Odjel za matematiku
Sveučilište Josipa Jurja Strossmayera u Osijeku

Department of Mathematics
J. J. Strossmayer University of Osijek

Serije: Rukopisi u pripremi

Series: Technical Reports

Weighted median of the data in solving least absolute deviations problems

Ivan Vazler, Kristian Sabo, Rudolf Scitovski

Weighted median of the data in solving least absolute deviations problems[1]

Ivan Vazler, Kristian Sabo, Rudolf Scitovski
Department of Mathematics, University of Osijek
Trg Lj. Gaja 6, HR – 31 000 Osijek, Croatia

Abstract. We consider the weighted median problem for a given set of data and analyze its main properties. As an illustration, an efficient method for searching for a weighted Least Absolute Deviations (LAD)-line is given, which is used as the basis for solving various linear and nonlinear LAD-problems occurring in applications. Our method is illustrated by an example of hourly natural gas consumption forecast.

Key words. median, weighted median, least absolute deviations, LAD, weighted LAD-line

MSC(2000). 65D10, 65C20, 62J05, 90C27, 90C56, 90B85, 34K29

1 Introduction

The problems of estimating Least Absolute Deviations (LAD)-parameters for a linear regression, determining a LAD-hyperplane, and searching for a LAD-solution of an overdetermined system of linear equations, are all equivalent, and solving them boils down essentially to solving the weighted median problem (Bartels et al., 1978; Bazaraa et al., 2006; Cadzow, 2002; Castillo et al., 2008; Schöbel, 2003; Scholz, 1978).

These problems arise frequently in various branches of applied research, i.e. robotics, neural networks, signal and image processing, etc. (Cupec et al., 2009; Hodge and Austin, 2004; Koch, 1996; Rousseeuw and Leroy, 2003; Schöbel, 1999; Wang and Peterson, 2008). Particularly, solving the weighted median problem is used in many methods for outlier detection, for example in fraud detection, loan application processing, intrusion detection, activity monitoring, network performance, fault diagnosis, structural defect detection, satellite image analysis, detection of novelties in images, motion segmentation, time-series monitoring, medical condition monitoring, pharmaceutical research, detection of novelties in text, detection of unexpected entries in databases, detection of mislabeled data in a training data set, etc. (see Hodge and Austin (2004), Rousseeuw and Leroy (2003)).

The LAD principle is attributed to Josip Rudjer Bošković (1711–1787), Croatian scientist (mathematician, physicist, astronomer and philosopher) born in Dubrovnik (see e.g. Dodge (1987), Schöbel (2003)). Due to today's powerful computers, this principle has become quite popular and raised a great interest, as can be seen in numerous journal papers as well as at various international conferences. A series of such conferences is dedicated to J. R. Bošković (Dodge, 1987).

This principle was also applied by the French mathematician and astronomer marquis Pierre-Simon de Laplace (1749–1827) in five volumes of his comprehensive work *Mécanique*

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

Céleste (1798–1825), which was translated into English in 1829 by N. Bewditch[2], and which *inter alia* analyzes the form of the Earth's ellipsoid of rotation.

In this paper we consider the weighted median problem, analyze its most important properties, and illustrate it by showing an efficient method for searching for a weighted LAD-line. Solving all previously mentioned problems is based on this method.

2 The weighted median problem

Let (w_i, y_i), $i = 1, \ldots, m$, $m \geq 1$, be some given data, where y_i are real numbers and $w_i > 0$ are the corresponding data weights. Throughout this paper we will use y and w to denote the m-tuples $(y_1, \ldots, y_m)$ and $(w_1, \ldots, w_m)$, respectively. The function $f\colon \mathbb{R} \to \mathbb{R}$

$$f(\alpha) = \sum_{i=1}^{m} w_i \, |y_i - \alpha|, \tag{1}$$

is convex and attains its global minimum. The set $\mathrm{Med}(w, y)$ of all global minimizers (i.e. points of global minima) of the function f is convex, and as the following lemma shows (Sabo and Scitovski, 2008), it can be a one-point set, in which case it is one of the y_i's, or a segment between two subsequent data. Any element of the set $\mathrm{Med}(w, y)$ is called a *weighted median of the data* and we denote any of these by $\mathrm{med}(w, y)$.

