Integrating Clustering Method in Compactly Supported Radial Basis Function for Surface Approximation

Khang Jie Liew*, Kah Heng Tee, Ahmad Ramli, and Wen Eng Ong

***Abstract*—Radial basis function, one of the mesh-free methods, makes it convenient to interpolate and approximate high dimensional data points. Interpolation fits the data points exactly, whereas approximation fits the data points approximately. The approximation method is more appropriate for large and noisy data points in comparison to the interpolation method. Compactly supported radial basis function is extensively discussed in the literature of approximation theory. It also becomes popular as a result of its computational advantages. Hence, in this study, Wendland's compactly supported radial basis function is chosen. The main contribution of this paper is integrating the clustering procedures to determine the recommended number of reference points, which is expected to provide as few reference points as possible for surface approximation. Three bivariate test functions are selected to generate a moderately large amount of data points followed by the process of adding different levels of noise in order to observe the effectiveness of the proposed method. The results obtained are further confirmed and analysed through error analysis. Finally, the experimental results show that the surfaces are well approximated from the recommended number of reference points gained from the proposed method.**

***Index Terms*—Compact support, Radial basis function, Clustering, Reference points, Surface approximation.**

I. Introduction

INTERPOLATION and approximation are prominently known as the most common approaches that can be adopted to solve engineering problems. In relation to this, one of the techniques that can be applied for data interpolation and approximation is known as radial basis functions (RBFs). RBF can be used to produce smoother curves or surfaces from a large number of data points in comparison to the other fitting methods. It also provides mesh-free method which allows large scattered data points to be handled in high dimensional space to overcome engineering problem that is related to unorganised data points. In addition, this method is able to be extended to high dimensional space due to the unnecessary tessellation of the geometric domain. [1] originally introduced the RBF method as a new analytical method that can be used to fit and represent the irregular surfaces using multiquadric function. Generally, RBFs are widely applied in engineering problems such as environmental modelling [2], neural networks [3], [4], [5], geography and digital terrain modelling [6], solving partial differential equations [7], surface reconstruction [8], [9], and surface denoising [10].

Furthermore, it should be observed that RBFs can be divided into two main groups, namely global RBFs and the compactly supported RBFs (CSRBFs). Gaussian, multiquadric, inversed multiquadric, and polyharmonic splines are examples of global RBF [11]. On the other hand, the CSRBF was introduced by [12]. Pertaining to this, global RBFs possess non-compact support which indicates the presence of infinite interval, whereas the CSRBFs contain the compact and locally-supported with the function value being zero outside a certain interval. In the aspect of approximation, global RBFs are insensitive to the distribution of data points; while the sensitivity of CSRBFs tends to be high towards the distribution of data points. The solution to the approximation and interpolation performed using global RBFs usually leads to the solving of the system of linear equation. However, the solution is unreliable when approximating a large amount of data points due to the dense and ill-conditioned matrix. Therefore, CSRBFs can be used as an alternative method to solve this particular problem due to sparse matrix. Sparse matrix is the matrix with a relatively large number of zeroes present in the system. Hence, decreased complexity and increased computational efficiency can be achieved by reducing the number of coefficients that needs to be solved. Moreover, the number of reference points has to be reduced, but a good approximation for the purpose of decreasing the computational time of CSRBFs must be maintained. It is important to acknowledge that the method to determine the recommended number of reference points for RBFs remains as an open problem, but several other methods such as randomly chosen, clustering methods, and furthest point algorithm can be used as replacement [13], [14]. [15] used the principal component analysis to cluster the data points by defining the computed centroid of the clusters as the reference points of CSRBFs. In another work of [14], the furthest point algorithm was used to select the reference points of CSRBFs before the surface is reconstructed from the unorganised and noisy data points. On the other hand, [16] tried placing the reference points in a uniform grid during the process of RBF approximation , but it was not the only possible method because the emphasis is on the capability of the placement to reflect the terrain of the surface.

Clustering is a technique that can be used to group objects based on their similarities or dissimilarities [17]. The most commonly used clustering methods are hierarchical and non-hierarchical. Hence, both clustering methods are proposed

Manuscript received February 24, 2018; revised August 15, 2018. This work was supported by Universiti Sains Malaysia under RUI grant no. 1001/PMATHS/8011014.

