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

Searching for a Best LAD-Solution of an Overdetermined System of Linear Equations Motivated by Searching for a Best LAD-Hyperplane on the Basis of Given Data

Kristian Sabo, Rudolf Scitovski,
Ivan Vazler

Searching for a Best LAD-Solution of an Overdetermined System of Linear Equations Motivated by Searching for a Best LAD-Hyperplane on the Basis of Given Data*

Kristian Sabo† Rudolf Scitovski‡ Ivan Vazler§

Abstract. We consider the problem of searching for a best LAD-solution of an overdetermined system of linear equations $\mathbf{X}\mathbf{a} = \mathbf{z}$, $\mathbf{X} \in \mathbb{R}^{m\times n}$, $m \geq n$, $\mathbf{a} \in \mathbb{R}^n, \mathbf{z} \in \mathbb{R}^m$. This problem is equivalent to the problem of determining a best LAD-hyperplane $\mathbf{x} \mapsto \mathbf{a}^T\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$ on the basis of given data $(\mathbf{x}_i, z_i)$, $\mathbf{x}_i = (x_1^{(i)}, \ldots, x_n^{(i)})^T \in \mathbb{R}^n$, $z_i \in \mathbb{R}$, $i = 1, \ldots, m$, whereby the minimizing functional is of the form

$$F(\mathbf{a}) = \|\mathbf{z} - \mathbf{X}\mathbf{a}\|_1 = \sum_{i=1}^{m} |z_i - \mathbf{a}^T\mathbf{x}_i|.$$

An iterative procedure is constructed as a sequence of weighted median problems, which gives the solution in finitely many steps. A criterion of optimality follows from the fact that the minimizing functional F is convex, and therefore the point $\mathbf{a}^* \in \mathbb{R}^n$ is the point of a global minimum of the functional F if and only if $\mathbf{0} \in \partial F(\mathbf{a}^*)$.

Motivation for the construction of the algorithm was found in a geometrically visible algorithm for determining a best LAD-plane $(x, y) \mapsto \alpha x + \beta y$, passing through the origin of the coordinate system, on the basis of the data (x_i, y_i, z_i), $i = 1, \ldots, m$.

Key words: LAD; least absolute deviations; overdetermined system of linear equations; l_1-norm approximation; weighted median problem; outliers; LAD-hyperplane

AMS Classification (2010): 65F20, 41A28, 57Q55, 68W25, 90C59

1 Introduction

We consider the problem of searching for a best Least Absolute Deviations (LAD) solution of an overdetermined system of linear equations (see e. g. [1, 2, 3, 4, 5, 6, 7, 8]):

Let $\mathbf{X}\mathbf{a} = \mathbf{z}$, where $\mathbf{X} \in \mathbb{R}^{m\times n}$, $m \geq n$, is a matrix of full column rank, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{z} \in \mathbb{R}^m$, and

$$F(\mathbf{a}) = \|\mathbf{z} - \mathbf{X}\mathbf{a}\|_1 = \sum_{i\in I} |r_i(\mathbf{a})|, \quad r_i(\mathbf{a}) = z_i - \mathbf{a}^T\mathbf{x}_i, \tag{1}$$

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

†K. Sabo, Department of Mathematics, University of Osijek, Croatia, e-mail: `ksabo@mathos.hr`

‡R. Scitovski (corresponding author), Department of Mathematics, University of Osijek, Croatia, e-mail: `scitowsk@mathos.hr`

§I. Vazler, Department of Mathematics, University of Osijek, Croatia, e-mail: `ivazler@mathos.hr`

where $I = \{1, \dots, m\}$, and $\mathbf{x}_i^T$ is the i-th row of the matrix $\mathbf{X}$. Functional F is convex and it attains its global minimum $\mathbf{a}^* \in \mathbb{R}^n$. This point is called a LAD-solution of an overdetermined system of linear equations $\mathbf{X}\mathbf{a} = \mathbf{z}$.

The same problem can be considered as the problem of estimation of optimal parameters of a best LAD-hyperplane $\mathbf{x} \mapsto \mathbf{a}^T\mathbf{x}$, $\mathbf{x} \in \mathbb{R}^n$, on the basis of the given set of experimental data $(\mathbf{x}_i, z_i)$, $\mathbf{x}_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n$, $z_i \in \mathbb{R}$, $i = 1, \dots, m$ (see e. g. [9, 10, 11, 12, 13]). In statistical literature this problem is also considered as the problem of estimating LAD-optimal parameters of linear regression (see e. g. [14, 15, 16]).

Example 1.1. *Let us consider the simplest case as an illustration: the system* $\mathbf{X}\alpha = \mathbf{z}$*,* $\mathbf{X} \in \mathbb{R}^{m\times 1}$*,* $\mathbf{z} \in \mathbb{R}^m$*.*

Searching for a best LAD-solution of this system can be considered as a problem of searching for a best LAD-line $z = \alpha x$, whose graph passes through the origin of the coordinate system, on the basis of the given set of data points $\Lambda = \{T_i = (x_i, z_i) : i \in I\}$, $\quad I = \{1, \dots, m\}$, $m \geq 1$ (see *Fig. 1*). With the notation $I' = \{i \in I : x_i = 0\}$, our problem is reduced to the minimization problem

$$\min_{\alpha\in\mathbb{R}} \sum_{i\in I} |z_i - \alpha x_i| = \sum_{i\in I'} |z_i| + \min_{\alpha\in\mathbb{R}} \sum_{i\in I\setminus I'} |x_i| \left| \frac{z_i}{x_i} - \alpha \right|, \tag{2}$$

known in literature as the *Weighted Median Problem* (see e.g. [5, 17, 18, 19]). The following lemma ([18]) gives properties and a solution of the weighted median problem.

Lemma 1.1. *Let* (w_i, y_i)*,* $i \in I = \{1, \dots, m\}$*,* $m \geq 2$*, be the data, where* $y_1 \leq y_2 \leq \dots \leq y_m$ *are real numbers and* $w_i > 0$ *corresponding data weights. Denote*

$$J = \{\nu \in I : 2\sum_{i=1}^{\nu} w_i - \sum_{i=1}^{m} w_i \leq 0\}.$$

For $J \neq \emptyset$*, let us denote* $\nu_0 = \max J$*. Furthermore, let* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a function defined by the formula*

$$\varphi(\alpha) = \sum_{i=1}^{m} w_i |y_i - \alpha|.$$

Then

(i) if $J = \emptyset$ *(i. e.* $2w_1 > \sum_{i=1}^m w_i$*), then the minimum of function* φ *is attained at the point* $\alpha^\star = y_1$*.*

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

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

By using *Lemma 1.1* we can carry out a simple analysis of our problem (2):

- if $I' = I$, then $\mathbf{X} = (0, \dots, 0)^T$, rank $(\mathbf{X}) = 0$, and each line $z = \alpha x$, $\alpha \in \mathbb{R}$ is a solution of problem (2) (see *Fig. 1.a*);

- if $I' = \emptyset$, whereby $x_1 = \cdots = x_m \neq 0$, then $\mathbf{X} = (x_1, \ldots, x_1)^T$, $\operatorname{rank}(\mathbf{X}) = 1$. Then (2) becomes (see *Fig. 1.b*)
$$\min_{\alpha\in\mathbb{R}} \sum_{i\in I} |z_i - \alpha x_i| = |x_1| \min_{\alpha\in\mathbb{R}} \sum_{i\in I} |\frac{z_i}{x_1} - \alpha|,$$
and $\alpha^* = \operatorname*{med}_{i\in I}(\frac{z_i}{x_1}) = \frac{1}{x_1} \operatorname*{med}_{i\in I} z_i$ is a solution of problem (2);

- if $I \setminus I' \neq \emptyset$, whereby $0 < x_1 < x_m$, then $\mathbf{X} = (x_1, \ldots, x_m)^T$, $\operatorname{rank}(\mathbf{X}) = 1$. In this case the solution of problem (2) is $\alpha^* = \operatorname*{med}_{i\in I\setminus I'} \left(|x_i|, \frac{z_i}{x_i}\right)$ (see *Fig. 1.c*).

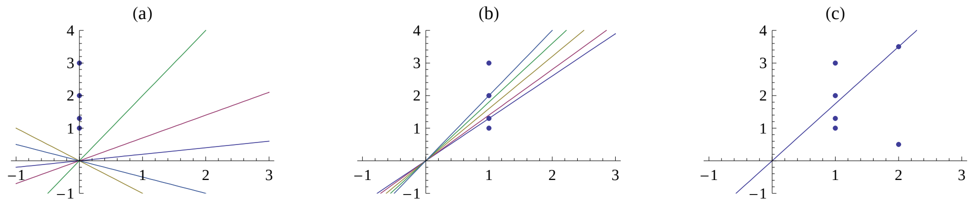

Figure 1: Best LAD-line passing through the origin

The best LAD-solution of an overdetermined system of linear equations is important in various fields of applied research, especially in the case if among the data a substantial amount of outliers (i.e. wild points) might appear (see e. g. [14, 20, 21, 5, 22, 23]).

This principle is considered to have been proposed by the Croatian mathematician J. R. Bošković in the mid-eighteenth century (see e. g. [14, 24]). The best LAD-solution has a property that it is less sensitive to extreme errors (outliers), it appears among the data, and points out the influence of the majority of data reflecting the real nature of the problem (see e. g. [24, 20, 17]).

Classical nondifferentiable minimization methods cannot be applied directly to searching the best LAD-solution since unreasonably long computing time would be necessary or we obtain a bad approximation of the solution. That is the reason why various specialized algorithms for solving this problem have been developed lately (see e. g. [2, 20, 15, 25, 10, 5, 26, 6, 8]).

