An outlier detection and recovery method based on moving least squares quasi-interpolation scheme and l_0-minimization problem

Sanpeng Zheng[a], Renzhong Feng[a,*], Aitong Huang[a]

[a] *School of Mathematicial Sciences, Beihang University, Beijing, China*

Abstract

In the scattered data used for function approximation, there may be some data that deviate greatly from their true values, which are called outliers. The existence of outliers significantly affects the accuracy of some approximation methods, such as moving least squares. How to quickly and accurately find these outliers and restore their true values has not been well resolved. In this paper, a new outlier detection and recovery method is proposed for data processing. The method uses the high accuracy of moving least squares quasi-interpolation scheme, its sensitivity to outliers and the sparse distribution of outliers to construct a l_0-minimization problem with an inequality constrain. Under certain assumptions, it is proved theoretically that the deviation vector corresponding to the outliers is a solution to the optimization problem. The classical orthogonal matching pursuit algorithm is introduced to solve the optimization problem efficiently. By solving the problem, the outliers are marked, and the deviations are also estimated approximately, which can restore the true values of outliers. The numerical experiments demonstrate that the proposed method has high computational efficiency, very high detection accuracy, and high recovery accuracy for the scattered data used for function approximation, so it is practical.

Keywords:
scattered data, outlier detection, moving least squares, quasi-interpolation, l_0-minimization problem
2000 MSC: 68P99, 33F05, 41A10, 62H30, 65K99

*Corresponding author: fengrz@buaa.edu.cn
Email address: fengrz@buaa.edu.cn (Renzhong Feng)

1. Introduction

Function approximation is a basic method in computational mathematics, and scattered data approximation is a general case in high dimensions. They play an important role in research and applications [1, 2]. Assuming the scattered data for approximation are sampled from an unknown function $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}^d$, at N points $X = \{\boldsymbol{x}_i\}_{i=1}^N$, and the exact sampling values are $\boldsymbol{F} = (f_1, \cdots, f_N)^\top$. In the processes of data collection, storage and application, some data that deviate greatly from their true values may be introduced, which are called outliers. The existence of outliers is unpredictable and it pollutes sampling values into $\tilde{\boldsymbol{F}} = (f_1 + e_1^*, \cdots, f_N + e_N^*)^\top = \boldsymbol{F} + \boldsymbol{E}^*$. In general, f and $\boldsymbol{F}$ are unknown, and $\boldsymbol{E}^*$ is sparse, which means that only a few components of $\boldsymbol{E}^*$ are large values and others are zero.

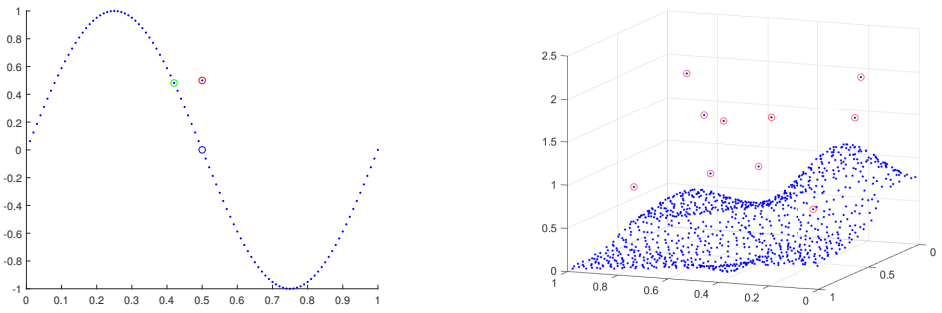
Most approximation methods are invalid for the data containing outliers. Take moving least squares (MLS), which is one of the most effective approximation methods for scattered data, as an example. MLS has low requirements for the distribution of X, is suitable for any dimension d, and can approximate the sampled function with high efficiency and accuracy [1–3]. Despite its high accuracy, Levin [4] points out that the MLS results have a large error around the outliers.

In order to fit the data containing outliers, some scholars modify the classical methods. Levin [4] combines MLS with the moving least l_1 method [5], which is insensitive to outliers but give the discontinuous approximation results, and proposes a moving least Hardy method (MLH). He prove that MLH can give continuous approximation results and is insensitive to outliers. However, MLH approximation requires solving a nonlinear system, which makes it less efficient.

Detecting outliers in the data in advance and handling those data in a special way is a more general method. The research on outlier detection has been continued, and many outlier detection algorithms are proposed for different types of outliers in different problems. Most of those algorithms are based on the statistical analysis or machine learning [6–8]. Generally, the classical outlier detection algorithms have certain assumptions or requirements for data distribution. For example, Liu et al. [9] assume that data can be classified into several linear subspaces in the data space. And then they cluster the data based on the low-rank representation (LRR) in the dictionary space and the sparsity of outliers to automatically detect outliers and restore their values. However, in approximation problems, the sampled function f may be highly nonlinear, so it is difficult to divide the data into several linear subspaces, which would make Liu's method fail. In addition, because the deviations of outliers exist only in one dimension, if the data set with

outliers is considered in $\mathbb{R}^{d+1}$, the abnormity of outliers would be weaken. Take $d = 1$ as an example. In Fig. 1(a), the point corresponding to the red circle is the outlier generated by adding 0.5 deviation to the point corresponding to the blue circle, which seems obvious. However, it should be pointed out that the distance between the outlier and the point corresponding to the green circle is only 0.0821, and the those points are distributed as a nonlinear curve in $\mathbb{R}^2$, so that it is difficult to detect the outlier only based on the data. If d is raised, the abnormity of outliers would be further hidden by points' complex nonlinear distribution, which can be observed from the another set of data which are sampled from the nonlinear Franke's function and contain ten outliers shown in Fig. 1(b). For any d, if the classical outlier detection algorithms are employed to deal with the data for approximation, it is necessary to set the thresholds and parameters in algorithms carefully, which are not easy to use. Therefore, we need to use some special outlier detection algorithms for the data for approximation.

Zheng et al. [10] propose a modified moving least squares (MMLS) method to ef-



(a) The data sampled from the sine function containing one outlier

(b) The data sampled from the Franke's function containing ten outliers

Fig. 1: Outliers in data for approximation

ficiently approximate the scattered data containing outliers by detecting and weakening outliers according to the feature that outliers cause the dramatic changes of the approximate gradient around them. It is theoretically proved that MMLS can accurately find all outliers when the points are dense enough. Numerical experiments indicate that the efficiency of MMLS is superior to that of MLH. However, due to the limitation of obtaining approximate gradient, MMLS is only applicable to $X \subset \mathbb{R}^2$.

Amir and Levin [11] further analyze the approximation error bound of MLS near outliers by using the MLS quasi-interpolation (MLS-QI) scheme, and propose

the quasi-interpolation and outliers removal (QIOR) method to iteratively remove outliers. Under certain assumptions, they prove that QIOR could accurately detect all the outliers in data. In QIOR, each iteration needs to calculate the MLS approximation at all non-marked points, only the point with the largest approximation error is marked as an outlier, and then the next iteration is carried out. If either the number of the points in X or the number of outliers is a little large, the computational efficiency of QIOR is not high, obviously.

Inspired by QIOR, we construct a l_0-minimization problem with an inequality constrain to detect the sparse distributed outliers based on the high accuracy of MLS-QI and its sensitivity to outliers. Under certain assumptions, we prove theoretically that the deviation vector $\boldsymbol{E}^*$ is a solution to the optimization problem. Since the problem is consistent with the classic noise model in compressed sensing (CS) [12, 13], we introduce the classical orthogonal matching pursuit (OMP) algorithm to solve the optimization problem, efficiently. Obviously, if $\boldsymbol{E}^*$ is obtained, the outliers can be marked and their exact values are restored easily. The way to detect outliers by solving the proposed l_0-minimization problem is named l_0-ODR method. Numerical experiments show that in the examples on different dimensions, solving the l_0-minimization problem by OMP can detect outliers quickly and accurately, and the data can be approximately restored in a high accuracy. In these examples, l_0-ODR performs better than QIOR in efficiency and the detection accuracy. Compared with the Novel Local Outlier Detection (NLOD) algorithm [14], which is an effective density-based outlier detection method with only a robust parameter, l_0-ODR is in a higher accuracy and more sensitive to the outliers in approximation problems.

The rest of the paper is organized as follows. Section 2 briefly introduces the important preliminary knowledge: the moving least squares method and its quasi-interpolation scheme, and the orthogonal matching pursuit algorithm for solving the l_0-minimization problem. In Section 3, the l_0-minimization problem for detecting and restoring outliers is proposed, and the corresponding theoretical analysis is given. The numerical experiments in Section 4 show the effectiveness and robustness of l_0-ODR in detecting and restoring outliers in the approximated data. Section 5 concludes this paper.

2. Preliminaries

2.1. Moving least squares method and its quasi-interpolation scheme

For a positive integer d, let $\Omega \subseteq \mathbb{R}^d$ be a nonempty and bounded set. The scattered point set X consists of N distinct points in Ω. The data vector $\boldsymbol{F} =$

$(f_1, \cdots, f_N)^\top$ is sampled at X from an unknown function f defined on Ω. Through a given weight function θ, define

$$E_{2,\theta,\boldsymbol{x}}(p) = \sum_{i=1}^{N} (f_i - p(\boldsymbol{x}_i))^2 \theta(\boldsymbol{x}, \boldsymbol{x}_i). \tag{2.1}$$

Let $\mathbb{P}_m^d$ denote the polynomial space containing all polynomials whose degrees are less than or equal to m on $\mathbb{R}^d$ and its dimension is $Q = C_{m+d}^d$. Suppose a basis for $\mathbb{P}_m^d$ is $\{p_i(\boldsymbol{x})\}_{i=1}^Q$. For an given point $\boldsymbol{x}$ in Ω, the moving least squares approximation (MLS) in $\mathbb{P}_m^d$ is obtained by minimizing (2.1), and the approximation value is

$$s_{f,X} = \sum_{i=1}^{Q} c_i(\boldsymbol{x}) p_i(\boldsymbol{x}).$$

Here, $\boldsymbol{c}(\boldsymbol{x})$ is a solution to

$$\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}) \boldsymbol{P} \boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}) \boldsymbol{F}, \tag{2.2}$$

where

$$\boldsymbol{P} = \begin{pmatrix} p_1(\boldsymbol{x}_1) & \cdots & p_Q(\boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ p_1(\boldsymbol{x}_N) & \cdots & p_Q(\boldsymbol{x}_N) \end{pmatrix}$$

and $\boldsymbol{W}(\boldsymbol{x}) = \mathrm{diag}(\theta(\boldsymbol{x}, \boldsymbol{x}_1), \cdots, \theta(\boldsymbol{x}, \boldsymbol{x}_N))$.
Generally, the compactly supported weight function is applied to improve computational efficiency. In this paper, let $\theta(\boldsymbol{x}, \boldsymbol{y}) = \Phi_\delta(\boldsymbol{x} - \boldsymbol{y}) = \Phi((\boldsymbol{x} - \boldsymbol{y})/\delta)$, where $\Phi(\boldsymbol{x}) = \phi(\|\boldsymbol{x}\|_2),\ \boldsymbol{x} \in \mathbb{R}^d$ with a univariate and nonnegative function $\phi : [0, \infty) \to \mathbb{R}$ that is positive on $[0, 1/2]$ and supported in $[0, 1]$. Further, for an given δ and any i, let $w_i(\boldsymbol{x}) = \Phi_\delta(\boldsymbol{x} - \boldsymbol{x}_i)$ which is supported in the ball $B(\boldsymbol{x}, \delta)$. For any $\boldsymbol{x} \in \mathbb{R}^d$, define the index set

$$I_{\delta,X}(\boldsymbol{x}) := \{j \in \{1, \cdots, N\} : \|\boldsymbol{x} - \boldsymbol{x}_j\|_2 < \delta, \boldsymbol{x}_j \in X\}. \tag{2.3}$$

In order to ensure that (2.2) has a unique solution, some assumptions should be added in the theoretical analysis of MLS. First, we introduce the definition of the unisolvent set.

Definition 2.1. *For a given set $X = \{\boldsymbol{x}_i\}_{i=1}^N$ with $N \geq Q$, if the zero polynomial is the only polynomial from $\mathbb{P}_m^d$, then the set X is called $\mathbb{P}_m^d$-unisolvent.*

Then we give the classical assumption in MLS as following.

Assumption 2.1. *With the given δ and X, for all $\boldsymbol{x} \in \Omega$, the point set $\{\boldsymbol{x}_j : j \in I_{\delta,X}(\boldsymbol{x})\}$ is $\mathbb{P}_m^d$-unisolvent.*

If Assumption 2.1 holds, $s_{f,X}(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x})(\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x})\boldsymbol{P})^{-1}\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x})\boldsymbol{F}$, where the vector $\boldsymbol{b}(\boldsymbol{x})$ is composed of the basis for $\mathbb{P}_m^d$, i.e., $\boldsymbol{b}(\boldsymbol{x}) = (p_1(\boldsymbol{x}), \cdots, p_Q(\boldsymbol{x}))$.

Remark 2.1. *For two different points $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$, $\boldsymbol{c}(\boldsymbol{x}_a)$ and $\boldsymbol{c}(\boldsymbol{x}_b)$ need to be solved by (2.2), respectively. Therefore, although the approximation function space is $\mathbb{P}_m^d$, $s_{f,X}(\boldsymbol{x}) \notin \mathbb{P}_m^d$, which is one of the meanings of "moving".*

Let $\boldsymbol{a}(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x})(\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x})\boldsymbol{P})^{-1}\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x})$, then

$$s_{f,X}(\boldsymbol{x}) = \sum_{i=1}^{N} a_i(\boldsymbol{x}) f_i. \tag{2.4}$$

The formula (2.4) is called the moving least squares quasi-interpolation (MLS-QI) scheme, where $a_i(\boldsymbol{x})$ is called the i-th shape function. The separation between the sample values $\boldsymbol{F}$ and the approximated point $\boldsymbol{x}$ makes MLS-QI widely used in theory and application.
Furthermore, for any i, there is

$$a_i(\boldsymbol{x}) = w_i(\boldsymbol{x})\boldsymbol{b}(\boldsymbol{x})(\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x})\boldsymbol{P})^{-1}\boldsymbol{b}(\boldsymbol{x}_i)^\top, \tag{2.5}$$

which means that $a_i(\boldsymbol{x})$ is compactly supported and its support is contained in that of $w_i(\boldsymbol{x})$. A brief analysis shows that for a given $\mathbb{P}_m^d$, the same $a_i(\boldsymbol{x})$ can be derived from the different bases $\boldsymbol{b}(\boldsymbol{x})$ and $\tilde{\boldsymbol{b}}(\boldsymbol{x})$. MLS and MLS-QI are essentially equivalent, but their expressions are different.
Because MLS can continuously approximate the data in arbitrary dimensional space in high accuracy, it is widely applied in many fields of research and engineering [15–18], especially in the study of meshless methods for the numerical solution of partial differential equations. More details about MLS and MLS-QI can be found in the books [1, 2].

2.2. Orthogonal matching pursuit algorithm

Compressed sensing (CS) [12, 13] is a famous technique in sparse signal recovery, whose main idea is to use nonlinear optimization method to restore sparse signals by sampling as little as possible. Its original recovery model is

$$\min_{\boldsymbol{c}\in\mathbb{C}^n} \|\boldsymbol{c}\|_0 \quad \text{s.t.} \quad \Phi\boldsymbol{c} = \boldsymbol{b}, \tag{2.6}$$

where $\|\boldsymbol{c}\|_0$ represents the number of non-zero elements in $\boldsymbol{c}$, $\Phi \in \mathbb{C}^{m\times n}$ is a measurement matrix, and $\boldsymbol{b} = (b_1, \cdots, b_m)^\top \in \mathbb{C}^m$ is an observation vector. Generally, the constraint $\Phi\boldsymbol{c} = \boldsymbol{b}$ can be an underdetermined system of equations. If there may be some noise in sampling, CS uses the following noisy recovery model

$$\min_{\boldsymbol{c}\in\mathbb{C}^n} \|\boldsymbol{c}\|_0 \quad \text{s.t.} \quad \|\Phi\boldsymbol{c} - \boldsymbol{b}\|_2 \leq \eta. \tag{2.7}$$

Unfortunately, both the optimization problem (2.6) and (2.7) are NP-hard problems. In order to solve them efficiently, the greedy algorithms are applied as the major method in CS. The orthogonal matching pursuit (OMP) algorithm is one of the most classical greedy algorithms to solve (2.6) and (2.7).