Lemma 1. *Let (w_i, y_i), $i \in I = \{1, \ldots, m\}$, $m \geq 2$, be some data, where $y_1 \leq y_2 \leq \ldots \leq y_m$ are real numbers and $w_i > 0$ are the corresponding data weights. Then there exists a $\mu \in I$, such that $y_\mu \in \mathrm{Med}(w, y)$. Therefore, by denoting*

$$J := \Big\{ \nu \in I : \sum_{i=1}^{\nu} w_i \leq \frac{W}{2} \Big\},$$

where $W := \sum_{i=1}^{m} w_i$, the following holds:

(a) *if $J = \emptyset$, then $\mathrm{Med}(w, y) = \{y_1\}$;*

(b) *if $J \neq \emptyset$ and $\nu_0 := \max J$, then*

  (i) *if $\sum_{i=1}^{\nu_0} w_i < \frac{W}{2}$, then $\mathrm{Med}(w, y) = \{y_{\nu_0+1}\}$;*

  (ii) *if $\sum_{i=1}^{\nu_0} w_i = \frac{W}{2}$, then $\mathrm{Med}(w, y) = [y_{\nu_0}, y_{\nu_0+1}]$.*

Note that, if especially $m = 1$ or $y_1 = \ldots = y_m$, then $\mathrm{Med}(w, y) = \{y_1\}$.

The following two corollaries are direct consequences of Lemma 1. The first one describes a special case of unweighted data, or, equivalently, the case when all weights are the same (Sabo and Scitovski, 2008). In this case, the set of global minimizers will be denoted by $\mathrm{Med}(y)$, and any of its elements is called *median of the data* and denoted by $\mathrm{med}(y)$.

Corollary 1. *Let $y_1 \leq y_2 \leq \cdots \leq y_m$ be some real numbers. Then there exists a $\mu \in \{1, \ldots, m\}$, such that $y_\mu \in \mathrm{Med}(y)$. Consequently,*

[2] Available at `http://www.archive.org/details/mcaniquecles02laplrich`

(i) *if* m *is odd* $(m = 2k+1)$*, then* $\mathrm{Med}(y) = \{y_{k+1}\}$*;*

(ii) *if* m *is even* $(m = 2k)$*, then* $\mathrm{Med}(y) = [y_k, y_{k+1}]$.

The second corollary describes the important *pseudo-halving property* (Schöbel, 1999, 2003).

Corollary 2. *Let* (w_i, y_i)*,* $i = 1, \ldots, m$*,* $m \geq 1$*, be some data, where* y_i *are real numbers and* $w_i > 0$ *are the corresponding data weights. Then for all* $u \in \mathrm{Med}(w, y)$ *the following so-called* pseudo-halving property*, holds:*

$$\sum_{y_i < u} w_i \leq \frac{W}{2} \quad \textit{and} \quad \sum_{y_i > u} w_i \leq \frac{W}{2}, \tag{2}$$

where $W = \sum_{i=1}^{m} w_i$*.*

According to Lemma 1 and Corollary 2, it can be said that the weighted data median is any number $u \in \mathbb{R}$ for which the pseudo-halving property (2) holds.

Let us mention another important property of the weighted median of the data, which follows directly from Corollary 2.

Let $y = (y_1, \ldots, y_m)$ be an $m-$tuple of real numbers $y_i \in [a, b]$, let $w = (w_1, \ldots, w_m)$, $w_i > 0$, be the corresponding weights, and let $\phi\colon [a, b] \to \mathbb{R}$ be a strictly monotone function and $\alpha > 0$. Then from Corollary 2 it readily follows that

$$\mathrm{Med}(\alpha w, (\phi(y_1), \ldots, \phi(y_m))) = \phi(\mathrm{Med}(\alpha w, y)) = \phi(\mathrm{Med}(w, y)), \tag{3}$$

and especially if the function ϕ is an affine function,

$$\mathrm{Med}(\alpha w, \beta y + \gamma\, e) = \beta\, \mathrm{Med}(w, y) + \gamma, \tag{4}$$

for all $\beta, \gamma \in \mathbb{R}$, where $e = (1, \ldots, 1)$.[3]

Note also that in case the number of data is large, calculation of the weighted median of the data may require a lot of computing time (Cupec et al., 2009). Several fast algorithms can be found in Gurwitz (1990). In numerical examples at the end of this paper, the weighted median of the data is calculated by a modification of Algorithm 1 proposed in Gurwitz (1990).

