Three points method for searching the best least absolute deviations plane

Robert Cupec, Ratko Grbić, Kristian Sabo, Rudolf Scitovski
Three points method for searching the best least absolute deviations plane

Robert Cupec, Ratko Grbić, Kristian Sabo, Rudolf Scitovski

Abstract. In this paper a new method for estimation of optimal parameters of a best least absolute deviations plane is proposed, which is based on the fact that there always exists a best least absolute deviations plane passing through at least three different data points. The proposed method leads to a solution in finitely many steps. Moreover, a modification of the aforementioned method is proposed that is especially adjusted to the case of a large number of data and the need to estimate parameters in real time. Both methods are illustrated by numerical examples on the basis of simulated data and by one practical example from the field of robotics.

Keyword. Least absolute deviations; LAD; $l_1$-norm approximation; Weighted median problem

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

1 Introduction

In this paper we consider the problem of a fast determination of a best least absolute deviations plane (LAD-plane) on the basis of the given set of points in space $\Lambda = \{ T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I \}$, where $I = \{1, \ldots, m\}$, $m \geq 3$, is a set of indices. It is thereby assumed that the number of data $m$ can be great and that among the

---

1This work was supported by the Ministry of Science, Education and Sports, Republic of Croatia, through research grants 235-2352818-1034, 165-0361621-2000 and 036-0363078-3018

2Faculty of Electrical Engineering, University of Osijek, Kneza Trpimira bb, HR-31000 Osijek, Croatia, E-mail: rcupec@etfos.hr

3Faculty of Electrical Engineering, University of Osijek, Kneza Trpimira bb, HR-31000 Osijek, Croatia, E-mail: rgrbic@etfos.hr

4Department of Mathematics, University of Osijek, Trg Lj. Gaja 6, HR-31 000 Osijek, Croatia, E-mail: ksabo@mathos.hr

5Department of Mathematics, University of Osijek, Trg Lj. Gaja 6, HR-31 000 Osijek, Croatia, E-mail: scitowsk@mathos.hr
data a substantial amount of outliers (i.e. wild points) might appear. This problem could be naturally extended to determination of the best LAD hyperplane. Such problems are found very often in applied research (see e.g. [7], [9], [13], [15], [26], [29], [30]).

Searching for the best least absolute deviations hyperplane on the basis of a given set of points is well-known in literature (see [2], [8], [12], [15], [20], [23], [32]). This principle is considered to have been proposed by the Croatian mathematician J. R. Bošković in the mid-eighteenth century (see [2], [9]).

The problem of determining optimal parameters in a mathematical model appears in all areas of applied research. Let us mention only applications in statistics, finance, engineering, bioinformatics, operations research, etc. The problem of estimating optimal parameters can be considered in several ways. Most frequently it is assumed that errors are normally distributed and that they can occur only in measured values of the dependent variable. In that case, it is generally taken that parameters are estimated by means of the Ordinary Least Squares method (see [8], [14]). If outliers can appear among the data, then the application of the $l_1$-norm is much more acceptable (see [3], [5], [9], [15], [18], [21], [23], [29]). In literature this approach is known as the Least Absolute Deviations problem (LAD-problem).

If errors are assumed to occur in measured values of both (dependent and independent) variables, the problem in question is the errors-in-variables problem (see e.g. [25]). Especially, if thereby the square of the $l_2$ norm is used, it is the Total Least Squares problem, and if the $l_1$ or the $l_2$ norm is used, then the problems in question are the orthogonal $l_1$ and the orthogonal $l_2$ problem, respectively (see e.g. [27], [31]).

In this paper a new method for estimating optimal parameters of a best LAD-plane is proposed, which is based on the fact that there always exists a best LAD-plane passing through at least three different data points. The proposed method leads to the solution of this LAD-problem in finitely many steps. Moreover, a modification of the aforementioned method is proposed that is especially adjusted to the case of a large number of data and the need to estimate parameters in real
time.

First, some important facts necessary for consideration and analysis of the problem are given. After that, basic properties of a best LAD-plane are considered, on the basis of which it is possible to design and analyze the proposed methods. Methods are illustrated and tested on several numerical examples. Finally, an example of applying these methods in robotics is given.

## 2 A best LAD-plane

Let us now define the LAD-plane problem. Let $I = \{1, \ldots, m\}$, $m \geq 3$ be a set of indices, and $\Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ a set of points in space with corresponding data weights $w_i > 0$. Thereby $z_i$'s are the measured values of the unknown function $g : \mathbb{R}^2 \to \mathbb{R}$ at points $P_i = (x_i, y_i)$, $i = 1, \ldots, m$ among which outliers appear. A best LAD-plane should be determined, i.e. optimal parameters $a^*, b^*, c^* \in \mathbb{R}$ of the function

\[
f(x, y; a, b, c) = ax + by + c,
\]

should be determined such that

\[
G(a^*, b^*, c^*) = \min_{(a, b, c) \in \mathbb{R}^3} G(a, b, c), \quad G(a, b, c) = \sum_{i=1}^{m} w_i |z_i - ax_i - by_i - c|.
\]

The problem of estimating parameters of a best LAD-plane is closely related to the notion and basic properties of the weighted median of the data. Therefore, this important notion is considered at the beginning (see [22], [23], [32]).

**Lemma 1** Let $(w_i, y_i)$, $i \in I$, $I = \{1, \ldots, m\}$, $m \geq 2$, be the data, where $y_1 \leq y_2 \leq \ldots \leq y_m$ 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 $F : \mathbb{R} \to \mathbb{R}$ be a function defined by the formula

$$F(\alpha) = \sum_{i=1}^{m} w_i |y_i - \alpha|.$$  \hfill (3)

Then

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

(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 $F$ is attained at the point $\alpha^* = y_{\nu_0 + 1}$.

(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 $F$ is attained at every point $\alpha^*$ from the segment $[y_{\nu_0}, y_{\nu_0 + 1}]$.