OMP iteratively selects some rows of Φ to approximate the unknown sparse vector $\boldsymbol{c}$. At every iteration, OMP calculates the absolute value of the inner product between the residual vector and the each columns in Φ, then the column with the maximum absolute value is added into the selected columns. Before doing next iteration, the residual vector is updated by the least square residuals of $\boldsymbol{c}$ in the space spanned by the selected columns of Φ. Generally, the termination condition of OMP can be set according to the sparsity s of the current approximation vector $\boldsymbol{c}^*$, or whether the norm of the residual vector reaches a given bound. No matter what the termination condition, the algorithm selects one column at each iteration so that the solution is naturally sparse.

As a kind of the greedy algorithm, OMP does not always obtain the exact solutiosn of the problem (2.6) and (2.7), theoretically. However, some studies [12, 19, 20] proves that under certain conditions, OMP can obtain the exact solutions of the problem (2.6) and (2.7) with high probability, which makes it one of the most classical methods in the field of CS.

3. Outlier detection and recovery

In this section, the analysis is based on the assumption that outlier points are sparse in the scattered point set X. Based on the sensitivity to outliers and high accuracy of MLS-QI, We construct a l_0-minimization problem. MLS-QI in this section employ the compactly supported weight function $\Phi_\delta(\cdot)$ whose radius of support is δ and is based on the polynomial space $\mathbb{P}_m^d$ which contains all polynomials whose degrees are less than or equal to m in $\mathbb{R}^d$. The error functional and the approximation function of MLS-QI is referred to (2.1) and (2.4) respectively.

In addition to proposing the l_0-minimization problem, we prove that the unknown deviation vector $\boldsymbol{E}$ is a solution to the problem under some conditions, which

means that solving the problem can detect outlier and restore their exact values, and name the way l_0-ODR method.

3.1. The error bound of MLS-QI

In the following main analysis, in addition to the preliminaries about MLS introduced in Section 2.1, we need to use some results about the error bound of MLS-QI. The most recent result on its error bound is given by Mirzaei [21] in fractional order Sobolev Spaces. However, fractional order Sobolev Spaces are not involved in this paper, so we introduce the classical error bound given by Wendland in Chapters 3-4 of the book [2] by local polynomial reproduction. To apply the results, we need to add some assumptions about the domain Ω and the scattered point set X.

Definition 3.1. *If there exists an angle $\theta \in (0,\ \pi/2)$ and a radius $r > 0$ such that for every $\boldsymbol{x} \in \Omega$ a unit vector $\boldsymbol{\xi}(\boldsymbol{x})$ exists such that the high-dimensional cone*

$$C(\boldsymbol{x}, \boldsymbol{\xi}(\boldsymbol{x}), \theta, r) := \{\boldsymbol{x} + \lambda \boldsymbol{y} : \boldsymbol{y} \in \mathbb{R}^d,\ \|\boldsymbol{y}\|_2 = 1,\ \boldsymbol{y}^\top \boldsymbol{\xi}(\boldsymbol{x}) \geq \cos\theta,\ \lambda \in [0,\ r]\}$$

is contained in Ω, then Ω is said to satisfy an interior cone condition with angle θ and radius r.

For a point set $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ in a bounded domain Ω, the fill distance is defined to be

$$h_{X,\Omega} := \sup_{\boldsymbol{x} \in \Omega} \min_{1 \leq i \leq N} \|\boldsymbol{x} - \boldsymbol{x}_i\|_2,$$

and the separation distance is defined by

$$q_X := \frac{1}{2} \min_{i \neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2.$$

Under the two distances, the point set X is said to be quasi-uniform with respect to a constant $c_{qu} > 0$ if

$$q_X \leq h_{X,\Omega} \leq c_{qu} q_X,$$

which means the two distances are equivalent and is beneficial in analysis.
With the above notations and definitions, Wendland [2] gives the following error bound of MLS in $\mathbb{P}_m^d$.

Lemma 3.1. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is compacted and satisfies an interior cone condition with angle $\theta \in (0,\ \pi/2)$ and radius $r > 0$. Fix $m \in \mathbb{N}$. There exists constants $C_1,\ C_2$ that can be calculated exactly. Let $h_0 = r/C_2$. Define Ω^* to be*

the closure of $\bigcup_{\boldsymbol{x}\in\Omega} B(\boldsymbol{x}, 2C_2h_0)$. *For any quasi-uniform* $X \subseteq \Omega$ *with* $h_{X,\Omega} \leq h_0$, *let* $\delta = 2C_2h_{X\Omega}$. *If* X *satisfies Assumption 2.1, then for all* $f \in C^{m+1}(\Omega^*)$ *there is*

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \leq C_1 h_{X,\Omega}^{m+1} |f|_{C^{m+1}(\Omega^*)},$$

where $s_{f,X}$ *is MLS approximation based on* $\mathbb{P}_m^d$ *and the compactly supported weight function whose radius of support is* δ, *and* $|f|_{C^{m+1}(\Omega^*)} := \max_{|\alpha|=m+1} \|D^\alpha f\|_{L_\infty(\Omega^*)}$.

Constant C_1 depends on the weight function and some constants such as the dimension d, and C_2 is

$$C_2 = \frac{16(1+\sin\theta)^2 m^2}{3\sin^2\theta}. \tag{3.1}$$

In following analysis, we need to use a special distance, defined as

Definition 3.2. *For a set of points* $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ *in a bounded domain* $\Omega \subseteq \mathbb{R}^d$, *the full fill distance is defined to be*

$$h_F = \max_{1\leq i\leq N} h_{X_i,\Omega},$$

where $h_{X_i,\Omega}$ *is the fill distance of* $X_i = X \setminus \{\boldsymbol{x}_i\}$ *in* Ω.

Obviously, $h_F \geq h_{X,\Omega}$ and intuitively if X is sufficiently dense, then h_F is not much different from $h_{X,\Omega}$. In order to make h_F strictly equivalent with $h_{X,\Omega}$, we define a new case.

Definition 3.3. *As* X *gets denser, if there exists a constant* $c_{sf} > 0$, *s.t.*

$$h_{X,\Omega} \leq h_F \leq c_{sf} h_{X,\Omega},$$

then the point set X *is said to be slowly filling in* Ω.

In order to show that the defined case makes sense, a sufficient condition satisfying the case in Definition 3.3 is given here.

Proposition 3.1. *If the domain* $\Omega \subseteq \mathbb{R}^d$ *is a convex set, then for all* $X \subseteq \Omega$ *there is*

$$h_{X,\Omega} \leq h_F \leq 3h_{X,\Omega}, \tag{3.2}$$

i.e., X *is slowly filling in* Ω.

Since Proposition 3.1 is not important for the following analysis, we place its proof in Appendix A.
Based on Lemma 3.1, we give an error bound of MLS described by h_F.

Corollary 3.1. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is compacted and satisfies an interior cone condition with angle $\theta \in (0,\ \pi/2)$ and radius $r > 0$. Fix $m \in \mathbb{N}$. There exists constants C_1, C_2 that can be calculated exactly. Let $h_0 = r/C_2$. Ω^* is defined in Lemma 3.1. For any quasi-uniform and slowly filling $X \subseteq \Omega$ with $h_F \leq h_0$, let $\delta = 2C_2 h_F$. If X satisfies Assumption 2.1, then for all $f \in C^{m+1}(\Omega^*)$ there is*

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \leq C_1 h_F^{m+1} |f|_{C^{m+1}(\Omega^*)}.$$

In Corollary 3.1, constant C_2 is the same as that in Lemma 3.1, which is (3.1). C_1 is a little bit different from that in Lemma 3.1, but it still only depends on the weight function and the same constants. Comparing Corollary 3.1 with Lemma 3.1, we point out that in addition to the used distance, the radius of support of the weight function in MLS is changed accordingly, too. The proof of Corollary 3.1 is consistent with that of Lemma 3.1 and is omitted here to avoid duplicating Wendland's work. The interested reader can refer to Chapters 3-4 of the book [2]. In the case that X is quasi-uniform and slowly filling in Ω, the number N of sampled points can be bound by h_F, as follow.

Proposition 3.2. *If X is quasi-uniform and slowly filling respect to constant c_{qu} and c_{sf} in Ω, then there exists a constant C_F such that*

$$N \leq \frac{C_F}{h_F^d}.$$

Let $d_\Omega = \sup_{\boldsymbol{x},\boldsymbol{y}\in\Omega} \|\boldsymbol{x} - \boldsymbol{y}\|_2$ and define $\Omega_M := \bigcup_{\boldsymbol{x}\in\Omega} B(\boldsymbol{x}, d_\Omega)$. It is clear that $\Omega \subseteq \Omega_M$. Due to $X \subseteq \Omega$ and $q_X \leq d_\Omega$, there is $\bigcup_{\boldsymbol{x}_i\in X} B(\boldsymbol{x}_i, q_X) \subseteq \Omega_M$.
Based on the definition of the separation distance, we know that if $i \neq j$, then $B(\boldsymbol{x}_i, q_X) \bigcap B(\boldsymbol{x}_j, q_X) = \emptyset$. It follows that

$$\begin{aligned} vol(\bigcup_{\boldsymbol{x}_i\in X} B(\boldsymbol{x}_i, q_X)) &= \sum_{i=1}^{N} vol(B(\boldsymbol{x}_i, q_X)) \\ &= N q_X^d \cdot vol(B(0,1)) \\ &\leq vol(\Omega_M), \end{aligned}$$

and

$$N \leq \frac{vol(\Omega_M)}{vol(B(0,1))} \cdot \frac{1}{q_X^d}.$$

Because X is quasi-uniform with respect to c_{qu} such that $h_{X,\Omega} \leq c_{qu} q_X$,

$$N \leq \frac{c_{qu}^d \cdot vol(\Omega_M)}{vol(B(0,1))} \cdot \frac{1}{h_{X,\Omega}^d}.$$

Further, because X is slowly filling with respect to c_{sf}, then

$$N \leq \frac{c_{qu}^d \cdot c_{sf}^d \cdot vol(\Omega_M)}{vol(B(0,1))} \cdot \frac{1}{h_F^d}.$$

Obviously, let $C_F = c_{qu}^d \cdot c_{sf}^d \cdot vol(\Omega_M)/vol(B(0,1))$, which completes the proof.

Next, we analyze the upper bound of the l_2 error of MLS-QI on X. According to MLS-QI (2.4), the vector composed of the MLS approximation on X can be written as

$$\boldsymbol{F_Q} = \begin{pmatrix} \boldsymbol{a}(\boldsymbol{x}_1)\boldsymbol{F} \\ \vdots \\ \boldsymbol{a}(\boldsymbol{x}_N)\boldsymbol{F} \end{pmatrix} = \boldsymbol{A}\boldsymbol{F}, \tag{3.3}$$

where matrix $\boldsymbol{A}$ is composed of the values of each shape function (2.5) on X, i.e., $(\boldsymbol{A})_{i,j} = a_j(\boldsymbol{x}_i) = w_j(\boldsymbol{x}_i)\boldsymbol{b}(\boldsymbol{x}_j)(\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}_i)\boldsymbol{P})^{-1}\boldsymbol{b}(\boldsymbol{x}_i)^\top$. Let $\boldsymbol{R}$ denote the error vector, i.e., $\boldsymbol{R} = \boldsymbol{F} - \boldsymbol{F_Q}$. Based on Corollary 3.1, there is

$$|R_i| \leq C_1 h_F^{m+1} |f|_{C^{m+1}(\Omega^*)}, \ \forall i = 1, \cdots N.$$

It follows from Proposition 3.2 that there is

$$\begin{aligned} \|\boldsymbol{R}\|_2 &\leq \sqrt{N} C_1 h_F^{m+1} |f|_{C^{m+1}(\Omega^*)} \\ &\leq C_1 \sqrt{C_F} h_F^{m+1-d/2} |f|_{C^{m+1}(\Omega^*)}. \end{aligned}$$

To avoid too many symbols, we let $\tilde{C}_1 = C_1\sqrt{C_F}$, then the l_2 bound is

$$\|\boldsymbol{R}\|_2 \leq \tilde{C}_1 h_F^{m+1-d/2} |f|_{C^{m+1}(\Omega^*)}. \tag{3.4}$$

3.2. The derivation and properties of the l_0-minimization problem

In this section, according to (3.4), we construct a l_0-minimization problem to detect outliers and restore their exact values.
A small number of outliers in scattered data pollute the sampling on X into $\tilde{\boldsymbol{F}} = (f_1 + e_1^*, \cdots, f_N + e_N^*)^\top = \boldsymbol{F} + \boldsymbol{E}^*$. Suppose the outlier point set is

$$\mathscr{O}_\varepsilon = \{\boldsymbol{x}_i \in X : |e_i^*| > \varepsilon\}, \tag{3.5}$$

and the corresponding index set is written as I_ε.
Generally, both which points are outliers and their deviation values are unknown,

which are the focuses of this paper. For the convenience of analysis, we assume that there is no small noise except outlier in this paper, i.e., $supp(\boldsymbol{E}^*) = I_\varepsilon$, where $supp(\boldsymbol{E}^*)$ denotes the index set of the non-zero elements of $\boldsymbol{E}^*$.
Because of the linearity of MLS-QI (3.3), its approximation vector $\tilde{\boldsymbol{F}}_{\boldsymbol{Q}}$ obtained by the polluted data $\tilde{\boldsymbol{F}}$ can be written as

$$\begin{aligned}\tilde{\boldsymbol{F}}_{\boldsymbol{Q}} &= \boldsymbol{A}\tilde{\boldsymbol{F}} \\ &= \boldsymbol{F}_{\boldsymbol{Q}} + \boldsymbol{A}\boldsymbol{E}^*.\end{aligned}$$

In this case, the error vector on the polluted data $\tilde{\boldsymbol{F}}$ is

$$\begin{aligned}\tilde{\boldsymbol{F}} - \tilde{\boldsymbol{F}}_{\boldsymbol{Q}} &= \tilde{\boldsymbol{F}} - \boldsymbol{A}\tilde{\boldsymbol{F}} \\ &= (\boldsymbol{F} - \boldsymbol{F}_{\boldsymbol{Q}}) + (\boldsymbol{E}^* - \boldsymbol{A}\boldsymbol{E}^*) \\ &= \boldsymbol{R} + (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{E}^*.\end{aligned} \tag{3.6}$$

Combining the second row with the third row in equation (3.6), we can get $(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{E}^* - (\boldsymbol{I} - \boldsymbol{A})\tilde{\boldsymbol{F}} = -\boldsymbol{R}$. It follows from (3.4) that there is

$$\|(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{E}^* - (\boldsymbol{I} - \boldsymbol{A})\tilde{\boldsymbol{F}}\|_2 \leq \tilde{C}_1 h_F^{m+1-d/2} |f|_{C^{m+1}(\Omega^*)}.$$

It is worth noting that all other quantities on the left-hand side of (3.2) except $\boldsymbol{E}^*$ are known after the problem is given. Because outliers are few in the data, $\boldsymbol{E}^*$ is a sparse vector. Based on the sparsity, we define the following l_0-optimization problem

$$\begin{aligned}\min_{\boldsymbol{E}\in\mathbb{R}^N} \quad & \|\boldsymbol{E}\|_0 \\ \text{subject to :} \quad & \left\|\boldsymbol{U}\boldsymbol{E} - \tilde{\boldsymbol{R}}\right\|_2 \leq \eta,\end{aligned} \tag{3.7}$$

where $\boldsymbol{U} = \boldsymbol{I} - \boldsymbol{A}$, $\tilde{\boldsymbol{R}} = (\boldsymbol{I} - \boldsymbol{A})\tilde{\boldsymbol{F}}$, $\|\boldsymbol{E}\|_0$ represents the number of non-zero elements in $\boldsymbol{E}$, and η is an upper bound of error set by users.

