# Alternative to Lagrange Multiplier Method

Kurbatov V.S.

Joint Institute for Nuclear Reaearch, JINR, RU-141980, Dubna, Russia

March 2024

**Аннотация**

In particle physics data analysis the so-called Lagrange Multiplier Method has been used for many years. It has been implemented in the 1960s by the famous Alvarez group for processing experimental data. Since then it is widely used in physical community. It is named after Lagrange who proposed the method for finding the minimum of functions of many variables under the requirement that they satisfy to some additional conditions (equalities, inequalities). The method uses some artificial variables called Lagrange multipliers having no physical meaning.

Another approach is described here, to find the minimum of a function (in our case it is either $\chi^2$ or logarithm of Likelihood Function) with the constraints. The proposed method is based on the linearization of the constraints during a suitable iteration procedure for the search for the minimum.

We propose a new method for selecting submatrices of partial derivatives Jacobi matrix in this paper.

## 1 Introduction

"The history of constrained optimization spans nearly three centuries. The principal warhorse, Lagrange multipliers, was discovered by Lagrange in the Statics section of his famous book on Mechanics from 1788 [1], by applying the idea of virtual velocities to problems in statics with constraints. The idea of virtual velocities, in turn, goes back to a letter of Johann Bernoulli from 1715 to Varignon, in which he announced a very simple rule for solving hundreds of Varignon's problems in the blink of an eye. Varignon then explains this rule in his book published in 1725 [2]. Half a century later, Bernoulli's rule was chosen by Lagrange as the general principle for the foundation of his mechanics, with the multipliers as the main tool for treating mechanical constraints"[3].

In particle physics data analysis Lagrange Multiplier Method has been used for many years. For processing experimental data it has been implemented in the 1960s by the famous Alvarez group [4, 5].

The problem may be formulated as follows [4, 5]: find such values of kinematical parameters $x_i$, $i = 1, \ldots, n_p$, that turn $\chi^2(\boldsymbol{x})$ (1) into the minimum

$$\chi^2(\boldsymbol{x}) = \sum_{i=1,\,j=1}^{n_p,\,n_p} (x_i - x_i^m)\; g_{i,j}\; \left(x_j - x_j^m\right) \tag{1}$$

and satisfy conservation law equations (constraints) (2)

$$f_\lambda(\boldsymbol{x}) = 0; \qquad \lambda = 1, \ldots, n_c, \tag{2}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n_p})^T$ is the column vector of kinematical parameters $x_i$, $n_p$ is their number, $\boldsymbol{G} = (g_{i,j})$ is the matrix, inverse to the error matrix, $x_j^m$ are the measured values of the parameters, and $n_c$ is the number of conservation law equations.

The authors have shown that if $x_j^m$ are distributed according to the Gauss law and $\boldsymbol{x}$ satisfy (2), form (1) has the $\chi^2$ distribution with the number of degrees of freedom ($n_{\mathrm{df}}$) equal to $n_c$ ($n_{\mathrm{df}} = n_c$).

The authors proposed to search for the minimum of (1) using iteration scheme for the Lagrange function (Lagrange multiplier method):

$$\chi^2(\boldsymbol{x}) = \sum_{i=1,\,j=1}^{n_p,\,n_p} (x_i - x_i^m)\; g_{i,j}\; \left(x_j - x_j^m\right) + 2\sum_{\lambda=1}^{nc} \alpha_\lambda \cdot f_\lambda(\boldsymbol{x}). \tag{3}$$

Here $\alpha_\lambda$ are the arbitrary multiplier to be found during the search for the minimum. During iterations both $x_i$ and $\alpha_\lambda$ are varied.

Later [6] we considered a more general case where instead (3) we proposed to search for minimum (as before in Lagrange Multiplier Method), minimizing the form

$$\chi^2(\boldsymbol{x}) = \sum_{i=1,\,j=1}^{n_f,\,n_f} (c_i(\boldsymbol{x}) - c_i^m)\; q_{i,j}\; \left(c_j(\boldsymbol{x}) - c_j^m\right) \tag{4}$$

and satisfying (2). Here $c_i(\boldsymbol{x})$ are the observables, i.e., the functions of kinematical parameters $\boldsymbol{x}$, $c_i^m$ are their measured values, $\boldsymbol{Q} = (q_{i,j})$ is the inverse matrix of the observables error matrix and $n_f$ is the number of measured functions. It was shown that if errors are distributed by the Gauss law and $\boldsymbol{x}$ satisfy (2), form (5) has $\chi^2$ distribution with $n_{\rm df} = n_f - n_p + n_c$. Here $n_p$ is the number of kinematical parameters or dimension of the vector $\boldsymbol{x}$. In the case where $n_f = n_p$, $n_{\rm df} = np$ as was in the original proposal of the method.

