ANZIAM J. 50(2008), 185–198
doi:10.1017/S1446181108000345

THE BEST LEAST ABSOLUTE DEVIATIONS LINE – PROPERTIES AND TWO EFFICIENT METHODS FOR ITS DERIVATION

KRISTIAN SABO[1] and RUDOLF SCITOVSKI✉[2]

(Received 2 November, 2007; revised 17 November, 2008)

Abstract

Given a set of points in the plane, the problem of existence and finding the least absolute deviations line is considered. The most important properties are stated and proved and two efficient methods for finding the best least absolute deviations line are proposed. Compared to other known methods, our proposed methods proved to be considerably more efficient.

2000 *Mathematics subject classification*: primary 65D10, 65C20, 62J05; secondary 90C27, 90C56, 90B85, 34K29.

Keywords and phrases: least absolute deviations, LAD, l_1-norm approximation, weighted median problem.

1. Introduction

Let $I = \{1, \ldots, m\}$, $m \geq 2$, be a set of indices, and let $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i \in I\}$ be a set of points in the plane. The problem of determining parameters a and b of the linear function $f(x; a, b) = ax + b$, in order to have its graph pass as close as possible (in some sense) to the given points, is an old problem which has been solved in various ways. Most frequently it is assumed that errors can occur only in measured values of the independent variable. In this case, if we use the l_2 norm, we have the *ordinary least squares* (OLS) problem and the OLS line. In many technical and other applications (where so-called 'outliers' can be expected) using the l_1 norm is much more interesting. In the literature this approach is better known as the *least absolute deviations* (LAD) problem of finding the LAD line (see, for example, [8, 11, 21]). For example, in calculating movements of robots, based on the data obtained from a stereographic camera, it is important to estimate the position of a straight line in the

[1]Department of Mathematics, University of Osijek, Trg Ljudevita Gaja 6, HR-31000 Osijek, Croatia; e-mail: ksabo@mathos.hr.
[2]Department of Mathematics, University of Osijek, Trg Ljudevita Gaja 6, HR-31000 Osijek, Croatia; e-mail: scitowsk@mathos.hr.

plane efficiently and in real time (see [7, 14]). At the same time, the outliers in the data should not affect the results obtained. It is known that problems of this kind can be solved by applying the LAD approach (see, for example, [2, 6, 8, 21, 26, 32]). In this paper we give an overview of basic properties and facts related to estimation of parameters of the best LAD line, and propose two methods for finding them.

If errors are assumed to occur in measured values of both (dependent and independent) variables, and then in general the l_p norm is used, we have the *total least p* problem (see, for example, [1, 28, 31]).

In this paper the most important properties of the LAD line will be stated and proved and two efficient methods will be proposed. The first method basically relies upon [17] and for it the convergence theorem is proved. The second method for finding the best LAD line is a particular Gauss–Newton iterative process, which is based on solving a series of weighted least squares (WLS) line problems. Therefore, first some general properties and facts involved in this approach are given.

Basic properties of the LAD line are already known (see, for example, [2, 5, 6, 8, 9, 11, 17, 19, 21, 22, 26, 34, 37]). In this paper we present new and constructive proofs of those properties which enable us to construct efficient methods and prove the convergence theorem. Our methods have been compared to numerous other methods known from literature, and have proved to be much more efficient.

2. The best WLS line

The best WLS line is the graph of the function

$$f(x; a^*, b^*) = a^*x + b^*, \quad a^*, b^* \in \mathbb{R}, \tag{2.1}$$

whose parameters a^*, b^* are determined by minimizing the functional

$$F_2(a, b) = \sum_{i=1}^{m} \omega_i (y_i - ax_i - b)^2, \tag{2.2}$$

where $T_i(x_i, y_i)$, $i = 1, \ldots, m$, $m \geq 2$, are points in the plane, and $\omega_i > 0$ are corresponding data weights.

Using some natural conditions on the data it can be shown (see, for example, [4]) that there exist unique optimal parameters a^*, b^* of the best WLS line.

The following lemma contains the well-known fact that the best l_2 weighted approximation of the measured data is the weighted arithmetical mean of the measured data. This lemma is used in proving the fact that the best WLS line passes through the centroid of the data, and its proof is trivial.

LEMMA 2.1. *Let (ω_i, y_i), $i = 1, \ldots, m$, be the data, where $y_1, \ldots, y_m$ are real numbers, and $\omega_i > 0$ corresponding data weights. Then*

$$\sum_{i=1}^{m} \omega_i (y_i - \lambda)^2 \geq \sum_{i=1}^{m} \omega_i (y_i - \overline{y})^2, \quad \overline{y} = \frac{1}{\omega} \sum_{i=1}^{m} \omega_i y_i, \quad \omega = \sum_{i=1}^{m} \omega_i, \tag{2.3}$$

where the equality holds if and only if $\lambda = \overline{y}$.

