Linear regression modeling in monitoring tasks based on the method of least absolute deviations

Cite as: AIP Conference Proceedings **2333**, 090013 (2021); https://doi.org/10.1063/5.0041859
Published Online: 08 March 2021

Alexander Tyrsin, and Aleksan Azaryan

Linear Regression Modeling in Monitoring Tasks Based on the Method of Least Absolute Deviations

Alexander Tyrsin[a)] and Aleksan Azaryan[b)]

Ural Federal University named after the First President of Russia B. N. Yeltsin, Yekaterinburg, Russia

[a)] at2001@yandex.ru
[b)] a.a.azaryan@gmail.com

Abstract. Algorithm for the exact solution of the problem of estimating the parameters of linear regression models by the least absolute deviations method is described. It is based on the descent through the nodal straight lines. This algorithm significantly outperforms other well-known methods of solving the problem and it can be effectively used in practice. The computational complexity of the descent algorithm through the nodal straight lines is assessed.

INTRODUCTION

In this paper, we consider the problem of resistant modeling of linear dependencies under stochastic heterogeneity. To create mathematical models using experimental data, for example for monitoring and diagnostic tasks, it is common to deal with stochastic heterogeneity. We will point out such features as: incomplete correspondence of some parts of the observations to the model; possible presence of outliers in samplings (not necessarily due to measurement errors); often non-experimental, heterogeneous nature of data; use of different groupings and rounding; possible dependence of the observation results [1]. One of the most common tasks in the statistical processing of experimental findings is to estimate the unknown parameters of multiple linear regression models in conditions of stochastic heterogeneity [2]:

$$\mathbf{y} = \mathbf{Xa} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$ – vector of the observable values of the dependent variable; $\mathbf{X} = \{x_{ij}\}_{n\times m} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1m} \\ 1 & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n2} & \dots & x_{nm} \end{pmatrix}$ – matrix of the observable values of explanatory variables; $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$ – random measurement discrepancy (errors); $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_m \end{pmatrix}$ – unknown parameters of multiple linear regression; n – the sample size; m – the dimension of the model.

The corresponding system of linear equations of model (1) is underdetermined (there are fewer equations than unknowns). Indeed, in addition to $\mathbf{a}$, random measurement discrepancy (errors) $\boldsymbol{\varepsilon}$ are also unknown. Therefore, it is necessary to solve the corresponding system of linear equations of model (1) statistically. Which means that we need to choose such values of unknown parameters $\mathbf{a}$ that the set of residuals

$$\varepsilon_i = y_i - \left(a_1 + a_2 x_{i2} + a_3 x_{i3} + \dots + a_m x_{im} \right), i = 1,2, \dots, n$$

Applications of Mathematics in Engineering and Economics (AMEE'20)
AIP Conf. Proc. 2333, 090013-1–090013-5; https://doi.org/10.1063/5.0041859
Published by AIP Publishing. 978-0-7354-4077-7/$30.00

can be characterized by some extremal properties. For this purpose, a certain function of residuals is usually created and minimized in the space of unknown parameters $a_1, a_2, \dots a_m$.

The most widely used method for estimating the parameters of model (1) is the ordinary least squares (OLS) method. However, the OLS estimator is consistent when the regressors are exogenous, and, by the Gauss–Markov theorem, optimal in the class of linear unbiased estimators when the errors $\boldsymbol{\varepsilon}$ are uncorrelated, have the same finite variances and expectation value of zero. Under the additional assumption that the errors are normally distributed, OLS is asymptotically efficient [2]. Nevertheless, in many problems, especially in dynamic problems of diagnostics and monitoring, we often have stochastically heterogeneous experimental data, the assumptions of the Gauss-Markov theorem are violated, and when observations contain large outliers the OLS estimates become invalid [2]. Under the circumstances, the estimation of the parameters of model (1) is required to perform using resistant methods. The most popular of which is the method of least absolute deviations (LAD) [3], which for model (1) minimizes the sum of the modules of the residuals

$$Q(\mathbf{a}) = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{m} a_j x_{ij} \right| = \sum_{i=1}^{n} |y_i - \langle \mathbf{x}_i, \mathbf{a} \rangle| \to \min_{\mathbf{a} \in \mathbf{R}^m}. \quad (2)$$

