An Algebraic solution of maximum likelihood function in case of Gaussian mixture distribution

Awange J.L[a,d,e], Paláncz B[b], Lewis R[c], Lovas T[b], Heck B[d], Fukuda Y[e]

[a] Western Australian Centre for Geodesy and The Institute for Geoscience Research, Curtin University, Perth, Australia

[b] Department of Photogrammetry and Geoinformatics, Budapest University of Technology and Economics 1521 Budapest, Hungary

[c] Department of Mathematics, Fordham University, New York, USA

[d] Geodetic Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany

[e] Department of Geophysics, Kyoto University, Kyoto, Japan

Abstract

Traditionally, least squares method (LSM) has been employed as a standard technique for parameter estimation and regression fitting of models to measured points in data sets in many engineering disciplines, geoscience fields as well as in geodesy. To get the optimal linear unbiased estimator, which provides minimum variance, the model error should follow a Gaussian distribution with zero mean. However, this may not always be the case due to contaminated data (i.e., the presence of outliers) or data from different sources with varying distributions. This study proposes an algebraic iterative method that approximates the error distribution model using a Gaussian mixture distribution, with the application of maximum likelihood estimation as a possible solution to the problem. The global maximization of the likelihood function is carried out through the computation of the global solution of a multivariate polynomial system using numerical Groebner basis in order to considerably reduce the running time. The novelty of the proposed method is the application of total least square (TLS) error model as opposed to ordinary least squares (OLS) and the maximization of the likelihood function of the Gaussian mixture via algebraic approach. Use of TLS error model rather than OLS enables errors in all the 3 coordinates of the model of a 3D plane (i.e., $z = \alpha x + \beta y + \gamma$) to be considered. The proposed method is illustrated by fitting a plane to real laser point cloud data containing outliers to test its robustness. Compared to the RANdom Sample Consensus (RANSAC) and Danish robust estimation methods, the results of the proposed algebraic method indicate its efficiency in terms of *computational time* and its *robustness* in managing outliers. The proposed approach thus offers an alternative method for solving mixture distribution problems geodesy.

Keywords: Robust parameter estimation, expectation maximization, maximum likelihood estimation, Groebner basis, outliers, point cloud, algebraic solution, Gaussian distributions, total least squares.

1. Introduction

In geodesy as is in many engineering disciplines, least squares method (LSM) is employed as a standard technique for parameter estimation and regression fitting of models to points of measured data sets (see, e.g., Grafarend and Awange 2012). Employing optimal unbiased linear LSM estimator providing minimum variance, one has to keep in mind that the model error distribution should follow a Gaussian distribution with zero mean. However, this may not always be the case due to contamination of the dataset (e.g., resulting from the presence of outliers) or having data that originates form different types of sources with different distributions. In either case, a mixed distribution has to be reckoned with (see, e.g., Lang et al., 1989, Xu 2005, Koch and Kargoll 2013, and Koch 2014).

In the emerging field of integrated geodesy, for example, where observations from global satellite navigation system (GNSS) and those of laser scanning, photogrammetry, and CAD modelling are integrated (e.g., Agnello and Brutto 2007; Borre 2006), such integration brings with it a mixture of different types of distributions that could be Gaussian or non-Gaussian. Furthermore, outliers that corrupt the laser scanned data could occur due, e.g., to occlusions, off-surface points and multiple reflectance, thereby limiting surface reconstruction using the point cloud. Further examples include the case where global positioning system (GPS) and Interferometric Synthetic Aperture Radar (InSAR) are related to a slip distribution model used in modelling co-seismic surface displacements (e.g., Sun et al 2011), GPS ambiguity resolution problem where the carrier phase observations are very precise but contain integer unknowns leading to a mixed observation model (e.g., Xu 1998), and assimilation of stream flow observations and satellite data in order to carry out hydrological model calibration (e.g., Eicker et al., 2014). Other disciplines where integrated data of mixed distributions are encountered include meteorology, oceanography, and seismology, where sampling data is imperfect and irregular (Nodet 2012). The foregoing discussions point to the need for robust fitting techniques that can manage the resulting outliers.

The problem of outlier management when the model error distribution does not satisfy the Gaussian with zero mean condition has been extensively discussed, e.g., in Zuliani (2012). The frequent solution to this problem is the application of robust estimation techniques, e.g., the Danish and the RANdom Sample Consensus (RANSAC) method (Krarup et al., 1980; Yaniv 2010), which eliminate outliers using noise thresholds. Robust statistical approaches of parameter estimation in case of contaminated data were pioneered by Huber (1964). Xu (2005) proposed the sign-constrained robust estimator with subjective breakdown point, which is methodologically different from methods discussed in any statistical literature. Other approaches include the principal component analysis (PCA, e.g., Huang and Tseng 2008), improved 3D Hough transform (Borrmann et al., 2011), Bayesian techniques (Diebel et al., 2006), elimination methods such as those based on the minimum covariance determinant (Russeeuw and Van Driessen 1999), employing the bifactor reduction model of weight elements (Yang et al., 2002, Chen and Stamos 2007), and data assimilation techniques using probability distributions and recursive Bayesian estimation and Kalman filtering (Guttman and Lin 1995; Sun et al 2011; and Elberg 2015).

In addition, robust estimations based on the expectation maximization (EM) of mixed distributions have been proposed (e.g., Lakaemper and Latecki 2006). For instance, Aitkin and Wilson (1980) applied a mixture of two normally distributed components, the first one for the observations with expected values defined by a linear model and the second one for an outlier with its own expected value. Thus, a mean-shift model is introduced and the unknown parameters estimated using the EM algorithm. Koch (2013) generalized this method by introducing a mixture of any number of normally distributed components, in case of two, the first one is for observations and the second one is for outliers. Furthermore, Lang et al., (1989) used the t-distribution to derive for a linear model a robust estimation by the EM algorithm. They introduced weights for the observations, which were small for outliers, thus using a variance-inflation model, and succeeded in obtaining an adaptive robust estimation. Koch and Kargoll (2013) suggested the use of variance-inflation model to detect the outliers and the mean-shift model to confirm them, a method that turned out to be very efficient. Later, Koch (2014) showed that the EM algorithm based on the mean-shift and variance-inflation model does not have to be restricted to a linear model but can also be applied to nonlinear models. The concept of break-down point, a point representing the maximum percentage of contaminated data beyond which the estimator can no longer produce meaningful solution was introduced by Donoho and Huber (1983), and improved by others (e.g., Rousseeuw 1984; Hampel at al. 1986), while Xu (2005) introduced a robust method with a highest possible break-down point.