Remark 3.1. *The optimization problem (3.7) is the same form as the noisy recovery model (2.7) in compressed sensing (CS) which is NP-hard. However, the $\boldsymbol{E}^*$ we need is a sparse solution to it, so in this paper, we only introduce the orthogonal matching pursuit (OMP) algorithm to solve it efficiently, without discussing the solvability of the problem. The detailed process for solving the problem (3.7) is in Section 3.4, and the feasibility and effect of our approach are demonstrated through numerical experiments in Section 4.*

Let $\mathscr{E}_\eta$ denote the solution set of the optimization problem (3.7). Since this problem is constructed, if $\eta = \tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)}$, then the constraint domain is nonempty and the optimal solution exists, i.e., $\mathscr{E}_\eta \neq \emptyset$. Next we show the conditions which can make the true unknown deviation vector $\boldsymbol{E}^*$ contained in $\mathscr{E}_\eta$.

Theorem 3.1. *If the conditions in Corollary 3.1 hold, we further suppose that the separation distance q_O of $\mathscr{O}_\varepsilon$ satisfies $q_O \geq \delta$ and there exists a constant $C_M < 1$ independent of h_F such that for every $i = 1, \cdots, N$ there is $0 \leq (\boldsymbol{A})_{i,i} \leq C_M$. Then if the upper bound of error, η, satisfies*

$$\tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)} \leq \eta < (1-C_M)\varepsilon - \tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)}, \tag{3.8}$$

the unknown deviation vector $\boldsymbol{E}^*$ is in $\mathscr{E}_\eta$.

As in the above analysis, the first inequality in (3.8) makes $\left\|\boldsymbol{U}\boldsymbol{E}^* - \tilde{\boldsymbol{R}}\right\|_2 \leq \eta$ and $\mathscr{E}_\eta \neq \emptyset$ true. Let $\boldsymbol{E}$ be the arbitrary one in $\mathscr{E}_\eta$. In the case of $\eta < (1-C_M)\varepsilon - \tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)}$, we suppose $\boldsymbol{E}^* \notin \mathscr{E}_\eta$. Obviously, $\left\|\boldsymbol{U}\boldsymbol{E} - \tilde{\boldsymbol{R}}\right\|_2 \leq \eta$ and $\|\boldsymbol{E}\|_0 < \|\boldsymbol{E}^*\|_0$. Without loss of generality, we suppose the number of outliers is K, then $\|\boldsymbol{E}\|_0 < \|\boldsymbol{E}^*\|_0 = K$. Due to $\delta \leq q_O$, $\{B(\boldsymbol{x}_j, \delta)\}_{j \in I_\varepsilon}$ are K disjoint open balls. Since $\|\boldsymbol{E}\|_0 < K$, there exists k such that $B(\boldsymbol{x}_k, \delta)$ do not contain the points corresponding to $\boldsymbol{E}$, i.e., $B(\boldsymbol{x}_k, \delta) \cap \{\boldsymbol{x}_i \in X : i \in supp(\boldsymbol{E})\} = \emptyset$.
Consider

$$\begin{aligned}\left\|\boldsymbol{U}\boldsymbol{E} - \tilde{\boldsymbol{R}}\right\|_2 &= \left\|\boldsymbol{U}\boldsymbol{E} - \boldsymbol{U}\tilde{\boldsymbol{F}}\right\|_2 \\ &= \|\boldsymbol{U}\boldsymbol{E} - \boldsymbol{U}(\boldsymbol{F} + \boldsymbol{E}^*)\|_2 \\ &\geq \|\boldsymbol{U}(\boldsymbol{E} - \boldsymbol{E}^*)\|_2 - \|\boldsymbol{R}\|_2.\end{aligned}$$

It is clear $\|\boldsymbol{U}(\boldsymbol{E} - \boldsymbol{E}^*)\|_2 \geq |\boldsymbol{U}_k(\boldsymbol{E} - \boldsymbol{E}^*)|$, where $\boldsymbol{U}_k$ denotes the k-th row of the matrix $\boldsymbol{U}$. Recall that $(\boldsymbol{A})_{k,j} = w_j(\boldsymbol{x}_k)\boldsymbol{b}(\boldsymbol{x}_j)(\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}_k)\boldsymbol{P})^{-1}\boldsymbol{b}(\boldsymbol{x}_k)^\top$, and it follows from $\boldsymbol{U} = \boldsymbol{I} - \boldsymbol{A}$ that if $\|\boldsymbol{x}_k - \boldsymbol{x}_j\|_2 > \delta$, then $(\boldsymbol{U})_{k,j} = 0$. If $\|\boldsymbol{x}_k - \boldsymbol{x}_j\|_2 \leq \delta$, due to $B(\boldsymbol{x}_k, \delta) \cap \{\boldsymbol{x}_i \in X : i \in supp(\boldsymbol{E})\} = \emptyset$, there is $E_j = 0$. Therefore, for every j, $(\boldsymbol{U})_{k,j}\boldsymbol{E}_j = 0$ always holds, which implies

$$\boldsymbol{U}_k \boldsymbol{E} = 0.$$

Based on the conditions of Theorem 3.1 and the definition of outliers (3.5), there is

$$\begin{aligned}\|\boldsymbol{U}_k \boldsymbol{E}^*\| &= \|(\boldsymbol{U})_{k,k} e_k^*\| \\ &= \|(1 - (\boldsymbol{A})_{k,k}) e_k^*\| \\ &\geq (1 - C_M)\varepsilon.\end{aligned}$$

It follows from (3.4), we get

$$\left\|\boldsymbol{U}\boldsymbol{E}-\tilde{\boldsymbol{R}}\right\|_2 \geq (1-C_M)\varepsilon - \tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)},$$

which implies $\eta \geq (1-C_M)\varepsilon - \tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)}$ and leads to a contradiction to (3.8).

For any given $\varepsilon > 0$ and $\eta > 0$, obviously, there exists a sufficiently small h_F such that (3.8) holds. Theorem 3.1 is a sufficient condition for $\boldsymbol{E}^* \in \mathscr{E}_\eta$, and the case $\tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)} > (1-C_M)\varepsilon - \tilde{C}_1 h_F^{m+1-d/2}|f|_{C^{m+1}(\Omega^*)}$ is beyond its scope. According to the above proof, the following corollary can be easily derived.

Corollary 3.2. *If the conditions in Theorem 3.1 hold, we further suppose that (3.8) holds. Then for all $\boldsymbol{E} \in \mathscr{E}_\eta$, the point set $\mathscr{O}' = \{\boldsymbol{x}_i \in X : i \in supp(\boldsymbol{E})\}$ contains the same number of points as $\mathscr{O}_\varepsilon$. Further, for all $\boldsymbol{x}_{i^*} \in \mathscr{O}_\varepsilon$, there exist only one point $\boldsymbol{x}_i \in \mathscr{O}'$ with $\|\boldsymbol{x}_{i^*} - \boldsymbol{x}_i\|_2 < \delta$.*

Corollary 3.2 implies that although there may be multiple solutions to (3.7), the number of outliers corresponding to these solutions is correct in the case that (3.8) holds, and each of these points is in the δ-neighborhood of a true outlier point. It is worth noting that in Corollary 3.1 there is $\delta = 2C_2 h_F$, so we can write Theorem 3.1 as the following limit form.

Corollary 3.3. *If the conditions in Theorem 3.1 hold, we further suppose that as $h_F \to 0$, q_O always satisfies $q_O \geq 2C_2 h_F$ and there exists a constant C_M independent of h_F such that $0 \leq (\boldsymbol{A})_{i,i} \leq C_M$ for all i. Then for all given $\delta_0 > 0$ and $\eta \in (0,\ (1-C_M)\varepsilon]$, there exists h_0 such that if $h_F \leq h_0$, then $\boldsymbol{E}^* \in \mathscr{E}_\eta$ and for all $\boldsymbol{E} \in \mathscr{E}_\eta$ there is $\|\boldsymbol{E}\|_0 = \|\boldsymbol{E}^*\|_0$. In the case, for all $\boldsymbol{x}_{i^*} \in \mathscr{O}_\varepsilon$, there exist only one points $\boldsymbol{x}_i \in \mathscr{O}'$ with $\|\boldsymbol{x}_{i^*} - \boldsymbol{x}_i\|_2 < \delta_0$.*

Based on Corollary 3.3, the assumption that outliers are few in data is essentially that the distribution of outlier points in $\mathbb{R}^d$ is sparse, which is slightly different from the definition of sparse vectors in CS, but they can be used without distinction in this paper.

Remark 3.2. *Corollary 3.3 shows that in the limit, we only need to let $\eta \in (0,\ (1-C_M)\varepsilon]$, then we can approximately detect outlier points by solving the optimization problem (3.7). However, in specific problems, the selection of η may require some experimentation. We discuss how to set η in Section 3.4.*

In Theorem 3.1, Corollary 3.2 and 3.3, we suppose that the diagonal elements of the matrix $\boldsymbol{A}$ have an upper bound C_M which is independent of h_F and strictly less than 1, which is not obvious but important for our analysis. Therefore we prove it in the next section. The process of the proof also reveals why we use the distance h_F instead of the more common $h_{X,\Omega}$.

3.3. The uniform bound of the diagonal elements of the matrix $\boldsymbol{A}$

In both Lemma 3.1 and Corollary 3.1 we suppose that Ω satisfies a certain inner cone condition. Some of its geometric properties are given by the following lemma.

Lemma 3.2. *Suppose that $C = C(\boldsymbol{x}, \boldsymbol{\xi}, \theta, r) \subseteq \mathbb{R}^d$ is a cone defined as in Definition 3.1. Then for every $h \leq r/(1+\sin\theta)$, the closed ball $B = B(\boldsymbol{y}, h\sin\theta)$ with center $\boldsymbol{y} = \boldsymbol{x} + h\boldsymbol{\xi}$ is contained in $C(\boldsymbol{x}, \boldsymbol{\xi}, \theta, r)$. Further, for all $\boldsymbol{z} \in B$, the line segment $\boldsymbol{x} + t(\boldsymbol{z}-\boldsymbol{x})/\|\boldsymbol{z}-\boldsymbol{x}\|_2$, $t \in [0,\ r]$ is contained in the cone.*

The proof of Lemma 3.2, which comes from Lemma 3.7 in [2], is not complicated, so we omit it.

As stated in Section 2.1, based on the given $\mathbb{P}_m^d$, the same shape function $a_i(\boldsymbol{x})$ defined by (2.5) can be derived from the different bases $\boldsymbol{b}(\boldsymbol{x})$, therefore we introduce the shifted and scaled polynomial basis (SSPB)

$$\left\{ \frac{(\boldsymbol{x}-\boldsymbol{z})^{\boldsymbol{\alpha}}}{h^{|\boldsymbol{\alpha}|}} \right\}_{0 \leq |\boldsymbol{\alpha}| \leq m} \tag{3.9}$$

which can be obtained from the normal power basis $\{\boldsymbol{x}^{\boldsymbol{\alpha}}\}_{0\leq|\boldsymbol{\alpha}|\leq m}$ shifting and scaling. Obviously, (3.9) is a basis for $\mathbb{P}_m^d$ and we call it $(\boldsymbol{z}, h)$-SSPB. In the following, we default the basis of $\mathbb{P}_m^d$, $\{p_i(\boldsymbol{x})\}_{i=1}^{Q}$, is as (3.9).

Define a matrix

$$\boldsymbol{M}(\boldsymbol{x}) = \boldsymbol{P}^{\top}\boldsymbol{W}(\boldsymbol{x})\boldsymbol{P}, \tag{3.10}$$

where $(\boldsymbol{M}(\boldsymbol{x}))_{i,j} = \sum_{k=1}^{N} w_k(\boldsymbol{x}) p_i(\boldsymbol{x}_k) p_j(\boldsymbol{x}_k)$ and $w_k(\boldsymbol{x})$ is the weight function at $\boldsymbol{x}_k$, i.e., $w_k(\boldsymbol{x}) = \Phi_\delta(\boldsymbol{x}-\boldsymbol{x}_k)$. Since $w_k(\boldsymbol{x})$ is compactly supported, the uses of $\sum_{k=1}^{N}$ and $\sum_{k\in I_{\delta,X}(\boldsymbol{x})}$ are equivalent, where $I_{\delta,X}(\boldsymbol{x})$ is the indices of the points in X that are in the δ-neighborhood of $\boldsymbol{x}$ and defined by (2.3).

Mirzaei [21] proves that if the radius of support of the weight function is $\delta = 2C_2 h_{X,\Omega}$, then the smallest eigenvalue of $\boldsymbol{M}(\hat{\boldsymbol{x}})$ under $(\hat{\boldsymbol{x}}, h_{X,\Omega})$-SSPB has a lower bound independent of $\hat{\boldsymbol{x}}$ and $h_{X,\Omega}$. In this paper, in order to use h_F and let $\delta = 2C_2 h_F$, we give the following theorem using the same technique as Mirzaei's.