3 An efficient method for searching a weighted LAD-line

To illustrate the median application, which is the basis for all aforementioned applications, we give an efficient algorithm for determining a weighted LAD-line. Let $\Lambda = \{T_i = (x_i, y_i) : i = 1, \ldots, m\}$ be a set of points in the plane with corresponding weights $w_i > 0$.

[3]For subsets $A, B \subseteq \mathbb{R}$ and real numbers α, β and γ, we denote $\alpha A + \beta B := \{\alpha a + \beta b : a \in A,\ b \in B\}$ and $A + \gamma := \{a + \gamma : a \in A\}$.

We are looking for a weighted LAD-line, i.e., we have to find the optimal parameters $a^*, b^* \in \mathbb{R}$ of the function $f(x; a, b) = ax + b$, such that

$$G(a^*, b^*) = \min_{(a,b)\in\mathbb{R}^2} G(a, b), \text{ where } G(a, b) = \sum_{i=1}^{m} w_i\,|y_i - ax_i - b|. \tag{5}$$

The functional G is convex and it always attains its global minimum on $\mathbb{R}^2$.

The algorithm we are going to construct is based on the fact that there always exists a weighted LAD-line passing through at least two points of Λ (see e.g. Bazaraa et al. (2006), Sabo and Scitovski (2008)). We first choose an initial point. It can be practically shown that choosing the centroid of the data set Λ is a good choice for the initial point (see e.g. Cupec et al. (2009)).

Next we choose a point in Λ in such a way that the line passing through these two points has the property that the sum over all points in Λ of absolute deviations of these points to the line is minimal. For this purpose the following lemma is used.

Lemma 2. *Let $\Lambda = \{T_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\}$, where $I = \{1, \dots, m\}$, $m \geq 2$, be a set of points in the plane such that $x_1 \leq \cdots \leq x_m$ and $x_1 < x_m$, and let $w_i > 0$ be the corresponding weights.*

Then for arbitrary $T_\mu = (x_\mu, y_\mu) \in \mathbb{R}^2$, there exists a $T_\nu = (x_\nu, y_\nu) \in \Lambda$, $x_\nu \neq x_\mu$, such that for $a^* = \frac{y_\nu - y_\mu}{x_\nu - x_\mu}$

$$\overline{G}(a; T_\mu) \geq \overline{G}(a^*; T_\mu), \quad \forall\, a \in \mathbb{R}, \tag{6}$$

where

$$\overline{G}(a; T_\mu) = \sum_{i=1}^{m} w_i\,|y_i - a(x_i - x_\mu) - y_\mu|. \tag{7}$$

Proof. Let $I_0 := \{i \in I : x_i = x_\mu\}$. Then

$$\overline{G}(a, T_\mu) = \sum_{i\in I_0} w_i\,|y_i - y_\mu| + \sum_{i\in I\setminus I_0} w_i\,|x_i - x_\mu| \left|\frac{y_i - y_\mu}{x_i - x_\mu} - a\right|,$$

and using Lemma 1 the assertion is proved. $\square$

Denote by $\mathcal{A}$ the algorithm which to any point $T_\mu = (x_\mu, y_\mu) \in \mathbb{R}^2$ associates the point $T_\nu \in \{(x_i, y_i) \in \Lambda : x_i \neq x_\mu\}$ and the slope $a^* = \frac{y_\nu - y_\mu}{x_\nu - x_\mu} \in \mathbb{R}$ as in Lemma 2. If there is more than one point that satisfies inequality (6), $\mathcal{A}$ should return the point and slope associated with the smallest index greater than μ or, if all indices are smaller, the one with the smallest index. We write this as $\mathcal{A}(T_\mu) = \{a^*, T_\nu\}$.

Generally, there may be an infinite number of weighted LAD-lines and not all of them must pass through at least two data points. Our goal is to construct an algorithm that will find a weighted LAD-line passing through at least two data points.

Now, based on a successive application of algorithm $\mathcal{A}$ and according to Lemma 2, we are going to define the *Weighted Two Points Algorithm (WTP)* for searching for a weighted LAD-line. In some iteration, the WTP algorithm starts from point $T_\mu \in \Lambda$ and by applying algorithm $\mathcal{A}$ it gives a slope a^* and a new point $T_\nu \in \Lambda$.

