Line of Best Fit by a Least Squares Modified Mixed Model

A. Fotiou

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

Summary: Least squares models for line fitting are presented and discussed. Emphasis is given on simplicity, clarity and ease of implemented accurate algorithms while some references to geodetic and surveying applications are mentioned. Analytical formulation is obtained using suitable equation forms, especially when data points have different precision, as happens with the so called EIV or total least square models. Special cases are investigated and measures of precision are also given. A new iterative algorithm is developed, called here the Modified Mixed Model, using the normal line form and based on a slight modification of the general adjustment model. By means of well-known data in the literature, the algorithm is tested and the results show full compatibility compared to other existed models and computational techniques.

Key words: line fitting, best fit, least squares, EIV models, total least squares, modified mixed model.

1. Introduction

The well-known problem of fitting a straight line to scattered data has been extensively studied and concentrated a great amount of literature over a century. The emphasis is giving according to the specific issue, experiment or application in various scientific fields, in conjunction with the functional and stochastic model of the related parameters and variables.

Certainly the presented models in this paper are suitable in almost all scientific fields. However, we underline some of the geodetic and surveying engineering applications, such as cadastral, GIS and map related works, where very often one is facing the problem to constrain points, expressed by coordinates, intended to be on a straight line representing e.g., a border or a street axis (e.g., Harvey 2009).

It is common in geodetic sciences to express points by orthogonal (Cartesian) coordinates, for their benefits in mathematical modeling and data processing. Grid or map coordinates have been recorded in catalogues and/or data bases, referred to maps, deposited plans, land redistributions and similar products or studies. Crucial

points are e.g., boundaries and marks of lots, private and public land parcels since they describe the geometry of an area and other important features. Such collinear points in theory (the plan as model) have to be coordinated and implemented or controlled in practice any time needed.

In general, the determination of point coordinates needs suitable measurements, such as angles, distances, azimuths, height differences, GPS/GNSS data and a reference frame that is accessed through included control points. In addition, coordinates can be produced by digitizing or scanning maps, images or by any other specific measuring procedure. The derived coordinates are subject to errors and characterized by uncertainties due to the unavoidable random errors, possibly existed systematic errors and even due to the less frequently outliers or mistakes. Therefore, their quality or accuracy has to be described. Usually, the quality is given in terms of precision or repeatability, i.e. by standard deviations, varying from a few mm to a few cm in the everyday engineering practice. It is acceptable and realistic in most of the cases to realize point coordinates as uncorrelated measurements, either directly or indirectly obtained, although correlations could be also included, if for instance, a very high accuracy is demanded.

Stand alone or special GIS software take care of best fitting algorithms. Least-squares criterion and/or robust techniques are usually applied, the latter mostly used in limiting the influence of outliers (e.g., Kiryati and Bruckstein 1992).

In this paper, we will start with a review of the standard least squares approaches based on the known Gauss-Markov model (GMM) for comparison reasons and completeness. Next, we will proceed on with a slight modification of the general or mixed least-squares model, called here MMM (Modified Mixed Model). On the other hand, the choice of the line equation form will be examined and some remarks against other existed models and techniques will be quoted as well. Though the idea of MMM is not new, as it is based on the known Gauss-Helmert Model (e.g., Jefferys 1980, Lybanon 1984, Dermanis 1987, Sneew et al. 2015), the MMM will be presented here by an easy and understandable layout. The MMM will be applied also using the normal line form, an algorithmic approach that to the author's knowledge is presented for first time.

The equation of a line can take different forms. Among them the most representative being the slope-intercept $y^{\alpha} = ax^{\alpha} + b$ and the general or standard form $Ax^{\alpha} + By^{\alpha} = C$ with the latter expressed also in the normal equation form $x^{\alpha}\cos t + y^{\alpha}\sin t = r$ and inversely (Fig.1), e.g. A=cost, B=sint and C=r.

Estimating line parameters from data points that are subject to errors some limitations and constraints have to be taken into account as far as the suitability and effectiveness of the functional model. A drawback of the slope-intercept equation is that it is not possible to describe vertical lines (x=constant) or when the intercept takes big values as the line tends to the vertical position, with uncontrollably com-

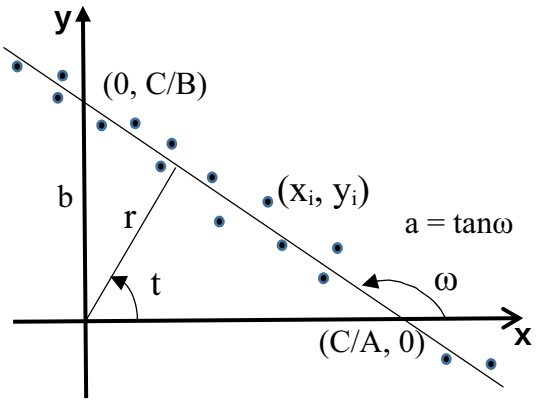

Fig. 1. *Data points and line parameters*

putational problems. An alternative to such dangerous cases would be the use of the equation $x^{\alpha} = cy^{\alpha} + d$, something reasonable in a few isolated cases. In contrast, when many best fit lines have to be determined as in land surveying projects, e.g. the determination of street axes in a city plan, a considerable number of them may be parallel or almost parallel to the y-axis. Due to homogeneity reasons the general line equation, either directly or through the normal form determined should be the preferable and right choice.

In any functional model, we distinguish the unknown (but fixed) line parameters, e.g., (a, b), (A, B, C) or (t, r) and the unknown (but fixed) variables x^{α}, y^{α} . All of them represent unknown (but fixed) true values that have to be properly estimated. Each one of x^{α}, y^{α} or both of them, can be treated as random variables, that is as observables, depending on the specific problem and the nature of the variables.

Given a number of data points (x_1, y_1), (x_2, y_2), ..., (x_n, y_n), considered in general as observations, the functional relation of the fitting line has to be determined. This demands for the estimation of the line parameters, e. g. $\hat{a}, \hat{b}$, where the adjusted points $(\hat{x}_i^{\alpha}, \hat{y}_i^{\alpha})$ are lying on the best fit line, i.e., $\hat{y}_i^{\alpha} = \hat{a}\hat{x}_i^{\alpha} + \hat{b}$. We underline that the stochastic model of the observational errors is considered the same as that of their associated observations.

Given a functional Least-Squares (LS) model the estimation of parameters can be obtained without the need of a priori statistical hypothesis for the behavior of errors. Naturally we expect that errors have small size and tend to cancel out in many repetitions of the observations. However, it is significant to account for their contribution on the estimated parameters and therefore some suitable hypotheses are needed. This is possible if we assume that the errors are random (stochastic) variables with zero mean for the infinite sample space (Expectation) and have a known or approximately known covariance matrix. With this stochastic model the applica-

tion of the error propagation law, called better the propagation law of covariances, is enabled and the estimation of covariance matrices of the estimated or adjusted parameters is achieved. In addition, if the random errors follow the normal (Gauss) distribution, the parameter estimates (not their estimated variances) are identical to those derived by the Maximum Likelihood Method.

2. Line fitting with one observable parameter

Considering one observable parameter, e.g. y^{α}, the 'slope-intercept' form $y^{\alpha} = ax^{\alpha} + b$ is used, where the slope is expressed by the parameter a ($a = \tan\omega$, ω reckoning anticlockwise from x-axis to the line, Fig.1) and the y-intercept by the parameter b. In this case, y_i-values are the corresponding (uncorrelated) observations, and $x_i = x_i^{\alpha}$ are constants (error free, fixed, absolutely known). This model is the well-known standard regression model of y^{α} (dependent variable) on x (independent variable) or the standard LS model. Exchanging the role of variables, regression of 'x^{α} on y' can easily be obtained with obviously different results.