Theorem 3.2. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is compacted and satisfies an interior cone condition with angle $\theta \in (0,\ \pi/2)$ and radius $r > 0$. Fix $m \in \mathbb{N}$. C_2 is defined in (3.1) and let $h_0 = r/C_2$. For any given $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\} \subseteq \Omega$ with $h_{X,\Omega} \leq h_0$ and any $h \in [h_{X,\Omega},\ h_0]$, let $\boldsymbol{\delta} = 2C_2 h$. If for all $\hat{\boldsymbol{x}} \in \Omega$ the point set $\{\boldsymbol{x}_j \in X :\ j \in I_{\delta,X}(\hat{\boldsymbol{x}})\}$ is $\mathbb{P}_m^d$-unisolvent, then there exist a positive constant C_λ such that for all $\hat{\boldsymbol{x}} \in \Omega$, the smallest eigenvalue of $\boldsymbol{M}(\hat{\boldsymbol{x}})$ under $(\hat{\boldsymbol{x}}, h)$-SSPB satisfies*

$$\lambda_{min}(\boldsymbol{M}(\hat{\boldsymbol{x}})) \geq C_\lambda > 0.$$

Let $\hat{\boldsymbol{x}}$ be a fixed but arbitrary point in Ω. According to definition of the smallest eigenvalue, there is

$$\lambda_{min}(\boldsymbol{M}(\hat{\boldsymbol{x}})) = \min_{\boldsymbol{v} \in \mathbb{R}^Q \setminus 0} \frac{\boldsymbol{v}^\top \boldsymbol{M}(\hat{\boldsymbol{x}}) \boldsymbol{v}}{\boldsymbol{v}^\top \boldsymbol{v}}. \tag{3.11}$$

Without loss of generality, we let $\boldsymbol{v}^*$ minimize (3.11) and define a polynomial

$$\boldsymbol{\pi}(\boldsymbol{x}) := \sum_{|\boldsymbol{\alpha}| \leq m} v_{\boldsymbol{\alpha}}^* \frac{(\boldsymbol{x} - \hat{\boldsymbol{x}})^{\boldsymbol{\alpha}}}{h^{|\boldsymbol{\alpha}|}} = \sum_{i=1}^{Q} v_i^* p_i(\boldsymbol{x}).$$

We suppose that the number of points in $\{\boldsymbol{x}_j \in X :\ j \in I_{\delta,X}(\hat{\boldsymbol{x}})\}$ is $\hat{N}$. Since $\{\boldsymbol{x}_j \in X :\ j \in I_{\delta,X}(\hat{\boldsymbol{x}})\}$ is $\mathbb{P}_m^d$-unisolvent, the functionals $Z = \{\delta_{\boldsymbol{x}_j} :\ j \in I_{\delta,X}(\hat{\boldsymbol{x}})\}$ form a norming set for $\mathbb{P}_m^d$, i.e., there exists an injective mapping $T : \mathbb{P}_m^d \to T(\mathbb{P}_m^d) \subseteq \mathbb{R}^{\hat{N}}$, where $T(p) = (p(\boldsymbol{x}_j))_{j \in I_{\delta,X}(\hat{\boldsymbol{x}})},\ \forall p \in \mathbb{P}_m^d$. We equip $T(\mathbb{P}_m^d)$ with a special norm

$$\|T(p)\|_{2,\boldsymbol{w}}^2 := \sum_{j \in I_{\delta,X}(\hat{\boldsymbol{x}})} w_j(\hat{\boldsymbol{x}}) p^2(\boldsymbol{x}_j).$$

Since $\{w_j(\hat{\boldsymbol{x}})\}_{j \in I_{\delta,X}(\hat{\boldsymbol{x}})}$ are all positive and T is injective, it is easy to be verified that $\|\cdot\|_{2,\boldsymbol{w}}^2$ meets the basic definition of norm.
Obviously, $\boldsymbol{\pi} \in \mathbb{P}_m^d$ and

$$\begin{aligned}
\|T(\boldsymbol{\pi})\|_{2,\boldsymbol{w}}^2 &= \sum_{j \in I_{\delta,X}(\hat{\boldsymbol{x}})} w_j(\hat{\boldsymbol{x}}) \boldsymbol{\pi}^2(\boldsymbol{x}_j) \\
&= \sum_{j=1}^{N} w_j(\hat{\boldsymbol{x}}) \boldsymbol{\pi}^2(\boldsymbol{x}_j) \\
&= \boldsymbol{v}^{*\top} \boldsymbol{M}(\hat{\boldsymbol{x}}) \boldsymbol{v}^*.
\end{aligned}$$

Because T is injective, there exists $T^{-1}: T(\mathbb{P}_m^d) \to \mathbb{P}_m^d$. We define

$$\|T^{-1}\| := \sup_{\boldsymbol{z}\in T(\mathbb{P}_m^d)\setminus 0} \frac{\|T^{-1}(\boldsymbol{z})\|_{L_\infty(B(\hat{\boldsymbol{x}},\delta))}}{\|\boldsymbol{z}\|_{2,\boldsymbol{w}}} = \sup_{p\in\mathbb{P}_m^d\setminus 0} \frac{\|p\|_{L_\infty(B(\hat{\boldsymbol{x}},\delta))}}{\|T(p)\|_{2,\boldsymbol{w}}}. \tag{3.12}$$

Further,

$$\|T(\boldsymbol{\pi})\|_{2,\boldsymbol{w}} \geq \frac{1}{\|T^{-1}\|}\|\boldsymbol{\pi}\|_{L_\infty(B(\hat{\boldsymbol{x}},\delta))}. \tag{3.13}$$

It follows from $\delta = 2C_2h > h$ that $\partial B(\hat{\boldsymbol{x}},h) = \{\boldsymbol{x} : \|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_2 = h\} \subseteq B(\hat{\boldsymbol{x}},\delta)$, which implies $\|\boldsymbol{\pi}\|_{L_\infty(B(\hat{\boldsymbol{x}},\delta))} \geq \|\boldsymbol{\pi}\|_{L_\infty(\partial B(\hat{\boldsymbol{x}},h))}$. Let $\boldsymbol{y} = (\boldsymbol{x}-\hat{\boldsymbol{x}})/h$ and define $\tilde{\boldsymbol{\pi}}(\boldsymbol{y}) := \sum_{|\boldsymbol{\alpha}|\leq m} v_{\boldsymbol{\alpha}}^* \boldsymbol{y}^{\boldsymbol{\alpha}} = \boldsymbol{\pi}(\boldsymbol{x})$. It is clear that $\|\boldsymbol{\pi}\|_{L_\infty(\partial B(\hat{\boldsymbol{x}},h))} = \|\tilde{\boldsymbol{\pi}}\|_{L_\infty(\partial B(0,1))}$.
Because the normed space $\langle \mathbb{P}_m^d, \|\cdot\|_{L_\infty(\partial B(0,1))}\rangle$ is Q-dimensional, it is isomorphic and homeomorphic to $\mathbb{R}^Q$, which implies the norms on them are equivalent. Therefore, there exists a constant C_π independent of $\boldsymbol{v}^*$, h and $\hat{\boldsymbol{x}}$ such that

$$\|\boldsymbol{\pi}\|_{L_\infty(\partial B(\hat{\boldsymbol{x}},h))} = \|\tilde{\boldsymbol{\pi}}\|_{L_\infty(\partial B(0,1))} \geq C_\pi \|\boldsymbol{v}^*\|_2. \tag{3.14}$$

Next we begin to bound $\|T^{-1}\|$.
First, consider the case $m \neq 0$. Because $\delta = 2C_2h \leq 2r$ holds and Ω satisfies an interior cone condition, there exist $\boldsymbol{\xi}_1$ such that the cone $C(\hat{\boldsymbol{x}}) = C(\hat{\boldsymbol{x}},\boldsymbol{\xi}_1,\theta,\delta/2)$ is contained in $\Omega \cap B(\hat{\boldsymbol{x}},\delta/2)$. Based on Lemma 3.2, let $h_1 = \delta/(2(1+\sin\theta))$ then the closed ball

$$B_1 := B(\tilde{\boldsymbol{x}}_1, r_1) \subset C(\hat{\boldsymbol{x}}),$$

where $r_1 = h_1 \sin\theta$, and $\tilde{\boldsymbol{x}}_1 = \hat{\boldsymbol{x}} + h_1\boldsymbol{\xi}_1$. According to $B_1 \subset C(\hat{\boldsymbol{x}}) \subset B(\hat{\boldsymbol{x}},\delta)$ and the Bernstein inequality, for all $p \in \mathbb{P}_m^d$, there is

$$\|p\|_{L_\infty(B(\hat{\boldsymbol{x}},\delta))} \leq \left(\frac{4(1+\sin\theta)}{\sin\theta}\right)^m \|p\|_{L_\infty(B_1)}. \tag{3.15}$$

The detailed derivation of formula (3.15) can refer to the proof of Lemma B.1 in book [22] given by Melenk.
For any given $p \in \mathbb{P}_m^d$, we suppose $\boldsymbol{x}_M \in B_1$ make $|p(\boldsymbol{x}_M)| = \|p\|_{L_\infty(B_1)}$. Because the ball whose radius is r_1 satisfies the interior cone condition with angle $\pi/3$ and radius r_1, there exists $\boldsymbol{\xi}_2$ such that $C(\boldsymbol{x}_M) = C(\boldsymbol{x}_M,\boldsymbol{\xi}_2,\pi/3,r_1) \subset B_1$. Let $h_2 = \frac{r_1}{4m^2(1+\sin\frac{\pi}{3})}$, which is obvious that $h_2 \leq \frac{r_1}{1+\sin\frac{\pi}{3}}$. Based on Lemma 3.2, the closed ball

$$B_2 := B(\tilde{\boldsymbol{x}}_2, r_2) \subset C(\boldsymbol{x}_M),$$

where $r_2 = h_2 \sin\frac{\pi}{3}$, and $\tilde{\boldsymbol{x}}_2 = \boldsymbol{x}_M + h_2\boldsymbol{\xi}_2$.
According to $C_2 = \frac{16(1+\sin\theta)^2 m^2}{3\sin^2\theta}$ and $\theta \in (0, \pi/2)$, it is clear that

$$\begin{aligned} r_2 &= \frac{r_1 \sin\frac{\pi}{3}}{4m^2(1+\sin\frac{\pi}{3})} \frac{\sin\theta}{1+\sin\theta} C_2 h \\ &= \frac{4}{3}\frac{\sqrt{3}}{2+\sqrt{3}}\frac{1+\sin\theta}{\sin\theta} h \\ &\geq h_{X,\Omega}. \end{aligned}$$

Based on the definition of $h_{X,\Omega}$, there exists $\boldsymbol{x}_k \in X \cap B_2$. Without loss of generality, we suppose $\boldsymbol{x}_k \neq \boldsymbol{x}_M$ and let $L = \|\boldsymbol{x}_k - \boldsymbol{x}_M\|_2$. It is clear that $L \leq \|\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_2\|_2 + \|\tilde{\boldsymbol{x}}_2 - \boldsymbol{x}_M\|_2 \leq r_2 + h_2$. Since $\boldsymbol{x}_k$ is in $C(\boldsymbol{x}_M)$, the segment line

$$\boldsymbol{x}_M + t\frac{\boldsymbol{x}_k - \boldsymbol{x}_M}{L}, \quad t \in [0,\ r_1],$$

is contained in $C(\boldsymbol{x}_M) \subset B_1$. And define a function

$$\tilde{p}(t) := p\left(\boldsymbol{x}_M + t\frac{\boldsymbol{x}_k - \boldsymbol{x}_M}{L}\right), \quad t \in [0,\ r_1].$$

Since p is in $\mathbb{P}^d_m$, $\tilde{p}$ is in $\mathbb{P}^1_m$. We introduce a simple form of Markov's inequality proposed in the proof of Theorem 3.8 in book [2], as

$$|\tilde{p}'(t)| \leq \frac{2}{r_1} m^2 \|\tilde{p}\|_{L_\infty([0,r_1])}, \quad t \in [0,\ r_1],$$

which holds for all $p \in \mathbb{P}^1_m$ and all $r_1 > 0$. Hence there is

$$\begin{aligned} |p(\boldsymbol{x}_M) - p(\boldsymbol{x}_k)| &= |\int_0^L \tilde{p}'(t)dt| \\ &\leq \int_0^L |\tilde{p}'(t)|dt \\ &\leq L\frac{2}{r_1}m^2\|\tilde{p}\|_{L_\infty([0,r_1])} \\ &\leq 2m^2(1+\sin\frac{\pi}{3})\frac{h_2}{r_1} \\ &\leq \frac{1}{2}|p(\boldsymbol{x}_M)|, \end{aligned}$$

which implies

$$|p(\boldsymbol{x}_k)| \geq \frac{1}{2}|p(\boldsymbol{x}_M)|. \tag{3.16}$$

If $\boldsymbol{x}_k = \boldsymbol{x}_M$, (3.16) holds obviously. Let $C_{\theta,m} := 2\left(\frac{4(1+\sin\theta)}{\sin\theta}\right)^m$ and substitute (3.16) into (3.15), there is

$$\|p\|_{L_\infty(B(\hat{\boldsymbol{x}},\delta))} \leq C_{\theta,m}|p(\boldsymbol{x}_k)|. \tag{3.17}$$

As stated in Section 2.1, the weight function $\Phi_\delta(\cdot)$ is generated by a univariate and nonnegative function ϕ which is positive on $[0,1/2]$ and supported in $[0,1]$, so we can define

$$w_{min} := \min_{s\in[0,\ 1/2]} \phi(s) > 0. \tag{3.18}$$

According to (3.17) and (3.18), there is

$$w_{min}\|p\|^2_{L_\infty(B(\hat{\boldsymbol{x}},\delta))} \leq C^2_{\theta,m} w_k(\hat{\boldsymbol{x}})|p(\boldsymbol{x}_k)|^2 \leq C^2_{\theta,m}\sum_{i=1}^{N} w_i(\hat{\boldsymbol{x}})|p(\boldsymbol{x}_i)|^2 = C^2_{\theta,m}\|T(p)\|_{2,\boldsymbol{w}},$$

for all $p \in \mathbb{P}^d_m$ with $m \neq 0$. Then, according to (3.12), it is clear that

$$\|T^{-1}\| \leq \frac{C_{\theta,m}}{\sqrt{w_{min}}}.$$

Consider the case of $m = 0$. Then $\mathbb{P}^d_m$ is composed of constant value functions, which implies

$$\|T^{-1}\| = \frac{1}{\sqrt{\sum_{j\in I_{\delta,X}(\hat{\boldsymbol{x}})} w_j(\hat{\boldsymbol{x}})}} \leq \frac{1}{\sqrt{w_{min}}}.$$

It is clear that $C_{\theta,m} \geq 1$ holds for all $m \in \mathbb{N}$, therefore $\|T^{-1}\| \leq \frac{C_{\theta,m}}{\sqrt{w_{min}}}$ always holds.

Recall (3.13) and (3.14), there is

$$\lambda_{min}(\boldsymbol{M}(\hat{\boldsymbol{x}})) = \frac{\boldsymbol{v}^{*\top}\boldsymbol{M}(\hat{\boldsymbol{x}})\boldsymbol{v}^*}{\boldsymbol{v}^{*\top}\boldsymbol{v}^*} = \frac{\|T(\pi)\|^2_{2,\boldsymbol{w}}}{\|\boldsymbol{v}^*\|^2_2} \geq \frac{\|\pi\|^2_{L_\infty(B(\hat{\boldsymbol{x}},\delta))}}{\|T^{-1}\|^2\|\boldsymbol{v}^*\|^2_2} \geq \frac{C^2_\pi}{\|T^{-1}\|^2} \geq \frac{w_{min}C^2_\pi}{C^2_{\theta,m}}.$$

It is worth noting that the above constants are independent of $\hat{\boldsymbol{x}}$. Therefore, let $C_\lambda := \frac{w_{min}C^2_\pi}{C^2_{\theta,m}}$, which completes the proof.

Theorem 3.2 allows us to expand the support radius of the weight function slightly and employ SSPB with other scaling sizes. For all $k \in \{1, \cdots, N\}$, consider the matrix $\boldsymbol{M}(\boldsymbol{x})$ under the point set $X_k = X \setminus \{\boldsymbol{x}_k\}$, and define it as $M_k(\boldsymbol{x})$, where

$$(\boldsymbol{M}_k(\boldsymbol{x}))_{i,j} = \sum_{l \in I_{\delta,X}(\boldsymbol{x}) \setminus \{k\}} w_l(\boldsymbol{x}) p_i(\boldsymbol{x}_l) p_j(\boldsymbol{x}_l).$$

It follows from Theorem 3.2 that the follow corollary is obtained.

Corollary 3.4. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is compacted and satisfies an interior cone condition with angle $\theta \in (0,\ \pi/2)$ and radius $r > 0$. Fix $m \in \mathbb{N}$. C_2 is defined in (3.1) and let $h_0 = r/C_2$. For all given $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\} \subseteq \Omega$ with $h_F \leq h_0$, let $\delta = 2C_2 h_F$. For all $\boldsymbol{x}_k \in X$, if for all $\boldsymbol{x} \in \Omega$ the point set $\{\boldsymbol{x}_j \in X :\ j \in I_{\delta,X}(\boldsymbol{x}),\ j \neq k\}$ is $\mathbb{P}_m^d$-unisolvent, then the smallest eigenvalue of $\boldsymbol{M}_k(\boldsymbol{x}_k)$ under $(\boldsymbol{x}_k, h_F)$-SSPB satisfies*

$$\lambda_{min}(\boldsymbol{M}_k(\boldsymbol{x}_k)) \geq C_\lambda > 0,$$

where C_λ is defined as in Theorem 3.2.