If in the next iteration, starting from point $T_\nu \in \Lambda$, we again obtain point T_μ (and naturally the same slope a^*), then

$$I_\nu := \{i \in I : a^*(x_i - x_\nu) + y_\nu = y_i\} = \{\mu, \nu\},$$

and the line $x \mapsto y_\mu + a^*(x - x_\mu)$ is a weighted LAD-line. This assertion will be proved in Theorem 1(i).

If $I_\nu \setminus \{\mu, \nu\} \neq \emptyset$, then the WTP algorithm will test points with indices from the set $I_\nu \setminus \{\mu, \nu\}$. That case is analyzed in Theorem 1(ii).

The Theorem 1 shows that parameters obtained by the WTP algoritm are best LAD-parameters.

WTP Algorithm[4]

Step 1: Input the set of points $T_i = (x_i, y_i)$, $i \in I = \{1, \dots, m\}$, $m \geq 2$, and
the corresponding data weights $w_i > 0$;
Determine the centroid of the data $T_p = (x_p, y_p)$, and
define the set $I_0 = \{i \in I : x_i = x_p\}$;

Step 2: According to Lemma 2, determine T_{i_1} and a_1 such that $\mathcal{A}(T_p) = \{a_1, T_{i_1}\}$, where $i_1 \in I \setminus I_0$;
Put $b_1 = -a_1 x_p - y_p$;
Set $j = 1$;

Step 3: Define a new set $I_0 = \{i \in I : x_i = x_{i_j}\}$, and
according to Lemma 2, determine $T_{i_{j+1}}$ and a_{j+1} such that $\mathcal{A}(T_{i_j}) = \{a_{j+1}, T_{i_{j+1}}\}$,
where $i_{j+1} \in I \setminus I_0$;
Put $b_{j+1} = -a_{j+1} x_{i_j} - y_{i_j}$;,

Step 4: If $i_{j+1} \notin \{i_1, i_2, \dots, i_j\}$, set $j = j + 1$ and go to Step 3;
Else, return (a_{j+1}, b_{j+1}) and STOP.

Theorem 1. *Let $\Lambda = \{T_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\}$, where $I = \{1, \dots, m\}$, $m \geq 2$, be a set of points in the plane such that $x_1 \leq \cdots \leq x_m$ and $x_1 < x_m$, let $w_i > 0$ be the corresponding weights, and let a_1^* and a_2^*, be such that $\mathcal{A}(T_\mu) = \{a_1^*, T_\nu\}$ and $\mathcal{A}(T_\nu) = \{a_2^*, T_k\}$. Denote $I_\nu := \{i \in I : y_\nu + a_2^*(x_i - x_\nu) = y_i\}$. Then*

(i) if $I_\nu = \{\mu, \nu\}$, then $T_k = T_\mu$ and $a_1^* = a_2^* =: a^*$, and $x \mapsto y_\mu + a^*(x - x_\mu)$ is a weighted LAD-line;

(ii) if $I_\nu \setminus \{\mu, \nu\} \neq \emptyset$ and for all $r \in I_\nu \setminus \{\mu, \nu\}$, $\mathcal{A}(T_r) = \{a_3^*, T_s\}$ with $a_3^* = \frac{y_s - y_r}{x_s - x_r}$, where $\overline{G}(a_2^*, T_\nu) = \overline{G}(a_3^*, T_r)$, then $x \mapsto y_\nu + a_2^*(x - x_\nu)$ is a weighted LAD-line.