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

LEMMA 2.2. *For the given data* (ω_i, x_i, y_i), $i = 1, \ldots, m$, $m \geq 2$, *in the plane with weights* $\omega_i > 0$, $i = 1, \ldots, m$, *the best WLS line passes through the centroid of the data* $(\overline{x}, \overline{y})$, *where*

$$\overline{x} = \frac{1}{\omega} \sum_{i=1}^{m} \omega_i x_i, \quad \overline{y} = \frac{1}{\omega} \sum_{i=1}^{m} \omega_i y_i, \quad \omega = \sum_{i=1}^{m} \omega_i.$$

If $\min x_i < \max x_i$, *then there exists a unique best WLS line whose parameters are given by*

$$a^* = \frac{\sum_{i=1}^{m} \omega_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{m} \omega_i (x_i - \overline{x})^2}, \quad b^* = \overline{y} - a^* \overline{x}. \tag{2.4}$$

If $x_1 = \cdots = x_m =: \xi$, *then there exist infinitely many best WLS lines, which take the form* $y = a^\star (x - \xi) + \overline{y}$, *where* $a^\star$ *is an arbitrary real number.*

PROOF. First note that there exist optimal parameters $a^*, b^* \in \mathbb{R}$ of the functional F_2 given by (2.2). Following [27] and according to Lemma 2.1, it is easy to see that

$$F_2(a^*, b^*) \geq F_2(a^*, \overline{y} - a^* \overline{x}),$$

from which we conclude that the best WLS line must pass through the centroid of the data, that is, the best WLS line should be searched for in the form $x \mapsto a(x - \overline{x}) + \overline{y}$ by minimizing the functional

$$\overline{F}_2(a) = \sum_{i=1}^{m} \omega_i [(y_i - \overline{y}) - a(x_i - \overline{x})]^2$$

whence, under condition $\min x_i < \max x_i$, it is easy to obtain formulae (2.4).

If $x_1 = \cdots = x_m =: \xi$, then according to Lemma 2.1,

$$F_2(a, b) = \sum_{i=1}^{m} \omega_i (y_i - a\xi - b)^2 \geq \sum_{i=1}^{m} \omega_i (y_i - \overline{y})^2,$$

where the equality holds if and only if $a\xi + b = \overline{y}$. Hence, in this case there exist infinitely many best WLS lines which take the form $\overline{f}(x; a^*) = f(x; a^*, -\xi a^* + \overline{y}) = a^\star (x - \xi) + \overline{y}$, where $a^\star$ is an arbitrary real number.

3. Weighted median of the data

Let us first prove the following lemma, which gives properties and a solution of the weighted median problem. This lemma is used in the proof of the existence of the best LAD line passing through at least two different points of the data and other important theorems.

LEMMA 3.1. *Let* (ω_i, y_i), $i \in I$, $I = \{1, \ldots, m\}$, $m \geq 2$, *be the data, where* $y_1 \leq y_2 \leq \cdots \leq y_m$ *are real numbers, and* $\omega_i > 0$ *corresponding data weights. Denote*

$$J = \left\{ \nu \in I \;\middle|\; 2 \sum_{i=1}^{\nu} \omega_i - \sum_{i=1}^{m} \omega_i \leq 0 \right\}.$$

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

For $J \neq \emptyset$, *let* $\nu_0 = \max J$. *Furthermore, let* $F : \mathbb{R} \to \mathbb{R}$ *be a function defined as*

$$F(\alpha) = \sum_{i=1}^{m} \omega_i |y_i - \alpha|. \tag{3.1}$$

(i) *If* $J = \emptyset$ *(that is,* $2\omega_1 > \sum_{i=1}^{m} \omega_i$*), then the minimum of* F *is attained at the point* $\alpha^\star = y_1$.

(ii) *If* $J \neq \emptyset$ *and* $2\sum_{i=1}^{\nu_0} \omega_i < \sum_{i=1}^{m} \omega_i$, *then the minimum of* F *is attained at the point* $\alpha^\star = y_{\nu_0+1}$.

(iii) *If* $J \neq \emptyset$ *and* $2\sum_{i=1}^{\nu_0} \omega_i = \sum_{i=1}^{m} \omega_i$, *then the minimum of* F *is attained at every point* $\alpha^\star$ *from the segment* $[y_{\nu_0}, y_{\nu_0+1}]$.

PROOF. Notice that on each interval

$$(-\infty, y_1), [y_1, y_2), \ldots, [y_{m-1}, y_m), [y_m, \infty),$$

F is a linear function and slopes of those linear functions are consecutively d_ν, $\nu = 0, \ldots, m$, where

$$d_0 = -\sum_{i=1}^{m} \omega_i, \quad d_m = \sum_{i=1}^{m} \omega_i,$$

$$d_\nu = 2\sum_{i=1}^{\nu} \omega_i - \sum_{i=1}^{m} \omega_i = d_{\nu-1} + 2\omega_\nu, \quad \nu = 1, \ldots, m-1.$$

If $J = \emptyset$, then for every $\nu = 1, \ldots, m$, $2\sum_{i=1}^{\nu} \omega_i - \sum_{i=1}^{m} \omega_i > 0$ and $d_0 < 0 < d_\nu$. It follows that F is strongly decreasing on $(-\infty, y_1)$ and strongly increasing on (y_1, ∞), therefore the minimum of F is attained for $\alpha^\star = y_1$.

If $J \neq \emptyset$, note that $\nu_0 = \max\{\nu \in I \mid d_\nu \leq 0\}$. Since $d_{\nu+1} - d_\nu = 2\omega_{\nu+1} > 0$ and $d_0 < 0$ and $d_m > 0$, the sequence (d_ν) is increasing and

$$d_0 < d_1 < \cdots < d_{\nu_0} \leq 0 < d_{\nu_0+1} < \cdots < d_m. \tag{3.2}$$

If $d_{\nu_0} < 0$, that is, $2\sum_{i=1}^{\nu_0} \omega_i < \sum_{i=1}^{m} \omega_i$, it follows from (3.2) that F is strongly decreasing on $(-\infty, y_{\nu_0+1})$ and strongly increasing on (y_{ν_0+1}, ∞), therefore the minimum of F is attained for $\alpha^\star = y_{\nu_0+1}$.

