# A Discussion on Least Squares Adjustment with Worked Examples


## A. Fotiou

*Professor, Aristotle University of Thessaloniki, Dept. of Geodesy and Surveying, Lab. of Geodetic Methods and Satellite Applications*
afotiou@topo.auth.gr



**Summary:** Some remarks on least squares adjustment topics are presented, often bypassed or given little attention. The minimum number of observable parameters, the relation between observables and other unknown parameters, the role of the functional model and the linearization effect are fundamental to build a correct adjustment model. Discussing simple worked examples from the fields of geodesy and surveying, with extensions to practical applications, a closer look is gained. Especially, a better understanding of the least squares principal concepts could be very helpful for students, young scientists and non-experts involved in problems requiring adjustment of observations.



**Key words:** adjustment of observations, parametric degree, functional model, adjustment methods, linearization effect, adjustment worked examples


### 1. Building the concept of the Least Squares adjustment model

A Least Squares (LS) adjustment problem includes observable parameters (true /unknown values), observations (arithmetic values), observational errors (true /unknown values) and often some other unknown parameters (true/unknown values). The relation between observables and/or unknown parameters form the functional adjustment model, expressed by non-linear equations in the general case. Moreover, the behavior of the errors and the observations is described by the same stochastic model.

The goal of the adjustment is the estimation of unknown parameters, given a set of observations or measurements, satisfying the minimization of squared errors. The concept of observation is the same with that of measurement, so they are interchangeably used throughout the paper. Applying an adjustment algorithm/method, the estimates of parameters and observables – adjusted/corrected values - satisfy also the functional (compatibility). In addition, the estimated observational errors, frequently called residuals, are the key to assess the adjustment results in a second necessary step. Without error control, nothing can be stated for the quality of the results. Since the number of observations is always greater than the minimum

needed for a solution, there can be more solutions by means of properly selected subsets of observations. To derive a unique solution and take advantage of all available observations simultaneously, a convenient criterion must be adopted. This solution should be preferably best in terms of quality/accuracy (precision and reliability) and possibly with a minimum cost.

Among various minimization/optimization criterions, the well-known LS criterion is one that has been proved efficient for best estimates since the early years of 1800s, the era of Legendre, Gauss and other famous scientists afterwards. As the magnitude of errors is the key for the quality control, its influence in the adjustment results should be investigated (design/optimization problems).

The LS criterion can be applied either in a simple form, where all errors equally effect the solution or in an extended form when the contribution of each error is counted regarding its importance, something that is called weight. Large weights lead generally to smaller error estimates or smaller weights tent to produce larger errors. This fact can be realized in the functional expression of the LS criterion with weights, where in the minimization process the weights act as zoom in or zoom out scale factors. In the most general case it is possible to count also for the interrelation of weights. The concept of a weight is proportionally related to the precision of the corresponding measurement or inversely proportional to its uncertainty; larger weights are associated with more precise measurements. Therefore, a relation is established between weight and precision.

Considering observational errors as being just small quantities that tend to cancel out as the number of observations increases more and more, applying the LS criterion a solution is achieved. This is the case of a deterministic adjustment model. Assigning proper weighs or generally a proper weight matrix demands for a stochastic error model. Under this requirement, errors are considered as random or stochastic variables, meaning that they are sample values of random variables that take different values when the observations are repeated under similar conditions. The same stochastic model holds also for the observations. In theory, an infinite number of repeated observations may be described by a common probability density function, from where the expectation and the covariance matrix of random errors can be determined. Even if it is not necessary the knowledge of the density function, it is possible to express the stochastic model by the expectation of random errors equal to zero and a covariance matrix that is determined and considered absolutely or a priori known. So far, the LS mathematical adjustment model or simply the adjustment model (functional + stochastic) has been built, allowing for unbiased estimations, i.e., those having mean values equal to their real ones for the infinite sample. The application of the law of covariance propagation, also known as the law of error propagation, provides the estimation of covariance matrices for any estimated parameter.

The elements of the error/observation covariance matrix are measures of precision. A weight matrix should be then preferably determined so that the estimated parameter have maximum precision, i.e., minimum variances. A more general criterion for unbiased estimations is expressed by the minimization of the variances of the estimated parameters. For mathematical simplicity, a compromise is made and linear estimations instead of non-linear are used. On the other hand, LS linear estimates are identical to the minimum - variance linear estimates in case the weight matrix is equal to the inverse of the error covariance matrix. This is the best possible choice for the weight matrix resulting in LS estimates characterized by maximum precision. Nonlinear LS estimates are identical to nonlinear estimates that have minimum - variance only if random errors follow the Normal (Gauss) distribution. Under the same stochastic behavior, LS linear estimates are also identical to the ones derived by the method of Maximum Likelihood Estimation. In conclusion, the LS adjustment of observations leads to Best Linear Unbiased Estimations (BLUE).

The final step of the adjustment is the assessment of the results where estimated covariance matrices are used together with error and other estimates. This is usually made through a statistical evaluation or reliability control. In this step, the distribution of errors must be known. Among various choices the Normal/Gauss distribution has been accepted for it is both a realistic and mathematically simple hypothesis. First, the reliability of the results is examined by hypothesis testing, verifying the validity of the a priori stated model hypotheses (null hypotheses). Usually, we try to detect and localize possible significant model errors, looking mainly in the estimated size in relation to their intercorrelation as given by the estimated covariance matrix. Ending with an acceptable set of observations and a realistic covariance matrix, in association with an efficient functional model, measures of marginal detectable errors or marginal effects on the estimated parameters can be also derived (reliability measures). At this point, computed measures of precision or uncertainty are also measures of accuracy/quality, like confidence intervals, areas, spaces. The assessment process could be carried out in a pre-analysis step, before the measurement campaign, in a trial and error optimization under certain standards and specification criteria set a priori (design criteria, optimization problems).

The accuracy or quality of an estimated parameter express the degree of closeness to its real (unknown) value; therefore, accuracy is always unknown. Moreover, the precision of an observation or an estimated parameter express the degree of closeness between repeated observations or estimated parameters. Precision of a parameter is given usually by the variance or standard deviation, also called r.m.s.e (root mean square error) for unbiased estimations. In case the observations are free of systematic errors (biases) and outliers (gross errors), measures of precision are

also measures of accuracy as errors are influenced only by the (unavoidable) random errors. Pre-processing and post-adjustment tests - usually statistical tests- refer to the reliability control.

Some authors characterize outliers as mistakes and define errors as being only random and systematic; obviously, mistakes have large values and normally could easily be detected and excluded. We should note that random errors may have any large value as illustrated by the normal distribution, though this is exceptionally not likely for a correct model. A more appropriate term for all types of errors might be the term model errors; the division to three or two categories supports mainly teaching requirements and an ease understanding with respect to their sources. Generally, it is difficult to distinguish the specific type of an estimated error, especially in models with many interrelated variables and parameters, for instance in the adjustment of geodetic and surveying networks, where a weak configuration in relation to the associated weights affects seriously error estimates, making thus difficult or even impossible a correct error localization.

Despite any statistical evaluation, there is not a hundred percent guaranty that all possible existed errors have been removed. Based on a statistical decision, there is always a probability (risk) to reject a correct null hypothesis - committing a type-I error - or to accept a wrong one - committing a type-II error. In the present paper, we will not deal with statistical assessment of the results. The goal is to discuss and make some useful remarks on basic topics, sometimes bypassed or completely omitted in literature. The number of the minimum observable parameters needed for an adjustment, the relation between the number of observables and other unknown parameters, the role of the adjustment model and the linearization process are hot and essential topics for any adjustment problem.

## 2. Formulating the functional model

Using matrix notation, a list of symbols for parameters, variables and constants is first given:

- the observable parameters or observables (unknown):

$$\mathbf{y}^{\alpha} = \begin{bmatrix} y_1^{\alpha} & y_2^{\alpha} & \dots & y_n^{\alpha} \end{bmatrix}^T, \quad i = 1, 2, ..., n$$

- the observations or measurements (arithmetic values):

$$\mathbf{y}^{b} = \begin{bmatrix} y_1^{b} & y_2^{b} & \dots & y_n^{b} \end{bmatrix}^T, \quad i = 1, 2, ..., n$$

- the errors (unknown):

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}^T, \quad i = 1, 2, ..., n$$

- the unknown parameters:

$$\mathbf{x}^{\alpha} = \left[ x_1^{\alpha} \quad x_2^{\alpha} \quad \dots \quad x_m^{\alpha} \right]^T, \quad i = 1, 2, ..., m$$

- the approximate values of the unknown parameters:

$$\mathbf{x}^{o} = \left[ x_1^{o} \quad x_2^{o} \quad \dots \quad x_m^{o} \right]^T, \quad i = 1, 2, ..., m$$

- the corrections to the unknown parameters:

$$\mathbf{x} = \mathbf{x}^{\alpha} - \mathbf{x}^{o} = \left[ x_1 \quad x_2 \quad \dots \quad x_m \right]^T, \quad i = 1, 2, ..., m$$