[4] All evaluations and illustrations were done by using *Mathematica* 6 on a PC (Intel Core 2 Duo, 2 GB) and our own code available at `http://www.mathos.hr/seminar/software/WTP.m`

Proof. (*i*) Since $I_\nu = \{\mu, \nu\}$, it is obvious that $a_1^* = a_2^* =: a^*$ and $T_k = T_\mu$. Therefore, for every $i \in I$

$$\operatorname{sign}(y_i - y_\mu - a^*(x_i - x_\mu)) = \operatorname{sign}(y_i - y_\nu - a^*(x_i - x_\nu)) =: \sigma_i.$$

Since a^* is a global minimizer of both convex functionals $\overline{G}(a; T_\mu)$ and $\overline{G}(a; T_\nu)$, the real number 0 has to belong to their subdifferentials (see e.g. Bazaraa et al. (2006), Ruszczynski (2006)), i.e.,

$$\begin{aligned} 0 \in \partial\overline{G}(a^*; T_\mu) &= \sum_{i\in\{\mu,\nu\}} w_i[-1,1](x_i - x_\mu) - \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i(x_i - x_\mu), \\ 0 \in \partial\overline{G}(a^*; T_\nu) &= \sum_{i\in\{\mu,\nu\}} w_i[-1,1](x_i - x_\nu) - \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i(x_i - x_\nu), \end{aligned}$$

where $[-1, 1] = \{x \in \mathbb{R} : -1 \le x \le 1\}$.

Therefore, there exist real numbers $v_\mu, v_\nu \in [-1, 1]$ such that

$$\begin{aligned} w_\nu v_\nu &= \frac{1}{x_\nu - x_\mu} \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i(x_i - x_\mu), \\ w_\mu v_\mu &= \frac{1}{x_\mu - x_\nu} \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i(x_i - x_\nu), \end{aligned}$$

and hence

$$\begin{aligned} w_\mu x_\mu v_\mu + w_\nu x_\nu v_\nu &= \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i x_i, \\ w_\mu v_\mu + w_\nu v_\nu &= \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i, \end{aligned}$$

which means that

$$(0,0) \in \partial G(a^*, b^*) = \sum_{i\in\{\mu,\nu\}} w_i[-1,1](x_i, 1) - \sum_{i\in I\setminus\{\mu,\nu\}} w_i\sigma_i(x_i, 1), \qquad b^* = y_\mu - a^* x_\mu.$$

Hence, (a^*, b^*) is the global minimizer of the functional G, i.e., $x \mapsto y_\mu + a^*(x - x_\mu)$ is a weighted LAD-line.

(*ii*) In case $x \mapsto y_\nu + a_2^*(x - x_\nu)$ is not a weighted LAD-line, there exists an $r \in I_\nu \setminus \{\mu, \nu\}$, such that, by Lemma 2, there exist $s \in I \setminus I_\nu$ such that

$$\overline{G}(a; T_r) > \overline{G}(a_3^*; T_r), \qquad a_3^* = \frac{y_s - y_r}{x_s - x_r}. \tag{8}$$

Furthermore, using (8) and $y_r - a_2^*(x_r - x_\nu) - y_\nu = 0$, we obtain

$$\begin{aligned}
\overline{G}(a_2^*; T_\nu) &= \sum_{i=1}^m w_i \left| y_i - a_2^*(x_i - x_\nu) - y_\nu \right| \\
&= \sum_{i=1}^m w_i \left| y_i - \frac{y_\nu - y_k}{x_\nu - x_k}(x_i - x_\nu) - y_\nu \right| \\
&= \sum_{i=1}^m w_i \left| y_i - \frac{y_r - y_k}{x_r - x_k}(x_i - x_r) - y_r \right| \\
&> \overline{G}(a_3^*, T_r).
\end{aligned}$$

□

In Example 1 the WTP Algorithm is illustrated on an example in which degeneration appears (see Yan (2003)).

Example 1. Using the WTP Algorithm we search for a weighted LAD-line for data points (x_i, y_i), $i = 1, \ldots, 7$, with weights $w_i > 0$, given below :

w_i	2	1	1	1	4	2	2
x_i	1	1	3	4	6	8	11
y_i	1	2	2	2	2	2	3

The flow of the iterative process is shown in Table 1, and Fig. 1 shows the data points and lines corresponding to the initial and the last iteration. Bigger black points correspond to data points with larger weights.