For a given k, consider to apply Theorem 3.2 for the point set X_k on the domain $\Omega \subseteq \mathbb{R}^d$. Since $h_F \geq h_{X_k,\Omega}$ satisfies the condition of Theorem 3.2, for all $\hat{\boldsymbol{x}} \in \Omega$, $\lambda_{min}(\boldsymbol{M}_k(\hat{\boldsymbol{x}})) \geq C_\lambda$ holds. Obviously, $\boldsymbol{x}_k \in X \subset \Omega$, which completes the proof.

If we suppose that for all k the conditions of Corollary 3.4 are satisfied, then the point set X satisfies Assumption 2.1 naturally, which implies the new assumption is stronger than Assumption 2.1. The strengthen is necessary for avoiding the cases which may cause $(\boldsymbol{A})_{k,k} = 1$. In practice, as long as the point set X is relatively dense, the stronger assumption is easy to be achieved. We propose the new assumption formally.

Assumption 3.1. *With the given δ and X, for all $\boldsymbol{x} \in \Omega$ and all $k \in \{1, \cdots, N\}$, the point set $\{\boldsymbol{x}_j : j \in I_{\delta,X}(\boldsymbol{x}),\ j \neq k\}$ is $\mathbb{P}_m^d$-unisolvent.*

According to Corollary 3.4 that the upper bound of $(\boldsymbol{A})_{k,k}$ which is strictly less than 1 can be obtained under Assumption 3.1.

Theorem 3.3. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is compacted and satisfies an interior cone condition with angle $\theta \in (0,\ \pi/2)$ and radius $r > 0$. Fix $m \in \mathbb{N}$. C_2 is defined in (3.1) and let $h_0 = r/C_2$. For all given $X = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\} \subseteq \Omega$ with $h_F \leq h_0$, let $\delta = 2C_2 h_F$. If X satisfies Assumption 3.1, then there exists a constant $C_M \in (0,\ 1)$ such that*

$$(\boldsymbol{A})_{k,k} \leq C_M.$$

Recalling (2.5) and (3.3), for every k, there is

$$(\boldsymbol{A})_{k,k} = w_k(\boldsymbol{x}_k)\boldsymbol{b}(\boldsymbol{x}_k)(\boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}_k)\boldsymbol{P})^{-1}\boldsymbol{b}(\boldsymbol{x}_k)^\top.$$

Let $\boldsymbol{b}_k := \boldsymbol{b}(\boldsymbol{x}_k) = (p_1(\boldsymbol{x}_k), \cdots, p_Q(\boldsymbol{x}_k))$. As stated above, under the given $\mathbb{P}_m^d$, the same $a_i(\boldsymbol{x})$ can be derived from the different bases, therefore we suppose that $(\boldsymbol{x}_k, h_F)$-BBPS $\{p_i(\boldsymbol{x})\}_{i=1}^Q$ defined as in (3.9) is employed. According to the expression, there is only one non-zero in the components of $\boldsymbol{b}(\boldsymbol{x}_k)$ and its value is 1, so $\|\boldsymbol{b}_k\|_2 = 1$.
Let $\tilde{I}_k = I_{\delta,X}(\boldsymbol{x}_k) \setminus \{k\}$. If Assumption 3.1 holds, $\{\boldsymbol{x}_j \in X : \ j \in \tilde{I}_k\}$ is $\mathbb{P}_m^d$-unisolvent. Further, because $\{w_l(\boldsymbol{x}_k)\}_{l\in\tilde{I}_k}$ are all positive, define an operator $< \cdot, \cdot >_{\tilde{w}}: \mathbb{P}_m^d \times \mathbb{P}_m^d \to \mathbb{R}$,

$$< f, g >_{\tilde{w}} = \sum_{l\in\tilde{I}_k} w_l(\boldsymbol{x}_k) f(\boldsymbol{x}_l) g(\boldsymbol{x}_l), \qquad \forall f,\ g \in \mathbb{P}_m^d.$$

It is east to be verified that $< \cdot, \cdot >_{\tilde{w}}$ is a inner product on $\mathbb{P}_m^d$, therefore there exists an unit orthogonal basis $\{\tilde{p}_i(\boldsymbol{x})\}_{i=1}^Q$ of $\mathbb{P}_m^d$ under $< \cdot, \cdot >_{\tilde{w}}$. Let $\tilde{\boldsymbol{b}}(\boldsymbol{x}) = (\tilde{p}_1(\boldsymbol{x}), \cdots, \tilde{p}_Q(\boldsymbol{x}))$, then there exists a matrix $\boldsymbol{Z}$ independent of $\boldsymbol{x}$ such that

$$\tilde{\boldsymbol{b}}(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x})\boldsymbol{Z}. \tag{3.19}$$

The matrix $\boldsymbol{P}_k$ represents the result after removing the k-th row of the matrix $\boldsymbol{P}$, i.e.,

$$\boldsymbol{P}_k = \begin{pmatrix} p_1(\boldsymbol{x}_1) & \cdots & p_Q(\boldsymbol{x}_1) \\ \vdots & \vdots & \vdots \\ p_1(\boldsymbol{x}_{k-1}) & \cdots & p_Q(\boldsymbol{x}_{k-1}) \\ p_1(\boldsymbol{x}_{k+1}) & \cdots & p_Q(\boldsymbol{x}_{k+1}) \\ \vdots & \vdots & \vdots \\ p_1(\boldsymbol{x}_N) & \cdots & p_Q(\boldsymbol{x}_N) \end{pmatrix}.$$

Let $\boldsymbol{W}_k = \mathrm{diag}(\{w_l(\boldsymbol{x}_k)\}_{l\in\tilde{I}_k})$. A routine computation can give $\boldsymbol{M}_k(\boldsymbol{x}_k) = \boldsymbol{P}_k^\top \boldsymbol{W}_k \boldsymbol{P}_k^\top$. According to (3.19) and the definition of $< \cdot, \cdot >_{\tilde{w}}$, there is $\boldsymbol{Z}\boldsymbol{Z}^\top = (\boldsymbol{M}_k(\boldsymbol{x}_k))^{-1}$. Consider $(\boldsymbol{A})_{k,k}$ under the basis $\{\tilde{p}_i(\boldsymbol{x})\}_{i=1}^Q$, there is

$$(\boldsymbol{A})_{k,k} = w_k(\boldsymbol{x}_k)\tilde{\boldsymbol{b}}_k(\boldsymbol{Z}^\top \boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}_k)\boldsymbol{P}\boldsymbol{Z})^{-1}\tilde{\boldsymbol{b}}_k^\top. \tag{3.20}$$

Let $\tilde{\boldsymbol{M}}_k = \boldsymbol{Z}^\top \boldsymbol{P}^\top \boldsymbol{W}(\boldsymbol{x}_k)\boldsymbol{P}\boldsymbol{Z}$, and it is clear that for all i, j, there is

$$\begin{aligned}(\tilde{\boldsymbol{M}}_k)_{i,j} &= \sum_{l\in I_{\delta,X}(\boldsymbol{x}_k)} w_l(\boldsymbol{x}_k)\tilde{p}_i(\boldsymbol{x}_l)\tilde{p}_j(\boldsymbol{x}_l) \\ &= w_k(\boldsymbol{x}_k)\tilde{p}_i(\boldsymbol{x}_k)\tilde{p}_j(\boldsymbol{x}_k) + \sum_{l\in\tilde{I}_k} w_l(\boldsymbol{x}_k)\tilde{p}_i(\boldsymbol{x}_l)\tilde{p}_j(\boldsymbol{x}_l) \\ &= w_k(\boldsymbol{x}_k)\tilde{p}_i(\boldsymbol{x}_k)\tilde{p}_j(\boldsymbol{x}_k) + <\tilde{p}_i,\tilde{p}_j>_{\tilde{w}}.\end{aligned}$$

It follows from the orthogonality that there is $\tilde{\boldsymbol{M}}_k = w_k(\boldsymbol{x}_k)\tilde{\boldsymbol{b}}_k^\top\tilde{\boldsymbol{b}}_k + \boldsymbol{I}$. It is easy to be checked that

$$(\tilde{\boldsymbol{M}}_k)^{-1} = \boldsymbol{I} - \frac{w_k(\boldsymbol{x}_k)}{1+w_k(\boldsymbol{x}_k)\|\tilde{\boldsymbol{b}}_k\|_2^2}. \tag{3.21}$$

By substituting (3.21) into (3.20), we obtain

$$(\boldsymbol{A})_{k,k} = \frac{w_k(\boldsymbol{x}_k)\|\tilde{\boldsymbol{b}}_k\|_2^2}{1+w_k(\boldsymbol{x}_k)\|\tilde{\boldsymbol{b}}_k\|_2^2}.$$

Since $w_k(\boldsymbol{x}_k) = \phi(0) > 0$, it is clear that the bigger $\|\tilde{\boldsymbol{b}}_k\|_2^2$ is, the bigger $(\boldsymbol{A})_{k,k}$ is. We know $\|\tilde{\boldsymbol{b}}_k\|_2^2 = \tilde{\boldsymbol{b}}_k\tilde{\boldsymbol{b}}_k^\top = \boldsymbol{b}_k\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{b}_k^\top$ and $\boldsymbol{Z}\boldsymbol{Z}^\top = (\boldsymbol{M}_k(\boldsymbol{x}_k))^{-1}$. If the conditions of Theorem 3.3 hold, Corollary 3.4 makes sure that the biggest eigenvalue of the positive semi-definite $\boldsymbol{Z}\boldsymbol{Z}^\top$ is strictly less than $1/C_\lambda$. At the beginning of the proof, we point out $\|\boldsymbol{b}_k\|_2 = 1$, therefore there is $\|\tilde{\boldsymbol{b}}_k\|_2^2 \le 1/C_\lambda$. It follows that

$$(\boldsymbol{A})_{k,k} \le \frac{\phi(0)}{C_\lambda + \phi(0)} < 1.$$

Obviously, let $C_M = \frac{\phi(0)}{C_\lambda+\phi(0)}$, and then the proof is completed.

So far, if the conditions of Theorem 3.3 hold, the diagonal elements of the matrix $\boldsymbol{A}$ have a positive uniform upper bound independent of the point set X, and the bound is strictly less than one. It indicates that all the results in Section 3.2 are true, if the conditions of Theorem 3.3 hold.

3.4. How to obtain and solve the l_0-optimization problem

In the last two sections, we propose a l_0-optimization problem (3.7) and analyze it in the theoretical case. In this section, we explain the detailed process of our l_0-ODR method, i.e., how to detects outliers and restores their true values in data $\{(\boldsymbol{x}_i,\tilde{f}_i)\}_{i=1}^N$ sampled from an unknown function f through solving (3.7).

There are two stages in l_0-ODR method. The first stage is obtaining the problem (3.7) by MLS-QI. Let $X = \{\boldsymbol{x}_i\}_{i=1}^N$ and $\tilde{\boldsymbol{F}} = (\tilde{f}_1, \cdots, \tilde{f}_N)^\top$. Set m in the polynomial space $\mathbb{P}_m^d$ small and bigger than $d/2$ according to (3.4). Assign the compactly supported weight function $\Phi_\delta(\cdot)$ of which an example is shown in Section 4 and recommended. Next, calculate the values of each shape function on X according to (2.5), and obtain matrix $\boldsymbol{A}$ referring to (3.3). Then, the problem (3.7) is obtained by letting $\boldsymbol{U} = \boldsymbol{I} - \boldsymbol{A}$ and $\tilde{\boldsymbol{R}} = (\boldsymbol{I} - \boldsymbol{A})\tilde{\boldsymbol{F}}$.
The second stage in l_0-ODR method is to solve problem (3.7) numerically. Because the problem (3.7) is the same form as the noisy recovery model (2.7), the orthogonal matching pursuit (OMP) algorithm in Section 2.2 is recommended, which can solve (2.7) efficiently. In solving, there is one problem that cannot be avoided, i.e., how to set η in (3.7). In Theorem 3.1, we give the theoretical scope of η, (3.8), and then we discuss how to set η in three common cases in outlier detection in practice. Further, how to solve the problem (3.7) through OMP in each case is also given.

(1) Knowing in advance the number of outliers, or an upper bound on the number K:
In this case, a solution to (3.7) that is at most K-sparse can be obtained by OMP which terminates after K iterations. The indices of non-zero elements in the solution correspond to K points in the data that are most likely to be outlier points. Namely, η is only used in analysis and does not need to be set in this case.

(2) Knowing in advance the lower bound of the outlier deviation ε:
In Corollary 3.3, we point out that if h_F is small enough, η can be set as any values in $(0,\ (1-C_M)\varepsilon]$. Setting $\eta = 0.5\cdot(1-C_M)\varepsilon$ seems to be reasonable with a small h_F. However, according the proof of Theorem 3.3, we can find out that the exact value of C_M can not be obtained easily. Therefore, in practice, we recommend

$$\eta_1 = U_m \cdot \varepsilon,$$

where $U_m := \min_i \|\boldsymbol{U}(:,i)\|_2$, and $\boldsymbol{U}(:,i)$ is the i-th column of matrix $\boldsymbol{U}$. $U_m \cdot \varepsilon$ is the lower bound of the MLS-QI error vector affected by single outlier.

(3) No prior information is available:
In some problems, we have to process data without priors. In this case,

we can estimate the lower bound of the outlier deviation ε according to the characteristics of the data.

Since we suppose the deviations of outliers are large and sparse, it is reasonable to assume that the outlier values hardly change the median of the data values, i.e.,

$$\text{median}(\boldsymbol{F}) \approxeq \text{median}(\tilde{\boldsymbol{F}}).$$

median$(\boldsymbol{F})$ shows roughly the magnitude of the data values. Further, we suppose $\varepsilon = \text{median}(\tilde{\boldsymbol{F}})$ in this case and recommend

$$\eta_2 = U_m \cdot \text{median}(\tilde{\boldsymbol{F}}).$$

The prior number of outliers is so strong in Case (1), that the standard OMP can solve (3.7) without η. The process of algorithm for Case (1) is presented in Algorithm 1.

For Case (2) and (3), the set of η not only affects the result of OMP but also

Algorithm 1 OMP Algorithm for the problem (3.7) in Case (1)

Input: measurement matrix $\boldsymbol{U}$, error vector $\tilde{\boldsymbol{R}}$, number of outliers K
Output: deviation vector $\boldsymbol{E}^*$
Initialization: $\boldsymbol{r}^0 = \tilde{\boldsymbol{R}}, \boldsymbol{E}^0 = \boldsymbol{0}, \Lambda^0 = \emptyset, k = 0$

```
while k < K do
  for all j do
    c^k(j) = U(:,j)^⊤ r^k / ||U(:,j)||_2
  end for
  Λ^{k+1} = Λ^k ∪ {arg max_j |c^k(j)|}
  E^{k+1} = arg min_{z:supp(z)⊂Λ^{k+1}} ||R̃ − Uz||_2
  r^{k+1} = R̃ − U E^{k+1}
  k = k + 1
end while
E* = E^k
```

changes the theoretical solution to (3.7). Therefore, the termination condition of OMP should be set more carefully. Considering that η_1 and η_2 are given in the case that h_F is small enough, we add an another termination condition in addition to the norm of the residual vector reaching η.

$$\left| \frac{\|\boldsymbol{r}^k\|_2 - \|\boldsymbol{r}^{k+1}\|_2}{\|\boldsymbol{r}^{k-1}\|_2 - \|\boldsymbol{r}^k\|_2} \right| \leq 5\%. \tag{3.22}$$

Inequality (3.22) means that the decline of the norm of the residual vector in OMP suddenly become flat, which implies all outliers may be removed. If inequality (3.22) holds or the norm of the residual vector reaches η, then OMP should be stopped. The details are shown in Algorithm 2.