If $d_{\nu_0} = 0$, that is, $2\sum_{i=1}^{\nu_0} \omega_i = \sum_{i=1}^{m} \omega_i$, it follows from (3.2) that F is strongly decreasing on $(-\infty, y_{\nu_0})$, is constant on $[y_{\nu_0}, y_{\nu_0+1}]$ and strongly increasing on (y_{ν_0+1}, ∞), therefore the minimum of F is attained at every point $\alpha^\star$ from the segment $[y_{\nu_0}, y_{\nu_0+1}]$.

COROLLARY 3.2. *Let* $y_1 \leq y_2 \leq \cdots \leq y_m$, $m > 1$, *be the data with weights* $\omega_1 = \cdots = \omega_m = 1$. *Then*

(i) *if* m *is odd* $(m = 2k+1)$, *then the minimum of* F *is attained at the point* $\alpha^\star = y_{k+1}$;

(ii) *if* m *is even* $(m = 2k)$, *the minimum of* F *is attained at every point* $\alpha^\star$ *from the segment* $[y_k, y_{k+1}]$.

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

PROOF. First, note that in this case the set J from Lemma 3.1 is always nonempty. Let $m = 2k + 1$. According to Lemma 3.1(ii),

$$\nu_0 = \max\{\nu \in I \mid 2\nu - m \le 0\} = \max\{\nu \in I \mid \nu \le k + \tfrac{1}{2}\} = k,$$
$$d_{\nu_0} = d_k = 2k - m = 2k - 2k - 1 < 0,$$

and therefore $\alpha^\star = y_{k+1}$.

Let $m = 2k$. According to Lemma 3.1(iii),

$$\nu_0 = \max\{\nu \in I \mid 2\nu - m \le 0\} = \max\{\nu \in I \mid \nu = k\} = k,$$
$$d_{\nu_0} = d_k = 2k - m = 2k - 2k = 0.$$

It follows that the minimum of F is attained at every point $\alpha^\star$ from the segment $[y_k, y_{k+1}]$.

REMARK 1. If the minimum of the functional F defined by (3.1) is attained at the real number α^*, then

$$F(\alpha) = \sum_{i=1}^{m} \omega_i |y_i - \alpha| \ge \sum_{i=1}^{m} \omega_i |y_i - \alpha^*|,$$

where the equality holds if and only if $\alpha = \alpha^*$.

4. The best LAD line

Let $I = \{1, \ldots, m\}$, $m \ge 2$, be a set of indices and $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i \in I\}$ a set of points in the plane. The best LAD line is to be determined, that is, optimal parameters $a^*, b^* \in \mathbb{R}$ of the function $f(x; a, b) = ax + b$ are to be determined such that

$$G(a^*, b^*) = \min_{(a,b)\in\mathbb{R}^2} G(a, b), \quad G(a, b) = \sum_{i=1}^{m} |y_i - ax_i - b|. \tag{4.1}$$

The following theorem is straightforward, and it can be proved by the principle applied in [9, 13, 15, 16].

THEOREM 4.1. *Let $I = \{1, \ldots, m\}$, $m \ge 2$, be a set of indices and $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i \in I\}$ a set of points in the plane. Then there is a best LAD line, that is, problem (4.1) has a solution. In particular, if $x_1 = \cdots = x_m =: \xi$, then there exist infinitely many best LAD lines, which take the form $y = a^\star(x - \xi) + \mu$, where $a^\star$ is an arbitrary real number, and μ is the median of the data $y_1, \ldots, y_m$.*

The following lemma shows that for the linear function whose graph passes through a certain point $T_0(x_0, y_0) \in \mathbb{R}^2$ there exists a best LAD line whose graph also passes through a certain point $T_i(x_i, y_i) \in \Lambda$, for which $x_i \ne x_0$. In particular, the point T_0 can be one of the points of the set Λ. In this case the lemma states that there exists a best LAD line whose graph passes through one more point $T_i(x_i, y_i) \in \Lambda$ for which $x_i \neq x_0$.

LEMMA 4.2. *Let $I = \{1, \ldots, m\}$, $m \geq 2$, be a set of indices and*

(i) $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i \in I\}$ *a set of points in the plane, such that* $(x_1 \leq \cdots \leq x_m)$ *and* $(x_1 < x_m)$*;*
(ii) $T_0(x_0, y_0) \in \mathbb{R}^2$*;*
(iii) $\overline{f}(x; a) = a(x - x_0) + y_0$, $a \in \mathbb{R}$, *a linear function whose graph passes through the point* $T_0(x_0, y_0) \in \mathbb{R}^2$.

Then there exists $a^* \in \mathbb{R}$ such that

$$\overline{G}(a^*) = \min_{a\in\mathbb{R}} \overline{G}(a), \quad \overline{G}(a) = \sum_{i=1}^{m} |y_i - \overline{f}(x_i; a)| = \sum_{i=1}^{m} |y_i - a(x_i - x_0) - y_0|, \tag{4.2}$$

and the graph of the linear function $x \mapsto \overline{f}(x; a^*)$ passes through at least one more point $T_\nu(x_\nu, y_\nu) \in \Lambda$, where $x_\nu \neq x_0$.