The objective function of the LAD $Q(\mathbf{a})$ is a convex, continuous, piecewise linear and bounded below function. However, it has a large number of kinks, moreover, these kinks are elongated, crossed and many of them are very close to each other, forming a so-called “bundle”. At the same time, the minimum of the function is inside this "bundle". The boundaries of these kinks are convex linear hyperplanes that gradually become almost parallel when the objective function approaches the minimum. This specificities of $Q(\mathbf{a})$ are the main reason of the problems which occur during solving task (2) by known methods (TABLE 1).

TABLE 1. Known methods of solving LAD task

Exact methods	Approximate methods
Brute-force search algorithm	Methods of searching for an unconditional zero-order extremum
Simplex-based methods	Variationally-weighted quadratic approximations’ algorithm

The brute-force search algorithm [4] enumerates all the singular points (where $Q(\mathbf{a})$ is not differentiable). This algorithm has $O\left(C_n^m \cdot (m^3 + m \cdot n)\right)$ computational complexity. Therefore, an exponential increase in computational costs associated with increases in sample size and the number of variables.

The implementation of different algorithms for finding the exact solution to task (2), based on the ideas of linear programming [5–10], is effective only for small dimensions of models and a limited sample size. The reasons are the accumulation of rounding errors and the requirement for an excessively large memory.

Due to the above-described specificities of the objective function, an increase in the accuracy of zero-order iterative optimization methods [11] leads to a rapid increase in computational costs. Another approximate method of solving task (2) is the method of iteratively reweighted least squares (Weiszfeld’s algorithm) [5, 12–13]. However, it is known that at the solution point of task (2) m residuals are equal to zero. To solve this problem, a regularization technique is used. But the question of convergence of the method of iteratively reweighted least squares remains open in the case of using a regularization.

Key Notation

Let $\boldsymbol{\Omega}$: $\{\Omega_1, \Omega_2, \dots, \Omega_n\}$ – is the set of all hyperplanes of the form

$$\Omega_i = \Omega(\mathbf{a}, \mathbf{x}_i, y_i) = y_i - \langle \mathbf{a}, \mathbf{x}_i \rangle = 0, \ (i = 1, 2, \dots, n). \quad (3)$$

Definition 1. Let us call the intersection point of m independent hyperplanes of the form (3), a nodal point:

$$\mathbf{u} = \bigcap_{s \in M} \Omega_s, \ M = \{k_1, \dots, k_m\}, \ k_1 < k_2 < \dots < k_m, \ k_l \in \{1, 2, \dots, n\}. \quad (4)$$

Let us denote U – the set of all nodal points.

Definition 2. Let us call the intersection straight line of $m-1$ independent hyperplanes of the form (3), a nodal straight line:

$$l_{(k_1,\ \dots\ ,\ k_{m-1})} = \bigcap_{i\in D} \Omega_i,\ D = \{k_1,\ \dots\ ,\ k_{m-1}\},\ k_1 < k_2 < \ \dots\ < k_{m-1},\ k_l \in \{1,\ 2,\ \dots,\ n\}.$$

It was proved in [14] that the minimum of the objective function $Q(\mathbf{a})$ belongs to the set U.

The Algorithm of Descent Through the Nodal Straight Lines

This algorithm for the exact solution of task (2) is based on the descent to the solution through the nodal straight lines.