Using the LS method of observation equations (method of parameters), the initial observation equations (functional model) are linear with respect to the unknown parameters (a, b), i.e.,

$$y_i^{\alpha} = ax_i + b = [x_i \quad 1]\begin{bmatrix} a \\ b \end{bmatrix}. \tag{2.1}$$

Taking the conventional for the sign relation $y_i = y_i^{\alpha} + v_i$, the linear observation equations are,

$$y_i = ax_i + b + v_i = [x_i \quad 1]\begin{bmatrix} a \\ b \end{bmatrix} + v_i. \tag{2.2}$$

In matrix notation,

$$\mathbf{y} = \mathbf{A}\mathbf{z}^{\alpha} + \mathbf{v} \tag{2.3}$$

or in matrix layout

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \tag{2.4}$$

where the errors v_i represent the so called 'vertical' distances along the y-axis, from any observed point to the line.

2.1. Equal measurement precision with one observable parameter

For n data points and assuming y_i of equal precision, i.e. $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$, their covariance matrix is $\mathbf{C} = \sigma^2 Q = \sigma^2 \mathbf{I}$ ($\mathbf{Q} = \mathbf{I}$) with σ^2 known or unknown. In both cases we can use the weight matrix $\mathbf{P} = \mathbf{Q}^{-1} = \mathbf{I}$. In case of an unknown σ^2 (equal weights with unknown precision) an unbiased estimate has to be determined, needed for the estimation of covariance matrices of the parameter estimates. Having thus equal weights for each observation y_i, i.e., $p = 1$, the LS solution is obtained by applying the criterion,

$$\sum pv_i^2 = \sum v_i^2 = \sum (y_i^\alpha - ax_i - b)^2 = \min \ . \tag{2.5}$$

The well-known LS algorithm of the method of observation equations (e.g., Dermanis and Fotiou 1992), results in the following best linear unbiased estimations (maximum precision) either σ^2 is known or unknown,

$$\hat{a} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{m_{xy}}{m_x^2} = \frac{s_{xy}}{s_{xx}} \tag{2.6}$$

$$\hat{b} = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \bar{y} - \frac{m_{xy}}{m_x^2}\bar{x} = \bar{y} - \frac{s_{xy}}{s_{xx}}\bar{x} = \bar{y} - \hat{a}\bar{x} \tag{2.7}$$

$$\hat{v}_i = y_i - (\hat{a}x_i + \hat{b}), \qquad \hat{y}_i^\alpha = y_i - \hat{v}_i = \hat{a}x_i + \hat{b} \ . \tag{2.8}$$

Note that the best fit line passes through the centroid ($\bar{x}$, $\bar{y}$), a point on the line. In case σ^2 is unknown an unbiased estimate, given by

$$\hat{\sigma}^2 = \frac{\sum \hat{v}_i^2}{n-2} \tag{2.9}$$

should be used instead of the unknown σ^2 and moreover it could be statistically tested against a priori σ_o^2 . The precision of the estimated parameters is given by

$$\hat{\sigma}_{\hat{a}}^2 = \sigma^2 \frac{1}{nm_x^2} = \sigma^2 \frac{1}{s_{xx}} \tag{2.10}$$

$$\hat{\sigma}_{\hat{b}}^2 = \sigma^2 \frac{m_x^2 + \bar{x}^2}{nm_x^2} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right) \tag{2.11}$$

$$\hat{\sigma}_{\hat{a}\hat{b}} = \sigma^2 \left(-\frac{\bar{x}}{nm_x^2}\right) = \sigma^2 \left(-\frac{\bar{x}}{s_{xx}}\right) \tag{2.12}$$

where, the mean values $\bar{x}$, $\bar{y}$, the dispersions m_x^2, m_y^2 and the scatter quantities s_{xx}, s_{xy}, are given by

$$\bar{x} = \frac{1}{n}\sum x_i\,, \quad \bar{y} = \frac{1}{n}\sum y_i \tag{2.13}$$

$$m_x^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2\,, \qquad s_{xx} = n m_x^2 = \sum (x_i - \bar{x})^2 \tag{2.14}$$

$$m_{xy} = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}\,,$$

$$s_{xy} = n m_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})\,. \tag{2.15}$$

A computational simplification results from the reduction of measurements to their centroid $(\bar{x}, \bar{y})$. Instead of the initial data (x_i, y_i) we can use the reduced values $x_i' = x_i - \bar{x}, y_i' = y_i - \bar{y}$, having thus $\sum x_i' = 0, \sum y_i' = 0$, and for the new centroid with $\bar{x} = \bar{y} = 0$, the previous expressions are accordingly modified, e.g., $\hat{b} = 0$.

2.2. Different measurement precision with one observable parameter

Considering now different precisions for each observation y_i, usually in the form $\sigma_i^2 = \sigma^2 q_i^2$, the diagonal covariance matrix is written as $\mathbf{C} = \sigma^2 \mathbf{Q}$, where $\mathbf{Q} = \mathrm{diag}(q_1^2, q_2^2, ..., q_n^2)$ is a known matrix and σ^2 a reference variance, known or unknown as previously stated. Therefore, the weight matrix **P** is

$$\mathbf{P} = \mathbf{Q}^{-1} = \mathrm{diag}(1/q_1^2, 1/q_2^2, ..., 1/q_n^2) = \mathrm{diag}(p_1, p_2, ..., p_n)\,.$$

In case of an unknown variance factor, a priori variance σ_o^2 could be (statistically) tested against the posteriori $\hat{\sigma}^2$ which is used instead of an unknown σ^2.

Applying the LS criterion,

$$\sum p_i v_i^2 = \sum p_i (y_i^{\alpha} - a x_i - b)^2 = \min\,. \tag{2.16}$$

It is obvious in (2.16) that measurements with less uncertainties (greater precision) or greater weights have greater influence on the estimations of parameters. For the sake of completeness, the corresponding estimations (as above) are given by,

$$\hat{a} = \frac{(\sum p_i)(\sum p_i x_i y_i) - (\sum p_i x_i)(\sum p_i y_i)}{(\sum p_i)(\sum p_i x_i^2) - \left(\sum p_i x_i\right)^2} = \frac{m_{pxy}}{m_{px}^2} = \frac{s_{pxy}}{s_{pxx}} \tag{2.17}$$

$$\hat{b} = \frac{(\sum p_i y_i)(\sum p_i x_i^2) - (\sum p_i x_i)(\sum p_i x_i y_i)}{(\sum p_i)(\sum p_i x_i^2) - \left(\sum p_i x_i\right)^2} = \bar{y} - \frac{m_{pxy}}{m_{px}^2}\bar{x} =$$

$$= \bar{y} - \frac{s_{pxy}}{s_{pxx}}\bar{x} = \bar{y} - \hat{a}\bar{x} \tag{2.18}$$

noting that weights have been incorporated, i.e.,

$$\bar{x} = \frac{1}{n}\sum p_i x_i \,, \quad \bar{y} = \frac{1}{n}\sum p_i y_i \tag{2.19}$$

$$m_{px}^2 = \frac{1}{\Sigma p_i}\sum p_i (x_i - \bar{x})^2 = \frac{1}{\Sigma p_i}\sum p_i x_i^2 - \bar{x}^2 \,,$$

$$s_{pxx} = \Sigma p_i \, m_{px}^2 = \sum p_i (x_i - \bar{x})^2 \tag{2.20}$$

$$m_{pxy} = \frac{1}{\Sigma p_i}\sum p_i (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{\Sigma p_i}\sum p_i x_i y_i - \bar{x}\,\bar{y} \tag{2.21}$$

$$s_{pxy} = \Sigma p_i \, m_{pxy} = \sum p_i (x_i - \bar{x})(y_i - \bar{y}) \,. \tag{2.22}$$

Moreover, precision estimates are given by

$$\hat{\sigma}^2 = \frac{\sum p_i \hat{v}_i^2}{n-2} \tag{2.23}$$

$$\hat{\sigma}_{\hat{a}}^2 = \sigma^2 \frac{1}{(\sum p_i) m_{px}^2} = \sigma^2 \frac{1}{s_{pxx}} \tag{2.24}$$

$$\hat{\sigma}_{\hat{b}}^2 = \sigma^2 \frac{m_{px}^2 + \bar{x}^2}{(\sum p_i) m_{px}^2} = \sigma^2 \left(\frac{1}{\sum p_i} + \frac{\bar{x}^2}{s_{pxx}}\right) \tag{2.25}$$

$$\hat{\sigma}_{\hat{a}\hat{b}} = \sigma^2 \left(-\frac{\bar{x}}{\sum p_i m_x^2}\right) = \sigma^2 \left(-\frac{\bar{x}}{s_{pxx}}\right) \tag{2.26}$$

where σ^2 is the known or the posteriori variance.