2 The System $m \times 2$ and the Plane $(x, y) \mapsto \alpha x + \beta y$

Searching for a best LAD-solution of the system
$$\mathbf{X}\mathbf{a} = \mathbf{z}, \qquad \mathbf{X} = \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_m & y_m \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \tag{3}$$
can be considered as a problem of searching for a best LAD-plane
$$z(x, y) = \alpha x + \beta y, \tag{4}$$
whose graph passes through the origin of the coordinate system, on the basis of the given data points
$$\Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}, \qquad I = \{1, \ldots, m\}, \quad m \geq 2.$$

In both cases the problem is reduced to minimizing the functional $F:\mathbb{R}^2\to\mathbb{R}$,

$$F(\alpha,\beta)=\sum_{i=1}^{m}|z_i-\alpha x_i-\beta y_i|, \tag{5}$$

which always attains its global minimum at $\mathbb{R}^2$. Solving this problem can be geometrically clearly represented creating in that way sound assumptions for solving a general problem. Searching for the aforementioned best LAD-plane is a generalization of the approach cited in [18], and a special case of the approach given in [10].

Let us introduce the following notations:

- $\mathcal{L}=\{P_i=(x_i,y_i)\in\mathbb{R}^2 : i\in I\}$ — projection of the set Λ on the (x,y)-plane;
- $M(O,T_\mu)$ — the plane containing a z-axis passing through the origin O and the point $T_\mu\in\Lambda$, i.e.

$$M(O,T_\mu)=\{(x,y,z)\in\mathbb{R}^3 : \begin{vmatrix} x_\mu & y_\mu \\ x & y \end{vmatrix}=0\}. \tag{6}$$

  Note that the point $T=(x,y,z)$ lies on the plane $M(O,T_\mu)$ if and only if vectors $(x,y)^T$ and $(x_\mu,y_\mu)^T$ are linearly dependent or if the point $P=(x,y)$ lies on the line determined by the points O and $P_\mu\in\mathcal{L}$;

- $I_\mu=\{i\in I : T_i\in M(O,T_\mu)\}$ — the set of indices of those points from Λ lying on the plane $M(O,T_\mu)$. Note that $i\in I\setminus I_\mu$ if and only if the point $T_i\in\Lambda$ does not lie on the plane $M(O,T_\mu)$, i.e. if the corresponding point $P_i\in\mathcal{L}$ does not lie on the line determined by the points O and P_μ. On the other hand, in this case the i-th row of the matrix X in (3) is linearly independent of its μ-th row.

Lemma 2.1. *Let $I=\{1,\dots,m\}$, $m\geq 2$, be a set of indices and let*

(i) $\Lambda=\{T_i=(x_i,y_i,z_i) : i\in I\}$ be a set of points in space such that its projection $\mathcal{L}$ on the (x,y)-plane does not lie on some line passing through the origin;

(ii) $T_\mu=(x_\mu,y_\mu,z_\mu)\in\Lambda$, such that $(x_\mu,y_\mu)\neq(0,0)$ and $I_\mu=\{i\in I : T_i\in M(O,T_\mu)\}$.

Then there exists $\nu\in I\setminus I_\mu$, such that a best LAD-plane of the form (4), containing points O and T_μ, also passes through the point $T_\nu\in\Lambda\setminus M(O,T_\mu)$.

Proof. The plane of the form (4) passing through the origin O and the point $T_\mu=(x_\mu,y_\mu,x_\mu)$ can be written in the form

$$z'=\frac{x}{x_\mu}(z_\mu-\beta' y_\mu)+\beta' y,\quad x_\mu\neq 0\quad \text{or}\quad z''=\alpha'' x+\frac{y}{y_\mu}(z_\mu-\alpha'' x_\mu),\quad y_\mu\neq 0, \tag{7}$$

whereby the parameter β' (i.e. parameter α'') can be found by minimizing the functional

$$\beta'\mapsto\sum_{i\in I\setminus I_\mu}\left|z_i x_\mu-x_i z_\mu-\beta'\left(y_i x_\mu-x_i y_\mu\right)\right|,\quad x_\mu\neq 0, \tag{8}$$

i.e. by minimizing the functional

$$\alpha''\mapsto\sum_{i\in I\setminus I_\mu}\left|y_i z_\mu-z_i y_\mu-\alpha''\left(y_i x_\mu-x_i y_\mu\right)\right|,\quad y_\mu\neq 0. \tag{9}$$

Due to condition *(i)*, the set $I \setminus I_\mu = \{i \in I : T_i \notin M(O, T_\mu)\} \neq \emptyset$ and the parameter β', i. e. α'', can be determined as a weighted median by minimizing functional (8), i.e. by minimizing functional (9). By the notation $w_i = \begin{vmatrix} x_\mu & y_\mu \\ x_i & y_i \end{vmatrix}$, there exists $\nu' \in I \setminus I_\mu$ such that

$$\beta' = \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \frac{1}{w_i} \begin{vmatrix} x_\mu & z_\mu \\ x_i & z_i \end{vmatrix} \right) = \frac{1}{w_{\nu'}} \begin{vmatrix} x_\mu & z_\mu \\ x_{\nu'} & z_{\nu'} \end{vmatrix}, \qquad \alpha' = \frac{1}{x_\mu}(z_\mu - y_\mu \beta'), \quad x_\mu \neq 0, \tag{10}$$

i.e. there exists $\nu'' \in I \setminus I_\mu$ such that

$$\alpha'' = \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \frac{1}{w_i} \begin{vmatrix} z_\mu & y_\mu \\ z_i & y_i \end{vmatrix} \right) = \frac{1}{w_{\nu''}} \begin{vmatrix} z_\mu & y_\mu \\ z_{\nu''} & y_{\nu''} \end{vmatrix}, \qquad \beta'' = \frac{1}{y_\mu}(z_\mu - x_\mu \alpha''), \quad y_\mu \neq 0. \tag{11}$$

□

Remark 2.1. Note that (α', β'), i.e. (α'', β''), is a solution of the system

$$\begin{bmatrix} x_\mu & y_\mu \\ x_{\nu'} & y_{\nu'} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} z_\mu \\ z_{\nu'} \end{bmatrix}, \qquad \text{i.e.} \quad \begin{bmatrix} x_\mu & y_\mu \\ x_{\nu''} & y_{\nu''} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} z_\mu \\ z_{\nu''} \end{bmatrix}, \tag{12}$$

whereby corresponding matrices are nonsingular.

Lemma 2.2. *By assumptions as in* Lemma 2.1 *there holds:*

(i) If $x_\mu = 0$, i.e. $y_\mu = 0$, then a best LAD-plane from Lemma 2.1*, passing through the point T_μ, is of the form $z'' = \alpha'' x + \frac{z_\mu}{y_\mu} y$, i.e of the form $z' = \frac{z_\mu}{x_\mu} x + \beta' y$;*

(ii) If $(x_\mu, y_\mu) \neq (0, 0)$, best LAD-planes z', z'' of the form (7) correspond.

Proof. Assertion (i) is evident. For the purpose of proving assertion (ii) it suffices to show that $\alpha' = \beta'$ and $\alpha'' = \beta''$. By using the following property of the median (see e. g. [17])

$$\operatorname*{med}_i (p_i, cu_i + v) = c \operatorname*{med}_i (p_i, u_i) + v, \qquad u_i, c, v \in \mathbb{R},\ p_i > 0, \tag{13}$$

we have

$$\begin{aligned} \alpha' = \tfrac{z_\mu}{x_\mu} - \tfrac{y_\mu}{x_\mu}\beta' &= \tfrac{z_\mu}{x_\mu} - \tfrac{y_\mu}{x_\mu} \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \tfrac{1}{w_i} \begin{vmatrix} x_\mu & z_\mu \\ x_i & z_i \end{vmatrix} \right) \\ &= \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \tfrac{z_\mu}{x_\mu} - \tfrac{y_\mu}{x_\mu} \tfrac{1}{w_i} \begin{vmatrix} x_\mu & z_\mu \\ x_i & z_i \end{vmatrix} \right) \\ &= \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \tfrac{1}{w_i} \begin{vmatrix} z_\mu & y_\mu \\ z_i & y_i \end{vmatrix} \right) = \alpha''. \end{aligned}$$

Similarly, it can be shown that $\beta' = \beta''$ holds. □

2.1 Searching for a Best LAD-plane

In accordance with *Lemma 2.1* we construct an algorithm that will search for a best LAD-plane passing through the origin on the basis of the given data points. First, the first point T_μ must be selected in the algorithm such that $(x_\mu, y_\mu) \neq (0, 0)$. After that, according to *Lemma 2.1*, we determine the next point $T_\nu \in \Lambda \setminus \{T_\mu\}$.

Next, and again according to *Lemma 2.1*, by the point T_ν we determine the following point $T_k \in \Lambda \setminus \{T_\nu\}$. If $T_k = T_\mu$, the procedure is finished; otherwise we repeat the procedure. Such stopping criterion of the iterative procedure can be seen for example in [9, 10]. However, if there exist more indices from the set $I \setminus I_\nu$ on which the weighted median is attained, i.e. if the plane passing through points T_μ and T_ν also contains some other points, e.g. point $T_{\nu'}$, the algorithm should continue with points T_μ and $T_{\nu'}$. Only after all points lying on the aforementioned plane have been used the algorithm can be stopped. Such situation cannot take place if the data have the property that no three points lie on the plane passing through the origin, i.e. if the augmented matrix $[\mathbf{X}; \mathbf{z}]$ satisfies the Haar condition (see e. g. [20]). By means of geometrical analysis in *Section 2.3* and by illustrative examples in *Section 4* we will illustrate all cases that might occur.