PROOF. Write $I_0 = \{i \in I \mid x_i = x_0\}$. It can be seen that $I \setminus I_0 \neq \emptyset$, that is, that there exists $i_0 \in I$ such that $x_{i_0} \neq x_0$. Otherwise $x_i = x_0$, for all $i \in I$, which contradicts assumption (i). $\overline{G}(a)$ in (4.2) can be written in the form

$$\overline{G}(a) = \sum_{i=1}^{m} |y_i - a(x_i - x_0) - y_0| = \sum_{i\in I_0} |y_i - y_0| + \sum_{i\in I\setminus I_0} |x_i - x_0| \left|\frac{y_i - y_0}{x_i - x_0} - a\right|. \tag{4.3}$$

Let $s = \operatorname{card}(I \setminus I_0)$. We rearrange the data $\{(|x_i - x_0|, (y_i - y_0)/(x_i - x_0)) : i \in I \setminus I_0\}$ so that the new data $\{(|x_i' - x_0|, (y_i' - y_0)/(x_i' - x_0)) : i = 1, \ldots, s\}$ are such that $((y_1' - y_0)/(x_1' - x_0), \ldots, (y_s' - y_0)/(x_s' - x_0))$ is an increasing sequence. Notice that it still remains the case that $x_i' - x_0 \neq 0$, for all $i = 1, \ldots, s$. Therefore, the problem of minimizing (4.3) is the weighted median problem which can be written in the form

$$\overline{G}(a) = \sum_{i\in I_0} |y_i - y_0| + \sum_{i=1}^{s} |x_i' - x_0| \left|\frac{y_i' - y_0}{x_i' - x_0} - a\right| \to \min. \tag{4.4}$$

Write $J = \{\nu : 2\sum_{i=1}^{\nu} |x_i' - x_0| - \sum_{i=1}^{s} |x_i' - x_0| \leq 0\}$. For $J \neq \emptyset$ let us choose $\nu_0 \in \{1, \ldots, s\}$, such that $\nu_0 = \max J$ and denote

$$W_0 = 2\sum_{i=1}^{\nu_0} |x_i' - x_0|, \quad W = \sum_{i=1}^{s} |x_i' - x_0|.$$

According to Lemma 3.1, the functional (4.4) always has a minimum; furthermore,

(i) if $J = \emptyset$, it is obtained for $a^* = (y_1' - y_0)/(x_1' - x_0)$, that is, the required linear function $\overline{f}$ is of the form

$$\begin{aligned}\overline{f}(x; a^*) &= a^*(x - x_0) + y_0 \\ &= \frac{y_1' - y_0}{x_1' - x_0}(x - x_0) + y_0,\end{aligned}$$

meaning that its graph also passes through the point $T_1'(x_1', y_1') \in \Lambda$, where $x_1' \neq x_0$;

(ii) if $J \neq \emptyset$ and $W_0 < W$, the minimum is obtained for $a^* = (y_{\nu_0+1}' - y_0)/(x_{\nu_0+1}' - x_0)$, that is, the required linear function $\overline{f}$ is of the form

$$\begin{aligned}\overline{f}(x; a^*) &= a^*(x - x_0) + y_0 \\ &= \frac{y_{\nu_0+1}' - y_0}{x_{\nu_0+1}' - x_0}(x - x_0) + y_0,\end{aligned}$$

meaning that its graph also passes through the point $T_{\nu_0+1}'(x_{\nu_0+1}', y_{\nu_0+1}') \in \Lambda$ for which $\nu_0 + 1 \in I \setminus I_0$, that is, $x_{\nu_0+1}' \neq x_0$;

(iii) if $J \neq \emptyset$ and $W_0 = W$, the minimum of functional $\overline{G}$ is attained at any point $a^* \in \big[(y_{\nu_0}' - y_0)/(x_{\nu_0}' - x_0), (y_{\nu_0+1}' - y_0)/(x_{\nu_0+1}' - x_0)\big]$, and similarly to previous cases it can be shown that if we choose $a^* = (y_{\nu_0}' - y_0)/(x_{\nu_0}' - x_0)$, then the graph of the searched linear function $\overline{f}$ also passes through the point $T_{\nu_0}'(x_{\nu_0}', y_{\nu_0}') \in \Lambda$ for which $x_{\nu_0} \neq x_0$, and if we choose $a^* = (y_{\nu_0+1}' - y_0)/(x_{\nu_0+1}' - x_0)$, then the graph of the required linear function $\overline{f}$ also passes through the point $T_{\nu_0+1}'(x_{\nu_0+1}', y_{\nu_0+1}')$ for which $x_{\nu_0+1}' \neq x_0$.

Thus, there always exists $a^* \in \mathbb{R}$ such that the graph of the function $x \mapsto \overline{f}(x; a^*)$ passes through at least one more point $T_\nu(x_\nu, y_\nu) \in \Lambda$ for which $x_\nu \neq x_0$.

THEOREM 4.3. *Let $I = \{1, \ldots, m\}$, $m \geq 2$, be a set of indices, and $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i \in I\}$ a set of points in the plane, such that $(x_1 \leq \cdots \leq x_m)$ and $(x_1 < x_m)$. Then there exists a best LAD line which passes through at least two different points from Λ.*

PROOF. According to Theorem 4.1, there exists a best LAD line $f(x; a^*, b^*) = a^*x + b^*$ with optimal parameters a^*, b^*. By means of a^* and data (x_i, y_i), $i \in I$, let us define the sequence

$$\Delta_m = (y_1 - a^*x_1, \ldots, y_m - a^*x_m),$$

which, without loss of generality, we can assume is increasing.