3. Two observable parameters with the slope-intercept form

This is the case where both x^α, y^α are observable parameters and therefore (x_i, y_i) are the measurements subjected to errors. Next we will see that this model is considerably more complex and has been too much analyzed and discussed, having moreover some drawbacks.

Models that account for errors on all observable parameters are also called EIV (Errors In Variables) models, usually in the statistical community or TLS models in the geodetic community, either being linear or non-linear. The solutions differ on the formulation of the target function that is minimized, the linearization process

and the computational methods and techniques. In geodetic terminology, let say, we have to do with an extension or modification of the general adjustment model of Gauss Helmert, called here MMM, as already stated above.

3.1. The modified mixed model with the slope-intercept form

The functional model is now not linear though looks so at a first sight. According to the LS method of mixed equations, also called the generalized LS method or total least squares, for any observed point a mixed equation is written, i.e.,

$$F_i = a x_i^{\alpha} + b - y_i^{\alpha} = 0, \qquad i = 1,\ldots, n. \tag{3.1}$$

Following the linearization process, a Taylor series expansion is considered, up to first order terms, around the approximate values a^o, b^o for the unknown parameters a, b and x_i^o, y_i^o (against the usual point x_i, y_i) for the unknown observable parameters as well. In this way, the linear system is formed as

$$F_i(a,b,x_i^{\alpha},y_i^{\alpha}) = F_i(a^o,b^o,x_i^o,y_i^o) + \left.\frac{\partial F_i}{\partial a}\right|_o (a-a^o) + \left.\frac{\partial F_i}{\partial b}\right|_o (b-b^o) +$$

$$+ \left.\frac{\partial F_i}{\partial x_i^{\alpha}}\right|_o (x_i^{\alpha} - x_i^o) + \left.\frac{\partial F_i}{\partial y_i^{\alpha}}\right|_o (y_i^{\alpha} - y_i^o) + \ldots = 0. \tag{3.2}$$

With $a = a^o + \delta a$, $b = b^o + \delta b$, $x_i^{\alpha} = x_i - v_{x_i}$, $y_i^{\alpha} = y_i - v_{y_i}$, $x_i^o = x_i - v_{x_i}^o$ and $y_i^o = y_i - v_{y_i}^o$, equation (3.2) becomes

$$[a^o(x_i - v_{x_i}^o) + b^o - (y_i - v_{y_i}^o)] + [(x_i - v_{x_i}^o) \quad 1]\begin{bmatrix}\delta a\\ \delta b\end{bmatrix} - [a^o - 1]\begin{bmatrix}v_{x_i}^o\\ v_{y_i}^o\end{bmatrix} +$$

$$+ [a^o - 1]\begin{bmatrix}x_i^{\alpha} - x_i^o\\ y_i^{\alpha} - y_i^o\end{bmatrix} = 0. \tag{3.3}$$

Noting that,

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

equation (3.3) results in

$$[a^o x_i + b^o - y_i] + [(x_i - v_{x_i}^o) \quad 1]\begin{bmatrix}\delta a\\ \delta b\end{bmatrix} - [a^o - 1]\begin{bmatrix}v_{x_i}\\ v_{y_i}\end{bmatrix} = 0 \tag{3.4}$$

We would arrive at the same expression (3.4) by means of an indirect linearization process, i.e., developing $a x_i^{\alpha} + b - y_i^{\alpha} = (a^o + \delta a)(x_i - v_{x_i}) + (b^o + \delta b) - (y_i - v_{y_i}) =$

$= (a^o + \delta a)(x_i^o + v_{x_i}^o - v_{x_i}) + (b^o + \delta b) - (y_i^o + v_{y_i}^o - v_{y_i}) = 0$, and making some algebraic arrangements as well as ignoring 2nd order terms.
From (3.4), the structure of the linear system is

$$\begin{bmatrix} a^o x_1 + b^o - y_1 \\ a^o x_2 + b^o - y_2 \\ \vdots \\ a^o x_n + b^o - y_n \end{bmatrix} + \begin{bmatrix} (x_1 - v_{x_1}^o) & 1 \\ (x_2 - v_{x_2}^o) & 1 \\ \vdots & \vdots \\ (x_n - v_{x_n}^o) & 1 \end{bmatrix} \begin{bmatrix} \delta a \\ \delta b \end{bmatrix} - \begin{bmatrix} a^o & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & a^o & \cdots & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a^o & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} v_{x_1} \\ v_{x_2} \\ \vdots \\ v_{x_n} \\ v_{y_1} \\ v_{y_2} \\ \vdots \\ v_{y_n} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.5}$$

or in brief,

$$\mathbf{w} + \mathbf{A}\mathbf{z} - \mathbf{B}\mathbf{v} = \mathbf{0}. \tag{3.6}$$

From (3.5) it is obvious that the design matrix **A** is partially affected by the errors of x_i while matrix **B,** as happens in this form of functional model, is unaffected of any observational error. Next, the LS solution is carried out through an iterative scheme, where at the same iteration the approximate values $a^o, b^o, v_{x_i}^o$ are properly updated as explained in below.

The LS solution is obtained by means of the traditional method of Lagrange multipliers, where an objective function is minimized under equality constraints, here,

$$\sum \left(p_{x_i} v_{x_i}^2 + p_{y_i} v_{y_i}^2 \right) = \mathbf{v}^T \mathbf{P} \mathbf{v} = \min, \quad \text{under} \quad F_i = a x_i^\alpha + b - y_i^\alpha = 0 \tag{3.7}$$

Remember that the weight matrix **P** is considered diagonal although the solution can go along with a full matrix.

In order to facilitate a simple computer program, a few details for the implementation of the MMM are given. Taking the (diagonal) covariance matrix

$$\mathbf{C} = \mathbf{P}^{-1} = \begin{bmatrix} \mathbf{C_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_y} \end{bmatrix} = \begin{bmatrix} \sigma^2 \mathbf{Q_x} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{Q}_y \end{bmatrix} \tag{3.8}$$

with the reference variance known or unknown, the weight matrix is,

$$\mathbf{P} = \begin{bmatrix} \mathbf{Q_x^{-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q_y^{-1}} \end{bmatrix} \tag{3.9}$$

reminding that the (posteriori) estimated variance $\hat{\sigma}^2$ is used instead of an unknown one. In order to formulate the normal equation system (**Nz = -u**), we go on

through matrix **B** and **M**, i.e.,

$$\mathbf{B} = \left[a^o \mathbf{I}_n \;\; -\mathbf{I}_n \right], \quad \mathbf{M} = \mathbf{B}\mathbf{P}^{-1}\mathbf{B}^T = \left[(a^o)^2 \mathbf{C_x} + \mathbf{C_y} \right] = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_n) \quad (3.10)$$

$$\mu_i = (a^o)^2 \sigma_{x_i}^2 + \sigma_{y_i}^2 = \frac{(a^o)^2}{p_{x_i}} + \frac{1}{p_{y_i}} \quad (3.11)$$

$$\mathbf{M}^{-1} = \mathrm{diag}(1/\mu_1, 1/\mu_2, ..., 1/\mu_n) = \mathrm{diag}(p_1, p_2, ..., p_n) \;, \qquad p_i = \frac{1}{\mu_i} \quad (3.12)$$

$$\mathbf{N} = \mathbf{A}^T\mathbf{M}^{-1}\mathbf{A} = \begin{bmatrix} \sum p_i \left(x_i - v_{x_i}^o \right)^2 & \sum p_i \left(x_i - v_{x_i}^o \right) \\ \sum p_i \left(x_i - v_{x_i}^o \right) & \sum p_i \end{bmatrix} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix}, \; (N_{21}=N_{12}) \quad (3.13)$$

$$\mathbf{u} = \mathbf{A}^T\mathbf{M}^{-1}\mathbf{w} = \begin{bmatrix} \sum p_i w_i \left(x_i - v_{x_i}^o \right) \\ \sum p_i w_i \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad w_i = a^o x_i + b^o - y_i \,. \quad (3.14)$$