Let us stress that that in Lagrange Multiplier Method **the dimension of the problem (or iteration procedure) is** $n_p + n_c$.

# 2 Alternative to Lagrange Multiplier Method

## 2.1 The idea of the method

At the time when we considered the so-called generalized [6] kinematical fit we came to the following: the search for the minimum of form (5)

$$\chi^2 = \sum_{i=1,j=1}^{nf,nf} (C_i(\boldsymbol{x}) - C_i^m) Q_{i,j} (C_j(\boldsymbol{x}) - C_j^m) \tag{5}$$

is usually performed by an iteration procedure, where near some initial values of the parameters $\boldsymbol{x} = \boldsymbol{x}_0$ function (5) is approximated by quadratic form

$$F(\boldsymbol{x}) = F_0 + \boldsymbol{g}^T \cdot \Delta\boldsymbol{x} + \frac{1}{2}\,\Delta\boldsymbol{x}^T \cdot \boldsymbol{Z} \cdot \Delta\boldsymbol{x}, \tag{6}$$

where $\boldsymbol{g}$ is the column vector of the derivatives and $\boldsymbol{Z}$ is the Hesse matrix of the second derivatives with respect to parameters $\boldsymbol{x}$. Near the same values $\boldsymbol{x}_0$ one can expand (2)

$$\boldsymbol{f}(\boldsymbol{x}) \cong \boldsymbol{f}(\boldsymbol{x}_0) + \boldsymbol{F}' \cdot \Delta\boldsymbol{x} = \boldsymbol{0}, \tag{7}$$

where $\boldsymbol{F}'$ is the rectangular matrix of the constraint derivatives with respect to parameters ($n_c$ rows and $n_p$ columns). Using this equation we may express some of the increments $\Delta(\boldsymbol{x})$ in terms of $\boldsymbol{f}(\boldsymbol{x}_0)$ and the matrix $\boldsymbol{F}'$ and substitute it into (6). After this substitution you get another quadratic form with the parameter vector $\boldsymbol{x}'$ having the dimension $n_p - n_c$. We have the reduction of the dimension of the problem. So it is the pure substitution method but for parameter increments! This approach was published in [7]. It was shown that under correct hypothesis average value $\overline{\chi^2} = n_f - n_p + n_c$. The expression for the error matrix of the vector $\boldsymbol{x}$ was also obtained.

Authors of [8] refer to the Reduced-Gradient-Type Methods for the solution of the minimization problems with nonlinear equalities constraints. A generalized reduced gradient (GRC) method was first proposed by J. Abadie and J. Carpentier (1965, 1969 [9]). The idea of our method is similar to the GRC approach.

We would like also to mention another approach, proposed by V.I. Moroz [10] in JINR in the time when bubble chamber was the main instrument in particle physics. He used a simple constraint accounting method by using a penalty function.

## 2.2 Necessary formulae

We may rewrite (7) as

$$\boldsymbol{f} \approx \boldsymbol{f}(\boldsymbol{x}_0) + \boldsymbol{F}'_f \cdot \Delta\boldsymbol{x}_f + \boldsymbol{F}'_c \cdot \Delta\boldsymbol{x}_c = \boldsymbol{0}. \tag{8}$$

Here $\boldsymbol{F}'_f$ and $\boldsymbol{F}'_c$ are submatrices of $\boldsymbol{F}'$, first has $n_c$ rows and $n_p - n_c$ columns, second $n_c$ rows and $n_c$ columns. We may express $\Delta\boldsymbol{x}_c$ as a function of $\Delta\boldsymbol{x}_f$ in the form

$$\Delta\boldsymbol{x}_c = \boldsymbol{r} + \boldsymbol{S} \cdot \Delta\boldsymbol{x}_f. \tag{9}$$

The subvector $\Delta\boldsymbol{x}_c$ in (6) may be changed according to (9), and we come to the new quadratic form depending only on $n_f = n_p - n_c$ increments $\Delta\boldsymbol{x}_f$:

$$F = \widetilde{F}_0 + \widetilde{\boldsymbol{g}}^T \cdot \Delta\boldsymbol{x}_f + \frac{1}{2}\Delta\boldsymbol{x}_f^T \cdot \widetilde{\boldsymbol{Z}} \cdot \Delta\boldsymbol{x}_f, \tag{10}$$

