know, or make a judgement on, two matters:

- Are there *generally* any gross errors left in this observation set?
- Is *this specific observation* in error?

Tableau 15.1 shows the role of testing in the whole planning and measurement process.

These questions will be discussed separately in the following sections.

15.4 Overall validation

15.4.1 The χ^2 distribution

Firstly we choose the following alternative hypothesis:

> H_a: somewhere in the measurement material (we do not yet know where) there is a gross error.

This kind of hypothesis can be tested using the χ^2 test. The method, and the tables belonging to it, can be found in statistics textbooks and on the Internet. The quantity to be tested is the length of the vector of residuals in the $\Sigma_{\ell\ell}$ or $Q_{\ell\ell}$ metric, its *norm*, squared:

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \|\underline{\mathrm{v}}\|^2_{\Sigma} \stackrel{\text{def}}{=} \underline{\mathrm{v}}^{\mathsf{T}}\Sigma^{-1}_{\ell\ell}\underline{\mathrm{v}} = \frac{1}{\sigma^2}\left(\underline{\mathrm{v}}^{\mathsf{T}}Q^{-1}_{\ell\ell}\underline{\mathrm{v}}\right) \stackrel{\text{def}}{=} \frac{1}{\sigma^2}\|\underline{\mathrm{v}}\|^2_{Q}\,. \tag{15.3}$$

painoyksikön varianssi

This quantity — note the scaling with the variance of unit weight — is distributed according to the χ^2_{n-m} distribution: the χ^2 distribution with $n-m$ degrees of freedom, figure 15.2. The number of *degrees of freedom* is the difference between the number of observations and the number of unknowns, also known as the *redundancy* $b = n - m$.

odotusarvo

Conceptually, a stochastic quantity with the χ^2_b distribution, for b degrees of freedom, is obtained as the sum of the squares of b independent, standard-normally distributed — that is, having an expectancy zero and a mean error one — stochastic quantities $\underline{n}_i$, $i = 1, \ldots, b$. As the expectancy of one such square equals the variance of the standard normal distribution, being one, it follows that the expectancy

$$E\{\chi^2_b\} = \sum_{i=1}^{b} E\{\underline{n}_i^2\} = \sum_{i=1}^{b} \mathrm{Var}\{\underline{n}_i\} = \sum_{i=1}^{b} 1 = b.$$

Yet another perspective is that the quantity

$$\widehat{\sigma^2} \stackrel{\text{def}}{=} \frac{\underline{\mathcal{E}}}{n-m} = \sigma^2\frac{\underline{\mathrm{v}}^{\mathsf{T}}\Sigma^{-1}_{\ell\ell}\underline{\mathrm{v}}}{n-m}$$

harhaton estimaattori

has the expectancy $E\{\widehat{\sigma^2}\} = \sigma^2$ and is thus an *unbiased estimator* of σ^2. It is called the *a posteriori* variance of unit weight. See also subsection 6.4.3. The ratio $\widehat{\sigma^2}\big/\sigma^2$ is expected to be close to unity if the null hypothesis is valid.

15.4.2 The overall test

By testing the above quantity 15.3, one may infer whether the measure-

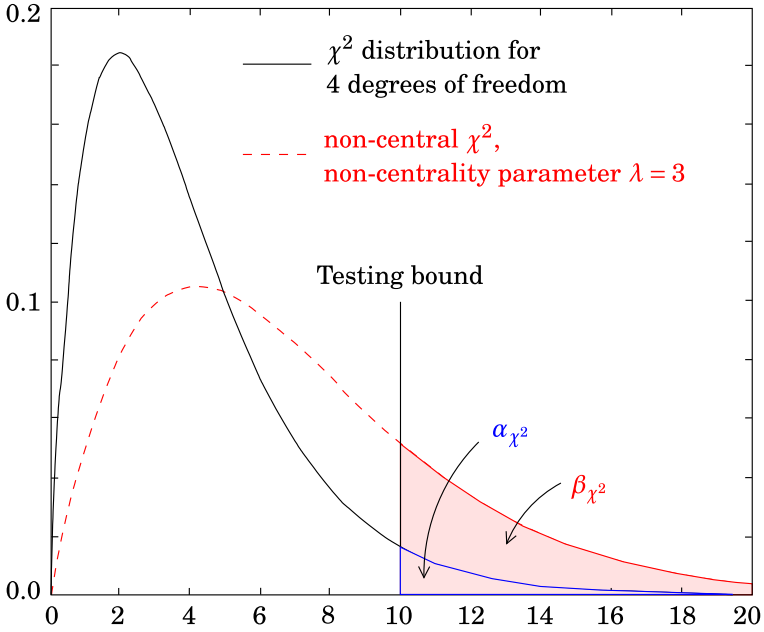


FIGURE 15.2. The χ^2 distribution with four degrees of freedom.

ment material contains *some* gross error or not,[5] without yet stating in which observation it might be found. [5]

MATLAB contains ready routines for applying the χ^2 method.

The quantity[6] $\underline{\mathcal{E}}/\sigma^2$ is distributed according to χ^2_{n-m} only in the case that the material contains no gross errors, that is, the *null hypothesis* H_0 *applies*. Then, the expectancy of the testing variate is [6]

$$E\left\{ \frac{1}{\sigma^2}\underline{\mathcal{E}} \,\middle|\, H_0 \right\} = E\left\{\chi^2_{n-m}\right\} = n - m.$$

Assume now instead however, that the observations contain one or more gross errors, taken together $\nabla\ell$: the *alternative hypothesis* H_a *applies*. The effect of this error vector on the *residuals* is ∇v. In this case, the distribution of the quantity $\underline{\mathcal{E}}/\sigma^2$ is the *non-central* χ^2, in figure 15.2 the red curve.

---

[5] The χ^2 test cannot distinguish between actual gross errors and the possible unsuitability of the functional model $\ell = Ax$ applied to the material. If it happens that the χ^2 test rejects the null hypothesis, but all observations appear to be okay, it might be that there is a problem with the functional model: some systematic effect may have been overlooked.

[6] In Baarda's terminology: "shifting variate". Willem Baarda (1917–2005) was a pioneer of modern adjustment theory and statistical geodesy.

15.4.3 Reliability

The interesting question is now how large the effect of $\nabla\ell$ on $\nabla\mathsf{v}$ is going to be. We may hope that it will be large, because then, the network is *reliable*.[7] Generally the whole of $\nabla\ell$ does not propagate into $\nabla\mathsf{v}$; the adjustment conveys part into the vector of unknowns, as a gross-error effect,[8] of magnitude $\nabla\mathsf{x}$. See also section 15.9.

In this situation, the vector of residuals is $\underline{\mathsf{v}} + \nabla\mathsf{v}$. Written out into terms, the testing variate becomes

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \|\underline{\mathsf{v}} + \nabla\mathsf{v}\|^2_\Sigma = \underline{\mathsf{v}}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\underline{\mathsf{v}} + \underline{\mathsf{v}}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\nabla\mathsf{v} + \nabla\mathsf{v}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\underline{\mathsf{v}} + \nabla\mathsf{v}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\nabla\mathsf{v}.$$

The expectancy of the testing variate becomes

$$\begin{aligned} E\left\{\frac{1}{\sigma^2}\underline{\mathcal{E}}\,\middle|\,H_\mathrm{a}\right\} = \\ = E\left\{\underline{\mathsf{v}}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\underline{\mathsf{v}}\right\} + E\left\{\underline{\mathsf{v}}^\mathsf{T}\right\}\Sigma^{-1}_{\ell\ell}\nabla\mathsf{v} + \nabla\mathsf{v}^\mathsf{T}\Sigma^{-1}_{\ell\ell}E\left\{\underline{\mathsf{v}}\right\} + \nabla\mathsf{v}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\nabla\mathsf{v} = \\ = E\left\{\chi^2_{n-m}\right\} + 0 + 0 + \nabla\mathsf{v}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\nabla\mathsf{v} = (n-m) + \lambda, \end{aligned}$$

in which $\lambda \overset{\text{def}}{=} \nabla\mathsf{v}^\mathsf{T}\Sigma^{-1}_{\ell\ell}\nabla\mathsf{v}$ is called the *non-centrality parameter* of the χ^2 distribution. It describes how far the effect of the assumed gross error on the residuals, $\nabla\mathsf{v}$, extends outside the uncertainty area of the observations as described by the matrix $\Sigma_{\ell\ell}$.