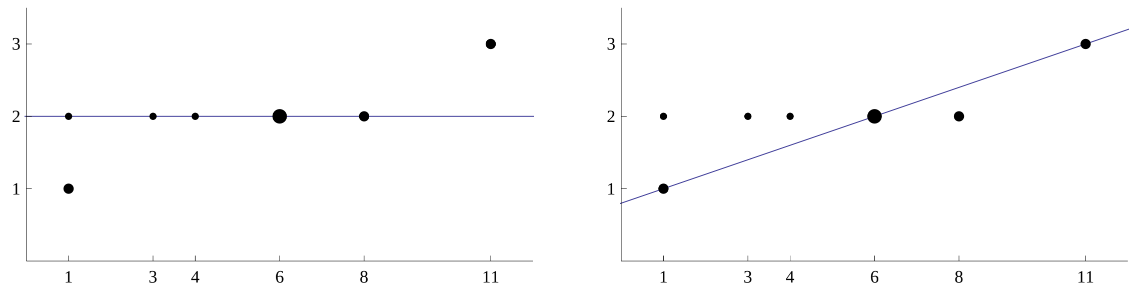


Figure 1: Data points and the initial and the last iteration

4 An application to the natural gas consumption forecast

Our algorithm for searching for a best LAD-line will be illustrated on the problem of natural gas consumption forecast. We consider the data (τ_i, y_i), $i = 1, \ldots, 24 \times m$, where

Iteration (i)	First Point	Second Point	a_i	b_i	$G(a_i, b_i)$
1	$(1, 2)$	$(3, 2)$	0	2	4
2	$(3, 2)$	$(4, 2)$	0	2	4
3	$(4, 2)$	$(6, 2)$	0	2	4
4	$(6, 2)$	$(11, 3)$	0.2	0.8	2.8
5	$(11, 3)$	$(1, 1)$	0.2	0.8	2.8
6	$(1, 1)$	$(6, 2)$	0.2	0.8	2.8

Table 1: Iterative process WTP

τ_i is the temperature and y_i is the consumption of natural gas per hour, during a certain m-day period in one part of the city of Osijek (Croatia) (see Fig. 2). We should forecast natural gas consumption Y_{m+1} for the next day, assuming that in that period no large industrial users were active.

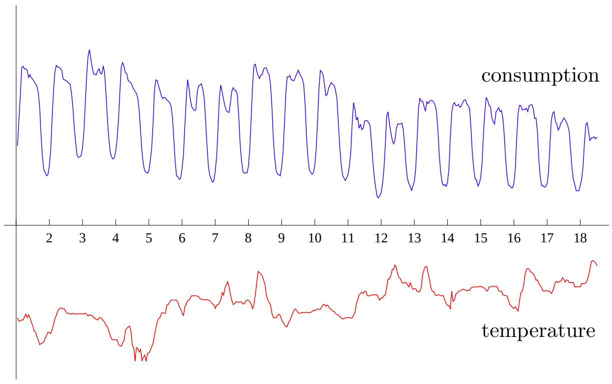


Figure 2: Hourly temperature and natural gas consumption

According to Brabec et al. (2008), there exists a functional relationship between natural gas consumption and the temperature, which may be expressed by Gompertz model function: $Y(\tau) = e^{a-be^{c\tau}}$, $a, b, c > 0$. Therefore, suppose that the total natural gas consumption on the $(m+1)$-th day depends on the average temperature $\overline{\tau}_m$ on the m-th day. On the basis of the data $(\overline{\tau}_i, Y_i)$, $i = 1, \ldots, m$, where $\overline{\tau}_i$ is the average temperature and Y_i is the total consumption on the i-th day, we determine the optimal parameters a^*, b^*, and c^* of the Gompertz model function (see Fig. 3).

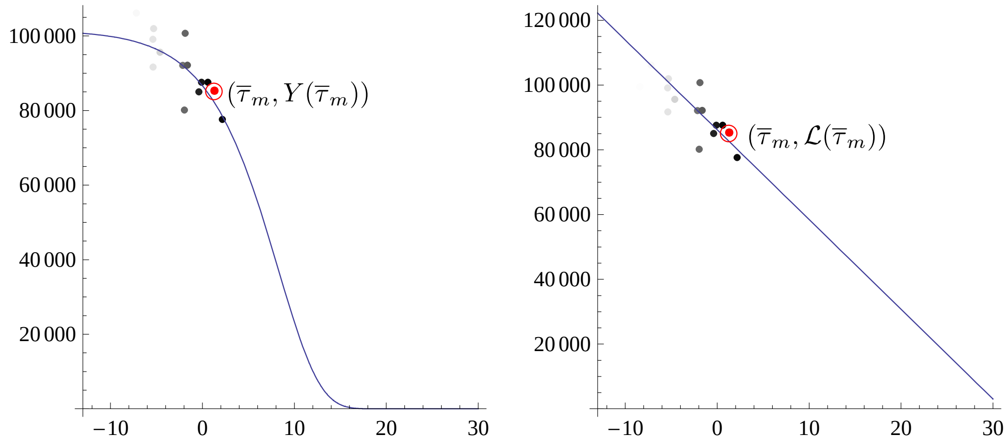
Since occasionally some large industrial users can become active, thereby causing outliers, we will apply the LAD-principle in such a way that the estimated parameters are more significantly influenced by data referring to consumption on those days when the average temperature was close to $\overline{\tau}_m$ (darker gray points), and less influenced by data for days when the average temperature was significantly different from $\overline{\tau}_m$ (lighter gray points). This will be achieved by minimizing the functional