Khang Jie Liew is with the Centre for American Education, Sunway University and the School of Mathematical Sciences, Universiti Sains Malaysia (e-mail: kenji_liewkj@yahoo.com.my / khangjiel@sunway.edu.my).

Kah Heng Tee, Ahmad Ramli, and Wen Eng Ong are with School of Mathematical Sciences, Universiti Sains Malaysia (email: royal_tee8127@hotmail.com, alaramli@usm.my, and weneng@usm.my).

in this study to determine the number of reference points for CSRBFs approximation. Furthermore, the hierarchical method is performed to obtain the optimal number of clusters based on the elbow rule, where the K-means method (a type of non–hierarchical method) is used to compute the cluster centroids which is defined as the reference points of CSRBFs once the optimal number of clusters is established. Therefore, it can be concluded that the reference points of CSRBFs are not the subset of the given data points. The mathematical background of CSRBFs approximation and cluster analysis with elbow rule are briefly provided. On top of that, the proposed algorithm that applies cluster analysis is described in Section 2, while Section 3 aims to present the related numerical results from the elbow rule with different noise levels, graphical results for CSRBFs approximation based on the recommended number of reference points, and the error analysis that validates the experimental results. A discussion will also be presented based on the results obtained from Section 3. Finally, a detailed conclusion will be provided in Section 4.

II. Materials and Methods

A. Compactly supported radial basis function

Compactly supported radial basis function (CSRBF) was initially introduced by [18], but the lowest degree of CSRSBFs remains as an open problem. In relation to this matter, [12] managed to solve the open problem by constructing the lowest degree of CSRBF. The main idea of CSRBF is to use a polynomial as a function of Euclidean norm $||\cdot||$ on d-dimensional, $\mathbb{R}^d$ which are denoted as r with the support on $[0,1]$. However, it can be scaled to $[0,\epsilon]$ by replacing the r with $\frac{r}{\epsilon}$ for some positive ϵ apart from CSRBFs with support on $[0,1]$. The ϵ is known as the scale factor or shape parameter. It set to one in this study. CSRBF is a strictly positive definite in $\mathbb{R}^d$ for all d less than or equal to some fixed value of d_0. Hence, the property of positive definite is crucial to allow the system of linear equation to be solved easily and successfully [19]. The general form of cut-off polynomial of CSRBF is presented in (1).

$$\phi_{d,\,k}(r) = \begin{cases} p_{d,\,k}(r), & 0 \le r \le 1 \\ 0, & r > 1 \end{cases} \tag{1}$$

where d refers to the dimension number, $2k$ represents the continuity (smoothness) of the function for some non-negative integer k, and $p_{d,\,k}(r)$ is described as a univariate polynomial of degree $\lfloor \frac{d}{2} \rfloor + 3k + 1$. The Wendland's recursive formula for the function $\phi_{d,\,k}$ for all d and $k = 0, 1, 2, \text{and } 3$ was formulated as an explicit formula in (2) [20].

$$\begin{aligned}
\phi_{d,0}(r) &= (1-r)_+^l, \\
\phi_{d,1}(r) &\doteq (1-r)_+^{l+1}[(l+1)r+1], \\
\phi_{d,2}(r) &\doteq (1-r)_+^{l+2}[(l^2+4l+3)r^2+(3l+6)r+3], \\
\phi_{d,3}(r) &\doteq (1-r)_+^{l+3}[(l^3+9l^2+23l+15)r^3 \\
&\quad + (6l^2+36l+45)r^2+(15l+45)r+15]
\end{aligned} \tag{2}$$

where $l = \lfloor \frac{d}{2} \rfloor + k + 1$, while the symbol $\doteq$ is referred as equality that is up to a multiplicative positive constant.

For example, let us assume $d = 3$. In this case, the Wendland's CSRBFs with the continuity of C^0, C^2, C^4, and C^6, respectively are given in (3).

$$\begin{aligned}
\phi_{3,0}(r) &= (1-r)_+^2, \\
\phi_{3,1}(r) &\doteq (1-r)_+^4(4r+1), \\
\phi_{3,2}(r) &\doteq (1-r)_+^6(35r^2+18r+3), \\
\phi_{3,3}(r) &\doteq (1-r)_+^8(32r^3+25r^2+8r+1).
\end{aligned} \tag{3}$$

Therefore, Wendland's basis function, $\phi_{3,1}(r)$ with C^2 continuity is selected for this study due its C^2 continuity.