Algorithm I.[1]

Step 1: Input the set of points $\Lambda = \{T_i = (x_i, y_i, z_i) : i \in I\}$, $I = \{1, \dots, m\}$ and check condition (i) from *Lemma 2.1*. Choose the point $T_\mu(x_\mu, y_\mu, z_\mu) \in \Lambda$ such that $(x_\mu, y_\mu) \neq (0, 0)$ and set $\nu = \mu$, $\gamma = 0$;

Step 2: If $\gamma = \nu$, STOP;

Else set $\gamma = \mu$, $\mu = \nu$ and define the set $I \setminus I_\mu$ and the numbers w_i in the following way:

$$i \in I \setminus I_\mu \Longleftrightarrow w_i = \begin{vmatrix} x_\mu & y_\mu \\ x_i & y_i \end{vmatrix} \neq 0,$$

Step 3: If $y_\mu \neq 0$, determine

$$\hat{\alpha} = \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \tfrac{1}{w_i} \begin{vmatrix} z_\mu & y_\mu \\ z_i & y_i \end{vmatrix} \right); \qquad \hat{\beta} = \tfrac{z_\mu}{y_\mu} - \tfrac{x_\mu}{y_\mu} \hat{\alpha}; \tag{14}$$

Else determine

$$\hat{\beta} = \operatorname*{med}_{i \in I \setminus I_\mu} \left(|w_i|, \tfrac{1}{w_i} \begin{vmatrix} x_\mu & z_\mu \\ x_i & z_i \end{vmatrix} \right); \qquad \hat{\alpha} = \tfrac{z_\mu}{x_\mu} - \tfrac{y_\mu}{x_\mu} \hat{\beta}, \tag{15}$$

by which $\nu \in I \setminus I_\mu$ is determined for which the median in (14), i.e. (15), is attained and go to *Step 2*.

Remark 2.2. Note that $(\hat{\alpha}, \hat{\beta})^T$ in some step of Algorithm I is a solution of the system (see also *Remark 2.1*)

$$\mathbf{B}\mathbf{a} = \begin{bmatrix} z_\mu \\ z_\nu \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} x_\mu & y_\mu \\ x_\nu & y_\nu \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{16}$$

and the corresponding plane passes through the points $T_\mu, T_\nu \in \Lambda$. Since $\nu \in I \setminus I_\mu$, the matrix $\mathbf{B}$ is nonsingular, so that the solution of system (16) is unique.

The following theorem shows how a sequence of approximations $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \dots$ from Algorithm I can be successively defined as an iterative process of the form

$$\bar{\mathbf{a}} = \mathbf{a} + \vartheta \mathbf{p}, \tag{17}$$

where $\mathbf{p} \in \mathbb{R}^2$ is the direction vector and $\vartheta \in \mathbb{R}$ step length. Thereby if the parameter vector $\mathbf{a}$ in *Lemma 2.1* is defined by the points T_μ, T_ν, then a new-better parameter vector $\bar{\mathbf{a}}$ can be determined according to *Lemma 2.1* only if we start from the point T_ν (drop the point T_μ). The following theorem

[1] All evaluations and illustrations were done using *Mathematica 6* on a PC (CPU: 2.00 GHz Intel Core 2 Duo processor, Memory: 1.99 GB DDR2) on the basis of our own software available at `http://www.mathos.hr/~scitowsk/Algorithms.nb` and `http://www.mathos.hr/~scitowsk/Algorithms.m`

shows that in iterative process (17) this corresponds to the choice of the first column of the matrix $\mathbf{B}^{-1}$ as a direction vector $\mathbf{p}$.

If we dropped the point T_ν, i.e. if we started from the point T_μ, then according to *Lemma 2.1* we would again obtain the point T_ν, and it means that in the iterative process the parameter vector $\mathbf{a}$ will not be changed. The following theorem shows that in iterative process (17) this corresponds to the choice of the second column of the matrix $\mathbf{B}^{-1}$ as a direction vector $\mathbf{p}$.

Theorem 2.1. *Let $I = \{1, \ldots, m\}$, $m \geq 2$, be the set of indices and let*

(i) $\Lambda = \{T_i = (x_i, y_i, z_i) : i \in I\}$ be a set of points in space such that its projection $\mathcal{L}$ on the (x, y)-plane does not lie on any line passing through the origin;

(ii) T_μ and T_ν be the first and the second point obtained by Algorithm I*, respectively, and $\mathbf{B}^{-1} = \frac{1}{x_\mu y_\nu - x_\nu y_\mu}\begin{bmatrix} y_\nu & -y_\mu \\ -x_\nu & x_\mu \end{bmatrix} =: [\mathbf{d}_1, \mathbf{d}_2]$.*

Then, if $\mathbf{a} = (\alpha, \beta)^T$ is the solution of system (16) for which a global minimum of the functional F given by (5) is not attained, then

I. Decreasing of functional values (5) can be attained by applying iterative process (17) in direction $\mathbf{d}_1$ of the first column of the matrix $\mathbf{B}^{-1}$, i.e. the next-better approximation obtained in Algorithm I *can be written as*

$$\bar{\mathbf{a}} = \mathbf{a} + \vartheta_1^* \mathbf{d}_1, \quad \mathbf{d}_1 = \frac{1}{x_\mu y_\nu - x_\nu y_\mu}\begin{bmatrix} y_\nu \\ -x_\nu \end{bmatrix}, \quad \vartheta_1^* = \operatorname*{med}_{i \in I \setminus I_1}\left(\left|\mathbf{d}_1^T \boldsymbol{\xi}_i\right|, \frac{z_i - \mathbf{a}^T \boldsymbol{\xi}_i}{\mathbf{d}_1^T \boldsymbol{\xi}_i}\right), \tag{18}$$

where $\boldsymbol{\xi}_i = (x_i, y_i)^T$ and $I_1 = \{i \in I : \mathbf{d}_1^T \boldsymbol{\xi}_i = 0\}$.

II. By choosing the second column $\mathbf{d}_2 = \frac{1}{x_\mu y_\nu - x_\nu y_\mu}(-y_\mu, x_\mu)^T$ of the matrix $\mathbf{B}^{-1}$ as a direction vector in iterative process (17) decreasing of values of the minimizing functional (5) will not be achieved, i.e. the step length in this direction is $\vartheta_2^* = \operatorname*{med}_{i \in I \setminus I_2}\left(\left|\mathbf{d}_2^T \boldsymbol{\xi}_i\right|, \frac{z_i - \mathbf{a}^T \boldsymbol{\xi}_i}{\mathbf{d}_2^T \boldsymbol{\xi}_i}\right) = 0$, whereby $I_2 = \left\{i \in I : \mathbf{d}_2^T \boldsymbol{\xi}_i = 0\right\}$.

Proof. I. By using the aforementioned property of median (13), in direction $\mathbf{d}_1$ we have

$$\begin{aligned}
\alpha + \vartheta_1^* \frac{y_\nu}{x_\mu y_\nu - x_\nu y_\mu} &= \operatorname*{med}_{i \in I \setminus I_1}\left(\left|\mathbf{d}_1^T \boldsymbol{\xi}_i\right|, y_\nu \frac{z_i - \alpha x_i - \beta y_i}{y_\nu x_i - x_\nu y_i} + \alpha\right) \\
&= \operatorname*{med}_{i \in I \setminus I_1}\left(\left|\mathbf{d}_1^T \boldsymbol{\xi}_i\right|, \frac{y_\nu z_i - y_i(\alpha x_\nu + \beta y_\nu)}{y_\nu x_i - x_\nu y_i}\right) \\
&= \operatorname*{med}_{i \in I \setminus I_1}\left(\left|\mathbf{d}_1^T \boldsymbol{\xi}_i\right|, \frac{y_\nu z_i - y_i z_\nu}{y_\nu x_i - x_\nu y_i}\right),
\end{aligned}$$

that according to *Lemma 2.1* corresponds to the optimal value α'' of the parameter α if we start from the point T_ν. Similarly, it can be shown that $\beta + \vartheta_1^* \frac{-x_\nu}{x_\mu y_\nu - x_\nu y_\mu}$ corresponds to the optimal value β'' of the parameter β if we start from the point T_ν.

II. In direction $\mathbf{d}_2$ of the second column of the matrix $\mathbf{B}^{-1}$ with step length ϑ_2^* we obtain

$$\begin{aligned}
\alpha + \vartheta_2^* \tfrac{-y_\mu}{x_\mu y_\nu - x_\nu y_\mu} &= \operatorname*{med}_{i \in I \setminus I_2}\left(\left|\mathbf{d}_2^T \boldsymbol{\xi}_i\right|, -y_\mu \tfrac{z_i - \alpha x_i - \beta y_i}{-y_\mu x_i + x_\mu y_i} + \alpha\right) \\
&= \operatorname*{med}_{i \in I \setminus I_2}\left(\left|\mathbf{d}_2^T \boldsymbol{\xi}_i\right|, \tfrac{-y_\mu z_i + y_i(\alpha x_\mu + \beta y_\mu)}{-y_\mu x_i + x_\mu y_i}\right) \\
&= \operatorname*{med}_{i \in I \setminus I_2}\left(\left|\mathbf{d}_2^T \boldsymbol{\xi}_i\right|, \tfrac{-y_\mu z_i + y_i z_\mu}{-y_\mu x_i + x_\mu y_i}\right) = \alpha.
\end{aligned}$$

Similarly, we obtain $\beta + \vartheta_2^* \frac{x_\mu}{x_\mu y_\nu - x_\nu y_\mu} = \beta$. $\square$

Remark 2.3. Note that direction vector $\mathbf{d}_1$ and direction vector $\mathbf{d}_2$ are perpendicular to the radius vector $\boldsymbol{\xi}_\nu = (x_\nu, y_\nu)^T$ of the point T_ν and the radius vector $\boldsymbol{\xi}_\mu = (x_\mu, y_\mu)^T$ of the point T_μ, respectively, i.e.

$$\mathbf{d}_1^T \boldsymbol{\xi}_\nu = 0, \quad \mathbf{d}_2^T \boldsymbol{\xi}_\mu = 0.$$

Note also that it is not necessary to take into account both directions $\mathbf{d}_1$ and $(-\mathbf{d}_1)$ because according to (13) the following holds: $\vartheta_i^*(-\mathbf{d}_i) = -\vartheta_i^*(\mathbf{d}_i)$, $i = 1, 2$.

2.2 Searching for a Best LAD-solution of a System of Equations

On the basis of Algorithm I we construct a more general algorithm for searching for a best LAD-solution of system (3). Note that condition (i) from *Theorem 2.1* is equivalent to the condition that the matrix $\mathbf{X}$ from (3) is of full column rank.