A quadratic quantity is always positive. Therefore, the χ^2 test is *one-sided*, unlike the later presented test for the normal distribution. So, because $\underline{\mathcal{E}}$ is a quadratic quantity, every gross error — and even systematic errors, that is, errors in the functional model used — will tend to increase it. *Each and every error tends to make χ^2 larger*, and makes noticing the error more likely. This makes the χ^2 test such a useful overall test.

In fact, the χ^2 test validates a lot more than just the observations. It assures that

- The observation set does not contain any (large) gross errors.
- The functional model used (the observation equations) is valid with sufficient accuracy.
- The assumed mean errors of the observations (and the possible assumption of non-correlatedness) are realistic.

[7] This is called *interior reliability*.

[8] This is called *exterior reliability*. A small effect means a good exterior reliability.

15.5 Locating gross errors

If we have inferred that the material presumably contains one or more gross errors, we next want to find out which observations are under suspicion. Let us assume for simplicity that a gross error occurs only in one observation, although there might well be errors in several observations simultaneously.

The simplest way to search for gross errors, or rather, to undertake *outlier detection*, is to *look at the residuals*. poikkeava havaintoarvo

Let the vector of residuals be

$$\underline{\mathsf{v}} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_i & \cdots & \underline{v}_{n-1} & \underline{v}_n \end{bmatrix}^{\mathsf{T}}.$$

The element $\underline{v}_i$ is the residual of observation number i, that is the observation $\underline{\ell}_i$. The element's variance is

$$\sigma_{v_i}^2 = \sigma^2 \left[Q_{\mathsf{vv}} \right]_{ii},$$

and its mean error σ_{v_i} is the square root of this.

Assume that the residuals $\underline{v}_i$ are *normally distributed*. Then, we may *test* every observation $i = 1, \ldots, n$:

$$\begin{aligned} \left| \underline{v}_i \right| > 1.96\,\sigma_{v_i} \quad &\Longrightarrow \quad \underline{\ell}_i \text{ is probably in error} \\ \left| \underline{v}_i \right| \leq 1.96\,\sigma_{v_i} \quad &\Longrightarrow \quad \underline{\ell}_i \text{ is presumably correct.} \end{aligned}$$

This two-sided test based on the normal distribution uses a *significance level* of 95 %: *Even if there is no gross error*, there nevertheless is a probability of 100 % − 95 % = 5 % that, based on the test, observation $\underline{\ell}_i$ will be rejected. merkitsevyystaso

Table 15.2 gives the rejection bounds for different significance levels in the two-sided test based on the standard normal distribution.

The method described here works correctly only if the observations do not correlate with each other, so that the matrix $Q_{\ell\ell}$ is a diagonal matrix. If it is not, the literature offers an adapted[9] testing method called *data snooping* (Baarda, 1968). 9

[9] The trick is simply that, instead of the residuals $\underline{\mathsf{v}}$, *weighted residuals* $\underline{\mathsf{w}} \stackrel{\text{def}}{=} Q_{\ell\ell}^{-1}\,\underline{\mathsf{v}}$ and their variances are used.
The logic is, that if we search for a gross error in observation number i, we look for the linear combination of residuals in which the error shows clearest. We calculate the *orthogonal projection* of $\underline{\mathsf{v}}$ (in the $Q_{\ell\ell}$ metric) on the direction of the assumed gross error

TABLE 15.2. Rejection bounds h for significance levels α_2 in a two-sided test based on the standard normal distribution, having mean error $\sigma = 1$ and expectancy $\mu = 0$. See figure 15.4.

$1-\alpha_2$, %	α_2, %	h
5	95	1.96
2.5	97.5	2.24
1	99	2.57
0.1	99.9	3.29

15.6 Calculation example: linear regression

Let us return to the linear regression example already used in subsection 14.5.3, see table 15.3.

Recall that the least-squares solution found was

$$\widehat{a} = 1.76 \quad \pm 1.25\,\sigma,$$
$$\widehat{b} = 0.495 \pm 0.349\,\sigma.$$

We compute the function values $\widehat{a} + \widehat{b}x_i$ of the fitted line, as well as its residuals $\underline{v}_i = \left(\widehat{a} + \widehat{b}x_i\right) - \underline{y}_i$. The condition $\sum_{i=1}^{n} \underline{v}_i = 0$ is a good sanity check.

TABLE 15.3. Example of linear regression.

	$i \rightarrow$	1	2	3	4	5	$\sum_{i=1}^{5}$
x_i		1.51	2.44	3.34	4.41	5.05	16.75
$\underline{y}_i$		2.32	3.12	3.57	3.93	4.15	17.09
$\widehat{a} + \widehat{b}x_i$		2.51	2.97	3.41	3.94	4.26	
$\underline{v}_i$		+0.19	−0.15	−0.16	+0.01	+0.11	0.00
$\underline{v}_i^2$		0.0361	0.0225	0.0256	0.0001	0.0121	0.0964

$\mathsf{e}_i \overset{\text{def}}{=} \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \end{bmatrix}^{\mathsf{T}}$ (where the element "1" is in place i):

$$\underline{w}_i \overset{\text{def}}{=} \langle \mathsf{e}_i \cdot \underline{\mathsf{v}} \rangle_Q = \mathsf{e}_i^{\mathsf{T}} Q_{\ell\ell}^{-1} \underline{\mathsf{v}}.$$

Together, the components $\underline{w}_i$ form the vector $\underline{\mathsf{w}}$ and they are optimally suited for discerning gross errors, or "outliers".

In the test, also the mean error σ_{w_i} of every individual $\underline{w}_i$ is needed, to be computed from the diagonal elements of the matrix (equation 15.2)

$$Q_{\mathsf{ww}} \overset{\text{def}}{=} Q_{\ell\ell}^{-1} Q_{\mathsf{vv}} Q_{\ell\ell}^{-1} = Q_{\ell\ell}^{-1} \left(Q_{\ell\ell} - A Q_{\mathsf{xx}} A^{\mathsf{T}} \right) Q_{\ell\ell}^{-1},$$

just as when testing using $\underline{\mathsf{v}}$.

TABLE 15.4. Values of the cumulative χ^2_3 distribution. α_{χ^2} is the significance level of the χ^2 test.

x	$\int_0^x \chi^2_3(x')\,dx'$ (α_{χ^2})	$\int_x^\infty \chi^2_3(x')\,dx'$ $(1-\alpha_{\chi^2})$
4.642	0.80	0.20
6.251	0.90	0.10
7.815	0.95	0.05
9.837	0.98	0.02
11.345	0.99	0.01
12.838	0.995	0.005
14.796	0.998	0.002
16.266	0.999	0.001

If the observations $\underline{y}_i$ have a variance matrix $\Sigma_{\ell\ell} = \sigma^2 I$, then the "shifting variate" to be tested is siirtosuure

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \underline{\mathsf{v}}^{\mathsf{T}}\Sigma_{\ell\ell}^{-1}\underline{\mathsf{v}},$$

in which $\underline{\mathsf{v}}$ is the vector formed by the residuals $\underline{v}_i$. We obtain

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\underline{v}_i^2 \Longrightarrow \underline{\mathcal{E}} = \sum_{i=1}^{n}\underline{v}_i^2.$$

If it is given *a priori* that $\sigma = \pm 0.15$, it follows that

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \frac{0.0964}{0.0225} = 4.28.$$

The quantity $\underline{\mathcal{E}}/\sigma^2$ is distributed according to χ^2_3: there are $n = 5$ observations and $m = 2$ unknowns (a and b), so the *number of degrees of freedom* (redundancy) is $n - m = 3$. According to table 15.4, the probability that under the null hypothesis $\chi^2_3 > 4.642$ is 20 %, so the value 4.28 is fully acceptable, at least on a significance level of 80%.