Corrections of parameter estimations are then given by,

$$\hat{\mathbf{z}} = \begin{bmatrix} \delta\hat{a} \\ \delta\hat{b} \end{bmatrix} = -\mathbf{N}^{-1}\mathbf{u} = -\frac{1}{\det(\mathbf{N})} \begin{bmatrix} N_{22} & -N_{12} \\ -N_{12} & N_{11} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \\ = -\frac{1}{\det(\mathbf{N})} \begin{bmatrix} N_{22}u_1 - N_{12}u_2 \\ -N_{12}u_1 + N_{11}u_2 \end{bmatrix} \quad (3.15)$$

where, $\det(\mathbf{N}) = N_{11}\,N_{22} - (N_{12})^2 \neq 0$, and the parameter estimations by

$$\hat{\mathbf{z}}^a = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \mathbf{z}^o + \hat{\mathbf{z}} = \begin{bmatrix} a^o + \delta\hat{a} \\ b^o + \delta\hat{b} \end{bmatrix} . \quad (3.16)$$

The estimation of errors and observable parameters (adjusted observations) are,

$$\hat{\mathbf{v}} = \begin{bmatrix} \hat{\mathbf{v}}_x \\ \hat{\mathbf{v}}_y \end{bmatrix} = \mathbf{P}^{-1}\mathbf{B}^T\mathbf{M}^{-1}(\mathbf{w} + \mathbf{A}\hat{\mathbf{z}}) \quad (3.17)$$

or analytically,

$$s_i = (\mathbf{w} + \mathbf{A}\hat{\mathbf{z}})_i = a^o x_i + b^o - y_i + (x_i - v_{x_i}^o)\delta a + \delta b \quad (3.18)$$

$$\hat{v}_{x_i} = (a^o \sigma_{x_i}^2 \, p_i)\, s_i, \qquad \hat{v}_{y_i} = -(\sigma_{y_i}^2 \, p_i)\, s_i \quad (3.19)$$

$$\hat{x}_i^a = x_i - \hat{v}_{x_i}, \qquad \hat{y}_i^a = y_i - \hat{v}_{y_i} \;. \quad (3.20)$$

Also, the estimation of an unknown variance factor, i.e. the posteriori variance, is

$$\hat{\sigma}^2 = \frac{\sum\left(p_{x_i}\hat{v}_{x_i}^2 + p_{y_i}\hat{v}_{y_i}^2\right)}{n-2} \ . \tag{3.21}$$

Lastly, the covariance matrices for different estimation can be derived applying the propagation law of covariances. Among them, of considerable importance are the covariance matrix of parameter estimations $\hat{a}, \hat{b}$, being the elements of the inverse matrix $\mathbf{N}^{-1}$ given above, that is,

$$\hat{\sigma}_{\hat{a}}^2 = \sigma^2 \frac{N_{22}}{\det(\mathbf{N})}, \ \hat{\sigma}_{\hat{b}}^2 = \sigma^2 \frac{N_{11}}{\det(\mathbf{N})}, \ \hat{\sigma}_{\hat{a}\hat{b}} = -\sigma^2 \frac{N_{12}}{\det(\mathbf{N})} \ . \tag{3.22}$$

The above estimations are Best Linear Unbiased Estimations (BLUE) according to the LS principles. For a statistical evaluation of the model, e.g. testing for outliers, the probability density function for the errors should be known. As usually, this is the Normal (Gauss) distribution, accepted mainly for mathematical simplicity against others and because it is a reasonable choice. Admitting the normal distribution, the estimations derived also by the Maximum Likelihood method are identical with the LS estimates.

An approximation to the above rigorous MMM solution starts from Taylor series expansion around the point $F_i(a^o, b^o, x_i, y_i)$ instead of $F_i(a^o, b^o, x_i^o, y_i^o)$, forming thus matrix **A** depending on observations x_i and not on the approximate observations x_i^o as stated previously.

This approximation is the usual application of the so called general/mixed model (Gauss-Helmert) in Geodesy and Surveying (e.g, Mikhail and Ackermann 1976, Dermanis 1987, Ghilani 1997), giving satisfactory results only if the approximate values a^o, b^o are sufficiently close to the true ones and that the observational errors have small size.

A drawback of this traditional mixed model is that in case of equal precision for the observed points the derived estimates of parameters (a, b) cannot be updated and improved, as it can easily be shown that they are independent of their approximate values. Actually they are identical with those derived from the simple/standard regression model, where only y_i -observations are subjected to errors (e.g., Dermanis and Fotiou 1992).

3.2. The iterative process of MMM with the slope-intercept form

The iterations for obtaining the MMM solution start with initial approximate values a^o, b^o computed by a suitable way, e.g. by the standard LS solution presented in chapter 2 or even by the orthogonal solution of the next chapter. Moreover, the initial values for the approximate errors in matrix **A** are taken as zero ($v_{x_i}^o = 0$). In

this way, approximate weights from (3.11) are computed and estimations $\delta\hat{a}^{(1)}$, $\delta\hat{b}^{(1)}$ given by (3.15), $\hat{a}^{(1)}$, $\hat{b}^{(1)}$ by (3.16) and $\hat{v}_{x_i}^{(1)}$ by (3.19) are obtained. Now, with these better estimates, as the new approximate values, the second iteration starts resulting in $\hat{a}^{(2)}$, $\hat{b}^{(2)}$, $\hat{v}_{x_i}^{(2)}$, and so on until the fulfillment of a convergence criterion. It should be noted that within the same iteration, though it is not necessary, other estimates, such as $\hat{x}_i^{a(1)} = x_i - \hat{v}_{x_i}^{(1)}$, $\hat{y}_i^{a(1)} = y_i - \hat{v}_{y_i}^{(1)}$, given by (3.20) and $\hat{\sigma}^{2(1)}$ given by (3.21), could be also updated instead of the end of the whole process, resulting in their final adjusted values.

A convergence criterion for the successive absolute differences is set, usually for the updated parameters (a, b) or even for all updated parameters. The tolerance depends on the degree of closeness of the initial approximate values to their true values and on the level of accuracy needed. For example, if the observations are UTM map coordinates given with ten significant figures and of cm-precision, a threshold ε_a = 1.0E-10 to 1.0E-11 for the slope a, $|\hat{a}^{(i+1)} - \hat{a}^{(i)}| \leq \varepsilon_a$, and ε_b = 1.0E-03 to 1.0E-04 for the intercept b, $|\hat{b}^{(i+1)} - \hat{b}^{(i)}| \leq \varepsilon_b$, could be an adequate choice. A general rule for the computations would be to use double precision arithmetic and round properly at the end of the process.

The iterated MMM algorithm is very easily understood and implemented, against other equivalent or similar existed models and iterated methods, as those given by Schaffrin and Wieser 2008, Shen et al. 2011, Simkooei and Jazaeri 2012, Pan et al. 2015, the latter being almost the same with MMM as far as the used line form. A completely different model has been given by Neri et al. (1989) where the best fitting line is estimated by minimizing the perpendicular distances to the line, which are weighted on the basis of error propagation law, without approximations and updating of parameters except in the final step of solving a quadratic equation. Without going into details, some of the rigorous referred models have computational difficulties and/or biases with the covariance estimations of parameters.