If $m = 2k + 1$, according to Corollary 3.2(i), the minimum of the functional

$$G(a^*, b) = \sum_{i=1}^{m} |y_i - a^*x_i - b| \tag{4.5}$$

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

is attained for $b^+ = y_{k+1} - a^* x_{k+1}$. Since

$$G(a^*, b^*) = \min_{b \in \mathbb{R}} G(a^*, b) = G(a^*, b^+),$$

at the point (a^*, b^+) the functional G also attains its minimum, and since $y_{k+1} = a^* x_{k+1} + b^+$, it means that there exists a best LAD line passing through the point $T_1(x_{k+1}, y_{k+1})$ and it can be written as

$$\overline{f}(x; a^*) := f(x; a^*, y_{k+1} - a^* x_{k+1}) = a^*(x - x_{k+1}) + y_{k+1}.$$

According to Lemma 4.2, then there exists at least one more point $T_2(x_\mu, y_\mu) \in \Lambda$, through which the best LAD line passes, and for which $x_\mu \neq x_{k+1}$.

If $m = 2k$, according to Corollary 3.2(ii), the minimum of functional (4.5) is attained at any real number from segment $[y_k - a^* x_k, y_{k+1} - a^* x_{k+1}]$. Let us choose $b^+ := y_{k+1} - a^* x_{k+1}$; by deduction similar to the previous case we conclude that at least one more point $T_2(x_\mu, y_\mu) \in \Lambda$ exists, through which the best LAD line passes, and for which $x_\mu \neq x_{k+1}$.

5. Methods for finding the best LAD line

To solve this problem general methods of minimization can be used that do not involve derivatives, such as differential evolutions, Nelder–Mead, random search, and simulated annealing (see, for example, [18, 20, 23, 38]). There are also a certain number of methods based on linear programming (see, for example, [2, 3, 24, 33]) or different special cases of the Gauss–Newton method (see, for example, [9, 25, 30, 35]). There are also some specialized methods for solving this problem (see, for example, [17, 21, 26, 37]). Numerous methods mentioned in the literature are listed in [8]. In this section two effective methods are proposed, and their properties and characteristics are given along with proofs of the convergence theorem.

5.1. Two-points method The method described in this subsection resulted from the improved method referred to in [17], for which we give theoretical justification and prove the convergence theorem. It will be shown that this method gives a solution of LAD problem (4.1) in finitely many steps (significantly less than m).

With the objective of developing an algorithm for finding the best LAD line, first note that Theorem 4.3 refers to the fact that the best LAD line should be searched for among those lines which pass through at least two different points of the sequence Λ. The following algorithm is based on that fact. To get as close as possible from the beginning to the best LAD line, we can choose the point $T_p(x_p, y_p)$ as the initial point, where

$$\begin{aligned} &x_p = \frac{1}{m}\sum_{i=1}^m x_i, \quad y_p = \frac{1}{m}\sum_{i=1}^m y_i \quad \text{(centroid of the data), or} \\ &x_p = \text{median}\,(x), \quad y_p = \text{median}\,(y) \quad \text{(median of the data),} \end{aligned} \tag{5.1}$$

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

which are quickly calculated and probably pass close to the best LAD line. Thereafter, in accordance with Lemma 4.2, we construct the following algorithm.

Algorithm [Two Points]

Step 1: Set $m \geq 2$, input points $T_i(x_i, y_i)$, $i \in I$, $I = \{1, \dots, m\}$, according to (5.1), determine the point $T_p(x_p, y_p)$ and define the set $I_0 = \{i \in I \mid x_i = x_p\}$.

Step 2: In accordance with Lemma 4.2:

- solve the LAD problem for the function $\overline{f}(x; a) = a(x - x_p) + y_p$ by minimizing functional (4.3), denote the solution by a_1 and determine a new point $T_{i_1}(x_{i_1}, y_{i_1}) \in \Lambda$ for which $x_{i_1} \neq x_p$;
- put $b_1 = -a_1 x_p - y_p$ and calculate $G_1 = \sum_{i=1}^{m} |y_i - a_1 x_i - b_1|$.

Step 3: Define the set $I_0 = \{i \in I \mid x_i = x_{i_1}\}$. In accordance with Lemma 4.2:

- solve the LAD problem for the function $\overline{f}(x; a) = a(x - x_{i_1}) + y_{i_1}$ by minimizing functional (4.3), denote the solution by a_2 and determine a new point $T_{i_2}(x_{i_2}, y_{i_2}) \in \Lambda$ for which $x_{i_2} \neq x_{i_1}$;
- put $b_2 = -a_2 x_{i_1} - y_{i_1}$ and calculate $G_2 = \sum_{i=1}^{m} |y_i - a_2 x_i - b_2|$.

Step 4: If $G_2 < G_1$, put $\{i_1 = i_2, G_1 = G_2\}$ and go to step 3; if not, stop.

The following theorem proves that the two-points algorithm leads to the best LAD line in finitely many steps. From the algorithm and Lemma 4.2 it is clear that the number of steps is less than the number of given points $T_1, \dots, T_m$. In practice, with a favourable choice of the initial point according to (5.1), the number of steps will be considerably less than m. In this way maximum efficiency of the algorithm is ensured, which is implementable in real time.