Next, the individual residuals are tested. Compute first the weight-coefficient matrix of the vector of residuals using equation 15.2:

$$Q_{\mathsf{vv}} = Q_{\ell\ell} - AQ_{\mathsf{xx}}A^{\mathsf{T}},$$

in which $Q_{\ell\ell} = I$,

$$Q_{\mathsf{xx}} = \begin{bmatrix} 1.5640 & -0.4072 \\ -0.4072 & 0.1215 \end{bmatrix}$$

TABLE 15.5. Example of linear regression. Computing the residuals, their variance-covariance matrix, and normalised residuals.

	$i \rightarrow$	1	2	3	4	5
v_i		+0.19	−0.15	−0.16	+0.01	+0.11
Q_{vv}		+0.3887	−0.4032	−0.2019	+0.0375	+0.1807
		−0.4032	+0.6998	−0.2006	−0.0821	−0.0112
		−0.2019	−0.2006	+0.8007	−0.1978	−0.1969
		+0.0375	−0.0821	−0.1978	+0.6646	−0.4178
		+0.1807	−0.0112	−0.1969	−0.4178	+0.4502
σ_{v_i}		0.0935	0.1255	0.1342	0.1223	0.1006
$\lvert \underline{v}_i \rvert / \sigma_{v_i}$		2.03	1.20	1.19	0.08	1.09

was already computed in subsection 14.5.3, and

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} 1 & 1.51 \\ 1 & 2.44 \\ 1 & 3.34 \\ 1 & 4.41 \\ 1 & 5.05 \end{bmatrix}.$$

After careful calculation (MATLAB!), we obtain the matrix Q_{vv}, shown in table 15.5. Of this matrix, it is mostly the *diagonal elements* that are interesting:

$$\sigma_{v_1} = \sigma\sqrt{[Q_{vv}]_{11}} = 0.15 \cdot \sqrt{0.3887} = 0.15 \cdot 0.623 = 0.0935,$$
$$\sigma_{v_2} = \sigma\sqrt{[Q_{vv}]_{22}} = 0.1255,$$

and so on. (Remember that $\sigma = 0.15$.) See table 15.5.[10]

As can be seen, all observations are acceptable, with the exception of $\underline{y}_1$, which, on the 95 % significance level, is barely rejected (rejection bound 1.96). However, already on a significance level of 97.5 %, it too is accepted.

15.7 Adding a gross error

karkea virhe

Next, we add to the observed value $\underline{y}_3$ a simulated gross error +1.0.

[10] Note how the mean errors of the residuals are *systematically smaller* than the mean errors of the observations $\sigma = \pm 0.15$, especially close to the edges! With a large number of points, this phenomenon vanishes and we may write $Q_{vv} \approx Q_{\ell\ell}$.
This is often done in any case. Then, gross errors in the edge points will not be noticed sufficiently well.

TABLE 15.6. Example of linear regression. A simulated gross error in point 3: original data, linear regression, residuals, testing.

	$i \to$	1	2	3	4	5	$\sum_{i=1}^{5}$
x_i		1.51	2.44	3.34	4.41	5.05	16.75
$\underline{y}_i$		2.32	3.12	4.57	3.93	4.15	18.09
x_i^2		2.28	5.95	11.16	19.45	25.50	64.34
$x_i \underline{y}_i$		3.50	7.61	15.26	17.33	20.96	64.66
$\widehat{a} + \widehat{b} x_i$		2.71	3.17	3.62	4.14	4.46	
$\underline{v}_i$		+0.39	+0.05	−0.95	+0.21	+0.31	0.01
$\underline{v}_i^2$		0.1521	0.0025	0.9025	0.0441	0.0961	1.1973
σ_{v_i}		0.0935	0.1255	0.1342	0.1223	0.1006	
$\lvert\underline{v}_i\rvert / \sigma_{v_i}$		4.17	0.40	7.08	1.72	3.08	
Rejection?		*		**		*	

Now, as the least-squares solution we obtain the result of table 15.6:

$$\widehat{b} = \frac{5 \cdot 64.66 - 16.75 \cdot 18.09}{5 \cdot 64.34 - 16.75^2} = \frac{20.2925}{41.1375} = 0.493,$$
$$\widehat{a} = \tfrac{1}{5}\left(18.09 - 16.75 \cdot \widehat{b}\right) = 1.97.$$

In table 15.6 the σ_{v_i} values have not changed.

Compute the "shifting variate"

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \frac{1}{\sigma^2}\sum_{i=1}^{n} \underline{v}_i^2 = \frac{1.1973}{0.0225} = 53.21!$$

There is something very wrong here...

Now look at table 15.6. The largest testing value by far, 7.08, is seen for the erroneous observation 3. But observations 1 and 5 are also rejected at even a 99% significance level! For this reason, one should proceed carefully. *Based on the test, one should reject only one observation at a time*, after which the whole least-squares computation should be repeated.

pienimmän neliösumman laskenta

15.8 Significance level of the test

15.8.1 Choice of rejection bound

When we test a certain alternative hypothesis against the null hypothesis using an assumedly normally distributed testing quantity or variate, one must choose a suitable *rejection bound*. If the variate to be tested exceeds

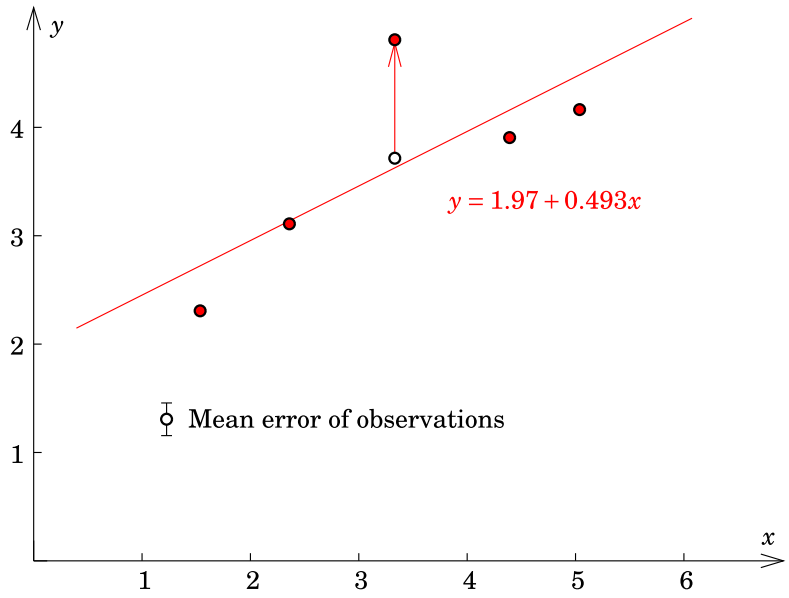


FIGURE 15.3. Example of linear regression, observation 3 contains a simulated gross error.

this bound, H_0 is rejected and H_a accepted. Choosing the rejection bound is an important strategic decision.

See figure 15.4. In the figure, the rejection bound chosen is $h = 2.5\sigma$: if the testing variate exceeds 2.5 times its own mean error σ, the null hypothesis H_0 is rejected and the alternative hypothesis H_a is accepted.

Now, the strategy may lead to two types of error:

ensimmäisen lajin virhe 11

- The null is rejected although it is valid. This is called an *error of the first kind*.[11] The probability of this error happening is the size of the vertically hatched (blue) area. In the case of normal distribution, it amounts to $1 - \alpha = 1.24\,\%$ (two-sided), if the rejection bound is

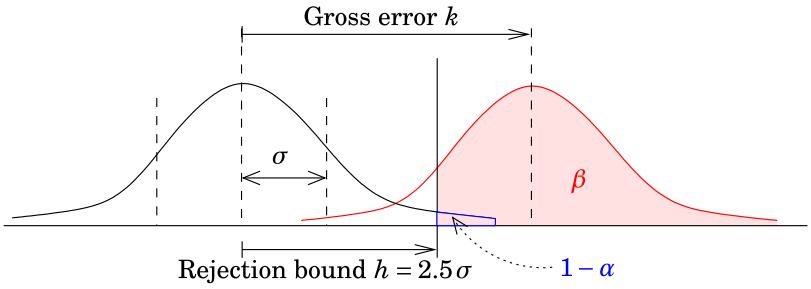


FIGURE 15.4. Statistical testing based on normal distribution.

