Robust moving total least squares: A technique for the reconstruction of measurement data in the presence of multiple outliers

Tianqi Gu[a], Hongxin Lin[a], Dawei Tang[b], Shuwen Lin[a] and Tianzhi Luo[c, d,*]

[a]School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

[b]Centre for Precision Technologies, University of Huddersfield, Huddersfield HD1 3DH, UK

[c]CAS Key Laboratory of Mechanical Behaviour and Design of Materials, Department of Modern Mechanics, University of Science and Technology of China, Hefei 230022, China

[d]Taizhou Luohua Biotechnology Ltd., Taizhou, 225306, China

Abstract

This article is concerned with the reconstruction of contaminated measurement data based on the moving total least squares (MTLS) method, which is extensively applied to many engineering and scientific fields. Traditional MTLS method is lack of robustness and sensitive to the outliers in measurement data. Based on the framework of MTLS method, we proposed a robust MTLS method called RMTLS method by introducing a two-step pre-process to detect and remove the anomalous nodes in the support domain. The first step is an iterative regression procedure that combines with k-medoids clustering to automatically reduce the weight of anomalous node for a regression-based reference (curve or surface). Based on the distances between reference and discrete points, the second step adopts a density function defined by a sorted distance sequence to select the normal points without setting a threshold artificially. After the two-step pre-process, weighted total least square is performed on the selected point set to obtain the estimation value. By disposing of the anomalous nodes in each independent support domain, multiple outliers can be suppressed within the whole domain. Furthermore, the suppression of multiple continual outliers is possible by adopting asymmetric support domain and introducing previous estimation points. The proposed method shows great robustness and accuracy in reconstructing the simulation and experiment data.

Corresponding author: tzluo@ustc.edu.cn

Keywords: Measurement data; Moving least squares; K-medoids clustering; Outlier

1. Introduction

Curve or surface reconstruction for discrete data is a key issue in many fields such as reverse engineering, measurement technique and computer vision [1-3]. Currently, reconstruction method can be generally categorized into explicit and implicit reconstruction. For explicit reconstruction, the reconstructed curve or surface can be characterized by an explicit function such as Bezier [4], B-Spline [5] and Non-uniform Rational B-Spline (NURBS) [6]. When the shape parameters of the function are determined, the reconstructed result can be obtained. However, it is not an easy task to obtain the optimal parameters. Many optimization methods have been incorporated to obtain the shape parameters, such as invasive weed [7], genetic algorithm [8], simulated annealing [9], particle swarm [10]. Implicit reconstruction methods define the reconstructed curve or surface as a zero isocontour of a scalar function. Moving least squares (MLS) [11], radial basis function (RBF) [12], signed distance function [13] and level-set [14] are widely used implicit reconstruction methods. These approaches approximate the implicit surface with different criterions to minimize the cost which represent different distance functions [15]. Compared with the explicit reconstruction method, the reconstructed result by implicit method can provide better topology information [16].

Among these reconstruction methods, MLS is a widely used reconstruction methods owing to its simplicity and computation efficiency. MLS is a point-wise estimation method proposed by Shepard [17], and Lancaster [18] promoted the MLS method to surface reconstruction. At each point, MLS gives the localized approximation by the weighted least square (WLS) with a compact weight function. MLS is firstly designed for reconstruction. After years of development, the characterization of localized approximation attracts many scholars to apply this method to meshless method [19-23]. In meshless method, a set of nodal points instead of meshes are used for the discretization of the field variables, which overcomes the shortcoming of triangulation and mesh refinement problem in finite element method (FEM) [24]. Belytschko [19] constructed the discrete weak form of the equilibrium equation by combining with MLS approximation and firstly proposed the element-free Galerkin (EFG) method. Based on Belytschko' work, many meshless methods such as improved element-free galerkin (IEFG) [20], meshless local

Petrov-Galerkin (MLPG) [21], complex variable element-free Galerkin (CVEFG) [22] and improved complex variable element-free Galerkin (ICVEFG) [23] method have been established.

Although there are great applications of MLS method in meshless methods, MLS method encounters some problems in the reconstruction of measurement data. In measurement field, random errors are inevitable and occur in all variables while the MLS method only considers the influence of the dependent variables with random errors. To cope with this shortcoming of MLS, Scitovski [25] proposed the MTLS method that uses weight total least square (WTLS) for the localized approximation. Furthermore, MLS and MTLS method are both non-robust reconstruction methods, which means that even a single outlier could greatly distort the reconstruction result. To suppress the outliers, one method is to assign a small weight to outlier to weaken its influence, such as redescending M-estimator [26] and robust Bayesian regression [27, 28]. The definition of the weight function for down-weighting outliers without affecting the weight of normal points is the key. An alternative way is to reject the outliers according to a certain criterion such as support vector machine (SVM) [29], least trimmed square (LTS) [30] and random sample consensus (RANSAC) [31] method. The determination of the pre-defined criterion is a challenging work for these methods.

In this article, a robust MTLS (RMTLS) method is proposed to improve the robustness of MTLS method to outliers without artificially setting threshold. The proposed RMTLS method adopts a two-step pre-process to select the normal points in the support domain. Firstly, a clustering-based iterative regression procedure is applied to obtain a regression-based reference within the support domain. Secondly, a density function based on the distribution of the discrete points is applied to select the normal points. Finally, WTLS is performed on the selected normal points with recalculated weights to obtain the estimation value. Through the validation of curve and surface reconstruction, the RMTLS method exhibits better robustness than the MLS and MTLS method. The outlines of this article are listed as follows. In section 2, the background knowledge of the MLS and MTLS method are introduced. In section 3, k-medoids clustering and the RMTLS method are illustrated in detail. In section 4, curve and surface reconstruction cases are used to investigate the effectiveness of the RMTLS method. In section 5, the robustness of RMTLS to different types of outliers is further discussed. In section 6, a brief conclusion on whole work is given.

2. Introduction to the fundamental theory

2.1 MLS method

In the meshless method, the representation of the MLS method can be divided into non-shifted polynomial basis scheme and shifted polynomial basis scheme. The latter one is more commonly used because it is more robust for moment matrix [32, 33]. In this paper, the MLS method with shifted polynomial basis scheme is adopted and will be introduced in this section. Suppose that $f(\mathbf{x})$ is a function defined in space $\mathbb{R}^d$. In the shifted polynomial basis scheme of MLS, the approximation $f^h(\mathbf{x},\mathbf{x}_c)$ of $f(\mathbf{x})$ is [34]

$$f(\mathbf{x}) \approx f^h(\mathbf{x},\mathbf{x}_c) = \sum_{i=1}^{m} p_i(\mathbf{x}-\mathbf{x}_c)\alpha_i(\mathbf{x}_c) = \mathbf{p}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_c)\boldsymbol{\alpha}(\mathbf{x}_c) \tag{1}$$

where $\mathbf{x}_c$ is the centre of support domain Ω_c in which the localized approximation is performed, $\boldsymbol{\alpha}(\mathbf{x}_c)$ is local coefficient vector to be determined, and $\mathbf{p}(\mathbf{x}–\mathbf{x}_c)$ is the shifted basis with the expression

$$\mathbf{p}(\mathbf{x}-\mathbf{x}_c) = \begin{cases} [1 \;\; x-x_c]^{\mathrm{T}} & ,m=1\; d=1 \\ [1 \;\; x-x_c \;\; y-y_c]^{\mathrm{T}} & ,m=1\; d=2 \end{cases} \tag{2}$$

Eq.(2) shows the linear shifted polynomial basis in different dimensions. The $\boldsymbol{\alpha}(\mathbf{x}_c)$ is obtained by minimizing the objective function

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \sum_{i=1}^{n} w(\mathbf{x}_i-\mathbf{x}_c)[f(\mathbf{x}_i)-\mathbf{p}^{\mathrm{T}}(\mathbf{x}_i-\mathbf{x}_c)\boldsymbol{\alpha}(\mathbf{x}_c)]^2 \\ &= [\mathbf{f}_{\Omega_c}-\mathbf{P}_{\Omega_c}{}^{\mathrm{T}}\boldsymbol{\alpha}(\mathbf{x}_c)]^{\mathrm{T}}\mathbf{W}_{\Omega_c}[\mathbf{f}_{\Omega_c}-\mathbf{P}_{\Omega_c}{}^{\mathrm{T}}\boldsymbol{\alpha}(\mathbf{x}_c)] \end{aligned} \tag{3}$$

with

$$\begin{cases} \mathbf{f}_{\Omega_c} = (f(\mathbf{x}_1) \;\; f(\mathbf{x}_2) \;\; \cdots \;\; f(\mathbf{x}_n))^{\mathrm{T}} \\ \mathbf{P}_{\Omega_c} = (\mathbf{p}(\mathbf{x}_1-\mathbf{x}_c) \;\; \mathbf{p}(\mathbf{x}_2-\mathbf{x}_c) \;\; \cdots \;\; \mathbf{p}(\mathbf{x}_n-\mathbf{x}_c))^{\mathrm{T}} \\ \mathbf{W}_{\Omega_c} = diag(w(\mathbf{x}_1-\mathbf{x}_c) \;\; w(\mathbf{x}_2-\mathbf{x}_c) \;\; \cdots \;\; w(\mathbf{x}_n-\mathbf{x}_c)) \end{cases} \tag{4}$$

where n is the total number of discrete points in Ω_c and $w(\mathbf{x}_i-\mathbf{x}_c)$ is a compact weight function. In this work, exponential weight function is used with the definition

$$w(\mathbf{x}_i-\mathbf{x}_c) = w(s) = \begin{cases} \dfrac{e^{-\theta s^2}-e^{-\theta}}{1-e^{-\theta}} & ,s\le 1 \\ 0 & ,s>1 \end{cases} \quad ,s = \frac{\|\mathbf{x}_i-\mathbf{x}_c\|}{\delta} \tag{5}$$

where δ is the radius of Ω_c and θ is a user-specified parameter that influences the distribution range of weight. θ is fixed to 16 in this article. Fig.1 shows the weight

function and its distribution when the centre of support domain moves from $\mathbf{x}_{i-2}$ to $\mathbf{x}_{i+2}$.