$$\widetilde{F}_0 = F_0 + \sum_{k=1}^{n_c} r_k \cdot \left[ g_{n_f+k} + \frac{1}{2} \cdot \sum_{l=1}^{n_c} z_{n_f+k,n_f+l}\, r_l \right], \tag{11}$$

$$\widetilde{g}_i = g_i + \sum_{k=1}^{n_c} g_{n_f+k} \cdot s_{k,i} + \sum_{k=1}^{n_c} r_k \left[ z_{nf+k,i} + \sum_{l=1}^{n_c} s_{l,i}\, z_{n_f+l,n_f+k} \right], \tag{12}$$

$$\widetilde{z}_{i,j} = z_{i,j} + \sum_{k=1}^{n_c} \left[ s_{k,i} \cdot z_{n_f+k,j} + s_{k,j} \cdot z_{i,n_f+k} \right] + \sum_{k=1,\, l=1}^{n_c,\, n_c} s_{k,i}\, z_{n_f+k,nf+l}\, s_{l,j}. \tag{13}$$

## 2.3 Selection of a submatrix $\boldsymbol{F}'_c$

There exist more principles of "correct" selection of a submatrix $\boldsymbol{F}'_c$ of $\boldsymbol{F}'$ in (8). If the selection process is defined by removing $n_p - n_c$ columns of the matrix $\boldsymbol{F}'$, we may consider, e.g., the next principles:

- select a submatrix $\boldsymbol{F}'_c$ with *maximum absolute value of the determinant* among all possible submatrices of size $n_c \times n_c$ of the matrix $\boldsymbol{F}'$ (to get the "maximal precision" in computation of vectors $\Delta\boldsymbol{x}_c$ from (8));
- remove the columns corresponding to the maximum absolute values in the upper triangle of the *correlation matrix* of $\boldsymbol{F}'$ columns (to remove the "maximal dependent" columns);
- remove the columns corresponding to the maximum absolute values in the upper triangle of the *cosine* matrix of $\boldsymbol{F}'$ columns (to keep the "maximal independent" columns).

In the general case, the last principle may be faster then the determinants evaluation. We do prefer the *maximum absolute value of the determinant* principle. Is some special cases, it is possible to determine the maximal determinant absolute value without the evaluation of all submatrices determinants. E.g., in some our computations, we have got the $4 \times 6$ matrices $\boldsymbol{F}'$ in the form:

$$\boldsymbol{F}' = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix},$$

for which the maximal possible determinant absolute value is equal to

$$|D|_{\max} = \max\left[|d_{14} - d_{11}|,\, |d_{15} - d_{12}|,\, |d_{16} - d_{13}|\right] = \max\left[|\Delta_{41}|,\, |\Delta_{51}|, |\Delta_{61}|\right] = \max_{k=4,5,6} |\Delta_{k1}|.$$

The column $k^* = \arg\max_{k=4,5,6} |\Delta_{k1}|$ is added to the columns 1, 2, and 3, to get the submatrix $\boldsymbol{F}'_c$ with the maximal absolute value of the determinant.

Finally, in this approach the dimension of the space of parameters, in which solutions are sought is $n_p - n_c$, compared with the dimension $n_p + n_c$ in the Lagrange Multiplier Method.

# 3 Realization of the method

The method was realized on the base of the algorithm of the Fumili program [11, 12]. The main idea of the Fumili algorithm is to use the approximate expression of the second derivative matrix $\boldsymbol{Z}$ in the formula (6). This approximate matrix is positively defined, which guarantees the motion to the minimum during iterations. The details of the Fumili algorithm in the simplest form are presented in [13].

## 3.1 Tutorial example on simulated data

A couple of years ago my two young colleagues asked me to give an example of usefulness of the kinematical fit, in particular of the proposed one – or in general case – constrained fit.

The suitable probability density function (pdf) was chosen as (14)

$$\text{pdf}\,(x,y) = (1+\alpha_1\cdot x+\alpha_2\cdot y)/(1+0.5\cdot\alpha_1+0.5\cdot\alpha_2), \tag{14}$$

i.e., a two-dimensional function defined on the area $0 < x < 1$ and $0 < y < 1$. The values of the parameters were $\alpha_1 = 0.5$ and $\alpha_2 = 0.8$. A constraint was just one equation, chosen in the simplest linear form (15):

$$\alpha_1+\alpha_2 = 1.3\,. \tag{15}$$

According to (14), $10^5$ events were generated and fitted in two approaches – fitting suitable likelihood function by the standard Fumili code without constraints and by the procedure proposed here.

| parameter | constrained option | unconstrained option |
|---|---|---|
| $\alpha_1$ | $0.501 \pm 0.013$ | $0.515 \pm 0.023$ |
| $\alpha_2$ | $0.799 \pm 0.013$ | $0.815 \pm 0.026$ |

Таблица 1: The values of the estimates

In Table 1, the results for the constrained and unconstrained cases are shown. Cited errors are those calculated by the program. It is clear that:

1. In both cases the estimates are within one calculated error of true values.
2. Calculated errors in the constrained option are two times smaller than in the unconstrained option.
3. Values of estimates in the constrained option are much nearer to the true ones.