[11] Also *rejection error*.

TABLE 15.7. Rejection bound h and probability of rejection in a test in the case of normal distribution. α_1 is the significance level of the one-sided, α_2 of the two-sided test.

h/σ	$1-\alpha_1$, %	$1-\alpha_2$, %
2.0	2.28	4.56
2.5	0.62	1.24
3.0	0.13	0.27
3.5	0.02	0.05

$h = 2.5\sigma$, see table 15.7. The quantity α is called the *significance level* of the test.

- The null is accepted although there is a gross error: H_0 is false, the alternative hypothesis H_a is true. This is an *error of the second kind*.[12] The probability of it happening depends on the size k of the gross error, more precisely, on the size of the normalised difference $(k-h)/\sigma$.

  toisen lajin virhe 12

  Its complement, the probability of rejection, is called the *power* β[13] of the test. In figure 15.4 it is the size of the pink area.

  erotuskyky 13

Choosing the *testing strategy*, choosing h, is thus always a *compromise*. It depends on the relative costs of errors of the first and second kind — including non-monetary, such as reputational, "costs". $h = 3\sigma$ is often used — the "three-sigma rule".

TABLE 15.8. Assumed size k of gross error and corresponding power β of the test. Normal distribution and rejection bound $h = 2.5\sigma$ assumed.

k/σ	$(k-h)/\sigma$	β, %
3.0	0.5	69.1
3.5	1.0	84.1
4.0	1.5	93.3
4.5	2.0	97.7
5.0	2.5	99.4
5.5	3.0	99.9
6.0	3.5	99.98

[12] Also *acceptance error*.

[13] So, the probability of an error of the second kind, if there is indeed a gross error in the observation, is $1-\beta$, or $100\,\% - \beta$.

15.8.2 Harmonisation of overall and per-observation tests

In testing the body of observations, there is a *link* between the overall validation test and the per-observation tests! The link is through the significance levels. See figure 15.5: if the significance level of the χ^2 test is α_{χ^2} and that of the test for a single observation is α, the connection is

$$\alpha_{\chi^2} = \alpha^{n-m},$$

[14] with $n-m$ the number of degrees of freedom.[14] In other words, the joint probability that all observations individually pass their tests must be the same as the probability of passing the overall validation. Only on that condition may it be expected that, if the common χ^2 test finds something "rotten", the tests for the individual observations will also point to the "guilty" observation.

Example If $\alpha_{\chi^2} = 95\,\%$ with ten degrees of freedom, it follows that

$$\alpha = \sqrt[n-m]{\alpha_{\chi^2}} = \sqrt[10]{0.95} \approx 0.99489 = 99.489\,\%,$$

some ten times closer to 100 %.

[15] After removing or correcting the "guilty" observation, the testing procedure is repeated, until the χ^2 test is passed.[15]

15.9 Reliability

15.9.1 Principle

The *reliability* of a measurement network is the property that gross errors are found easily, and are found even if they are relatively small.

Reliability corresponds to the network being "strong". It is however not the same kind of strength as when the network is precise.

See figure 15.6. From points A, B, and C are measured the directions to a fourth point. Error ellipses for three different cases are drawn:

I when the point is far from the points A and B

II when the point is in a location where the directions to points A and B are perpendicular to each other, and

[14] This procedure is similar to the well-known *Bonferroni correction*, Wikipedia, Bonferroni correction.

[15] If the test still fails, perhaps the other models used should be checked, such as the assumed precisions of the observations and so on.

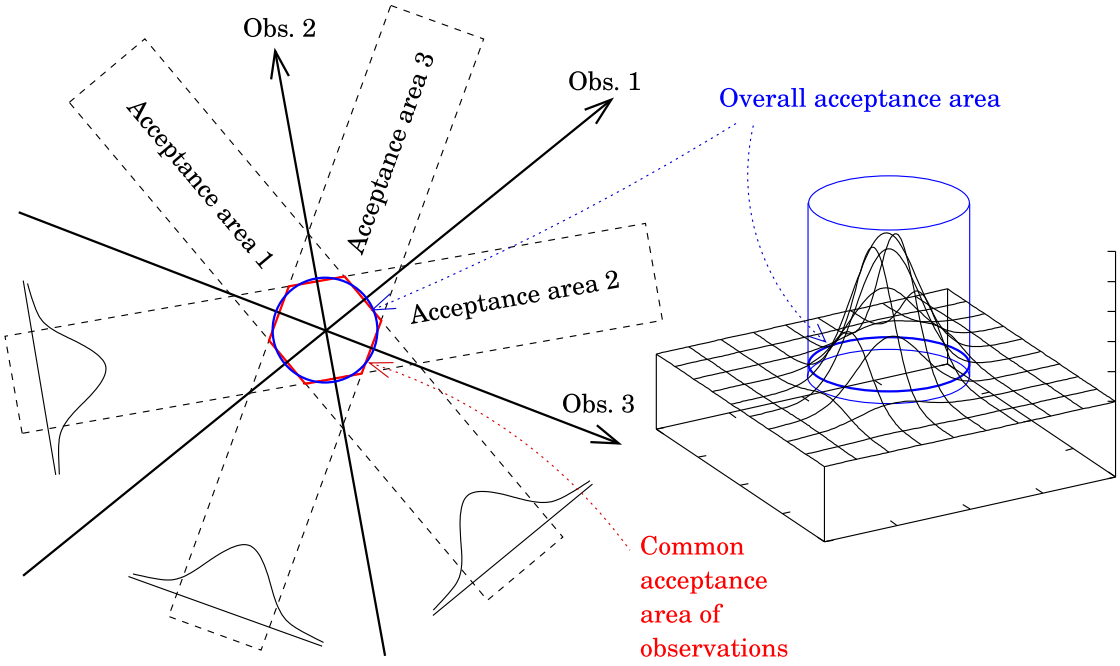


FIGURE 15.5. Harmonisation of the significance levels of the overall validation and per-observation tests.

III when the point is between points A and B.

As can be seen, the most precise result is obtained in case III. The error ellipse is the smallest.

However, *reliability is poor* (non-existent) in case III. If the measurement made from point C contains a gross error (dashed line), we still

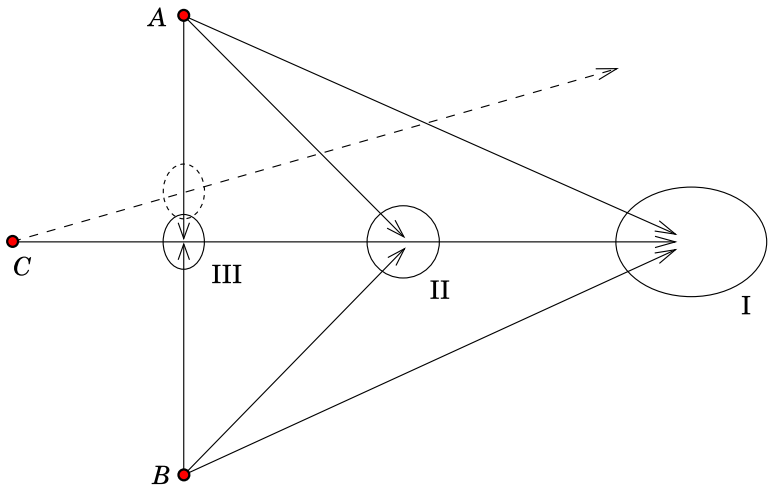


FIGURE 15.6. An example of reliability.