$$F(a, b, c) = \sum_{i=1}^{m} w_i \left| Y_i - e^{a-be^{c\overline{\tau}_i}} \right|, \quad a, b, c > 0, \tag{9}$$

where $w_i > 0$ are the data-weights defined in the following way (Scitovski et al., 1998):

$$w_i = W\left(\frac{|\overline{\tau}_i - \overline{\tau}_m|}{r_i}\right), \text{ with } W(u) := \begin{cases} (1-u^3)^3, & 0 \le u \le 1, \\ 0, & u > 1. \end{cases}, \tag{10}$$

where the parameter $r_i > 0$ defines the range of influence of the i-th datum.



(a) Gompertz model function (b) Linear model function

Figure 3: Data points and Gompertz and linear model function

Using the optimal parameters a^*, b^*, and c^*, we estimate natural gas consumption on the following $(m+1)$-th day as $Y(\overline{\tau}_m) = e^{a^* - b^* e^{c^* \overline{\tau}_m}}$. Fig. 3a shows the data $(\overline{\tau}_i, Y_i)$, $i = 1, \ldots, m$ $(m = 15)$, and the corresponding Gompertz model function. The forecast of natural gas consumption for the $(m+1)$-th day, that is linked to the average temperature $\overline{\tau}_m$, is emphasized by a circle.

In practice, we are interested only in the behavior of gas consumption for the average temperature close to $\overline{\tau}_m$, so the Gompertz model function can be approximated by a linear model function $\mathcal{L}(\overline{\tau}) = \alpha\overline{\tau} + \beta$, whose parameters α, β can be determined by minimizing the functional

$$\Phi(a, b, c) = \sum_{i=1}^{m} w_i \, |Y_i - \alpha\overline{\tau}_i - \beta|, \tag{11}$$

where data weights $w_i > 0$ can also be determined from (10). Hence, the problem is reduced to the problem of determining a weighted LAD-line.

Fig. 3b shows the data and the corresponding linear model function. The forecast of natural gas consumption for the $(m+1)$-th day, that is linked to the average temperature $\overline{\tau}_m$, is emphasized by a circle.

Table 2 lists LAD parameters for the Gompertz and linear model function used to estimate natural gas consumption on the 14-th, 16-th, and the 18-th day. Table 3 gives a comparison of the forecast of natural gas consumption on the 14-th, 16-th, and the 18-th day, based on the Gompertz model function $Y(\overline{\tau}_m)$ and the corresponding linear approximation $\mathcal{L}(\overline{\tau}_m)$, with actual consumption Y_{m+1} on these days. From the above illustrations and the experiment performed one can conclude that, for practical purposes, linear approximation gives an acceptable forecast.

		Gompertz model function			Linear model function	
Day	$\overline{\tau}_m$	a^*	b^*	c^*	α^*	β^*
14	-0.1^oC	11.5351	0.16077	0.216611	-1924.68	88738.8
16	1.3^oC	11.5288	0.158752	0.222853	-2771.97	86188.4
18	2.8^oC	11.4653	0.0853622	0.387744	-3792.5	86005.4

Table 2: Weighted LAD parameters of the Gompertz and linear model function

			Gompertz model		Linear model	
Day	$\overline{\tau}_m$	Y_{m+1}	$Y(\overline{\tau}_m)$	Rel. err. (in %)	$\mathcal{Y}(\overline{\tau}_m)$	Rel. err. (in %)
14	-0.1^oC	2.10163×10^6	2.09633×10^6	0.3	2.13416×10^6	1.5
16	1.3^oC	2.04778×10^6	1.97234×10^6	3.7	1.98204×10^6	3.2
18	2.8^oC	1.73542×10^6	1.76911×10^6	1.9	1.8051×10^6	4.0