## 3.2 In the processing of experimental data

In fact, the above-mentioned approach was born during our experiments in Protvino [14, 15].

The other field of application of this method was experiments in a COSY Accelerator Center (Germany).

During many years a lot of nuclear physics experiments were performed in a COSY. Many processes were measured and analyzed within this program. In particular, processes with the production of the diproton (i.e., two-proton system ($^1S_0$ state) with small binding energy) were extensively investigated. There is a lot of the papers where proposed method was used. An example is [16], where the process of $pp \to pp\,(^1S_0) + \pi^0$ was studied. It was found that the differential cross section of the production of the $^1S_0$ system in the forward direction had resonance-like behavior in the region $\sqrt{s} \approx 2.65$ GeV.

# 4 Conclusion

New approach to the estimation of particle parameters by the minimization of $\chi^2$ or logarithmic functions with the constraints is presented. Since its origination it required more than 30 years before we understood that we came to "complete" solution of a problem in a frame of original proposal. In

a sense this approach may be considered as an alternative to Lagrange Multiplier Method, proposed nearly seventy years ago and actively used since then.

We are not going to compare proposed method with Lagrange Multiplier Method in terms of computer speed, simplicity and so on. Our aim was just to show that **there is another option for the solution of similar problem!** As for its advantages or disadvantages we would prefer to do such a conclusions by the reader.

The only thing which we may say that the dimension of the problem diminishes compared with Lagrange Multiplier Method,but on the other hand there is a problem of the selection of so-called "free" parameters.

## Acknowledgement

Discussions with Dr. Ján Buša from MLIT JINR are gratefully acknowledged.

## Список литературы

[1] G.L. de la Grange (J.-L. Lagrange), *Méchanigue Analitique*, Paris, (1788)

[2] P. Varignon, *Elemens de Mathematique*, Amsterdam, (1734)

[3] *History of Lagrange Multiplier and constrained Optimization*, January 22, (2016), https://helpdeskkfc.blogspot.com/2016/01/history-of-lagrange-multiplier-and.html

[4] J.P. Berge, F.T. Solmitz, H.D. Taft, *Kinematical Analysis of Interaction Vertices From Bubble Chamber Data*, The Rev. Sci. Instr. **32**, No. 5 (1961)

[5] R. Bock, CERN 60-30 (1960)

[6] A.J. Ketikian, ..., V.S. Kurbatov, ..., *Generalized kinematical fit in event reconstruction*, NIM **A314** (1992) 572–577,

[7] A.J. Ketikian, ..., V.S. Kurbatov, *New algorithm for minimizing $\chi^2$ functionals with constraints*, NIM **A314** (1992) 578–583

[8] P.E. Gill, W. Murray, M.H. Wright, *Practical Optimization*, Academic Press, Printed in Great Britain by The Bath Press, Bath (1981)

[9] J. Abadie, J. Carpentier, *Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints*, Optimization (Sympos., Univ. Keele, Keele, 1968) (R. Fletcher, ed.), Academic Press, London, with discussion, pp. 37–47 (1969)

[10] V.I. Moroz, Preprint JINR, P-1958, Dubna (1965)

[11] S.N. Sokolov, I.N. Silin, Preprint JINR D-810, Dubna (1961)

[12] I.N. Silin, CERN Progam Library D 510, FUMILI (1983)

[13] S.N. Dymov, V.S. Kurbatov, ..., *Constrained minimization in the C++ environment*, NIM **A440** (2000) 431–437

[14] В. Артёмов, ..., В. Курбатов и др., *Определение параметров формфакторов $\lambda_+, \lambda_0$ в $K_{\mu 3}$ и $K_{e3}$- распадах, Ядерная физика,* **60** *(1997) 2205*

[15] В. Артёмов, ..., В. Курбатов и др., *Определение параметров формфакторов $\lambda_+, \lambda_0$ в $K_{\mu 3}$-распаде, Ядерная физика,* **60** *(1997) 277*

[16] V. Kurbatov, ..., et al *Observation of resonance-like behavior of the $pp \to pp_s\pi^0$ reaction around $\sqrt{s} = 2.65$ GeV*, EPJ Web of Conferences **204**, 08008 (2019)