THEOREM 5.1. *Let $I = \{1, \dots, m\}$, $m \geq 2$, be a set of indices, and $\Lambda = \{T_i(x_i, y_i) \mid i \in I\}$ a set of points in the plane, such that $(x_1 \leq \cdots \leq x_m)$ and $(x_1 < x_m)$. Then the sequence (a_n, b_n), defined by the iterative method in the two-points algorithm in $N \leq m$ steps leads to the solution of the LAD problem*

$$G(a, b) = \sum_{i=1}^{m} |y_i - a x_i - b| \to \min, \quad \textit{where} \tag{5.2}$$
$$G(a_{n+1}, b_{n+1}) < G(a_n, b_n), \quad n = 1, \dots, N.$$

PROOF. If we determine the coefficient b of the linear function $f(x) = ax + b$ such that its graph passes through the point $T_p(x_p, y_p)$ chosen as in (5.1), the functional G given by (5.2) can be written as

$$\overline{G}(a) = G(a, -a x_p + y_p) = \sum_{i=1}^{m} |y_i - a(x_i - x_p) - y_p|,$$

and according to Lemma 4.2, we can solve the corresponding LAD problem, the solution of which is denoted by a_1. The value of G at the point (a_1, b_1) is $G(a_1, b_1) = G(a_1, -a_1 x_p + y_p) = \overline{G}(a_1)$. According to Lemma 4.2, in this way the linear

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

function $\overline{f}_1(x) = a_1(x - x_p) + y_p$ is defined, whose graph passes through one more new point $T_{i_1}(x_{i_1}, y_{i_1}) \in \Lambda$ for which $x_{i_1} \neq x_p$. Therefore, function $\overline{f}_1$ can be written as

$$\overline{f}_1(x) = a_1(x - x_{i_1}) + y_{i_1}.$$

Again according to Lemma 4.2, for this function we solve the LAD problem and denote the solution by a_2. If (a_2, b_2), $b_2 = -a_2 x_{i_1} + y_{i_1}$ is not the point of the minimum of functional G, it is

$$G(a_1, b_1) > \sum_{i=1}^{m} |y_i - a_2(x_i - x_{i_1}) - y_{i_1}| = G(a_2, b_2).$$

Since according to Lemma 4.2 the graph of the function $\overline{f}_2(x) = a_2(x - x_{i_1}) + y_{i_1}$ also passes through a new point $T_{i_2}(x_{i_2}, y_{i_2}) \in \Lambda$, for which $x_{i_2} \neq x_{i_1}$, we can write

$$\overline{f}_2(x) = a_2(x - x_{i_2}) + y_{i_2}.$$

Now again according to Lemma 4.2, for this function we solve the LAD problem and denote the solution by a_3. If (a_3, b_3), $b_3 = -a_3 x_{i_2} + y_{i_2}$ is not the point of the minimum of functional G, it is

$$G(a_2, b_2) > \sum_{i=1}^{m} |y_i - a_3(x_i - x_{i_2}) - y_{i_2}| = G(a_3, b_3).$$

By repeating this procedure, since the set Λ has finitely many points, this iterative method will end in finitely many steps.

5.2. Iterative reweighted least squares method The idea of the method presented in this subsection appears in different versions in the literature (see, for example, [2, 9, 24, 25, 29]). A general description of the method on which our iterative reweighted least squares (IRLS) method is also based can be found in, for example, [9, 24, 29, 36].

The idea is to write the minimizing functional (4.1) in the form

$$G(a, b) = \sum_{i=1}^{m} |y_i - a x_i - b| = \sum_{i=1}^{m} \frac{1}{|y_i - a x_i - b|}(y_i - a x_i - b)^2 \tag{5.3}$$

and construct the iterative method so that in every step we solve one WLS problem according to Lemma 2.2. The iterative method is specified in the following algorithm. Since the best LAD line must pass through at least two data points (Theorem 4.3), then at least two weights in step 2 near the solution can become infinitely large. This fact will be used in step 3 to stop the iterative process.

Algorithm [IRLS]

Step 1: Set $m \geq 2$ and $M > 0$ and input points $T_i(x_i, y_i)$, $i \in I$, $I = \{1, \ldots, m\}$.

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

Step 2: By using explicit formulae for solving the OLS problem (see Lemma 2.2) determine $a_0, b_0 \in \mathbb{R}$:

$$a_0 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}, \quad b_0 = \overline{y} - a_0 \overline{x},$$

$$\text{where } \overline{x} = \frac{1}{m} \sum x_i, \quad \overline{y} = \frac{1}{m} \sum y_i.$$

Step 3: Calculate $\omega_i = 1/(|y_i - a_0 x_i - b_0|)$, solve the WLS problem

$$G(a, b) = \sum \omega_i (y_i - a x_i - b)^2 \to \min$$

by using explicit formulae from Lemma 2.2 and denote the solution by a_1, b_1:

$$a_1 = \frac{\sum \omega_i (x_i - x_p)(y_i - y_p)}{\sum \omega_i (x_i - x_p)^2}, \quad b_1 = y_p - a_1 x_p, \quad \text{where}$$

$$x_p = \frac{1}{\omega} \sum \omega_i x_i, \quad y_p = \frac{1}{\omega} \sum \omega_i y_i \quad \text{and} \quad \omega = \sum \omega_i.$$

Step 4: If $\max_{i=1,m} \omega_i < M$, put $a_0 = a_1$, $b_0 = b_1$ and go to step 3; otherwise stop.

As shown in [9, 29], this algorithm defines a sequence of approximations $(\theta^{(k)}) = (a_k, b_k)^T$, for which the sequence $G_k = G(\theta^{(k)})$ is strictly decreasing. It can be shown that this is actually a special case of the well-known Gauss–Newton method (see, for example, [9, 10, 29])

$$\theta^{(k+1)} = \theta^{(k)} - \lambda_k p^{(k)}, \quad k = 0, 1, \ldots,$$

where λ_k is a length of steps in the direction of vector $p^{(k)}$, which is a solution of the WLS problem

$$\sqrt{\Omega^{(k)}} J^{(k)} p \simeq \sqrt{\Omega^{(k)}} r^{(k)},$$

where $J^{(k)}$ is a Jacobian of the function $f(x; a, b) = ax + b$ at the point $\theta^{(k)}$, $r^{(k)}$ a vector of residuals at the point $\theta^{(k)}$, and $\Omega^{(k)}$ a diagonal matrix with elements $\omega_{ii}^{(k)} = 1/(|y_i - a_k x_i - b_k|)$.