All the quoted iterated models use different precision for each one of the observations and use also the same well-known example, whose data (x_i, y_i) has been given by Pearson (1901) completed with York's weights (1966). The presented here MMM, used the same data and gave identical results both for the estimations (parameters, errors, observables, a posteriori variance) and the covariances of the estimated parameters when base on LS principles.

Another group of EIV models and solutions found in the literature is based on the Maximum Likelihood Method (e.g., Madansky 1959, York 1966, Reed 1992, York et al. 2004, Cantrell 2008). Although the estimation of parameters agrees with those given by the presented MMM and the other equivalent referred models, criticism has been raised on the stochastic behavior and interpretation of the parameter estimates, and also on some computational difficulties.

3.3. Orthogonal fitting with equal precision and the slope intercept form

The previous MMM algorithm works also with equal precision or equal weights,

$$\sigma_{x_i}^2 = \sigma_{y_i}^2 = \sigma^2, \quad (P = I,\ p_{x_i} = p_{y_i} = 1) \tag{3.23}$$

Orthogonal fitting (geometric fitting, orthogonal regression) is a special case that has been studied since the late of 1800's (Adcock 1987) and has been taken closed analytical solutions, based on non-linear LS. For that reason, and because this case is interesting for many applications, it is here presented.

The squares of the errors that are minimized represent the squares of the perpendicular distances from any point to the fitting line. According to the LS criterion, the associated target function S is equivalently written as,

$$S(a,b) = \frac{1}{1+a^2}\sum\left(y_i - ax_i - b\right)^2 = \min. \tag{3.24}$$

The minimization criterion in this form was given for the first time by Adcock 1877, obtaining a solution given by a somehow inelegant (but correct) formula.

Zeroing the first derivatives with respect to (a) and (b), the estimations $\hat{a}, \hat{b}$ are obtained. Setting first $(\partial S(a,b)/\partial b) = 0$, we have

$$b = \overline{y} - a\,\overline{x} \tag{3.25}$$

meaning that the fit line passes through the centroid $(\overline{x}, \overline{y})$. Substituting (3.25) in (3.24), S depends only in a,

$$S(a) = \frac{s_{yy} - 2as_{xy} + s_{xx}a^2}{1+a^2} = \min. \tag{3.26}$$

From (3.26), taking $(\partial S(a)/\partial a) = 0$, we arrive at the quadratic equation

$$s_{xy}\hat{a}^2 - (s_{yy} - s_{xx})\hat{a} - s_{xy} = 0 \tag{3.27}$$

which has generally two solutions. The correct solution, that minimizes (3.26) (2^{nd} derivative>0) is that with the + sign before the square root of the discriminant, i.e.,

$$\hat{a} = \frac{s_{yy} - s_{xx} + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}} \tag{3.28}$$

(see also, Dermanis 1988, Dermanis and Fotiou 1992, Chevron 2010). Finally,

$$\hat{b} = \overline{y} - \hat{a}\,\overline{x}\ . \tag{3.29}$$

An equivalent and simple to apply formula given by Pearson (1901), is

$$\tan(2\hat{\theta}) = \frac{2s_{xy}}{s_{xx} - s_{yy}} \tag{3.30}$$

where θ stands for the orientation line angle, although (3.30) can be derived by a simple transformation of (3.28), e.g. Dermanis (1988), and inversely (3.28) can be derived from (3.3) if (tan2θ) is expressed as a function of θ. From (3.30), with a little investigation, ($2\hat{\theta}$) is computed at first and the reduced to the proper quadrant, accounting for both signs (+ or -) of the numerator and denominator. Then $\hat{\theta}$ ($0 \le \theta \le \pi$) and the slope $\hat{a} = \tan(\hat{\theta})$ are obtained,

The above parameter estimations by means of orthogonal fitting have the significant characteristic that they are independent of the position and the orientation of the coordinate system. Therefore, the best fit line remains the same (invariant) under translation and rotation of the (orthogonal) coordinate frame, but not invariant under scaling transformation unless the scale parameters for both coordinates are the same (e.g., Chernov 2010).

3.3.1. Special cases

Obviously, a solution for $\hat{a}$ from (3.28) is possible if $s_{xy} \neq 0$. In case $s_{xy} = 0$, the variables x and y are completely independent and this condition can hold only if the line is either horizontal $(\hat{a} = 0, s_{yy} = 0,)$ or vertical $(\hat{a} = \infty, s_{xx} = 0)$. The correct solution depends on the comparison between (s_{xx}, s_{yy}). We have a horizontal line if $s_{xx} > s_{yy} (= 0)$ or a vertical line if $s_{yy} > s_{xx} (= 0)$.

Another special case is when all data points are collinear (perfect fit). This (theoretical) situation means that the orthogonal distances from points to the line are zero, therefore, from (3.26) the sum S(a) = 0 or $s_{yy} - 2as_{xy} + s_{xx}a^2 = 0$. Because there is only one solution, the discriminant is zero, that is $4s_{xy}^2 - 4s_{xx}s_{yy} = 0$ or $s_{xy}^2 - s_{xx}s_{yy} = 0$ which is the condition for collinearity. The associated solution is then (s_{xy} / s_{xx}) that is identical to the simple linear regression (eq. 2.6).

Apart from the unique solution, i.e., the estimation of a unique best fit line, there are some cases, although seldom in practice, where infinitely may solutions of best fit lines exist (Chernov et al. 2013). This happens if and only if $(s_{xy=0}, s_{xx} = s_{yy})$, a condition that is a perfect circular symmetry (isomorphism) under arbitrary rotation. An example of such lines is when the data points are placed on the vertices of a regular polygon, where any line passing through its centroid is a best fit line.

3.4. Fitting with a known ratio of precision with the slope-intercept form

In a similar way with the previous development, we consider now the case where the ratio k of the precision for any observed point is known, that is, $k = \sigma_x / \sigma_y$, or

$k^2 = (\sigma_x^2 / \sigma_y^2) = (p_y / p_x)$. In this case we'll have (e.g., Chernov 2010),

$$S(a,b) = \frac{1}{1+k^2a^2}\sum\left(y_i - ax_i - b\right)^2 = \min.$$

$$\hat{a} = \frac{k^2 s_{yy} - s_{xx} + \sqrt{(k^2 s_{yy} - s_{xx})^2 + 4k^2 s_{xy}^2}}{2k^2 s_{xy}} \tag{3.31}$$

$$\hat{b} = \bar{y} - \hat{a}\,\bar{x}. \tag{3.32}$$

It has to be pointed out that in this model the minimized distances are not perpendicular to the fit line but have a direction along the vector $(\hat{k}a, -1)$. With a ratio k different from point to point, a closed solution is not possible and iterative methods have to be applied, as those above. For k=1 we are reduced to the orthogonal regression (equal weights).

4. Two observable parameters with the general equation form

The general form is free from the drawbacks or inconsistencies of the slope-intercept form. The general line equation,

$$Ax^{\alpha} + By^{\alpha} = C \tag{4.1}$$

is applicable with the restriction that A, B should not both equal zero, or equivalently $A^2 + B^2 > 0$. Then, for A≠0 the line cuts of intercept (C/A) on x-axis while for B≠0 the line cuts of intercepts (C/B) on y-axis and line slope (-A/B). As special cases, we have a horizontal line (y = C/B) when A=0 and a vertical line (x = C/A) when B=0. In addition, for A=0 and B=0 there is not exist a line if C ≠ 0 and there is an infinite number of solutions-lines if C = 0.