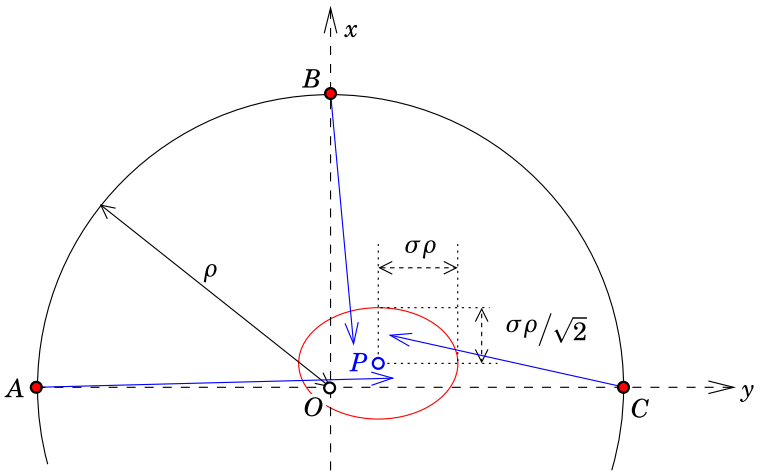


FIGURE 15.7. Another example of reliability.

obtain in case III a seemingly good — precise — but *erroneous* result. See the dashed error ellipse.

In cases I and II, a gross error in the observation from C will produce a contradictory result. It is not possible to find a location for the target point that is compatible with the direction measurements from all three points. This is a *good thing*, because it enables the detection of the gross error. The network is then called *reliable*.

In network planning, attention must be paid to reliability, of course in addition to precision. The network must be designed with appropriate redundancy: it must contain measurements that check each other. Common sense helps a great deal here. There are mathematical and software tools for evaluating the reliability of a network. One must always ask, "what if this or that observation were in error... would I notice?"

15.9.2 Another example

In this example, figure 15.7, the observation points A, B, and C are located on the edge of a circle, and point P, the direction to which is being measured, is located near the origin (centre point of the circle) O.

The observation equations are obtained by looking at the geometry:

$$\overbrace{\begin{bmatrix} \underline{\theta}_{AP} - \theta_{AO} \\ \underline{\theta}_{BP} - \theta_{BO} \\ \underline{\theta}_{CP} - \theta_{CO} \end{bmatrix}}^{\underline{\ell}} + \overbrace{\begin{bmatrix} \underline{v}_1 \\ \underline{v}_2 \\ \underline{v}_3 \end{bmatrix}}^{\underline{\mathrm{v}}} = \overbrace{\frac{1}{\rho}\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} \widehat{x}_P \\ \widehat{y}_P \end{bmatrix}}^{\widehat{\mathrm{x}}}.$$

Symbolically

$$\underline{\ell} + \underline{v} = A\hat{x},$$

in which

$$\underline{\ell} = \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ \underline{\ell}_3 \end{bmatrix} = \begin{bmatrix} \underline{\theta}_{AP} - \theta_{AO} \\ \underline{\theta}_{BP} - \theta_{BO} \\ \underline{\theta}_{CP} - \theta_{CO} \end{bmatrix}, \qquad A = \frac{1}{\rho} \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The least-squares solution is

$$\hat{x} = \left(A^{\mathsf{T}}A\right)^{-1} A^{\mathsf{T}}\underline{\ell} = \rho \cdot \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \underline{\ell}_3 - \underline{\ell}_1 \\ -\underline{\ell}_2 \end{bmatrix} = \rho \cdot \begin{bmatrix} \frac{1}{2}(\underline{\ell}_3 - \underline{\ell}_1) \\ -\underline{\ell}_2 \end{bmatrix}.$$

From this are obtained the residuals

$$\underline{v} = A\hat{x} - \underline{\ell} = \begin{bmatrix} -\frac{1}{2}(\underline{\ell}_3 - \underline{\ell}_1) \\ \underline{\ell}_2 \\ \frac{1}{2}(\underline{\ell}_3 - \underline{\ell}_1) \end{bmatrix} - \begin{bmatrix} \underline{\ell}_1 \\ \underline{\ell}_2 \\ \underline{\ell}_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\underline{\ell}_1 - \frac{1}{2}\underline{\ell}_3 \\ 0 \\ -\frac{1}{2}\underline{\ell}_1 - \frac{1}{2}\underline{\ell}_3 \end{bmatrix}. \qquad (15.4)$$

Note 1 As can be seen, the observation $\underline{\ell}_2$ has vanished from the residuals! If $\underline{\ell}_2 = \underline{\theta}_{BP} - \theta_{BO}$ contains a gross error, we are never going to notice it as an overly large residual.

Note 2 From the residuals it cannot be seen whether a gross error comes from observation $\underline{\ell}_1$ or observation $\underline{\ell}_3$. In the residuals, their coefficients are identical.

We may also write equation 15.4 as

$$\underline{v} = A\hat{x} - \underline{\ell} = A\left(A^{\mathsf{T}}A\right)^{-1}A^{\mathsf{T}}\underline{\ell} - \underline{\ell} = -R\underline{\ell},$$

with the *redundancy matrix*

$$R \stackrel{\text{def}}{=} I - A\left(A^{\mathsf{T}}A\right)^{-1}A^{\mathsf{T}} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

Each diagonal element of the redundancy matrix is a rough-and-ready measure for how well the geometry controls for a gross error in the corresponding observation. $R_{11} = R_{33} = \frac{1}{2}$ tells us that observations $\underline{\ell}_1$ and $\underline{\ell}_3$ are somewhat controlled, but $R_{22} = 0$ tells us that $\underline{\ell}_2$ is not controlled at all. A sensible requirement is that all $R_{ii} \gtrsim 0.5$.

Next, compute the shifting variate

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \underline{v}^{\mathsf{T}}\Sigma_{\ell\ell}^{-1}\underline{v}.$$

Here, $\Sigma_{\ell\ell}$ is the variance matrix of the observations. Assume that the observations do not correlate with each other and that their mean error is σ. Then

$$\Sigma_{\ell\ell} = \sigma^2 \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}.$$

We obtain (H_0 is the null hypothesis):

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} \,\Big|\, H_0 = \frac{1}{\sigma^2}\sum_{i=1}^{3} \underline{v}_i^2 = \frac{1}{2\sigma^2}\left(\underline{\ell}_1 + \underline{\ell}_3\right)^2.$$

Because the mean errors of both $\underline{\ell}_1$ and $\underline{\ell}_3$ are σ and they do not correlate, the mean error of the sum $\underline{\ell}_1 + \underline{\ell}_3$ is $\sigma\sqrt{2}$ and its variance $2\sigma^2$. The number of degrees of freedom is 1 and the variate $\underline{\mathcal{E}}/\sigma^2$ is distributed according to χ_1^2, as should be the case according to theory.

By comparing the value $\underline{\mathcal{E}}/\sigma^2$ computed from the observations with the values from the χ_1^2 table, one can test whether the observations might contain a gross error. If all observations are free of gross errors, the expectancy of $\underline{\mathcal{E}}/\sigma^2$ is 1.

Nevertheless, as already pointed out above, we cannot observe any gross
16 errors in $\underline{\ell}_2$ at all. We say[16] that the measurement geometry is *reliable* for observations $\underline{\ell}_1$ and $\underline{\ell}_3$, but *unreliable* for observation $\underline{\ell}_2$. If observation $\underline{\ell}_2$ contained a gross error of size ∇, it would slip *in its entirety* into the
17 co-ordinate $\widehat{y}_P$ as an error $-\rho \cdot \nabla$! We also say[17] that the measurement geometry is *unreliable* for unknown $\widehat{y}_P$, but *reliable* for unknown $\widehat{x}_P$.

A sufficiently large gross error ∇ in observations $\underline{\ell}_1$ or $\underline{\ell}_3$ would again be detected as an overly large value for the shifting variate (alternative hypothesis H_a):

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} \,\Big|\, H_\mathrm{a} = \frac{1}{2\sigma^2}\left(\underline{\ell}_1 + \underline{\ell}_3 + \nabla\right)^2,$$

the expectancy of which is $1 + \frac{1}{2}\left(\nabla/\sigma\right)^2$, for a non-centrality parameter $\lambda = \frac{1}{2}\left(\nabla/\sigma\right)^2$, see figure 15.2. If $\nabla \gg \sigma$, this would be detected with considerable confidence.