First, we choose two linearly independent rows of the matrix $\mathbf{X}$, by means of which we define a square nonsingular matrix $\mathbf{B}$, calculate $\mathbf{B}^{-1} =: [\mathbf{d}_1, \mathbf{d}_2]$ and define the initial approximation $\mathbf{a}_0$ of the solution. If the step length ϑ_1^* in direction $\mathbf{d}_1$ and the step length ϑ_2^* in direction $\mathbf{d}_2$ vanish, we suppose that we have achieved a best LAD-solution. Otherwise we search for the next approximation.

Such stopping criterion is analogous to the stopping criterion from Algorithm I and it is often mentioned in literature (see e. g. [9, 10]). A more detailed geometrical analysis in *Section 2.3* and illustrative examples in *Section 4* will show that it is not always implied that the point of the global minimum is attained. For further analysis of the problem the term subdifferential of the functional F (see *Section 3*) is necessary.

Algorithm II.

Step 1: Input matrix $\mathbf{X} = [\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m]^T$ and vector $\mathbf{z}$ and among the rows of matrix $\mathbf{X}$ choose two linearly independent rows: $\boldsymbol{\xi}_\mu = (x_\mu, y_\mu)^T$, $\boldsymbol{\xi}_\nu = (x_\nu, y_\nu)^T$;
Define matrix $\mathbf{B} = \begin{bmatrix} x_\mu & y_\mu \\ x_\nu & y_\nu \end{bmatrix}$ and calculate $\mathbf{B}^{-1} =: [\mathbf{d}_1, \mathbf{d}_2]$ and $\mathbf{a} := \mathbf{B}^{-1} \begin{bmatrix} z_\mu \\ z_\nu \end{bmatrix}$;

Step 2: Define $I_1 = \{i \in I : \mathbf{d}_1^T \boldsymbol{\xi}_i = 0\}$ and calculate $\vartheta_1^* = \operatorname*{med}_{i \in I \setminus I_1} \left(\left| \mathbf{d}_1^T \boldsymbol{\xi}_i \right|, \frac{z_i - \mathbf{a}^T \boldsymbol{\xi}_i}{\mathbf{d}_1^T \boldsymbol{\xi}_i} \right)$.
If $\vartheta_1^* = 0$, go to Step 3;
Otherwise, determine $k \in I \setminus I_1$;
Set $\mathbf{d} := \mathbf{d}_1$, $\vartheta^* = \vartheta_1^*$ and go to Step 4;

Step 3: Define set $I_2 = \left\{ i \in I : \mathbf{d}_2^T \boldsymbol{\xi}_i = 0 \right\}$ and calculate $\vartheta_2^* = \operatorname*{med}_{i \in I \setminus I_2} \left(\left| \mathbf{d}_2^T \boldsymbol{\xi}_i \right|, \frac{z_i - \mathbf{a}^T \boldsymbol{\xi}_i}{\mathbf{d}_2^T \boldsymbol{\xi}_i} \right)$.
If $\vartheta_2^* = 0$, STOP;
Otherwise, determine $k \in I \setminus I_2$;
Set $\mathbf{d} := \mathbf{d}_2$, $\vartheta^* = \vartheta_2^*$ and go to Step 4;

Step 4: Calculate $\mathbf{a} = \mathbf{a} + \vartheta^* \mathbf{d}$;
Set $\mu = \nu$, $\nu = k$ and define matrix $\mathbf{B} = \begin{bmatrix} x_\mu & y_\mu \\ x_\nu & y_\nu \end{bmatrix}$;

Step 5: Calculate $\mathbf{B}^{-1} =: [\mathbf{d}_1, \mathbf{d}_2]$ by using the previously mentioned inverse matrix and go to *Step 2*.

Remark 2.4. Note:

1. According to *Lemma 1.1*, in some cases the median of the data can be attained in every point of some interval $[y_{\nu_0}, y_{\nu_0+1}]$. In order to make assertions and calculations consequent, in that case the left edge y_{ν_0} of the interval should be taken for the median of data.

2. The set $I \setminus I_1$ in *Step 2* is not empty because of at least $\mu \in I \setminus I_1$. If that is a unique element of that set, then $\vartheta_1^* = 0$. Similar also holds for the set $I \setminus I_2$ from *Step 3*.

3. Matrix $\mathbf{B}$ from *Step 4* is always nonsingular because of $\det \mathbf{B} = c\,\mathbf{d}_1^T \boldsymbol{\xi}_\nu$, where c is a determinant of the matrix $\mathbf{B}$ from the preceding step.

2.3 Geometrical Analysis of the Method

To every data point $(x_i, y_i, z_i) \in \mathbb{R}^3$, $i = 1, \ldots, m$ we correspond the line

$$p_i = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha x_i + \beta y_i = z_i\}.$$

Note that $\boldsymbol{\xi}_i = (x_i, y_i)^T$ is a normal vector to the line p_i.

Because of the fact that there always exists a best LAD-plane passing through at least two different data points, the global minimum of the functional F is attained on the set

$$\Omega = \{(\alpha, \beta) : (\alpha, \beta) \in p_i \cap p_j,\ i \neq j, \quad i, j = 1, \ldots, m\}.$$

Suppose we selected two data points T_μ and T_ν. Lines p_μ and p_ν (see Fig. 2) are assigned to these points. Let $(\alpha_0, \beta_0) = p_\mu \cap p_\nu$ and $\mathbf{a}_0 = (\alpha_0, \beta_0)^T$.

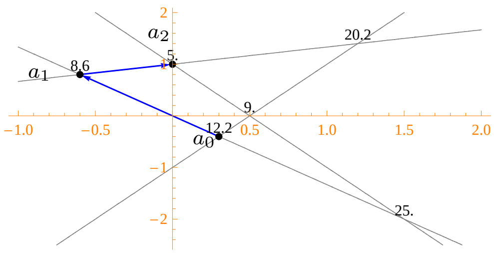

Figure 2: Geometry of the iterative process

In accordance with Algorithm II, the next approximation $\mathbf{a}_1 = (\alpha_1, \beta_1)^T$ is searched for such that from the point $\mathbf{a}_0$ we start either in direction $\mathbf{d}_\mu$ perpendicular to the vector $\boldsymbol{\xi}_\mu = (x_\mu, y_\mu)^T$ or in direction $\mathbf{d}_\nu$ perpendicular to the vector $\boldsymbol{\xi}_\nu = (x_\nu, y_\nu)^T$, i.e. either along the line p_μ or along the line p_ν. Thus, it holds $\mathbf{d}_\mu^T \boldsymbol{\xi}_\mu = 0$ and $\mathbf{d}_\nu^T \boldsymbol{\xi}_\nu = 0$.

The point $(\alpha_0, \beta_0) = p_\mu \cap p_\nu$, i.e. the vector $\mathbf{a}_0$, is obtained as a solution of system (16), and vectors $\mathbf{d}_\mu, \mathbf{d}_\nu$ should be perpendicular to the rows of the matrix $\mathbf{B}$. Columns of the matrix $\mathbf{B}^{-1}$ have such property, so that they can be used as possible direction vectors. The corresponding step length ϑ^* is determined by solving a weighted median problem. In such way we find the point of intersection on this line in which the functional F attains the minimal value. A significant difference of the proposed method is reflected in this in relation to methods given in e.g. [2, 3, 14, 6, 8]. Namely, in those papers and books it is always the first nearest point that is searched for in which decreasing of the value of functional F is attained, whereby sometimes direction $\mathbf{d}$ and $(-\mathbf{d})$ with a positive step length is especially considered.

Fig. 2 shows the flow of the iterative process for searching for a best LAD-solution of the system $\mathbf{Xa} = \mathbf{z}$, where

$$\mathbf{X}^T = \begin{bmatrix} 8 & 2 & 4 & -2 \\ 4 & -1 & 3 & 6 \end{bmatrix}, \qquad \mathbf{z} = (4, 1, 0, 6)^T.$$

The sequence of approximations $\mathbf{a}_0, \mathbf{a}_1, \dots$ and a direction of movement are denoted by black dots and blue arrows, respectively. In addition, beside every intersection point the value of the minimizing functional F in this intersection point is given.

3 LAD-solution of an Overdetermined System of Linear Equations

The problem of determining a best LAD-solution of an overdetermined system of linear equations $\mathbf{Xa} = \mathbf{z}$, $\mathbf{X} \in \mathbb{R}^{m\times n}$, $\mathbf{z} \in \mathbb{R}^m$, $m \ge n$, i.e. the problem of determining a best LAD-hyperplane $\mathbf{x} \mapsto \mathbf{a}^T\mathbf{x}$, on the basis of experimental data $(\mathbf{x}_i, z_i)$, $\mathbf{x}_i = (x_1^{(i)}, \dots, x_n^{(i)})^T \in \mathbb{R}^n$, $z_i \in \mathbb{R}$, $i = 1, \dots, m$ is reduced to minimization of the convex functional $F : \mathbb{R}^n \to \mathbb{R}$,

$$F(\mathbf{a}) = \sum_{i=1}^{m} |r_i(\mathbf{a})|, \quad r_i(\mathbf{a}) = z_i - \mathbf{a}^T\mathbf{x}_i, \tag{19}$$

which always attains its global minimum on $\mathbb{R}^n$. A subdifferential of the functional F (see e. g. [27, 28, 4, 29]) is of special importance in the analysis of this problem.

Lemma 3.1. *For some $\mathbf{a} \in \mathbb{R}^n$ a subdifferential of the functional F given by (19) is*

$$\partial F(\mathbf{a}) = \sum_{i\in I_0} [-1, 1]\mathbf{x}_i - \sum_{i\in I\setminus I_0} \sigma_i(\mathbf{a})\mathbf{x}_i, \quad \sigma_i(\mathbf{a}) = \operatorname{sign}(r_i(\mathbf{a})), \tag{20}$$

where $[-1, 1] = \{\lambda \in \mathbb{R} : -1 \le \lambda \le 1\}$ *and* $I_0 = \{i \in I : r_i(\mathbf{a}) = 0\}$.