In contributing to the expectation maximization robust based methods, the present contribution proposes an alternative algebraic method that is iterative in nature, but which defines the error model in terms of *total least squares* (TLS) as opposed to the ordinary least squares method commonly used in most of the studies above. The method applies a likelihood function, where the distribution of model error is approximated using a Gaussian mixture distribution computed by the EM algorithm. Using this approach, a linear parameter estimation model is considered, where the global maximization of the likelihood function is carried out by solving a multivariate polynomial system using numerical Groebner basis that considerably reduces the computation time. The advantages inherent in using total least squares error model compared to ordinary least squares is that it is able to take into consideration all the measurement errors in all the 3 coordinates (x, y, z) of a 3D plane model such as $z = \alpha x + \beta y + \gamma$. The rest of the study is organised as follows. In section 2, the likelihood function for standard LSM is presented followed by a discussion of the expectation maximization method in section 3. Section 4 considers the likelihood function for a Gaussian mixture before presenting the proposed iterative algorithm with the embedded algebraic solution in section 5. Section 6 presents an illustrative example based on a real laser scanned data obtained from a site in Budapest (Hungary) while section 7 concludes the study.

2. Maximum Likelihood Estimation

Generally, to carry out a regression procedure using maximum likelihood method (ML), one needs to have a model $\mathcal{M}$ (x, y, z: $\boldsymbol{\theta}$) = 0, a model error definition $e_{\mathcal{M}_i}(x_i$, y_i,z_i: $\boldsymbol{\theta}$), as well as, the probability density function of the error PDF ($e_{\mathcal{M}}$(x, y, z: $\boldsymbol{\theta}$)). The linear model then becomes

$$\mathcal{M}(x, y, z : \boldsymbol{\theta}) = \alpha x + \beta y + \gamma - z, \tag{1}$$

where the terms of the parameter $\boldsymbol{\theta} = (\alpha, \beta, \gamma)$ are to be determined. The error model corresponds to the shortest distance of a point P_i from its perpendicular projection to a hyperplane,

$$e_{\mathcal{M}_i}\left(x_i, y_i, z_i : \boldsymbol{\theta}\right)) = \frac{z_i - x_i\alpha - y_i\beta - \gamma}{\sqrt{1 + \alpha^2 + \beta^2}}. \tag{2}$$

One has to mention that a mathematically equivalent error-in-variable (EIV) model can be given employing a nonlinear adjustment model (see, e.g., Xu 2012). The probability density function of the error model is considered as a Gaussian error distribution of $\mathcal{N}$(0, σ) given by

$$PDF\left(e_{\mathcal{M}}(x, y, z : \boldsymbol{\theta})\right) = \frac{e^{-\frac{(e_{\mathcal{M}})^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}. \tag{3}$$

Considering a set $\{(x_1, y_1), (x_2, y_2) \ldots, (x_N, y_N)\}$ as measurement points, the maximum likelihood method aims at finding the parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood of the joint error distribution. Assuming that the errors are independent, one should maximize,

$$\mathcal{L} = \prod_{i=1}^{N} \frac{e^{-\frac{\left(e_{\mathcal{M}_i}\right)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}. \tag{4}$$

In order to use the sum instead of product, the logarithm of Eq. (4), i.e.,

$$\mathrm{Log}\mathcal{L} = \mathrm{Log}\left(\prod_{i=1}^{N} PDF(e_{\mathcal{M}})\right) = \sum_{i=1}^{\mathbb{N}} \mathrm{Log}(PDF(e_{\mathcal{M}})), \tag{5}$$

is used. If the Gaussian error distribution is considered, the function to be now minimized becomes

$$\text{-}\mathrm{Log}\mathcal{L}(\alpha, \beta, \gamma) = N\mathrm{Log}\left(\sqrt{2\pi}\sigma\right) + \frac{1}{2\sigma^2} \sum_{i=1}^{N} \frac{\left(z_i - x_i\alpha - y_i\beta - \gamma\right)^2}{1 + \alpha^2 + \beta^2}, \tag{6}$$

which is practically the optimal least squares method. Forming the necessary conditions of the optimum through partial derivatives as

$$\mathrm{eq}_1 = \frac{\partial \mathrm{Log}\mathcal{L}}{\partial \alpha} = 0, \mathrm{eq}_2 = \frac{\partial \mathrm{Log}\mathcal{L}}{\partial \beta} = 0, \mathrm{eq}_3 = \frac{\partial \mathrm{Log}\mathcal{L}}{\partial \gamma} = 0, \tag{7}$$

one obtains the following system of three multivariate polynomial equations,

$$\left.\begin{array}{l} \mathrm{eq}_1 = i - b\alpha + h\alpha - i\alpha^2 - e\beta - 2g\alpha\beta + e\alpha^2\beta + i\beta^2 - \\ b\alpha\beta^2 + d\alpha\beta^2 - e\beta^3 - a\gamma - 2f\alpha\gamma + a\alpha^2\gamma + 2c\alpha\beta\gamma - a\beta^2\gamma + N\alpha\gamma^2 \\ \mathrm{eq}_2 = g - e\alpha + g\alpha^2 - e\alpha^3 - d\beta + h\beta - 2i\alpha\beta + b\alpha^2\beta - \\ d\alpha^2\beta - g\beta^2 + e\alpha\beta^2 - c\gamma - c\alpha^2\gamma - 2f\beta\gamma + 2a\alpha\beta\gamma + c\beta^2\gamma + N\beta\gamma^2 \\ \mathrm{eq}_3 = f - a\alpha - c\beta - N\gamma \end{array}\right], \tag{8}$$

where the constants $(a, b, c, d, e, f, g, h, i)$ depend on the measured values (x_i, y_i, z_i), i= 1,2,...,N.

The solutions of this system of polynomial equations are the possible optimums of Eq.(6), and can be solved, for example, using the Sylvester resultant (e.g., Awange and Grafarend 2005; Awange et al., 2010; Awange and Palancz 2016) or the Dixon resultant (Lewis et al., 2014). Since the last expression of Eq. (8) is linear, it can be expressed in terms of γ and then substituted in the first two equations of Eq. (8) leading to a system of two equations in two unknowns (α and β), which can be solved using reduced Groebner basis (Awange and Grafarend 2005; Awange et al., 2010; Awange and Palancz 2016) to yield a univariate polynomial of seventh order in α and β (see, e.g., Awange et al., 2014). Once α and β have been solved, they are substituted in the last equation of Eq. (8) to yield γ. The triplet (α, β, and γ) of the solution is considered as a proper global maximum solution if it is real, and in comparison to other triplets, minimizes Eq. (6). To illustrate the situation, let us consider Fig. (1, left), where inliers (blue points) and outliers (red points) are considered together as data points and Eq. (8) is solved. The model error can then be computed with the known parameters (α,β,γ), see Fig 2. This distribution has a "tail" on the right-hand-side indicating that the histogram does not represent a Gaussian distribution with zero mean. It can even be considered as a mixture of two general Gaussians.