Each line in the general form (4.1) can be represented by an infinite number of vectors (A, B, C), all being proportional to each other, for instance lines $3x^{\alpha} + 4y^{\alpha} = 9$ and $6x^{\alpha} + 8y^{\alpha} = 18$ express the same line. In order to eliminate this 'ambiguity', it is necessary to impose a constraint, a suitable form being,

$$A^2 + B^2 = 1 \tag{4.2}$$

which satisfies the above restriction ($A^2 + B^2 > 0$) and has a clear geometric interpretation as we will see below.

Following the idea of the presented MMM in chapter 3, the target function to be minimized is

$$S(A,B,C) = \sum\left(p_{x_i} v_{x_i}^2 + p_{y_i} v_{y_i}^2\right) = \mathbf{v}^T\mathbf{P}\mathbf{v} = \min, \tag{4.3}$$

under the conditions and constraint

$$G_i = Ax_i^{\alpha} + By_i^{\alpha} - C = 0 \quad \text{and} \quad h = A^2 + B^2 - 1 = 0\,. \tag{4.4}$$

We can proceed on using as above the MMM model with constraints, based on the well-known general/mixed model with constraints, following the Least Squares solution with Lagrange multipliers and taking care on proper updating of parameters in the iterative algorithm.

However, a simplification to the functional model is possible by taking a parameter (t) instead of the two parameters A and B, so that

$$A = \cos t, \quad B = \sin t \tag{4.5}$$

fulfilling thus automatically the constraint $A^2 + B^2 = 1$. In doing so, the constraint is eliminated, having two independent parameters (t, C). Equation (4.1) becomes,

$$x^{\alpha}\cos t + y^{\alpha}\sin t = r \tag{4.6}$$

where C = r, and requiring at the end of a best fit process $r \geq 0$, parameter r represents the distance from the origin of the reference system to the line while t is the orientation angle of distance r reckoned positive from x-axis and anticlockwise.

Obviously angle (t) is related to the orientation line angle ω, i.e. a=tanω, a = –A / B = –cost/sint and b=r/sint. Noting that $0 \leq t < 2\pi$ and $0 \leq \omega < \pi$, it is easily derived the relation $\omega = t + \pi/2$, or $\omega = t + 90^o$ (Fig.1). If $\omega \geq 2\pi$ then ω is reduced to ω1= (ω -2π). If now, $\omega 1 \geq \pi$ then a new reduction ω2= (ω1 -π) gives the final correct value. These procedures can be work inversely, that is having ω to compute t, e.g. when an approximate value for (t) is needed from line slope (a) derived by a simple regression model (see next chapters 4.1, 4.2.).

From here on we will use the normal form of the line and the MMM solution/adjustment algorithm will be presented.

4.1. The modified mixed model with the normal equation form

The functional model is obviously not linear. According to the LS method of mixed equations, for any observed point a mixed equation is written, i.e.,

$$G_i = x_i^{\alpha}\cos t + y_i^{\alpha}\sin t = 0\,, \qquad i = i,\ldots, n. \tag{4.7}$$

Following, as above, the known linearization up to first order terms, around the approximate values t^o, r^o and x_i^o, y_i^o, the linear system is formed, and for any point we have,

$$G_i(t, r, x_i^{\alpha}, y_i^{\alpha}) = G_i(t^o, r^o, x_i^o, y_i^o) + \left.\frac{\partial G_i}{\partial t}\right|_o (t - t^o) + \left.\frac{\partial G_i}{\partial r}\right|_o (r - r^o) +$$

$$+\left.\frac{\partial G_i}{\partial x_i^{\alpha}}\right|_o (x_i^{\alpha} - x_i^{o}) + \left.\frac{\partial G_i}{\partial y_i^{\alpha}}\right|_o (y_i^{\alpha} - y_i^{o}) + \ldots = 0 \tag{4.8}$$

With $t = t^o + \delta t$, $r = r^o + \delta r$, $x_i^{\alpha} = x_i - v_{x_i}$, $y_i^{\alpha} = y_i - v_{y_i}$, $x_i^{o} = x_i - v_{x_i}^{o}$, $y_i^{o} = y_i - v_{y_i}^{o}$, and noting that, $x_i^{\alpha} - x_i^{o} = (-v_{x_i} + v_{x_i}^{o})$, $y_i^{\alpha} - y_i^{o} = (-v_{y_i} + v_{y_i}^{o})$, equation (4.8) becomes,

$$(x_i^{o}\cos t^{o} + y_i^{o}\sin t^{o} - r^{o}) + [(-x_i^{o}\sin t^{o} + y_i^{o}\cos t^{o}) \quad -1]\begin{bmatrix}\delta t\\ \delta r\end{bmatrix} +$$

$$+[\cos t^{o} \quad \sin t^{o}]\begin{bmatrix}-v_{x_i} + v_{x_i}^{o}\\ -v_{y_i} + v_{y_i}^{o}\end{bmatrix} = 0 \tag{4.9}$$

or,

$$[x_i\cos t^{o} + y_i\sin t^{o} - r^{o}] + [(-x_i^{o}\sin t^{o} + y_i^{o}\cos t^{o}) \quad -1]\begin{bmatrix}\delta t\\ \delta r\end{bmatrix} +$$

$$-[\cos t^{o} \quad \sin t^{o}]\begin{bmatrix}v_{x_i}\\ v_{x_i}\end{bmatrix} = 0\,. \tag{4.10}$$

The same expression (4.10) is also derived by an indirect linearization process as stated in chapter 3.1.

From (4.10), the linear system in matrix layout is written as,

$$\begin{bmatrix} x_1\cos t^{o} + y_1\sin t^{o} - r^{o}\\ x_2\cos t^{o} + y_2\sin t^{o} - r^{o}\\ \vdots\\ x_n\cos t^{o} + y_n\sin t^{o} - r^{o}\end{bmatrix} + \begin{bmatrix}(-x_1^{o}\sin t^{o} + y_1^{o}\cos t^{o}) & -1\\ (-x_2^{o}\sin t^{o} + y_2^{o}\cos t^{o}) & -1\\ \vdots & \vdots\\ (-x_n^{o}\sin t^{o} + y_n^{o}\cos t^{o}) & -1\end{bmatrix}\begin{bmatrix}\delta t\\ \delta r\end{bmatrix}$$

$$-\begin{bmatrix}\cos t^{o} & 0 & \cdots & 0 & \sin t^{o} & 0 & \cdots & 0\\ 0 & \cos t^{o} & \cdots & 0 & 0 & \sin t^{o} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \cos t^{o} & 0 & 0 & \cdots & \sin t^{o}\end{bmatrix}\begin{bmatrix}v_{x_1}\\ v_{x_2}\\ \vdots\\ v_{x_n}\\ v_{y_1}\\ v_{y_2}\\ \vdots\\ v_{y_n}\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix} \tag{4.11}$$

or in brief,

$$\mathbf{w} + \mathbf{A}\mathbf{z} - \mathbf{B}\mathbf{v} = \mathbf{0}\,. \tag{4.12}$$

Again, it is evident that matrix **A** is now fully affected by the errors of both observations x_i and y_i.

As in chapter 3.1 the LS solution is carried out in an iterative scheme, where at each iteration the approximate values $t^o, r^o, v_{x_i}^o, v_{y_i}^o$ are properly updated.

The LS solution follows the traditional method of Lagrange multipliers, where the objective function is minimized under equality conditions, i.e.,

$$\sum\left(p_{x_i} v_{x_i}^2 + p_{y_i} v_{y_i}^2\right) = \mathbf{v}^T\mathbf{P}\mathbf{v} = \min, \quad \text{under} \quad G_i = x_i^\alpha \cos t + y_i^\alpha \sin t - r = 0 \tag{4.13}$$

or,

$$\Phi = \mathbf{v}^T\mathbf{P}\mathbf{v} + 2\boldsymbol{\lambda}^T (x_i^\alpha \cos t + y_i^\alpha \sin t - r) = \min. \tag{4.14}$$

with $\boldsymbol{\lambda}$ being the Lagrange multipliers.