Note Reliability has nothing to do with *precision*! The precision of the

[16] This is called interior reliability.

[17] Exterior reliability.

unknowns $\widehat{\mathrm{x}} = \begin{bmatrix} \widehat{x}_P & \widehat{y}_P \end{bmatrix}^{\mathsf{T}}$ is described by their variance matrix

$$\mathrm{Var}\{\widehat{\mathrm{x}}\} = \sigma^2 \left(A^{\mathsf{T}} Q_{\ell\ell}^{-1} A\right)^{-1} = \sigma^2 \rho^2 \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix},$$

so, the mean error of $\widehat{x}_P$ is $\frac{1}{2}\sigma\rho\sqrt{2}$, and that of $\widehat{y}_P$ is $\sigma\rho$, and they are uncorrelated with each other. See the error ellipse in figure 15.7.

However, a good mean error gives no solace if the co-ordinate solution $\widehat{y}_P$ contains a gross error...

15.9.3 The meaning of redundancy

Even though the reliability of a measurement network is good, we may still ask whether it is easy to identify the observation in which the gross error has occurred. It this is not easy, we end up measuring all suspect observations again, or throwing them out. This is not good.

From the viewpoints of both good reliability and identifiability of gross errors, the *degree of redundancy* of a geodetic measurement network should not be too low. If the number of observations is n and the number of unknowns m, then the number of conditions, or *degrees of freedom*, is $n - m$. The degree of redundancy is then $(n-m)/n$. This is often stated as a percentage. For example, linear regression of a straight line through five points: $n = 5$, $m = 2$, so a degree of redundancy of $\frac{3}{5} = 60\,\%$. On the other hand, a levelling line of ten points between two known points: $n = 11$, $m = 10$, the degree of redundancy being $\frac{1}{11} = 9\,\%$ — weak, but unfortunately common. By measuring in both directions we obtain $n = 22$, $m = 10$, so a degree of redundancy of $\frac{12}{22} \approx 55\,\%$, which is already good. A good rule of thumb is that a degree of redundancy of 50 % is desirable.

15.10 Deformation analysis

Deformation analysis is one practical application of statistical testing. The null hypothesis H_0 in these tests is, that no observable deformation has happened. There may be many different alternative hypotheses H_a, from the hypothesis that some unspecified deformation took place, to many concrete hypotheses about the precise nature of the deformation sought.

Deformation analysis is also an application that involves the time dimension: measurements collected in two or more measurement epochs

are compared. The deformations studied may be natural, like deformations in the Earth's crust brought about by tectonic movements or by varying glacial loading; or they may be brought about by human activity, like the subsidence caused by mineral extraction (petroleum, natural gas, irrigation water, ...). The object of study may be the Earth's crust in an area, or a building or other built structure like a reservoir dam. The possibilities are very broad.

maan vajoaminen

vesiallas

Deformation analysis is discussed in the textbooks by Cooper (1987) pages 331–352 and Vaníček and Krakiwsky (1986) pages 611–659.

15.10.1 Height deformation analysis

In one dimension, *height deformation analysis* studies vertical movement, for example using levelling. In the simplest case, the same levelling line or network of n points has been measured twice:

$$\underline{H}_i(t_1), \quad \underline{H}_i(t_2), \qquad i = 1,\dots,n,$$

and the variance matrices of the heights, $\Sigma(t_1)$ and $\Sigma(t_2)$, are also available.

Clearly, comparison is possible only if both measurements are first reduced to the same reference or *datum point*. We choose the first network point, point 1, as the datum point:

$$H_1^{(1)}(t_1) = H_1^{(1)}(t_2) \qquad (= \text{some agreed value}).$$

After this, the variance matrices for both measurement times or *epochs* are only of size $(n-1)\times(n-1)$, because now point 1 is *known* and no longer has (co-)variances.

$$\Sigma^{(1)}(t_1) = \begin{bmatrix} \sigma_{22}^{(1)}(t_1) & \sigma_{23}^{(1)}(t_1) & \cdots & \sigma_{2n}^{(1)}(t_1) \\ \sigma_{32}^{(1)}(t_1) & \sigma_{33}^{(1)}(t_1) & \cdots & \sigma_{3n}^{(1)}(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n2}^{(1)}(t_1) & \sigma_{n3}^{(1)}(t_1) & \cdots & \sigma_{nn}^{(1)}(t_1) \end{bmatrix},$$

and the same for $\Sigma^{(1)}(t_2)$. Here

$$\left.\begin{aligned} \sigma_{ii}^{(1)}(t_\ell) &= \mathrm{Var}\{H_i^{(1)}(t_\ell)\}, \\ \sigma_{ij}^{(1)}(t_\ell) &= \mathrm{Cov}\{H_i^{(1)}(t_\ell), H_j^{(1)}(t_\ell)\}, \end{aligned}\right\} \qquad \ell = 1,2; \quad i,j = 2,\dots,n.$$

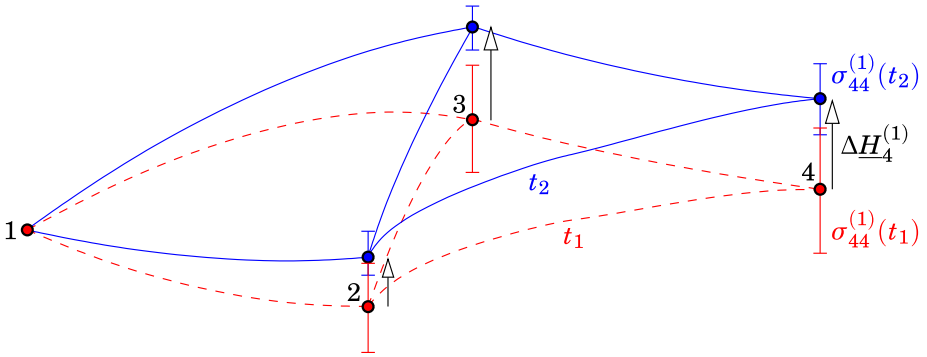


FIGURE 15.8. Height deformation monitoring network for the epochs t_1 (red) and t_2 (blue). Realistic error bars.

Now, calculate the height displacements between the two measurement epochs and their joint variance matrix, assuming that the measurements made at times t_1 and t_2 are statistically independent of each other:

$$\Delta\underline{H}^{(1)}_i \stackrel{\text{def}}{=} \underline{H}^{(1)}_i(t_2) - \underline{H}^{(1)}_i(t_1), \qquad i = 2,\ldots,n,$$
$$\Sigma^{(1)}_{\Delta\mathsf{H}\Delta\mathsf{H}} = \Sigma^{(1)}(t_1) + \Sigma^{(1)}(t_2).$$

After this it is intuitively clear — provided that both sets of height measurements are multi-normally distributed — that the following quantity, the *shifting variate*, has the χ^2_{n-1} distribution:

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \left(\Delta\underline{\mathsf{H}}^{(1)}\right)^{\mathsf{T}} \left(\Sigma^{(1)}_{\Delta\mathsf{H}\Delta\mathsf{H}}\right)^{-1} \Delta\underline{\mathsf{H}}^{(1)},$$

in which

$$\Delta\underline{\mathsf{H}}^{(1)} \stackrel{\text{def}}{=} \begin{bmatrix} \underline{H}^{(1)}_2(t_2) - \underline{H}^{(1)}_2(t_1) \\ \underline{H}^{(1)}_3(t_2) - \underline{H}^{(1)}_3(t_1) \\ \vdots \\ \underline{H}^{(1)}_n(t_2) - \underline{H}^{(1)}_n(t_1) \end{bmatrix} = \begin{bmatrix} \Delta\underline{H}^{(1)}_2 \\ \Delta\underline{H}^{(1)}_3 \\ \vdots \\ \Delta\underline{H}^{(1)}_n \end{bmatrix}$$

is the (abstract) vector of height differences.

Statistical testing for deformation is based on this variate $\underline{\mathcal{E}}$.

15.10.2 Horizontal deformation analysis

In two dimensions we proceed in the same way as in the one-dimensional case, except that

1. It is tempting to write the plane co-ordinates as *complex numbers*.
2. There are now *two* datum points, the co-ordinates of which are considered identical between the two epochs.