Table 3: Forecast of natural gas consumption

5 Conclusion

The method for searching for a weighted LAD-line, which is given in this paper, is geometrically motivated and it includes both the nondegenerate and the degenerate case. An efficient WTP algorithm is constructed based on the mentioned method. Thereby it was taken into consideration that the number of data may be very large, and that the solution should be reached in a very short period of time.

In numerous applied researches (e.g. the forecast of natural gas consumption, Brabec et al. (2008), or the movement of robots, Cupec et al. (2009)) it is extremely important to have such an algorithm, which always gives a solution to the fundamental problem, reliably and in real time.

Acknowledgement. We would like to thank an anonymous referee for useful comments and remarks, which helped us improve the paper significantly.

References

R. H. Bartels, A. R. Conn, J. W. Sinclair, *Minimization techniques for piecewise differentiable functions: The l_1 solution to an overdetermined linear system*, SIAM J. Numer. Anal. **15**(1978), 224–241

M. S. Bazaraa, H. D. Sherali, C. M. Shetty, *Nonlinear Programming. Theory and Algorithms. 3^{rd} Edition*, (Wiley, New Jersey, 2006).

M. Brabec, O. Konár, E. Pelikán, M. Malý, *A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers*, International Journal of Forecasting **24**(2008), 659–678

J. A. Cadzow, *Minimum l_1, l_2 and $l\infty$ Norm Approximate Solutions to an Overdetermined System of Linear Equations*, Digital Signal Processing **12**(2002), 524–560

E. Castillo, R. Mínguez, C. Castillo, A. S. Cofinño, *Dealing with the multiplicity of solutions of the l_1 and l_∞ regression models*, European J. Oper. Res. **188**(2008), 460–484

R. Cupec, R. Grbić, K. Sabo, R. Scitovski, *Three points method for searching the best least absolute deviations plane*, Applied Mathematics and Computation, **215**(2009) 983–994

Y. Dodge, *An introduction to L_1-norm based statistical data analysis*, Computational Statistics & Data Analysis **5**(1987), 239–253

C. Gurwitz, *Weighted median algorithms for L_1 approximation*, BIT **30**(1990), 301–310

V. J. Hodge, J. Austin, *A survey of outlier detection methodologies*, Artificial Intelligence Review **22**(2004), 85–126

I. Koch, *On the asymptotic performance of median smoothers in image analysis and nonparametric regression*, The Annals of Statistics **24**(1996), 1648–1666

P. J. Rousseeuw, A. M. Leroy, *Robust Regression and Outlier Detection*, (Wiley, New York, 2003).

A. Ruszczynski, *Nonlinear Optimization*, (Princeton University Press, Princeton and Oxford, 2006).

K. Sabo, R. Scitovski, *The best least absolute deviations line – properties and two efficient methods*, ANZIAM Journal **50**(2008), 185–198

A. Schöbel, *Locating Lines and Hyperplanes: Theory and Algorithms*, (Springer-Verlag, Berlin, 1999).

A. Schöbel, *Anchored hyperplane location problems*, Discrete & Computational Geometry **29**(2003), 229–238

F. W. Scholz, *Weighted median regression estimates*, The Annals of Statistics **6**(1978), 603–609

R. Scitovski, Š. Ungar, D. Jukić, *Approximating surface by moving total least squares method*, Applied Mathematics and Computation **93**(1998), 219–232

M. Shi, M. A. Lukas, *An L_1 estimation algorithm with degeneracy and linear constraints*, Computational Statistics & Data Analysis **39**(2002), 35–55

Z. Wang, B. S. Peterson, *Constrained least absolute deviation neural networks*, IEEE Transactions on Neural Networks **19**(2008), 273–283

G. A. Watson, *Approximation theory and numerical methods*, (John Wiley & Sons, Chichester, 1980).

S. Yan, *A Base-point descent algorithm for solving the linear l_1 problem*, Intern. J. Computer Math., **80**(2003), 367–380