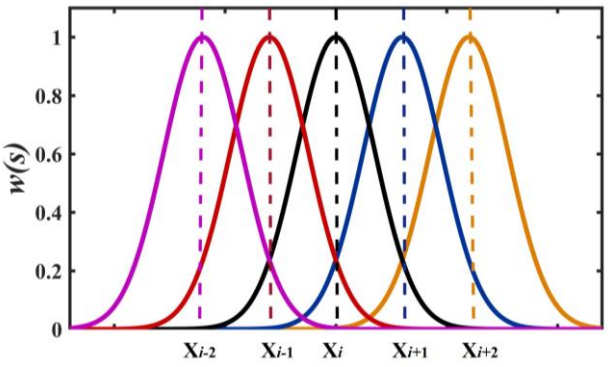


Fig. 1. The distribution of weight at different estimation points

The minimum of $J(\boldsymbol{\alpha})$ can be calculated by letting the partial differential of $J(\boldsymbol{\alpha})$ to $\boldsymbol{\alpha}$ be 0

$$\frac{\partial J}{\partial \boldsymbol{\alpha}} = \mathbf{M}_{\Omega_c}^{-1}\boldsymbol{\alpha}(\mathbf{x}_c) - \mathbf{P}_{\Omega_c}^{\mathrm{T}}\mathbf{W}_{\Omega_c}\mathbf{f}_{\Omega_c} = 0 \tag{6}$$

Then, $\boldsymbol{\alpha}(\mathbf{x}_c)$ can be obtained by

$$\boldsymbol{\alpha}(\mathbf{x}_c) = \mathbf{M}_{\Omega_c}^{-1}\mathbf{P}_{\Omega_c}^{\mathrm{T}}\mathbf{W}_{\Omega_c}\mathbf{f}_{\Omega_c} \tag{7}$$

with

$$\mathbf{M}_{\Omega_c} = \mathbf{P}_{\Omega_c}^{\mathrm{T}}\mathbf{W}_{\Omega_c}\mathbf{P}_{\Omega_c} \tag{8}$$

where $\mathbf{M}_{\Omega c}$ is the moment matrix. The approximation of $f(\mathbf{x})$ can be calculated by substituting Eq.(7-8) into Eq.(1)

$$f^h(\mathbf{x},\mathbf{x}_c) = \mathbf{p}^{\mathrm{T}}(\mathbf{x}-\mathbf{x}_c)\mathbf{M}_{\Omega_c}^{-1}\mathbf{P}_{\Omega_c}^{\mathrm{T}}\mathbf{W}_{\Omega_c}\mathbf{f}_{\Omega_c} \tag{9}$$

The above expression gives MLS approximation of $f(\mathbf{x})$ when the centre of support domain is located at $\mathbf{x}_c$. When applying the MLS method to reconstructing measurement data, the estimation point $\mathbf{x}$ is chosen as the centre of the support domain, that is, $\mathbf{x} = \mathbf{x}_c$. Then the Eq.(9) can be further rewritten as

$$f^h(\mathbf{x}_c) = \mathbf{p}^{\mathrm{T}}(\mathbf{0})\mathbf{M}_{\Omega_c}^{-1}\mathbf{P}_{\Omega_c}^{\mathrm{T}}\mathbf{W}_{\Omega_c}\mathbf{f}_{\Omega_c} \tag{10}$$

where

$$\mathbf{p}(\mathbf{0}) = \begin{cases} [1\ \ 0]^{\mathrm{T}} & ,m=1\ d=1 \\ [1\ \ 0\ \ 0]^{\mathrm{T}} & ,m=1\ d=2 \end{cases} \tag{11}$$

2.2 MTLS method

The representation of MTLS method with shifted polynomial basis scheme will be introduced in this section. In 1981, Lancaster formally proposed the MLS method and solved the problem that ordinary least square (OLS) is not feasible to fit the data with complex shape. MLS adopts WLS method in the support domain to obtain the estimation value. Then, Scitovski proposed the MTLS method by replacing WLS estimation with WTLS estimation. MTLS can be seen as an evolution of MLS, which further considers both dependent and independent variables.

WTLS method is established based on the errors in variable (EIV) model [35] which can be formulated as

$$(\mathbf{P}_{\Omega_c}+\mathbf{E}_{\mathbf{x}})^{\mathrm{T}}\boldsymbol{\alpha}(\mathbf{x}_c)=\mathbf{f}_{\Omega_c}+\mathbf{e}_{\mathbf{f}} \tag{12}$$

where $\mathbf{E}_{\mathbf{x}}$ is an unknown error matrix of $\mathbf{P}_{\Omega c}$ and $\mathbf{e}_{\mathbf{f}}$ is an unknown error vector of $\mathbf{f}_{\Omega c}$. The dimension of $\mathbf{P}_{\Omega c}$ (or $\mathbf{E}_{\mathbf{x}}$) is $n\times(m+1)$ and the dimension of $\mathbf{f}_{\Omega c}$ (or $\mathbf{e}_{\mathbf{f}}$) is $n\times 1$. In WLTS, to obtain the estimation of $\boldsymbol{\alpha}(\mathbf{x}_c)$ is to minimize the function

$$\underset{\mathbf{E}_{\mathbf{x}},\mathbf{e}_{\mathbf{f}}}{\text{minimize}}\,\|\mathbf{D}[\mathbf{E}_{\mathbf{x}}\,|\,\mathbf{e}_{\mathbf{f}}]\mathbf{T}\|_F^2=\underset{\mathbf{E}_{\mathbf{x}},\mathbf{e}_{\mathbf{f}}}{\text{minimize}}\left\|\begin{array}{ccc|c}d_1e_{\mathbf{X}11}t_1 & \cdots & d_1e_{\mathbf{X}1(m+1)}t_{m+1} & d_1e_{\mathbf{f}1}t_{m+2}\\ \cdots & \cdots & \cdots & \cdots\\ d_ne_{\mathbf{X}n1}t_1 & \cdots & d_ne_{\mathbf{X}n(m+1)}t_{m+1} & d_ne_{\mathbf{f}n}t_{m+2}\end{array}\right\|_F^2, \text{ s.t. Eq.(12)} \tag{13}$$

where $\|\cdot\|_{\mathrm{F}}^2$ is the Frobenius norm defined as $\|\mathbf{A}_{m\times n}\|_{\mathrm{F}}^2=\Sigma_{i=1}^{m}\Sigma_{j=1}^{n}|a_{ij}|^2$, and $\mathbf{D}$ (or $\mathbf{T}$) is a diagonal weight matrix with dimension of n (or $m+2$). MTLS adopts a simplified WTLS estimation in which the weight matrix $\mathbf{D}$ is set as compact weight matrix $\mathbf{W}_{\Omega c}$ and $\mathbf{T}$ is assumed to be $\mathbf{I}_{m+2}$, which means that the error terms on each row of augmented matrix $\mathbf{D}[\mathbf{E}_{\mathbf{x}}\,|\,\mathbf{e}_{\mathbf{f}}]$ have the same weight.

By left multiplying the matrix $\mathbf{D}$ on both side of Eq.(12), Eq.(12) can be rewritten as

$$\left\{\mathbf{D}[\mathbf{P}_{\Omega_c}\,|\,\mathbf{f}_{\Omega_c}]+\mathbf{D}[\mathbf{E}_{\mathbf{x}}\,|\,\mathbf{e}_{\mathbf{f}}]\right\}\begin{bmatrix}\boldsymbol{\alpha}(\mathbf{x}_c)\\ -1\end{bmatrix}=\mathbf{0}_{n\times 1} \tag{14}$$

Then, the minimization of Eq.(13) is equivalent to find a matrix $\mathbf{D}[\mathbf{E}_{\mathbf{x}}\,|\,\mathbf{e}_{\mathbf{f}}]$ with the lowest Frobenius norm that reduces $\mathbf{D}[\mathbf{P}_{\Omega c}\,|\,\mathbf{f}_{\Omega c}]$ with rank $m+2$ to $\{\mathbf{D}[\mathbf{P}_{\Omega c}\,|\,\mathbf{f}_{\Omega c}]+\mathbf{D}[\mathbf{E}_{\mathbf{x}}\,|\,\mathbf{e}_{\mathbf{f}}]\}$ with rank $m+1$, which leads to a best rank-$m+1$ approximation problem. Let the singular decomposition of the augmented matrix $\mathbf{G}:=\mathbf{D}[\mathbf{P}_{\Omega c}\,|\,\mathbf{f}_{\Omega c}]$ be

$$\mathbf{G}:=\mathbf{D}[\mathbf{P}_{\Omega_c}\,|\,\mathbf{f}_{\Omega_c}]=[\mathbf{U}_{n\times(m+1)}\quad\mathbf{u}_{n\times 1}]\begin{bmatrix}\boldsymbol{\Sigma}_{m+1} & \mathbf{0}_{(m+1)\times 1}\\ \mathbf{0}_{1\times(m+1)} & \sigma_{m+2}\end{bmatrix}\left[\begin{array}{c|c}\mathbf{V}_{m+1} & \mathbf{v}_{(m+1)\times 1}\\ \hline \mathbf{v}'_{1\times(m+1)} & v\end{array}\right] \tag{15}$$

According to the low-rank approximation theory, the best rank-m+1 approximation of $\mathbf{D}[\mathbf{P}_{\Omega c} \mid \mathbf{f}_{\Omega c}]$, i.e. $\{\mathbf{D}[\mathbf{P}_{\Omega c} \mid \mathbf{f}_{\Omega c}]+\mathbf{D}[\mathbf{E}_{\mathbf{x}} \mid \mathbf{e}_{\mathbf{f}}]\}$, can be obtained by dropping the singular value σ_{m+2} of $\mathbf{G}$, and the minimization is unique when $\sigma_{m+1} \neq \sigma_{m+2}$. Then $\{\mathbf{D}[\mathbf{P}_{\Omega c} \mid \mathbf{f}_{\Omega c}]+\mathbf{D}[\mathbf{E}_{\mathbf{x}} \mid \mathbf{e}_{\mathbf{f}}]\}$ can be formulated as

$$\{\mathbf{D}[\mathbf{P}_{\Omega_c} \mid \mathbf{f}_{\Omega_c}]+\mathbf{D}[\mathbf{E}_{\mathbf{x}} \mid \mathbf{e}_{\mathbf{f}}]\}=[\mathbf{U}_{n\times(m+1)} \quad \mathbf{u}_{n\times 1}]\begin{bmatrix}\boldsymbol{\Sigma}_{m+1} & \mathbf{0}_{(m+1)\times 1}\\ \mathbf{0}_{1\times(m+1)} & 0\end{bmatrix}\left[\begin{array}{c|c}\mathbf{V}_{m+1} & \mathbf{v}_{(m+1)\times 1}\\ \hline \mathbf{v}'_{1\times(m+1)} & v\end{array}\right] \tag{16}$$

Based on the above, $\mathbf{D}[\mathbf{E}_{\mathbf{x}} \mid \mathbf{e}_{\mathbf{f}}]$ can be obtained by

$$\mathbf{D}[\mathbf{E}_{\mathbf{x}} \mid \mathbf{e}_{\mathbf{f}}]=\{\mathbf{D}[\mathbf{P}_{\Omega_c} \mid \mathbf{f}_{\Omega_c}]+\mathbf{D}[\mathbf{E}_{\mathbf{x}} \mid \mathbf{e}_{\mathbf{f}}]\}-\mathbf{G}=-\mathbf{D}[\mathbf{P}_{\Omega_c} \mid \mathbf{f}_{\Omega_c}]\begin{bmatrix}\mathbf{v}_{(m+1)\times 1}\\ v\end{bmatrix}\begin{bmatrix}\mathbf{v}_{(m+1)\times 1}\\ v\end{bmatrix}^{\mathrm{T}} \tag{17}$$

Combined with Eq.(14) and Eq.(17), the solution of $\boldsymbol{\alpha}(\mathbf{x}_c)$ can be expressed as

$$\boldsymbol{\alpha}(\mathbf{x}_c)=-\mathbf{v}_{(m+1)\times 1}/v \tag{18}$$

3. Proposed RMTLS method

3.1 K-medoids clustering method

In the RMTLS method, k-medoids clustering is introduced to construct the weight function based on the geometry feature of dataset itself. Hence, before illustrating the proposed RMTLS method, it is needed to give a brief introduction to k-medoids clustering. K-medoids clustering is an improved k-means clustering method with higher robustness to outliers [36]. For a given point set $\mathbf{X}=\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ in $\mathbb{R}^d$, the objective of k-medoids clustering is to find k subsets $\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_k$ subject to $\cup_{i=1}^{k}\mathbf{C}_i=\mathbf{X}$ and $\mathbf{C}_i \cap \mathbf{C}_{i'}=\phi$, $i \neq i'$ and satisfy the minimum cost defined by [37]

$$E=\sum_{i=1}^{k}\sum_{\mathbf{x}_j\in\mathbf{C}_i}\left\|\mathbf{x}_j-\mathbf{p}_i\right\|^2 \tag{19}$$

where $\mathbf{p}_i$ is the medoid of the cluster $\mathbf{C}_i$ which can be defined as the most centrally located point in cluster $\mathbf{C}_i$, and $\|\cdot\|$ denotes the Euclidian norm. To obtain the global minima of Eq.(19) is a time consuming work and it is considered to be as a NP-hard problem. The most popular realization of k-medoids clustering is the partition around medoids (PAM) algorithm proposed by Kaufman [38], which is listed as follows:

Step 1: Randomly set k points in $\mathbf{X}$ as the initial medoids.