Considering again a diagonal covariance matrix of errors or observations (3.8, 3.9), the MMM algorithm is as follows: In order to formulate the system of normal equations (**Nz = -u**) we have to determine step by step,

$$\mathbf{B} = \begin{bmatrix} \cos t^o \mathbf{I}_n & \sin t^o \mathbf{I}_n \end{bmatrix} \tag{4.15}$$

$$\mathbf{M} = \mathbf{B}\mathbf{P}^{-1}\mathbf{B}^T = \left[(\cos t^o)^2 \mathbf{C}_\mathbf{x} + (\sin t^o)^2 \mathbf{C}_\mathbf{y} \right] = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_n) \tag{4.16}$$

$$\mu_i = (\cos t^o)^2 \sigma_{x_i}^2 + (\sin t^o)^2 \sigma_{y_i}^2 = \frac{(\cos t^o)^2}{p_{x_i}} + \frac{(\sin t^o)^2}{p_{y_i}}, \qquad p_i = \frac{1}{\mu_i} \tag{4.17}$$

$$\mathbf{M}^{-1} = \mathrm{diag}(1/\mu_1, 1/\mu_2, ..., 1/\mu_n) = \mathrm{diag}(p_1, p_2, ..., p_n) \tag{4.18}$$

$$\mathbf{N} = \mathbf{A}^T\mathbf{M}^{-1}\mathbf{A} = \begin{bmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{bmatrix} = \qquad (N_{21}=N_{12})$$

$$= \begin{bmatrix} \sum p_i \left(-x_i^o \sin t^o + y_i^o \cos t^o\right)^2 & -\sum p_i \left(-x_i^o \sin t^o + y_i^o \cos t^o\right) \\ -\sum p_i \left(-x_i^o \sin t^o + y_i^o \cos t^o\right) & \sum p_i \end{bmatrix} \tag{4.19}$$

$$\mathbf{u} = \mathbf{A}^T\mathbf{M}^{-1}\mathbf{w} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} =$$

$$= \begin{bmatrix} \sum p_i w_i \left(-x_i^o \sin t^o + y_i^o \cos t^o\right) \\ -\sum p_i w_i \end{bmatrix}, \qquad w_i = x_i \cos t^o + y_i \sin t^o - r^o. \tag{4.20}$$

Then, the estimation of corrections and parameters are,

$$\hat{\mathbf{z}} = \begin{bmatrix} \hat{\delta t} \\ \hat{\delta r} \end{bmatrix} = -\mathbf{N}^{-1}\mathbf{u} = -\frac{1}{\det(\mathbf{N})}\begin{bmatrix} N_{22} & -N_{12} \\ -N_{12} & N_{11} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} =$$

$$= -\frac{1}{\det(\mathbf{N})}\begin{bmatrix} N_{22}u_1 - N_{12}u_2 \\ -N_{12}u_1 + N_{11}u_2 \end{bmatrix} \tag{4.21}$$

where, $\det(\mathbf{N}) = N_{11}\,N_{22} - (N_{12})^2 \neq 0$, and

$$\hat{\mathbf{z}}^a = \begin{bmatrix} \hat{t} \\ \hat{r} \end{bmatrix} = \mathbf{z}^o + \hat{\mathbf{z}} = \begin{bmatrix} t^o + \hat{\delta t} \\ r^o + \hat{\delta r} \end{bmatrix} . \tag{4.22}$$

The estimation of errors and observable parameters are given by

$$\hat{\mathbf{v}} = \begin{bmatrix} \hat{\mathbf{v}}_x \\ \hat{\mathbf{v}}_y \end{bmatrix} = \mathbf{P}^{-1}\mathbf{B}^T\mathbf{M}^{-1}(\mathbf{w} + \mathbf{A}\hat{\mathbf{z}}) \tag{4.23}$$

or, analytically,

$$s_i = (\mathbf{w} + \mathbf{A}\hat{\mathbf{z}})_i = x_i \cos t^o + y_i \sin t^o - r_i^o + (-x_i^o \sin t^o + y_i^o \cos t^o)\delta t - \delta r \tag{4.24}$$

$$\hat{v}_{x_i} = (\cos t^o \sigma_{x_i}^2 p_i)\, s_i, \quad \hat{v}_{y_i} = (\sin t^o \sigma_{y_i}^2 p_i)\, s_i \tag{4.25}$$

$$\hat{x}_i^a = x_i - \hat{v}_{x_i}, \quad \hat{y}_i^a = y_i - \hat{v}_{y_i} . \tag{4.26}$$

Also, the estimation of an unknown variance is

$$\hat{\sigma}^2 = \frac{\sum\left(p_{x_i}\hat{v}_{x_i}^2 + p_{y_i}\hat{v}_{y_i}^2\right)}{n-2} \tag{4.27}$$

and the covariances of the estimations $\hat{t}, \hat{r}$, being the elements of the inverse matrix $\mathbf{N}^{-1}$, are

$$\hat{\sigma}_{\hat{t}}^2 = \sigma^2 \frac{N_{22}}{\det(\mathbf{N})}, \quad \hat{\sigma}_{\hat{r}}^2 = \sigma^2 \frac{N_{11}}{\det(\mathbf{N})}, \quad \hat{\sigma}_{\hat{t}\hat{r}} = -\sigma^2 \frac{N_{12}}{\det(\mathbf{N})} . \tag{4.28}$$

The above estimations are Best Linear Unbiased Estimations (BLUE). For a statistical evaluation of the model, we admit the normal distribution of errors. Moreover, with this stochastic hypothesis the estimations of parameters are identical to those derived by the Maximum Likelihood method. We underline again that an approximation to the above rigorous MMM solution starts from a Taylor series expansion around the point $F_i(t^o, r^o, x_i, y_i)$, instead of $F_i(t^o, r^o, x_i^o, y_i^o)$, as happens in the application of the usual general/mixed model, where satisfactory results are obtained in case good approximate values and small errors exist.

In contrast to the significant weakness of the slope-intercept model with the famil-

iar general/mixed adjustment model, the normal equation form gives accurate results in case of equal precision with the normal form and almost satisfactory results in case of different precision, as showed numerical tests. However, the MMM serves all cases and is preferable, particularly with data of different precision.

4.2. The iterative process of the modified mixed model with the normal form

At the beginning of the computations, initial approximate values t^o, r^o are computed by a simple method and moreover, $v_{x_i}^o = 0, v_{y_i}^o = 0$. Starting from (4.17), the estimations $\delta\hat{t}^{(1)}$, $\delta\hat{r}^{(1)}$ from (4.21), $\hat{t}^{(1)}$, $\hat{r}^{(1)}$ from (4.22) and $\hat{v}_{x_i}^{(1)}, \hat{v}_{y_i}^{(1)}$ from (4.25) are obtained. With the new estimates as the new approximate values the second iteration starts resulting in $\hat{t}^{(2)}$, $\hat{r}^{(2)}$, $\hat{v}_{x_i}^{(2)}$, $\hat{v}_{y_i}^{(2)}$, and so on until a convergence criterion is fulfilled. Within the same iteration, though it is not necessary, other estimates, e.g., $\hat{x}_i^{a(1)} = x_i - \hat{v}_{x_i}^{(1)}$, $\hat{y}_i^{a(1)} = y_i - \hat{v}_{y_i}^{(1)}$ from (4.26) and $\hat{\sigma}^{2(1)}$ from (4.27) can be also updated. In this parallel computations all estimations end with their adjusted values.

A convergence criterion for the successive absolute differences is set. It depends on the initial approximate values and the level of accuracy needed. For example, if the observations are Cartesian map coordinates given with ten significant figures and of cm-precision, reasonable thresholds could be ε_t = 1.0E-10 to 1.0E-12 for the angle t in radians, $|\hat{t}^{(i+1)} - \hat{t}^{(i)}| \le \varepsilon_t$, and ε_r = 1.0E-03 for the distance r, $|\hat{r}^{(i+1)} - \hat{r}^{(i)}| \le \varepsilon_r$ in meters.