Remark 1 The point $\alpha^*$ in which function (3) attains its minimum is called the weighted median of the data $(w_i, y_i), i \in I$. In the case of a large number of data, calculation of the weighted median of the data may require a long computing time. Several fast algorithms can be seen in [12]. In numerical examples at the end of this paper the weighted median of the data is calculated by a modification of the algorithm proposed in [12].

Remark 2 Note that as a consequence of the Lemma 1 the pseudo–halving property of data $(w_i, y_i), i \in I$, follows directly, so it can be said that the weighted median of the data is every number $\alpha^*$ for which it holds

$$\sum_{y_i < \alpha^*} w_i \leq \frac{W}{2} \quad \text{and} \quad \sum_{y_i > \alpha^*} w_i \leq \frac{W}{2}, \quad W = \sum_{i=1}^{m} w_i.$$  \hfill (4)

Also from Lemma 1 it can be easily seen that the following holds (see e.g. [2], [23])

$$\left| \sum_{y_i < \alpha^*} w_i - \sum_{y_i > \alpha^*} w_i \right| \leq \sum_{y_i = \alpha^*} w_i.$$  \hfill (5)
Theorem 1 (Existence theorem of a best LAD-plane) Let \( I = \{1, \ldots, m\}, m \geq 3 \), be a set of indices, and \( \Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\} \) a set of points in space with corresponding data weights \( w_i > 0 \). Then there exists a best LAD-plane, i.e. problem (2) has a solution.

Proof of the existence theorem for problem (2) is simple (see e.g. [13] or [32]). In general, a best weighted LAD-plane does not have to be unique (see [2], [3], [5], [20], [29], [32]). For example, it easy to see that if data \((x_i, y_i), i = 1, \ldots, m\) lie on some line in the plane, there exist infinitely many LAD-planes. Let us mention that there also exist other cases in which LAD-plane is not unique.

The following lemma shows how among all planes passing through point \( T_0 = (x_0, y_0, z_0) \in \mathbb{R}^3 \) and being parallel to the x-axis (or to the y-axis, respectively) to select a best LAD-plane which contains at least one other data point from set \( \Lambda \setminus \{T_0\} \). Especially, point \( T_0 \) can be from set \( \Lambda \). This lemma will be used for initialization of the proposed algorithm.

Lemma 2 Let \( I = \{1, \ldots, m\}, m \geq 3 \), be a set of indices and

(i) \( \Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\} \) a set of points in space with corresponding data weights \( w_i > 0 \), such that a set of points \( \mathcal{L} = \{P_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\} \) does not lie on any line,

(ii) \( T_0 = (x_0, y_0, z_0) \in \mathbb{R}^3 \).

Then

(a) there exist \( \mu \in I \setminus \{i \in I : y_i = y_0\} \) and \( b^* = \frac{z_{\mu} - z_0}{y_{\mu} - y_0} \in \mathbb{R} \) such that \( \overline{G}(b^*) = \min_{b \in \mathbb{R}} \overline{G}(b) \), where

\[
\overline{G}(b) = \sum_{i=1}^{m} w_i |z_i - z_0 - b(y_i - y_0)|. \tag{6}
\]

Furthermore, the plane passing through point \( T_0 \) and defined by the affine function \( \overline{f}_x(x, y; b^*) = z_0 + b^*(y - y_0) \) also passes through at least point \( T_\mu = (x_\mu, y_\mu, z_\mu) \in \Lambda \setminus \{T_0\} \).
there exists ν ∈ I \ {i ∈ I : x_i = x_0} and a∗ = \frac{z_ν - z_0}{x_ν - y_0} ∈ ℝ such that G(a∗) = \min_{a ∈ ℝ} G(a), where
\[ G(a) = \sum_{i=1}^{m} w_i |z_i - z_0 - a(x_i - x_0)|, \] (7)

and the plane passing through point T_0 and defined by the affine function
\[ f_y(x, y; a^*) = z_0 + a^*(x - x_0) \] also passes through at least point \( T_ν = (x_ν, y_ν, z_ν) \) ∈ Λ \ {T_0}.

**PROOF.** (a) Let us first note that if there exists \( i_0 ∈ I \) such that \( T_{i_0} = T_0 \), then the term of sum (6) for \( i = i_0 \) vanishes. Denote \( I' = I \setminus \{i_0\} \) and \( I_0 = \{i ∈ I' : y_i = y_0\} \). Note that due to condition (i) there holds \( I' \setminus I_0 \neq \emptyset \).

According to Lemma 1, there exists \( µ ∈ I' \setminus I_0 \) such that
\[ f_x(x_µ, y_µ; b^*) = z_µ, \] where \( b^* = \frac{z_µ - z_0}{y_µ - y_0} \).

The proof of statement (b) is similar.

The following lemma occupies a central position in the proposed algorithm and shows how among all planes passing through two different points \( T_0, T_µ ∈ \mathbb{R}^3 \) and \( T_µ ∈ Λ \) to select a best LAD-plane, which also passes through at least one new point \( T_ν ∈ Λ \setminus \{T_0, T_µ\} \) that does not lie in the plane \( M(T_0, T_µ) \), given by
\[ M(T_0, T_µ) := \{(x, y, z) ∈ \mathbb{R}^3 : (y - y_µ)(x_0 - x_µ) - (x - x_µ)(y_0 - y_µ) = 0\}, \] (8)
which is parallel to the z-axis and passes through points \( T_0 \) and \( T_µ \). Especially, point \( T_0 \) can be from set \( Λ \). A similar statement can also be found in [24].