- the approximate values of the observations:

$$\mathbf{y}^{o} = \mathbf{f}(\mathbf{x}^{o}) = \left[ y_1^{o} \quad y_2^{o} \quad \dots \quad y_n^{o} \right]^T, \quad i = 1, 2, ..., n$$

The relation between observables, observations and errors is given by the conventional with respect to the sign (+ or –) relation,

$$\mathbf{y}^{b} = \mathbf{y}^{\alpha} + \mathbf{v} \tag{2.1}$$

With **P,** the weight matrix and **C** the variance covariance matrix of errors or equally of observations, we usually have two types of stochastic models:

$$\mathbf{C} = \sigma^2 \mathbf{Q}, \ \mathbf{P} = \mathbf{C}^{-1} = (\sigma^2 \mathbf{Q})^{-1} \quad (\sigma^2 \text{ known or a priori known and } \mathbf{Q} \text{ known}) \tag{2.2}$$

$$\mathbf{C} = \sigma^2 \mathbf{Q}, \ \mathbf{P} = \mathbf{Q}^{-1} \quad (\sigma^2 \text{ unknown } (\hat{\sigma}^2 \text{ is used}) \text{ and } \mathbf{Q} \text{ known}) \tag{2.3}$$

Model (2.2) reflects an absolutely known observation accuracy, that is a theoretical case not so realistic as (2.3). On the other hand, the first choice offers much more mathematical simplicity in estimating measures of accuracy and reliability and is preferred in many software packages. The choice depends on the user's opinion in relation to the specific adjustment problem.

The LS-criterion, the minimization of the quantity φ, is expressed by

$$\varphi = \sum v_i^2 = \mathbf{v}^T \mathbf{v} = \min., \ \mathbf{P} = \mathbf{I} \ \text{(simple form).} \tag{2.4}$$

$$\varphi = \sum p_i v_i^2 = \mathbf{v}^T \mathbf{P} \mathbf{v} = \min., \ \mathbf{P} = \text{diag}(p_1, p_2, ..., p_n) \ \text{(extended form)} \tag{2.5}$$

$$\varphi = \mathbf{v}^T \mathbf{P} \mathbf{v} = \min., \quad \mathbf{P} \text{ full matrix} \quad \text{(general form)} \tag{2.6}$$

In any LS adjustment problem, there must be a least number of observable parameters that give a solution, i.e., making possible the determination of all unknown parameters. We will call this minimum number, the parametric degree r.

The necessary and sufficient condition to characterize a problem as an adjustment problem is that the number of observations must be greater than the parametric degree ($n > r$) and at the same time a proper subset of observations $n_o$ equal to r ($n_o = r$) exists. Otherwise, the problem is not properly described, even if $n > r$, and

a solution does not exist. Always in an adjustment problem there are more than one proper selections of $n_o$ and therefore more than one solutions but not the LS-solution.

The general form of the functional model is,

$$\mathbf{u}(\mathbf{y}^{\alpha}, \mathbf{x}^{\alpha}) = \mathbf{0} \tag{2.7}$$

The compatibility between estimates $(\hat{\mathbf{y}}^{\alpha}, \hat{\mathbf{x}}^{\alpha})$ and the functional model is expressed by $\mathbf{u}(\hat{\mathbf{y}}^{\alpha}, \hat{\mathbf{x}}^{\alpha}) = \mathbf{0}$. Understanding that the number m of unknown parameters in (2.7) is (n + m) while r is only required for a solution, there must be a number s of independent mathematical conditions/equations in the model,

$$s = (n + m) - r = (n - r) + m = f + m \tag{2.8}$$

where $f$ is the degrees of freedom representing the redundant observations.

In geodetic adjustment problems, as it is the case of trigonometric, levelling and GPS/GNSS network adjustments, r is a constant regardless of the number n of observations. As the number n, can be as greater as possible, we can only select the number m of the unknown parameters with respect to r. This choice determines the specific form of the functional model (2.7) and thus the number s of equations, something fundamental for the adjustment method that should be followed.

## 3. Basic observation adjustment methods

All the adjustment methods are equal, meaning that they lead to the same adjustment results regardless the specific method. The choice of one or another method depends on the type of the functional model. In the following, we will examine the three basic adjustment methods, namely: the method of observation equations (the method of parameters or the method of indirect observations), the method of condition equations (the method of direct observations) and the method of mixed or compound equations (the general method), see e.g., see, e.g., Wells and Krakiwsky 1971, Mikhail and Ackermann 1976, Dermanis 1986/1987, Koch 1987, Dermanis and Fotiou 1992, Rossikopoulos 1999, Fotiou 2007, Sneew et al. 2015

### *3.1 The method of observation equations*

This is the most common adjustment method including unknown parameters $\mathbf{x}^{\alpha}$ and having an obvious number of (independent) equations.

The choice m = r, and consequently s = n, makes possible (2.7) expressed by

$$\mathbf{y}^{\alpha} = \mathbf{f}(\mathbf{x}^{\alpha}) \tag{3.1}$$

Noting that each one of the observables is written as a function of only the unknown parameters (constants may be included); no observable can be present on the right side of (3.1). Because it is generally difficult to deal with non-linear systems and estimations, as a rule in all methods, linearization is adopted. Expanding (3.1) in a Taylor series around the approximate point $\mathbf{x}^o$, computed by any suitable means - often using a subset of observations - and keeping first order terms, we have,

$$\mathbf{y}^\alpha = \mathbf{f}(\mathbf{x}^o) + \frac{\partial \mathbf{f}}{\partial \mathbf{x}^\alpha}|_o (\mathbf{x}^\alpha - \mathbf{x}^o) = \mathbf{y}^o + \mathbf{A}\mathbf{x}\,,\ \mathbf{y}^b - \mathbf{y}^o = \mathbf{A}\mathbf{x} + \mathbf{v} \tag{3.2}$$

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{v} \tag{3.3}$$

where, the (n x m) matrix **A** is called the design matrix, having elements the partial derivatives computed by means of $\mathbf{x}^o$ and **b** is the (n x 1) vector of the reduced observations noting that $\mathbf{y}^o = \mathbf{f}(\mathbf{x}^o)$ is accurately computed.

Having thus (3.3), the linear system of observations, we note that this is also an underdetermined system for there are s = n equations with (m + n) > n unknowns. System (3.3) has generally an infinite number of solutions and the question arising is which is the best one. Among all solutions, there exists one that satisfies the LS criterion, and this is a best solution. Under the LS optimization constraint, and without going into proofs, the (m x m) or (r x r) normal equations system is formed,

$$(\mathbf{A}^\mathbf{T}\mathbf{P}\mathbf{A})\hat{\mathbf{x}} = \mathbf{A}^\mathbf{T}\mathbf{P}\mathbf{b}\,,\quad \text{or} \quad \mathbf{N}\hat{\mathbf{x}} = \mathbf{u} \tag{3.4}$$

where **N** is the (symmetric) normal equation matrix and **u** the (m x 1) constant vector. From (3.4) we get the LS solution/estimates,

$$\hat{\mathbf{x}} = (\mathbf{A}^\mathbf{T}\mathbf{P}\mathbf{A})^{-1}(\mathbf{A}^\mathbf{T}\mathbf{P}\mathbf{b}) = \mathbf{N}^{-1}\mathbf{u} \quad [\det(\mathbf{N}\neq 0)], \quad \hat{\mathbf{x}}^\alpha = \hat{\mathbf{x}}^o + \hat{\mathbf{x}} \tag{3.5}$$

$$\hat{\mathbf{v}} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}, \quad \hat{\mathbf{y}}^\alpha = \mathbf{y}^b - \hat{\mathbf{v}} \quad \text{or} \quad \hat{\mathbf{y}}^\alpha = \mathrm{f}(\hat{\mathbf{x}}^\alpha) \tag{3.6}$$

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{v}}^\mathrm{T}\mathbf{P}\hat{\mathbf{v}}}{\mathrm{n}-\mathrm{r}} = \frac{\hat{\mathbf{v}}^\mathrm{T}\mathbf{P}\hat{\mathbf{v}}}{\mathrm{n}-\mathrm{m}} = \frac{\hat{\mathbf{v}}^\mathrm{T}\mathbf{P}\hat{\mathbf{v}}}{\mathrm{f}} \tag{3.7}$$

The associated estimates of covariance matrices, are obtained by applying the covariance propagation law, i.e.,

$$\mathbf{C}_{\hat{\mathbf{x}}} = \mathbf{C}_{\hat{\mathbf{x}}^\alpha} = \mathbf{N}^{-1} \quad \text{if } \mathbf{P} = \mathbf{C}^{-1},\ \hat{\mathbf{C}}_{\hat{\mathbf{x}}} = \hat{\mathbf{C}}_{\hat{\mathbf{x}}^\alpha} = \hat{\sigma}^2\mathbf{N}^{-1} \text{ if } \mathbf{P} = \mathbf{Q}^{-1} \tag{3.8}$$

$$\mathbf{C}_{\hat{\mathbf{v}}} = \mathbf{P}^{-1} - \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^\mathrm{T} = \mathbf{C} - \mathbf{C}_{\hat{\mathbf{y}}^\alpha} \quad \text{if } \mathbf{P} = \mathbf{C}^{-1},\ \hat{\mathbf{C}}_{\hat{\mathbf{v}}} = \hat{\sigma}^2\mathbf{C}_{\hat{\mathbf{v}}} \quad \text{if } \mathbf{P} = \mathbf{Q}^{-1} \tag{3.9}$$

Any other estimate based on the above best estimates is also a best estimate.

