Comput Optim Appl
DOI 10.1007/s10589-014-9643-2

Robust registration of point sets using iteratively reweighted least squares

Per Bergström · Ove Edlund

Received: 27 April 2013

Abstract Registration of point sets is done by finding a rotation and translation that produces a best fit between a set of data points and a set of model points. We use robust M-estimation techniques to limit the influence of outliers, more specifically a modified version of the iterative closest point algorithm where we use iteratively re-weighed least squares to incorporate the robustness. We prove convergence with respect to the value of the objective function for this algorithm. A comparison is also done of different criterion functions to figure out their abilities to do appropriate point set fits, when the sets of data points contains outliers. The robust methods prove to be superior to least squares minimization in this setting.

Keywords Convergence · ICP · IRLS · M-estimation · Registration · Robust

1 Introduction

The problem of registering point sets, considered in this paper, is that of finding a rigid body transformation, consisting of a rotation and translation, such that it transforms a set of data points to a set of model points producing a best fit between the point sets. A common application where this kind of registration problem needs to be solved, is image reconstruction of shape measurements, see e.g. [22,23], where partially overlapping images acquired from different views are to be merged. Another similar application is free-form surface inspection, see e.g. [4,32], where a set of data

P. Bergström (✉) · O. Edlund
Division of Mathematical Sciences, Department of Engineering Sciences and Mathematics,
Luleå University of Technology, 971 87 Luleå, Sweden
e-mail: per.bergstrom@ltu.se

O. Edlund
e-mail: ove.edlund@ltu.se

Published online: 21 February 2014

points representing a measured shape of an object is to be aligned to its ideal shape represented by a CAD-model. There are applications in other areas as well. Finger print verification is one, see e.g. [11]. One use in chemistry, is described in [1,14], where protein structures are aligned.

The most commonly used method for doing the registration of point sets, are all based on the iterative closest point (ICP) algorithm [3,30,31], where the closest model points to the data points are to be found in each iteration. In most application these points are in $\mathbb{R}^2$ or $\mathbb{R}^3$ but higher dimensions are possible as well.

Real-world applications that are prone to measurement errors and disturbing background, benefit from the use of robust methods, as considered in e.g. [5,16,18,21,25,26,33]. If robust methods are not used, the registration runs the risk of being spoiled by deviant and incorrect data, since least squares algorithms are highly sensitive to gross errors in the data

A least median of squares (LMS or LMedS) version of the ICP algorithm is used in [12,15,16]. The LMS estimates are very robust and are not affected by up to 50 % outliers. A drawback is the multitude of local minima that the method introduces.

Weighting of point-pairs is one of the robust registration methods discussed in [6,21]. The method is based on assigning different weights to different point-pairs and a weighted least squares problem is then solved. Lower weights are assigned to point-pairs that seem to be outliers. These weights may be based on the distance between the points or some quality values of the data points obtained from their source.

Another method discussed in [21] is rejection of point-pairs. That method is quite similar to weighting of point-pairs but the weights are either one or zero. These rejected point-pairs can be point-pairs with an inner distance larger than a given distance apart [27] or the n % of point-pairs with largest distance [20]. This distance must be determined somehow in the first case and computational time must be spent determining which of the point-pairs that belongs to these n % in the second case. Another rejection method discussed in [27] is based on rejection of point-pairs with a model point on the boundary to the model. It gives a geometry dependent rejection method. Using a rejection method "good" data might also be rejected.

A method using the Levenberg–Marquardt algorithm, LM-ICP, is presented in [5]. It is applied directly on the squared sum of shortest distances, so the Levenberg–Marquardt step is used for the rigid body transformation step. What is required though is the gradient at the iteration point, which either must be approximated at considerable computational cost, or estimated from a precomputed customized data structure. A good point of the LM-ICP algorithm, is that it makes it is easy to introduce robust estimators.

In our paper we show that is easy to introduce robust M-estimators in the traditional ICP algorithm, and still have a convergent algorithm.

The algorithm that we analyze is a version of the ICP algorithm that uses iteratively re-weighed least squares (IRLS), which has previously been touched upon in e.g. [25]. IRLS is widely used in robust linear regression, see e.g. [9,28,29]. From a computational point of view this method is very similar to the method of weighting of point-pairs, but the weights are obtained from an M-estimation criterion. A scaling of the residuals is needed in order to define what a large residual is. In each iteration we are determining a scale parameter for fixed residuals. An update of the rigid body

transformation is then found while the scale parameter is kept unchanged. The standard deviation of the error of the "good" data that represents what it is supposed to represent pretty well is assumed to be known. It should be given from knowledge about the acquirement of the data or by some other assumptions.

2 Robust registration of point sets

In the following we will consider robust registration of point sets in $\mathbb{R}^3$, but the reasoning is valid for other dimensions as well. We want to find a rigid body transformation consisting of a rotation matrix $\mathbf{R} \in \{\mathbb{R}^{3\times 3} | \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = +1\}$ and a translation vector $\mathbf{t} \in \mathbb{R}^3$ that fits a set of N data points, $\{\mathbf{p}_i\}_{i=1}^N$ to a set of model points X, i.e. we want to solve a registration problem. A robust M-estimate of the rigid body transformation using a robust criterion function $\varrho : \mathbb{R} \to [0, \infty)$ is obtained by solving

$$\min_{\mathbf{R},\mathbf{t}} f(\mathbf{R}, \mathbf{t}) , \tag{1}$$

where

$$f(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N} \varrho(\mathrm{d}(\mathbf{R}\mathbf{p}_i + \mathbf{t}, X)) , \tag{2}$$

and d is a distance operator defined by

$$\mathrm{d}(\mathbf{p}, X) = \min_{\mathbf{y}\in X} \|\mathbf{p} - \mathbf{y}\|_2 .$$

An essential property of a robust criterion function ϱ is that it reduces undesired influence of data with gross errors on the estimation. We make a formal definition of a set of criterion functions $\mathcal{Q}$ for which we are proving convergence. The derivative of ϱ is denoted with ψ, $\psi(r) = \varrho'(r)$, which is the standard notation in robust statistics. A criterion function ϱ belongs to the set $\mathcal{Q}$ if and only if all of the following conditions are fulfilled:

1. $\varrho(r)$ is an even function and $\boldsymbol{C}^1$-continuous on $\mathbb{R}$, and $\varrho(0) = 0$,
2. $\varrho(r)$ is monotonically increasing on $[0, \infty)$,
3. $\psi(r)/r$ is monotonically decreasing and bounded above on $(0, \infty)$.

Since the function ϱ is even and continuously differentiable, its derivative at zero satisfies $\psi(0) = 0$. The second derivative $\varrho''(0) = \psi'(0)$ is always defined and bounded above. Criterion functions in $\mathcal{Q}$ are monotonically increasing on $[0, \infty)$ so that $\psi(r) \geq 0$ is satisfied for all $r \geq 0$.

A weight function w is defined as

$$w(r) = \begin{cases} \dfrac{\psi(r)}{r} & \text{if } r \neq 0, \\ \lim\limits_{r\to 0} \dfrac{\psi(r)}{r} = \psi'(0) & \text{if } r = 0. \end{cases} \tag{3}$$

From the definition of $\mathcal{Q}$ it follows that the weight function w associated to $\varrho \in \mathcal{Q}$ is an even and non-negative function which is bounded above and monotonically decreasing on $[0, \infty)$. A property of $w(r)$ is that $w(r)r = \psi(r)$ holds for all $r \in \mathbb{R}$.