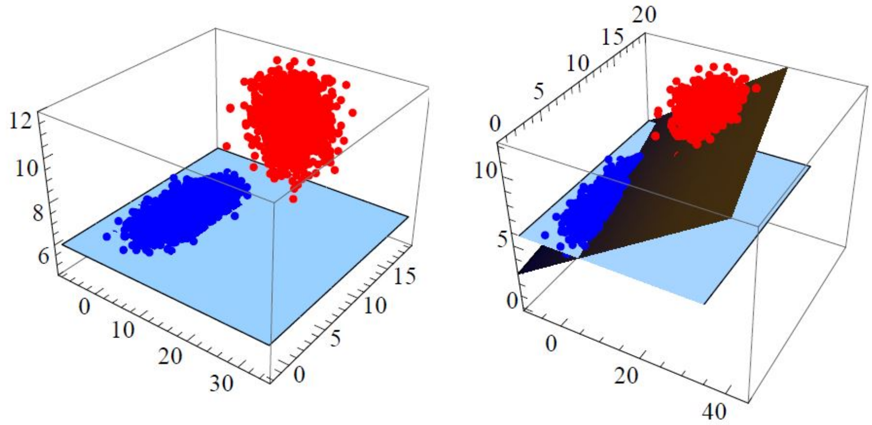

Figure 1: The inliers (blue points) and outliers (red points) considered together as one data set (left figure), and the plane with black color resulted from the solution of Eq. 8 (right figure)

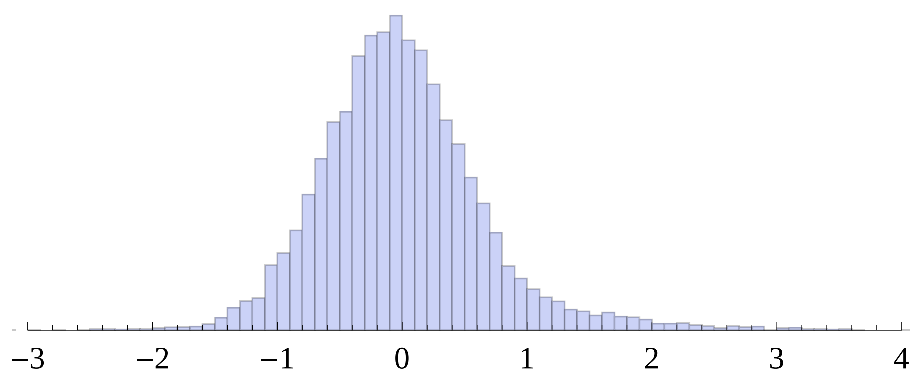

Figure 2: The histogram of the error distribution computed with parameters α, β, and γ from Eq. 8

3. Expectation Maximization Algorithm

Now, the question that arises from the error distribution in Fig. 2 is how it can be approximated by two Gaussians, or alternatively, how can the distribution be separated into two Gaussian ones. This separation can be done by the EM algorithm (Hastie and Tibshirani 2008). This algorithm appeared in the geodetic literature, e.g., in the paper of Luxen and Brunn (2003) extracting straight lines and parabolas from digital images as well as in the paper of Peng (2009) employing EM for robust estimation of parameters and variance components. Recently the method was also employed by Koch (2014) as well as Koch and Kargoll (2013). One should mention that robust methods with high breakdown point can also be applied to solve the problem, see e.g., Xu (2005). Let us consider a two-component Gaussian mixture represented by the mixture model in the following form,

$$\mathcal{N}_{12}(x) = \eta_1 \mathcal{N}\left(\mu_1, \sigma_1, x\right) + \eta_2 \mathcal{N}\left(\mu_2, \sigma_2, x\right), \tag{9}$$

where

$$\mathcal{N}\left(\mu_i, \sigma_i, x\right) = \frac{e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi}\sigma_i}, i = 1, 2, \tag{10}$$

and η_i's are the membership weights constrained by

$$\eta_1 + \eta_2 = 1. \tag{11}$$

The parameters being sought are (μ_1, σ_1) and (μ_2, σ_2). The log-likelihood function in case of Nsamples is

$$\mathrm{Log}\mathcal{L}\left(x_i, \theta\right) = \sum_{i=1}^{N} \mathrm{Log}\left(\mathcal{N}_{12}\left(x_i, \theta\right)\right) = \sum_{i=1}^{N} \mathrm{Log}\left(\eta_1 \mathcal{N}\left(\mu_1, \sigma_1, x_i\right) + \eta_2 \mathcal{N}\left(\mu_2, \sigma_2, x_i\right)\right), \tag{12}$$

where $\theta = (\mu_1,\ \sigma_1, \mu_2, \sigma_2)$ are the parameters of the Gaussian distributions. The problem is the direct maximization of this function due to the sum of terms inside the logarithm. In order to solve this problem, let us introduce the following alternative log-likelihood function:

$$\begin{gathered}\mathrm{Log}\mathcal{L}\left(x_i, \theta, \Delta\right) = \textstyle\sum_{i=1}^{N} \left(1 - \Delta_i\right) \mathrm{Log}\left(\mathcal{N}\left(\mu_1, \sigma_1, x_i\right)\right) + \Delta_i \mathrm{Log}\left(\mathcal{N}\left(\mu_2, \sigma_2, x_i\right)\right) + \\ \textstyle\sum_{i=1}^{N} \left(1 - \Delta_i\right) \mathrm{Log}\left(\eta_1\right) + \Delta_i \mathrm{Log}\left(\eta_2\right).\end{gathered} \tag{13}$$

Here Δ_i's are considered as unobserved latent variables taking values 0 or 1. If x_i belongs to the first component, then Δ_i= 0, so

$$\mathrm{Log}\mathcal{L}\left(x_i, \theta, \Delta\right) = \sum_{i \in N_1(\Delta)} \mathrm{Log}\left(\mathcal{N}\left(\mu_1, \sigma_1, x_i\right)\right) + N_1 \mathrm{Log}\left(\eta_1\right) \tag{14}$$

otherwise x_i belongs to the second component then Δ_i=1, leading to

$$\mathrm{Log}\mathcal{L}\left(x_i, \theta, \Delta\right) = \sum_{i \in N_2(\Delta)} \mathrm{Log}\left(\mathcal{N}\left(\mu_2, \sigma_2, x_i\right)\right) + N_2 \mathrm{Log}\left(\eta_2\right), \tag{15}$$

where N_1 and N_2 are the numbers of the elements of the mixture, which belong to the first and to the second components, respectively.