B. Compactly supported radial basis function approximation

Generally, the available data points of 2D or 3D are normally assumed to be scattered and contaminated with noise. Hence, the approximation method is considered as a more suitable mean compared to the interpolation method in order to fit these data points. The scattered data points can be described as the given set of N distinct data points, $X = \{x_i\}_{i=1}^N \subseteq \mathbb{R}^d$, where d refers to the d-dimensional space. Furthermore, every point x_i is associated with a set of function value $\{f_i\}_{i=1}^N \subseteq \mathbb{R}$, whereas, the approximation function, s is denoted as $s : \mathbb{R}^d \to \mathbb{R}$, whereby $s(x_i) \approx f(x_i)$.

The general formula for RBFs approximation function is given in (4).

$$s(X) = q(X) + \sum_{j=1}^{K} \lambda_j \phi(||X - \xi_j||) \tag{4}$$

where $q(X)$ refers to the low degree polynomial function, λ_j represents the weight value that corresponds to ξ_j which is known as the reference point, and K is the number of reference points. Commonly, the polynomial function $q(X)$ is in linear form; hence, it is able to ensure the stability and solvability of RBFs approximation function [21].

In this study, the CSRBFs approximation function for data of two variables with corresponding function values is denoted as $s : \mathbb{R}^2 \to \mathbb{R}$; hence:

$$\begin{aligned}
s(x_i, y_i) &= a_1 + a_2 x_i + a_3 y_i \\
&\quad + \sum_{j=1}^{K} \lambda_j \phi\left(\sqrt{(x_i - \xi_{x_j})^2 + (y_i - \xi_{y_j})^2)}\right) \\
&= a_1 + a_2 x_i + a_3 y_i + \sum_{j=1}^{K} \lambda_j \phi(r_{i,j})
\end{aligned} \tag{5}$$

for $i = 1, 2, 3, \ldots, N$ and the x-coordinate and y-coordinate of the reference points are denoted as ξ_{x_j} and ξ_{y_j}, respectively. Hence, the linear system of equations of (5) can be written in the matrix form as shown in (6).

$$\begin{bmatrix} 1 & x_1 & y_1 & \phi(r_{1,1}) & \phi(r_{1,2}) & \cdots & \phi(r_{1,K}) \\ 1 & x_2 & y_2 & \phi(r_{2,1}) & \phi(r_{2,2}) & \cdots & \phi(r_{2,K}) \\ 1 & x_3 & y_3 & \phi(r_{3,1}) & \phi(r_{3,2}) & \cdots & \phi(r_{3,K}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_N & y_N & \phi(r_{N,1}) & \phi(r_{N,2}) & \cdots & \phi(r_{N,K}) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_K \end{bmatrix} = \begin{bmatrix} f(x_1, y_1) \\ f(x_2, y_2) \\ f(x_3, y_3) \\ \vdots \\ f(x_N, y_N) \end{bmatrix}. \tag{6}$$

The system in (6) is overdetermined as a result of $(K+3) < N$. Hence, there are $(K+3)$ unknown weight values $(a_1, a_2, a_3, \lambda_1, \lambda_1, \ldots, \lambda_K)^T$ need to be determined. In regards to this, the system can be solved using QR decomposition instead of applying the normal equations of the least square method due to the possibility of sparse matrix in the system of CSRBFs. Let the system in (6) is represented as $A\hat{\mathbf{x}} = \mathbf{f}$ and the QR decomposition $A = QR$, then the system has a unique least square solution given in (7).