Huber's function ϱ_{Hu}, Cauchy's function ϱ_{Ca}, and Tukey's bi-weight function ϱ_{Tu} are examples of robust criterion functions in $\mathcal{Q}$. These functions, their derivatives, and corresponding weight functions are as follows:

– Huber's function

$$\varrho_{\mathrm{Hu}}(r) = \begin{cases} \frac{1}{2}r^2 & \text{if} |r| \le \kappa_{\mathrm{Hu}} \\ \kappa_{\mathrm{Hu}}|r| - \frac{\kappa_{\mathrm{Hu}}^2}{2} & \text{if} |r| > \kappa_{\mathrm{Hu}} \end{cases}$$

$$\psi_{\mathrm{Hu}}(r) = \begin{cases} r & \text{if} |r| \le \kappa_{\mathrm{Hu}} \\ \mathrm{sgn}(r)\kappa_{\mathrm{Hu}} & \text{if} |r| > \kappa_{\mathrm{Hu}} \end{cases}$$

$$w_{\mathrm{Hu}}(r) = \begin{cases} 1 & \text{if} |r| \le \kappa_{\mathrm{Hu}} \\ \frac{\kappa_{\mathrm{Hu}}}{|r|} & \text{if} |r| > \kappa_{\mathrm{Hu}} \end{cases}$$

– Cauchy's function

$$\varrho_{\mathrm{Ca}}(r) = \frac{\kappa_{\mathrm{Ca}}^2}{2} \log\left(1 + \left(\frac{r}{\kappa_{\mathrm{Ca}}}\right)^2\right)$$

$$\psi_{\mathrm{Ca}}(r) = \frac{r}{1 + \left(\frac{r}{\kappa_{\mathrm{Ca}}}\right)^2}$$

$$w_{\mathrm{Ca}}(r) = \frac{1}{1 + \left(\frac{r}{\kappa_{\mathrm{Ca}}}\right)^2}$$

– Tukey's bi-weight function

$$\varrho_{\mathrm{Tu}}(r) = \begin{cases} \frac{\kappa_{\mathrm{Tu}}^2}{6}\left\{1 - \left(1 - \frac{r^2}{\kappa_{\mathrm{Tu}}^2}\right)^3\right\} & \text{if } |r| \le \kappa_{\mathrm{Tu}} \\ \frac{\kappa_{\mathrm{Tu}}^2}{6} & \text{if } |r| > \kappa_{\mathrm{Tu}} \end{cases}$$

$$\psi_{\mathrm{Tu}}(r) = \begin{cases} r\left(1 - \frac{r^2}{\kappa_{\mathrm{Tu}}^2}\right)^2 & \text{if } |r| \le \kappa_{\mathrm{Tu}} \\ 0 & \text{if } |r| > \kappa_{\mathrm{Tu}} \end{cases}$$

$$w_{\mathrm{Tu}}(r) = \begin{cases} \left(1 - \frac{r^2}{\kappa_{\mathrm{Tu}}^2}\right)^2 & \text{if } |r| \le \kappa_{\mathrm{Tu}} \\ 0 & \text{if } |r| > \kappa_{\mathrm{Tu}} \end{cases}$$

These three criterion functions, which are plotted in Fig. 1, have different robust properties with respect to handling data with occasional gross errors, when estimating parameters. Tukey's bi-weight criterion function has the strongest protection of high influence caused by this kind of undesired data, which are excluded in the minimization. Cauchy's criterion function also has a strong protection against influence from

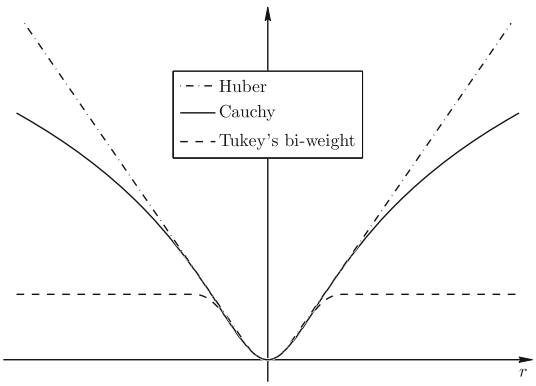

Fig. 1 Three robust criterion functions, Huber's function, Cauchy's function, and Tukey's bi-weight function

Algorithm 1 The IRLS-ICP algorithm

/ Operations are performed for all $i = 1, \ldots, N$ */*
Require: $\{\mathbf{p}_i\}$, X, weight function w, $\mathbf{R}^{(0)} = \mathbf{I}$, $\mathbf{t}^{(0)} = \mathbf{0}$
  $k = 0$, $\mathbf{p}_i^{(0)} = \mathbf{R}^{(0)}\mathbf{p}_i + \mathbf{t}^{(0)}$
  repeat
    $k = k + 1$
    $\mathbf{y}_i^{(k-1)} = \mathcal{C}(\mathbf{p}_i^{(k-1)}, X)$
    $w_i = w(\|\mathbf{p}_i^{(k-1)} - \mathbf{y}_i^{(k-1)}\|_2)$
    if any $w_i > 0$ **then**
      $\left[\mathbf{R}^*, \mathbf{t}^*\right] = \arg\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} w_i \|\mathbf{R}\mathbf{p}_i^{(k-1)} + \mathbf{t} - \mathbf{y}_i^{(k-1)}\|_2^2$
      $\mathbf{p}_i^{(k)} = \mathbf{R}^*\mathbf{p}_i^{(k-1)} + \mathbf{t}^*$
      $\mathbf{R}^{(k)} = \mathbf{R}^*\mathbf{R}^{(k-1)}$
      $\mathbf{t}^{(k)} = \mathbf{R}^*\mathbf{t}^{(k-1)} + \mathbf{t}^*$
    else */* all $w_i = 0$ */*
      $\mathbf{p}_i^{(k)} = \mathbf{p}_i^{(k-1)}$
      $\mathbf{R}^{(k)} = \mathbf{R}^{(k-1)}$
      $\mathbf{t}^{(k)} = \mathbf{t}^{(k-1)}$
      Terminate iterations
    end if
  until convergence
  return $\mathbf{R}^{(k)}, \mathbf{t}^{(k)}$

outliers. Huber's criterion function is convex, and thus does not introduce local minima to the same extent. While it still is robust, it does not completely dis-regard highly deviating points, as Tukey's function does.

We solve (1) using criterion functions from $\mathcal{Q}$, and associated weight function w defined in (3), using the IRLS-ICP algorithm, presented in Algorithm 1. It operates in two principal stages in each iteration. These stages are; finding closest model points to the data points, and finding a rigid body transformation such that the data points are fitted to the closest model points. An initial rigid body transformation, $(\mathbf{R}^{(0)}, \mathbf{t}^{(0)})$, is required. In the testing phase we use $\mathbf{R}^{(0)} = \mathbf{I}$ and $\mathbf{t}^{(0)} = \mathbf{0}$, but if there is any prior knowledge that can be utilized, the algorithm benefits greatly from it. In Algorithm 1

we use the notation $\mathcal{C}$ to denote a closest point operator. That is $\mathbf{y}_i^{(k-1)} = \mathcal{C}(\mathbf{p}_i^{(k-1)}, X)$ is the closest model point in X to a transformed data point, $\mathbf{p}_i^{(k-1)} = \mathbf{R}^{(k-1)}\mathbf{p}_i + \mathbf{t}^{(k-1)}$.