**Lemma 3** Let \( I = \{1, \ldots, m\}, m ≥ 3, \) be a set of indices and

(i) \( Λ = \{T_i = (x_i, y_i, z_i) ∈ \mathbb{R}^3 : i ∈ I\} \) a set of points in space with corresponding data weights \( w_i > 0, \) such that the set of points \( L = \{P_i = (x_i, y_i) ∈ \mathbb{R}^2 : i ∈ I\} \) does not lie on any line;
(ii) \( T_0 = (x_0, y_0, z_0) \in \mathbb{R}^3 \) and \( T_\mu = (x_\mu, y_\mu, z_\mu) \in \Lambda \setminus \{T_0\} \) such that \((x_\mu, y_\mu) \neq (x_0, y_0)\);

(iii) Let \( I_0 = \{i \in I : T_i \in M(T_0, T_\mu)\}\), where \( M(T_0, T_\mu) \) is a plane given by (8).

Then there exists a third point \( T_\nu \in \Lambda \setminus M(T_0, T_\mu) \) such that a best LAD-plane containing \( T_0 \) and \( T_\mu \) passes through these three points. Thereby,

(a) if \( x_0 \neq x_\mu \), then a best LAD-plane is determined by the affine function

\[
 f_1(x, y; \beta^*) = z_\mu + \left(\frac{z_0 - z_\mu}{x_0 - x_\mu} - \beta^* \frac{y_0 - y_\mu}{x_0 - x_\mu}\right)(x - x_\mu) + \beta^*(y - y_\mu),
\]

where \( \beta^* \) is obtained by solving the weighted median problem

\[
 \sum_{i \in I \setminus I_0} w_i \left| z_i - z_\mu - \frac{z_0 - z_\mu}{x_0 - x_\mu}(x_i - x_\mu) - \beta \left( y_i - y_\mu - \frac{y_0 - y_\mu}{x_0 - x_\mu}(x_i - x_\mu)\right) \right| \longrightarrow \min_{\beta},
\]

which also defines the third point \( T_\nu \in \Lambda \setminus M(T_0, T_\mu) \), through which this best LAD-plane passes.

(b) if \( y_0 \neq y_\mu \), then a best LAD-plane is determined by affine function

\[
 f_2(x, y; \alpha^*) = z_\mu + \alpha^*(x - x_\mu) + \left(\frac{z_0 - z_\mu}{y_0 - y_\mu} - \alpha^* \frac{x_0 - x_\mu}{y_0 - y_\mu}\right)(y - y_\mu),
\]

where \( \alpha^* \) is obtained by solving the weighted median problem

\[
 \sum_{i \in I \setminus I_0} w_i \left| z_i - z_\mu - \frac{z_0 - z_\mu}{y_0 - y_\mu}(y_i - y_\mu) - \alpha \left( x_i - x_\mu - \frac{x_0 - x_\mu}{y_0 - y_\mu}(y_i - y_\mu)\right) \right| \longrightarrow \min_{\alpha},
\]

which also defines the third point \( T_\nu \in \Lambda \setminus M(T_0, T_\mu) \), through which this best LAD-plane passes.

**Proof.** First note that, because of (i), it is possible to choose point \( T_\mu \in \Lambda \setminus \{T_0\} \) such that \((x_\mu, y_\mu) \neq (x_0, y_0)\).
(a) Assume that \( x_0 \neq x_\mu \). The affine function passing through points \( T_0, T_\mu \) is of the form

\[
f_1(x, y; \beta) = z_\mu + \frac{z_0 - z_\mu}{x_0 - x_\mu}(x - x_\mu) + \beta \left( y - y_\mu - \frac{y_0 - y_\mu}{x_0 - x_\mu}(x - x_\mu) \right).
\] (13)

Note that \( \mu \in I_0 \), and if there exists \( i_0 \in I_0 \), such that \( T_{i_0} = T_0 \), then also \( i_0 \in I_0 \), and due to \( (i) \), \( I_0 \neq \emptyset \). Note also that for \( i \in I_0 \), term \( y_i - y_\mu - \frac{y_0 - y_\mu}{x_0 - x_\mu}(x_i - x_\mu) \) vanishes. Therefore, the optimal value of parameter \( \beta \) will be found by minimizing the functional

\[
\Psi_1(\beta) = \sum_{i \in I} w_i |z_i - f_1(x_i, y_i; \beta)| = \sum_{i \in I_0} w_i \left| z_i - z_\mu - \frac{z_0 - z_\mu}{x_0 - x_\mu}(x_i - x_\mu) \right| \\
+ \sum_{i \in I \setminus I_0} w_i \left| z_i - z_\mu - \frac{z_0 - z_\mu}{x_0 - x_\mu}(x_i - x_\mu) - \beta \left( y_i - y_\mu - \frac{y_0 - y_\mu}{x_0 - x_\mu}(x_i - x_\mu) \right) \right|.
\]

According to Lemma 1, there exist \( \nu \in I \setminus I_0 \) (which also defines the third point \( T_\nu \in \Lambda \setminus M(T_0, T_\mu) \), through which this best LAD-plane passes) and the optimal value \( \beta^* \) which is obtained by solving the weightedmedian problem (10).

The proof of statement (b) is similar.

\[ \square \]

In the next theorem two important properties of a best LAD-plane are given. These results will be used for the purpose of constructing an efficient algorithm for finding a best LAD-plane.

**Theorem 2 (Properties of a best LAD-plane)** Let \( I = \{1, \ldots, m\}, m \geq 3, \) be a set of indices and \( \Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\} \) a set of points in space with corresponding data weights \( w_i > 0 \), such that the set of points \( \mathcal{L} = \{P_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\} \) does not lie on any line.

(i) Then there exists a best LAD-plane which passes through at least three different points from \( \Lambda \).

(ii) A best weighted LAD-plane is pseudo–halving.
PROOF. The proof of statement (i) can be seen in e.g. [2], [23], [32].