Step 2: Assign the other points into their nearest medoids to form k clusters.

Step 3: Calculate current cost E.

Step 4: For each medoid $\mathbf{p}_i$ ($1\leq i\leq k$), swap $\mathbf{p}_i$ with all other non-medoid points $\mathbf{x}_j$ in turn and calculated the associated cost $E_{i,j}$.

Step 5: If the minimum of $E_{i,j}$ is smaller than current cost E, swap the medoid $\mathbf{p}_i$ with point $\mathbf{x}_j$.

Step 6: Repeat Step 2 – Step 5 until the minimum of $E_{i,j}$ is larger than current cost E.

3.2 RMTLS method

3.2.1 Iterative regression procedure

The proposed RMTLS method contains a two-step pre-process. The first step is to obtain a regression-based reference by iterative regression procedure and the second step is to define and select the normal points based on the reference.

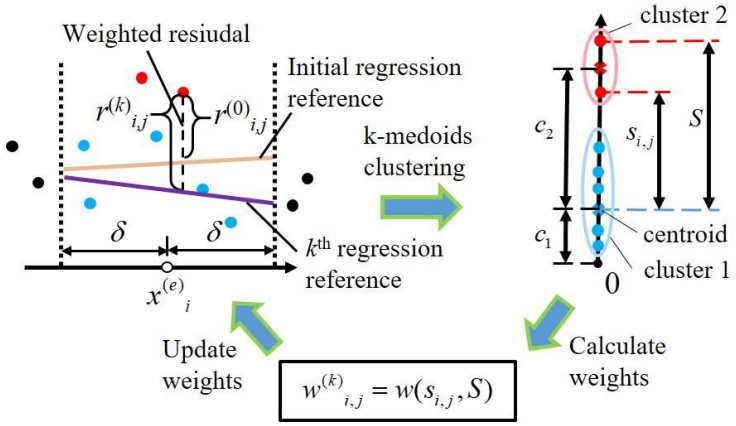


Fig. 2. The principle of iterative regression procedure

Fig.2 shows the principle of iterative regression procedure. In Fig.2, $r_{i,j}$ is the residual of j^{th} point in support domain Ω_i centred at $x^{(e)}_i$, and it is defined by $r_{i,j}=|y_j-\hat{y}_j|$ where $\hat{y}_j$ is the fitting value at x_j. The iterative procedure starts with the initial weighted residuals $\{r^{(0)}_{i,j}\}_{j=1}^{n}$, which are obtained by performing WLS method within the support domain Ω_i with the uniform initial weights $\{w^{(0)}_{i,j}\}_{j=1}^{n}=1$. In this paper, WLS method with third order basis is adopted to fit the local shape of the curve in a support domain. Then, the weight residuals $\{r^{(0)}_{i,j}\}_{j=1}^{n}$ are divided into two clusters through k-medoids clustering. Cluster 1 is defined as the cluster with lower centroid value while cluster 2 has a higher centroid. Based on clustering result, a weight function is proposed to calculate the weight of points with the expression

$$w^{(1)}_{i,j} = \begin{cases} 1 & , r^{(0)}_{i,j} \in \text{cluster 1} \\ 1-\dfrac{s_{i,j}}{S} & , r^{(0)}_{i,j} \in \text{cluster 2} \end{cases} \tag{20}$$

where $s_{i,j}$ is the difference between $r^{(0)}{}_{i,j}$ in cluster 2 and centroid 1, and S is the difference between centroid 1 and maximum of $\{r^{(0)}{}_{i,j}\}_{j=1}^{n}$. In k^{th} iteration ($k \geq 1$), through the WLS method with the weights $\{w^{(k)}{}_{i,j}\}_{j=1}^{n}$, the weight residuals $\{r^{(k)}{}_{i,j}\}_{j=1}^{n}$ can be obtained with the expression

$$r^{(k)}{}_{i,j} = r_{i,j} \cdot w^{(k)}{}_{i,j} \tag{21}$$

where

$$w^{(k)}{}_{i,j} = \begin{cases} w^{(k-1)}{}_{i,j} & , r^{(k-1)}{}_{i,j} \in \text{cluster 1} \\ w^{(k-1)}{}_{i,j} \cdot (1 - \dfrac{s_{i,j}}{S}) & , r^{(k-1)}{}_{i,j} \in \text{cluster 2} \end{cases} \tag{22}$$

In this way, after the iterative procedure, the influence of the outlier can be significantly weakened. It should be noticed that with the increase of iteration, the sum of weighted residual rapidly converges to near 0 and the reference barely changes. To quit the iteration and save computation cost, a criterion c_1/c_2 based on the clustering result is proposed. In Fig.2, c_1 is the difference between centroid 1 and zero, and c_2 is the difference between two centroids. If the difference between the sum of weighted residual set of two iterations is lower than the criterion c_1/c_2, the iterative procedure will be terminated and the last regression is considered as the regression-based reference which serves as the input of selection procedure. The details of the algorithm are listed as follows.

Algorithm 1 Iterative regression process

Input: discrete point set $\mathbf{X}$, estimation point $\mathbf{x}^{(e)}{}_i$, radius of support domain δ

Output: Regression-based reference l

Determine the support domain $\Omega_i = \{ \mathbf{x} \in \mathbf{X} \mid \ || \mathbf{x} - \mathbf{x}^{(e)}{}_i || < \delta \}$

Set $k = 0$

Set $\{w^{(k)}{}_{i,j}\}_{j=1}^{n} = 1$

while $((\text{sum}\{r^{(k-2)}{}_{i,j}\}_{j=1}^{n} - \text{sum}\{r^{(k-1)}{}_{i,j}\}_{j=1}^{n}) > c_1/c_2, k > 1)$ or $(k \leq 1)$ **do**

Obtain the regression $l^{(k)}$ in Ω_i by WLS method with $\{w^{(k)}{}_{i,j}\}_{j=1}^{n}$

Compute the weighted residuals $\{r^{(k)}{}_{i,j}\}_{j=1}^{n}$ with $\{w^{(k)}{}_{i,j}\}_{j=1}^{n}$ by Eq.(21)

Perform k-medoids clustering on $\{r^{(k)}{}_{i,j}\}_{j=1}^{n}$

$k = k+1$

Update the weights $\{w^{(k)}{}_{i,j}\}_{j=1}^{n}$ by Eq.(22)

end while

return Regression-based reference l

3.2.2 Density-based selection procedure

In the second step of pre-process, a density-based selection procedure is proposed to select the normal points in the support domain according to the regression-based

reference determined by the iterative regression procedure.

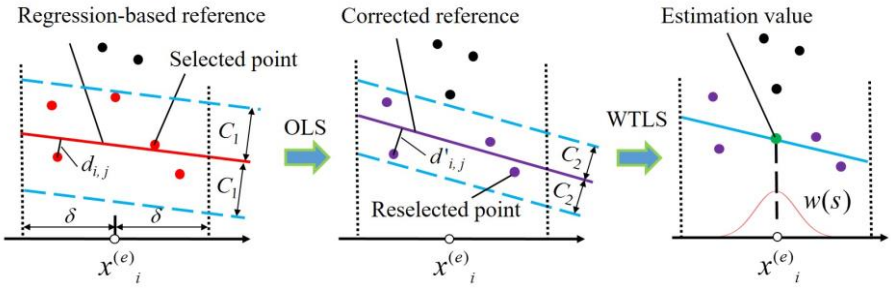


Fig. 3. The principle of density-based selection procedure

Fig.3 shows the principle of the density-based selection procedure. The distances $\{d_{i,1}, d_{i,2}, \cdots, d_{i,n}\}$ between each discrete points and regression-based reference are firstly calculated. The density function are defined based on the sorted set $\mathbf{D}_s = \{d_{i,(1)}, d_{i,(2)}, \cdots, d_{i,(n)}\}$, $d_{i,(1)} \le d_{i,(2)} \le \cdots \le d_{i,(n)}$ with the expression

$$\rho(j)^{-1} = \frac{d_{i,(j)}}{j}, \quad 1 \le j \le n \tag{23}$$

where ρ denotes the number of points within per unit length. The median of density can be defined as

$$\rho_{median}{}^{-1} = \frac{d_{i,(n/2)}}{n/2} \tag{24}$$

To ensure the result of $n/2$ is an integer, $n/2$ is rounded when n is an odd number. Then, the selected point set $\mathbf{X}_s$ can be determined through

$$\begin{cases} \mathbf{X}_s = \{d_{i,(j)} \mid d_{i,(j)} < C_1\} \\ C_1 = d_{i,(j')},\ j' = \arg\min_{k>n/2}(\rho(k)^{-1} > p_1 \cdot \rho_{median}{}^{-1}) \end{cases} \tag{25}$$

By letting $\rho(k)^{-1} = p_1 \cdot \rho_{median}{}^{-1}$, the outlier criterion for k^{th} distance $d_{i,(k)}$ in $\mathbf{D}_s$ can be further obtained with the following expression

$$criterion\ of\ \ d_{i,(k)} = \frac{d_{i,(n/2)}}{n/2} \cdot p_1 k = \rho_{median}{}^{-1} \cdot p_1 k\ , k > n/2 \tag{26}$$

where p_1 is a parameter related to the selection range. Suppose that normal points obey the Gauss distribution along regression-based reference with zero means and variance σ^2. Then, $d_{i,(n/2)}$ can be calculated approximately by $Z_{0.75} \cdot \sigma \approx 0.68\sigma$ according to cumulative Z-Table of standard normal distribution, and $d_{i,(n)}$ can be considered as 3σ. Then the ratio $\rho(n)^{-1}/\rho(n/2)^{-1} = (3\sigma/n)/(0.68\sigma/n/2) \approx 2.2$ can be obtained. Parameter p_1 is set to 2 in this paper.

Subsequently, the ordinary least square (OLS) method is applied on the point set $\mathbf{X}_s$ to obtain a corrected reference because that in the iterative regression method, the weights of part of normal points are influenced, resulting in the slight deviation of regression-based reference. Then, the distances $\{d'_{i,j}\}_{j=1}^{n}$ between corrected reference and discrete points are calculated, and re-selected point set $\mathbf{X}'_s$ is determined by performing the selection procedure with sorted distance set $\mathbf{D}'_s$ and parameter p_2. The determination of parameter p_2 is the same as p_1. However, the number of discrete points within support domain is small especially in curve reconstruction situation. Based on insufficient statistical information, the median distance $d'_{i,(n/2)}$ may deviates from the approximation value 0.68σ so that some normal points will be excluded from the re-selection step. Therefore, for curve reconstruction, p_2 is set to 4 and for surface reconstruction, p_2 is set to 2 in this paper.