Proof. A subdifferential of the function $\mathbf{a} \mapsto |r_i(\mathbf{a})|$ is given by

$$\partial(|r_i(\mathbf{a})|) = \begin{cases} -\mathbf{x}_i, & r_i(\mathbf{a}) > 0 \\ \mathbf{x}_i, & r_i(\mathbf{a}) < 0 \\ [-1, 1]\mathbf{x}_i, & r_i(\mathbf{a}) = 0 \end{cases} = \begin{cases} -\operatorname{sign}(r_i(\mathbf{a}))\mathbf{x}_i, & r_i(\mathbf{a}) \ne 0 \\ [-1, 1]\mathbf{x}_i, & r_i(\mathbf{a}) = 0 \end{cases}$$

from which there follows (20). ☐

Definition 3.1. Let the global minimum $\hat{\mathbf{a}} \in \mathbb{R}^n$ of the functional F given by (19) be searched by the iterative procedure of the form

$$\bar{\mathbf{a}} = \mathbf{a} + \vartheta\mathbf{p}, \qquad \mathbf{p} \in \mathbb{R}^n, \quad \vartheta \in \mathbb{R}.$$

Optimal step length ϑ^* in direction $\mathbf{p}$ implies

$$\vartheta^* = \operatorname*{argmin}_{\vartheta\in\mathbb{R}} \varphi(\vartheta), \qquad \varphi(\vartheta) = F(\mathbf{a} + \vartheta\mathbf{p}) - F(\mathbf{a}).$$

Theorem 3.1. *Let the global minimum of the functional F given by (19) be searched for by the iterative procedure of the form $\bar{\mathbf{a}} = \mathbf{a} + \vartheta\mathbf{p}$. Then the optimal step length ϑ^* in direction $\mathbf{p}$ is given by*

$$\vartheta^* = \operatorname*{med}_{i\in I\setminus I_1} \left(\left|\mathbf{p}^T\mathbf{x}_i\right|, \frac{z_i - \mathbf{a}^T\mathbf{x}_i}{\mathbf{p}^T\mathbf{x}_i} \right), \qquad I_1 = \left\{ i \in I : \mathbf{p}^T\mathbf{x}_i = 0 \right\}. \tag{21}$$

Proof. There holds

$$\begin{aligned}\varphi(\vartheta) = F(\mathbf{a}+\vartheta\mathbf{p}) - F(\mathbf{a}) &= \sum_{i\in I\setminus I_1} \left|z_i - \mathbf{a}^T\mathbf{x}_i - \mathbf{p}^T\mathbf{x}_i\vartheta\right| - \sum_{i\in I\setminus I_1} \left|z_i - \mathbf{a}^T\mathbf{x}_i\right| \\ &\geq \sum_{i\in I\setminus I_1} \left|\mathbf{p}^T\mathbf{x}_i\right| \left|\tfrac{z_i-\mathbf{a}^T\mathbf{x}_i}{\mathbf{p}^T\mathbf{x}_i} - \vartheta^*\right| - \sum_{i\in I\setminus I_1} \left|z_i - \mathbf{a}^T\mathbf{x}_i\right|,\end{aligned}$$

whereby the equality holds if and only if ϑ^* is given by (21). □

Lemma 3.2. *Let for some* $\mathbf{a}\in\mathbb{R}^n$ *and* $\mathbf{p}\in\mathbb{R}^n$

$$\begin{aligned}I_0 &:= \{i\in I : r_i(\mathbf{a}) = 0\},\\ J &:= \{i \in I\setminus I_0 : \mathbf{p}^T\mathbf{x}_i \neq 0\}.\end{aligned}$$

Then there exists $\varepsilon > 0$ *such that for every* $\vartheta\in(-\varepsilon,\varepsilon)$ *the following holds*

$$F(\mathbf{a}+\vartheta^*\mathbf{p}) - F(\mathbf{a}) \leq -\vartheta\mathbf{p}^T\mathbf{h}(\mathbf{a}) + |\vartheta| \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i|, \tag{22}$$

where $\mathbf{h}(\mathbf{a}) := \sum_{i\in I\setminus I_0} \sigma_i(\mathbf{a})\mathbf{x}_i \in \partial F(\mathbf{a})$ *and*

$$\vartheta^* = \operatorname*{med}_{i\in I_0\cup J} (w_i,\rho_i), \qquad w_i = |\mathbf{p}^T\mathbf{x}_i|,\ i\in I_0\cup J, \quad \rho_i = \begin{cases} 0, & i\in I_0 \\ \frac{r_i(\mathbf{a})}{\mathbf{p}^T\mathbf{x}_i}, & i \in J. \end{cases} \tag{23}$$

Proof. Because of $r_i(\mathbf{a}) = 0$, $\forall i\in I_0$, according to *Theorem 3.1*, for every $\vartheta\in\mathbb{R}$ the following holds

$$\begin{aligned}F(\mathbf{a}+\vartheta^*\mathbf{p}) - F(\mathbf{a}) &= \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i||0-\vartheta^*| + \sum_{i\in J} |\mathbf{p}^T\mathbf{x}_i| \left|\tfrac{r_i(\mathbf{a})}{\mathbf{p}^T\mathbf{x}_i} - \vartheta^*\right| - \sum_{i\in J} |r_i(\mathbf{a})| \\ &\leq |\vartheta| \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i| + \sum_{i\in J} |\mathbf{p}^T\mathbf{x}_i| \left|\tfrac{r_i(\mathbf{a})}{\mathbf{p}^T\mathbf{x}_i} - \vartheta\right| - \sum_{i\in J} |r_i(\mathbf{a})| \\ &= |\vartheta| \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i| + \sum_{i\in J} (r_i(\mathbf{a}) - \vartheta\mathbf{p}^T\mathbf{x}_i)\operatorname{sign}(r_i(\mathbf{a}) - \vartheta\mathbf{p}^T\mathbf{x}_i) - \sum_{i\in J} |r_i(\mathbf{a})|.\end{aligned}$$

Note that $\sum_{i\in J} \mathbf{p}^T\mathbf{x}_i = \sum_{i\in I\setminus I_0} \mathbf{p}^T\mathbf{x}_i$ and that there always exists $\varepsilon > 0$ such that for every $\vartheta\in(-\varepsilon,\varepsilon)$ there holds

$$\operatorname{sign}(r_i(\mathbf{a}) - \vartheta\mathbf{p}^T\mathbf{x}_i) = \operatorname{sign}(r_i(\mathbf{a})) = \sigma_i(\mathbf{a}), \quad \forall i\in J. \tag{24}$$

Therefore

$$\begin{aligned}F(\mathbf{a}+\vartheta^*\mathbf{p}) - F(\mathbf{a}) &\leq |\vartheta| \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i| + \sum_{i\in J} r_i(\mathbf{a})\sigma_i(\mathbf{a}) - \vartheta \sum_{i\in I\setminus I_0} (\mathbf{p}^T\mathbf{x}_i)\sigma_i(\mathbf{a}) - \sum_{i\in J} |r_i(\mathbf{a})| \\ &= |\vartheta| \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i| + \sum_{i\in J} |r_i(\mathbf{a})| - \vartheta \sum_{i\in I\setminus I_0} (\mathbf{p}^T\mathbf{x}_i)\sigma_i(\mathbf{a}) - \sum_{i\in J} |r_i(\mathbf{a})| \\ &= -\vartheta\mathbf{p}^T\mathbf{h}(\mathbf{a}) + |\vartheta| \sum_{i\in I_0} |\mathbf{p}^T\mathbf{x}_i|.\end{aligned}$$

□

Theorem 3.2. *Let* $\mathbf{X}\in\mathbb{R}^{m\times n}$, $m\geq n$, *be the matrix of full column rank,* $\mathbf{z}\in\mathbb{R}^m$, $\mathbf{B} = [\mathbf{x}_{i_1},\ldots,\mathbf{x}_{i_n}]^T \in \mathbb{R}^{n\times n}$, *a nonsingular square submatrix of the matrix* $\mathbf{X}$, $\mathbf{z}_B = (z_{i_1},\ldots,z_{i_n})^T$, $I_B = \{i_1,\ldots,i_n\}$, *and let*

(i) $\hat{\mathbf{a}}\in\mathbb{R}^n$ *be the solution of the system* $\mathbf{B}\mathbf{a} = \mathbf{z}_B$, *and* $\mathbf{B}^{-1} = [\mathbf{d}_1,\ldots,\mathbf{d}_n]$,

(ii) $I_0 = \{i\in I : r_i(\hat{\mathbf{a}}) = 0\}$,

(iii) $h(\hat{\mathbf{a}}) = \sum_{i\in I\setminus I_0} \sigma_i(\hat{\mathbf{a}})\mathbf{x}_i \in \partial F(\hat{\mathbf{a}})$.