For the purpose of proving statement (ii), note that due to Lemma 1 the following holds:
\[
\sum_{i=1}^{m} w_i |z_i - a^* x_i - b^* y_i - c| \geq \sum_{i=1}^{m} w_i |z_i - a^* x_i - b^* y_i - c^*|,
\]
where the equality holds if and only if \(c^*\) is the weighted median of the data \((w_i, z_i - a^* x_i - b^* y_i)\), \(i = 1, \ldots, m\). According to Remark 2, by denoting \(W := \sum_{i=1}^{m} w_i\) this means
\[
\sum_{z_i < a^* x_i + b^* y_i + c^*} w_i \leq \frac{W}{2} \quad \text{and} \quad \sum_{z_i > a^* x_i + b^* y_i + c^*} w_i \leq \frac{W}{2}.
\] (14)

It is very difficult to obtain sufficient optimality conditions for such problems. However, for a broad class of problems Theorem 4 gives sufficient optimality conditions. Prior to that, let us give a Lemma 4 whose proof follows directly from the following theorem (Theorem 6.1, p. 118 in [32]):

**Theorem 3** Let \(X \in \mathbb{R}^{m \times n}\) and \(z \in \mathbb{R}^m\). A vector \(u^* \in \mathbb{R}^n\) solves the problem
\[
\min_{u \in \mathbb{R}^n} \|Xu - z\|_1,
\]
if and only if there exists \(v \in V(u^*)\) such that \(X^T v = 0\), where
\[
V(u^*) = \{v \in \mathbb{R}^m : |v_i| \leq 1; v_i = \text{sign}(r_i(u^*)), i \notin I_0\}, \quad r(u^*) = Xu^* - z,
\]
and \(I_0 = \{i \in I : r_i(u^*) = 0\}, I = \{1, \ldots, m\}\).

**Lemma 4** Vector \((a^*, b^*, c^*)^T \in \mathbb{R}^3\) is a solution to the LAD-problem (2) if and only if there exists vector \(v \in \mathbb{R}^m\) with components
\[
v_i = \begin{cases} \text{sign}(r_i^*), & r_i^* \neq 0, \\ \text{any number from interval } [-1, 1], & \text{else} \end{cases}, \quad i = 1, \ldots, m
\]
where \(r_i^* = a^* x_i + b^* y_i + c^* - z_i\) is such that \(X^T v = 0\), where
\[
X = \begin{bmatrix} w_1 x_1 & w_1 y_1 & w_1 \\ \vdots & \vdots & \vdots \\ w_m x_m & w_m y_m & w_m \end{bmatrix}.
\]
Theorem 4 Let \( I = \{1, \ldots, m\}, m \geq 3, \) be a set of indices and \( \Lambda = \{ T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I \} \) a set of points in space with corresponding data weights \( w_i > 0, \) such that the set of points \( \mathcal{L} = \{ P_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I \} \) does not lie on any line.

Plane \( z = a^*x + b^*y + c^*, \) which passes through exactly three data points \( T_\mu, T_\nu, T_k \in \Lambda, \) is a best LAD-plane if and only if the following three inequalities hold

\[
\begin{align*}
(1) \quad |D_1| & \leq |D|, \\
(2) \quad |D_2| & \leq |D|, \\
(3) \quad |D_3| & \leq |D|,
\end{align*}
\]

where \( D = \begin{vmatrix} w_\mu & w_\nu & w_k \\
w_\mu x_\mu & w_\nu x_\nu & w_k x_k \\
w_\mu y_\mu & w_\nu y_\nu & w_k y_k \end{vmatrix}, \) and determinants \( D_i \) are obtained by replacing the \( i \)-th column of determinant \( D \) by vector

\[
\left( \sum_{i \in I'} w_i \text{sign}(r_i^*), \sum_{i \in I'} w_ix_i \text{sign}(r_i^*), \sum_{i \in I'} w_iy_i \text{sign}(r_i^*) \right)^T,
\]

where \( r_i^* = a^*x_i + b^*y_i + c^* - z_i, \quad i \in I'' = \{1, \ldots, m\} \setminus \{\mu, \nu, k\}. \)

**PROOF.** Let \( z = a^*x + b^*y + c^* \) be a plane passing through exactly three data points \( T_\mu, T_\nu, T_k \in \Lambda. \) According to Lemma 4, that plane is a best LAD-plane if and only if there exist \( v_\mu, v_\nu, v_k \in [-1, 1], \) such that

\[
\begin{align*}
w_\mu x_\mu v_\mu + w_\nu x_\nu v_\nu + w_k x_kv_k &= -\sum_{i \in I''} w_ix_i \text{sign}(r_i^*) \\
w_\mu y_\mu v_\mu + w_\nu y_\nu v_\nu + w_k y_kv_k &= -\sum_{i \in I''} w_iy_i \text{sign}(r_i^*) \\
w_\mu v_\mu + w_\nu v_\nu + w_kv_k &= -\sum_{i \in I''} w_i \text{sign}(r_i^*). \tag{16}
\end{align*}
\]

By solving system (16) with respect to \( v_\mu, v_\nu \) and \( v_k \) and by using condition \( v_\mu, v_\nu, v_k \in [-1, 1], \) we obtain the theorem assertion. \( \square \)
3 Methods of searching for a best LAD-plane

To search for optimal parameters of a best LAD-plane, general methods of minimization without using derivatives can be used (see [14], [16]), but their efficiency in the case of a large number of data is very low. Efficient methods for solving this problem are based on Linear Programming (see e.g. [1], [2], [5], [20], [32]). There are also various specializations of the Gauss-Newton method (see e.g. [5], [8], [12], [20], [32]), special methods of combinatorial optimization (see [18], [33]), as well as methods which tend to smooth the LAD-problem by means of M-estimators (see [8]).

The methods we propose respond to specific demands of the problem: a large number of data and reaching a solution in real time.