Algorithm 2 OMP Algorithm for the problem (3.7) in Case (2) and (3)

Input: measurement matrix $\boldsymbol{U}$, error vector $\tilde{\boldsymbol{R}}$, η
Output: deviation vector $\boldsymbol{E}^*$
Initialization: $\boldsymbol{r}^0 = \tilde{\boldsymbol{R}}, \boldsymbol{E}^0 = \boldsymbol{0}, \Lambda^0 = \emptyset, k = 0$

```
while ||r^k||_2 > η do
  for all j do
    c^k(j) = U(:,j)^⊤ r^k / ||U(:,j)||_2
  end for
  Λ^{k+1} = Λ^k ∪ {arg max_j |c^k(j)|}
  E^{k+1} = arg min_{z:supp(z)⊂Λ^{k+1}} ||R̃ − Uz||_2
  r^{k+1} = R̃ − U E^{k+1}
  if k > 0 then
    if | (||r^k||_2 − ||r^{k+1}||_2) / (||r^{k−1}||_2 − ||r^k||_2) | ≤ 5% then
      break while
    end if
  end if
  k = k + 1
end while
E* = E^k
```

Remark 3.3. *In (3.22), the value 5% is an empirical value in this type of termination condition. Because the decline of the norm of the residual vector in OMP is similar to "L", modifying the value slightly does not affect the results of Algorithm 2. If the value is too big, inequality (3.22) is easy to be achieved, which makes the algorithm terminate prematurely, and only a few outliers can be marked. If the value is too small, inequality (3.22) is hard to be achieved, which makes the algorithm terminate only according to the another condition, and some fake outliers may be marked. In order to detect as many outliers as possible, only the small values are recommended.*

Through the above analysis and algorithms, it can be found that Case (1) is

easier to solve than the other two cases, so we do not verify this case in the numerical experiments. Case (2) requires some prior information of the data. For Case (3), which may exist more widely in various problems, the scheme we give is to estimate ε based on the data, and then we give η_2 in a similar way to Case (2), which implies that η_2 may be less reliable than η_1. The proposal of η_2 is only to show an idea. Users can estimate ε by other ways and give the corresponding η which may be more reasonable. Algorithm 2 is always recommended no matter how to set η.

After $\boldsymbol{E}^*$ is obtained by OMP, the outlier points can be can be obtained by analyzing the position of non-zero elements in $\boldsymbol{E}^*$. The polluted sampling values in $\tilde{\boldsymbol{F}}$ can be restored by the deviations in $\boldsymbol{E}^*$. The above is the whole process for l_0-ODR method.

In numerical experiments, we would demonstrate the effect of l_0-ODR method under Algorithm 2 on outlier detection and recovery, and further compare the influence of η_1 and η_2 on the results.

4. Numerical experiments

In the previous sections, we propose the l_0-ODR method to detect outliers and restore their true values by solving the l_0-minimization problem (3.7), and introduce the orthogonal matching pursuit (OMP) algorithm to solve the problem effectively. According to the analysis in Section 3.2, we point out that the results of l_0-ODR are related to the given upper bound η. In this section we demonstrate the effect of l_0-ODR method under Algorithm 2 equipped with η_1 and η_2 in Section 3.4 by some numerical examples respectively, and compare them.

In our numerical experiments, the data containing outliers are generated by randomly selecting some data points in the continuous function sampling or real data and adding large deviations. In sampling, we calculate the values on Halton [23] point set which is quasi-scattered in $[0,\ 1]^d$. It is worth noting that in the analysis in Section 3.2, we suppose that the separation distance of $\mathscr{O}_\varepsilon$ satisfies $q_O \geq \delta$. In fact, this condition is too strong. In the experiments, we simply select some data points from data at random and add large deviations to them in order to generate outliers

In addition to QIOR [11] which is also a outlier detection method in approximation, the Novel Local Outlier Detection (NLOD) [14] algorithm is employed to be compared with our l_0-ODR. NLOD is based on the very famous peak density clustering [24] to do outlier detection. Density peak clustering can quantify the possibility that each data point is cluster center and outlier, and it is suitable for

data with various distributions. However, the density peak clustering still needs to be set some thresholds to classify all points, which is not convenient for use. Du et al. set the threshold by relying on Chebyshev's theorem in statistics and propose NLOD with only one robust parameter. The review [6] describes NLOD as "a simple but effective density-based approach". In our experiments, the robust parameter of NLOD is be set as $k = 2$. To compare, we show the accuracy and the computational time of all methods in outlier detection. At the same time, we show the recovery accuracy of l_0-ODR for outlier values, which is unique among the three methods.

Both l_0-ODR and QIOR are based on MLS, so we equip them with same settings. The compactly supported C^4 Wendland's function $\phi(r) = (1-r)_+^6(35r^2+18r+3)$ is employed to deduce the weight function whose radius of support is $\delta = 2mh_F$. The used polynomial space is $\mathbb{P}_3^d$, i.e., the polynomial space containing all polynomials whose degrees are less than or equal to 3 on $\mathbb{R}^d$.

In order to demonstrate the accuracy of each method in outlier detection, we introduce the true positive (TP) and the false positive (FP) in image segmentation, which are respectively defined by the following equation

$$\begin{aligned} \mathrm{TP} &= |\mathscr{O}' \cap \mathscr{O}_\varepsilon|, \\ \mathrm{FP} &= |\mathscr{O}' \setminus \mathscr{O}_\varepsilon|, \end{aligned}$$

where $\mathscr{O}'$ is the outlier point set obtained by the corresponding method, and $|\cdot|$ denotes the number of elements in a set. Obviously, a larger TP and a smaller FP mean that $\mathscr{O}'$ is closer to $\mathscr{O}_\varepsilon$, which implies the method is in a higher accuracy. To show the recovery accuracy for outlier values, we define a recovery error, as

$$\mathrm{Er} = \frac{\|\boldsymbol{E} - \boldsymbol{E}^*\|_2}{\|\boldsymbol{E}\|_0},$$

where $\boldsymbol{E}$ is the solution to the problem (3.7) by OMP. Er represents the average error of $\boldsymbol{E}$ for each non-zero deviation in $\boldsymbol{E}^*$.

All experiments in this section are carried out on our PC, the CPU is Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, and the programming software is Matlab R2019a.

4.1. Simulation data—Franke's test function data

In this section, the sampled function is the classical Franke's test function

$$f(x,y) = \frac{3}{4}\exp(\frac{(9x-2)^2+(9y-2)^2}{4}) + \frac{3}{4}\exp(-\frac{(9x-2)^2}{49} - \frac{(9y+1)^2}{10}) - \frac{1}{5}\exp(-(9x-4)^2-(9y-7)^2) + \frac{1}{2}\exp(\frac{(9x-7)^2+(9y-3)^2}{4}),$$

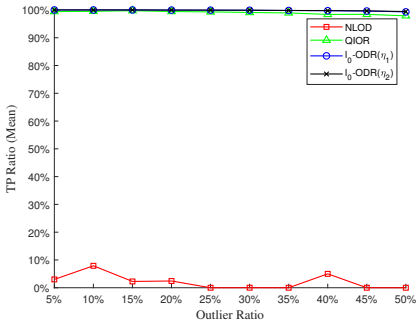
which is define in $\mathbb{R}^2$. The range of $f(x,y)$ in $[0,\ 1]\times[0,\ 1]$ is $[0,\ 1.3]$. The data are generated by sampling $f(x,y)$ on Halton point set whose size is N.

4.1.1. Results under different ratios of outlier

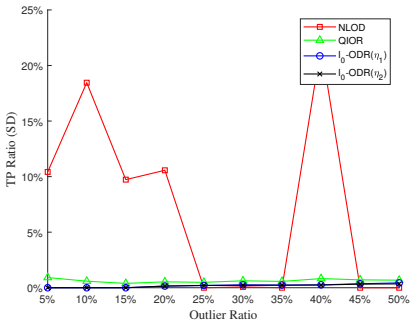
To show the effects of our l_0-ODR on the data containing different ratios of outlier, we keep $N = 1000$, randomly select $K = \alpha \cdot N$ points from the point set and add deviations with a range $[1,\ 2]$ to make them outliers, where α is the outlier ratio and is set to from 5% to 50% in the experiment. Under each outlier ratio, we use NLOD, QIOR and l_0-ODR equipped with η_1 and η_2 to detect outliers independently and repeatedly 100 times in the data with random outliers. In each experiment, we calculate the TP ratio (TP/K), the FP ratio (FP/K), the computational time and the recovery error (Er) of each method, and then calculate the means and the standard deviations (SD) of them under each outlier ratio. The results about the TP ratio, the FP ratio, the computational time and Er are respectively shown in Fig. 2, Fig. 3, Fig. 4 and Fig. 5. Detailed numerics in the four figures are presented in Appendix B.

It can be known from Fig. 2 that the means of the TP ratio obtained by QIOR and l_0-ODR are all above 95% and the SDs of them are all below 1%, which means that QIOR and l_0-ODR can detect the most of outliers in the most of experiments under the outlier ratios from 5% to 50%. In fact, under the outlier ratios from 5% to 15%, the means of the TP ratio by l_0-ODR equipped with two ηs are 100% and the SDs are 0, which means that l_0-ODR can detect all outliers in all experiments under these outlier ratios. The means of TP ratio obtained by NLOD are always below 10%, which implies that this density-based approach can not detect outliers in the scattered data for approximation.

It can be known from Fig. 3 that the means of TP ratio obtained by QIOR are much bigger than those obtained by l_0-ODR under the outlier ratios from 5% to 50%. Although as the outlier ratio increases, the means of TP ratio by l_0-ODR

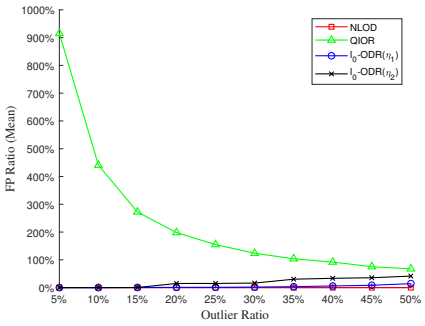


(a) The mean of the TP ratios obtained by NLOD, QIOR and l_0-ODR respectively

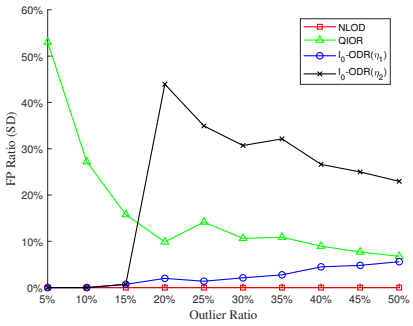


(b) The SD of the TP ratios obtained by NLOD, QIOR and l_0-ODR respectively

Fig. 2: Under the outlier ratios from 5% to 50%, the mean and the SD of the TP ratios obtained by NLOD, QIOR and l_0-ODR respectively

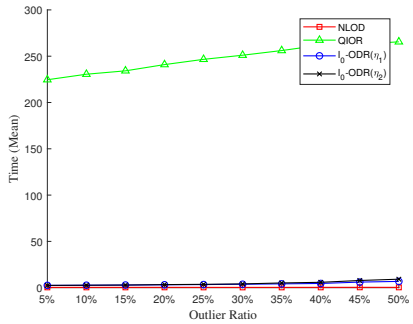


(a) The mean of the FP ratios obtained by NLOD, QIOR and l_0-ODR respectively

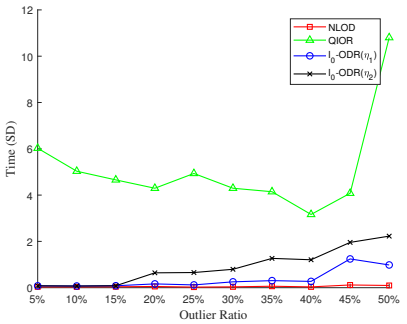


(b) The SD of the FP ratios obtained by NLOD, QIOR and l_0-ODR respectively

Fig. 3: Under the outlier ratios from 5% to 50%, the mean and the SD of the FP ratios obtained by NLOD, QIOR and l_0-ODR respectively

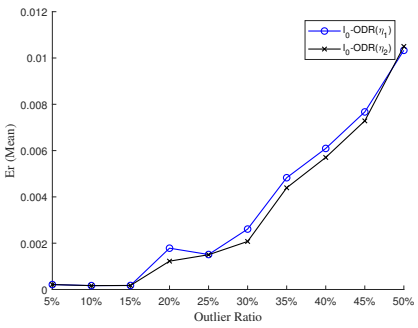


(a) The mean of the computational time of NLOD, QIOR and l_0-ODR respectively

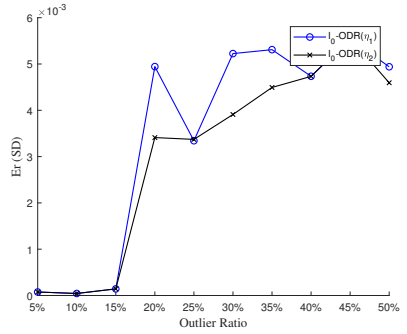


(b) The SD of the computational time of NLOD, QIOR and l_0-ODR respectively

Fig. 4: Under the outlier ratios from 5% to 50%, the mean and SD of the computational time of NLOD, QIOR and l_0-ODR respectively



(a) The mean of Er obtained by l_0-ODR equipped with η_1 and η_1 respectively



(b) The SD of Er obtained by l_0-ODR equipped with η_1 and η_1 respectively

Fig. 5: Under the outlier ratios from 5% to 50%, the mean and the SD of Er obtained by l_0-ODR equipped with η_1 and η_1 respectively

equipped with two ηs also increases, the mean of TP ratio by η_1 and η_2 is not above 15% and 50% respectively. Under the low outlier ratios, the mean of TP ratio by η_1 and η_1 are all very small. In fact, if the outlier ratio is 5% or 10%, the means of FP ratio by η_1 and η_1 are 0%. And if the outlier ratio is 10%, the means of FP ratio by them are 0.37%. These means that our l_0-ODR equipped with η_1 and η_2 has almost no misclassification when the outlier ratio is lower than 15%.
According to the means and the SDs in Fig. 4, there is no doubt that the computational time of QIOR is always much longer than that of l_0-ODR and NLOD. Though the computational time of NLOD is shortest, it fails according to the results in Fig. 2. It can be known from Fig. 5 that if the outlier ratio is lower than 15%, the recovery errors are surprisingly low and little changed. Although as the outlier ratio increases, the recovery error also increases, the magnitude of the errors are acceptable for the high outlier ratio. The magnitude of the SDs in Fig. 4(b) means that the recovery for the deviation vector by our l_0-ODR is relatively stable under different outlier ratios.
Combined with the results in Fig. 2, 3, 4, it can be found that l_0-ODR equipped with two ηs can always detect the most outliers with high efficiency under the outlier ratios from 5% to 50%. If the outlier ratio is low, l_0-ODR equipped with two ηs has no misclassification and the results are stable.

Remark 4.1. *In outlier detection, it is general to assume that the outlier ratio is low, because a large number of outliers can obscure the relationship between the true data. Although our l_0-ODR performs well at high outlier ratios, we can not guarantee such excellent results for arbitrary data, because outlier points are required to be sparse in the theory. Therefore, we would set the outlier ratio low and change other in the following experiments. Because the results of each method, especially our l_0-ODR, are quite stable under the low outlier ratio, we would only show the results of a single experiment in the following.*

4.1.2. Effects of the deviation range and small noise on the results

In this section, the effects of the deviation range and small noise on the results by QIOR, NLOD and l_0-ODR would be discussed. Firstly, we randomly select $K = 5\%N$ points from each set of data and add deviations with a range $[1,\ 2]$ to make them outliers, where N is set to 1000, 2000, 4000 respectively. The results of NLOD, QIOR, and l_0-ODR on outlier detection are shown in Table 1. The values in $(\cdot)$ in Table 1 are the results after adding small noise to the data containing outliers. The noise ratio is 2%, which means that for all i, there exists a random number in $[0,\ 0.02]$ such that $\tilde{F}_i = (1 \pm r_i) \cdot f_i$.