The estimation value of $\mathbf{x}^{(e)}_i$ can be determined by performing the WTLS method on $\mathbf{X}'_s$ with the compact weight function $w(s)$. The algorithm of density-based selection procedure is listed as follows.

Algorithm 2 Density-based selection procedure

Input: regression-based reference l, support domain Ω_i, parameter p_1, p_2

Output: point set $\mathbf{X}'_s$

Compute the orthogonal distances $\{d_{i,j}\}_{j=1}^{n}$ between discrete points in Ω_i and l

Set $\mathbf{D}_s$ = sort $(\{d_{i,j}\}_{j=1}^{n}) = \{d_{i,(j)}\}_{j=1}^{n}$, $d_{i,(1)} \leq d_{i,(2)} \leq \cdots \leq d_{i,(n)}$

Determine the point set $\mathbf{X}_s$ with C_1 and p_1 by Eq.(25)

Obtain the corrected reference l' by performing OLS method on $\mathbf{X}_s$

Compute the orthogonal distances $\{d'_{i,j}\}_{j=1}^{n}$ between discrete points in Ω_i and l'

Set $\mathbf{D}'_s$ = sort $(\{d'_{i,j}\}_{j=1}^{n})$

Determine the reselected point set $\mathbf{X}'_s$ with C_2 and p_2

return point set $\mathbf{X}'_s$

4. Simulation and experiment validation

4.1. Reconstruction of simulated data

In this section, two simulated reconstruction instances are used to validate the effectiveness of the proposed RMTLS method. 1D and 2D function are taken as the curve and surface reconstruction objective, respectively. Firstly, a series of equal spaced estimation points are used to obtain the discrete data of the function. Then, outliers and random errors which obey normal distribution with zero means and the variance of σ_x^2, σ_y^2 are added to these discrete data to form the simulated data.

The 1D function

$$y = 1.1(1 - x + 2x^2)e^{-\frac{x^2}{4.5}}, \quad x \in \Omega \tag{27}$$

where $\Omega=\{x|\ x\in[0,10]\ \}$ is taken as the simulated curve reconstruction case. In this case, 201 estimation points are used. The radius of support domain δ is set to $\Omega_{length}\times 5/100$. Random errors with $\sigma_x=\sigma_y=10^{-3}$ are added and five outliers E_1, ⋯, E_5 are added to the different locations of the function.

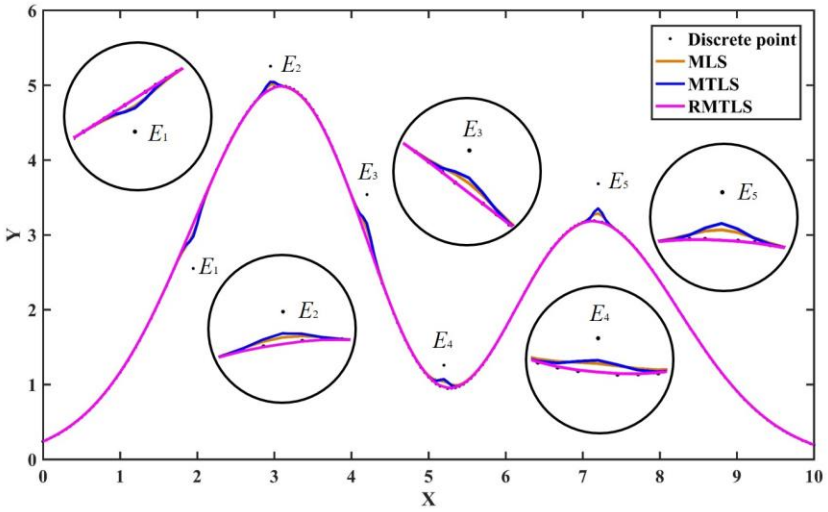


Fig. 4. The reconstructed curves in Case 1D

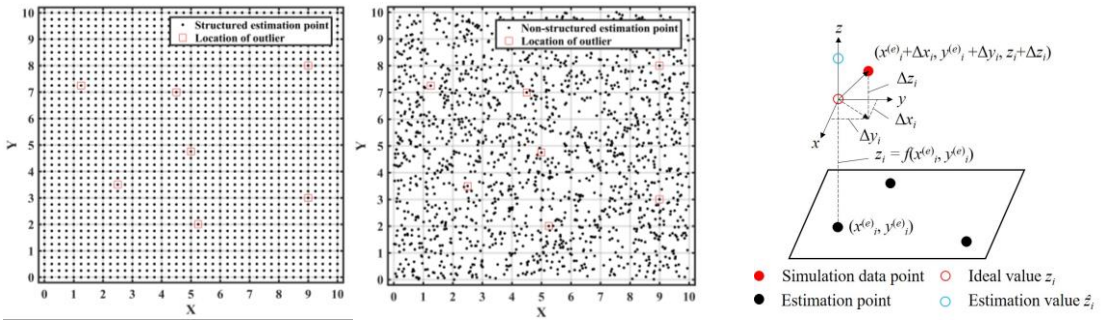
Fig.4 shows the reconstructed curves by three methods, namely MLS, MTLS and RMTLS in a typical simulation run. From the local enlargement graph in Fig.4, the presence of outliers distorts the reconstructed curves in the MLS and MTLS method. The poor estimation is adverse to the application of further analysis techniques such as roughness estimation and geometry characterization. By applying the RMTLS method to Case 1D, we obtain a smooth reconstructed curve (purple line) and all the outliers are successfully suppressed as shown in Fig.4.

To study the effectiveness of the RMTLS method on surface reconstruction, the 2D function [39]

$$z = \frac{3}{4}e^{(-\frac{1}{4}((0.9x-2)^2+(0.9y-2)^2))} + \frac{3}{4}e^{(-\frac{1}{49}(0.9x+1)^2-\frac{1}{10}(0.9y+1)^2))} + \frac{1}{2}e^{(-\frac{1}{4}(0.9x-7)^2-\frac{1}{4}(0.9y-3)^2))} - \frac{1}{5}e^{(-(0.9x-4)^2-(0.9y-7)^2))} \tag{28}$$

defined in $\Omega=\{(x,\ y)|(x,\ y)\in[0,10]\times[0,10]\}$ is taken as the simulated surface reconstruction case. In this case, two types of estimation points are investigated, i.e., structured and non-structured estimation points as shown in Fig.5(a) and (b). The number of both estimation points are 1681. In Fig.5(a), estimation points lie on a structured grid with 41 columns and 41 rows. For Fig.5(b), estimation points lie on a non-structured grid which is generated by sampling from random variable T_x and T_y which both obey a uniform distribution in [0,10] and the sampled values (t_x, t_y) are

taken as the coordinates of estimation points. Fig.5(c) shows the generation of 2D simulation data. For estimation point $(x^{(e)}_i, y^{(e)}_i)$, the ideal value z_i can be obtained through Eq.(28). Then, three random error terms ($\Delta x \sim N(0,\sigma_x^2)$, $\Delta y \sim N(0,\sigma_y^2)$, $\Delta z \sim N(0,\sigma_z^2)$) are added to form the simulation data point $(x^{(e)}_i+\Delta x_i, y^{(e)}_i+\Delta y_i, z_i+\Delta z_i)$, $1 \le i \le n$. Let $\delta=(\Omega_{width}+\Omega_{length})\times 5/100$. In this case, random errors are set to $\sigma_x=\sigma_y=\sigma_z=10^{-3}$ and seven outliers are added. Let $\delta=(\Omega_{width}+\Omega_{length})\times 5/100$. In this case, random errors are set to $\sigma_x=\sigma_y=\sigma_z=10^{-3}$ and seven outliers are added.

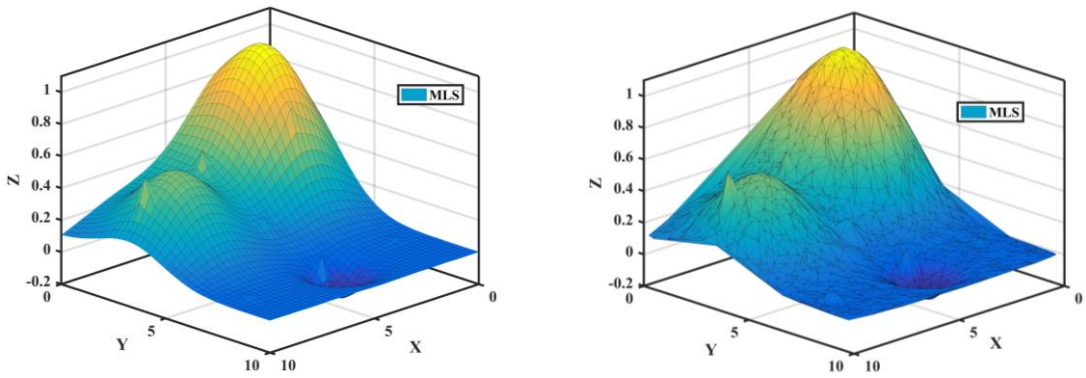


(a) Structured grid (b) Non-structured grid (c) Generation of simulation data

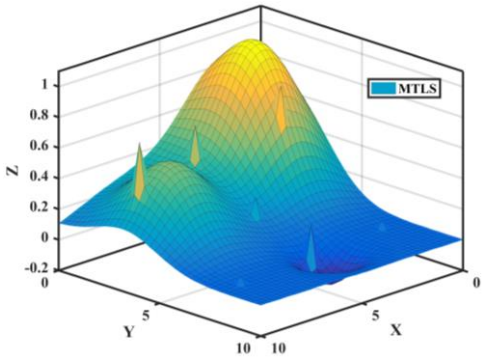
Fig. 5. The arrangement of estimation points in Case 2D

Fig.6 shows the reconstructed surfaces by three methods. On a structured grid, for the MLS and MTLS method, it can be found that reconstruction surfaces are both distorted near the outlier because these two methods treat outlier as a sharp feature and thus give ill result around the outlier. In contrast, RMTLS method removes the outlier and gives smooth reconstructed surface. On a non-structured grid, similar conclusion can be made that RMTLS has higher robustness to outliers. Fig.6(d1) and (d2) further give the error distribution of reconstructed surfaces by RMTLS on two types of grids. The errors in the marked regions of Fig.6(d1) are significantly smaller than that of Fig.6(d2).

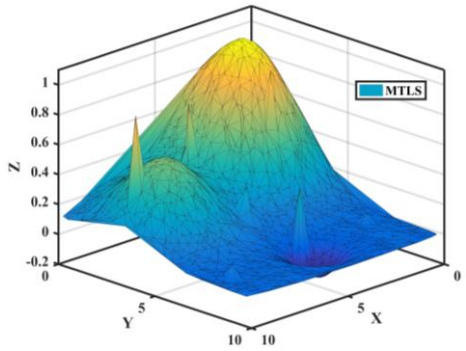
The reconstruction result of simulation data in Case 1D and 2D both prove that the RMTLS method appears the highest robustness among three methods.