We start the process with an arbitrary nodal point $\mathbf{u}^{(0)} = \mathbf{u}_{(k_1,\ \dots\ ,\ k_m)}$, which is the intersection of m arbitrary different hyperplanes $\Omega_{k_1}, \dots, \Omega_{k_m}$. Excluding one of the hyperplanes, we obtain a nodal straight line $l_{(k_1,\ \dots\ ,\ k_{m-1})}$ (m nodal straight lines pass through any nodal point). Then we find all the nodal points lying on this nodal straight line and sort them. At $\mathbf{u}^{(0)}$ we calculate the value of the objective function and the directional derivative of the objective function along the nodal straight line $l_{(k_1,\ \dots\ ,\ k_{m-1})}$. Then we move from the nodal point $\mathbf{u}^{(0)}$ in the direction of derivative's decreasing. This derivative is discontinuous only at nodal points. Therefore, we determine the derivative from different sides at the next nodal point $\mathbf{u}^{(1)} = (u_1^{(1)}, u_2^{(1)}, \dots, u_m^{(1)})$. Each of the derivatives is equal to the sum of n terms

$$\frac{\partial Q(\mathbf{u}^{(1)})}{\partial \vec{l}_{(k_1,\ \dots\ ,\ k_{m-1})}} = \sum_{i=1}^{n} (l_1 + x_{i2}l_2 + \dots + x_{im}l_m) \cdot \operatorname{sign}\left(\left(\sum_{j=1}^{m} u_j^{(1)} x_{ij}\right) - y_i\right), \tag{5}$$

where $\vec{l}_{(k_1,\ \dots\ ,\ k_{m-1})} = (l_1, l_2, \dots, l_m)$ is the directing vector of the nodal straight line $l_{(k_1,\ \dots\ ,\ k_{m-1})}$.

When passing through a nodal point, only one of these terms changes (the number of the term corresponds to the number of the hyperplane crossing at this nodal point $\mathbf{u}^{(1)}$ the line $l_{(k_1,\ \dots\ ,\ k_{m-1})}$). If at $\mathbf{u}^{(1)}$ the left derivative $\left.\frac{\partial Q(\mathbf{u}^{(1)})}{\partial \vec{l}_{(k_1,\ \dots\ ,\ k_{m-1})}}\right|_{\mathbf{u}_-^{(1)}}$ is negative and the right derivative $\left.\frac{\partial Q(\mathbf{u}^{(1)})}{\partial \vec{l}_{(k_1,\ \dots\ ,\ k_{m-1})}}\right|_{\mathbf{u}_+^{(1)}}$ is positive, then this point is the point of minimum of the objective function along the nodal straight line. Then instead of $\mathbf{u}^{(0)}$ we fix $\mathbf{u}^{(1)}$. If on a given line the objective function takes the minimum value at the initial nodal point, then we proceed to the next nodal straight line passing through this nodal point.

When instead of $\mathbf{u}^{(0)}$ we fix $\mathbf{u}^{(1)}$, then at $\mathbf{u}^{(1)}$ we proceed to another nodal straight line passing through it and continue descending along it according to the same principle.

As a result, a nodal point will be found, descent from which is impossible. This nodal point will be the exact solution of task (2). It has been established that the descent through the nodal straight lines always converges to an exact solution in a finite number of steps [14].

Computational Complexity of the Developed Algorithm

The computational efficiency of the descent algorithm can be improved by using sparse matrices. Indeed, moving along the nodal straight line $l_{(k_1,\ \dots\ ,\ k_{m-1})}$, in order to find the nodal points belonging to this nodal straight line, it is necessary to solve for each point a linear system of m equations:

$$\begin{cases} a_1 + a_2 x_{k_1,2} + a_3 x_{k_1,3} + \dots + a_m x_{k_1,m} = y_{k_1} \\ a_1 + a_2 x_{k_2,2} + a_3 x_{k_2,3} + \dots + a_m x_{k_2,m} = y_{k_2} \\ \dots \\ a_1 + a_2 x_{k_{m-1},2} + a_3 x_{k_{m-1},3} + \dots + a_m x_{k_{m-1},m} = y_{k_{m-1}} \\ a_1 + a_2 x_{i,2} + a_3 x_{i,3} + \dots + a_m x_{i,m} = y_i \end{cases}, \tag{6}$$

where $k_1 < k_2 < \ \dots\ < k_{m-1}, i \in \{1, 2, \dots\ ,\ n\}, i \notin \{k_1, k_2, \dots\ ,\ k_{m-1}\}$.