3.1 Three Points (TP) method

The method proposed and described in this section is based upon the fact that a best LAD-plane passes through at least three different data points (see Theorem 2). Therefore, it will be called the Three Points (TP) method. Let $I = \{1, \ldots, m\}$, $m \geq 3$, be a set of indices, and $\Lambda = \{T_i = (x_i, y_i, z_i) \in \mathbb{R}^3 : i \in I\}$ a set of points in space with corresponding data weights $w_i > 0$. Let us also assume that a set of points $L = \{P_i = (x_i, y_i) \in \mathbb{R}^2 : i \in I\}$ does not lie on any line.

First, let us mention some remarks in relation to construction of a TP algorithm.

1. In order to get to a best LAD-plane as close as possible from the start, the centroid of the data can be selected as the initial point. Namely, the centroid of the data is probably positioned close to a best LAD-plane passing, and its coordinates are calculated quickly.

2. The second point is taken based upon Lemma 2. Thereby either statement (a) or (b) can be taken into account depending on whether there are less points whose ordinate (i.e. abscissa) corresponds to $y_0$ (i.e. $x_0$).
3. Parameters \((a_1, b_1)\) and the third point \(T_\nu \in \Lambda \setminus M(T_0, T_\mu)\) in Step 3 are chosen according to Lemma 3 in the following way:

   (i) if \(x_0 \neq x_\mu\), \(b_1\) is the solution of the weighted median problem (10), which is attained for \(\nu \in I \setminus I_0\), and
   \[ a_1 = \frac{z_0 - z_\mu}{x_0 - x_\mu} - b_1 \frac{y_0 - y_\mu}{x_0 - x_\mu}; \]

   (ii) if \(y_0 \neq y_\mu\), \(a_1\) is the solution of the weighted median problem (12), which is attained for \(\nu \in I \setminus I_0\), and
   \[ b_1 = \frac{z_0 - z_\mu}{y_0 - y_\mu} - a_1 \frac{x_0 - x_\mu}{y_0 - y_\mu}; \]

Parameters \((a_2, b_2)\) and point \(T_k \in \Lambda \setminus M(T_\mu, T_\nu)\) in Step 4 are determined analogously.

4. In the first loop through the Algorithm (Step 3) the third point is chosen according to Lemma 3 without additional restrictions.

   In all other loops (Step 4) it has to be taken into consideration that a new point does not correspond to the point that was dropped in the previous loop. Furthermore, out of three available points \(T_0, T_\mu,\) and \(T_\nu\) there really exist only two different ways for the choice of two points necessary in Lemma 3 (either \(\{T_\mu, T_\nu\}\) or \(\{T_0, T_\nu\}\)). If Lemma 3 gives a new point (different from \(T_0\) or \(T_\mu\)), then in the algorithm decreasing of the value of the minimizing function is attained; otherwise, the optimum is attained.

5. Bearing in mind a real time problem of estimating optimal parameters of a best LAD-plane on the basis of a large number of data, special attention is to be paid to the application of an efficient algorithm for calculating the weighted median of the data in Step 2, Step 3 and Step 4. Several fast algorithms for that can be seen in [12].

6. Note that the TP algorithm is constructed such that in every step the value of the minimizing functional decreases and since minimizing functional (2) is convex, the optimum is attained in finitely many steps. Namely, sequence \(G(a_1, b_1, c_1), G(a_2, b_2, c_2), \ldots\) is decreasing and finite.

**Algorithm Three Points**
Step 0. (Input) $I = \{1, \ldots, m\}, \Lambda = \{T_i = (x_i, y_i, z_i) : i \in I\}, \{w_i > 0, i \in I\};$

Step 1. (Determining the first point – centroid of the data) Calculate
\[ W = \sum_{i=1}^{m} w_i, \quad x_0 = \frac{1}{W} \sum_{i=1}^{m} w_i x_i, \quad y_0 = \frac{1}{W} \sum_{i=1}^{m} w_i y_i, \quad z_0 = \frac{1}{W} \sum_{i=1}^{m} w_i z_i; \]
and set $T_0 = (x_0, y_0, z_0);$

Step 2. (Searching for the second point and initial approximation) According to Lemma 2, determine the second point $T_\mu \in \Lambda \setminus \{T_0\};$

Step 3. (Searching for the third point) According to Lemma 3, determine the third point $T_\nu \in \Lambda \setminus M(T_0, T_\mu)$ and $(a_1, b_1) \in \mathbb{R}^2.$ Calculate $c_1 = z_\nu - a_1 x_\nu - b_1 y_\nu$ and $G_1 = G(a_1, b_1, c_1);$

Step 4. (Searching for a new point) According to Lemma 3, determine a new point $T_k \in \Lambda \setminus M(T_\mu, T_\nu)$ and $(a_2, b_2) \in \mathbb{R}^2$ and set $T'_\mu = T_\mu;$

If $T_k = T_0$, set $T'_\mu = T_0$ and, according to Lemma 3, determine a new point $T_k \in \Lambda \setminus M(T_\mu, T_\nu)$ and $(a_2, b_2) \in \mathbb{R}^2;

Step 5. (Preparation for a new loop)
If $[T_k \neq T'_\mu]$, calculate $c_2 = z_k - a_2 x_k - b_2 y_k$ and $G_2 = G(a_2, b_2, c_2)$
and set: $T_0 = T'_\mu; \quad T_\mu = T_\nu; \quad T_\nu = T_k;$ and go to Step 4,
else STOP]

Remark 3 Note that the Three Points algorithm can be understood as an algorithm for searching for three different points from set $\Lambda$, which do not lie in the plane parallel with the z-axis and which determine a best LAD-plane. It means that by steps of the algorithm it is necessary to calculate neither parameters $(a_i, b_i, c_i)$ nor values of the minimizing functional $G(a_i, b_i, c_i).$ Since the algorithm is finite, at the end of the algorithm we obtain three points which determine a best LAD-plane.
Note also that the first two initial points can be selected randomly from the set of given points \( \Lambda \). Such selection of initial points also leads to the optimum, but the duration of the iterative process depends heavily on the selected initial points.