The MMM iterated algorithm with the normal line form is accurate and easily implemented, as its counterpart in chapter 3. A computational test, by means of a fortran program in double precision arithmetic and with the same "Pearson's-York's" data showed that after about 10 to 12 iterations (depending on the approximate values) the results were identical with the previous MMM or other rigorous approaches. Note that the estimations of (t, r) should be converted properly in order to be comparable with those of (a, b).

4.3. Orthogonal fitting with equal precision and the normal form

The previous MMM algorithm (chapter 4.1) works also with equal precision or equal weights. Orthogonal fitting using the normal form leads to closed analytical solutions, as also happens with the general line equation model and should be preferable.

The squares of the errors of the perpendicular distances from any point to the fit line, that is minimized, is written as

$$S(t,r) = \sum \left(x_i \cos t + y_i \sin t - r \right)^2 = \min. \qquad (4.29)$$

Zeroing the first derivatives with respect to (t) and (r) the estimations $(\hat{t}, \hat{b})$ are

derived. Setting $(\partial S(t,r)/\partial r)=0$ results in

$$r=\bar{x}\cos t+\bar{y}\sin t \tag{4.30}$$

meaning that the line passes through the centroid $(\bar{x},\bar{y})$, a point on the fit line. Substituting (4.30) in (4.29), the function S depends only in (t), i.e.,

$$S(t)=\sum\left((x_i-\bar{x})\cos t+(y_i-\bar{y})\sin t\right)^2=$$

$$=\cos^2 t\,S_{xx}+2\sin t\cos t\,S_{xy}+\sin^2 t\,S_{yy}=\min. \tag{4.31}$$

where, S_{xx}, S_{yy}, S_{xy} has been given above (e.g. eq. 2.14). From (4.31), setting $(\partial S(t)/\partial t)=0$, we arrive at the quadratic equation

$$(\cos^2 t-\sin^2 t)s_{xy}+\sin t\cos t(s_{yy}-s_{xx})=0\,. \tag{4.32}$$

Dividing by cost, (4.32) becomes

$$\tan^2 t\,s_{xy}-\tan t\,(s_{yy}-s_{xx})-s_{xy}=0 \tag{4.33}$$

a solution that at a first glance is identical with (3.28). Dividing by sint, (4.32) becomes,

$$s_{xy}\cot^2 t\;+(s_{yy}-s_{xx})\cot(t)-s_{xy}=0 \tag{4.34}$$

an equation that is almost identical to that given by Alciatore and Miranda (1995) by means of a similar process.

Furthermore, using the identities sin(2t) = 2sintcost, cos(2t) = ($\cos^2$t - $\sin^2$t), and dividing (4.32) by cos(2t), we arrive at the solution,

$$\tan(2t)=\frac{-2S_{xy}}{s_{yy}-s_{xx}} \tag{4.35}$$

which is the same as that given by (3.29). The same solution is given also by Munoz et al. (2014).

Equations (4.33) and (4.34) are quadratic and each one have two solutions, the correct one being that with the (+) sign before the square root of the discriminants. In order to compute the direction angle (t), a proper quadrant reduction is needed, as with the case of (4.35). Special cases have to be also examined, as presented above, noting that the normal form and the associated general form can safely represent vertical lines.

Our interest is mainly (sint, cost) for the best fit line, from which A and B coefficients, and C or r, can be derived. For example, (4.33) gives the solution (3.28) and thus, $A=2s_{xy}$ and $B=s_{yy}-s_{xx}+\sqrt{(s_{yy}-s_{xx})^2+4s_{xy}^2}$.

With a similar, as previously, development, we can easily derive closed expressions when the ratio k of the precision for any observed point is known, that is, $k = \sigma_x / \sigma_y$ or $k^2 = (\sigma_x^2 / \sigma_y^2) = (p_y / p_x)$ is known. The corresponding solution in terms of (t, r), goes similar to that of chapter 3.4, based on the present formulation. Apart of the closed solutions presented previously, the orthogonal fitting with equal precision can be equivalently applied either by MMM or by the familiar general/mixed adjustment model.

5. Concluding remarks

The problem of a best fit line, though it is a well-known solved problem over a century now, could still show certain pitfalls and inconsistencies, especially when both observations are subject to errors and the slope-intercept form is used.

The traditional implementation of the general adjustment model does not lead in all cases to the correct or satisfactory solution except when used with the normal line form and certainly with data of equal precision.

A new accurate iterative algorithm is presented using the normal form and a least squares modified mixed model (MMM), characterized by simplicity, clarity and ease of implementation.

Many of the EIV or total least squares models, existed in the literature, present some computational difficulties and/or biases in the estimations. The presented MMM dos not suffer from such problems and could be the preferable model.

The general line equation parameters obtained by the normal equation model adjustment process, or even directly using proper constraints, should be used as their values can be bounded.

References

Adcock, R.J. (1878). A problem in least squares. *Analyst*, London, Vol. 5, 53 – 54.

Alciatore, D. and Miranda, R. (1995). *The Best Least-Squares Line Fit*. In Graphics Gems, V.A.W. Paeth (Ed.), Academic Press, Waltham, MA, USA, 91 – 97.

Cantrell, C.A. (2008). Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problem. *Atmos. Chem. Phys.*, 8, 5477 – 5487.

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

Chernov, N., Q. Huang and H. Ma (2013). Is the Best Fitting Curve Always Unique? *Journal of Mathematics*, Volume 2013, 5 pages.

Harvey, B.R. (2009). Constraint Equations in Cadastral Modelling. *Spatial Science*, Vol. 54, No 2, 1 – 12.

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

Dermanis, A. (1988). Best fitting of a line to points in the plain (in Greek). *Eratosthenes*, 21, 5 – 24.

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

Jefferys, H. W. (1980). On the method of least squares. *The Astronomical Journal*, Vol. 85, N0. 2, 177 – 181.

Kyriati, N. and A.M. Bruckstein (1992). What's in a set of Points? *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 14, No. 4, 496 – 500.

Lybanon, M. (1984). A better least-squares method when both variables have uncertainties. *Am. J. Phys.*, 52(1), 22 – 26.

Madansky, A. (1959). The Fitting of Straight Lines When both Variables are Subject to Error. *Journal of the American Statistical Association*, Vol. 54, No. 285, 173 – 205.

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

Munoz, L.R., M. G. Villanueva and C. G. Suarez (2014). A tutorial on the total least squares method for fitting a straight line and a plane. *Revista de Ciencia e Ingen. del Inst. Techol. Superior de Coatzacoalcos*. An. 1, No. 1, 167 – 173.

Neri F., G. Saitta and S. Chiofalo (1989). An accurate and straightforward approach to line regression analysis of error-affected experimental data. *J. Phys. Ser E: Sci. Instrum*., 22, 215 – 217.

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

Pearson, K. (1901). On lines and planes of closest fit to system of points in space. *Philosophical Magazine*, 2:559 – 572.

Reed, C.B. (1992). Linear least-squares fits with errors in both coordinates, II: Comments on parameter variances. *Am. J. Phys*. 60, 59 – 62.

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

Shen Y., Li B., Chen Y. (2011). An iterative solution of weighted total least-squares adjustment. *J. Geod.* 85, 229 – 238.

Simkooei A. A. and S. Jazaeri (2012). Weighted total least squares formulated by standard least squares theory. *Journal of Geodetic Science*, 2(2), 113 – 124.

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

Wolf, P. R. and C. D. Ghilani (1997). *Adjustment Computations. Statistics and Least Squares in Surveying and GIS*. John Wiley & Sons, Inc.

York, D (1966). Least-squares fitting of a straight line. *Can. J. Phys*., 44, 1079 – 1086.

York, D., N. M. Evensen, M. L. Martinez and Delgado J de B. (2004). Unified equations for the slope, intercept, and standard errors of the best straight line. *Am. J. Phys.* 72, 367 – 375.