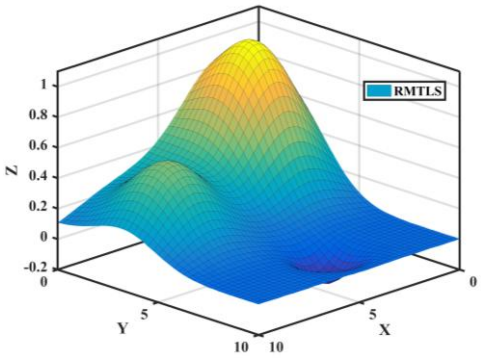
(a1) MLS - Structured grid



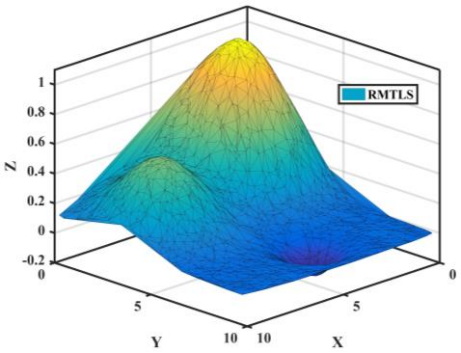
(a2) MLS with - Non-structured grid



(b1) MTLS - Structured grid

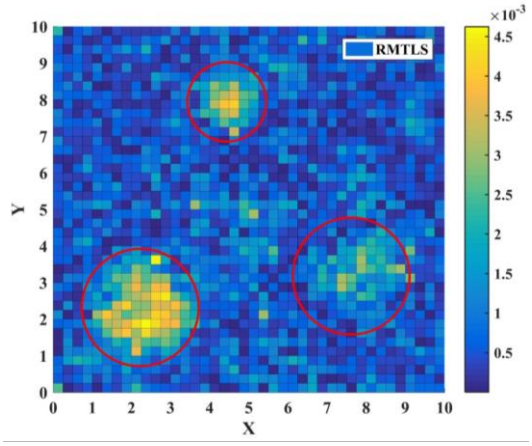


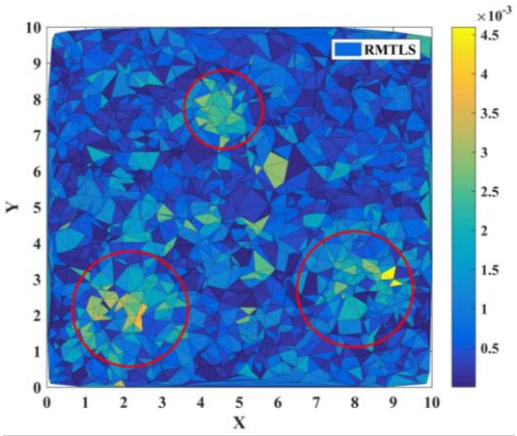
(b2) MTLS - Non-structured grid



(c1) RMTLS - Structured grid

(c2) RMTLS - Non-structured grid





(d1) Error distribution of (c1)

(d2) Error distribution of (c2)

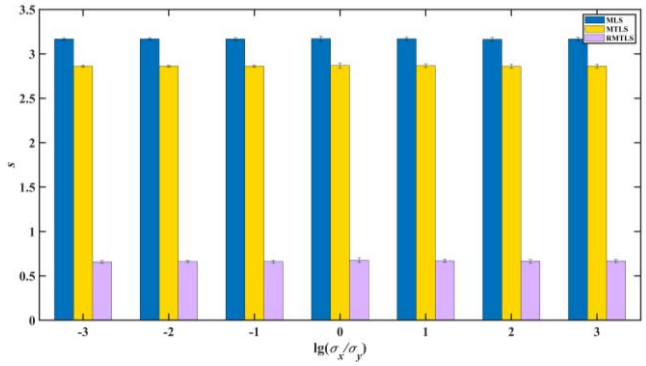
Fig. 6. Reconstructed surfaces and error distribution in Case 2D

To further investigate the performance of RMTLS method under different random errors, outliers in Case 1D and 2D, and random errors with the ratio σ_x/σ_y or $\sigma_{x(y)}/\sigma_z$ of being $10^{-6}/10^{-3}$, $10^{-5}/10^{-3}$, $10^{-3}/10^{-3}$, $10^{-3}/10^{-4}$, $10^{-3}/10^{-4}$, $10^{-3}/10^{-4}$, $10^{-3}/10^{-6}$ are added to both simulation cases, respectively. Moreover, in order to quantify the difference between different reconstructed curves and illustrate the reconstruction

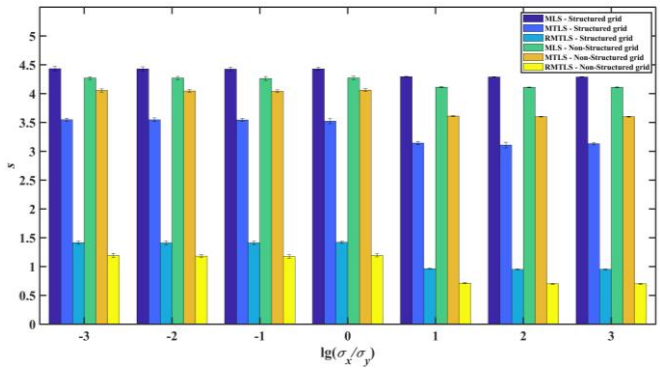
effectiveness, the reconstruction error defined by the sum of absolute residual of all estimation points s is adopted with the expression

$$s = \begin{cases} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| & \text{, in case 1D} \\ \sum_{i=1}^{n} \left| z_i - \hat{z}_i \right| & \text{, in case 2D} \end{cases} \tag{29}$$

where y_i and z_i are ideal value, and $\hat{y}_i$ and $\hat{z}_i$ is estimation value. Fig.7 shows the s values by three methods with different random errors. It can be found that in Fig.4 and Fig.6, the reconstructed curve and surface by the MTLS method distort more seriously near outliers than that by the MLS method, meaning higher estimation error of the MTLS method in these areas. However, from Fig.7(a) and (b), the reconstruction errors of the MTLS method in Case 1D and 2D with different random errors are all lower than that of the MLS method, which illustrates that the MTLS method achieves higher reconstruction accuracy in the areas without the presence of outliers compared to the MLS method. Then, by applying the robust pre-process, the s values by the RMTLS method are further reduced compared to the MTLS method because of the suppression of outliers within whole parameter domain. The comparison results show that the RMTLS method has the lowest reconstruction error and outstanding robustness to outliers.



(a) Case 1D



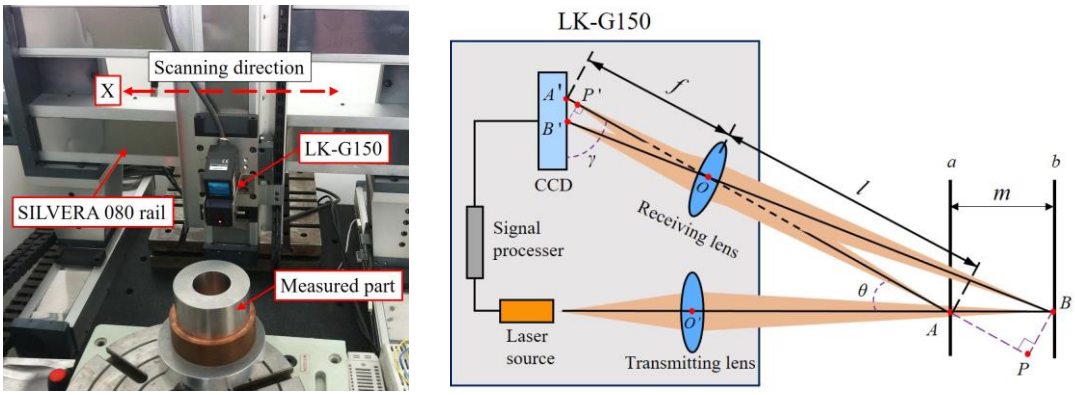
(b) Case 2D

Fig. 7. The *s* values in Case 1D and 2D with different random errors

4.2 Arrangement of measurement experiment

In order to apply the RMTLS method to the reconstruction of measurement data and investigate the reconstruction effectiveness, measurement experiments with two instruments were used to obtain the measurement data. In this section, a brief introduction to the instruments used in the experiment is given.

For the first measurement experiment, a self-built measuring platform was used to acquire curve measurement data. Fig.8(a) shows the self-built measurement platform that was mainly composed by a three-axis moving platform and a laser displacement sensor. The moving platform used a SILVERA 080 precision rail as the moving component controlled by Parker 1505 controller and LK-G150 laser displacement sensor from KEYENCE was adopted. In this measurement platform, LK-G150 was clamped on the Z-axis rail and the measuring object was mounted on the workbench. The repetitive positioning error of the moving platform is about 15μm and the repetitive measurement error of LK-G150 is about 5μm. The measurement data are obtained by scanning the tested sample through the translation platform and the scanning space was set to 0.05mm.



(a) Self-built measurement platform (b) Principle of laser displacement sensor LK-G150

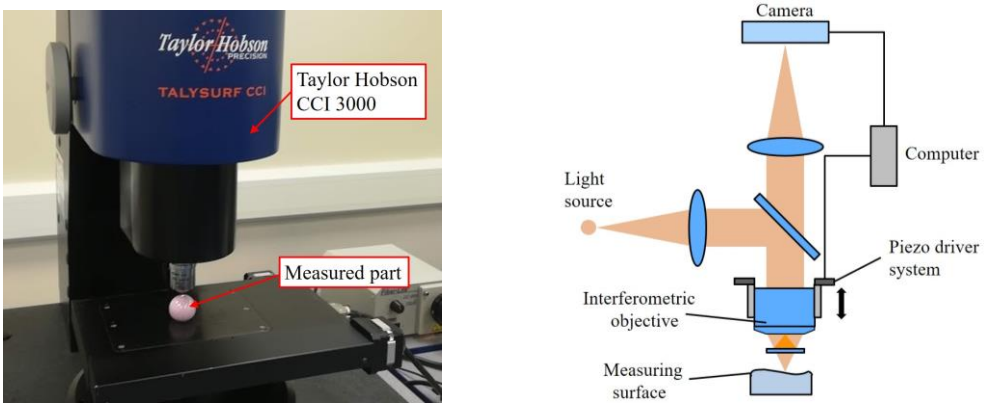
Fig. 8. The self-built measurement platform

Fig.8(b) gives the inner structure of the laser displacement sensor. LK-G150 measures the object at position *b* according to a reference position *a*. In Fig.8(b), $\triangle A'B'P'$ and $\triangle ABP$ are similar, and the relative distance *m* can be calculated through [40]

$$m = \frac{l \sin(\gamma)\overline{A'B'}}{f \sin(\theta) - l \sin(\theta+\gamma)\overline{A'B'}} \tag{30}$$

where l is the object distance, f is the focal distance, θ is the angle between transmitting beam and receiving beam and γ is the angle between receiving beam and CCD.

For the second measurement experiment, a non-contact measurement instrument, white light interferometer (WLI) - Taylor Hobson CCI 3000, is adopted. As shown in Fig.9(a), a sphere part is fixed on the movable workbench. The obtained 2D surface measurement data has a uniform space of 1.7969μm on x and y axis. Fig.9(b) shows the structure of the WLI system. A broadband light beam is firstly reflected by a beamsplitter and passes through an interferometric objective. The reflected beams from the reference mirror and tested surface were focused onto a CMOS camera. Interference fringes occur when the optical path difference (OPD) of two arms of the interferometer is within the coherence length, and the visibility of the fringes reaches maximum when the OPD equals zero. Through scanning the objective along the depth direction, a sequence of interferograms can be obtained. Afterwards, the surface topography within the field of view of the objective can be constructed by tracking coherence peaks in each interferogram and corresponding PZT positions.