\[
\begin{array}{cccc}
1 & 0.88 & 4.99 & -45.29 \\
2 & 10.17 & 4.53 & 40.62 \\
3 & 2.4 & 9.82 & -32.77 \\
4 & 0.9 & 3.98 & -59.12 \\
5 & 7.53 & 4.2 & 56.47 \\
6 & 2.39 & 12.17 & 61.62 \\
7 & 11.9 & 12.46 & 52.91 \\
8 & 2.84 & 12.86 & 26.18 \\
9 & 6.39 & 1.02 & -32.77 \\
10 & 12.95 & 6.68 & 50.44 \\
11 & 6.35 & 1.9 & 24.41 \\
12 & 6.82 & 8.87 & 33.37 \\
13 & 1.11 & 2.77 & 12.18 \\
14 & 10.69 & 10.2 & 49.51 \\
15 & 11.87 & 12.61 & 51.61 \\
16 & 9.65 & 1.64 & 37.2 \\
17 & 6.14 & 3.7 & 27.39 \\
18 & 8.66 & 13.65 & 39.9 \\
19 & 7.53 & 13.55 & 40.41 \\
20 & 11.63 & 0.5 & 36.68 \\
21 & 0.89 & 9.02 & 17.12 \\
22 & 12.86 & 12.76 & 55.97 \\
23 & 2.96 & 3.91 & 19.33 \\
24 & 8.93 & 7.68 & 35.63 \\
25 & 13.9 & 2.3 & 49.91 \\
26 & 0.76 & 7.69 & 13.53 \\
27 & 13.15 & 0.89 & 46.34 \\
28 & 2.53 & 1.81 & 13.81 \\
29 & 4.24 & 12.11 & 29.59 \\
30 & 4.36 & 6.01 & 23.51 \\
31 & 8.71 & 0.11 & 11.12 \\
32 & 3.76 & 7.24 & 20.46 \\
33 & 1.28 & 0.11 & 9.59 \\
34 & 1.28 & 0.11 & 9.59 \\
35 & 12.84 & 8.77 & 52.43 \\
36 & 10.85 & 1.68 & 38.53 \\
37 & 7.14 & 13.17 & 36.62 \\
38 & 3.7 & 2.01 & 18.04 \\
39 & 11.28 & 6.14 & 45.31 \\
40 & 7.53 & 13.55 & 40.41 \\
41 & 11.28 & 6.14 & 45.31 \\
42 & 1.28 & 0.11 & 9.59 \\
43 & 3.7 & 2.01 & 18.04 \\
44 & 7.53 & 13.55 & 40.41 \\
45 & 11.28 & 6.14 & 45.31 \\
\end{array}
\]

Table 1: The data (outliers are denoted bold)

Example 1 Let the set of points \( \Lambda = \{ T_i = (x_i, y_i, z_i) : i = 1, \ldots, 45 \} \) be given, where the data \( (x_i, y_i, z_i) \) are given in Table 1 and \( w_1 = \cdots = w_{45} = 1 \). Data points \( (x_i, y_i), i = 1, \ldots, 45 \) are selected randomly on the set \([0, 15] \times [0, 15]\) and \( z_i = 3x_i + y_i + 4 + \varepsilon_i \), where \( \varepsilon_i \sim N(0, 4) \), whereby a random outlier from the interval \((-80, 80)\) is added to 6 randomly selected data.

The proposed TP algorithm will be illustrated on the basis of these data. The flow of the iterative process\(^6\) is shown in Table 2. The initial point (centroid of the data) is \( T_0 = (7.067, 6.68, 29.54) \).

Additionally, in every step parameter values and the value of the minimizing functional \( G \) are computed. Also, on the basis of (14), the pseudo-halving property is checked in every step (see Column 6): the top (i.e. bottom) number represents the number of points above (i.e. below) the plane. Note that the pseudo-halving

\(^6\)All evaluations were done by MS Visual C++ on the basis of our own software on a PC (CPU: 1.73 GHz Intel Celeron M, Memory: 1.5 GB DDR2).
property is satisfied before the optimum is attained, which means that this criterion may be used only as a necessary condition of optimality, as claimed in Theorem 2. Condition (15) from Theorem 4 is also checked on this example. Results can be seen in Column 7. Number "1" in the $i$-th place means here that condition $(i)$ from (15) is fulfilled; number "0" means that this condition is not fulfilled. Both on this and on numerous other examples it is shown that condition (15) is fulfilled no sooner than optimal values of parameters are attained. It means that in accordance with Theorem 4 this condition can be considered a sufficient condition of optimality with reserve that a optimal LAD-plane passes through exactly three points. Meanings of values $\varphi_k$ and $\psi_k$ shown in Column 8 and Column 9 will be mentioned later in a description of a modified Three Points algorithm.

The first four iterations of the proposed TP algorithm are also illustrated in Figure 1.

### 3.2 Modified Three Points (MTP) Method

In the case of a large number of data, such as in robotics (see [7], [11], [26]), the number of iterations necessary for obtaining an optimal solution can be great. If the results are to be obtained in real time, then a sufficiently good approximate solution would suffice. As confirmed by numerous examples, the algorithm comes close to the optimal solution very fast, after which a greater number of minor corrections is necessary in order to attain the optimum.
In order to achieve the shortest execution time of the iterative process, it is important to answer how to define a suitable criterion for stopping the iterative process before the optimum is attained, so that approximation of a solution is satisfactory, and at the same time the least number of iterations is carried out with minimum calculations.

For that purpose, let us denote by \( \mathbf{n}_1 \) and \( \mathbf{n}_2 \) unit normal vectors on planes \( z = a_1 x + b_1 y + c_1 \) and \( z = a_2 x + b_2 y + c_2 \), respectively. Assuming that the angle \( \varphi \) between them is small, there holds

\[
\varphi \approx \sin \varphi \approx \frac{||\mathbf{n}_2 - \mathbf{n}_1||_2}{||\mathbf{n}_1||_2} = ||\mathbf{n}_2 - \mathbf{n}_1||_2.
\]
In Table 2, Column 8 shows behavior of the angle $\varphi_k$ between normals of neighboring approximation planes.