So, if there are n points, the size of the variance matrix is now $(n-2)\times(n-2)$. The variance matrix is now also complex valued, and *Hermitian*: its transpose is its complex conjugate.

liittoluku

18 The testing variate is again the shifting variate[18]

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \left(\underline{\mathbf{d}}^{(AB)}\right)^{\dagger}\left(\boldsymbol{\Sigma}_{\mathrm{dd}}^{(AB)}\right)^{-1}\underline{\mathbf{d}}^{(AB)},$$

in which $\underline{\mathbf{d}}$ is the complex vector of all co-ordinate differences, or *displacement vector*:

$$\begin{aligned}\underline{\mathbf{d}}^{(AB)} &\stackrel{\mathrm{def}}{=} \\ &= \begin{bmatrix} \underline{x}_3^{(AB)}(t_2) - \underline{x}_3^{(AB)}(t_1) + i\left(\underline{y}_3^{(AB)}(t_2) - \underline{y}_3^{(AB)}(t_1)\right) \\ \underline{x}_4^{(AB)}(t_2) - \underline{x}_4^{(AB)}(t_1) + i\left(\underline{y}_4^{(AB)}(t_2) - \underline{y}_4^{(AB)}(t_1)\right) \\ \vdots \\ \underline{x}_n^{(AB)}(t_2) - \underline{x}_n^{(AB)}(t_1) + i\left(\underline{y}_n^{(AB)}(t_2) - \underline{y}_n^{(AB)}(t_1)\right) \end{bmatrix} = \begin{bmatrix} \Delta\mathbf{z}_3^{(AB)} \\ \Delta\mathbf{z}_4^{(AB)} \\ \vdots \\ \Delta\mathbf{z}_n^{(AB)} \end{bmatrix},\end{aligned}$$

with

$$\begin{aligned}\Delta\mathbf{z}_i^{(AB)} &= \underline{\mathbf{z}}_i^{(AB)}(t_2) - \underline{\mathbf{z}}_i^{(AB)}(t_1) = \\ &= \left(\underline{x}_i^{(AB)}(t_2) - \underline{x}_i^{(AB)}(t_1)\right) + i\left(\underline{y}_i^{(AB)}(t_2) - \underline{y}_i^{(AB)}(t_1)\right), \quad i = 3,\ldots,n.\end{aligned}$$

AB is the chosen datum or starting point for both epochs t_1 and t_2. The other points are numbered 3, 4, …, n. The symbol † signifies both

19 transposition and complex conjugate, the *Hermitian*[19] *conjugate*:

$$A^{\dagger} \stackrel{\mathrm{def}}{=} \overline{A^{\mathsf{T}}} = \overline{A}^{\mathsf{T}}.$$

15.10.3 Example

Let the adjusted co-ordinates $\underline{\mathbf{x}}_i(t_1)$, $i = 3, \ldots, 6$ of the deformation

20 network from the first measurement epoch be given[20] in table 15.9a,

---

[18] *Warning:* in Cooper's book (Cooper, 1987, page 335) there is a mistake under equation (9.52): the correct equation is (inverse, not transpose):

$$\Omega = \widehat{\mathbf{d}}^{\mathrm{t}} Q_{\mathrm{d}}^{-1} \widehat{\mathbf{d}}.$$

[19] Charles Hermite (1822–1901) FRS FRSE was a French mathematician and number theorist.

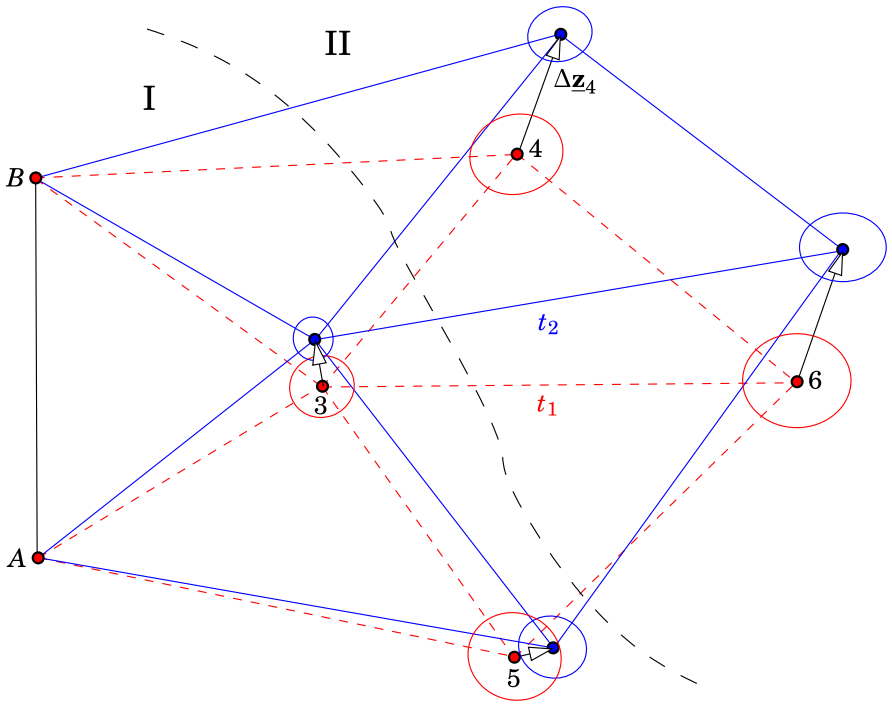


FIGURE 15.9. Horizontal deformation monitoring network for the epochs t_1 (red) and t_2 (blue). Realistic error ellipses. The alternative hypothesis that domain II moves with respect to domain I is also indicated.

and the co-ordinates of the second measurement epoch $\underline{\mathbf{x}}_i(t_2)$, $i = 3, \ldots, 6$ be given in table 15.9b.

Compute the inter-epoch differences vector $\underline{\mathbf{d}}$, table 15.9c.

Using real numbers, with the definition

$$\underline{\mathrm{d}} \stackrel{\text{def}}{=} \underline{\mathrm{x}}(t_2) - \underline{\mathrm{x}}(t_1) = \left[\begin{array}{c} \underline{x}_3(t_2) - \underline{x}_3(t_1) \\ \underline{y}_3(t_2) - \underline{y}_3(t_1) \\ \hline \underline{x}_4(t_2) - \underline{x}_4(t_1) \\ \underline{y}_4(t_2) - \underline{y}_4(t_1) \\ \hline \underline{x}_5(t_2) - \underline{x}_5(t_1) \\ \underline{y}_5(t_2) - \underline{y}_5(t_1) \\ \hline \underline{x}_6(t_2) - \underline{x}_6(t_1) \\ \underline{y}_6(t_2) - \underline{y}_6(t_1) \end{array}\right] = \left[\begin{array}{c} \Delta\underline{x}_3 \\ \Delta\underline{y}_3 \\ \hline \Delta\underline{x}_4 \\ \Delta\underline{y}_4 \\ \hline \Delta\underline{x}_5 \\ \Delta\underline{y}_5 \\ \hline \Delta\underline{x}_6 \\ \Delta\underline{y}_6 \end{array}\right],$$

[20] These are *only* the co-ordinates of the points to be tested. They are assumed to be connected, for both epochs, to the same two unnamed datum points — say, points 1 and 2, or A and B — outside the area, which are assumed to be motionless.

TABLE 15.9. Horizontal deformation analysis, co-ordinates. Unit metre.