(a) WLI-Taylor Hobson CCI 3000 (b) Principle of the WLI system

Fig. 9. The principle of the WLI system

4.3. Reconstruction of measurement data

In the first measurement experiment, the self-built measurement platform was used to measure the profile of a cylindrical part. Through calibration by PGI 1240 profilometer, the radius of the cylindrical part is considered as 40.1840 mm. We take

a segment of measurement data which contains 805 measurement points. Let $\delta=(\Omega_{length})\times 5.5/100$ where $\Omega=\{x|\ x\in[-22.5,22.5]\}$.

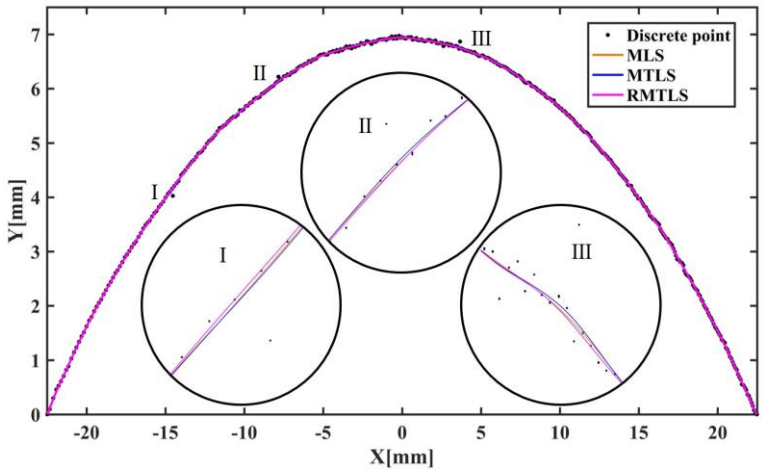
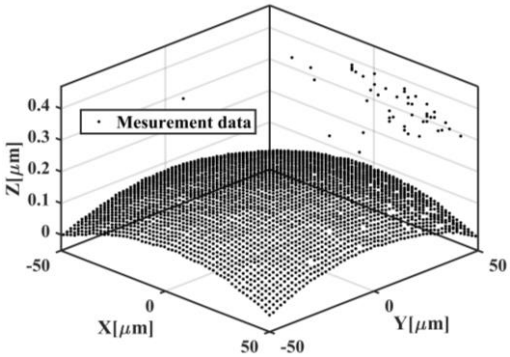


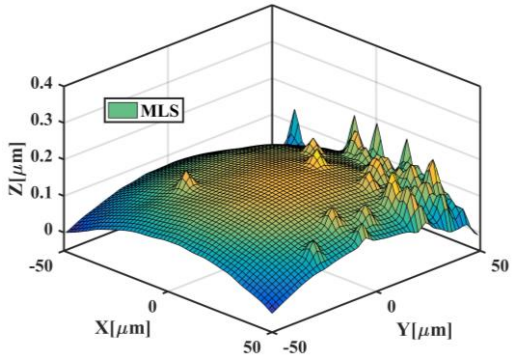
Fig. 10. The reconstructed curves of curve measurement data

Fig.10 shows the reconstructed curves by three methods. From the enlargement graph in Fig.10, the reconstructed curves by the MLS and MTLS method slightly distort toward the outlier compared to that by the RMTLS method. The reason of the small difference between the reconstructed curve by the RMTLS method and two other methods may attribute to the fact that the outlier is relatively close to the normal points.

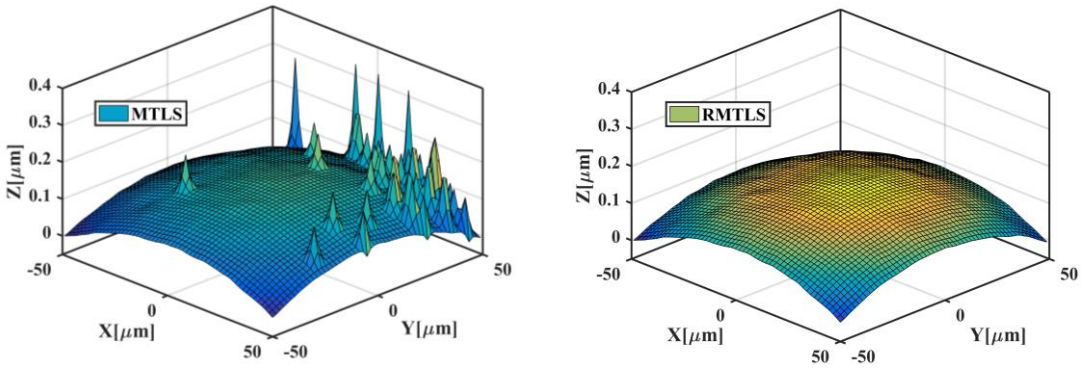
In the second experiment, a spherical surface was measured by a WLI system with a radius of 14.402mm obtained by the WLI system. Let $u=v=56$, $n=uv=3136$, $\delta=(\Omega_{width}+\Omega_{length})\times 2/100$ where $\Omega=\{(x,\ y)|(x,\ y)\in[-50,50]\times[-50,50]\}$. Fig.11(a) gives the measurement data from WLI system. Plenty of outliers on the right corner of the data can be observed. As shown in Fig.11(b) and (c), the reconstructed surfaces by the MLS and MTLS method distorted seriously by outliers. In contrast, the RMTLS method successfully suppresses the influence of all outliers and gives a smooth reconstructed surface as shown in Fig.11(d).



(a) Measurement data



(b) Reconstructed surface by the MLS method



(c) Reconstructed surface by the MTLS method (d) Reconstructed surface by the RMTLS method

Fig. 11. Reconstructed surfaces of surface measurement data

To further quantify the reconstruction effectiveness of the RMTLS method in this section, simulated annealing algorithm [41] is adopted to obtain a regression circle (or sphere

e) based on the reconstructed curve (or surface) with minimum sum of residuals. By comparing the regression radius and the reference radius, the reconstruct effectiveness can be evaluated. As shown in Table 1, for the curve measurement data, the regression radius by the RMTLS method is closest to the calibration radius 40.1840 mm, indicating that the RMTLS method achieves the best reconstruct effectiveness among these methods. For the surface measurement data, similar conclusion can also be drawn. Both experiments also prove that the RMTLS method has highest robustness to outliers and it is more suitable for the reconstruction of the measurement data.

Table 1. Regression radius of reconstructed curves and surfaces [mm]

Experiment No.	Reference radius	MLS	MTLS	RMTLS
1	40.184	39.932	40.006	40.044
2	14.402	14.594	14.631	14.567

5. Discussion

5.1 Suppression of multiple discrete outliers

Through Case 1D and 2D, the RMTLS method has showed great robustness to single outlier in a support domain. And the reconstruction result in the second experiment preliminarily shows the ability of the RMTLS method in handling multiple outliers. To quantitatively investigate the robustness of RMTLS in the

presence of multiple discrete outliers, the outliers with different percentages in the support domain are added.

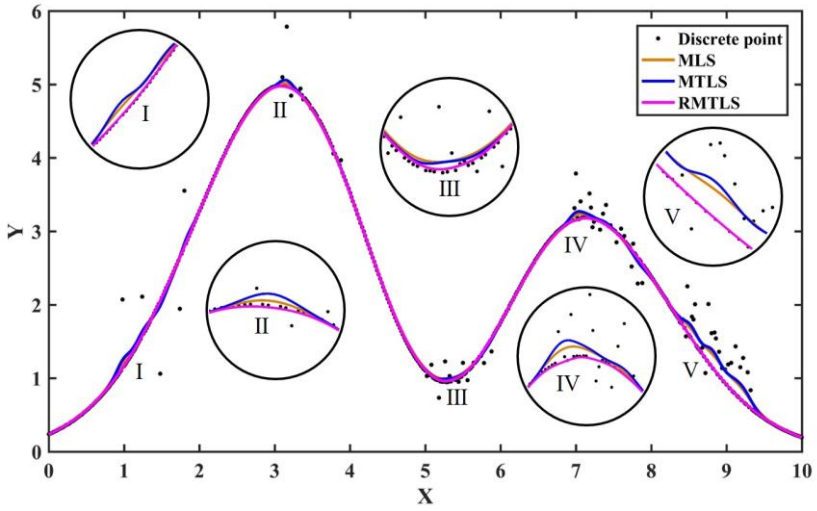


Fig. 12. The reconstructed curves of Case 1D with multiple discrete outliers

Take Case 1D as an example and increase the total number of estimation points to 500. Let $r=\Omega_{length}\times 5/100$. The outliers with percentages ranging from 10% to 48% in the support domain are added in different regions of the Case 1D. In Fig.12, the support domain in location I, II, III, IV, V includes the outlier with the percentage of 10%, 20%, 30%, 40%, 48%, respectively. From the reconstructed curves by three methods, it can be found that RMTLS is able to effectively suppress multiple discrete outliers. When the outliers account over 50% of total number of the discrete points in a support domain, according to the principle of selection procedure, $d_{i,(n/2)}$ will be considered as a normal point while it is actually an outlier in this situation. Consequently, more outliers will be selected owing to the poor estimation of selection range. Therefore, in this situation, even if all outliers are successfully suppressed in the iterative regression procedure, the selection procedure will still include some outliers wrongly, resulting in the deviation of estimation value.

5.2 Suppression of multiple continual outliers

As shown above, RMTLS is able to suppress the multiple discrete outliers in the support domain. However, if continual outliers exist in support domain, the iterative regression procedure will give an ill reference that biases to the outliers as shown in the left part of Fig.13. Different from multiple discrete outliers, there are no normal points between continuous outliers, which is difficult for iterative regression procedure to suppress them especially when continuous outliers exist on the edge of

support domain. An applicable solution is to introduce the previous estimation points which can be seen as normal points to the iterative regression procedure to transfer the continuous outliers into discrete outliers as shown in section 5.1.

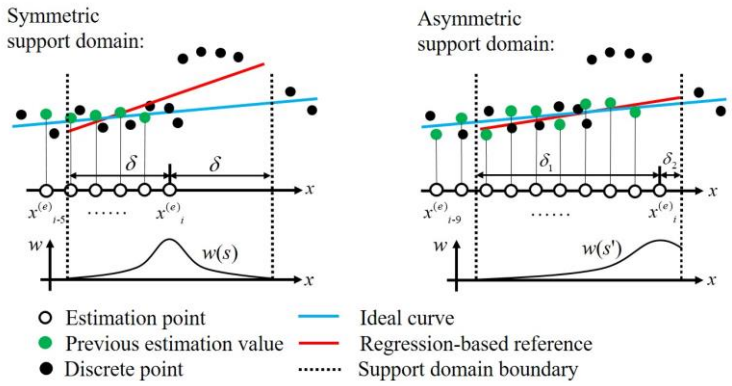


Fig. 13. The comparison between symmetric and asymmetric support domain

In a symmetric support domain, i.e., the estimation point locates at the center of support domain. If all the previous estimation points are taken into the iterative regression procedure as shown in the left part of Fig.13, there are still no normal points in the front of current estimation point $x^{(e)}_i$. To solve this problem, the location of the estimation point $x^{(e)}_i$ should be moved forward to the frontal edge of the support domain as shown in the right part of Fig.13, which leads to an asymmetric support domain. In this way, by introducing the previous estimation points in asymmetric support domain to the iterative regression procedure, continual outliers can be transferred into discrete outliers. Moreover, in asymmetric support domain, the compact weight function becomes

$$w(s') = \begin{cases} \dfrac{e^{-\theta s'^2} - e^{-\theta}}{1 - e^{-\theta}} & , s \le 1 \\ 0 & , s > 1 \end{cases} \quad , s' = \frac{\left\| \mathbf{x}_i - (\mathbf{x}_c + (\delta_1 - \mathbf{x}_c)) \right\|}{\max\{\delta_1, \delta_2\}} \tag{31}$$

where δ_1 is the distance between left boundary of support domain and estimation point $x^{(e)}_i$, and δ_2 is the distance of $x^{(e)}_i$ to the right boundary of support domain as shown in Fig.13.