If there exists $j_0 \in I_B$ *such that*

$$|\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}})| > 1 + \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|, \tag{25}$$

then

$$\vartheta^* = \operatorname*{med}_{i\in I_0\cup J}(w_i, \rho_i) \neq 0, \qquad \textit{where} \quad w_i = |\mathbf{d}_{j_0}^T \mathbf{x}_i|, \quad \rho_i = \begin{cases} 0, & i \in I_0 \\ \frac{r_i(\hat{\mathbf{a}})}{\mathbf{d}_{j_0}^T \mathbf{x}_i}, & i \in J \end{cases}, \tag{26}$$

where $J = \{i \in I \setminus I_B : \mathbf{d}_{j_0}^T \mathbf{x}_i \neq 0\}$*, and the following holds*

$$F(\hat{\mathbf{a}} + \vartheta^* \mathbf{d}_{j_0}) < F(\hat{\mathbf{a}}). \tag{27}$$

Proof. Let us first show that $\vartheta^* \neq 0$. Suppose contrary, i.e. that $\vartheta^* = 0$. For that purpose define an auxiliary function $\psi : \mathbb{R} \to \mathbb{R}$,

$$\psi(\vartheta) = \sum_{i\in I_0\cup I_1} w_i|\rho_i - \vartheta|,$$

which attains its global minimum for $\vartheta^* = \operatorname*{med}_{i\in I_0\cup J}(w_i, \rho_i)$. Thereby $\vartheta^* = 0$ if and only if $0 \in \partial\psi(0)$, where $\partial\psi(0)$ is a subdifferential of the function ψ in the point 0

$$\partial\psi(0) = \sum_{i\in I_0} [-1, 1]\mathbf{d}_{j_0}^T \mathbf{x}_i - \sum_{i\in J} \sigma_i(\hat{\mathbf{a}})\mathbf{d}_{j_0}^T \mathbf{x}_i.$$

Since generally $I_B \subseteq I_0$ and $\mathbf{d}_{j_0}^T \mathbf{x}_i = \delta_{ij_0}$, $\forall i \in I_B$, the condition mentioned will be fulfilled if and only if $\exists\, \lambda_0, \gamma_i \in [-1, 1]$, such that

$$\lambda_0 + \sum_{i\in I_0\setminus I_B} \gamma_i \mathbf{d}_{j_0}^T \mathbf{x}_i - \sum_{i\in J} \sigma_i(\hat{\mathbf{a}})\mathbf{d}_{j_0}^T \mathbf{x}_i = 0,$$

i.e. since $\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}) = \sum_{i\in J} \sigma_i(\hat{\mathbf{a}})\mathbf{d}_{j_0}^T \mathbf{x}_i$, if and only if

$$\lambda_0 + \sum_{i\in I_0\setminus I_B} \gamma_i \mathbf{d}_{j_0}^T \mathbf{x}_i = \mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}). \tag{28}$$

Note that

$$-1 - \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i| \leq \lambda_0 + \sum_{i\in I_0\setminus I_B} \gamma_i \mathbf{d}_{j_0}^T \mathbf{x}_i \leq 1 + \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|.$$

According to (28), this means that

$$|\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}})| \leq 1 + \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|,$$

which contradicts assumption (25).

For proving (27) let us first notice that

$$\sum_{i\in I_0} |\mathbf{d}_{j_0}^T \mathbf{x}_i| = |\mathbf{d}_{j_0}^T \mathbf{x}_{j_0}| + \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i| = 1 + \sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|.$$

Due to (25), according to *Lemma 3.2*, there exists $\varepsilon > 0$ such that for every $\vartheta \in (-\varepsilon, \varepsilon)$ the following holds

$$\begin{aligned} F(\hat{\mathbf{a}} + \vartheta^* \mathbf{d}_{j_0}) - F(\hat{\mathbf{a}}) \leq -\vartheta \mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}) + |\vartheta| \textstyle\sum_{i\in I_0} |\mathbf{d}_{j_0}^T \mathbf{x}_i| \quad &\leq \quad -\vartheta \mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}) + |\vartheta| + |\vartheta| \textstyle\sum_{i\in I_0\setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i| \\ &< \quad -\vartheta \mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}) + |\vartheta| |\mathbf{h}^T \mathbf{d}_{j_0}| =: A, \end{aligned} \tag{29}$$

where ϑ^* is given by (26). If $\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}) > 0$, then $A = 0$ for $\vartheta \in (0, \varepsilon)$, and if $\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}}) < 0$, then $A = 0$ for $\vartheta \in (-\varepsilon, 0)$.

Hence, decreasing of the value of the minimizing functional F is attained in the direction $\mathbf{d}_{j_0}$ with the step length $\vartheta^* \neq 0$. □

Theorem 3.3. *By the assumption as in* Theorem 3.2*, let* $I_0 = I_B$*. Then,*

I. Functional F attains its global minimum for $\hat{\mathbf{a}} = \mathbf{B}^{-1}\mathbf{z}_B$ if and only if $|\mathbf{d}_j^T \mathbf{h}(\hat{\mathbf{a}})| \leq 1 \ \forall j \in I_B$.

II. If there exists $j_0 \in I_B$ such that $|\mathbf{d}_{j_0}^T \mathbf{h}(\hat{\mathbf{a}})| > 1$, then

$$F(\hat{\mathbf{a}} + \vartheta^* \mathbf{d}_{j_0}) < F(\hat{\mathbf{a}}),$$

where ϑ^* is given by (26).

Proof. I. According to *Lemma 3.1*, the subdifferential of the functional F in the point $\mathbf{a} \in \mathbb{R}^n$ can be written as

$$\partial F(\mathbf{a}) = \sum_{i\in I_B} [-1, 1]\mathbf{x}_i - \sum_{i\in I\setminus I_B} \sigma_i(\mathbf{a})\mathbf{x}_i. \tag{30}$$

The point $\hat{\mathbf{a}} \in \mathbb{R}^n$ is the point of the global minimum of the functional F if and only if $\mathbf{0} \in \partial F(\hat{\mathbf{a}})$ (see e.g. [4]). Since the matrix $\mathbf{B}$ is nonsingular, then from (30) it follows that $\hat{\mathbf{a}} = \mathbf{B}^{-1}\mathbf{z}_B$ is the point of the global minimum of the functional F if and only if there exists $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\|\boldsymbol{\lambda}\|_\infty \leq 1$, such that

$$\boldsymbol{\lambda} = \sum_{i\in I\setminus I_B} \sigma_i(\hat{\mathbf{a}}) \left(\mathbf{B}^{-1}\right)^T \mathbf{x}_i, \tag{31}$$

i.e. if and only if

$$|\lambda_j| = \left| \sum_{i\in I\setminus I_B} \sigma_i(\hat{\mathbf{a}}) \mathbf{d}_j^T \mathbf{x}_i \right| = |\mathbf{d}_j^T \mathbf{h}(\hat{\mathbf{a}})| \leq 1, \quad \forall j \in I_B. \tag{32}$$

II. The proof of this assertion follows directly as a special case of *Theorem 3.2*. □

3.1 Algorithm for Searching for a Best LAD-solution of an Overdetermined System of Linear Equations

Analogously to Algorithm II we construct an algorithm for searching for a best LAD-solution of an overdetermined system of linear equations $\mathbf{X}\mathbf{a} = \mathbf{z}$, where $\mathbf{X} \in \mathbb{R}^{m\times n}$, $m \geq n$, is the matrix of full column rank, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{z} \in \mathbb{R}^m$, by minimizing the functional

$$F(\mathbf{a}) = \sum_{i\in I} |z_i - \mathbf{a}^T \mathbf{x}_i|, \quad I = \{1, \ldots, m\},$$

where $\mathbf{x}_i^T$ is the i-th row of the matrix $\mathbf{X}$. The algorithm is constructed as an iterative process of the form

$$\bar{\mathbf{a}} = \mathbf{a} + \vartheta\, \mathbf{p},$$

where $\mathbf{p} \in \mathbb{R}^n$ is the direction vector, and ϑ is the step length in this direction, which are determined in accordance with *Theorem 3.2*, i.e. *Theorem 3.3*.

Remark 3.1. The approximation $\hat{\mathbf{a}}$ of the solution is considered to be optimal if and only if $\mathbf{0} \in \partial F(\hat{\mathbf{a}})$ (see e. g. [27, 28, 4, 29]). With notations $I_0 = \{i \in I : z_i - \hat{\mathbf{a}}\mathbf{x}_i = 0\} = \{i_i, \ldots, i_l\}$, $\mathbf{X}_0 = [\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_l}]$, $\mathbf{h}(\hat{\mathbf{a}}) = \sum\limits_{i \in I \setminus I_0} \sigma_i(\hat{\mathbf{a}})\mathbf{x}_i$, the conditions mentioned in *Theorem 3.2* and *Theorem 3.3* will be fulfilled if and only if the system

$$\mathbf{X}_0 \boldsymbol{\lambda} = \mathbf{h}(\hat{\mathbf{a}}) \quad \text{with condition} \quad \|\boldsymbol{\lambda}\|_\infty \leq 1, \tag{33}$$

has a solution. If specially $I_0 = I_B$, then the system from (33) becomes

$$\mathbf{B}^T \boldsymbol{\lambda} = \mathbf{h}(\hat{\mathbf{a}}),$$

whose solution is

$$\lambda_j = \mathbf{d}_j^T \mathbf{h}(\hat{\mathbf{a}}), \qquad j \in I_B,$$

so that (33) will have a solution if and only if $|\mathbf{d}_j^T \mathbf{h}(\hat{\mathbf{a}})| < 1, \forall j \in I_B$, which is in accordance with Theorem 3.3. For the example from *Section 2.3* we obtain $\hat{\mathbf{a}} = (0, 1)^T$, $I_0 = I_B = \{1, 4\}$, and $\boldsymbol{\lambda} = (-\frac{5}{14}, -\frac{3}{7})^T$.

Algorithm III.

Step 0 $I = \{1, \ldots, m\}$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]^T$, $\mathbf{x}_i \in \mathbb{R}^n$, $\mathbf{z} = (z_1, \ldots, z_m)^T \in \mathbb{R}^m$;

Step 1: Choose n linearly independent rows in $\mathbf{X}$ with ordinal numbers $I_B = \{i_1, \ldots, i_n\}$ and set:
$\mathbf{B} = [\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_n}]^T, \quad \mathbf{z}_B = (z_{i_1}, \ldots, z_{i_n})^T,$

$\mathbf{B}^{-1} = [\mathbf{d}_1, \ldots, \mathbf{d}_n],$

$\mathbf{a} = \mathbf{B}^{-1}\mathbf{z}_B, \quad I_0 = \{i \in I : z_i - \mathbf{a}^T\mathbf{x}_i = 0\},$

$\mathbf{h} = \sum\limits_{i \in I \setminus I_0} \operatorname{sign}(z_i - \mathbf{a}^T\mathbf{x}_i)\mathbf{x}_i;$

Step 2: If there exists $j_0 \in I_B$ such that $|\mathbf{d}_{j_0}^T \mathbf{h}| > 1 + \sum_{i \in I_0 \setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|$, go to *Step 3*;
Else go to *Step 5*.