Table 1: Under the scattered data containing outliers whose deviations is in $[1,\ 2]$, the results of NLOD, QIOR, and l_0-ODR on outlier detection

(N,K)	Method	TP	FP	Time(s)	Er
(1000,50)	NLOD	24(24)	0(0)	0.3010(0.2616)	–
	QIOR	50(50)	430(655)	168(195)	–
	l_0-ODR(η_1)	50(50)	0(0)	2(2)	2.3703E-04(3.5994E-04)
	l_0-ODR(η_2)	50(50)	0(0)	2(2)	2.3703E-04(3.5994E-04)
(2000,100)	NLOD	0(3)	0(0)	0.9880(0.9490)	–
	QIOR	100(100)	995(1538)	1337(1490)	–
	l_0-ODR(η_1)	100(100)	0(0)	9(9)	8.9025E-05(1.5835E-04)
	l_0-ODR(η_2)	100(100)	0(0)	9(9)	8.9025E-05(1.5835E-04)
(4000,200)	NLOD	1(0)	0(0)	6(6)	–
	QIOR	200(200)	634(3116)	5575(11361)	–
	l_0-ODR(η_1)	200(200)	0(0)	40(40)	3.0314E-06(1.3170E-04)
	l_0-ODR(η_2)	200(200)	0(0)	40(40)	3.0314E-06(1.3170E-04)

According to Table 1, we can find that NLOD takes the least time among the three methods, but it hardly detects enough outliers. QIOR and l_0-ODR equipped with the two ηs can successfully detect all outliers with or without small noise. Especially, FPs of l_0-ODR are all zero, which indicates that there is no misclassification. FPs of QIOR are all non-zero and rise sharply after adding noise, which also makes the computational time increase significantly. All these show that l_0-ODR is better than QIOR in the accuracy of outlier detection, and has anti-noise ability. l_0-ODR takes almost equal time under the two ηs, which is hardly affected by small noise and significantly less than the time used by QIOR. With the increase of data points, although the number of outliers also increases, the recovery error shows a obvious downward trend. In the cases without small noise, the magnitudes of the recovery error show that l_0-ODR does restore the polluted data in a high accuracy. After adding noise, although the recovery errors have corresponding rises, the magnitudes of them are still relative small.

We think that the small deviations of outliers leads to the poor performance of NLOD. Then we raise the deviation range to $[10,\ 15]$, and the others are not changed. The effects of NLOD, QIOR, and l_0-ODR are shown in Table 2.

After the deviation range is raised to $[10,\ 15]$, NLOD successes in detecting all outliers with the least computational time, but the results of l_0-ODRs equipped with the two ηs and QIOR are almost the same as those in Table 1. Such results show that l_0-ODR and QIOR are more sensitive to outliers in the scattered data used for function approximation, but the computational efficiency is not as good as NLOD. It is worth mentioning that after raising the deviations, the recovery er-

Table 2: Under the scattered data containing outliers whose deviations is in $[10,\ 15]$, the results of NLOD, QIOR, and l_0-ODR on outlier detection

(N,K)	Method	TP	FP	Time(s)	Er
(1000,50)	NLOD	50(50)	0(0)	0.3639(0.3631)	–
	QIOR	50(50)	430(655)	176(200)	–
	l_0-ODR(η_1)	50(50)	0(0)	2(2)	2.3703E-04(3.5994E-04)
	l_0-ODR(η_2)	50(50)	0(0)	2(2)	2.3703E-04(3.5994E-04)
(2000,100)	NLOD	100(100)	0(0)	2(1)	–
	QIOR	100(100)	995(1538)	1427(1554)	–
	l_0-ODR(η_1)	100(100)	0(0)	8(8)	8.9025E-05(1.5835E-04)
	l_0-ODR(η_2)	100(100)	0(0)	8(8)	8.9025E-05(1.5835E-04)
(4000,200)	NLOD	200(200)	0(0)	12(13)	–
	QIOR	200(200)	634(3116)	5800(11734)	–
	l_0-ODR(η_1)	200(200)	0(0)	37(37)	3.0314E-06(1.3170E-04)
	l_0-ODR(η_2)	200(200)	0(0)	37(37)	3.0314E-06(1.3170E-04)

rors of l_0-ODR are not affected, which implies that the recovery error of l_0-ODR is independent of the magnitudes of the deviation.
Combined the results of Table 1 with those of Table 2, it can be found that l_0-ODR can detect and restore outliers in high accuracy in a moderate computational efficiency. l_0-ODR equipped with the two ηs can detect all outliers in the cases of small deviations, big deviations, noise and no noise, and there is no misclassification, which implies that l_0-ODR is robust.

Remark 4.2. *Since the range of Franke's test function in* $[0,\ 1]\times[0,\ 1]$ *is* $[0,\ 1.3]$*, detecting outliers with deviations in* $[10,\ 15]$ *is easy and can be also done by many statistical methods. Therefore, although NLOD performs well in the case of big deviations, we control the deviations in a relative small range in the following experiments to show the sensitivity to outliers, which is a aspect of practicality.*

4.2. Simulation data— the exponential test function data

In this section, we show and compare the effects of NLOD, QIOR, and l_0-ODR on the scattered point set X with $d = 3,\ 4$. The sampled function is the exponential test function referred to [11],

$$g_d(\boldsymbol{x}) = \exp(\sum_{i=1}^{d} x_i).$$

The sampled point sets are still Halton point sets with $N = 1000, 2000, 4000$ in $[0,\ 1]^d$. Since the range of $g_d(\boldsymbol{x})$ is $[1,\ \exp(d)]$ in $[0,\ 1]^d$, keeping $K = 5\%N$

fixed, the range of deviations is $[5,\ 10]$ in the case of $d = 3$, and the range of deviations is $[15,\ 30]$ in the case of $d = 4$. The results of the three method under the two ds are shown in Table 3 and 4, respectively. The values in $(\cdot)$ are the results under the same data added to small noise whose noise ratio is 2%.

According to Table 3 and 4, for such exponential data, we point out that l_0-ODR

Table 3: In the case fo $d = 3$, the results of NLOD, QIOR, and l_0-ODR on outlier detection

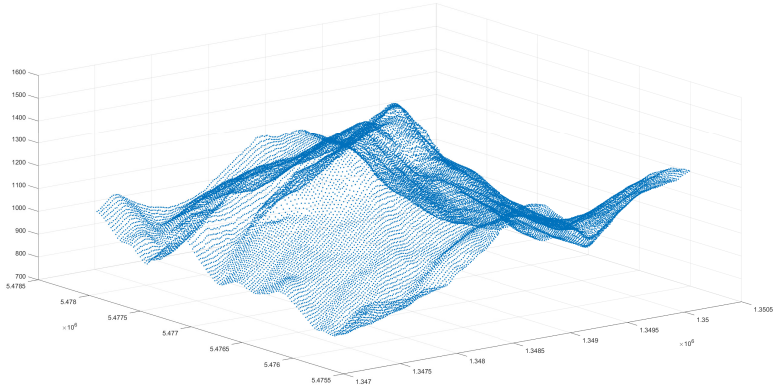
(N,K)	Method	TP	FP	Time(s)	Er
(1000,50)	NLOD	23(23)	0(0)	0.5261(0.3194)	–
	QIOR	50(50)	0(686)	63(449)	–
	l_0-ODR(η_1)	50(50)	0(0)	3(3)	1.4145E-04(3.3528E-03)
	l_0-ODR(η_2)	50(50)	0(0)	3(3)	1.4145E-04(3.3528E-03)
(2000,100)	NLOD	5(5)	0(0)	1(1)	–
	QIOR	100(100)	0(1514)	412(297)	–
	l_0-ODR(η_1)	100(100)	0(0)	12(12)	4.4980E-05(1.8762E-03)
	l_0-ODR(η_2)	100(100)	0(0)	12(12)	4.4980E-05(1.8762E-03)
(4000,200)	NLOD	2(2)	0(0)	6(6)	–
	QIOR	200(200)	0(3192)	2468(18006)	–
	l_0-ODR(η_1)	200(200)	0(0)	47(47)	8.8174E-06(1.1096E-03)
	l_0-ODR(η_2)	200(200)	0(0)	47(47)	8.8174E-06(1.1096E-03)

Table 4: In the case fo $d = 4$, the results of NLOD, QIOR, and l_0-ODR on outlier detection

(N,K)	Method	TP	FP	Time(s)	Er
(1000,50)	NLOD	0(0)	0(0)	0.5017(0.3881)	–
	QIOR	50(50)	42(487)	225(885)	–
	l_0-ODR(η_1)	50(50)	0(0)	7(7)	1.7247E-03(4.5962E-03)
	l_0-ODR(η_2)	50(50)	0(0)	7(7)	1.7247E-03(4.5962E-03)
(2000,100)	NLOD	47(47)	0(0)	2(2)	–
	QIOR	100(100)	97(1269)	2130(8572)	–
	l_0-ODR(η_1)	100(100)	0(0)	31(31)	9.2674E-04(3.0241E-03)
	l_0-ODR(η_2)	100(100)	0(0)	31(31)	9.2674E-04(3.0241E-03)
(4000,200)	NLOD	3(3)	0(0)	7(8)	–
	QIOR	200(200)	66(2846)	17547(79235)	–
	l_0-ODR(η_1)	200(200)	0(0)	104(104)	3.9407E-04(1.6748E-03)
	l_0-ODR(η_2)	200(200)	0(0)	104(104)	3.9407E-04(1.6748E-03)

can detect all outliers without misclassification in the cases of the two ds, with noise and without noise, which indicates that l_0-ODR work for the case of $d > 2$. Although the added outliers is intolerable for the function approximation problem, NLOD is still invalid. QIOR can detect all outliers, but it has larger FPs and time-consuming in all cases.

Fig. 6: The points in Vel'ký Rozsutec in the Malá Fatra dataset



Combining with the results in Section 4.1, we think that for the simulation data, among the three methods compared, the results of l_0-ODR equpped with the two ηs are relatively stable, and it can efficiently detect all outliers in high accuracy and restore the polluted data in relative high accuracy, which implies that the two recommended ηs are effective. Overall, we think l_0-ODR performs better in the compared methods. Next we show the effects of the three methods on outlier detection under real data.

4.3. Real data— GPS data of mount Vel'ký Rozsutec in the Malá Fatra

In this section, the original data are the GPS data of mount Vel'ký Rozsutec in the Malá Fatra, which are proposed in Section 5.3 of [25] and can be obtained through the online tool GPSVisualizer[1]. The dataset consists of 24,190 points in $\mathbb{R}^3$, which are shown in Fig. 6, and the ranges of the three coordinates is $[1.3472E+06,\ 1.3501E+06]$, $[5.4756E+06,\ 5.4782E+06]$ and $[764.2,\ 1583]$, respectively.

The dataset is not preprocessed and is only added some outliers to compare the three methods. Since the height coordinates of the mountain are large, the outlier deviation is set to $[250,\ 500]$. In this experiment, we let N fixed and the number of outliers $1\%N$, $3\%N$, $5\%N$, respectively. For such real data, there may be some machine errors in the sampling, so we do not add small noise. The results of NLOD, QIOR, and l_0-ODR on outlier detection under such real data are shown

[1] http://www.gpsvisualizer.com/elevation

in Table 5.

It follows from Table 5 that the computational time of the three method are

Table 5: The results of NLOD, QIOR, and l_0-ODR on outlier detection under Vel'ký Rozsutec in the Malá Fatra dataset

(N,K)	Method	TP	FP	Time(s)	Er
(21490,241)	NLOD	1	0	316	–
	QIOR	238	0	70747	–
	l_0-ODR(η_1)	241	0	1219	9.0455E-02
	l_0-ODR(η_2)	236	0	1204	2.4580E+00
(21490,725)	NLOD	5	0	296	–
	QIOR	723	2	205784	–
	l_0-ODR(η_1)	725	0	1977	4.7077E-02
	l_0-ODR(η_2)	719	0	1948	8.8299E-01
(21490,1209)	NLOD	8	0	314	–
	QIOR	1202	2	347161	–
	l_0-ODR(η_1)	1209	0	2975	4.8277E-02
	l_0-ODR(η_2)	1203	0	2854	5.3621E-01

significantly increased as the magnitude of data used is increased, but the size relationships among them are the same as in Section 4.1, 4.2. Similar to the results under simulation data, although NLOD takes the least time, it can not detect all outliers. QIOR can detect most of outliers, but there may be missed and misclassification. What is more, the computational time of QIOR is much higher that those of the other two methods in all cases.

l_0-ODR equipped with η_1 detects all outliers accurately in a low computational time and restores the polluted data in an acceptable accuracy relative to the magnitude of the height, which maintains the good performance in Section 4.1, 4.2. The performance of l_0-ODR equipped with η_2 goes bad, but it can detect most of outliers without FNs in about the same time as η_1. Because of the missing, the recovery errors of η_2 are bigger than those of η_1, but the magnitude of the errors is still small enough relative to the magnitude of the height. According to the discussion in Section 3.4, we believe that a slight modification of η_2 can improve the results.

Combining the results of the three methods, we believe that under the two ηs recommended in the paper, l_0-ODR can deal with the real data whose distribution is in a wide range and amount is large. l_0-ODR should be more practical than QIOR and NLOD for the data used for approximation.

5. Conclusion

In order to efficiently and accurately detect outliers in scattered data for approximation, this paper proposes a l_0-minimization problem with an inequality constrain based on the sensibility to outliers and high approximation accuracy of the moving least squares method and propose the l_0-ODR method by solving the l_0-minimization problem. We theoretically prove a sufficient condition for that the true outlier deviation vector is contained in the solution set of the proposed problem, which means that solving the problem can detect outliers and approximately restore their true values. For the parameter in the proposed problem, we not only analyze it in theory, but also propose two useful recommended values. For three common cases in outlier detection, we introduce the orthogonal matching algorithm in compressed sensing to solve the proposed problems efficiently, and give the specific approach respectively.

We compare our l_0-ODR with the quasi-interpolation and outliers removal (QIOR) method and the Novel Local Outlier Detection (NLOD) method. Numerical experiments show that no matter there is small noise or not, l_0-ODR detects outliers more accurate than other two methods and can restore the polluted data in high accuracy relative to their values, under the simulation data whose sampling dimension is from 2 to 4 respectively. The performance of l_0-ODR under the real data with wide distribution and a large amount is similar to that under the simulation data. The computational efficiency of l_0-ODR is much higher than that of QIOR but slightly lower than that of NLOD. Due to those results, it is reasonable to think that l_0-ODR has an anti-noise ability to small noise, and it is practical in outlier detection in scattered data for approximation.

Our work still has some room for improvement. For example, the bound of the parameter of the proposed l_0-minimization problem we give in Theorem 3.1 would be able to be further improved, which might bring some better setting schemes of the parameter. And the properties of other solutions to the proposed problem is worthy of being studied.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.12071019).

Appendix A. Proof of Proposition 3.1

Since $h_{X,\Omega} \leq h_F$ is obvious, we need only prove that the second inequality in 3.2 is true.

Suppose $h_F > 3h_{X,\Omega}$. Without loss of generality, let $h_F = h_{X_i,\Omega}$, then $\exists \boldsymbol{x}^* \in \bar{\Omega}$ such that $h_{X_i,\Omega} = \min_{\substack{1\leq k\leq N,\\ k\neq i}} \|\boldsymbol{x}^* - \boldsymbol{x}_k\|_2$, where $\bar{\Omega}$ denotes the closure of Ω. Further, we suppose $k = j$ reaches the minimum, i.e., $h_F = h_{X_i,\Omega} = \|\boldsymbol{x}^* - \boldsymbol{x}_j\|_2$.