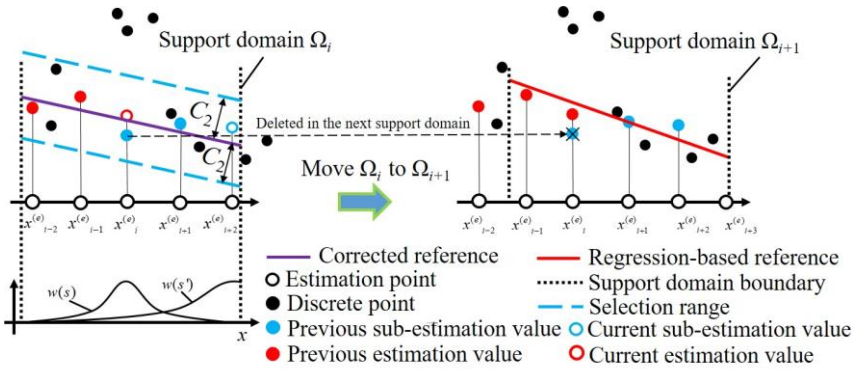


Fig. 14. The application of asymmetric support domain in the RMTLS method

The left of Fig.14 shows the WTLS estimation step in support domain Ω_i. WTLS is applied to obtain the estimation value at the central estimation point $x=x^{(e)}_i$ with weight $w(s)$ and the sub-estimation value at the most frontal estimation point $x=x^{(e)}_{i+2}$ with weight $w(s')$. It is noted that after WTLS estimation, there are two estimation values at $x=x^{(e)}_i$. To ensure one estimation value at each estimation point, the previous sub-estimation value at $x=x^{(e)}_i$ will be deleted in the next support domain. Then, in the next support domain Ω_{i+1}, discrete points, previous estimation and sub-estimation points are taken as the input of iterative regression procedure to obtain a regression-based reference.

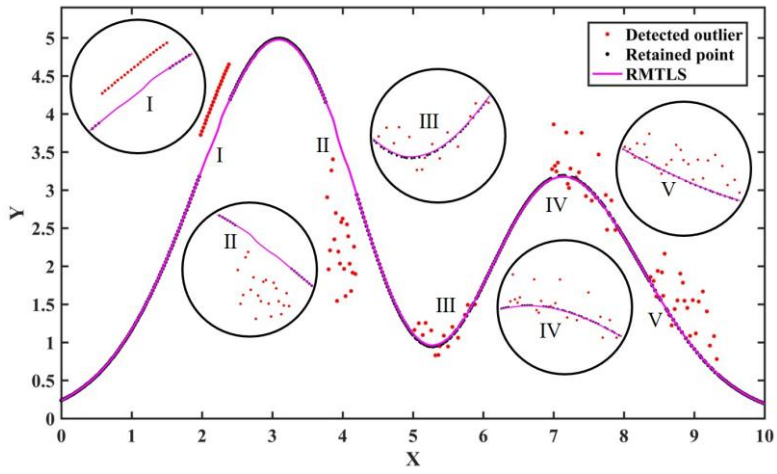


Fig. 15. The reconstructed curve by the RMTLS method with asymmetric support domain

To validate the effectiveness of the RMTLS method with asymmetric support domain, some continual outliers and discrete outliers are both added to different locations of Case 1D. Fig.15 shows the reconstructed curve. Detected outliers in Fig.15 represent those points that have not been selected by the pre-process in any support domains. From Fig.15, it can be seen that most of continual and discrete

outliers are detected. The reconstructed curve proves that the robustness of RMTLS can be further improved by adopting an asymmetric support domain.

5.3 Suppression of random errors

In this section, another situation that no outliers exist in the support domain will be discussed. Under this situation, basically all the points will be selected thorough pre-process so that RMTLS method will give the same estimation result with MTLS method. To further suppress the random errors, a clustering-based weight function is applied to adjust the weight of points within the support domain. After the two-step pre-process, the selected point set $\mathbf{X}'_s$ and corresponding distance set $\{d'_{i,j} \mid \mathbf{x}_{i,j} \in \mathbf{X}'_s\}$ can be obtained as stated in section 3.2.

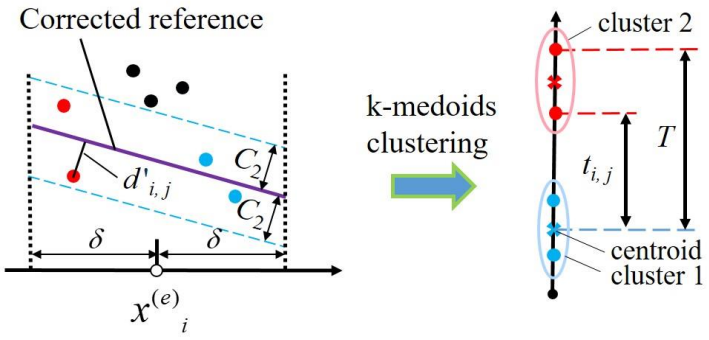
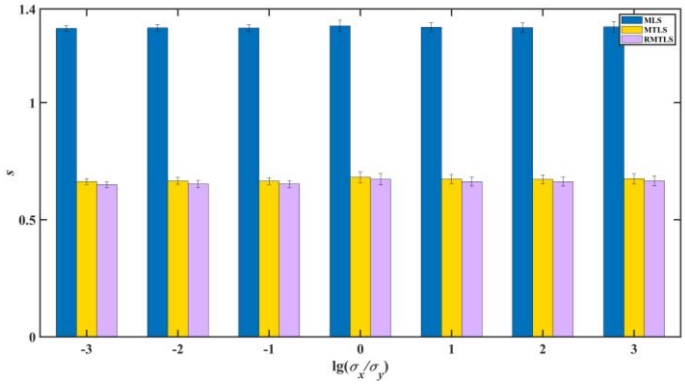


Fig. 16. The construction of clustering-based weight function

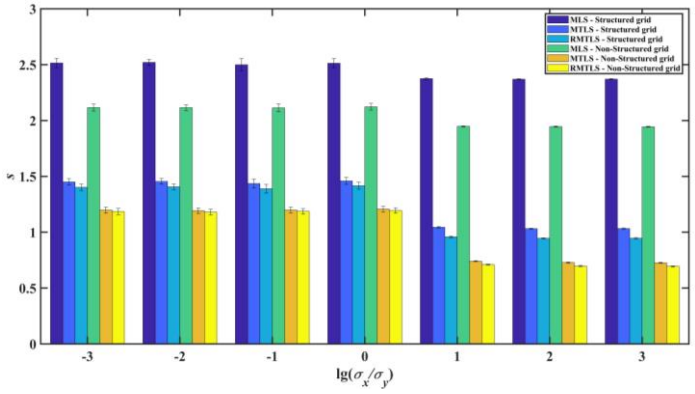
By performing k-medoids clustering on the distance set $\{d'_{i,j} \mid \mathbf{x}_{i,j} \in \mathbf{X}'_s\}$ as shown in Fig.16, clustering result can be used to define a weight function

$$w^{(d)}_{i,j} = \begin{cases} 1 & , d'_{i,j} \in \text{cluster 1} \\ 1+\lambda \cdot \dfrac{t_{i,j}}{T} & , d'_{i,j} \in \text{cluster 2} \end{cases}, \quad \lambda \in [-1,1] \tag{32}$$

where $t_{i,j}$ is the difference between $d'_{i,j}$ and centroid 1, T is the difference between centroid 1 and maximum of distance set, and λ is a parameter which is associated with weight of point in cluster 2. When λ is greater than 0, the influence of the point with large random errors will be increased and vice versa. A large value of $|\lambda|$ indicates that a sharp transition of weight between the points in two clusters. The estimation value can be determined through the WTLS method with the combination of compact weight function $w(s)$ and clustering-based weight function $w^{(d)}$.



(a) Case 1D without outliers



(b) Case 2D without outliers

Fig. 17. The *s* values in previous cases only with different random errors

Previous simulated curve and surface reconstruction case with different random errors are adopted to validate the effectiveness of clustering-based weight function on suppressing random errors. Fig.17(a) and (b) show the *s* values in Case 1D and 2D only with different random errors, respectively. The comparison results show that RMTLS method with clustering-based weight function achieves the highest accuracy.

5.4 Computation efficiency

In this section, the computation efficiency of three methods is investigated. Three methods are run in MATLAB and the CPU is Intel i5-7500 3.6GHz. Time consumption is used to evaluate the computation efficiency of each method. Two sets of simulations are designed based on Case 2D with a structured grid, and the first simulation keeps the number of points in support domain $n\times\delta^2$ fixed (n is the number of total points and δ is the radius of support domain) while n gradually increases. In the second simulation, n is fixed and $n\times\delta^2$ gradually increases. Table 2 and 3 give the results of two simulations.

Table 2. Comparison of time consumption by three methods with fixed $n\times\delta^2$

n	32^2	48^2	64^2	96^2	128^2	192^2	256^2
δ	1	2/3	1/2	1/3	1/4	1/6	1/8
t_{MLS}[*second*]	0.209	0.522	1.175	3.659	10.730	45.591	133.243
t_{MTLS}[*second*]	0.200	0.484	1.091	3.645	10.433	44.609	132.602
t_{RMTLS}[*second*]	1.303	2.904	5.186	12.709	30.063	87.967	213.867
$(t_{RMTLS}-t_{MLS})/t_{MLS}$	5.234	4.563	3.414	2.473	1.802	0.929	0.605

Table 3. Comparison of time consumption by three methods with fixed n

n	41^2	41^2	41^2	41^2	41^2	41^2	41^2
δ	0.8	1	1.2	1.4	1.8	2.2	2.6
t_{MLS}[*second*]	0.456	0.544	0.766	0.923	1.506	2.158	2.965
t_{MTLS}[*second*]	0.443	0.478	0.690	0.878	1.478	2.133	2.782
t_{RMTLS}[*second*]	2.378	2.609	3.693	5.420	12.589	21.983	36.220
$(t_{RMTLS}-t_{MLS})/t_{MLS}$	4.215	3.796	3.821	4.872	7.359	9.187	11.216

In Table 2, when the number of points in support domain is fixed, with the increase of n, the relative difference between t_{RMTLS} and t_{MLS}, i.e., $(t_{RMTLS}-t_{MLS})/t_{MLS}$, gradually decreases. While in Table 3, when the number of points in each support increases, the relative difference between t_{RMTLS} and t_{MLS} increases rapidly. The two simulations indicate that the efficiency of RMTLS method is more sensitive to $n\times\delta^2$ (number of points in support domain) while less sensitive to number of total points n. Based on the simulation results, the selection of $n\times\delta^2$ should be mainly considered to allow the application of RMTLS on reconstructing large scale data with a small δ.