Step 3: Define $J = \{i \in I \setminus I_0 : \mathbf{d}_{j_0}^T \mathbf{x}_i \neq 0\}$ and

$\forall i \in I_0 \cup J$ define $w_i = |\mathbf{d}_{j_0}^T \mathbf{x}_i|, \quad \rho_i = \begin{cases} 0, & i \in I_0 \\ \frac{z_i - \mathbf{a}^T\mathbf{x}_i}{\mathbf{d}_{j_0}^T \mathbf{x}_i}, & i \in J \end{cases},$

and determine $\nu \in I_0 \cup J$ on which $\vartheta^* = \operatorname*{med}\limits_{i \in I_0 \cup J} (w_i, \rho_i)$ is attained;

Set $\mathbf{a} = \mathbf{a} + \vartheta^* \mathbf{d}_{j_0}$.

Step 4: Define a new matrix B, which is made from the old one by replacing the j_0-th row by the ν-th row;

Set $I_B = \mathit{ReplacePart}[I_B, j_0 \to \nu], \quad I_0 = \{i \in I : z_i - \mathbf{a}^T\mathbf{x}_i = 0\},$

determine a new matrix $\mathbf{B}^{-1} = [\mathbf{d}_1, \ldots, \mathbf{d}_n]$, calculate

$\mathbf{h} = \sum\limits_{i \in I \setminus I_0} \operatorname{sign}(z_i - \mathbf{a}^T\mathbf{x}_i)\mathbf{x}_i$ and go to *Step 2*.

Step 5: Define $\mathbf{X}_0 = [\mathbf{x}_{j_0}, \ldots, \mathbf{x}_{j_l}]^T$, where $\{j_0, \ldots, j_l\} = I_0$.

If the system $\mathbf{X}_0\boldsymbol{\lambda} = \mathbf{h}$ subject to $\|\boldsymbol{\lambda}\|_\infty \leq 1$ has a solution, STOP;

Else set $Ind = \{\}$ and go to *Step 6*.

Step 6: Choose $j_0 \in (I_0 \setminus Ind) \setminus I_B$, set $Ind = Ind \cup \{j_0\}$ and

define $\bar{\mathbf{B}} = [\mathbf{x}_{j_0}, \mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_n}]^T$;

Do QR factorization with pivoting $\bar{\mathbf{B}} = QR$ such that the position of the first row is not changed;

Define a new set of indices $I_B = \{i_1, \ldots, i_n\}$ (i_1 is a new row).

If there exists $j_0 \in I_B$ such that $|\mathbf{d}_{j_0}^T \mathbf{h}| > 1 + \sum_{i \in I_0 \setminus I_B} |\mathbf{d}_{j_0}^T \mathbf{x}_i|$, go to *Step 3*;

Else repeat *Step 6*.

Remark 3.2. The initial approximation in the algorithm is obtained by choosing a nonsingular submatrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ of the matrix $\mathbf{X}$, which can be determined by applying QR factorization with column pivoting (see e.g. [2, 30]), although other approaches can be found in literature as well (see e. g. [6]).

Since in every step of the algorithm the matrix $\mathbf{B}$ changes in only one row, calculation of the inverse matrix $\mathbf{B}^{-1}$ may be simplified significantly also by applying QR factorization (see e. g. [2, 31, 32] (that is also used in our algorithm) or by applying the Sherman-Morrison formula (see e. g. [14, 6].

As mentioned previously in *Remark 2.4*, the set J from *Step 3* is not empty since at least $j_0 \in J$. If j_0 is a unique element of the set J, then $\vartheta^* = 0$. Also, matrix $\mathbf{B}$ from *Step 4* always remains nonsingular.

Checking whether the system $\mathbf{X}_0 \boldsymbol{\lambda} = \mathbf{h}$ with condition $\|\boldsymbol{\lambda}\|_\infty \leq 1$ has a solution is carried out by a *Mathematica*-instruction `FindInstance`, which is based upon the Buchberger's algorithm and the Gröbner system (see e. g. [33]).

4 Illustrative Examples and Numerical Experiments

The aforementioned algorithms will first be illustrated on a 10×2 example in which different nondegenerate and degenerate situations appear, and which could be visually well observed.

Example 4.1. *Given is the system* $\mathbf{Xa} = \mathbf{z}$, $\mathbf{X} \in \mathbb{R}^{10 \times 2}$, $\mathbf{z} \in \mathbb{R}^{10}$, *where*

$$\mathbf{X}^T = \begin{bmatrix} -2 & -1 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ -1 & 0 & -3 & -1 & -2 & -1 & 1 & -3 & -2 & 0 \end{bmatrix}, \qquad \mathbf{z} = (-5, -2, -9, -1, -2, 0, 4, -1, 2, 2)^T.$$

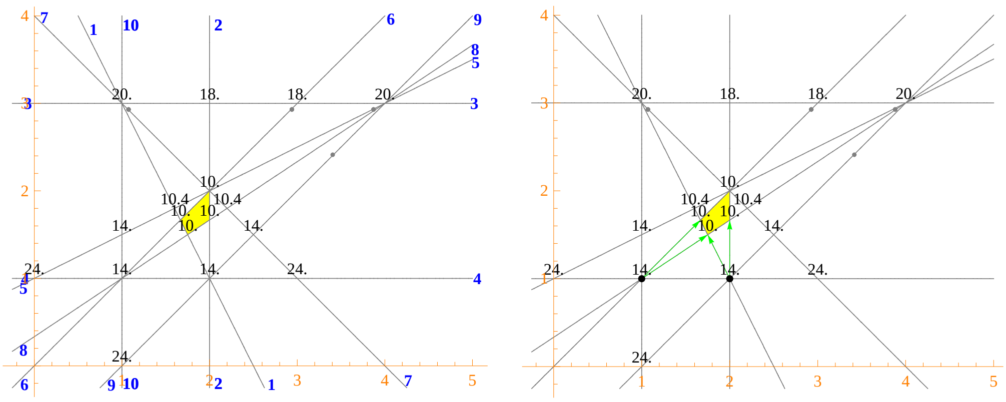

Figure 3: Iterative process

In accordance with *Section 2.3* in *Fig. 3* left, the system is shown by the lines marked with numbers $i = 1, \ldots, 10$. In the same figure, beside each intersection of the lines, the value of the minimizing functional F is denoted, whereby a yellow polygon denotes the area on which the functional F attains its global minimum.

No	Initial equations	Initial appr.	Iteration 1	Iteration 2	$\boldsymbol{\lambda}^*(I_0)$
Fig. 3 right	$\{4,10\}$	$\{(1,1),14\}$	$\{(1,1),14\}$	$\{(\frac{5}{3},\frac{5}{3}),10\}$	$\{-1,1\}$
	$\{4,6\}$	$\{(1,1),14\}$	$\{(\frac{5}{3},\frac{5}{3}),10\}$	—	$\{-1,1\}$
	$\{4,8\}$	$\{(1,1),14\}$	$\{(\frac{7}{4},\frac{3}{2}),10\}$	—	$\{-1,-1\}$
Fig. 4 left	$\{3,9\}$	$\{(4,3),20\}$	$\{(\frac{5}{2},\frac{3}{2}),14\}$	$\{(2,2),10\}$	$\{1,1,1,-1\}$
	$\{3,5\}$	$\{(4,3),20\}$	$\{(2,2),10\}$	—	$\{1,1,1,-1\}$
	$\{3,8\}$	$\{(4,3),20\}$	$\{(2,\frac{5}{3}),10\}$	—	$\{1,-1\}$
Fig. 4 right	$\{3,10\}$	$\{(1,3),20\}$	$\{(1,\frac{3}{2}),14\}$	$\{(2,2),10\}$	$\{1,1,1,-1\}$
	$\{3,7\}$	$\{(1,3),20\}$	$\{(2,2),10\}$	—	$\{1,1,1,-1\}$
	$\{3,1\}$	$\{(1,3),20\}$	$\{(\frac{5}{3},\frac{5}{3}),10\}$	—	$\{-1,1\}$

Table 1: Iterative process

If we choose the intersection of the lines $\{4,10\}$, i.e. the point $(1,1)$, as the initial approximation, then Algorithm I and Algorithm II cannot be run since along these lines there does not exist a smaller value of the functional F. Algorithm III in *Step 5* detects that it is not the point of the global minimum and in this point it selects a new direction along line 6, which after that leads to a solution in one single step (a green arrow from the point $(1,1)$ to the point $(\frac{5}{3},\frac{5}{3})$). We have a similar situation if we choose the intersection of the lines $\{4,9\}$, i.e. the point $(2,1)$, as the initial approximation. Algorithm III in *Step 5* detects that it is not the point of the global minimum and in this point it selects a new direction along line 1, which after that leads to a solution in one single step (a green arrow from the point $(2,1)$ to the point $(\frac{7}{4},\frac{3}{2})$). These situations are shown in *Fig. 3* right. Thereby green arrows also show the direction of optimal strategy of movement from the point $(1,1)$, i.e. from the point $(2,1)$. Corresponding data can be seen in *Table 1.* The column denoted by $\boldsymbol{\lambda}^*(I_0)$ shows values of parameters λ_i from (33) in the optimal point.

If we choose the intersection of the lines $\{3,9\}$, i.e. the point $(4,3)$, as the initial approximation, then as a solution all algorithms give the point $(2,2)$, in which the lines $\{2,5,6,7\}$ intersect. This situation is shown in *Fig. 4* left by blue arrows, and the corresponding flow of the algorithm is also given in *Table 1.* Green arrows show directions of optimal strategy of movement from the point $(4,3)$.

If we choose the intersection of the lines $\{3,10\}$, i.e. the point $(1,3)$, as the initial approximation, then as a solution all algorithms give again the point $(2,2)$, in which the lines $\{2,5,6,7\}$ intersect. The flow of Algorithm III is shown in *Fig. 4* right by blue arrows, and it is also shown in *Table 1.* Green arrows show directions of optimal strategy of movement from the point $(1,3)$.