It is obvious that system of linear equations of two different nodal points belonging to the same line differ in only one equation. Therefore, the computational efficiency of the descent algorithm will significantly increase if, to find the nodal points that lie on the line $l_{(k_1,\dots,k_{m-1})}$, the first $(m-1)$ rows of the augmented matrix corresponding to (6), with using elementary transformations, we preliminarily convert to a stepwise form:

$$\mathbf{A'}_{l_{(k_1,\dots,k_{m-1})}} = \begin{pmatrix} 1 & x_{k_1,2} & x_{k_1,3} & \dots & x_{k_1,m-1} & x_{k_1,m} & y_{k_1} \\ 0 & 1 & x'_{k_2,3} & \dots & x'_{k_2,m-1} & x'_{k_2,m} & y'_{k_2} \\ 0 & 0 & 1 & \dots & x'_{k_3,m-1} & x'_{k_3,m} & y'_{k_3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & x'_{k_{m-1},m} & y'_{k_{m-1}} \end{pmatrix}.$$

Using matrix $\mathbf{A'}_{l_{(k_1,\dots,k_{m-1})}}$ allows to significantly reduce the computational cost of finding all the nodal points belonging to the line $l_{(k_1,\dots,k_{m-1})}$: for each i-th nodal point we have an expanded matrix

$$\mathbf{A}_{u_{(k_1,\dots,k_{m-1},i)}} = \begin{pmatrix} 1 & x_{k_1,2} & x_{k_1,3} & \dots & x_{k_1,m-1} & x_{k_1,m} & y_{k_1} \\ 0 & 1 & x'_{k_2,3} & \dots & x'_{k_2,m-1} & x'_{k_2,m} & y'_{k_2} \\ 0 & 0 & 1 & \dots & x'_{k_3,m-1} & x'_{k_3,m} & y'_{k_3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & x'_{k_{m-1},m} & y'_{k_{m-1}} \\ 1 & x_{i,2} & x_{i,3} & \dots & x_{i,m-1} & x_{i,m} & y_i \end{pmatrix}, \tag{7}$$

where $k_1 < k_2 < \dots < k_{m-1}, i \in \{1,2,\dots,n\}, i \notin \{k_1,k_2,\dots,k_{m-1}\}$. To find i-th nodal point, it remains only to use back-substitution. Varying i, we find all the nodal points belonging to the line $l_{(k_1,\dots,k_{m-1})}$.

Since all the other rows of matrix (7) remain the same, we also do not accumulate rounding errors.

It is known that the algorithm of descent through nodal straight lines using sparse matrices and considering the direction of descent has an $O(m^2n^2 + m^4n\ln n + m^2n\ln^2 n)$ average computational complexity [15].

Now, the effect on the computational complexity by the proposed modification shall be highlighted. It is no longer necessary to calculate the values of the objective function at all nodal points belonging to the considered nodal straight line. In the proposed modification of the algorithm, for each considered nodal straight line, the value of the objective function is calculated only at the first considered nodal point, according to the formula:

$$Q(\mathbf{a}) = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{m} a_j x_{ij} \right|.$$

In addition, the value of the directional derivative of the objective function along this line is calculated (at the same nodal point), according to formula (5).

To calculate these values, about $3m + 4mn + n + 2mn = O(mn)$ operations are performed.

Then, at each considered nodal point on this line, we check whether the derivative changes sign.

Since for this modification we do not consider all m nodal straight lines at each nodal point, the average number of transitions from one nodal straight line to another also changes. Using a Monte Carlo method [16], the average number of transitions from one nodal straight line to another (P) was estimated for this modification of the algorithm. An analysis of the results for different n and m showed that the number of transitions from one nodal straight line to another is $P \sim O\left(m^2 n^{\frac{1}{4}}\right)$.