To make the notation simpler we will drop the superscript with the iteration index where it is convenient. In Algorithm 1 the update $(\mathbf{R}^*, \mathbf{t}^*)$ of the transformation is obtained from

$$\left[\mathbf{R}^*, \mathbf{t}^*\right] = \arg\min_{\mathbf{R},\mathbf{t}} g(\mathbf{R}, \mathbf{t}) \,, \tag{4}$$

where

$$g(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N} w_i \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{y}_i\|_2^2. \tag{5}$$

This is a weighted rigid body transformation problem considered in e.g. [17]. There always exists a solution set for this problem since $w_i \geq 0$ for $i = 1, \ldots, N$. The un-weighted rigid body transformation problem is discussed in e.g. [2,8,24]. Let

$$\hat{w} = \sum_{i=1}^{N} w_i \,, \;\; \bar{\mathbf{p}} = \frac{1}{\hat{w}} \sum_{i=1}^{N} w_i \mathbf{p}_i \,, \;\; \bar{\mathbf{y}} = \frac{1}{\hat{w}} \sum_{i=1}^{N} w_i \mathbf{y}_i \,,$$

where $\bar{\mathbf{p}}$ and $\bar{\mathbf{y}}$ are weighted arithmetical mean values of the data points $\{\mathbf{p}_i\}$ and the closest model points $\{\mathbf{y}_i\}$. We replace $\mathbf{t}$ with a new independent vector $\mathbf{u}$ by the change of variables $\mathbf{t} = -\mathbf{R}\bar{\mathbf{p}} + \bar{\mathbf{y}} + \mathbf{u}$. Thus the function g in (5) expressed in terms of $\mathbf{R}$ and $\mathbf{u}$ is

$$\sum_{i=1}^{N} w_i \|\mathbf{R}\mathbf{p}_i - \mathbf{R}\bar{\mathbf{p}} + \bar{\mathbf{y}} + \mathbf{u} - \mathbf{y}_i\|_2^2, \tag{6}$$

which can be expressed in terms of scalar products as

$$\sum_{i=1}^{N} w_i (\mathbf{R}(\mathbf{p}_i - \bar{\mathbf{p}}) - (\mathbf{y}_i - \bar{\mathbf{y}}) + \mathbf{u})^{\mathrm{T}} (\mathbf{R}(\mathbf{p}_i - \bar{\mathbf{p}}) - (\mathbf{y}_i - \bar{\mathbf{y}}) + \mathbf{u}) \,.$$

Since both the sum of all $w_i(\mathbf{p}_i - \bar{\mathbf{p}})$ and the sum of all $w_i(\mathbf{y}_i - \bar{\mathbf{y}})$ are equal to the zero vector their scalar product with $\mathbf{u}$ vanish. The expression above can therefore be simplified to

$$\sum_{i=1}^{N} w_i \|\mathbf{p}_i - \bar{\mathbf{p}}\|_2^2 + \sum_{i=1}^{N} w_i \|\mathbf{y}_i - \bar{\mathbf{y}}\|_2^2 + \hat{w}\|\mathbf{u}\|_2^2 - 2\hat{w}\,\mathrm{trace}(\mathbf{R}\mathbf{C}) \,, \tag{7}$$

where

$$\mathbf{C} = \frac{1}{\hat{w}} \sum_{i=1}^{N} \left[w_i \mathbf{p}_i \mathbf{y}_i^{\mathrm{T}} \right] - \bar{\mathbf{p}} \bar{\mathbf{y}}^{\mathrm{T}} \in \mathbb{R}^{3\times 3} .$$

Obviously this is minimized by $\mathbf{u}^* = \mathbf{0}$ and by searching for an $\mathbf{R}$ that maximizes trace($\mathbf{RC}$). Finding this $\mathbf{R}^*$ is analogous to the un-weighted rigid body transformation problem, see e.g. [2,8,24]. The rotation matrix $\mathbf{R}^*$ is usually found from a singular value decomposition of $\mathbf{C}$, $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, from which we form $\mathbf{R}^* = \mathbf{V}\mathbf{U}^{\mathrm{T}}$. In some rare degenerate cases this may however produce a reflection matrix ($\det(\mathbf{V}\mathbf{U}^{\mathrm{T}}) = -1$), which we correct by reevaluating the matrix as $\mathbf{R}^* = \mathbf{V}\,\mathrm{diag}(1, 1, -1)\,\mathbf{U}^{\mathrm{T}}$. From the calculated $\mathbf{R}^*$ and $\mathbf{u}^* = \mathbf{0}$, in the change of variables expression, we obtain the translation vector $\mathbf{t}^* = \bar{\mathbf{y}} - \mathbf{R}^*\bar{\mathbf{p}}$.

3 Convergence analysis

We define a function

$$\phi(u) = \varrho(\sqrt{u}) ,$$

where $u > 0$. A first order Taylor approximation $\varphi(u)$ of $\phi(u)$ about $u = v$ is $\varphi(u) = \phi(v) + \phi'(v)(u - v)$, which is equivalent to

$$\varphi(u) = \phi(v) - \phi'(v)v + \phi'(v)u.$$

The derivative above is

$$\phi'(v) = \frac{\psi(\sqrt{v})}{2\sqrt{v}}.$$

We have that $\varrho(s) = \phi(s^2)$ and a quadratic approximation to $\varrho(s)$ about $s = r$ is

$$\eta(s, r) = \begin{cases} \varrho(r) - \frac{\psi(r)}{2} r + \frac{\psi(r)}{2r} s^2 & \text{if } r \neq 0, \\ \frac{\psi'(0)}{2} s^2 & \text{if } r = 0, \end{cases} \tag{8}$$

which satisfies $\eta(s, r) = \varrho(r)$ and $\frac{\partial}{\partial s}\eta(s, r) = \psi(r)$ at $s = r$, and $\frac{\partial}{\partial s}\eta(s, r) = w(r)s \;\; \forall s, r \in \mathbb{R}$.

Lemma 1 *For all criterion functions $\varrho \in \mathcal{Q}$ the corresponding quadratic approximation $\eta(s, r)$ fulfills $\eta(s, r) \geq \varrho(s) \;\; \forall s, r \in \mathbb{R}$.*

Proof We introduce a function $\delta(s) = \eta(s, r) - \varrho(s)$ and we will prove that this $\boldsymbol{C}^1$-continuous function $\delta(s)$ satisfies $\delta(s) \geq 0$. Since $\eta(r, r) = \varrho(r)$ is fulfilled it follows that $\delta(r) = 0$ and $\eta(s, r) \geq \varrho(s)$ is true when $s = r$. Now assume that $s, r \geq 0$ and

$s \neq r$. From the mean value theorem and since $\delta(s) - \delta(r) = \delta(s)$ it follows that there must exist a point $q > 0$ between r and s such that

$$\delta(s) = \delta'(q)(s - r)\,,$$

where the derivative $\delta'(q)$ can be written as

$$\delta'(q) = q\big(w(r) - w(q)\big).$$

If $0 \leq r < q < s$ then $(s-r) > 0$ and from the third condition in the definition of $\mathcal{Q}$ it follows that $w(r) - w(q) \geq 0$ and hence, $\delta'(q) \geq 0$. If $0 \leq s < q < r$ then $(s-r) < 0$ and $\delta'(q) \leq 0$ using the same argument as before. Thus, $\delta(s) = \delta'(q)(s-r) \geq 0$ and it is proved that $\eta(s, r) \geq \varrho(s)$ for all $s, r \geq 0$. Both $\eta(s, r)$ and $\varrho(s)$ are even with respect to s. Therefore, we have $\eta(s, r) \geq \varrho(s)$ for all $r \geq 0$ and for all $s \in \mathbb{R}$. Since the derivative ψ is an odd function the function $\eta(s, r)$ is even with respect to r as well. Hence, we also have that $\eta(s, r) \geq \varrho(s)$ holds for all $r \leq 0$ and for all $s \in \mathbb{R}$. It follows that $\eta(s, r) \geq \varrho(s)$ $\forall s, r \in \mathbb{R}$. □

