1. Denotations and assumptions

N – number of points;
$\mathbf{a}$ – r element vector of the parameters sought;
$\boldsymbol{\xi}$ – n element vector of all the quantities measured (true values, not known); when w denotes the number of dimensions of each measuring point and N denotes the number of points then $n = wN$;
if the measuring points are two-dimensional, the quantities may be sorted in the way that ξ_{2j} and ξ_{2j+1} denote the first and the second coordinate of the point with the number j;
$\mathbf{x}$ – n element vector of all the measurement results, sorted in the same way as the vector $\boldsymbol{\xi}$; (i.e. if there are 10 two-dimensional measuring points, n=10· 2=20),
$\boldsymbol{\varepsilon}$ – n element vector of measurement errors, not known:

$$\mathbf{x}=\boldsymbol{\xi}+\boldsymbol{\varepsilon} \tag{1}$$

Expected value and standard deviation of ε_j is equal

$$E(\varepsilon_j)=0, \quad \sigma^2(\varepsilon_j)=\sigma_j^2 ; \quad \text{for } j=1, \ldots,n. \tag{2}$$

$\mathbf{Cx}$ – $n\times n$-dimensional covariance matrix (σ_j^2 are the diagonal elements of this matrix),
$\mathbf{Gx} = \mathbf{Cx}^{-1}$.
$\mathbf{a_o}$ – the first approximation of the parameter vector $\mathbf{a}$,
$\boldsymbol{\xi o}$ – the first approximation of the vector $\boldsymbol{\xi}$ of the quantities measured; we also give:

$$\boldsymbol{\xi o} = \mathbf{x}. \tag{3}$$

We assume that the parameters $\mathbf{a}$ and true values $\boldsymbol{\xi}$ of the quantities measured are connected with each other by a system of m equations:

$$f_k(\mathbf{a},\boldsymbol{\xi}) = 0 , \quad \text{for } k=1, \ldots,m. \tag{4}$$

Each function f_k must be obtained by carrying over all the expressions and all the variables to one side of the equal sign. For the simple case, most often occurring, when one function F is applied to each two-dimensional measuring point, the function $f_k=F(\mathbf{a}, \xi_{2k} , \xi_{2k+1})$ depends on the entire parameter vector $\mathbf{a}$ and on the coordinates ξ_{2k} and ξ_{2k+1} of point with the number k only. In such case, the number of measuring points is equal the number of functions, i.e.

$$m = N \qquad \text{(not always !!!)}$$

However, we can imagine some more complicated situations, when there are many functions F with different formulae for different sets of points (N=m but the function $f_k=F_i(\mathbf{a}, \xi_{2k} , \xi_{2k+1})$ may not depend on the entire parameter vector $\mathbf{a}$) or for one measuring point there is more than one function connecting its coordinates (N<m, for example one point is common for the two point sets with two different functions F_l) or the function f_k enclose more than one measuring point (the last case is not included in LSM program).
We also assume that each function f_k is sufficiently linear in the neighbourhood of ($\mathbf{a_o}$, $\boldsymbol{\xi o}$). The radius of the neighbourhood of ξo_j is ε_j and therefore the errors $\boldsymbol{\varepsilon}$ must be small.

2. The method

Let each function f_k be expanded in the power series with an accuracy of the first power:

$$f_k(\mathbf{a},\boldsymbol{\xi}) \approx f_k(\mathbf{a_o},\boldsymbol{\xi_o}) + \left.\frac{\partial f_k}{\partial a_1}\right|_{\mathbf{ao},\boldsymbol{\xi o}} (a_1 - a_{o_1}) + \ \ldots \ + \left.\frac{\partial f_k}{\partial a_r}\right|_{\mathbf{ao},\boldsymbol{\xi o}} (a_r - a_{o_r}) +$$

$$+ \left.\frac{\partial f_k}{\partial \xi_1}\right|_{\mathbf{ao},\boldsymbol{\xi o}} (\xi - \xi_{o_1}) + \ \ldots \ + \left.\frac{\partial f_k}{\partial \xi_n}\right|_{\mathbf{ao},\boldsymbol{\xi o}} (\xi_n - \xi_{o_n}) . \tag{5}$$