Since the values of the Δ_i's are unknown, an iterative procedure is adopted. Substituting for each Δ_i, its expected value becomes,

$$\xi_i(\theta) = E\left(\Delta_i|\theta, x\right) = \Pr\left(\Delta_i = 1|\theta, x\right) \approx \frac{\eta_2 \mathcal{N}\left(\mu_2, \sigma_2, x_i\right)}{\left(1-\eta_2\right)\mathcal{N}\left(\mu_1, \sigma_1, x_i\right) + \eta_2 \mathcal{N}\left(\mu_2, \sigma_2, x_i\right)}. \quad (16)$$

This expression is also called the responsibility of component 2 for observation i. Then, the EM algorithm for the two components of the Gaussian mixture is as follows:

Take the initial guess for the parameters: $\theta = (\tilde{\mu_1}, \tilde{\sigma_1}, \tilde{\mu_2}, \tilde{\sigma_2})$ and for $\tilde{\eta_2}$

Expectation Step: compute the responsibilities:

$$\tilde{\xi_i} = \frac{\tilde{\eta_2} \mathcal{N}\left(\tilde{\mu_2}, \tilde{\sigma_2}, x_i\right)}{\left(1-\tilde{\eta_2}\right)\mathcal{N}\left(\tilde{\mu_1}, \tilde{\sigma_1}, x_i\right) + \tilde{\eta_2} \mathcal{N}\left(\tilde{\mu_2}, \tilde{\sigma_2}, x_i\right)}, \text{ for } i = 1, 2, ..., N. \quad (17)$$

Maximization Step: compute the weighted means and variances for the two components:

$$\begin{gathered} \tilde{\mu_1} = \sum_{i=1}^N \left(1-\tilde{\xi_i}\right) x_i \Big/ \sum_{i=1}^N \left(1-\tilde{\xi_i}\right) \\ \tilde{\sigma_1} = \sum_{i=1}^N \left(1-\tilde{\xi_i}\right) \left(x_i - \tilde{\mu_1}\right)^2 \Big/ \sum_{i=1}^N \left(1-\tilde{\xi_i}\right) \\ \tilde{\mu_2} = \sum_{i=1}^N \tilde{\xi_i} x_i \Big/ \sum_{i=1}^N \tilde{\xi_i} \\ \tilde{\sigma_2} = \sum_{i=1}^N \tilde{\xi_i} \left(x_i - \tilde{\mu_1}\right)^2 \Big/ \sum_{i=1}^N \tilde{\xi_i}, \end{gathered} \quad (18)$$

and the mixing probability

$$\tilde{\eta_2} = \sum_{i=1}^N \tilde{\xi_i} \Big/ N. \quad (19)$$

Assuming two Gaussian distributions, this method provides not only the means and standard deviations $\{\mu_1, \sigma_1\}$, and $\{\mu_2, \sigma_2\}$ of the distributions, but also the membership functions for every data point $\{\eta_1, \eta_2\}$. Consequently, the data belonging to the two different distributions can be identified (see Fig. 3, left). It can be seen that some sample elements are misclassified. Using these parameters $\{\mu_1, \mu_2, \sigma_1, \sigma_2, \eta_1, \texttt{and}\, \eta_2\}$, let us compute new parameters of the plane (α, β, γ). To achieve that, ML estimation is now employed using a Gaussian mixture as a type of distribution.

Therefore, we turn to the modified ML function involving a Gaussian mixture as an error distribution.

4. Maximum Likelihood Estimation for a Gaussian Mixture

Maximum likelihood estimation (MLE) models having different probability distributions than standard Gaussian, $\mathcal{N}(0, \sigma)$ can be found in literature. For example Uhler

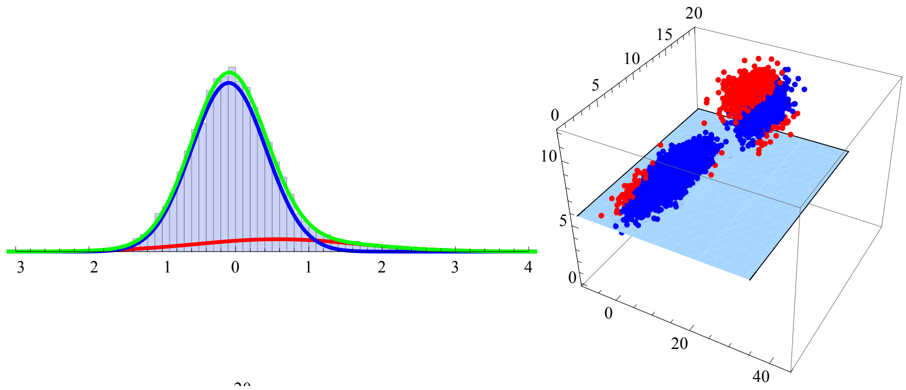

Figure 3: The histograms and the data points of the error distributions resulting from the application of EM algorithm (left). Two distributions with the inliers (blue points) and outliers (red points) are shown in the right figure.

(2011) employed multivariate Gaussian while Rose (2000) considered the probability density function different from Gaussian and solved MLE in symbolical form. There are also examples for using multivariate algebraic polynomial solution with mixture of distributions, e.g., by Drton (2006) who applied seemingly unrelated regression (SUR), and recently Batselier et al. (2012) who used discrete statistical model of mixture. Here, a multivariate polynomial solution of the MLE in case of two component Gaussian mixture of continuous probability variables is adopted, where in ideal case one of the components can represent the inliers points, while the other component is for the outliers. The likelihood function for Gaussian mixture is

$$\begin{gathered}\mathrm{Log}\mathcal{L}\left(x_i,\theta\right)=\sum_{i\in N_1}\mathrm{Log}\left(\mathcal{N}\left(\mu_1,\sigma_1,x_i,\theta\right)\right)+\\ \sum_{i\in N_2}\mathrm{Log}\left(\mathcal{N}\left(\mu_2,\sigma_2,x_i,\theta\right)\right)+N_1\mathrm{Log}\left(\eta_1\right)+N_2\mathrm{Log}\left(\eta_2\right),\end{gathered} \tag{20}$$

where the index “1” refers to the first component while “2” corresponds to the second component. Considering Eq. (6) for both components and forming the necessary conditions of the optimum, the corresponding polynomial form of the maximization problem is developed in similar manner to the single component distribution, see Eq. (7). Details of the algebraic derivation of these equations in symbolic way can be found in Paláncz (2014).