Also, by applying formula (18) we can observe the angle $\psi_k$ between a LAD-plane obtained in the $k$-th iteration and the plane by means of which data are generated (see Column 9 in Table 2). It can be noted that after only first several iterations satisfactory correspondence is reached.

These are all reasons due to which for some practical purposes the TP algorithm might be modified such that with some $\epsilon > 0$ instead of Step 5 we introduce

\textbf{Step 5’}. (Preparation for a new loop)

\begin{itemize}
  \item Calculate $c_2 = z_k - a_2x_k - b_2y_k$, $u_1 = \sqrt{a_1^2 + b_1^2 + 1}$, $u_2 = \sqrt{a_2^2 + b_2^2 + 1}$;
  \item Calculate $\varphi = \frac{1}{u_1u_2} \sqrt{(a_1u_2 - a_2u_1)^2 + (b_1u_2 - b_2u_1)^2 + (u_2 - u_1)^2}$
  \item If $\varphi \geq \epsilon$, set: $T_0 = T_\mu$; $T_\mu = T_\nu$; $T_\nu = T_k$; and go to Step 4
  \item else STOP
\end{itemize}

\textbf{Remark 4} The iterative procedure will stop if the angle $\varphi$ between neighboring approximation planes becomes sufficiently small. Note that in the MTP algorithm constructed in this way it is not necessary to calculate the value of the minimizing functional $G$, but only parameter values $(a_i, b_i, c_i)$.

In the following two examples we will examine the MTP algorithm in the case of a large number of data among which a significant number of outliers might appear.

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|l|l|l|l|}
\hline
1 & 2 & 3 & 4 & 5 & 6 \\
\hline
m & (a^*, b^*, c^*) & $G(a^*, b^*, c^*)$ & $\psi$(in deg) & it & time[ms] \\
\hline
10 000 & (0.790, 1.981,113.945) & 0.22 & 7(17) & 12(26) \\
20 000 & (0.800, 1.988,111.784) & 0.15 & 6(16) & 21(54) \\
30 000 & (0.796, 1.974,114.526) & 0.29 & 7(29) & 37(151) \\
40 000 & (0.812, 1.975,112.395) & 0.54 & 7(32) & 51(214) \\
\hline
\end{tabular}
\caption{Comparison of the Three Points and the Modified Three Points algorithm for 20\% outliers}
\end{table}
Example 2 Let the set of points $\Lambda = \{ T_i = (x_i, y_i, z_i) : i = 1, \ldots, m \}$ be given, where data points $(x_i, y_i), i = 1, \ldots, m$ are selected randomly on the set $[0, 320] \times [0, 240]$ and $z_i = .8x_i + 2y_i + 110 + \varepsilon_i$, where $\varepsilon_i \sim N(0, 256)$, whereby a random outlier from the interval $(0, 1000)$ is added to a greater number of randomly selected data. Thereby $w_1 = \cdots = w_m = 1$.

In Table 3 we will compare results obtained by the TP algorithm and the MTP algorithm if to 20% of $m$ data we add random outliers from the interval $(0, 1000)$. In Table 4 we will compare results obtained by the TP algorithm and the MTP algorithm if among $m = 20000$ data there appear 10%, 20%, 30% or 40% random outliers from the interval $(0, 1000)$. Results will be compared on the basis of the following indicators: value of the minimizing functional $G$ in Column 3, angle $\psi_k$ between a best LAD-plane obtained by the MTP method and the plane $z = .8x + 2y + 110$ (by using (18)) in Column 4, the number of iteration in Column 5 and computing time in Column 6. Thereby, indicators referring to a best LAD-plane obtained by the TP algorithm are given inside brackets, whereas indicators referring to the MTP algorithm with $\epsilon = .004$ are given outside brackets. The weighted median of the data is calculated by a modification of the algorithm proposed in [12].

<table>
<thead>
<tr>
<th>% outliers</th>
<th>$(a^<em>, b^</em>, c^*)$</th>
<th>$G(a^<em>, b^</em>, c^*)$</th>
<th>$\psi$ (in deg)</th>
<th>it</th>
<th>time[ms]</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>(0.808,1.990,110.285)</td>
<td>788301</td>
<td>0.27(0.23)</td>
<td>9(27)</td>
<td>30(88)</td>
</tr>
<tr>
<td>20%</td>
<td>(0.786,1.960,117.053)</td>
<td>1323180</td>
<td>0.44(0.21)</td>
<td>6(20)</td>
<td>32(71)</td>
</tr>
<tr>
<td>30%</td>
<td>(0.789,1.963,116.809)</td>
<td>1833125</td>
<td>0.41(0.20)</td>
<td>8(17)</td>
<td>26(58)</td>
</tr>
<tr>
<td>40%</td>
<td>(0.723,1.992,127.392)</td>
<td>2371698</td>
<td>1.71(0.25)</td>
<td>6(25)</td>
<td>21(84)</td>
</tr>
</tbody>
</table>

Table 4: Comparison of the Three Points and the Modified Three Points algorithm for 20 000 data points
4 Application in robotics

Detection of planar surfaces is an important problem in robot vision. Most mobile robots today are constructed to operate in indoor environment moving on a flat horizontal surface called herein the ground plane. Safe obstacle avoidance can be performed by classifying all environment points into two sets. The first represents all points lying sufficiently close to the ground plane so that the robot can safely roll over them. The second set contains all other points representing the obstacles which must be avoided. Furthermore, dominant surfaces in indoor environments, such as floor, walls, doors, cupboard surfaces, etc. are planar or approximately planar. Hence, detection of planar surfaces can play an important role in an advanced environment perception system of a mobile robot.