In case of initial linear equations, $\mathbf{y}^\alpha = \mathbf{A}\mathbf{x}^\alpha + \mathbf{t}$ or $\mathbf{b} = \mathbf{A}\mathbf{x}^\alpha + \mathbf{v}$, where a constant

vector **t** is possibly present. Comparing with (3.3), $\mathbf{y}^{b} - \mathbf{t} \equiv \mathbf{b}$ and $\mathbf{x}^{\alpha} \equiv \mathbf{x}$ meaning that $\mathbf{x}^{o} = \mathbf{0}$ (no approximate values are needed), while matrices **A**, **N** remain the same. The approximate values $\mathbf{x}^{o}$ should be close enough to their correct estimates, otherwise serious linearization errors will affect the solution. In any case, the adjustment must be repeated with approximate values the previous estimates until a negligible difference, between consecutive results, e.g. between the unknown parameter estimates, is achieved.

### *3.2 The method of condition equations*

In this method, m = 0 and s = n – r = f. As we have only observables in the functional model, (2.7) becomes

$$\mathbf{g}(\mathbf{y}^{\alpha}) = \mathbf{0} \tag{3.10}$$

having s independent condition equations. Following the linearization, and taken as approximate values of observations the values of the observations, something that is followed in the classical approach,

$$\mathbf{g}(\mathbf{y}^{\alpha}) = \mathbf{g}(\mathbf{y}^{b}) + \frac{\partial \mathbf{g}}{\partial \mathbf{y}^{\alpha}}|_{b}(\mathbf{y}^{\alpha} - \mathbf{y}^{b}) = \mathbf{0}, \quad \text{or} \quad \mathbf{B}\mathbf{v} = \mathbf{w} \tag{3.11}$$

with obvious substitutions. **B** is the (s x s) matrix of partial derivatives computed by means of $\mathbf{y}^{b}$ and $\mathbf{w} = \mathbf{g}(\mathbf{y}^{b})$ the (s x 1) vector of the so-called closing errors.

The linear system (3.10) is again an underdetermined system as there are s = (n –r) equations with n > s unknowns. Imposing the LS criterion, the derived best LS solution is given by the (s x s) normal equations system,

$$(\mathbf{B}\mathbf{P}^{-1}\mathbf{B}^{\mathbf{T}})\hat{\mathbf{k}} = \mathbf{w}, \quad \mathbf{M}\hat{\mathbf{k}} = \mathbf{w}, \quad \hat{\mathbf{k}} = \mathbf{M}^{-1}\mathbf{w}, \quad \hat{\mathbf{v}} = \mathbf{P}^{-1}\mathbf{B}^{\mathrm{T}}\hat{\mathbf{k}} = \mathbf{P}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{w} \tag{3.12}$$

where **M** is the (symmetric) normal matrix ($\det(\mathbf{M} \neq 0)$). The adjusted observations, and the posteriori variance factor when needed, are given by

$$\hat{\mathbf{y}}^{\alpha} = \mathbf{y}^{b} - \hat{\mathbf{v}}, \quad \hat{\sigma}^{2} = \frac{\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{v}}}{n-r} = \frac{\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{v}}}{s} = \frac{\hat{\mathbf{v}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{v}}}{f} \tag{3.13}$$

Compatibility between adjusted observations and the functional model is expressed by, $\mathrm{g}(\hat{\mathbf{y}}^{\alpha}) = \mathbf{0}$, $\mathbf{B}\hat{\mathbf{v}} = \mathbf{w}$. Also, the associated covariance matrices are given by

$$\mathbf{C}_{\hat{\mathbf{v}}} = \mathbf{P}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{B}\mathbf{P}^{-1} = \mathbf{P}^{-1} - \mathbf{C}_{\hat{\mathbf{y}}^{\alpha}} \quad (\mathbf{P} = \mathbf{C}^{-1}) \quad \text{or} \quad \hat{\mathbf{C}}_{\hat{\mathbf{v}}} = \hat{\sigma}^{2}\mathbf{C}_{\hat{\mathbf{v}}} \quad (\mathbf{P} = \mathbf{Q}^{-1}) \tag{3.14}$$

We remind that a complete elimination of the unknown parameters from the observation equations (3.1) results in condition equations (3.10).

### 3.3 The method of mixed equations

This is the general adjustment method, realizing that it is a method of observation equations with respect to the unknown parameters $\mathbf{x}^{\alpha}$ and simultaneously a method of condition equations with respect to the observables.

Here, $s = n + m - r$ with $0 < m \leq r$ and thus $n - r < s \leq n$. We see that there are m unknown parameters, generally less than the number of r $(m < r)$. In case $m = r$, the model can be written in the form of the observation equations method but for some reason, e.g. due to mathematical complication, the method of mixed equations may be preferred. Since $m < r$, some of the parameters $\mathbf{x}^{\alpha}$ have been eliminated from (3.1); if all of them had been eliminated, then $m = 0$ and conditions (3.10) would be derived.

The mixed model $\mathbf{u}(\mathbf{y}^{\alpha}, \mathbf{x}^{\alpha}) = \mathbf{0}$ is expanded in a Taylor series around the point $(\mathbf{x}^{o}, \mathbf{y}^{b})$, as it is followed in the classical approach (discussed later). Keeping first order terms, we will have,

$$\mathbf{u}(\mathbf{y}^{\alpha}, \mathbf{x}^{\alpha}) = \mathbf{u}(\mathbf{y}^{b}, \mathbf{x}^{o}) + \frac{\partial \mathbf{u}}{\partial \mathbf{x}^{\alpha}}|_{o,b}(\mathbf{x}^{\alpha} - \mathbf{x}^{o}) + \frac{\partial \mathbf{u}}{\partial \mathbf{y}^{\alpha}}|_{o,b}(\mathbf{y}^{\alpha} - \mathbf{y}^{b}) = \mathbf{0}, \quad \text{or} \tag{3.15}$$

$$\mathbf{w} + \mathbf{A}\mathbf{x} - \mathbf{B}\mathbf{v} = \mathbf{0} \tag{3.16}$$

where, the matrices (s x m) **A**, (s x n) **B** and the (s x 1) vector $\mathbf{w} = \mathbf{u}(\mathbf{y}^{b}, \mathbf{x}^{o})$ are evaluated at $(\mathbf{x}^{o}, \mathbf{y}^{b})$ while **x** and **v** are estimated from the LS algorithm. As previously, (3.16) expresses an underdetermined system with s mixed equations and $(m + n) > s$ unknowns. The LS solution is given by the normal equation system,

$$(\mathbf{A}^{T}\mathbf{M}^{-1}\mathbf{A})\hat{\mathbf{x}} = -\mathbf{A}^{T}\mathbf{M}^{-1}\mathbf{w}, \quad \text{or} \quad \mathbf{N}\hat{\mathbf{x}} = -\mathbf{u} \rightarrow \hat{\mathbf{x}} = -\mathbf{N}^{-1}\mathbf{u}, \ [\det(\mathbf{N} \neq 0)] \tag{3.17}$$

$$\hat{\mathbf{x}}^{\alpha} = \hat{\mathbf{x}}^{o} + \hat{\mathbf{x}}, \quad \hat{\mathbf{v}} = \mathbf{P}^{-1}\mathbf{B}^{T}\mathbf{M}^{-1}(\mathbf{w} + \mathbf{A}\hat{\mathbf{x}}), \quad \hat{\mathbf{y}}^{\alpha} = \mathbf{y}^{b} - \hat{\mathbf{v}}, \quad \hat{\sigma}^{2} = \frac{\hat{\mathbf{v}}^{T}\mathbf{P}\hat{\mathbf{v}}}{s - m} \tag{3.18}$$

The covariance matrices for the above estimates are then given by

$$\mathbf{C}_{\hat{\mathbf{x}}} = \mathbf{C}_{\hat{\mathbf{x}}^{\alpha}} = \mathbf{N}^{-1} \quad \text{if} \quad \mathbf{P} = \mathbf{C}^{-1}, \quad \hat{\mathbf{C}}_{\hat{\mathbf{x}}} = \hat{\mathbf{C}}_{\hat{\mathbf{x}}^{\alpha}} = \hat{\sigma}^{2}\mathbf{N}^{-1} \quad \text{if} \quad \mathbf{P} = \mathbf{Q}^{-1} \tag{3.19}$$

$$\mathbf{C}_{\hat{\mathbf{v}}} = \mathbf{P}^{-1}\mathbf{B}^{T}\mathbf{M}^{-1}[\mathbf{I} - \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^{T}\mathbf{M}^{-1}]\mathbf{B}\mathbf{P}^{-1} \ (\mathbf{P} = \mathbf{C}^{-1}), \ \hat{\mathbf{C}}_{\hat{\mathbf{v}}} = \hat{\sigma}^{2}\mathbf{C}_{\hat{\mathbf{v}}} \ (\mathbf{P} = \mathbf{Q}^{-1}) \tag{3.20}$$