$$\sigma_2^2\left(-\frac{\mathbb{N}_1\left(-2\alpha\gamma^2-2\alpha\gamma\mu_1\sqrt{1+\alpha^2+\beta^2}\right)}{2}+\right.$$

$$\left(2\alpha^2\gamma-\left(1+\alpha^2+\beta^2\right)\gamma\right)a_1+b_1\left(\alpha^3-\alpha\left(1+\alpha^2+\beta^2\right)\right)+$$

$$
\begin{aligned}
&e_1\left(2\alpha^2\beta-\beta\left(1+\alpha^2+\beta^2\right)\right)+i_1\left(\left(1+\alpha^2+\beta^2\right)-2\alpha^2\right)+\\
&a_1\left(\alpha^2\mu_1\sqrt{1+\alpha^2+\beta^2}-\mu_1\left(1+\alpha^2+\beta^2\right)^{3/2}\right)+2\alpha\beta\gamma c_1+\alpha\beta^2 d_1-2\alpha\gamma f_1+\\
&\alpha h_1-2\alpha\beta g_1+\alpha\beta\mu_1\sqrt{1+\alpha^2+\beta^2}-f_1\alpha\mu_1\sqrt{1+\alpha^2+\beta^2}\Bigg)+\\
&\sigma_1^2\Bigg(-\frac{\mathbb{N}_2\left(-2\alpha\gamma^2-2\alpha\gamma\mu_2\sqrt{1+\alpha^2+\beta^2}\right)}{2}+\\
&\left(2\alpha^2\gamma-\left(1+\alpha^2+\beta^2\right)\gamma\right)a_2+b_2\left(\alpha^3-\alpha\left(1+\alpha^2+\beta^2\right)\right)+\\
&e_2\left(2\alpha^2\beta-\beta\left(1+\alpha^2+\beta^2\right)\right)+i_2\left(\left(1+\alpha^2+\beta^2\right)-2\alpha^2\right)+\\
&a_2\left(\alpha^2\mu_2\sqrt{1+\alpha^2+\beta^2}-\mu_2\left(1+\alpha^2+\beta^2\right)^{3/2}\right)+2\alpha\beta\gamma c_2+\alpha\beta^2 d_2-2\alpha\gamma f_2+\\
&\alpha h_2-2\alpha\beta g_2+\alpha\beta\mu_2\sqrt{1+\alpha^2+\beta^2}-f_2\alpha\mu_2\sqrt{1+\alpha^2+\beta^2}\Bigg), \qquad (21)
\end{aligned}
$$

similarly, the second polynomial

$$
\begin{aligned}
&\sigma_2^2\Bigg(-\frac{\mathbb{N}_1\left(-2\beta\gamma^2-2\beta\gamma\mu_1\sqrt{1+\alpha^2+\beta^2}\right)}{2}+\\
&c_1\left(2\beta^2\gamma-\gamma\left(1+\alpha^2+\beta^2\right)\right)+e_1\left(2\alpha\beta^2-\left(1+\alpha^2+\beta^2\right)\alpha\right)+\\
&d_1\left(\beta^3-\left(1+\alpha^2+\beta^2\right)\beta\right)+g_1\left(-2\beta^2+\left(1+\alpha^2+\beta^2\right)\right)+\\
&c_1\left(\beta^2\mu_1\sqrt{1+\alpha^2+\beta^2}-\mu_1\left(1+\alpha^2+\beta^2\right)^{3/2}\right)+2\alpha\beta\gamma a_1+\alpha^2\beta b_1-2\beta\gamma f_1+\\
&\beta h_1-2\alpha\beta i_1+\alpha\beta\mu_1\sqrt{1+\alpha^2+\beta^2}-\beta\mu_1 f_1\sqrt{1+\alpha^2+\beta^2}\Bigg)+\\
&\sigma_1^2\Bigg(-\frac{\mathbb{N}_2\left(-2\beta\gamma^2-2\beta\gamma\mu_2\sqrt{1+\alpha^2+\beta^2}\right)}{2}+\\
&c_2\left(2\beta^2\gamma-\gamma\left(1+\alpha^2+\beta^2\right)\right)+e_2\left(2\alpha\beta^2-\left(1+\alpha^2+\beta^2\right)\alpha\right)+\\
&d_2\left(\beta^3-\left(1+\alpha^2+\beta^2\right)\beta\right)+g_2\left(-2\beta^2+\left(1+\alpha^2+\beta^2\right)\right)+\\
&c_2\left(\beta^2\mu_2\sqrt{1+\alpha^2+\beta^2}-\mu_2\left(1+\alpha^2+\beta^2\right)^{3/2}\right)+2\alpha\beta\gamma a_2+\alpha^2\beta b_2-2\beta\gamma f_2+\\
&\beta h_2-2\alpha\beta i_2+\alpha\beta\mu_2\sqrt{1+\alpha^2+\beta^2}-\beta\mu_2 f_2\sqrt{1+\alpha^2+\beta^2}\Bigg), \qquad (22)
\end{aligned}
$$

and the third one as

$$
\begin{aligned}
&\sigma_2^2\left(-\frac{\mathbb{N}_1\left(2\gamma+2\mu_1\sqrt{1+\alpha^2+\beta^2}\right)}{2}-\alpha a_1-\beta c_1+f_1\right)+\\
&\sigma_1^2\left(-\frac{\mathbb{N}_2\left(2\gamma+2\mu_2\sqrt{1+\alpha^2+\beta^2}\right)}{2}-\alpha a_2-\beta c_2+f_2\right). \qquad (23)
\end{aligned}
$$

The unknowns are the model parameters of the linear model to be fitted (α, β, and γ), while the others are known (constant) parameters, partly computed from data points and partly via EM algorithm as $(\mu_1, \mu_2, \sigma_1, \sigma_2)$. To solve this system using numerical Groebner basis is feasible, but to solve it in symbolic way is very difficult. However, Dixon's method implemented using Early Discovery of Factors heuristic algorithm can be applied (see, e.g., Lewis et al., 2014).

5. The Proposed Algebraic Solution

The steps of the algorithm are as follows (see the flow-chart in Fig. 4):

1) Step 1: Employ likelihood function developed for least squares in section 2 using Eq. 8.

2) Step 2: Having the values of the parameters, compute the model error distribution.

3) Step 3: Employing EM algorithm, compute the parameters of the Gaussians representing the two components in the mixture (see Eqs. 9-19).

4) Step 4: Using likelihood function developed for Gaussian mixture, see Eq. 21 in section 4, compute the new model parameters via numerical Groebner basis.

5) Step 5: Repeat steps 2, 3 and 4 above until the change of the values of the model parameters (α, β, and γ) are less than a given threshold of the error limit.

6. Illustrative Example

The application of the proposed algorithm is illustrated by fitting a plane to a slope having dense vegetation represented by real laser scanner data set. Outdoor laser scanning measurements were carried out on a hilly Park in Budapest (Hungary) using a Faro Focus 3D terrestrial laser scanner (Fig. 5, left). The test area was on a steep slope covered by dense but low vegetation (Fig. 5, right). The vegetation are bushes, which are natural part of the slope side and low compared to trees. The test also aimed at investigating tie points' (i.e., markers with known positions and sizes) detection capability of the scanner's processing software. This necessitated the deployment of different types of tie points (spheres in this case) all over the test area. In case of multiple scanning positions, these spheres were used for registering the point cloud (Fig. 5, left). The measurement range of the scanner is 120 m, and according to the manufacturer's technical specification, the ranging error is ± 2 mm .