To eventually arrive at a decreasing sequence of objective function values, we need to establish some properties that follow from the quadratic approximation. Let s_i be a function of a transformation $(\mathbf{R}, \mathbf{t})$, defined by $s_i = s_i(\mathbf{R}, \mathbf{t}) = \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{y}_i\|_2$, and let r_i here be a fixed value of s_i given by $r_i = s_i(\mathbf{I}, \mathbf{0})$. It follows from (3) and (8) that the problem

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} \eta(s_i(\mathbf{R}, \mathbf{t}), r_i)\,, \tag{9}$$

is exactly the same as the problem

$$\min_{\mathbf{R},\mathbf{t}} \left\{ \sum_{i=1}^{N} \left(\varrho(r_i) - \frac{\psi(r_i)}{2} r_i \right) + \frac{1}{2} \sum_{i=1}^{N} w(r_i) s_i^2(\mathbf{R}, \mathbf{t}) \right\}.$$

We can easily see that the problem

$$\min_{\mathbf{R},\mathbf{t}} \sum_{i=1}^{N} w(r_i) s_i^2(\mathbf{R}, \mathbf{t})\,, \tag{10}$$

must share the same solution set and have the same properties as the two previous problems since the objective function differs only with a constant factor of one half and a sum which is independent of the transformation $(\mathbf{R}, \mathbf{t})$. The minimization problem in (10) is the same as the minimization problem in (4). There always exists a solution set for these minimization problems since $w \geq 0$.

A rigid body transformation in $\mathbb{R}^3$ can be parameterized by six variables. The optimality conditions are investigated by introducing the parameterization

$$\mathbf{z} = [\omega_1, \omega_2, \omega_3, t_1, t_2, t_3]^{\mathrm{T}},$$

where $\omega_1, \omega_2, \omega_3$ are angles of rotations about the coordinate axes and t_1, t_2, t_3 are translations in x, y, z-directions respectively. The residual in term i, $r_i = \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{y}_i\|_2$, can be written as a function of $\mathbf{z}$, i.e. $r_i = r_i(\mathbf{z})$. Hence, the functions f and g can both be expressed by $\mathbf{z}$ as $f = f(\mathbf{z})$ and $g = g(\mathbf{z})$. We consider the derivative of r_i with respect to z_j, $j = 1, \ldots, 6$, as zero if $r_i = 0$. The optimality condition for f is

$$\nabla f = \sum_{i=1}^{N} \psi(r_i) \nabla r_i = \mathbf{0}, \tag{11}$$

and the optimality condition for g is

$$\nabla g = 2 \sum_{i=1}^{N} w_i r_i \nabla r_i = \mathbf{0}. \tag{12}$$

If the weight function $w(r)$ in (3) is used, these two optimality conditions are equivalent since $w(r_i) r_i = \psi(r_i)$ for all r_i, thus validating our choice of weight function. Also note that after applying (4), the condition $\nabla g = \mathbf{0}$, will be fulfilled.

The derivative of r_i with respect to z_j is undefined when $r_i = 0$. This does not cause any troubles though, since ψ is continuous and $\psi(0) = 0$, and thus the terms of ∇f and ∇g for the zero residuals vanish.

Theorem 1 *The IRLS-ICP algorithm, presented in Algorithm 1, converges to a local extremum of the objective function f in (2) and the sequence of function values f_k over the iterations $k = 1, 2, \ldots$ is monotonically decreasing to the value at the extremum for all criterion functions $\varrho \in \mathcal{Q}$ using the weight function $w(r)$ defined in (3).*

Proof We let the index i goes from 1 to N, i.e. $i = 1, \ldots, N$. The set of model points X and the set of data points $\{\mathbf{p}_i^{(k-1)}\}$ from iteration $k-1$ are given. Compute the set of closest model points, $\mathbf{y}_i^{(k-1)} = \mathcal{C}(\mathbf{p}_i^{(k-1)}, X)$, residuals $r_i^{(k-1)} = \|\mathbf{p}_i^{(k-1)} - \mathbf{y}_i^{(k-1)}\|_2$, and weights, $w_i = w(r_i^{(k-1)})$. The value f_{k-1} of the objective function f for these residuals is

$$f_{k-1} = \sum_{i=1}^{N} \varrho\left(r_i^{(k-1)}\right).$$

Since $\eta\left(r_i^{(k-1)}, r_i^{(k-1)}\right) = \varrho\left(r_i^{(k-1)}\right)$ for all i it follows that

$$\sum_{i=1}^{N} \eta\left(r_i^{(k-1)}, r_i^{(k-1)}\right) = f_{k-1}.$$

If all weights w_i are zero we have reached a local maximum and we will not find any better solution using the IRLS-ICP algorithm, given in Algorithm 1. Then we are terminating the iterations and the current transformation is returned. If any of the weights w_i are greater than zero then the update $(\mathbf{R}^*, \mathbf{t}^*)$ of the rigid body transformation, obtained from (4), is applied to $\mathbf{p}_i^{(k-1)}$ giving the update of the data points, $\mathbf{p}_i^{(k)} = \mathbf{R}^*\mathbf{p}_i^{(k-1)} + \mathbf{t}^*$ and new residuals $s_i^{(k)} = \|\mathbf{p}_i^{(k)} - \mathbf{y}_i^{(k-1)}\|_2$ are computed. From this update we define

$$e_k = \sum_{i=1}^{N} \varrho\left(s_i^{(k)}\right) .$$

The rigid body transformation $(\mathbf{R}^*, \mathbf{t}^*)$ gives the least value of the function g in (5) resulting in

$$\sum_{i=1}^{N} \eta\left(s_i^{(k)}, r_i^{(k-1)}\right) \leq \sum_{i=1}^{N} \eta\left(r_i^{(k-1)}, r_i^{(k-1)}\right) .$$

Accordion to Lemma 1 we have $\eta\left(s_i^{(k)}, r_i^{(k-1)}\right) \geq \varrho\left(s_i^{(k)}\right)$ for all i yielding

$$e_k \leq \sum_{i=1}^{N} \eta\left(s_i^{(k)}, r_i^{(k-1)}\right) ,$$

which gives that $e_k \leq f_{k-1}$.

In the next iteration closest model points $\mathbf{y}_i^{(k)} = \mathcal{C}\left(\mathbf{p}_i^{(k)}, X\right)$, residuals $r_i^{(k)} = \|\mathbf{p}_i^{(k)} - \mathbf{y}_i^{(k)}\|_2$, and new weights, $w_i = w(r_i^{(k)})$ are computed. The value f_k of the objective function f for these residuals is

$$f_k = \sum_{i=1}^{N} \varrho\left(r_i^{(k)}\right) .$$

The points $\mathbf{y}_i^{(k)}$ are the closest model points to $\mathbf{p}_i^{(k)}$ so it is clear that

$$\|\mathbf{p}_i^{(k)} - \mathbf{y}_i^{(k)}\|_2 \leq \|\mathbf{p}_i^{(k)} - \mathbf{y}_i^{(k-1)}\|_2 ,$$

i.e. $r_i^{(k)} \leq s_i^{(k)}$, holds for all i. Since ϱ is a monotonically increasing function for positive residuals we have $f_k \leq e_k$. The equality is satisfied if and only if $\mathbf{y}_i^{(k)} = \mathbf{y}_i^{(k-1)}$ for all i.