Example 4.2. *Similarly to [25], for the function* $f:[1,2]\to\mathbb{R}$, $f(x)=e^x+\begin{cases}5, & x\in(1.2,1.4)\\ 0, & x\notin(1.2,1.4)\end{cases}$ *we will search for a best LAD-polynomial of the* $(n-1)$*-th degree* $P_{n-1}(x)=\alpha_1+\alpha_2x+\cdots+\alpha_nx^{n-1}$ *on the basis of given data* (x_i,z_i), $i=1,\dots,m$, *where*

$$x_i=1+\frac{i}{m},\qquad z_i=f(x_i)+\varepsilon_i,\quad \varepsilon_i\sim\mathcal{N}(0,\sigma^2),\quad \sigma=0.5.$$

For $n=4$ *and* $m=18$ *the problem is reduced to searching for a best LAD-solution of the system* $\mathbf{X}\mathbf{a}=\mathbf{z}$, *where* $\mathbf{X}_{ij}=x_i^{j-1}$, $\quad i=1,\dots,m,\quad j=1,\dots n.$

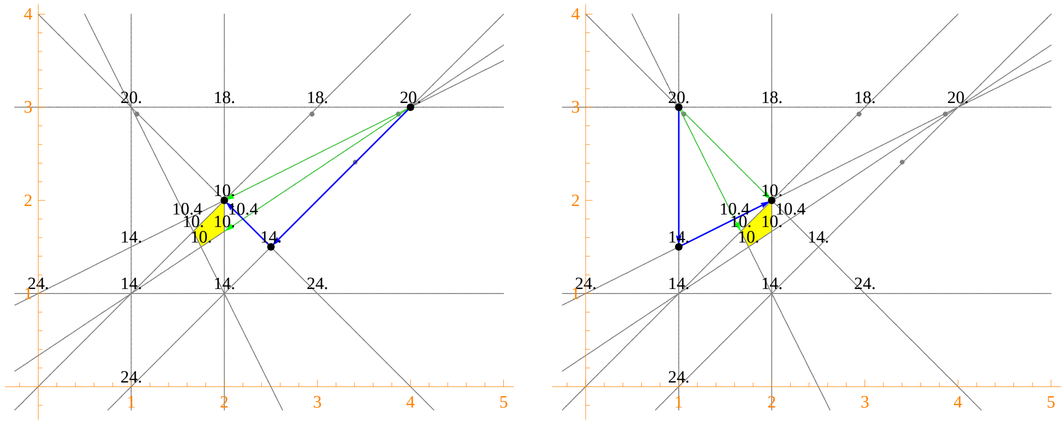

Figure 4: Iterative process

k	I_B	$\mathbf{a}_k^T$				$F(\mathbf{a}_k)$
0	$\{18, 1, 10, 5\}$	$(-405.091,$	$852.626,$	$-573.812,$	$125.323)$	35.6137
1	$\{1, 10, 5, 17\}$	$(-426.959,$	$904.515,$	$-614.345,$	$135.746)$	31.6462
2	$\{1, 5, 17, 13\}$	$(-294.730,$	$607.759,$	$-398.664,$	$85.3265)$	22.4969
3	$\{1, 17, 13, 2\}$	$(-90.766,$	$191.204,$	$-126.184,$	$27.625)$	21.5901
4	$\{1, 17, 2, 11\}$	$(-93.870,$	$198.533,$	$-131.779,$	$28.9859)$	21.4035
5	$\{1, 17, 11, 14\}$	$(-133.619,$	$281.305,$	$-187.208,$	$41.0065)$	21.1829

Table 2: Iterative process

The flow of the iterative process and the corresponding approximate polynomials are shown in *Table 2* and *Fig. 5*, respectively. Note that the graph of each approximate polynomial passes through 4 data points, which is in accordance with the described theory and Algorithm III.

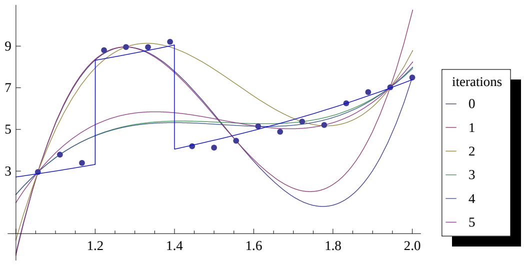

Figure 5: Polynomial LAD-approximation of the function

Example 4.3. Algorithm III *will also be tested on large systems. For that purpose we consider the problem given in* Example 4.2 *for* $n = 5, 10$ *and* $m = 50, 100, 200, 500, 1000$.

The number of iterations is shown in *Table 3*, whereby below every number of iterations a value of the minimizing functional obtained by Algorithm III and a value of the minimizing functional obtained by the *Mathematica*-module `NMinimize` are shown. In this way it can be seen that Algorithm III is dominant.

	$m = 50$	$m = 100$	$m = 200$	$m = 500$	$m = 1000$
$n = 5$	8	27	19	21	30
Algorithm III	48.9	107.9	212.9	536.4	1077.1
`NMinimize`	58.8	119.9	232.6	572.8	1151.1
$n = 10$	23	33	34	43	65
Algorithm III	24.7	48.1	104.6	261.8	523.6
`NMinimize`	90.6	122.5	236.6	563.6	1082.4

Table 3: Testing Algorithm III on large systems

5 Concluding Remarks

In this paper we consider the problem of searching for a best LAD-solution of an overdetermined system of linear equations $\mathbf{Xa} = \mathbf{z}$, where $\mathbf{X} \in \mathbb{R}^{m\times n}$, $m \geq n$, is a matrix of full column rank, $\mathbf{a} \in \mathbb{R}^n$, and $\mathbf{z} \in \mathbb{R}^m$ (see e. g. [2, 4, 5, 6, 7, 8]). Motivated by an efficient method for solving the problem of estimation of optimal parameters of a best LAD-plane $(x, y) \mapsto \alpha x + \beta y$ on the basis of the given set of experimental data (x_i, y_i, z_i), $i = 1, \dots, m$ (see e. g. [9, 10, 11, 12, 13]), which can easily be geometrically visualized, we define an iterative procedure for searching for a best LAD-solution of an overdetermined system of linear equations. For the mentioned iterative procedure we construct an appropriate algorithm and give a few illustrative examples for nondegenerate and degenerate situations. The examples with large systems show efficiency of the given method, which can be easily expanded on an overdetermined system with linear constraints. The methodology used in the paper could also be used for solving a more difficult and numerically more demanding orthogonal distance linear regression problem (see e. g. [13]).

References

[1] Barrodale, I., Roberts, F.D.K.: Algorithm 552: Solution of the constrained l_1 approximation problem [F4]. ACM Trans. Math. Software **6**, 231–235 (1980)

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

[3] Bartels, R.H., Conn, A.R.: Linearly constrained discrete l_1 problems. ACM Trans. Math. Software **6**, 594–608 (1980)

[4] Gonin, R., Money, A.H.: Nonlinear L_p-norm Estimation. CRC Press (1989)

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

[6] Shi, M., Lukas, M.A.: An L_1 estimation algorithm with degeneracy and linear constraints. Comput. Statist. Data Anal. **39**, 35–55 (2002)

[7] Watson, G.A.: Approximation Theory and Numerical Methods. John Wiley & Sons, Chichester (1980)

[8] Yan, S.: A base-point descent algorithm for solving the linear l_1 problem. Int. J. Comput. Math. **80**, 367–380 (2003)

[9] Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming. Theory and Algorithms. Wiley, New Jersey (2006)

[10] Cupec, R., Grbić, R., Sabo, K., Scitovski, R.: Three points method for searching the best least absolute deviations plane. Appl. Math. Comput. **215**, 983–994 (2009)

[11] Plastria, F., Carrizosa, E.: Gauge Distances and Median Hyperplanes. J. Optim. Theory Appl. **110**, 173–182 (2001)

[12] Schöbel, A.: Locating Lines and Hyperplanes: Theory and Algorithms. Springer Verlag, Berlin (1999)

[13] Schöbel, A.: Anhored hyperplane location problems. Discrete Comput. Geom. **29**, 229–238 (2003)

[14] Birkes, D., Dodge, Y.: Alternative Methods of Regression. Wiley, New York (1993)

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

[16] Dodge, Y.: An introduction to L_1-norm based statistical data analysis. Comput. Statist. Data Anal. **5**, 239–253 (1987)

[17] Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley, New York (2003)

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

[19] Vazler, I., Sabo, K., Scitovski, R.: Weighted median of the data in solving least absolute deviations problems. Comm. Statist. Theory Methods (to appear)

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

[21] Dasgupta, S., Mishra, S.K.: Least absolute deviation estimation of linear econometric models: A literature review. Munich Personal RePEc Archive, 1-26, (2004)

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

[23] Xia, Y., Kamel, M.S.: A generalized least absolute deviation method for parameter estimation of autoregressive signals. IEEE Trans. on Neural Networks **19**, 107–118 (2008)

[24] Bloomfield, P., Steiger, W.: Least Absolute Deviations: Theory, Applications, and Algorithms. Birkhauser, Boston (1983)

[25] Coleman, T.F., Li, Y.: A globally and quadratically convergent affine scaling method for linear l_1 problems. Math. Program. **56**, 189–222 (1992)

[26] Hoffman, K.L., Shier, R.R.: A test problem generator for discrete linear L_1 approximation problems. ACM Trans. Math. Software **6**, 587–593 (1980)

[27] Avriel, M.: Nonlinear Programming: Analysis and Methods. Dover Publications, Inc. Mineola, New York (2003)

[28] Демянов, В.Ф., Васиљев, Л.В.: Недифференцируемая Оптимизация. Наука, Моцква (1981)

[29] Ruszczynski, A.: Nonlinear Optimization. Princeton University Press, Princeton and Oxford (2006)

[30] Golub, G.H., Van Loan, C.F.: Matrix Computation. The John Hopkins University Press, London (1996)

[31] Daniel, J.W., Gragg, W.B., Kaufman, L., Stewart, G.W.: Reorthogonalization and stable algorithms for updating the Gram-Schmidt QR factorization. Math. Comp. **30**, 772–795 (1976)

[32] Escudero, L.F.: On QR factorization updatings. Trabajos de Estadistica y de Investigacion Operativa **35**, 17-31 (1984)

[33] Yokoyama, K., Noro, M., Takeshima, T.: Solutions of systems of algebraic equations and linear maps on residue class rings. J. Symbolic Comput. **14**, 399–417 (1992)