The scanning parameters were set to 1/2 resolution, which equals to 3 mm/10m point spacing. The terrestrial laser scanner (Fig. 5) produced about 179 millions points acquired within five and half minutes. In order to reduce running time, the number of these points were reduced (automatically) to 33292 via random but proportional extraction, saving the original structure of data set. The final data set comprised of 33292

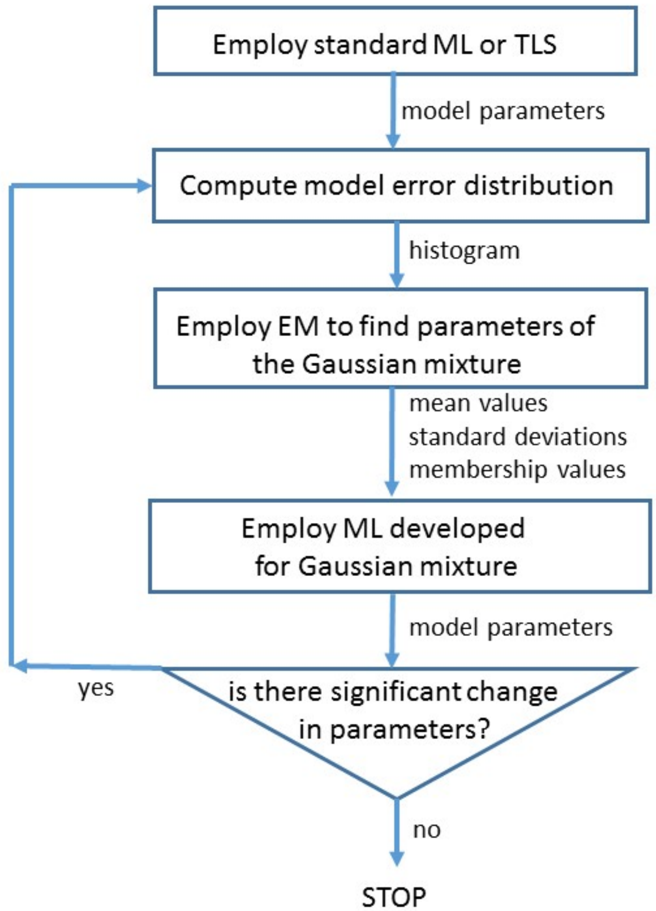

Figure 4: Steps of the iterative algorithm with implemented algebraic solution (see Eq. 21) in the fourth step.

points in ASCII format, where only the x, y, z coordinates were kept (no intensity values). The measured coloured laser scanning point cloud and the extracted test point cloud are presented in Fig. (6). Figure 7 shows the results of the different iteration steps. Details of the implementation of the proposed algorithm in *Mathematica* can be found in Paláncz (2014).

Tables 1 and 2 show the numerical results of the iteration process. In Table 1, the changing parameter values of the Gaussian components can be seen, i.e., the mean value (μ), standard deviation (σ) as well as the number of the data points (η, $\mathbb{N}$) belonging to inliers and outliers respectively. Concerning this last parameter, in reality, more information is known, since the identity (membership function) of every data point is also provided by EM algorithm. Table 2 shows the progress of the corresponding parameter value of the linear model (here, a plane). The convergence of the determined parameters after the 11th iteration is noticeable. Table 3 shows the computation times of the global maximization of the likelihood function with Gaussian mixture (Eq. 20) in every iteration step. It can be seen that the global algebraic solution of Eq. (21) using numerical Groebner basis is faster in nearly every step than the stochastic global optimization techniques, which can never find a truly global optimal solution but only an improved solution to truly global optimization methods of deterministic types, see e.g., Xu (2003). A comparison of the algebraic solution to those of three robust estimation methods in Table 4 indicates that the algebraic method had the smallest maximum error and standard deviation.

Figure 5: The test area covered by dense, but low vegetation. The left image indicates the tie points (i.e., reference points with known positions and radiuses) in white and the scanner while the right image shows the test area.

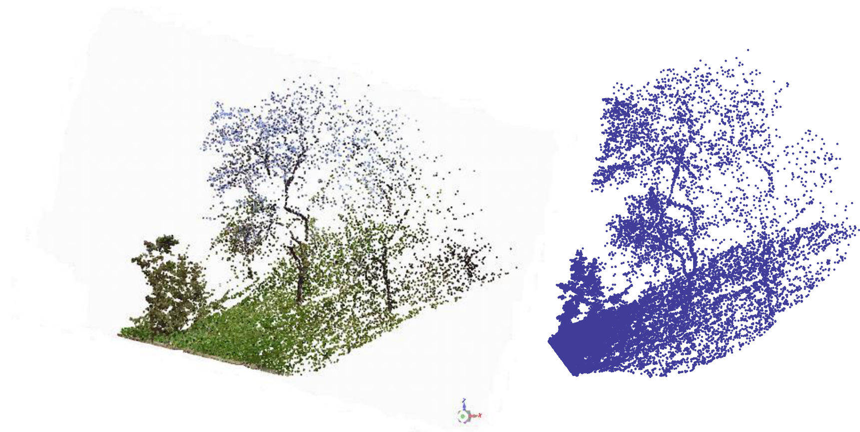

Figure 6: The colored laser scanning point cloud (left) and the extracted test point clouds (right).