To summarize we can conclude that

$$0 \leq f_k \leq e_k \leq f_{k-1},$$

is fulfilled. The lower bound is zero since a sum of non-negative values cannot be negative. It gives that the sequences of f_k and e_k are monotonically decreasing and bounded below. Then they will converge to a real value, which is the same for both sequences. There are just two possibilities for the updating transformation that do not change the values of f_k and e_k to their updates in the subsequent iteration. Either the updating transformation is equal to the identity transformation, $(\mathbf{R}^*, \mathbf{t}^*) = (\mathbf{I}, \mathbf{0})$, and that occurs when the function g has a unique minimum. For this transformation all the residuals are preserved so that $s_i = r_i$ for all i and hence, all the terms of ∇g are unchanged. The other possibility for the updating transformation is that it rotates the data points around a line or a point and that occurs when the function g does not have a unique minimum. For this transformation all the residuals associated with positive weights are preserved. The residuals corresponding to weights that are equal to zero can however change but that does not affect ∇g, since the terms of ∇g corresponding to zero valued weights are zero anyway. Since g is minimized and all terms of ∇f and $\frac{1}{2}\nabla g$ are equal we have that the gradients of f and g satisfy $\nabla f = \frac{1}{2}\nabla g = \mathbf{0}$. Therefore, the optimality condition given by (11) is fulfilled. Suppose that the sequence of f_k will converge to a value that is larger than the value of the objective function f at the local extremum. Then this would contradict the operation of the closest point operator $\mathcal{C}$. We can do the conclusion that the IRLS-ICP algorithm converges monotonically to a local extremum with respect to the value of the objective function f in (2). □

From Theorem 1 we know that the IRLS-ICP algorithm converges to a local extremum for all criterion functions ϱ in $\mathcal{Q}$. There is a possibility that the IRLS-ICP algorithm will return a transformation $(\mathbf{R}, \mathbf{t})$ which gives a maximum of f. It can happen if the used criterion function ϱ has a constant value for inputs of large magnitude and all initial residuals $\|\mathbf{p}_i^{(0)} - \mathbf{y}_i^{(0)}\|_2$ are of significant size. Then the value of ψ is zero for these residuals so that all weights w_i are equal to zero. It results in that the function g is a constant equal to zero, so there is nothing to minimize in (4). Thus, the IRLS-ICP algorithm do not have any possibility to find a transformation that gives a better fit between the point sets in such cases. The value of the criterion function ϱ can not be greater than its value at these residuals and therefore, the objective function f attains a maximum. It can also happen, but it is extremely unlikely, that the IRLS-ICP algorithm will end up in a local maximum or a saddle point after some iterations. We can not exclude that possibility, but it will only happen in some very rare cases. In general, we can expect that the IRLS-ICP algorithm will reach a local minimum. If for example the initial orientation of the data points is turned the wrong way, being upside-down, or having some other apparent misalignment in comparison with the model points we will probably not obtain the global minimum. The same thing can also happen if the shape of the model points has many similar features. However, if the initial rigid body transformation, $(\mathbf{R}^{(0)}, \mathbf{t}^{(0)})$, is carefully chosen we can expect that the reached local extremum is the global minimum.

4 Scaling

So far we have just been discussing criterion functions as they originally are defined. Let say that we have a positive scale parameter, $\sigma_* > 0$, and define a criterion function

Algorithm 2 The IRLS-ICP algorithm, Algorithm 1, using re-scaled residuals

Require: $\ldots, \sigma_0 > \sigma_*$
repeat
$\cdots$
$w_i = w(\|\mathbf{p}_i^{(k-1)} - \mathbf{y}_i^{(k-1)}\|_2 / \sigma_{k-1})$
$\cdots$
Update σ_{k-1} to σ_k
until convergence

using ϱ as

$$\tilde{\varrho}(r) = \varrho\left(\frac{r}{\sigma_*}\right).$$

If the function ϱ is in the set $\mathcal{Q}$ then the function $\tilde{\varrho}$ is also in the set $\mathcal{Q}$. Therefore, the problem

$$\min_{\mathbf{R},\mathbf{t}} \tilde{f}(\mathbf{R}, \mathbf{t}), \tag{13}$$

where

$$\tilde{f}(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N} \varrho\left(\frac{\mathrm{d}(\mathbf{R}\mathbf{p}_i + \mathbf{t}, X)}{\sigma_*}\right), \tag{14}$$

can be solved using the IRLS-ICP algorithm with the weight function $\tilde{w}(r) = w(r/\sigma_*)$ and convergence is ensured according to Theorem 1.

If the value of σ_* is small and ψ is close or equal to zero for large residuals the criterion function $\tilde{\varrho}$ will have a narrow descend at zero. It might result in a slow convergence or convergence to a local extremum far from the global minimum. A way to reduce this problem is to use a quite large value of σ, greater than σ_*, in the first iterations and then gradually decrease its value to the desired value σ_*. An extended version of the IRLS-ICP algorithm is given in Algorithm 2, where a scaling is used and is updated in the iterations. In Algorithm 2 most of the steps are the same as in Algorithm 1 and therefore they are removed to make it shorter.

Lemma 2 *Assume that* $\sigma_0 > \sigma_1 > 0$. *For all criterion functions* $\varrho \in \mathcal{Q}$ *and residuals* $r \in \mathbb{R}$ *we have*

$$\varrho\left(\frac{r}{\sigma_0}\right) \leq \varrho\left(\frac{r}{\sigma_1}\right) \leq \frac{\sigma_0^2}{\sigma_1^2}\varrho\left(\frac{r}{\sigma_0}\right).$$

Proof It is obvious that the two inequalities in the lemma hold with equality if $r = 0$ since $\varrho(0) = 0$. Now let r be positive. Then we have $0 < r/\sigma_0 < r/\sigma_1$ resulting in $\varrho(r/\sigma_0) \leq \varrho(r/\sigma_1)$, which follows from the second condition in the definition of $\mathcal{Q}$. Thus, the first inequality in the lemma holds for positive r. We introduce a function

$$\delta(r) = \frac{\sigma_0^2}{\sigma_1^2}\varrho\left(\frac{r}{\sigma_0}\right) - \varrho\left(\frac{r}{\sigma_1}\right),$$

and we have $\delta(0) = 0$. The derivative of δ, δ', can be written as

$$\delta'(r) = \frac{r}{\sigma_1^2}\left(w\left(\frac{r}{\sigma_0}\right) - w\left(\frac{r}{\sigma_1}\right)\right).$$

The weight function w is monotonically decreasing on $[0, \infty)$ and $0 < r/\sigma_0 < r/\sigma_1$ so $w(r/\sigma_0) - w(r/\sigma_1) \geq 0$. The factor r/σ_1^2 is positive and therefore, the derivative δ' on $(0, \infty)$ is non-negative and thus, δ is monotonically increasing on $[0, \infty)$. Hence, $\delta \geq 0$ on $[0, \infty)$ and we have proved that the second inequality in the lemma holds for all $r \geq 0$. Since the function ϱ is even it follows that the inequality relation in the lemma holds for all $r \in \mathbb{R}$. □

Theorem 2 *Consider the IRLS-ICP algorithm presented in Algorithm 2, where a scaling $\sigma_k \geq \sigma_* > 0$ is used that is updated in the iterations $k = 1, 2, \ldots$, and let $\varrho \in \mathcal{Q}$. The values $\tilde{f}_k$ of the objective function $\tilde{f}$ in (14) from each iteration of Algorithm 2 satisfy*

$$\sum_{i=1}^{N} \varrho\left(\frac{r_i}{\sigma_k}\right) \leq \tilde{f}_k \leq \frac{\sigma_k^2}{\sigma_*^2} \sum_{i=1}^{N} \varrho\left(\frac{r_i}{\sigma_k}\right).$$

If the sequence of σ_k is monotonically decreasing then the upper limit of the above inequality relation is also monotonically decreasing. The IRLS-ICP algorithm converges to a local extremum of $\tilde{f}$ if the sequence of σ_k approaches to σ_*.