### 3.4 Constraints on the unknown parameters

An extension to the above (2.7) and (3.1) models, that both contain unknown parameters $\mathbf{x}^{\alpha}$, is to set constraints on all or some of the unknown parameters. This is the case when k equations $\mathbf{h}(\mathbf{x}^{\alpha}) = \mathbf{0}$ must be satisfied. In consequence, the linearized constraints,

$$\mathbf{h}(\mathbf{x}^{\alpha}) = \mathbf{h}(\mathbf{x}^{o}) + \frac{\partial \mathbf{h}}{\partial \mathbf{x}^{\alpha}}|_{o}(\mathbf{x}^{\alpha} - \mathbf{x}^{o}) = -\mathbf{z} + \mathbf{H}\mathbf{x} = \mathbf{0}\,, \quad \text{or} \quad \mathbf{H}\mathbf{x} = \mathbf{z} \tag{3.21}$$

must be included, as a new subset of equations, in the corresponding linear systems. For the related adjustment algorithms not given here, we underline some issues. The number of k-constraints mean that one or more of the m parameters are not independent and therefore not fundamental for the description of the problem. The independent (m – k) parameters, (m – k) = r < m, if they are known, the rest k parameters are determined; thus, we must include k equations in the functional model.

In the method of observation equations with constraints, the new total number of equations is given by $s' = n + m - r = n + k$ where $m - r = k$ or $m = r + k$. Likewise, in the method of mixed equations, the new total number of equations is, $s' = (n + m - r) + k = s + k$, where $m - r \leq k$ or $m \leq r + k$.

Sometimes it is better to eliminate a part or even all the constraints. A simple case of constraints that are eliminated, refers to fix a number of point coordinates in a geodetic network, in order to define the coordinate system or datum of the adjustment.

## 4. An introductory example

The role of the parametric degree r is fundamental in any adjustment model. More over the selection of the adjustment method and the linearization are important issues. With the help of worked examples we will try to clarify some related issues.

Consider a problem where the shape of a triangle on a horizontal plane must be determined. For this purpose, we measured just two angles of the triangle. Obviously, the third angle is the difference of their sum from 180º. Geometrically, the triangle's shape is determined by taking an arbitrary length of the side with the two angles at each end and intersecting the two lines formed by each angle (the third vertex is defined). Note that there are two solutions, two intersections at each halfplane; we choose one of them if there is some knowledge of the relative orientation. However, it is not possible with two measured angles to have error control; any serious error in the observations affects directly the position of the third vertex resulting in an erroneous shape. Realizing that the shape is determined by a minimum number of two angles (observables), the parametric degree r = 2. Also, n = 2 and therefore, the condition n > r for an LS adjustment is not fulfilled.

Let's modify the design and measure also the third angle, in total the three angles $(\omega_A, \omega_B, \omega_C)$. Suppose that measurements have known precision given by their standard deviations $\sigma_{\omega_A}, \sigma_{\omega_B}, \sigma_{\omega_C}$. While r remains constant (r = 2), regardless the increase of observations, we will have, n = 3 > 2. In addition, there is a minimum

suitable number of observables, $n_o = r = 2$ that gives a solution, e.g. any pair of angles determines the shape. Since the necessary and sufficient condition is fulfilled, the LS-best shape of the triangle can be estimated. Of course, any empirical solution can be obtained but this is not a best LS-solution.

In the following, we will present the formulation of the functional LS model in each one of the three basic adjustment methods, underling that the adjusted/estimated results from any method are the same.

### *4.1 Using the method of observation equation*

First, we must select m = r = 2 unknown parameters that fully describe the shape of the triangle and then formulate the three observation equations. One choice is to take any pair of the observables as being also unknown parameters. According to this choice, we'll have,

$$\mathbf{y}^{\alpha} = \begin{bmatrix} \omega_A & \omega_B & \omega_C \end{bmatrix}^T, \quad \mathbf{y}^{b} = \begin{bmatrix} \omega'_A & \omega'_B & \omega'_C \end{bmatrix}^T, \quad \mathbf{v} = \begin{bmatrix} v_{\omega_A} & v_{\omega_B} & v_{\omega_C} \end{bmatrix}^T$$

and the unknown parameters, $\mathbf{x}^{\alpha} = \begin{bmatrix} z_A & z_B \end{bmatrix}^T$.

Covariance and the weight matrices of the errors/observations are

$$\mathbf{C} = \mathrm{diag}(\sigma^2_{\omega_A}, \sigma^2_{\omega_B}, \sigma^2_{\omega_C}), \quad \mathbf{P} = \mathbf{C}^{-1} = \mathrm{diag}(1/\sigma^2_{\omega_A}, 1/\sigma^2_{\omega_B}, 1/\sigma^2_{\omega_C})$$

The observation equations $\mathbf{y}^{\alpha} = \mathbf{f}(\mathbf{x}^{\alpha})$ are then given by

$$\omega_A = z_A \quad \omega_B = z_B \quad \omega_C = 180^\circ - (z_A + z_B)$$

Although, these equations are directly linear and one should take advantage of it (see below), we will treat the model as if it was nonlinear.

With $\mathbf{x}^{o} = \begin{bmatrix} z^o_A & z^o_B \end{bmatrix}^T$, e.g., $z^o_A = \omega'_A$, $z^o_B = \omega'_B$, $\mathbf{x} = \mathbf{x}^{\alpha} - \mathbf{x}^{o} = \begin{bmatrix} \delta z_A & \delta z_B \end{bmatrix}^T$

and $\mathbf{y}^{o} = \mathbf{f}(\mathbf{x}^{o}) = \begin{bmatrix} \omega^o_A & \omega^o_B & \omega^o_C \end{bmatrix}^T = \begin{bmatrix} z^o_A & z^o_B & 180^\circ - (z^o_A + z^o_B) \end{bmatrix}^T$

the algorithm is applied as given above, noting that correct units should be used throughout the computational steps. Applying the algorithm, we begin with the evaluation of matrices **A**, **b** of the linear system $\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{v}$:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad \mathbf{b} = \mathbf{y}^{b} - \mathbf{y}^{0} = \begin{bmatrix} \omega'_A - z^o_A \\ \omega'_B - z^o_B \\ \omega'_C - \omega^o_C \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \omega'_C - (180^\circ - (\omega'_A + \omega'_B)) \end{bmatrix}$$

The easily formed (2x2) normal matrix **N** must be inverted and all the relevant es-

timates $\{\delta\hat{z}_A, \delta\hat{z}_B, \hat{z}_A, \hat{z}_B, \hat{v}_{\omega_A}, \hat{v}_{\omega_B}, \hat{v}_{\omega_c}, \hat{\omega}_A, \hat{\omega}_B, \hat{\omega}_C\}$ are then computed. In this example, having initially linear equations, we could alternatively proceed on with a direct estimation of $(\hat{z}_A, \hat{z}_B)$ instead of their corrections. In this case matrix **A** is the same while **b** is different. The same holds for **N** and **u**.

*4.1.1 Using coordinates as model parameters*

An alternative approach to formulate the functional model is to use (2-d) coordinates as model parameters. Due to the linearization, approximate coordinates for points under estimation are computed by any suitable way, usually by a subset of observations and some known point coordinates, given or arbitrary defined. Coordinates are preferably used in geodetic and surveying adjustment problems as they offer ease to express equations and compute any quantity. Coordinates cannot be derived from classical measurements like angles and distances. Therefore, in our measurement scheme, we must include or assign known coordinates to some points, either with respect to an official or arbitrary defined coordinate system.

Assume a geodetic network with many vertices, the triangle being a trivial case, with a number of measured angles, more than the parametric degree r. Assigning coordinates to points-vertices, the shape, size, orientation, and the position of the network are fully defined. On the other hand, the observables (here angles) define only the shape of the network. Therefore, to describe fully the problem, we must formulate, by some means, the rest information carried by the coordinates, i.e. we must define the scale/size, orientation and position with respect to the coordinate system. This is possible by means of imposing constraints on point coordinates. A simple, usually followed, type of constraints is to keep fixed some coordinates, such that they do not influence what is defined by the observables (minimal constraints) - here the shape of the network/triangle. In a 2-d network with four degrees of freedom in general, the position is defined by fixing the coordinates of one point. The orientation and the size are also defined by fixing the coordinates of another point. Generally, in a 2-d angular network any four coordinates, often two per point, eliminate the four degrees of freedom in plane; otherwise there is a so-called datum deficiency.