(a) Epoch 1			(b) Epoch 2			(c) Displacements		
i	$\underline{x}_i(t_1)$	$\underline{y}_i(t_1)$	i	$\underline{x}_i(t_2)$	$\underline{y}_i(t_2)$	i	$\Delta\underline{x}_i$	$\Delta\underline{y}_i$
3	1234.123	2134.453	3	1234.159	2134.448	3	+0.036	−0.005
4	1681.045	2507.487	4	1681.123	2507.516	4	+0.078	+0.029
5	755.495	2623.456	5	755.507	2623.487	5	+0.012	+0.031
6	1248.865	3051.775	6	1248.951	3051.807	6	+0.086	+0.032

we find by computation

$$\underline{\mathrm{d}}^{\mathsf{T}}\underline{\mathrm{d}} = \sum_{i=3}^{6}\left(\left(\underline{x}_i(t_2)-\underline{x}_i(t_1)\right)^2+\left(\underline{y}_i(t_2)-\underline{y}_i(t_1)\right)^2\right) = 0.017771\,\mathrm{m}^2.$$

Similarly with complex numbers, with the definition

$$\underline{\mathbf{d}} \overset{\text{def}}{=} \begin{bmatrix} \underline{\mathbf{z}}_3(t_2)-\underline{\mathbf{z}}_3(t_1) \\ \underline{\mathbf{z}}_4(t_2)-\underline{\mathbf{z}}_4(t_1) \\ \underline{\mathbf{z}}_5(t_2)-\underline{\mathbf{z}}_5(t_1) \\ \underline{\mathbf{z}}_6(t_2)-\underline{\mathbf{z}}_6(t_1) \end{bmatrix} = \begin{bmatrix} \Delta\underline{\mathbf{z}}_3 \\ \Delta\underline{\mathbf{z}}_4 \\ \Delta\underline{\mathbf{z}}_5 \\ \Delta\underline{\mathbf{z}}_6 \end{bmatrix}$$

we obtain in the same way by computation

$$\underline{\mathbf{d}}^{\dagger}\underline{\mathbf{d}} = \sum_{i=3}^{6}\left(\overline{\underline{\mathbf{z}}}_i(t_2)-\overline{\underline{\mathbf{z}}}_i(t_1)\right)\left(\underline{\mathbf{z}}_i(t_2)-\underline{\mathbf{z}}_i(t_1)\right) = 0.017771\,\mathrm{m}^2.$$

Here, $\underline{\mathbf{z}}_i \overset{\text{def}}{=} \underline{x}_i + i\underline{y}_i$, and $\overline{\underline{\mathbf{z}}}_i \overset{\text{def}}{=} \underline{x}_i - i\underline{y}_i$ is its complex conjugate.

Let the precisions (mean co-ordinate errors) of the co-ordinates $\underline{x}_i(t_1)$ and $\underline{y}_i(t_1)$ measured at the first epoch be $\sigma_1 = \pm 5\,\mathrm{cm}$, and the precisions of the co-ordinates $\underline{x}_i(t_2)$ and $\underline{y}_i(t_2)$ of the second epoch $\sigma_2 = \pm 1\,\mathrm{cm}$ — *for every point*, and furthermore the co-ordinates are assumed to be uncorrelated.[21] The variance matrices of the co-ordinate vectors are thus

$$\Sigma_1 = \sigma_1^2 I, \qquad \Sigma_2 = \sigma_2^2 I.$$

We compute the mean error σ_Δ of a *single co-ordinate displacement* $\Delta\underline{x}_i = \underline{x}_i(t_2) - \underline{x}_i(t_1)$, or, equivalently, $\Delta\underline{y}_i = \underline{y}_i(t_2) - \underline{y}_i(t_1)$. Propagation of

[21] This is obviously unrealistic: in real networks, the point error grows with the distance from the datum points, and the co-ordinate errors within each epoch are strongly correlated.

variances yields

$$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2 = (25+1)\,\mathrm{cm}^2 = 26\,\mathrm{cm}^2.$$

Now, the variance matrix of the co-ordinate *displacements* is

$$\Sigma_{\mathsf{dd}} = \Sigma_1 + \Sigma_2 = \sigma_\Delta^2 I,$$

with $\sigma_\Delta = \sqrt{26}\,\mathrm{cm} = 5.1\,\mathrm{cm} = 0.051\,\mathrm{m}$.

Now, we choose for the mean error of unit weight just this value: $\sigma \overset{\text{def}}{=} \sigma_\Delta$. Then we may also write

$$\Sigma_{\mathsf{dd}} = \sigma_\Delta^2 Q_{\mathsf{dd}} = \sigma^2 Q_{\mathsf{dd}} = \sigma^2 I,$$

so the weight-coefficient matrix is the unit matrix.

Compute the deformation's *testing variate*, the shifting variate:

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \underline{\mathsf{d}}^{\mathsf{T}}\Sigma_{\mathsf{dd}}^{-1}\underline{\mathsf{d}} = \frac{\underline{\mathsf{d}}^{\mathsf{T}}Q_{\mathsf{dd}}^{-1}\underline{\mathsf{d}}}{\sigma^2} = \frac{\underline{\mathsf{d}}^{\mathsf{T}}\underline{\mathsf{d}}}{\sigma^2}.$$

Here, $\underline{\mathsf{d}} = \underline{\mathsf{x}}(t_2) - \underline{\mathsf{x}}(t_1)$ is the displacement vector, the abstract vector of co-ordinate differences between the epochs. Because we assume that both co-ordinate sets are given in the same, common datum, the definition points of which nevertheless do *not* belong to the set 3–6, we may assume that all co-ordinates are free. In that case, the number of degrees of freedom is $b = 2n = 8$, where n is the number of points. The variance matrix of the components of the displacement vector, or vector of co-ordinate differences, $\underline{\mathsf{d}}$, is $\sigma^2 I$. We obtain

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \frac{1}{0.0026\,\mathrm{m}^2}\left(\underline{\mathsf{d}}^{\mathsf{T}}\underline{\mathsf{d}}\right) = \frac{0.017771\,\mathrm{m}^2}{0.0026\,\mathrm{m}^2} = 6.835.$$

Question The quantity $\underline{\mathcal{E}}/\sigma^2$ is distributed according to the χ_8^2 distribution. If the limit value of this distribution for a significance level of 95 % is 15.51 (see Cooper (1987) page 355), has a deformation probably taken place in this case?

Answer No, it has not. $6.835 < 15.51$.

Question If, however, the assumed precisions were $\sigma_1 = \sigma_2 = \pm 1\mathrm{cm}$, would then a deformation have probably taken place, at a significance level of 95 %?

Answer Yes, it would. $\sigma^2 = (1+1)\,\mathrm{cm}^2 = 0.0002\,\mathrm{m}^2$ and

$$\frac{1}{\sigma^2}\underline{\mathcal{E}} = \frac{1}{0.0002\,\mathrm{m}^2}\left(\underline{\mathsf{d}}^{\mathsf{T}}\underline{\mathsf{d}}\right) = \frac{0.017771\,\mathrm{m}^2}{0.0002\,\mathrm{m}^2} = 88.9 > 15.51.$$

———

Self-test questions

1. What is the relationship, and difference, between the variance matrix $\Sigma_{\ell\ell}$ and the weight-coefficient matrix $Q_{\ell\ell}$ of the observations?
2. What is the relationship between the *a priori* variance of unit weight σ^2 and the *a posteriori* one $\widehat{\sigma^2}$?
3. What are errors of the first kind and errors of the second kind?
4. What is the power of a statistical test?
5. What is the relationship between the significance level α_{χ^2} of the overall validation test and the significance level α of the outlier test on the individual observations? Why?
6. What is redundancy, and why is it important?
7. What is interior and what is exterior reliability?
8. What are the steps in planning and measuring a geodetic network?
9. What is the Hermitian conjugate of a matrix?
10. Heathrow airport, UK, receives about 35 million incoming international passengers per year. A fancy new system is proposed to be installed that, by analysing the behaviour of people from closed-circuit video, can "flag" them as potential terrorists. The rate of "false positives", or errors of the first kind, is $1-\alpha = 1\,\%$. The rate of errors of the second kind, false negatives or justified but not-called alarms, $1-\beta$, is believed to be small, less than $50\,\%$.

    The background to this is that since 1970 there have been some 4000 deaths due to terrorism in the UK.

    How would you handle the passengers flagged by the system, and why?

    (a) Have them all killed.

    (b) Arrest and investigate them.

    (c) Send them back to where they came from.

    (d) Have a chat with them and informally look into their backgrounds before doing anything.

    (e) That system is worthless.