6. Numerical experiments

In the first example below we graphically illustrate the behaviour of the two-points algorithm; in the second, for a large set of points, we will compare the efficiency of the two-points and IRLS algorithms by using some other methods which can be found in the aforementioned references.

EXAMPLE 1. We are given the set of points $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, 8\}$, where

x_i	1	2	3	4	5	6	7	8
y_i	7	14	10	17	15	21	26	23

https://doi.org/10.1017/S1446181108000345 Published online by Cambridge University Press

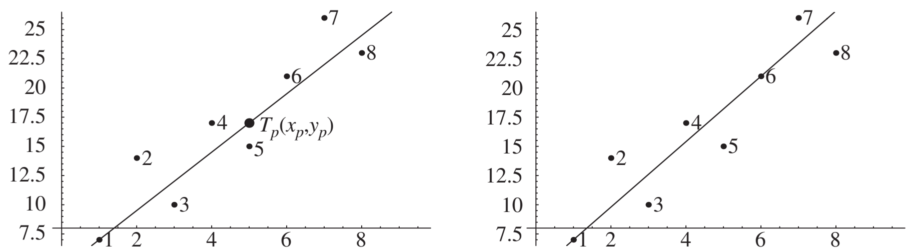

FIGURE 1. Illustration of the two-points algorithm.

TABLE 1. Comparison of algorithms for solving the LAD problem.

m	10	100	500	1000	5000	10 000
Two-points algorithm	2(0)	3(0.03)	3(0.42)	2(1)	2(24.5)	3(75)
IRLS algorithm	16(0)	18(0.05)	32(0.1)	17(0.11)	7(0.22)	44(1.4)
(Li–Arce, [21])	2(0)	2(0.04)	3(1.03)	4(4.53)	1(105.8)	1(529)
(Wesolowsky, [37])	3(0)	3(0.04)	3(1.09)	3(4.34)	3(127.8)	3(635)
Differential Evolution	(0.3)	(0.44)	(1.86)	(6.95)	(194)	(825)
Nelder–Mead	(0.14)	(0.16)	(1.36)	(6.42)	(186)	(775)
Random Search	(0.25)	(0.47)	(3.2)	(14.75)	(322)	(1168)
Simulated Annealing	(0.25)	(0.17)	(1.66)	(6.53)	(209)	(852)

The two-points algorithm is initialized so that first the centroid of the data is calculated: $T_p(5, 17) \notin \Lambda$. Then the algorithm chooses the first point $T_1(1, 7) \in \Lambda$ and determines the linear function $f_1(x) = 2.5x + 4.5$, whose graph passes through those two points (see the left-hand illustration in Figure 1). The sum of absolute deviations is $G_1 = 18$.

Then the algorithm fixes the point $T_1(1, 7)$ and chooses a new point $T_6(6, 21)$ and a new linear function $f_2(x) = 2.8x + 4.2$, whose graph passes through those two points (see the right-hand illustration in Figure 1). The sum of absolute deviations is now $G_2 = 17.4$, which is a global minimum for this problem.

Of course, the same result is obtained by applying the module `NMinimize` using the *Nelder–Mead* method from the Mathematica software (see also [20, 23, 38]).

EXAMPLE 2. We are given the set of points $\Lambda = \{T_i(x_i, y_i) \in \mathbb{R}^2 \mid i = 1, \ldots, m\}$, where

$$x_i = \frac{10i}{m}, \quad i = 1, \ldots, m,$$
$$y_i = 3x_i + 2 + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2).$$

The efficiency of the two-points and IRLS algorithms will be compared to some special methods for finding the best LAD line [21, 37], but also to general minimization algorithms, which are also included in Mathematica: `DifferentialEvolution`, `NelderMead`, `RandomSearch`, and `SimulatedAnnealing` (see also [18, 20, 23, 38]).

For $m = 10$, 100, 500, 1000, 5000, 10 000, Table 1 shows (in parentheses) the time in seconds, and for the two-points, IRLS, Wesolowsky and Li–Arce algoirithms the number of iterations[1]. It has been shown that the two-points algorithm does not have significantly different properties if it is initialized with the centroid of the data $(\overline{x}, \overline{y})$ or with the point (M_x, M_y), where M_x (M_y) is the median of the sequence x (y), as proposed in (5.1). Note that for the purpose of computing time efficiency the centroid of the data would be a better choice.

As can be seen, the two-points algorithm reaches a global minimum in only few steps (most often 2–3 steps), while the IRLS algorithm requires a few more iterations. The IRLS algorithm requires a significantly shorter time than the other algorithms considered. The time required by the two-points algorithm is a bit longer, which is probably the result of a direct application of Mathematica module `Sort` (see [38]). Several fast `Sort` algorithms can be found in [12]. From Table 1 it can be seen that both proposed algorithms show significant superiority over other methods, which can be further improved by careful programming.

Acknowledgement

This work was supported by the Ministry of Science, Education and Sports, Republic of Croatia, through research grant 235-2352818-1034.

References

[1] A. Atieg and G. A. Watson, "Use of l_p norms in fitting curves and surfaces to data", *ANZIAM J.* **45** (2004) C187–C200.

[2] A. Bargiela and J. K. Hartley, "Orthogonal linear regression algorithm based on augmented matrix formulation", *Comput. Oper. Res.* **20** (1993) 829–836.