According to the definition of the fill distance $h_{X,\Omega}$, there is $\min_{1\leq k\leq N} \|\boldsymbol{x}^* - \boldsymbol{x}_k\|_2 \leq h_{X,\Omega}$. Combined with the following analysis, there is $\min_{1\leq k\leq N} \|\boldsymbol{x}^* - \boldsymbol{x}_k\|_2 = \min\{\|\boldsymbol{x}^* - \boldsymbol{x}_i\|_2, h_F\}$.

If $\min_{1\leq k\leq N} \|\boldsymbol{x}^* - \boldsymbol{x}_k\|_2 = h_F \leq h_{X,\Omega}$, it leads to a contradiction to $h_F > 3h_{X,\Omega}$. To avoid the contradiction, we suppose $\min_{1\leq k\leq N} \|\boldsymbol{x}^* - \boldsymbol{x}_k\|_2 = \|\boldsymbol{x}^* - \boldsymbol{x}_i\|_2 < h_F$. Then the closed ball $\bar{B}(\boldsymbol{x}^*, h_F)$ meets $\{\boldsymbol{x}_i, \boldsymbol{x}_j\} \subseteq (\bar{B}(\boldsymbol{x}^*, h_F) \cap X)$. Let $\bar{\boldsymbol{x}} = (\boldsymbol{x}_i + \boldsymbol{x}_j)/2$. Because Ω is a convex set and $\boldsymbol{x}_i, \boldsymbol{x}_j \in \Omega$, $\bar{\boldsymbol{x}}$ is in Ω, too. Due to $\|\boldsymbol{x}^* - \boldsymbol{x}_j\|_2 = h_F > 3h_{X,\Omega}$ and $\|\boldsymbol{x}^* - \boldsymbol{x}_i\|_2 \leq h_{X,\Omega}$, there is

$$\begin{aligned}
\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 &= \|\boldsymbol{x}_i - \boldsymbol{x}^* + \boldsymbol{x}^* - \boldsymbol{x}_j\|_2 \\
&\geq \|\boldsymbol{x}^* - \boldsymbol{x}_j\|_2 - \|\boldsymbol{x}^* - \boldsymbol{x}_i\|_2 \\
&> 2h_{X,\Omega}.
\end{aligned}$$

Therefore, $\|\bar{\boldsymbol{x}} - \boldsymbol{x}_i\|_2 = \|\bar{\boldsymbol{x}} - \boldsymbol{x}_j\|_2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2/2 > h_{X,\Omega}$. For every $\boldsymbol{x}_l \in X \setminus \{\boldsymbol{x}_i, \boldsymbol{x}_j\}$, because $\|\boldsymbol{x}^* - \boldsymbol{x}_j\|_2 = h_{X_i,\Omega} = h_F$, there is $\|\boldsymbol{x}^* - \boldsymbol{x}_l\|_2 \geq h_F > 3h_{X,\Omega}$. Further, there is

$$\begin{aligned}
\|\bar{\boldsymbol{x}} - \boldsymbol{x}_l\|_2 &= \|\bar{\boldsymbol{x}} - \boldsymbol{x}^* + \boldsymbol{x}^* - \boldsymbol{x}_l\|_2 \\
&= \|\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x}_j - \boldsymbol{x}^*) + (\boldsymbol{x}^* - \boldsymbol{x}_l)\|_2 \\
&\geq \|\boldsymbol{x}^* - \boldsymbol{x}_l\|_2 - \frac{1}{2}\|\boldsymbol{x}^* - \boldsymbol{x}_j\|_2 - \frac{1}{2}\|\boldsymbol{x}^* - \boldsymbol{x}_i\|_2 \\
&\geq \|\boldsymbol{x}^* - \boldsymbol{x}_l\|_2 - \frac{1}{2}h_F - \frac{1}{2}h_{X,\Omega} \\
&> h_{X,\Omega}.
\end{aligned}$$

Those implies $\min_{1\leq k\leq N} \|\bar{\boldsymbol{x}} - \boldsymbol{x}_k\|_2 > h_{X,\Omega}$, which leads to a contradiction to the definition of $h_{X,\Omega}$.

In conclusion, $h_F \leq 3h_{X,\Omega}$ must be true.

Appendix B. The specific numbers contained in the figures in Section 4.1

The specific results in Fig. 2, 3, 4 and 5 are shown in Table B.6, B.7, B.8 and B.9 respectively. The experimental conclusions are discussed in Section 4.1. These tables are presented for readers to know the results clear.

Table B.6: Under the outlier ratios from 5% to 50%, the mean and the SD of the TP ratios obtained by NLOD, QIOR and l_0-ODR respectively

Method	Outlier Ratio							
	5		10		15		20	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NLOD	3.00%	1.0421E-01	7.90%	1.8452E-01	2.27%	9.7283E-02	2.43%	1.0570E-01
QIOR	99.40%	9.1652E-03	99.50%	5.9161E-03	99.73%	3.8873E-03	99.43%	5.3092E-03
l_0-ODR(η_1)	100.00%	0.0000E+00	100.00%	0.0000E+00	100.00%	0.0000E+00	99.95%	1.5000E-03
l_0-ODR(η_2)	100.00%	0.0000E+00	100.00%	0.0000E+00	100.00%	0.0000E+00	99.95%	1.5000E-03
Method	Outlier Ratio							
	25		30		35		40	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NLOD	0.00%	0.0000E+00	0.02%	7.2648E-04	0.00%	0.0000E+00	5.00%	2.1794E-01
QIOR	99.30%	4.8785E-03	99.08%	6.2249E-03	98.93%	5.7054E-03	98.39%	8.2339E-03
l_0-ODR(η_1)	99.92%	2.0396E-03	99.88%	2.6405E-03	99.80%	2.4075E-03	99.73%	2.6101E-03
l_0-ODR(η_2)	99.92%	2.0396E-03	99.90%	2.1344E-03	99.81%	2.2633E-03	99.76%	2.4335E-03
Method	Outlier Ratio							
	45		50					
	Mean	SD	Mean	SD				
NLOD	0.00%	0.0000E+00	0.00%	0.0000E+00				
QIOR	98.41%	7.1068E-03	97.88%	6.7941E-03				
l_0-ODR(η_1)	99.61%	3.5746E-03	99.30%	4.3589E-03				
l_0-ODR(η_2)	99.67%	3.2584E-03	99.42%	3.4583E-03				

Table B.7: Under the outlier ratios from 5% to 50%, the mean and the SD of the FP ratios obtained by NLOD, QIOR and l_0-ODR respectively

Method	Outlier Ratio							
	5		10		15		20	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NLOD	0.00%	0.0000E+00	0.00%	0.0000E+00	0.00%	0.0000E+00	0.00%	0.0000E+00
QIOR	915.00%	5.3019E-01	441.00%	2.7212E-01	272.33%	1.5768E-01	198.03%	9.8647E-02
l_0-ODR(η_1)	0.00%	0.0000E+00	0.00%	0.0000E+00	0.37%	7.1414E-03	0.98%	1.9967E-02
l_0-ODR(η_2)	0.00%	0.0000E+00	0.00%	0.0000E+00	0.37%	7.1414E-03	14.98%	4.3933E-01
Method	Outlier Ratio							
	25		30		35		40	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NLOD	0.00%	0.0000E+00	0.00%	0.0000E+00	0.00%	0.0000E+00	0.00%	0.0000E+00
QIOR	154.88%	1.4162E-01	123.72%	1.0626E-01	103.80%	1.0895E-01	92.01%	8.9546E-02
l_0-ODR(η_1)	1.10%	1.3791E-02	2.00%	2.1134E-02	3.34%	2.7614E-02	5.91%	4.4877E-02
l_0-ODR(η_2)	15.06%	3.4926E-01	16.40%	3.0695E-01	30.31%	3.2130E-01	33.56%	2.6615E-01
Method	Outlier Ratio							
	45		50					
	Mean	SD	Mean	SD				
NLOD	0.00%	0.0000E+00	0.00%	0.0000E+00				
QIOR	75.36%	7.6989E-02	67.56%	6.7562E-02				
l_0-ODR(η_1)	8.47%	4.8007E-02	14.34%	5.5993E-02				
l_0-ODR(η_2)	35.60%	2.4987E-01	41.47%	2.2957E-01				

Table B.8: Under the outlier ratios from 5% to 50%, the mean and the SD of the computational time of NLOD, QIOR and l_0-ODR respectively

Method	Outlier Ratio							
	5		10		15		20	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NLOD	3.54E-01	3.60E-02	3.62E-01	3.35E-02	3.65E-01	3.94E-02	3.87E-01	5.26E-02
QIOR	2.25E+02	6.02E+00	2.31E+02	5.03E+00	2.34E+02	4.66E+00	2.41E+02	4.30E+00
l_0-ODR(η_1)	2.48E+00	8.00E-02	2.65E+00	7.26E-02	2.86E+00	8.37E-02	3.13E+00	1.64E-01
l_0-ODR(η_2)	2.48E+00	8.17E-02	2.66E+00	7.22E-02	2.86E+00	8.37E-02	3.34E+00	6.42E-01
Method	Outlier Ratio							
	25		30		35		40	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NLOD	3.49E-01	2.58E-02	3.27E-01	3.72E-02	3.47E-01	6.30E-02	3.72E-01	3.54E-02
QIOR	2.47E+02	4.93E+00	2.51E+02	4.30E+00	2.56E+02	4.15E+00	2.62E+02	3.16E+00
l_0-ODR(η_1)	3.37E+00	1.20E-01	3.72E+00	2.53E-01	4.15E+00	3.05E-01	4.56E+00	2.72E-01
l_0-ODR(η_2)	3.63E+00	6.55E-01	4.09E+00	7.95E-01	5.12E+00	1.27E+00	5.81E+00	1.21E+00
Method	Outlier Ratio							
	45		50					
	Mean	SD	Mean	SD				
NLOD	4.17E-01	1.17E-01	3.80E-01	9.59E-02				
QIOR	2.63E+02	4.08E+00	2.66E+02	1.08E+01				
l_0-ODR(η_1)	6.05E+00	1.24E+00	6.81E+00	9.82E-01				
l_0-ODR(η_2)	7.78E+00	1.96E+00	9.18E+00	2.23E+00				

Table B.9: Under the outlier ratios from 5% to 50%, the mean and the SD of Er obtained by l_0-ODR equipped with η_1 and η_1 respectively

Method	Outlier Ratio							
	5		10		15		20	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
l_0-ODR(η_1)	2.1276E-04	7.5326E-05	1.6732E-04	4.3528E-05	1.7345E-04	1.4274E-04	1.7845E-03	4.9434E-03
l_0-ODR(η_2)	2.1276E-04	7.5326E-05	1.6732E-04	4.3528E-05	1.7345E-04	1.4274E-04	1.2241E-03	3.4109E-03
Method	Outlier Ratio							
	25		30		35		40	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
l_0-ODR(η_1)	1.5111E-03	3.3424E-03	2.6097E-03	5.2247E-03	4.8279E-03	5.3097E-03	6.0934E-03	4.7345E-03
l_0-ODR(η_2)	1.5011E-03	3.3713E-03	2.0753E-03	3.9096E-03	4.3928E-03	4.4933E-03	5.7041E-03	4.7325E-03
Method	Outlier Ratio							
	45		50					
	Mean	SD	Mean	SD				
l_0-ODR(η_1)	7.6703E-03	5.5438E-03	1.0331E-02	4.9398E-03				
l_0-ODR(η_2)	7.2834E-03	5.5563E-03	1.0499E-02	4.5950E-03				

References

[1] G. E. Fasshauer, Meshfree Approximation Methods with Matlab, WORLD SCIENTIFIC, 2007. doi:10.1142/6437.

[2] H. Wendland, Scattered Data Approximation, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2004. doi:10.1017/cbo9780511617539.

[3] D. Levin, The approximation power of moving least-squares, Mathematics of Computation 67 (224) (1998) 1517–1531. doi:10.1090/s0025-5718-98-00974-0.

[4] D. Levin, Between moving least-squares and moving least-l_1, BIT Numerical Mathematics 55 (3) (2014) 781–796. doi:10.1007/s10543-014-0522-0.

[5] T. E. Dielman, Least absolute value regression: recent contributions, Journal of Statistical Computation and Simulation 75 (4) (2005) 263–286. doi:10.1080/00949650420002236 80.

[6] A. Boukerche, L. Zheng, O. Alfandi, Outlier Detection: Methods, Models, and Classification, ACM Computing Surveys 53 (3) (2020) 1–37. doi:10.1145/3381028.

[7] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM Computing Surveys 41 (3) (2009) 1–58. doi:10.1145/1541880.1541882.

[8] V. Hodge, J. Austin, A Survey of Outlier Detection Methodologies, Artificial Intelligence Review 22 (2) (2004) 85–126. doi:10.1023/b:aire.0000045502.10941.a9.

[9] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust Recovery of Subspace Structures by Low-Rank Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013) 171–184. doi:10.1109/tpami.2012.88.

[10] S. Zheng, R. Feng, A. Huang, A modified moving least-squares suitable for scattered data fitting with outliers, Journal of Computational and Applied Mathematics 370 (2020) 112655. doi:10.1016/j.cam.2019.112655.

[11] A. Amir, D. Levin, Quasi-interpolation and outliers removal, Numerical Algorithms 78 (3) (2017) 805–825. doi:10.1007/s11075-017-0401-2.

[12] Z. Xu, Compressed sensing: a survey, SCIENTIA SINICA Mathematica 42 (9) (2012) 865–877. doi:10.1360/012011-1043.

[13] Z. Xu, T. Zhou, On Sparse Interpolation and the Design of Deterministic Interpolation Points, SIAM Journal on Scientific Computing 36 (4) (2014) A1752–A1769. doi:10.1137/13094596x.

[14] H. Du, S. Zhao, D. Zhang, J. Wu, Novel clustering-based approach for Local Outlier Detection, in: 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2016, pp. 802–811. doi:10.1109/infcomw.2016.7562187.

[15] M. Dehghan, M. Abbaszadeh, Interpolating stabilized moving least squares (MLS) approximation for 2D elliptic interface problems, Computer Methods in Applied Mechanics and Engineering 328 (2018) 775–803. doi:10.1016/j.cma.2017.09.002.

[16] L. Zhang, T. Gu, J. Zhao, S. Ji, Q. Sun, M. Hu, An adaptive moving total least squares method for curve fitting, Measurement 49 (2014) 107–112. doi:10.1016/j.measurement.2013.11.050.

[17] X. Li, S. Li, Analysis of the complex moving least squares approximation and the associated element-free Galerkin method, Applied Mathematical Modelling 47 (2017) 45–62. doi:10.1016/j.apm.2017.03.019.

[18] R. Salehi, M. Dehghan, A generalized moving least square reproducing kernel method, Journal of Computational and Applied Mathematics 249 (2013) 120–132. doi:10.1016/j.cam.2013.02.005.

[19] J. A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Transactions on Information Theory 50 (10) (2004) 2231–2242. doi:10.1109/tit.2004.834793.

[20] J. A. Tropp, A. C. Gilbert, Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, IEEE Transactions on Information Theory 53 (12) (2007) 4655–4666. doi:10.1109/tit.2007.909108.

[21] D. Mirzaei, Analysis of moving least squares approximation revisited, Journal of Computational and Applied Mathematics 282 (2015) 237–250. doi:10.1016/j.cam.2015.01.007.

[22] J. M. Melenk, On Approximation in Meshless Methods, in: A. W. Blowey, James F.and Craig (Ed.), Frontiers of Numerical Analysis, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 65–141. doi:10.1007/3-540-28884-8_2.

[23] J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, Numerische Mathematik 2 (1) (1960) 84–90. doi:10.1007/bf01386213.

[24] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, Science 344 (6191) (2014) 1492–1496. doi:10.1126/science.1242072.

[25] Z. Majdisova, V. Skala, Radial basis function approximations: comparison and applications, Applied Mathematical Modelling 51 (2017) 728–743. doi:10.1016/j.apm.2017.07.033.