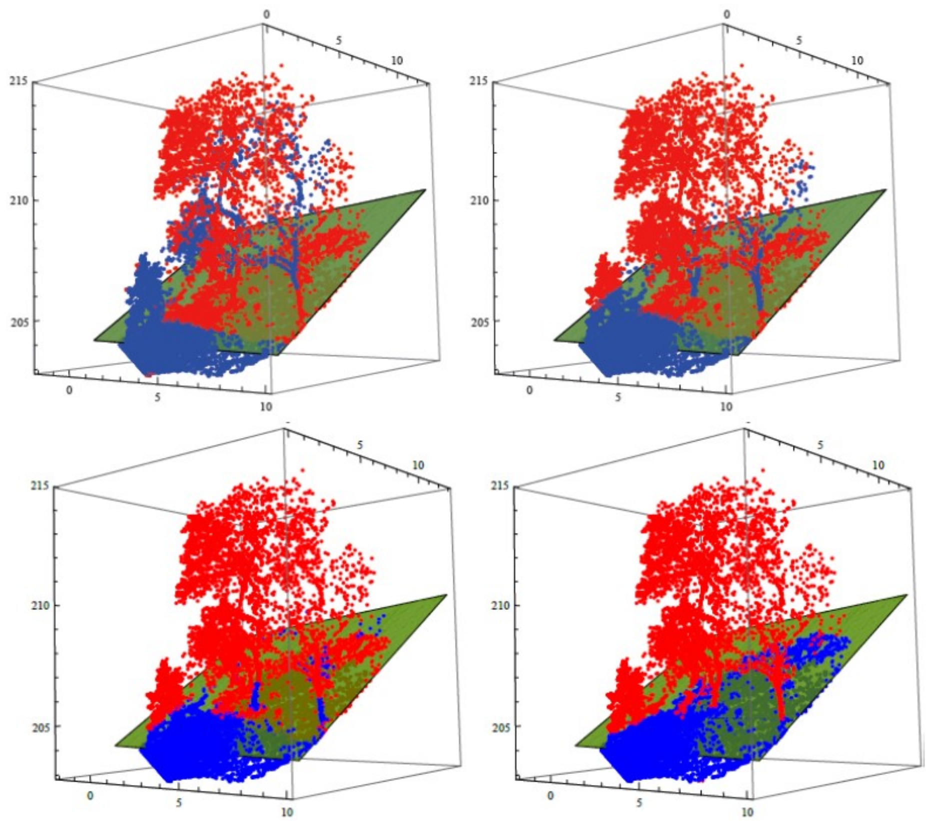

Figure 7: Isolation of outliers (red points) in the subsequent iteration steps leaving the desired good data (inliers, blue points). First iteration (top, left), second iteration (top, right), third iteration (bottom, left), and fourth iteration (bottom, right). After the fourth iteration (see the convergence in Table 2), it can be seen that the proposed algorithm successfully isolates the red outlying points.

Table 1: Parameters of the two-component Gaussian during the iteration processing steps.

	μ_1	σ_1	μ_2	σ_2	$\mathbb{N}_1$	$\mathbb{N}_2$
1	−0.262556	1.991	0.183204	0.441901	9959	23333
2	−0.0116018	2.04777	0.107134	0.403876	10192	23100
3	0.263474	2.11406	0.038595	0.369776	10357	22935
4	0.560364	2.16896	−0.0120112	0.317158	10808	22484
5	0.843549	2.22209	−0.0527925	0.265089	11195	22097
6	1.14661	2.27333	−0.0889218	0.216683	11529	21763
7	1.51548	2.32325	−0.119917	0.156959	11696	21596
8	1.87306	2.359	−0.140957	0.109529	11497	21795
9	2.2991	2.35912	−0.157437	0.0922262	10621	22671
10	2.92102	2.2328	−0.173098	0.0849716	8868	24424
11	2.97683	2.24941	−0.174499	0.0651785	9000	24292
12	2.97717	2.25054	−0.174496	0.0647087	9009	24283
13	2.97713	2.25058	−0.174432	0.0646964	9009	24283
14	2.9772	2.25058	−0.174368	0.0646943	9009	24283
15	2.97727	2.25059	−0.174304	0.0646923	9009	24283

Table 2: Model parameters during the iteration processing steps

	α	β	γ
1	0.427919	1.41389	199.663
2	0.359091	1.23481	200.339
3	0.306114	1.06996	200.908
4	0.265764	0.93406	201.363
5	0.226564	0.820129	201.752
6	0.181955	0.708478	202.141
7	0.149965	0.621019	202.436
8	0.131545	0.567046	202.615
9	0.119698	0.535416	202.728
10	0.111722	0.512999	202.815
11	0.111727	0.511941	202.817
12	0.111777	0.51188	202.817
13	0.111775	0.511876	202.817
14	0.111774	0.511873	202.817
15	0.111772	0.511869	202.817

Table 3: Comparison of the computation times of each iteration steps (in sec) in the algebraic solution and in the direct global optimization of the likelihood function.

i	Algebraic Solution	Global Optimization
1	0.2184	0.2340
2	0.1248	0.2184
3	0.1248	0.2184
4	0.2652	0.2340
5	0.1560	0.2184
6	0.1240	0.2340
7	0.1248	0.2184
8	0.1716	0.2184
9	0.1248	0.2340
10	0.1404	0.2184
11	0.1872	0.2340
12	0.1404	0.2340
13	0.1404	0.2184
14	0.1560	0.2340
15	0.1872	0.2340

Table 4 Results of the computation in case of real data obtained from laser scanning of the test area in Fig. 5.

Method	Number of Inlier *Set*	α	β	γ	Min of error (cm)	Max of error (cm)	Standard deviation (cm)
RANSAC	24382	0.106	0.503	202.66	−22.4	28.3	6.4
Danish	24576	0.106	0.505	202.66	−22.0	37.0	7.0
PCA	26089	0.103	0.567	202.54	−46.0	94.6	18.6
Algebraic solution	24283	0.107	0.503	202.66	−22.0	25.0	6.2

7. Conclusion

This study has presented an iterative algorithm using an embedded algebraic solution for the parameter estimation of a linear model in cases where the distribution of model error does not follow the criteria of a distribution of Gaussian with zero mean. To find the model parameters of a linear model, one can employ ML estimation developed for a two component Gaussian mixture. The maximization problem of this likelihood function can be converted into the task of solving a multivariate polynomial system. In order to obtain the parameters of the Gaussian distributions, EM algorithm was employed. To demonstrate the suggested algorithm, an outdoor area was laser scanned; with the acquired point cloud consisting of both inliers (i.e., points reflecting from the ground) and outliers (i.e., points reflecting from vegetation). The results were compared to those of robust estimation methods; RANSAC, Danish and PCA. The results indicate that the quality of the parameter estimation from the proposed algebraic method - smaller maximum value and standard deviation of the fitting error - proved to be better. Future studies will consider heterogeneous data originating from different sources resulting in different error distributions as another possible application of the suggested algorithm.

Acknowledgement

J. Awange is grateful for the Alexander von Humboldt Foundation, which provided the financial support that facilitated his stay at the Geodetic Institute, Karlsruhe Institute of Technology (KIT), Germany and Japan Society of Promotion of Science (JSPS), which provided the financial support that facilitated his stay at the Geophysics Institute, Kyoto University, Japan. B. Paláncz is grateful to the visiting grant provided by TIGeR and The Department of Spatial Science, Curtin University (Australia).

8. References

Agnello, F. and Lo Brutto M., " Integrated surveying techniques in cultural heritage documentation" , In: Remondino, F. and El-Hakim, S. (eds., 2007): Proceedings of the 2nd ISPRS International Workshop 3DARCH 2007, ETH Zurich, International Archives of Photogrammetry, (2007). www.isprs.org/proceedings/XXXVI/5-W47/pdf/agnello_lobrutto.pdf

Aitkin M and Wilson GT (1980) Mixture models, outliers and The EM algorithm. Technometrics 22: 325-331

Awange JL, Paláncz B (2016) Geospatial algebraic computation. Theory and applications. Springer. Heidelberg.

Awange JL, Paláncz B, Lewis R (2014) Maximizing likelihood function for parameter estimation in point clouds via Groebner Basis. In Mathematical Software – ICMS 2014, ed. Hong, Hoon, Yap, Chee, 359-366. Berlin: Springer.