[3] A. Barrodale and F. D. K. Roberts, "An improved algorithm for discrete l_1 linear approximation", *SIAM J. Numer. Anal.* **10** (1973) 839–848.

[4] Å. Björck, *Numerical methods for least squares problems* (SIAM, Philadelphia, PA, 1996).

[5] P. Bloomfield and W. Steiger, *Least absolute deviations: theory, applications, and algorithms* (Birkhäuser, Boston, 1983).

[6] J. Brimberg, H. Juel and A. Schöbel, "Properties of three-dimensional median line location models", *Ann. Oper. Res.* **122** (2003) 71–85.

[7] R. Cupec, *Scene reconstruction and free space representation for biped walking robots* (VDI Verlag, Düsseldorf, 2005).

[8] M. Dasgupta and S. K. Mishra, Least absolute deviation estimation of linear econometric models: A literature review. http:// mpra.ub.uni-muenchen.de/ 1781/.

[9] E. Z. Demidenko, *Optimization and regression* (Nauka, Moscow, 1989) (in Russian).

[10] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations* (SIAM, Philadelphia, PA, 1996).

[11] Y. Dodge (ed.), "Statistical data analysis based on the L_1-norm and related methods", in *Proceedings of the Third International Conference on Statistical Data Analysis Based on the L_1-norm and Related Methods* (Elsevier, Neuchâtel, 1997).

[1] The corresponding Mathematica modules were written and the Wesolowsky and Li–Arce algorithms tested by I. Kuzmanović, Department of Mathematics, University of Osijek.

[12] C. Gurwitz, "Weighted median algorithms for L_1 approximation", *BIT* **30** (1990) 101–110.

[13] K. P. Hadeler, D. Jukić and K. Sabo, "Least squares problems for Michaelis Menten kinetics", *Math. Methods Appl. Sci.* **30** (2007) 1231–1241.

[14] K. Hirai, M. Hirose, Y. Haikawa and T. Takenaka, "The development of Honda humanoid robot", in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, Leuven, Belgium, 1998) 1321–1326.

[15] D. Jukić, K. Sabo and R. Scitovski, "Total least squares fitting Michaelis–Menten enzyme kinetic model function", *J. Comput. Appl. Math.* **201** (2007) 230–246.

[16] D. Jukić and R. Scitovski, "The best least squares approximation problem for a 3-parametric exponential regression model", *ANZIAM J.* **42** (2000) 254–266.

[17] O. J. Karst, "Linear curve fitting using least deviations", *J. Amer. Statist. Assoc.* **53** (1958) 118–132.

[18] C. T. Kelley, *Iterative methods for optimization* (SIAM, Philadelphia, PA, 1999).

[19] N. M. Korneenko and H. Martini, "Hyperplane approximation and related topics", in *New trends in discrete and computational geometry* (ed. J. Pach), (Springer, Berlin, 1993).

[20] R. M. Lewis, V. Torczon and M. W. Trosset, "Direct search methods: then and now", *J. Comput. Appl. Math.* **124** (2000) 191–207.

[21] Y. Li and G. R. Arce, "A maximum likelihood approach to least absolute deviation regression", *EURASIP J. Appl. Signal Process.* **12** (2004) 1762–1769.

[22] K. Madsen and H. B. Nielsen, "A finite smoothing algorithm for linear l_1 estimation", *SIAM J. Optim.* **3** (1993) 223–235.

[23] J. A. Nelder and R. Mead, "A simplex method for function minimization", *Comput. J.* **7** (1965) 308–313.

[24] M. R. Osborne, *Finite algorithms in optimization and data analysis* (Wiley, Chichester, 1985).

[25] E. J. Schlossmacher, "An iterative technique for absolute deviations curve fitting", *J. Amer. Statist. Assoc.* **68** (1973) 857–859.

[26] A. Schöbel, *Locating lines and hyperplanes: theory and algorithms* (Springer, Berlin, 1999).

[27] R. Scitovski, Š. Ungar and D. Jukić, "Approximating surface by moving total least squares method", *Appl. Math. Comput.* **93** (1998) 219–232.

[28] H. Späth, "On discrete linear orthogonal L_p approximation", *ZAMM* **62** (1982) 354–355.

[29] H. Späth, *Mathematical algorithms for linear regression* (Academic Press Professional, San Diego, 1992).

[30] H. Späth, "Identifying spatial point sets", *Math. Commun.* **8** (2003) 69–76.

[31] H. Späth and G. A. Watson, "On orthogonal linear l_1 approximation", *Numer. Math.* **51** (1987) 531–543.

[32] A. Tarantola, *Inverse problem theory and methods for model parameter estimation* (SIAM, Philadelphia, PA, 2005).

[33] N. T. Trendafilov and G. A. Watson, "The l_1 oblique Procrustes problem", *Statist. Comput.* **14** (2004) 39–51.

[34] G. A. Watson, *Approximation theory and numerical methods* (Wiley, Chichester, 1980).

[35] G. A. Watson, "Aspects of approximation with emphasis on the univariate case", in *The state of the art in numerical analysis, IMA Conference Series* (eds. I. S. Duff and G. A. Watson), (Oxford University Press Inc., New York, 1997).

[36] G. A. Watson, "On the Gauss–Newton method for l_1 orthogonal distance regression", *IMA J. Numer. Anal.* **22** (2002) 345–357.

[37] G. O. Wesolowsky, "A new descent algorithm for the least absolute value regression problem", *Comm. Statist. Simulation Comput.* **B10** (1981) 479–491.

[38] S. Wolfram, *The mathematica book* (Wolfram Media, Champaign, IL, 2007).