In the triangle with the three measured angles, the position, size and orientation are defined by keeping fixed the coordinates for two vertices, e.g., A and B. By considering arbitrary values, e.g., ($x_A = 0$ m, $y_A = 0$ m) and arbitrary orientation, e.g., plane azimuth $\alpha_{AB} = 90^o$ and scale, e.g., length $d_{AB} = 100$ m, then ($x_B$=100 m, $y_B$=0 m) are obviously obtained. Keeping A and B fixed their coordinates are absolutely known, i.e., constants. Applying the LS-algorithm, with unknown parameters and approximate values

$$\mathbf{x}^{\alpha} = \begin{bmatrix} x_C & y_C \end{bmatrix}^T, \quad \mathbf{x}^o = \begin{bmatrix} x_C^o & y_C^o \end{bmatrix}^T \text{ (computed from simple trigonometry)}$$

the three non-linear equations $\mathbf{y}^{\alpha} = \mathbf{f}(\mathbf{x}^{\alpha})$ are given by,

$$\omega_A = \arctan\frac{y_C}{x_C}, \quad \omega_B = \arctan\frac{y_C}{x_B - x_C}, \quad \omega_C = \arctan\frac{x_C}{y_C} + \arctan\frac{x_B - x_C}{y_C}$$

Also, the analytical structure of **A** and **b** is,

$$\mathbf{A} = \begin{bmatrix} \frac{\partial\omega_A}{\partial x_C}|_o & \frac{\partial\omega_A}{\partial y_C}|_o \\ \frac{\partial\omega_B}{\partial x_C}|_o & \frac{\partial\omega_B}{\partial y_C}|_o \\ \frac{\partial\omega_C}{\partial x_C}|_o & \frac{\partial\omega_C}{\partial y_C}|_o \end{bmatrix} = \begin{bmatrix} -\frac{y_C^o}{(d_{AC}^o)^2} & \frac{x_C^o}{(d_{AC}^o)^2} \\ \frac{y_C^o}{(d_{BC}^o)^2} & -\frac{x_C^o}{(d_{BC}^o)^2} \\ -\frac{y_C^o}{(d_{BC}^o)^2} + \frac{y_C^o}{(d_{AC}^o)^2} & \frac{x_C^o}{(d_{BC}^o)^2} - \frac{x_C^o}{(d_{AC}^o)^2} \end{bmatrix}$$

$$\mathbf{b} = \mathbf{y}^b - \mathbf{y}^0 = \begin{bmatrix} \omega'_A - \omega_A^0 & \omega'_B - \omega_B^0 & \omega'_C - \omega_C^0 \end{bmatrix}^T$$

### *4.2 Using the method of condition equations*

From the observation equations, eliminating the two unknown parameters, we get one independent condition equation (s = n – r = 3 – 2 = 1), i.e.,

$$g(\mathbf{y}^{\alpha}) = \omega_A + \omega_B + \omega_C - 180° = 0$$

which is also linear. This equation could easily be written directly realizing the geometric angle condition. The matrices of the linear system are then formed,

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \omega'_A + \omega'_B + \omega'_C - 180° \end{bmatrix}$$

Generally, the number of independent conditions is easily determined if the parametric degree r had been previously determined (s = n – r), e.g. thinking in terms of the observation equations method. To find the correct r and mainly to formulate the condition equations, it may be a simple matter for some problems having a simple geometric configuration/design; otherwise it is rather difficult with a possible risk for an incorrect model.

In this example, matrix **M** to be inverted is one (1x1)-one element. Generally, in view of the computational 'difficulty' compared to the method of observation equation where **N** is (mxm) or (rxr), we note that in cases that s = n – r < r or n < 2r the condition method is preferable. This is mainly the reason for using the condition method in the past when computing means were limited or even absent. On the contrary, formulating conditions is rather a difficult task in a complicated configuration like a classical geodetic trigonometric network. Understanding that in geo-

detic networks coordinates estimation is of primary importance, the ease of expressing the observation equations and the direct estimation of covariance matrices of unknown parameters, in relation to the plenty of computing power, makes the observation equation method almost the standard choice since some decades.

### 4.3 Using the method of mixed equations

In this method, m≤ r. Because m = r = 2 and s = n + m – r = 3, the functional mixed model is obtained directly by simply writing the observation equations as

$$u_1 = \omega_A - z_A = 0\,, \quad u_2 = \omega_B - z_B = 0\,, \quad u_3 = \omega_C + z_A + z_B - 180° = 0$$

from where,

$$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I}\,, \quad \mathbf{w} = \begin{bmatrix} \omega'_A - z_A^0 \\ \omega'_B - z_B^0 \\ \omega'_C + z_A^0 + z_B^0 - 180° \end{bmatrix}$$

Applying the LS algorithm best estimates are then derived.

In the common case where m<r, we would have had a different functional model. For instance, by eliminating the parameter $z_B$ : from equation $u_2$, $z_B = \omega_B$ and substituting in $u_3$ results in $u_3 = \omega_C + z_A + \omega_C - 180° = 0$. Finally, in the remaining two equations $u_1$ and $u_3$, we'll have only one unknown parameter ( $z_A$ ), noting that $u_3$ is of a mixed type and cannot be expressed as $\mathbf{y}^\alpha = \mathbf{f}(\mathbf{x}^\alpha)$, i.e., we have,

$$u_1 = \omega_A - z_A = 0\,, \quad u_3 = z_A + \omega_B + \omega_C - 180° = 0$$

The application of the algorithm results in $\hat{z}_A$. After the end of the adjustment algorithm the estimate $\hat{z}_B = \hat{\omega}_B$ is also obtained.

We can also apply the mixed equations method using coordinates as previously described. As far as the algorithmic and computing ease, we will see that in some problems such as the best fit of a function to data points (see below), the mixed equation method takes precedence.

### 4.4 Including observations of different type

Apart from the three measured (horizontal) angles, let's also measure the horizontal distance $d_{AB}$, having in total $n = 4$ observations. In this case the observations define the shape and the scale/size of the triangle. The minimum number of observables for a solution (shape + scale) is now three (r = 3), e.g., two angles to define the shape and the measured distance to define the scale of the triangle.

In the method of observation equations with s = 4 and m = r = 3, we can proceed

on as previously having now one more unknown parameter ( $z_{AB}$ ) and thus one more equation, e.g., $d_{AB} = z_{AB}$. Using coordinates as parameters, we must realize that the minimal constraints are three; two for the position of the triangle, e.g., fixing $(x_A, y_A)$ and one for the orientation, e.g., by fixing $y_B$ or $x_B$. Considering $(x_B, x_C, y_C)$ as the three unknown parameters (m=3), the additional observation equation is $d_{AB} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$ . The functional model is easily formed when more distance observations are included, noting that m = r = 3 in any case.

In the method of condition equations, s = n – r = 4 – 3 = 1 and the obvious condition is $g(\mathbf{y}^{\alpha}) = g_l = \omega_A + \omega_B + \omega_C - 180^\circ = 0$, leaving ´out of the model´ the distance $d_{AB}$, meaning that the scale/size of the triangle is not adjusted and simply is defined from the distance observation without error control on it. The same holds also for the method of observation equations, not directly shown as in the condition equation. Should this be avoided, one more or two distance observations should be observed. With one more distance observation, n=5, s = n – r = 5 – 3 = 2, and the second condition equation comes from the law of sines where two measured distances are included. With three distance observations, n = 3, s = n – r = 6 – 3 = 3, and the third condition equation comes again from the law of sines or even from the law of cosines.

If we would prefer to use the method of mixed equations with m<r, one at least of the unknown parameters should be eliminated from the equations of the method of observation equations.

## 5. Introducing redundant constraints on unknown parameters

In geodetic and surveying networks coordinates are used as unknown parameters. In the previous example, we discussed the definition of the coordinate system through minimal constraints on a number of point coordinates that are kept fixed. Minimal constraints do not affect the estimated errors and observable parameters, regardless of their choice; introducing them in the adjustment model, statistical tests on the quality of observations are meaningful. On the other hand, they affect the estimates of the unknown parameters/coordinates and their covariance matrices.

Very often, compliance with technical specifications or operational requirements demand for redundant constraints based on the supposed higher accuracy of the control points included in the network and kept fixed, to increase the overall accuracy. This is normally true but sometimes, a significant problem may have destroyed the (assumed) superior quality of the control points. Redundant constraints do affect everything in the adjustment results. Normally, their quality must be tested, usually statistically; otherwise the reason for unreliable results cannot be

safely detected, i.e. if the reason is the quality of the observations or the quality of the redundant constraints or both.