6. Conclusions

A robust pre-process for the support domain is proposed to improve the robustness of the MTLS method to outliers in this paper. The pre-process includes two steps. The first step is to obtain a regression-based reference within support domain. By adopting the k-medoids clustering, the weight of the outlier can be automatically weakened without setting any threshold values. The second step is to select the normal points along the reference. Simulated curve and surface reconstruction cases are used to validate the effectiveness of the RMTLS method. The sum of absolute residuals as well as root mean square (RMS) both prove that the RMTLS method has

higher robustness to outliers compared to the MLS and MTLS method. Then, the RMTLS method is applied to the reconstruction of measurement data derived from a self-built measuring platform and a WLI measuring system respectively. The RMTLS method successfully suppresses all the outliers and gives the reconstructed result which is the closest to the calibration result. Furthermore, by adopting asymmetric support domain and including the previous estimation points into the pre-process, multiple continual outliers can be suppressed. The random error can also be suppressed by introducing the clustering-based weight function.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 51605094 and 11572316), the Science Project of Anhui Province in China (Grant No. 201903a07020019), and the Fundamental Research Funds for the Central Universities (Grant No. WK2480000006).

References

[1] A. Hashemian, S.F. Hosseini, An integrated fitting and fairing approach for object reconstruction using smooth NURBS curves and surfaces, Comput. Math. with Appl. 76 (2018) 1555-1575. https://doi.org/10.1016/j.camwa.2018.07.007.

[2] M. Lhuillier, Surface reconstruction from a sparse point cloud by enforcing visibility consistency and topology constraints, Comput. Vis. Image Underst. 175 (2018) 52-71. https://doi.org/10.1016/j.cviu.2018.09.007.

[3] T. Gu, S. Ji, S. Lin, T. Luo, Curve and surface reconstruction method for measurement data, Measurement. 78 (2016) 278-282. https://doi.org/10.1016/j.measurement.2015.10.011.

[4] A. Gálvez, A. Iglesias, A. Avila, Immunological-based Approach for Accurate Fitting of 3D Noisy Data Points with Bézier Surfaces, Procedia Comput. Sci. 18 (2013) 50-59. https://doi.org/10.1016/j.procs.2013.05.168.

[5] M.P. Limongelli, The Surface Interpolation Method for damage localization in plates, Mech. Syst. Signal Process. 118 (2019) 171-194. https://doi.org/10.1016/j.ymssp.2018.08.032.

[6] D. Jiang, L.C. Wang, An algorithm of NURBS surface fitting for reverse engineering, Int. J.

Adv. Manuf. Technol. 31 (2006) 92-97. https://doi.org/10.1007/s00170-005-0161-3.

[7] K. Uyar, E. Ülker, B-spline curve fitting with invasive weed optimization, Appl. Math. Model. 52 (2017) 320-340. https://doi.org/10.1016/j.apm.2017.07.047.

[8] X. Zhao, C. Zhang, L. Xu, B. Yang, Z. Feng, IGA-based point cloud fitting using B-spline surfaces for reverse engineering, Inf. Sci. (Ny). 245 (2013) 276-289. https://doi.org/10.1016/j.ins.2013.04.022.

[9] E.K. Ueda, M. de Sales Guerra Tsuzuki, R.Y. Takimoto, A.K. Sato, T. de Castro Martins, P.E. Miyagi, R.S. Ubertino Rosso, Piecewise Bézier Curve Fitting by Multiobjective Simulated Annealing, IFAC-PapersOnLine. 49 (2016) 49-54. https://doi.org/10.1016/j.ifacol.2016.12.160.

[10] A. Gálvez, A. Iglesias, Particle swarm optimization for non-uniform rational B-spline surface reconstruction from clouds of 3D data points, Inf. Sci. (Ny). 192 (2012) 174-192. https://doi.org/10.1016/j.ins.2010.11.007.

[11] D. Mirzaei, Analysis of moving least squares approximation revisited, J. Comput. Appl. Math. 282 (2015) 237-250. https://doi.org/10.1016/j.cam.2015.01.007.

[12] E.J. Kansa, Multiquadrics - A scattered data approximation scheme with applications to computational fluid-dynamics - I surface approximations and partial derivative estimates, Comput. Math. with Appl. 19 (1990) 127-145. https://doi.org/10.1016/0898-1221(90)90270-T.

[13] L.M. Zhu, X.M. Zhang, H. Ding, Y.L. Xiong, Geometry of Signed Point-to-Surface Distance Function and Its Application to Surface Approximation, J. Comput. Inf. Sci. Eng. 10 (2010) 041003. https://doi.org/10.1115/1.3510588.

[14] H.K. Zhao, S. Osher, B. Merriman, M. Kang, Implicit and Nonparametric Shape Reconstruction from Unorganized Data Using a Variational Level Set Method, Comput. Vis. Image Underst. 80 (2000) 295–314. https://doi.org/10.1006/cviu.2000.0875.

[15] T. Ni, Z. Ma, A fast surface reconstruction algorithm for 3D unorganized points, in: 2010 2nd Int. Conf. Comput. Eng. Technol., 2010: pp. V7-15-V7-18. https://doi.org/10.1109/ICCET.2010.5485908.

[16] S.P. Lim, H. Haron, Surface reconstruction techniques: a review, Artif. Intell. Rev. 42 (2014) 59-78. https://doi.org/10.1007/s10462-012-9329-z.

[17] D. Shepard, A Two-Dimensional Interpolation Function for Irregularly-Spaced Data, in: Proc. 1968 23rd ACM Natl. Conf., Association for Computing Machinery, New York, NY, USA, 1968: pp. 517-524. https://doi.org/10.1145/800186.810616.

[18] P. Lancaster, K. Salkauskas, Surfaces Generated by Moving Least Squares Methods, Math.

Comput. 37 (1981) 141-158. https://doi.org/10.2307/2007507.

[19] T. Belytschko, Y.Y. Lu, L. Gu, Element-free Galerkin methods, Int. J. Numer. Methods Eng. 37 (1994) 229-256. https://doi.org/10.1002/nme.1620370205.

[20] Z. Zhang, D.M. Li, Y.M. Cheng, K.M. Liew, The improved element-free Galerkin method for three-dimensional wave equation, Acta Mech. Sin. 28 (2012) 808-818. https://doi.org/10.1007/s10409-012-0083-x.

[21] S.N. Atluri, T. Zhu, A new Meshless Local Petrov-Galerkin (MLPG) approach in computational mechanics, Comput. Mech. 22 (1998) 117-127. https://doi.org/10.1007/s004660050346.

[22] X. Li, Three-dimensional complex variable element-free Galerkin method, Appl. Math. Model. 63 (2018) 148-171. https://doi.org/10.1016/j.apm.2018.06.040.

[23] D.M. Li, K.M. Liew, Y. Cheng, An improved complex variable element-free Galerkin method for two-dimensional large deformation elastoplasticity problems, Comput. Methods Appl. Mech. Eng. 269 (2014) 72-86. https://doi.org/10.1016/j.cma.2013.10.018.

[24] S. Soghrati, A. Nagarajan, B. Liang, Conforming to interface structured adaptive mesh refinement: New technique for the automated modeling of materials with complex microstructures, Finite Elem. Anal. Des. 125 (2017) 24-40. https://doi.org/10.1016/j.finel.2016.11.003.

[25] R. Scitovski, Š. Ungar, D. Jukić, Approximating surfaces by moving total least squares method, Appl. Math. Comput. 93 (1998) 219-232. https://doi.org/10.1016/S0096-3003(97)10077-7.

[26] R. Hoseinnezhad, A. Bab-Hadiashar, An M-estimator for high breakdown robust estimation in computer vision, Comput. Vis. Image Underst. 115 (2011) 1145-1156. https://doi.org/10.1016/j.cviu.2011.03.007.

[27] K.W. Fornalski, Applications of the robust Bayesian regression analysis. Int. J. Soc. Syst. Sci. 7 (2015) 314-333. https://doi.org/10.1504/IJSSS.2015.073223.

[28] T. Baldacchino, K. Worden, J. Rowson, Robust nonlinear system identification: Bayesian mixture of experts using the t-distribution, Mech. Syst. Signal Process. 85 (2017) 977-992. https://doi.org/10.1016/j.ymssp.2016.08.045.

[29] C.C. Chuang, Z.J. Lee, Hybrid robust support vector machines for regression with outliers, Appl. Soft Comput. 11 (2011) 64-72. https://doi.org/10.1016/j.asoc.2009.10.017.

[30] D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, A.Y. Wu, On the Least Trimmed Squares Estimator, Algorithmica. 69 (2014) 148–183. https://doi.org/10.1007/s00453-012-9721-8.

[31] C.M. Cheng, S.H. Lai, A consensus sampling technique for fast and robust model fitting, Pattern Recognit. 42 (2009) 1318-1329. https://doi.org/10.1016/j.patcog.2009.01.007.

[32] X. Li, S. Li, On the stability of the moving least squares approximation and the element-free Galerkin method, Comput. Math. with Appl. 72 (2016) 1515-1531. https://doi.org/10.1016/j.camwa.2016.06.047.

[33] L. Cueto-Felgueroso, I. Colominas, X. Nogueira, F. Navarrina, M. Casteleiro, Finite volume solvers and Moving Least-Squares approximations for the compressible Navier-Stokes equations on unstructured grids, Comput. Methods Appl. Mech. Eng. 196 (2007) 4712-4736. https://doi.org/10.1016/j.cma.2007.06.003.

[34] J. Zhao, W. Wang, Y. Zhou, J. Zhao, Robust high precision multi-frame motion detection for PMLSMs' mover based on local upsampling moving least square method, Mech. Syst. Signal Process. 159 (2021) 107803. https://doi.org/https://doi.org/10.1016/j.ymssp.2021.107803.

[35] B. Kim, T. Lee, T.B.M.J. Ouarda, Total least square method applied to rating curves, Hydrol. Process. 28 (2014) 4057-4066. https://doi.org/10.1002/hyp.9944.

[36] H.S. Park, C.H. Jun, A simple and fast algorithm for K-medoids clustering, Expert Syst. Appl. 36 (2009) 3336-3341. https://doi.org/10.1016/j.eswa.2008.01.039.

[37] P. Arora, Deepali, S. Varshney, Analysis of K-Means and K-Medoids Algorithm For Big Data, Procedia Comput. Sci. 78 (2016) 507-512. https://doi.org/10.1016/j.procs.2016.02.095.

[38] L. Kaufman, P.J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, J. R. Stat. Soc. C-Appl. 40 (1991) 486-487. https://doi.org/10.1002/9780470316801.

[39] W. Li, G. Song, G. Yao, Piece-wise moving least squares approximation, Appl. Numer. Math. 115 (2017) 68-81. https://doi.org/10.1016/j.apnum.2017.01.001.

[40] J.H. Kalivas, Optimization using variations of simulated annealing, Chemom. Intell. Lab. Syst. 15 (1992) 1-12. https://doi.org/10.1016/0169-7439(92)80022-V.

[41] Z. Dong, X. Sun, C. Chen, H. Yang, L. Yang, An improved signal processing method for the laser displacement sensor in mechanical systems, Mech. Syst. Signal Process. 122 (2019) 403-418. https://doi.org/10.1016/j.ymssp.2018.12.018.