Awange JL, Grafarend EW (2005) Solving algebraic computational problems in geodesy and geoinformatics. Springer, Berlin

Awange JL, Grafarend EW, Paláncz B, Zaletnyik P. (2010) Algebraic geodesy and geoinformatics. Springer, Berlin

Batselier K, Dreesen P and De Moor B (2012) Maximum Likelihood Estimation and Polynomial System Solving, ESANN 2012 Proceedings, Bruges (Belgium) pp. 369-374

Borre K (2006) Mathematical foundation of geodesy. Selected papers of Torben Krarup. Springer, Berlin, pp. 289-299.

Borrmann D, Elseberg J, Lingemann K, Nüchter A (2011) The 3D Hough transform for plane detection in point clouds: A review and a new accumulator design, 3D Res. 02, 02003, 3DR EXPRESS.

Buot M-L G, Richards (2006) Counting and locating the solution of polynomial systems of maximum likelihood equation, I., J of Symbolic Computation, Vol. 41., pp.234-244

Chen CC, Stamos I (2007) Range image segmentation for modeling and object detection in urban scenes, 3DIM2007.1

Diebel JR, Thrun S and Brunig M (2006) A Bayesian method for probable surface reconstruction and decimation, ACM Transactions on Graphics (TOG).

Donoho DL and Huber PJ (1983) The notation of breakdown point. In:Bickel PJ, Doksum KA and Hodges JL Jr (eds) A Festschrift for Erich L. Lehmann Wadsworth, Belmont, pp 157-184

Drton M (2006) Computing all roots of the likelihood equatons of seemingly unrelated regression, J. of Symbolic Computations, 41, 2, pp245-254.

Eicker et al (2014) Calibration data assimilation approach for integrating GRACE data into the WaterGAP global hydrology model (WGHM) using an ensemble Kalman filter: first results, Sureys in Geophysics, 35, pp. 1285-1309

Elbern H. (2015) Data Assimilation Basics, esa Dragon Advanced Training Course in Atmossphere Remote Sensing, https://earth.esa.int/dragon/D5_L1_Elbern.pdf, downloaded 21 03 2015

Grafarend E, Awange JL (2012) Applications of linear and nonlinear models. Springer, Berlin, Heidelberg.

Guttman, I., and Lin, D.K.J. (1995) Robust recursive estimation for correlated observations. Statistics and Probability 33: 79-92.

Hampel FR, Ronchetti E, Rousseeuw PJ and Stahel WA (1986) Robust statistics: approach based on influence functions. Wiley, New York

Hastie T, Tibshirani R (2008) Gaussian mixture models, Lecture for Statistical Learning and Data Mining, Stanford Uni. http://statweb.stanford.edu/ tibs/stat315a/LECTURES/em.pdf

Huang CM, Tseng YH (2008) Plane fitting methods of Lidar point cloud, Dept. of Geomatics, National Cheng Kung Uni. Taiwan.

Huber PJ (1964) Robust estimation of a local parameter. Ann Math Stat 35:73-101

Koch KR (2013) Robust estimation by expectation maximization algorithm. J Geod 87:107-116

Koch KR and Kargoll B (2013) Expectation maximization algorithm for the variance-inflation model by applying the t-distribution. J Appl. Geod. 7:217-225

Koch KR (2014) Robust estimations for the nonlinear Gauss Helmert model by the expectation maximization algorithm. J Geod 88: 263-271

Krarup T, Kubik K, Juhl J (1980) Götterdammerung over least squares. In: Proceedings of Int. Soc. for Photogrammetry 14-th Congress, Hamburgpp. 370-378

Lange KL Little RJA and Taylor JMG (1989) Robust statistical modeling using the t-distribution. J Am Statist Ass 84(408):881-896

Lakaemper R, Latecki LJ (2006) Extended EM for planar approximation of 3D Data, IEEE Int. Conf. on Robotics and Automation (ICRa), Orlando, Florida, May 2006.

Lewis R, Paláncz B, Awange JL (2014) Application of Dixon resultant to maximization of the likelihood function of Gaussian mixture distribution. Applied Computer Algebra Conference, New York, USA,

Luxen M and Brunn A (2003) Parameterschatzung aus unvollstandigen Beobachtungsdaten mittels des EM-Algorithmus. ZfV-Z Geodasie, Geoinformation und Landmanagement 128:71-79

Paláncz B (2014) Algebraic robust estimation employing expectation maximization, e-publication, Wolfram Research, Wolfram Information Center, Mathematica Technology, Applied Mathematics, Numerical Methods, http://library.wolfram.com/infocenter/MathSource/8887

Peng J (2009) Jointly robust estimation of unknown parameters and variance components based on Expectation-Maximization algorithm. J Surv Eng 135:1-9

Riel B, Simons M, Agram P and Zhan Z (2014) Detecting transient signals in geodetic time series using sparse estimation techniques, Journal of Geophysical Research: Solid Earth, Research Article, 10.1002/2014JB011077, pp. 1-21.

Rose C (2000) Symbolic maximum likelihood estimation with *Mathematica*, The Statistician 49, Part 2, pp.229 -240

Rousseeuw PJ (1984) Least median of squares regression. J Am Stat Assoc 79:871-880

Russeeuw PJ, Van Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator, Technometrics, Vol. 41. No. 3. pp. 212-223.

Sun, J., Johnson, K.M., Cao, Z., Shen, Z., Bürgmann, R. and Xu, X (2011) Mechanical constraints on inversion of coseismic geodetic data for fault slip and geometry: Example from InSAR observation of the 6 October 2008 Mw 6.3 Dangxiong-Yangyi (Tibet) earthquake. Journal Of Geophysical Research 116: B01406, doi:10.1029/2010JB007849.

Uhler C (2011) Geometry of maximum likelihood estimation in Gaussian graphical models, PhD. dissertation, California, Berkeley

Xu, P (1998) Mixed Integer Geodetic Observation Models and Integer Programming with Applications to GPS Ambiguity Resolution. Journal of the Geodetic Society of Japan 44 (3): 169-187.

Xu P (2005) Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. J Geod 79:146-159

Xu P, Liu J and Shi C (2012) Total least squares adjustment in partial errors-in-variables models: algorithm and statistical analysis. J Geod 86:661-675

Xu P (2003) Numerical solution for bounding feasible point sets. J Comp Appl Math 155:423-446

Yaniv Z (2010) Random Sample Consensus (RANSAC) Algorithm, A Generic Implementation. Georgetown University Medical Center, Washington, DC, USA, http://yanivresearch.info/writtenMaterial/RANSAC.pdf [Accessed 29/5/2014]

Zuliani M (2012) RANSAC for Dummies, vision.ece.ucsb.edu/ ~zuliani.