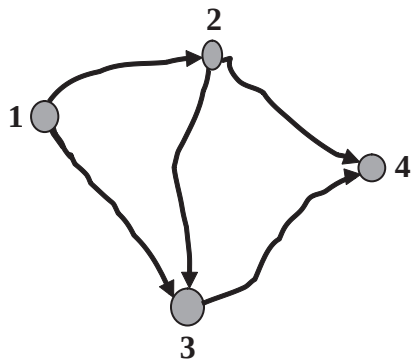
### *5.1 A worked example with and without constraints*

In a levelling network with four points (1, 2, 3, 4), five height differences were measured (n=5),

$$\mathbf{y}^{\alpha} = \left[\delta h_{12}\ \delta h_{13}\ \delta h_{23}\ \delta h_{24}\ \delta h_{34}\ \right]^{T}, \quad \mathbf{y}^{b} = \left[\delta h'_{12}\ \delta h'_{13}\ \delta h'_{23}\ \delta h'_{24}\ \delta h'_{34}\ \right]^{T}$$

The goal is to determine the best estimates of height differences between any pair of points under the condition $\delta h_{14} = q_{14} = \text{constant}$ .

The observations, obviously, define height differences. Ignoring, for the moment, the condition that should be satisfied, we need only three (r = 3) observables for a solution, e.g., $(\delta h_{12}\ \delta h_{13}\ \delta h_{24})$: from $(\delta h_{12}\ \delta h_{13})$ and the loop-condition on (1-2-3-1), $\delta h_{12} + \delta h_{23} - \delta h_{13} = 0$ , $\delta h_{23}$ is defined, while from the loop-condition on (2-4-3-2), $\delta h_{23} + \delta h_{34} - \delta h_{24} = 0$ , $\delta h_{34}$ is also defined. We see that n (=4) > r (=3), and a suitable number $n_o = r = 3$ of observables exist; therefore, we have an adjustment problem.




Using the observation equations method, the problem can be solved by choosing three of the observables (m = r = 3) as also being unknown parameters and subsequently formulate the respective equations, as we did in the previous worked example. Alternatively, we can choose heights to describe the problem and the three unknown parameters, e.g., the heights of points (2, 3, 4), $\mathbf{x}^{\alpha} = \left[h_2\ h_3\ h_4\right]^{T}$, while a fixed height will be given to the fourth point, i.e., $h_1 = q_1 = \text{constant}$ , in order to define the height system/datum by minimal constraints (k=1). From three heights among five points and one known height any other height difference is well defined.

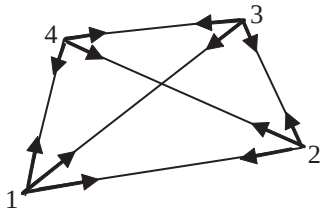
Including the constraint, $\delta h_{14} = q_{14}$, we easily notice that the parametric degree changes and becomes r = 2, as we need only two observables to obtain a solution, e.g., $(\delta h_{12}\ \delta h_{13})$. In the case of coordinates, we have only two known heights (m=2), because, from $\delta h_{14} = h_4 - h_1 = q_{14} = \text{constant}$ we get one more known height, $h_4 = h_1 + \delta h_{14} = h_1 + q_{14} = \text{constant}$, eliminating thus the additional constraint equation by fixing two heights. The adjustment algorithm, i.e. observation equation without constraints, can be normally applied keeping in mind that the applied two constraints are redundant and should be tested.

For the implementation of the condition method, we notice that without constraints, s = 5 – 3 = 2, and the two (independent) conditions are directly formed by any pair of equations among the loops (1-2-3-1), (2-3-4-2) and (1-2-4-3-1). Setting the constraint, we have, s = 5 – 2 = 3. The three (independent) conditions are directly formed by any pair of equations among the loops (1-2-3-1), (2-3-4-2) and (1-2-4-3-1) plus one condition reflecting the need that one given height difference is constant. The last condition could be formed by a loop-equation that includes the constant height difference, e.g., $\delta h_{12} + \delta h_{24} - q_{14} = 0$.

The mixed equation method can also be applied considering the above remarks.

## 6. Derived observations

Sometimes, instead of the original observations, derived observations are used. The advantage could be a simpler functional model but with a probably more complicated stochastic model and a demanding software. A simple example will clarify these concepts. Consider a measured plane (convex) quadrilateral (1-2-3-4) with observables the 12 horizontal directions, three from each vertex to the other ones.




A horizontal direction $\beta_{ij}$ is an angular quantity measured from a point (i) to another point (j) with respect to a reference direction $\theta_{io}$, e.g. the zero direction of the horizontal circle of the theodolite. The reference direction is an orientation parameter that must be defined, among other parameters, in the method of observation equations (or even in the mixed model); otherwise the concept of a direction is meaningless and the problem is not correctly described. A series of measured di-

rections refers to the directions from one point to other points, having all of them the same orientation direction. On the other hand, the difference between any two directions form the respective horizontal angle, $\omega_{ijk} = \beta_{ik} - \beta_{ij}$, where i → k is the right sighting and i → j the left sighting. From a horizontal angle and one of its directions the other direction is directly derived.

Directions as observable parameters define not only the shape of the quadrilateral but also the four orientation parameters, one at each point. The parametric degree r = 8 and this should be explained, through a construction scheme, as it is not so clear: Let's take an arbitrary length e.g., for the distance $d_{12}$ with an arbitrary orientation. From point 1 we use $\beta_{14}$ as an orientation direction and draw $\beta_{13}$, $\beta_{12}$. From point 2 take as reference $\beta_{21}$ and draw $\beta_{24}$, $\beta_{23}$ so that point 3 is determined by the intersection with $\beta_{13}$. The intersection of the respective directions defines points 3 and 4. Up to now we have used six directions so that the shape of the quadrilateral is well defined; any other angle can be derived using simple geometry. However, there is something missing, and this is the possibility to describe directions from points 3 and 4. This can be feasible by using a reference direction for point 3, e.g., $\beta_{32}$ and one for point 4, e.g., $\beta_{43}$. At this point the problem has been fully described by a least number of eight observables, hence the parametric degree r = 8 and the degrees of freedom f = 12 – 8 = 4. Other equivalent subsets of eight observables are possible.

Having n = 12 and r = 8, we must choose m = r = 8 unknown parameters in case the method of observation equations is used. One option is to take a suitable set of eight unknown parameters among the set of the twelve observables and form the function model accordingly (some geometric manipulations needed). The other option, almost always preferred, is to use point coordinates to describe the problem. In this case the reference system/datum should be also defined, e.g., by fixing coordinates ($x_1$=0, $y_1$=0) and ($x_2$=q, $y_2$=0), q being a constant (minimal constraints). Therefore, with four unknown parameters, being ($x_3$, $y_3$), ($x_4$, $y_4$) and with four unknown orientation parameters ($\theta_1$, $\theta_2$, $\theta_3$, $\theta_4$), eight unknown parameters are used in total. We remind that the observation equation of a horizontal direction is given by,

$$\beta_{ij} = \alpha_{ij} - \theta_i = \arctan\left(\frac{x_j - x_i}{y - y_i}\right) - \theta_i$$

where, $\alpha_{ij}$ is the plane azimuth (angle from the grid north to the direction i→j). Usually, in geodetic networks, $\theta_i$ coincides with the most left measured direction in a series from a station point; its approximate value is given by the azimuth computed from approximate coordinates.

In case of condition equations, the number of the independent equations should be

s = n – r = 12 – 8 = 4; three simple angular equations – one for each of three out of four formed triangles, and one 'side equation' which is a bit complicated and will not be given here. Because the condition equations use the directions as pairs of differences between them (angles), no orientation parameter is needed.

### *6.1 Using derived angles instead of the initial directions*

A different functional model makes use of the eight horizontal angles derived from the above twelve measured directions. In terms angles as observables, we need at least four angles (r = 4) to define the shape, e.g. the four angles at the end points of side 1–2 (any other angle is well defined). This is the case of the so-called derived observations (here derived angles) instead of the original measured observations (here directions). No orientation parameters exist in the respective eight angle observation equations, as expressed with respect to the four unknown parameters (m = r = 4), in the method of observation equations. In the method of condition equations, the number of conditions s = n – r = 8 – 4 = 4 should be formed, three angle equations and one side equation (like the directions scheme). The results between the adjustments, directions against angles, will be the same only if we will use the correct stochastic model for the derived angles; correlation exist between consecutive angles as their derivation uses a common direction. For instance, supposing uncorrelated directions, and taking $\sigma^2_{\beta_{12}} = \sigma^2_{\beta_{13}} = \sigma^2_{\beta_{14}} = \sigma^2_1$, the derived consecutive angles

$$\omega_{132}(=\beta_{12}-\beta_{13}) \quad \text{and} \quad \omega_{143}(=\beta_{13}-\beta_{14})$$

will have,

$$\sigma^2_{\omega_{132}} = \sigma^2_{\omega_{143}} = 2\sigma^2_1 \quad \text{and} \quad \sigma_{\omega_{132}\omega_{143}} = -\sigma^2_1,$$

as obtained by the error propagation law.

## 7. Best fitting problems with an example for a best fit circle

In problems, such as the best fit of a function to data points expressed by measured coordinates, e.g., line, plane curve, plane, circle, sphere, ellipsoid, some difficulties arise. Among them the selection of the functional model and method, the linearization process and the implementation of the algorithm are of primary importance. Next, we will try to point out some of these issues through a worked example of a best fit circle.