Using the above expansion, the system of m equations (4) may be denoted by matrices **A**, **B** and vectors $\boldsymbol{\alpha}$, $\boldsymbol{\delta}$ and **c**:

$$\mathbf{A}\boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\delta} + \mathbf{c} \approx 0, \tag{6}$$

where

$$A_{k,l} = \left.\frac{\partial f_k}{\partial a_l}\right|_{\mathbf{ao},\boldsymbol{\xi}\mathbf{o}}, \quad \text{for } k=1, \ldots,m, \quad l=1, \ldots,r, \tag{7}$$

$$B_{k,j} = \left.\frac{\partial f_k}{\partial \xi_j}\right|_{\mathbf{ao},\boldsymbol{\xi}\mathbf{o}}, \quad \text{for } k=1, \ldots,m, \quad j=1, \ldots,n, \tag{8}$$

$$c_k = f_k(\mathbf{a_o}, \boldsymbol{\xi}_\mathbf{o}) , \tag{9}$$

$$\boldsymbol{\alpha} = \mathbf{a} - \mathbf{a_o} , \qquad \boldsymbol{\delta} = \boldsymbol{\xi} - \boldsymbol{\xi}_\mathbf{o} . \tag{10}$$

From the **maximum likelihood method** it follows that the estimates of the parameters **a** and the estimates of the quantities measured $\boldsymbol{\xi}$ can be found from equation

$$M = \boldsymbol{\delta}^\mathrm{T}\mathbf{G_x}\boldsymbol{\delta} = \min. \tag{11}$$

If measuring points are two-dimensional and the covariance matrix $\mathbf{C_x}$ consists of the diagonal elements (variances) only, the minimum function M is given by:

$$M = \sum_{k=1}^{m}\left[\left(\frac{\xi_{2k} - \xi_{o_{2k}}}{\sigma_{2k}}\right)^2 + \left(\frac{\xi_{2k+1} - \xi_{o_{2k+1}}}{\sigma_{2k+1}}\right)^2\right] = \sum_{v=1}^{N}\left[\left(\frac{x_v - xo_v}{\sigma_{x_v}}\right)^2 + \left(\frac{y_v - yo_v}{\sigma_{y_v}}\right)^2\right] \tag{12}$$

Applying the **Lagrange multipliers method**, instead of function $M=\boldsymbol{\delta}^\mathrm{T}\mathbf{G_x}\boldsymbol{\delta}$, we can use the Lagrange function

$$L = \boldsymbol{\delta}^\mathrm{T}\mathbf{G_x}\boldsymbol{\delta} + 2\boldsymbol{\mu}^\mathrm{T}(\mathbf{A}\boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\delta} + \mathbf{c}) \tag{13}$$

as a minimum function, where $\boldsymbol{\mu}$ is an m-dimensional vector of Lagrange multipliers. Both the functions M and L differ by the addend approximately equal zero, according to Eq. (6). The addend is exactly equal zero when the functions f_k in Eqs. (4) are linear or the initial values $\mathbf{a_o}$, $\boldsymbol{\xi}\mathbf{o}$ are selected to be equal the true values **a**, $\boldsymbol{\xi}$, according to Eqs. (9) and (10).

Equating the function L derivatives with zero, with respect to both $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$, we obtain two equations:

$$\mathbf{G_x}\boldsymbol{\delta} + \boldsymbol{\mu}^\mathrm{T}\mathbf{B} = \mathbf{G_x}\boldsymbol{\delta} + \mathbf{B}^\mathrm{T}\boldsymbol{\mu} = 0 \tag{14}$$

$$\boldsymbol{\mu}^\mathrm{T}\mathbf{A} = \mathbf{A}^\mathrm{T}\boldsymbol{\mu} = 0. \tag{15}$$