Proof It is given that $\sigma_k \geq \sigma_*$. According to Lemma 2 each of the N terms in the given sums, and the sum of $\tilde{f}$, satisfy the inequalities. Hence the same must also hold for the sums. Now assume that the sequence of σ_k is monotonically decreasing, i.e. $\sigma_* \leq \sigma_k \leq \sigma_{k-1}$. From Lemma 2 and Theorem 1 we can deduce,

$$\frac{\sigma_k^2}{\sigma_*^2} \sum_{i=1}^{N} \varrho\left(\frac{r_i^{(k)}}{\sigma_k}\right) \leq \frac{\sigma_k^2}{\sigma_*^2}\frac{\sigma_{k-1}^2}{\sigma_k^2} \sum_{i=1}^{N} \varrho\left(\frac{r_i^{(k)}}{\sigma_{k-1}}\right) =$$
$$\frac{\sigma_{k-1}^2}{\sigma_*^2} \sum_{i=1}^{N} \varrho\left(\frac{r_i^{(k)}}{\sigma_{k-1}}\right) \leq \frac{\sigma_{k-1}^2}{\sigma_*^2} \sum_{i=1}^{N} \varrho\left(\frac{r_i^{(k-1)}}{\sigma_{k-1}}\right).$$

Therefore, the upper bound in the inequality relation given in the theorem is monotonically decreasing if the sequence of σ_k is monotonically decreasing. The function ϱ is continuous which is stated in the first condition in the definition of $\mathcal{Q}$ resulting in

$$\sigma_k \to \sigma_* \;\Rightarrow\; \varrho\left(r/\sigma_k\right) \to \varrho\left(r/\sigma_*\right).$$

Hence, the sequence of $\sum_{i=1}^{N} \varrho\left(r_i/\sigma_k\right)$ over the iterations $k = 1, 2, \ldots$ converges to the objective function $\tilde{f}$ and $\sum_{i=1}^{N} \varrho\left(r_i/\sigma_k\right)$ converges to a local extremum as

a consequence of Theorem 1 since $\varrho_k(r) = \varrho(r/\sigma_k) \in \mathcal{Q}$. Thus, the IRLS-ICP algorithm, Algorithm 2, converges to a local extremum of $\tilde{f}$ if the sequence of σ_k approaches to σ_*. □

The scaling parameter σ_* in (14) should not be chosen to be smaller than the expected standard deviation of the error of the "good" data, which is the data that represents what it is supposed to represent pretty well. After a scaling using an appropriate value of σ_k, the residuals of the "good" data that are entered into the criterion function can be assumed to have approximate the same deviations in each iteration k. We want to get good solutions if the data is free of outliers and normally distributed. To achieve this, the parameters κ of the used criterion functions are chosen in such a way that the asymptotic variance

$$V(\psi, F) = \frac{\int \psi^2 dF}{\left(\int \psi' dF\right)^2},$$

(see e.g. [7,10,13]) is close to 1, given an assumption of normal distribution $\mathcal{N}(0, 1)$ for the errors. The asymptotic variance $V(\psi, F)$ is always greater than 1 for robust criterion functions. By choosing it to be 1.01, we get the following values of the parameters:

$$\begin{aligned} \kappa_{\text{Hu}} &= 2.0138, \\ \kappa_{\text{Ca}} &= 4.3040, \\ \kappa_{\text{Tu}} &= 7.0589. \end{aligned} \tag{15}$$

In the first iterations of the IRLS-ICP algorithm the scaling σ_k should be chosen so that the algorithm gives fast convergence to a minimum and not get stuck at a maximum outside the region of convergence. During the iterations the value of σ_k should approach to σ_* from above. A way of choosing this scaling is to let

$$\sigma_k = \begin{cases} 1.90 \text{ median } \left(\{r_i\}_{i=1}^N\right) & \text{if } k = 0, \\ \xi(\sigma_{k-1} - \sigma_*) + \sigma_* & \text{if } k > 0, \end{cases} \tag{16}$$

where ξ is a parameter satisfying $0 \leq \xi < 1$. The value of ξ should be chosen so it match the expected rate of convergence. If the residuals $\{r_i\}_{i=1}^N$ are positive observations from a normal distribution $\mathcal{N}(0, 1)$, then approximately 80 % of the residuals will be less than σ_0 in the first iteration. Using the median in the estimation of σ_0 provides less vulnerability to outliers than if the scaling were based on the actual upper residuals. We can also be sure that not all weights w_i are equal to zero in the first iteration, so that the IRLS-ICP algorithm will start in a region of convergence. Since the median is used only in the first iteration time for computing it is negligible.

5 Numerical comparisons

Here we conduct numerical experiments in order to test the performance of the IRLS-ICP algorithm, presented in Algorithm 2, when solving surface registration problems

given by (13), where X is a set of surface points and most of the data points $\mathbf{p}_i$ represent the shape of the surface. The set of data points is also containing some errors so the capability of the robust criterion functions can be examined. A least squares solution, i.e. where we have $\varrho(r) = r^2/2$, is also computed for comparison.

Six different smooth surfaces are used in the experiment, which are shown in Fig. 2. The first four surfaces might seem to have sharp edges and corners but they are actually rounded. The last two surfaces are kinematic surfaces [19] which are invariant under some rigid body transformations so no unique solution of (13) does exists for these surfaces. From each of these six surfaces we have sampled 100000 points. A local quadratic surface approximation is acquired at each of these points to represent the surface as the set of model points X. We are also sampling a set of $N_\mathrm{g} = 1000$ ("good") non disturbed data points $\{\mathbf{p}_i\}_{i=1}^{N_\mathrm{g}}$, and $N_\mathrm{b} = 500$ ("bad") disturbed data points $\{\mathbf{p}_i\}_{i=N_\mathrm{g}+1}^{N_\mathrm{g}+N_\mathrm{b}}$ from the surfaces, and all these $N = N_\mathrm{g} + N_\mathrm{b}$ data points are transformed by a rigid body transformation. The disturbed data points are generated randomly from a normal distribution in the 3D-space. The problem is then to estimate a rigid body transformation for fitting the set of data points $\{\mathbf{p}_i\}_{i=1}^{N}$ to the set of points X by solving (13). The sets of data points are also shown in Fig. 2.

The search for the closest point in X to an arbitrary data point, written as the operator $\mathcal{C}$ in Algorithm 1, is done by first finding the closest point among all the 100000 sampled surface points in X. The corresponding local quadratic surface approximation is then used to find an even better approximation of the closest point on the surface.