In the experiment presented in this section, the proposed Three Points algorithm has been applied for estimation of the ground plane using the data obtained by stereo vision. Stereo vision [10] is a commonly used tool for three-dimensional reconstruction of robot’s environment from images acquired by cameras. A stereo vision system consists of two or more cameras and a computer which processes the camera images using an appropriate stereo reconstruction software. The result obtained by a stereo vision system is a set of 3D points representing some of the visible surfaces in the observed scene. Stereo vision has already been used to reconstruct the ground plane on which a robot moves [4], [7], [26] and other important planar structures in the robot’s environment [19].

In the experiment presented in the sequel, a commercially available stereo vision system [17] has been used. A stereo camera pair has been mounted on a mobile robot and directed towards the floor surface in front of the robot. The image acquired by the left camera is shown in Fig. 2 (left). For each point of the image acquired by the left camera, a stereo reconstruction algorithm searches for the point in the right camera image which corresponds to the same point in the scene. If the search is successful, the left image point is assigned a value representing the distance between its position and the position of the corresponding right image point. This value is called disparity. The image obtained by assigning a disparity value to each point of
a camera image is called a disparity map. The position of a point relative to the camera system can be computed from the position of its image projection and the corresponding disparity [10]. Hence, the disparity map actually represents a three-dimensional reconstruction of the camera image. An example of a disparity map is shown in Fig. 2 (middle), where the darker points have smaller disparities and the lighter points have greater disparities. The points for which the stereo algorithm could not find a match in the right image are denoted by black color and these points are not assigned a disparity.

From the points with an assigned disparity, a set $\Lambda = \{T_i = (x_i, y_i, z_i) : i = 1, \ldots, m\}$ is formed where $x_i$ and $y_i$ are the coordinates of the $i^{th}$ point in the image expressed in pixels and $z_i$ is the disparity assigned to the point expressed in 1/16 of a pixel. The space of such points is called the disparity space. It can be easily shown that a plane in a Euclidean 3D space maps to a plane in the disparity space (see [6]). In the considered experiment, the detection of the ground surface is performed by searching for a best LAD-plane in the disparity space. The points lying on the estimated plane within the tolerance of 32 are represented by gray color in Fig. 2 (right).

The three-dimensional information provided by stereo vision is often corrupted by a substantial amount of outliers. Many of them are caused by the ambiguity in stereo matching. Furthermore, in cases where the input data represent multiple structures, e.g. two or more objects or surfaces in a scene, the inliers of one structure
represent outliers of another structure. In order to evaluate the performance of the proposed Three Points algorithm with respect to outliers, two objects, a book and a box, have been placed onto the floor surface and the algorithm has been applied to disparity maps obtained by the stereo vision system. The left camera images, the corresponding disparity maps and the points lying on the estimated planes within the tolerance of 32 are shown in Figs. 3 and 4. Numerical results of the performed experiment are presented in Table 5. By comparing Fig. 3 (right) to Fig. 2 (right) it can be noticed that most of the points representing the book are denoted as not belonging to the estimated ground plane, while most of the ground points are correctly assigned to the ground plane. Furthermore, additional outliers introduced by placing the book on the floor surface did not cause a significant change in the
Table 5: Evaluation of the MTP algorithm and RANSAC of data obtained by stereo vision

<table>
<thead>
<tr>
<th>Sample</th>
<th>MTP ((a^<em>, b^</em>, c^*))</th>
<th>RANSAC ((a^<em>, b^</em>, c^*))</th>
</tr>
</thead>
<tbody>
<tr>
<td>floor</td>
<td>((-0.043, 4.915, -116.08))</td>
<td>((0.000, 4.915, -122.53))</td>
</tr>
<tr>
<td>book</td>
<td>((-0.018, 4.864, -109.59))</td>
<td>((0.004, 4.900, -119.57))</td>
</tr>
<tr>
<td>box</td>
<td>((-0.190, 1.181, 550.49))</td>
<td>((0.002, 4.891, -118.27))</td>
</tr>
</tbody>
</table>

estimated ground plane parameters, as can be seen from the results presented in Table 5. However, placing a box on the floor surface so that it covers approximately one half of the image data resulted in a false ground plane estimation, as can be seen from Fig. 4 (right) and Table 5.

We compared performance of the MTP algorithm with RANSAC [11], a method widely used in robotics to fit a model to a set of data corrupted by outliers. The number of iterations is chosen to be 35, because for this number of iterations probability of randomly selecting 3 points sufficient to define a plane in the case of data corrupted by 50% of outliers is greater than 99%. A discussion about a sufficient number of RANSAC iterations can be found in [11] together with other details about RANSAC. In cases where a smaller amount of outliers is expected, as in examples presented in Figs. 2 and 3, a RANSAC procedure with less iterations, and thus with a lower execution time could be applied. Nevertheless, for the presented examples, computational efficiency of the novel MTP algorithm is comparable with that of RANSAC. In the example with the box presented in Fig. 4, RANSAC gives a better result. However, RANSAC is a randomized algorithm, which means that it can give a false result in a certain number of trials, while MTP provides a result which is arbitrary close to the optimal solution according to the \(l_1\) norm.

It can be concluded that in the case of data corrupted with a moderate amount of outliers, the ground plane estimation can be performed by searching for a best LAD-plane. However, this method has its limitations. It can be shown that the performance of the method depends not only on the amount of outliers but also on their distribution. Inherent robustness against outliers and a rather high computational
efficiency, which can be seen from the execution times presented in Tables 3–5 indicate the potential of the proposed Three Points algorithm for real-time applications, e.g. in robotics. A proper combination of the proposed algorithm with a method robust to outliers such as RANSAC could potentially provide a good solution to problems such as stereo and range image segmentation, detection of dominant planar surfaces, etc.

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

**References**