We can evaluate $\boldsymbol{\delta}$ from Eq. (14):

$$\boldsymbol{\delta} = -\,\mathbf{G_x}^{-1}\mathbf{B}^\mathrm{T}\boldsymbol{\mu} , \tag{16}$$

insert it to Eq. (6) and then evaluate $\boldsymbol{\mu}$**:**

$$\boldsymbol{\mu} = \mathbf{G_B}(\mathbf{A}\boldsymbol{\alpha}+\mathbf{c}), \tag{17}$$

where

$$\mathbf{G_B} = (\mathbf{B}\mathbf{C_x}\mathbf{B}^\mathrm{T})^{-1}. \tag{18}$$

Inserting $\boldsymbol{\mu}$ from Eq. (17) to Eq. (15) yields estimate $\boldsymbol{\alpha}$

$$\boldsymbol{\alpha} = -\,\mathbf{C_a}\mathbf{A}^\mathrm{T}\mathbf{G_B}\mathbf{c}, \tag{19}$$

where

$$\mathbf{C_a} = (\mathbf{A}^\mathrm{T}\mathbf{G_B}\mathbf{A})^{-1}. \tag{20}$$

After insertion $\boldsymbol{\alpha}$ from Eq. (19) to Eq. (17) and then to Eq. (16), we obtain estimate $\boldsymbol{\delta}$

$$\boldsymbol{\delta} = - \mathbf{C_x B^T G_B (A\alpha + c)} = - \mathbf{C_x B^T G_B (c - A\, C_a A^T G_B c)} \tag{21}$$

We get the estimates of parameters and quantities measured from equations

$$\tilde{\mathbf{a}} = \mathbf{a_o} + \boldsymbol{\alpha} \tag{22}$$

$$\tilde{\boldsymbol{\xi}} = \mathbf{x} + \boldsymbol{\delta} \tag{23}$$

If equations (4) are linear, the solutions (22) and (23) are obtained after one iteration step, otherwise we continue the iteration, setting $\tilde{\mathbf{a}}$ and $\tilde{\boldsymbol{\xi}}$ as new starting vectors $\mathbf{a_o}$ and $\boldsymbol{\xi_o}$.
Covariance matrixes of the estimates of $\mathbf{a}$ and $\boldsymbol{\xi}$ can be found after the last iteration step from the law of error transfer:

$$\mathbf{C}_{\tilde{\mathbf{a}}} = (\mathbf{A^T G_B A})^{-1} \tag{24}$$

$$\mathbf{C}_{\tilde{\xi}} = \mathbf{C_x (I + B^T G_B (-I + A C_a A^T G_B) B C_x)} \tag{25}$$

From the **maximum likelihood method** it follows that if each function f_k is sufficiently linear in the neighbourhood of ($\mathbf{a_o}$, $\boldsymbol{\xi o}$), and the errors $\boldsymbol{\varepsilon}$ have the normal distribution, then function

$$\tilde{M} = \left(\mathbf{B}\tilde{\boldsymbol{\varepsilon}}\right)^{\mathrm{T}} \mathbf{G_B} \left(\mathbf{B}\tilde{\boldsymbol{\varepsilon}}\right), \qquad \text{where} \quad \tilde{\boldsymbol{\varepsilon}} = \mathbf{x} - \tilde{\boldsymbol{\xi}} \tag{26}$$

have the chi-square distribution with degree of freedom m-r. Therefore expected value and standard deviation of M are equal:

$$\mathrm{E}(M) = m\text{-}r, \qquad \sigma(M) = \sqrt{2\left(m - r\right)}\,.$$

The last fact may be used as a criterion of quality of fitting. The value $\tilde{M}$ computed should not differ from E(M) more than $\sigma(M)$. Or more precisely, the probability that M would be less then the value $\tilde{M}$ computed should be less than the confidence level assumed.
However, the other practice is possible. If we are sure of the ratios of measuring uncertainties only instead of the uncertainties themselves, we can enter any uncertainties to computation taking into account these ratios, and then calculate the estimates of uncertainties of points and parameters by dividing the uncertainties by the square root of E(M)/M.