Therefore, this modification of the algorithm of descent through the nodal straight lines has an

$$W = O\left\{m^2 n^{\frac{1}{4}} \cdot \left[\left(m^3 + (n-m)\cdot m^2 + (n-m)\cdot \ln(n-m) + \frac{nm}{m^2 n^{\frac{1}{4}}} \cdot m^2 + mn\right)\right]\right\} =$$
$$= O\left(m^4 n^{\frac{5}{4}} + m^2 n^{\frac{5}{4}} \ln n\right) = O\left(m^2 n^{\frac{5}{4}} \cdot \max(m^2, \ln n)\right)$$

average computational complexity.

Note that for $\gg m$ $W = O(m^2 n^{\frac{5}{4}} \ln n) \leq O(m^2 n^{1,5})$. Thus, the proposed modification of the algorithm significantly increases its computational efficiency.

ACKNOWLEDGEMENTS

The research was supported by a grant from the Russian Fund of Fundamental Research, project no. 20-41-660008.

REFERENCES

1. S.A Ayvazyan, I.S. Yenyukov, L.D. Meshalkin, “The applied statistics. Researches of Relations”. Finances and statistics. Moskow. 488p. 1985.
2. E.Z Demidenko. “Linear and Nonlinear Regression”. Moscow. Finance and Statistics. 302p. 1981.
3. P. Bloomfield and W.L. Steiger, “Least absolute deviations: theory, applications, and algorithms”. Boston-Basel-Stuttgart: Birkhauser. 349p. 1983.
4. J. Kleinberg and E. Tardos, “Algorithm Design”. Cornell University. 864p. 2005.
5. V.I. Mudrov and V.L. Kushko, “Methods of Measurement Processing. Quasi Plausible Estimates”. Moscow, Radio and Communication. 304p. 1983.
6. S.I. Zukhovitski and L.I. Avdeeva, “Linear and Convex Programming, 2nd edition, revised and updated”. Moscow. Science Fizmathlit. 460p. 1967.
7. W.D. Fisher, “A note on curve fitting with minimum deviations by linear programming “. Journal of the American Statistical Association. **56**. pp. 359–362. 1961.
8. I. Barrodale and F.D.K. Roberts, “An improved algorithm for discrete L1 linear approximation”. SIAM Journal on Numerical Analysis. **10**. pp. 839–848. 1973.
9. S.C. Narula and J.F. Wellington, “Algorithm AS108: Multiple linear regression with minimum sum of absolute errors”. Applied Statistics. **26**. pp. 106–111. 1977.
10. R.D. Armstrong and D.S. Kung, “Algorithm AS132: Least absolute value estimates for a simple linear regression problem”. Applied Statistics. **27**. pp. 363–366. 1978.
11. A. V. Panteleev and T. A. Letova, “Optimization Methods in Examples and Tasks”. Moscow. High school.. 544p. 2002.
12. E. Weiszfeld “On the point for which the sum of the distances to n given points is minimum” Annals of Operations Research. **167, № 1**. P. 7–41. 2008. Translated from the French original [Tohoku Mathematics Journal, 1937. **43**. pp. 355–386] and annotated by Frank Plastria
13. P. A. Akimov and A.I. Matasov, “Levels of Non-optimality of the Weisfeld’s Algorithm for the Method of Least Absolute Deviations”. Moscow. Automation and Remote Control. **2**. pp. 4–16. 2010.
14. A. N. Tyrsin and A. A. Azaryan, “Resistant Linear Model Fitting Methods Based on the Descent Through the Nodal Straight Lines”. Models, systems, networks in economics, technology, nature and society. **1 (25)**. pp. 188–202. 2018.
15. A. N. Tyrsin and A. A. Azaryan, “Exact algorithms for the implementation of the least absolute deviations method based on the descent through the nodal straight lines”. Ulan-Ude: «Bulletin of the Buryat State University», series «Mathematics. Informatics». **4**. pp. 21–32. 2017.
16. G.A. Mikhailov “Numerical Statistical Modeling. Monte Carlo methods”. Moscow. Academia. 368p. 2006.