Given N data points $(x'_i, y'_i)$, i=1, 2, …N, measured along the circumference of a circle $(x_c, y_c, R)$, with $(x_c, y_c)$ the coordinates of the circle center and R its radius. The familiar circle equation is given by,

$$(x_i - x_c)^2 + (y_i - y_c)^2 - R^2 = 0 \tag{7.1}$$

Due to errors, all points do not satisfy the same equation; any triplet of data points determine a different circle. The goal is to find the best solution, i.e. the best circle under some fitting criterion. Generally, in fitting problems, data points should have a suitable distribution, here along a large arc and preferably along the whole circle, so that the circle is correctly defined avoiding probable inconsistencies and bad approximate values. However, small arcs should be avoided as considerable errors may grossly affect estimates of circle parameters.

Noting with $(x_i, y_i)$ the observables and $(v_{x_i}, v_{y_i})$ the errors, an efficient fitting criterion, is based on the minimization of squares of both errors. This can be expressed by the minimization of the geometric distances of the observed points from the best fit circle,

$$\sum d_i^2 = \sum \left( \sqrt{(x_i' - x_c)^2 + (y_i' - y_c)^2} - R \right)^2 \tag{7.2}$$

Such distances are reckoned along the directions from any data point to the center of the circle in case the errors/observations have the same uncertainty, a common situation in practice (geometric fit, orthogonal fit).

A criterion not so strict but competitive and often preferable to the previous one for its mathematical simplicity, is based on the minimization of the squares of the so-called algebraic distances, i.e.,

$$\sum d_i^2 = \sum \left( (x_i' - x_c)^2 + (y_i' - y_c)^2 - R^2 \right)^2 \tag{7.3}$$

In literature (see, e.g., Chernov 2010), there are many approaches and model modifications to both above model implementations, e.g., either using the model

$A(x_i^2 + y_i^2) + Bx_i + Cy_i + D = 0$ (under constraint on A, B, C, D)

or $z_i + Bx_i + Cy_i + D = 0$, from where, with the substitutions,

$$B = -2x_c, \quad C = -2y_c, \quad D = x_c^2 + y_c^2, \quad z_i = x_i^2 + y_i^2$$

we derive a linear model with respect to the unknown parameters with simple and fast computational algorithms. We do not intend here to go into details; instead we will point out how this problem could be faced with each one of the three basic adjustment methods and how one could face the effect of the linearization to obtain a best or at least a sufficient solution.

### *7.1 Using the method of observation equations*

The model $\mathbf{y}^\alpha = \mathbf{f}(\mathbf{x}^\alpha)$ for a circle, can be described by

$$x_i = x_c + R\cos\theta_i, \quad y_i = y_c + R\sin\theta_i \tag{7.4}$$

where, $\theta_i$ the angle from the x-axis anticlockwise to the point on the circumference. In total, we have m = r = 3 + N unknown parameters with n = 2N observations. For N > 3 so that n = 2N > r, an adjustment problem of observations is evident. The adjustment can be normally applied, noting that the normal equation matrix is (mxm) and that good starting values for the approximate coordinates are easily computed by suitable triplet/triplets of data points, for the center and the radius. In addition, from each data point an approximate angle θ is also obtained. Constraints on the unknown parameters may be included, such as the circle must have a known radius, the circle is passing from a known point or the circle is tangential to a known line.

A worthwhile notice for the parametric degree is that r (= N + 3) depends on the number of observations, something that do not happen in geodetic and surveying network adjustment problems. The same dependence of the parametric degree on data points is true in all best fitting problems, where given/measured coordinates are treated as observables. The same situation exists in the coordinate transformation problems between different systems, as it is the well-known similarity transformation, where coordinates of common points in both systems are affected by errors.

### *7.2 Using the method of mixed equations*

Eliminating all the $\theta_i$ parameters from (7.4), by adding the squares of coordinate differences, we will obtain a different functional model,

$$u_i = (x_i - x_c)^2 + (y_i - y_c)^2 - R^2 = 0 \tag{7.5}$$

that describe mixed equations $\mathbf{u}(\mathbf{y}^\alpha, \mathbf{x}^\alpha) = \mathbf{0}$ . In this model, we have m = 3 unknown parameters (m < r (= 3 + N)) and s = N mixed equations (one per point). For uncorrelated observations, a common practice, the advantage with this method is that the normal equation matrix 3x3 instead of (3+N)x(3+N) in the method of observation equations; certainly, there are some matrix-partitioned techniques reducing the computation effort for many data points in the method of observation equations. Overall, the computational ease makes the mixed equation method the preferred one. Constraints can also be included.

### *7.3 Using the method of condition equations*

Even though the condition method does not offer any serious advantage, for the shake of completeness, we note that the functional model, $\mathbf{g}(\mathbf{y}^\alpha) = \mathbf{0}$, can be derived by complete elimination of the unknown parameters. From the mixed equa-

tions (7.5), we take three of them and analytically express the coordinates of the circle and its radius in terms of the coordinates of the three used points. Next, we substitute these coordinates to the remaining (N–3) equations and in this form no unknown parameter except the observables is included. Matrix **M** to be inversed is (N–3)x(N–3) and nothing notable is gained. Moreover, equation complexity is obvious. Therefore, the mixed equation model is again the preferred one among the three basic adjustment methods.

### *7.4 The linearization effect*

LS adjustment follows normally the expansion of the initial functional model to a Taylor series to derive linear equations. In some problems, if an exact non-linear solution is possible, it should be preferred or at least compared to the obtained one through linearization, avoiding possible divergence of a correct solution.

Partial derivatives in the method of observation equations (matrix **A**), depend only on the approximate values of unknown parameters $\mathbf{x}^{\alpha}$. Starting with good initial values $\mathbf{x}^{o}$, some iterations are generally needed to achieve convergence and a correct solution.

In the method of condition equations where there are no (unknown) parameters the point of expansion in most of the classical treatments is around $\mathbf{y}^{b}$ and not $\mathbf{y}^{o}(=\mathbf{f}(\mathbf{x}^{o}))$. This fact imposes some questions for the correct solution. The algorithmic iterations in the adjustment process start with initial values –the classical approach- and then repeated accordingly, in a similar treatment as described in the following more general case of the mixed model. However, the linearization error (neglecting higher order terms) may affect the solution to a false point despite an achieved convergence. Moreover, matrix **B** acquires variability/stochastic properties under the theoretical infinite repetitions of the observations, as it depends on the observations. To avoid stochastic problems, since the experiment is usually carried out once, we suppose that $\mathbf{y}^{o}$ 'happens' to be identical with $\mathbf{y}^{b}$.

Similarly, in the mixed equations model, with

$$\mathbf{x}^{\alpha}=\begin{bmatrix} x_c & y_c & R \end{bmatrix}^T, \quad \mathbf{x}=\begin{bmatrix} \delta x_c & \delta y_c & \delta R \end{bmatrix}^T$$

$$\mathbf{y}^{\alpha}=\begin{bmatrix} x_1 y_1 & x_2 y_2 & \cdots & x_N y_N \end{bmatrix}^T, \ \mathbf{y}^{b}=\begin{bmatrix} x_1' y_1' & x_2' y_2' & \cdots & x_N' y_N' \end{bmatrix}^T$$

the expansion is traditionally developed around the point $(\mathbf{x}^{o},\mathbf{y}^{b})$ instead of $(\mathbf{x}^{o},\mathbf{y}^{o})$. Again, we have a similar risk as mentioned above, as matrices **B** and **A** depend also on the observations.

In case the initial equations of the functional model are linear or partly linear the algorithm allows for simplifications and similarities with the classical approaches.

We must face the linearization effects by either a non-linear solution, if possible, or by iterating properly the linear one, introducing some modification on the general adjustment model of Gauss-Helmert.