The aim of the experiment is to test the performance of the criterion functions and make a comparison of them. Hence, we need a comparable measure of efficiency. In iteration k we are using

$$\ell_k = \frac{1}{N_\mathrm{g}} \sum_{i=1}^{N_\mathrm{g}} r_i^2 \,, \tag{17}$$

as a measure of efficiency, which is the mean squared residual between the non-disturbed data points and the surface. We have that $\ell_k \geq 0$ and ℓ_k should eventually approach to a value close to zero. It is not exactly zero since the search for the closest surface point is just approximate by the usage of the 100000 sampled surface points and the corresponding local quadratic surface approximation. The values of the parameters κ of the robust criterion functions given in (15) are used making the criterion functions comparative. The scaling parameter σ_k defined in (16) is used to scale the residuals. It is still arbitrary, but by some thought we let $\sigma_* = a/1000$, where $a = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ is the diagonal of the distribution of the surface points in x, y, z-directions. The value of the parameter ξ for surface 3 and 4 is set to 0.95 and for the other surfaces the value of ξ is set to 0.85. The result from our experiment is shown in Fig. 3.

From Fig. 3 it is clear that the least square solution is inferior to the others. The large residuals of the "bad" data influence the surface registration to such an extent that is simply fails. The robust methods on the other hand, manage to find much better solutions, measured by ℓ_k from (17). Tukey's bi-weight function seems to give the best

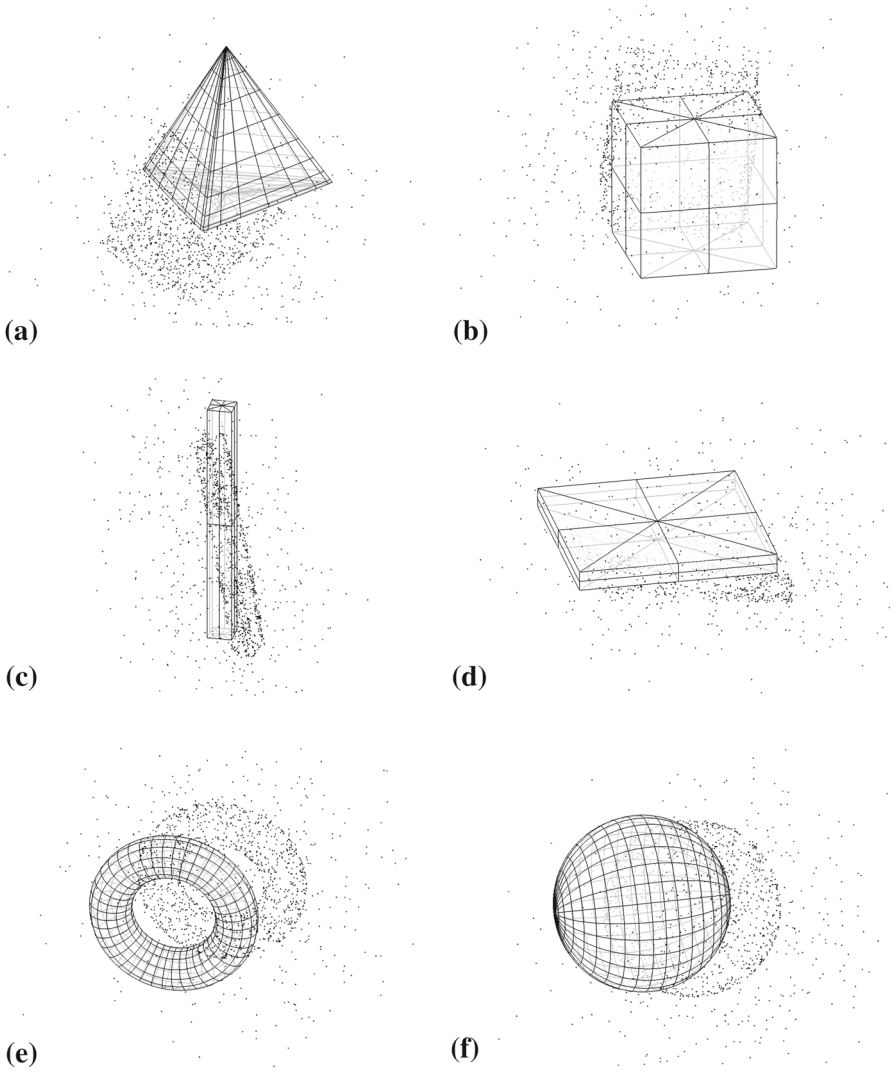

Fig. 2 Surfaces and data points, **a** rounded tetrahedron, **b** rounded cube, **c** rounded cuboid, **d** rounded cuboid, **e** torus, **f** sphere

results in this tests. Second best is Cauchy's function, then comes Huber's function, and finally the quadratic function. From Fig. 1 we can see that this order coincides with their ability to reduce strong influence of the outliers.

We can also observe from Fig. 3 that in the registration using surface 3 and 4, the two rounded cuboids, we have a very slow convergence compared to the other surfaces. Surface 3 is very drawn-out in one direction which gives a slow convergence in that direction. Surface 4 is much flattened; almost like a plane, having the same difficulties of convergence as points on a plane. The other surfaces have more distinct shapes

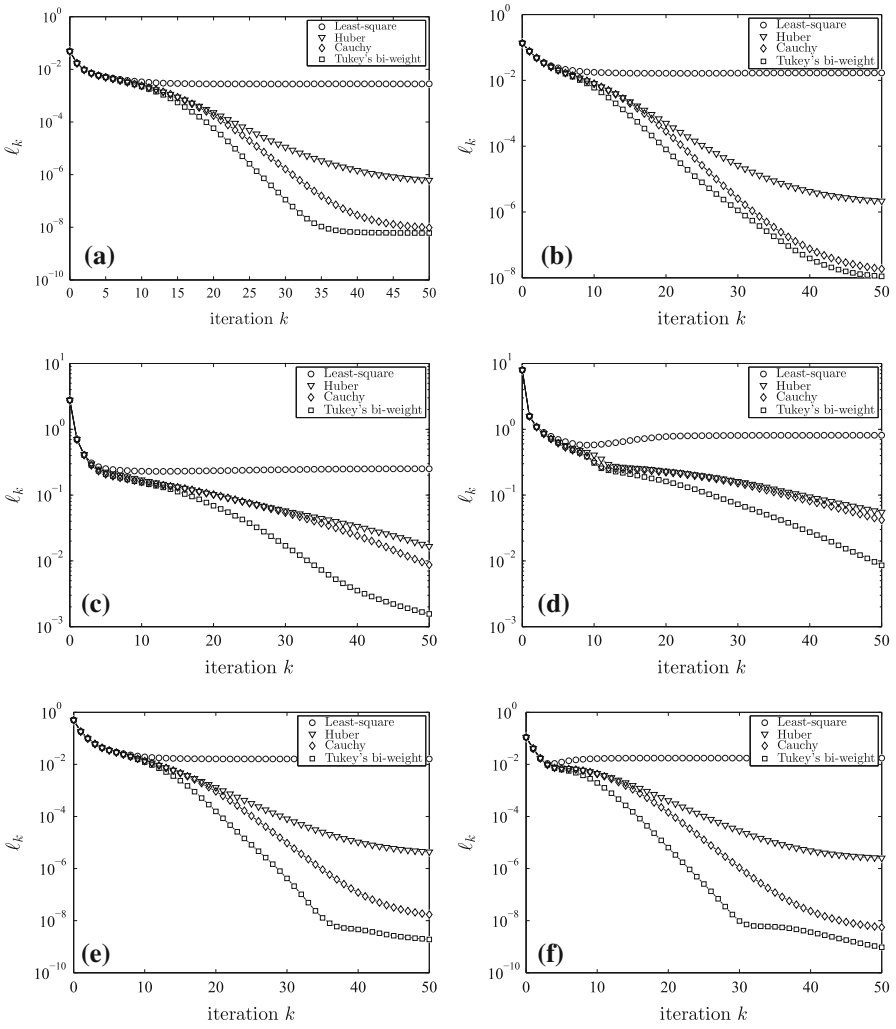

Fig. 3 Observed values of ℓ_k in iteration k given by (17) for the **a** rounded tetrahedron, **b** rounded cube, **c** rounded cuboid, **d** rounded cuboid, **e** torus, **f** sphere

giving faster convergence. We do not have any unique solution of (13) for the last two surfaces, the kinematic surfaces, since these surfaces are invariant under some rigid body transformation. However, in comparison we get quite fast convergence of ℓ_k for these two surfaces.

6 Discussion

The non-convex criterion functions are the clear winners in our experiments. However, in traditional regression problems, those criterion functions are known to introduce

local minima, causing all kinds of troubles. None of that could be seen in our experiments here, but it would be advisable to be cautious. A remedy to this problem could be to start by using Huber's criterion function for some iterations and then change to a more robust criterion function.

The sequence of scaling parameters σ_k, defined in (16), is exponentially decreasing to σ_*. The values of σ might be chosen in many different ways but exponential decreasing seems to be a good choice from observations of non-published experiments. In our experiment we have set the exponential base ξ to 0.85 and 0.95 dependent of the shape of the surfaces which seems to be appropriate values. If the value of σ_* is not known at all some method of estimating it is necessary to be added to the algorithm.

References

1. Andreani, R., Martínez, J.M., Martínez, L., Yano, F.: Continuous optimization methods for structure alignments. Math. Progr. **112**(1), 93–124 (2008)
2. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3-D point sets. IEEE Trans. Pattern Anal. Mach. Intell. **9**(5), 698–700 (1987)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
4. Bispo, E.M., Fisher, R.B.: Free-form surface matching for surface inspection. In: Proceedings of the 6th IMA Conference on the Mathematics of Surfaces, pp. 119–136. Clarendon Press, New York, NY, USA (1996)
5. Fitzgibbon, A.W.: Robust registration of 2D and 3D point sets. Image Vis. Comput. **21**(13–14), 1145–1153 (2003)
6. Godin, G., Rioux, M., Baribeau, R.: Three-dimensional registration using range and intensity information. In: El-Hakim, S.F. (ed.) Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, vol. 2350, pp. 279–290 (1994)
7. Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust statistics. Wiley, New York (1986)
8. Hanson, R.J., Norris, M.J.: Analysis of measurements based on the singular value decomposition. SIAM J. Sci. Stat. Comput. **2**(3), 363–373 (1981)
9. Holland, P.W., Welsch, R.E.: Robust regression using iteratively reweighted least-squares. Commun. Stat. Theory Methods **6**(9), 813–827 (1977)
10. Huber, P.J.: Robust statistics. Wiley, New York (1981)
11. Jain, A., Hong, L., Bolle, R.: On-line fingerprint verification. IEEE Trans. Pattern Anal. Mach. Intell. **19**(4), 302–314 (1997)
12. Li, Z., Xu, Z., Cen, M., Ding, X.: Robust surface matching for automated detection of local deformations using least-median-of-squares estimator. Photogramm. Eng. Remote Sens. **67**(11), 1283–1292 (2001)
13. Maronna, R.A., Martin, D.R., Yohai, V.J.: Robust statistics: theory and methods. Wiley, New York (2006)
14. Martínez, L., Andreani, R., Martínez, J.: Convergent algorithms for protein structural alignment. BMC Bioinf. **8**(1), 306 (2007)
15. Masuda, T., Sakaue, K., Yokoya, N.: Registration and integration of multiple range images for 3-D model construction. In: ICPR '96 Proceedings of the 1996 International Conference on Pattern Recognition (ICPR '96) Vol. I, pp. 879–883. IEEE Computer Society, Washington, DC, USA (1996)
16. Masuda, T., Yokoya, N.: A robust method for registration and segmentation of multiple range images. CAD-Based Vision Workshop, 1994, Proceedings of the 1994 Second pp. 106–113 (1994)
17. Maurer Jr, C.R., Aboutanos, G., Dawant, B., Maciunas, R., Fitzpatrick, J.: Registration of 3-D images using weighted geometrical features. IEEE Trans. Med. Imaging **15**(6), 836–849 (1996)
18. Phillips, J.M., Liu, R., Tomasi, C.: Outlier robust ICP for minimizing fractional RMSD. In: IEEE International Conference on 3-D Digital Imaging and Modeling, 2007. 3DIM '07., pp. 427–434 (2007)
19. Pottmann, H., Wallner, J.: Computational line geometry. Springer, Berlin (2010)

20. Pulli, K.: Multiview registration for large data sets. In: IEEE international conference on 3-D digital imaging and modeling, 1999, pp. 160–168 (1999)
21. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proceedings of the Third international conference on 3D digital imaging and modeling, pp. 145–152 (2001)
22. Salvi, J., Matabosch, C., Fofi, D., Forest, J.: A review of recent range image registration methods with accuracy evaluation. Image Vis. Comput. **25**(5), 578–596 (2007)
23. Santamaría, J., Cordón, O., Damas, S.: A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. Comput. Vis. Image Underst. **115**(9), 1340–1354 (2011)
24. Söderkvist, I.: Perturbation analysis of the orthogonal procrustes problem. BIT **33**, 687–694 (1993)
25. Stewart, C., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Trans. Med. Imaging **22**(11), 1379–1394 (2003)
26. Trucco, E., Fusiello, A., Roberto, V.: Robust motion and correspondence of noisy 3-D point sets with missing data. Pattern Recognit. Lett. **20**(9), 889–898 (1999)
27. Turk, G., Levoy, M.: Zippered polygon meshes from range images. In: SIGGRAPH '94: Proceedings of the 21st annual conference on computer graphics and interactive techniques, pp. 311–318. ACM Press, New York, NY, USA (1994)
28. Wolke, R.: Iteratively reweighted least squares: a comparison of several single step algorithms for linear models. BIT **32**(3), 506–524 (1992)
29. Wolke, R., Schwetlick, H.: Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons. SIAM J. Sci. Stat. Comput. **9**(5), 907–921 (1988)
30. Zhang, Z.: Iterative point matching for registration of free-form curves. Tech. Rep. RR-1658, INRIA Sophia-Antipolis, France (1992)
31. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vis. **13**(2), 119–152 (1994)
32. Zhu, L., Barhak, J., Srivatsan, V., Katz, R.: Efficient registration for precision inspection of free-form surfaces. Int. J. Adv. Manuf. Technol. **32**(5–6), 505–515 (2006)
33. Zinßer, T., Schmidt, J., Niemann, H.: A refined ICP algorithm for robust 3-D correspondence estimation. In: Proceedings of the international conference on image processing, vol. 2, pp. II- 695–698 vol. 3. Barcelona, Spain (2003)