To show the effect of the model modification, we'll take the general/mixed model for the circle fit. The linearization, taking only the first order terms, has as follows:

$$u_i(x_c, y_c, R, x_i, y_i) = u_i(x_c^o, y_c^o, R^o, x_i^o, y_i^o) + \left.\frac{\partial u_i}{\partial x_c}\right|_o (x_c - x_c^o) + \left.\frac{\partial u_i}{\partial y_c}\right|_o (y_c - y_c^o) +$$

$$+ \left.\frac{\partial u_i}{\partial R}\right|_o (R - R^o) + \left.\frac{\partial u_i}{\partial x_i}\right|_o (x_i - x_i^o) + \left.\frac{\partial u_i}{\partial y_i}\right|_o (y_i - y_i^o) + \ldots = 0 \,. \qquad (7.6)$$

Noting that

$$x_c = x_c^o + \delta x_c \,, \quad y_c = y_c^o + \delta y_c \,, \quad R = R^o + \delta R \,,$$

and

$$x_i = x_i' - v_{x_i} \,, \quad y_i = y_i' - v_{y_i} \,, \quad x_i^o = x_i' - v_{x_i}^o \,, \quad y_i^o = y_i' - v_{y_i}^o \,,$$

from where,

$$x_i - x_i^o = (x_i' - v_{x_i}) - (x_i' - v_{x_i}^o) = -v_{x_i} + v_{x_i}^o \,, \quad y_i - y_i^o = -v_{y_i} + v_{y_i}^o \,,$$

we have,

$$u_i(x_c^o, y_c^o, R^o, x_i^o, y_i^o) = (x_i^o - x_c^o)^2 + (y_i^o - y_c^o)^2 - (R^o)^2 =$$

$$= (x_i' - v_{x_i}^o - x_c^o)^2 + (y'_i - v_{y_i}^o - y_c^o)^2 - (R^o)^2$$

$$\left.\frac{\partial u_i}{\partial x_c}\right|_o (x_c - x_c^o) + \left.\frac{\partial u_i}{\partial y_c}\right|_o (y_c - y_c^o) + \left.\frac{\partial u_i}{\partial R}\right|_o (R - R^o) =$$

$$= -2(x_i^o - x_c^o)\delta x_c - 2(y_i^o - y_c^o)\delta y_c - 2R^o \delta R$$

$$= -2\begin{bmatrix} (x_i' - v_{x_i}^o - x_c^o) & (y'_i - v_{y_i}^o - y_c^o) & R^o \end{bmatrix} \begin{bmatrix} \delta x_c \\ \delta y_c \\ \delta R \end{bmatrix} = \mathbf{A}_i \mathbf{x}_i$$

$$\left.\frac{\partial u_i}{\partial x_i}\right|_o (x_i - x_i^o) + \left.\frac{\partial u_i}{\partial y_i}\right|_o (y_i - y_i^o) = 2(x_i^o - x_c^o)(x_i - x_i^o) + 2(y_i^o - y_c^o)(y_i - y_i^o) =$$

$$= 2(x_i' - v_{x_i}^o - x_c^o)(x_i' - v_{x_i} - v_{x_i}^o) + 2(y'_i - v_{y_i}^o - y_c^o)(y'_i - v_{y_i} - v_{y_i}^o) =$$

$$= -2\begin{bmatrix} (x_i' - v_{x_i}^o - x_c^o) & (y'_i - v_{y_i}^o - y_c^o) \end{bmatrix} \begin{bmatrix} v_{x_i} \\ v_{y_i} \end{bmatrix} +$$

$$+2\left[(x'_i - v^o_{x_i} - x^o_c) \quad (y'_i - v^o_{y_i} - y^o_c)\right]\begin{bmatrix} x'_i - v^o_{x_i} \\ y'_i - v^o_{y_i} \end{bmatrix} = -\mathbf{B}_i\mathbf{v}_i + \mathbf{B}_i\begin{bmatrix} x'_i - v^o_{x_i} \\ y'_i - v^o_{y_i} \end{bmatrix}.$$

With

$$\mathbf{w}_i = (x'_i - v^o_{x_i} - x^o_c)^2 + (y'_i - v^o_{y_i} - y^o_c)^2 - (R^o)^2 + \mathbf{B}_i\begin{bmatrix} x'_i - v^o_{x_i} \\ y'_i - v^o_{y_i} \end{bmatrix}$$

we obtain the linear system,

$$\mathbf{w} + \mathbf{A}\mathbf{x} - \mathbf{B}\mathbf{v} = \mathbf{0} \tag{7.7}$$

A clarification on the iterating LS adjustment algorithm until convergence is achieved. We begin with the zeroth iteration or the first solution, by choosing $\mathbf{y}^o = \mathbf{y}^b$ from where $\mathbf{v}^o$ is zeroed and matrices **A**, **B** and **w** are functions of $(\mathbf{y}^b, \mathbf{x}^o)$ as happens in the classical treatment. Therefore, the LS solution gives $(\hat{\mathbf{v}}^o, \hat{\mathbf{x}}^o)$ from where $(\hat{\mathbf{y}}^o = \mathbf{y}^b - \hat{\mathbf{v}}^o, \hat{\mathbf{x}}^{\alpha(o)} = \mathbf{x}^o + \hat{\mathbf{x}}^o)$. With the new approximate values $(\hat{\mathbf{y}}^o, \hat{\mathbf{x}}^{\alpha(o)})$, the first iteration or 2nd solution gives $(\hat{\mathbf{v}}^{o(1)}, \hat{\mathbf{x}}^{o(1)})$ and $(\hat{\mathbf{y}}^{o(1)} = \mathbf{y}^b - \hat{\mathbf{v}}^{o(1)}, \hat{\mathbf{x}}^{\alpha(1)} = \hat{\mathbf{x}}^{\alpha(o)} + \hat{\mathbf{x}}^{o(1)})$. Similarly, the $n^{th}$ iteration (n+1 solution) results in $(\hat{\mathbf{v}}^{o(n)}, \hat{\mathbf{x}}^{o(n)})$ and $(\hat{\mathbf{y}}^{o(n)} = \mathbf{y}^b - \hat{\mathbf{v}}^{o(n)}, \hat{\mathbf{x}}^{\alpha(n)} = \hat{\mathbf{x}}^{\alpha(n-1)} + \hat{\mathbf{x}}^{o(n)})$. Setting a converge limit on the (absolute) differences between two successive estimations, i.e., $(|\hat{\mathbf{v}}^{o(n)} - \hat{\mathbf{v}}^{o(n-1)}| \le \mathbf{e}_\mathbf{v}, |\hat{\mathbf{x}}^{o(n)} - \hat{\mathbf{x}}^{o(n-1)}| \le \mathbf{e}_\mathbf{x})$, the final LS-estimates are obtained. To prevent many iterations and probably a diverse solution to a local minimum, starting values $\mathbf{x}^o$ should be computed as close as possible to the correct estimates, usually by a subset of observations.

Details for the previous scheme of iterations are extensively discussed by Pope (1972) where some pitfalls are also underlined. Some other examples and approaches are given, e.g., in Schaffrin and Wieser (2008), Pan et al. (2015). A simplified mathematical and algorithmic implementation of the above Modified Mixed Model (MMM) in practical fitting problems is presented in Fotiou (2017a, 2017b).

## 8. Concluding remarks

A means to study the real world and to obtain quantitative determination for various physical problems is provided by the analysis of observations in relation to a mathematical model that reflects the unknown reality. For that purpose, the adjustment of observations consists an indispensable part in geodetic sciences, geomatics and other related fields.

Central LS adjustment topics, about the correct choice of the adjustment model/method and the efficient algorithmic implementation for parameter esti-

mates and error control, need a closer look as the results may be crucial in many practical applications. A better understanding of the involved concepts, models and methods can be gained through simple worked examples presented in this paper, building step by step acquaintance and developing a good and effective theoretical framework.

## References


Chernov, N. (2010). *Circular and Linear Regression: Fitting Circles and Lines by Least Squares*. CRC Press, A Chapman &Hall Book.

Dermanis, A. (1986). Adjustment of Observations and Estimation Theory, Vol. I *(in greek)*. Ziti editions, Thessaloniki.

Dermanis, A. (1987). Adjustment of Observations and Estimation Theory, Vol. II *(in greek)*. Ziti editions, Thessaloniki.

Dermanis, A. and A. Fotiou (1992). *Methods and Application of Observation Adjustment (in greek)*. Ziti Publisher, Thessaloniki.

Fotiou, A. (2007). *Geometric Geodesy-Theory and Practice*. Ziti editions, Thessaloniki (in greek).

Fotiou, A. (2017a). *Line of Best Fit by a Least Squares Modified Mixed Model.* Volume '*Living with GIS'* in honour of the memory of Prof. I. Paraschakis, School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, 57-80.

Fotiou, A. (2017b). *Computationally Efficient Methods and Solutions with Least Squares Similarity Transformation Models.* Volume '*Living with GIS'* in honour of the memory of Prof. I. Paraschakis, School of Rural and Surveying Engineering, Aristotle University of Thessaloniki, 107-128.

Koch, K.R. (1987). Parameter Estimation and Hypothesis Testing in Linear Models. Springer-Verlang.

Mikhail, E.M. and F. Ackermann (1976). *Observations and Least Squares*. IEP-A Dun-Donnelley Publisher.

Pan, G., Y. Zhou, H. Sun, and W. Guo (2015). Linear observation based total least squares. *Survey Review*, 47:340, 18-27.

Pope, A. J. (1972). *Some pitfalls to be avoided in the iterative adjustment of nonlinear problems*. In Proc.: 38th Ann. Meeting, American Society of Photogrammetry, Washington, D.C., 449-473.

Rossikopoulos, D. (1999). Surveying Networks and Computations (in greek). Ziti editions.

Schaffrin, B. and A. Wieser (2008). On weighted total least-squares adjustment for linear regression. *J Geod.*, 82, 415– 421.

Sneew, N., F. Krumm, M. Roth (2015). *Adjustment Theory*. Lecture Notes, Rev. 4.43, Geod. Inst. Univ. Stuttgart.

Wells, D.E. and E.J. Krakiwsky (1971). The Method of Least Squares. Lecture Notes, Department of Geodesy and Geomatics Engineering, U.N.B.