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{f} \tag{7}$$

C. Cluster analysis

Cluster analysis is a method which involves grouping a set of objects. This set of objects is characterised based on certain measurements. Moreover, it possesses a wide range of applications in several domains such as artificial intelligence, life sciences, medical sciences, social sciences, and others [22]. However, no assumptions can be made regarding the number of groups that allows a set of subjects to be grouped based on the measurement of distance. An example of this measurement of distance is the Euclidean distance which is normally preferred for clustering [17]. As mentioned, the two types of clustering method are known as hierarchical and non-hierarchical. One of the examples of hierarchical method is agglomerative method, whereby each object is defined as a cluster in with the most similar clusters being combined as new cluster. Additonally, the initial clusters will be further merged according to their similarities. The process of merging is repeated until a single cluster managed to be obtained. A few types of approaches can be used to combine the most similar clusters into a new cluster. In the context of this study, centroid approach is chosen as the most appropriate method because the clusters are merged based on the shortest distance between cluster centroids defined by the squared Euclidean distance [23].

One of the examples of non-hierarchical method is K-means method. In this case, the number of clusters, K needs to be specified at the initial stage, which means that the unknown K is the limitation of K-means. Thus, in hierarchical method, the elbow rule is a rule that will be used to determine the optimal value of K in this study, and will be further discussed in the following section. Random seed points are required to start the algorithm. It is considered as a faster method that can be applied to a large number of datasets compared to the hierarchical method. According to [24], the K-means process consists of three steps. The first step of the process begins with the partitioning of objects into K initial clusters together with the specified K centroids (random seed points). Next, each of the objects is assigned to the cluster with the closet centroid based on the Euclidean distance. Finally, the recalculation of K centroids is performed for the cluster receiving new objects and the cluster losing its objects. This step is repeated until no allocation of the objects is needed.

D. Clustering method in compactly supported radial basis function for surface approximation

In this study, the elbow rule is applied to determine the optimal number of K clusters from a set of noisy data points. This K value will be served as the recommended numbers of reference point. As a result, the K-means method will calculate the K centroids of the clusters once the K is specified. It is important to note that these centroids are not the subset of the data points, but they serve as the reference points of CSRBFs for surface approximation. [25] mentions elbow rule is performed to find an optimal K clusters, while the process of clustering is carried out by K-means. Furthermore, the process to determine the optimal K clusters is initiated by the agglomerative hierarchical method using the centroid approach, while the distance is measured based on the squared Euclidean distance. On top of that, the process to combine the objects at every stage is based on the agglomeration schedule produced using Predictive Analytics Software (PASW). Consequently, the distance between two clusters for each stage number is produced from the agglomeration schedule. The agglomeration schedule is a table that presents the objects or clusters combined at each stage based on distance and clustering method [26]. In addition, the indices or stage numbers in the agglomeration schedule is one less than the number of data points. Next, the distance for each stage number is observed until a stage number that corresponds to the large jumps in the distance value. In this context, the particular stage number is known as the elbow point, while the large jump can be explained as two clusters with different similarities that start to merge together. The optimal K can be calculated as shown in (8).

$$K = \text{number of data points} - \text{elbow point} \tag{8}$$

The noise with different noise levels is added to the noise-free 3D data points in order to resemble the data points to real life data and to test the approximation scheme of CSRBFs. The noise is described as the variation in a set of data points. The proposed method, which involves the integration of clustering method in CSRBFs for surface approximation is computed using Mathematica and PASW and described in Algorithm 1.

III. Results and Discussion

In this section, the noise-free data points are randomly generated by the test functions. In the case of generating data points, only three bivariate test functions are selected in this study out of all test functions that can be used to verify the effectiveness of the proposed method. The three bivariate test

Algorithm 1 Integrating clustering method in CSRBFs for surface approximation

1: **Input:** A sample of noise-free 3D data points
2: **Output:** Approximated CSRBF surface
3: Add noise with different noise levels on noise-free data points.
4: Perform agglomerative hierarchical method using centroid approach and squared Euclidean distance.
5: Determine the optimal number of K clusters from the agglomeration schedule using elbow rule and (8).
6: Calculate the centroid of each clusters using K-means non-hierarchical method.
7: Use the calculated centroids as the reference points ξ_j of CSRBF.
8: Set up the linear system of equations based on the equation shown in (6) using Wendland's basis function $\phi_{3,1}(r)$.
9: Solve the overdetermined linear system using QR decomposition.
10: Substitute the obtained value in (5).
11: Perform error analysis and plot the approximated CSRBF surface.

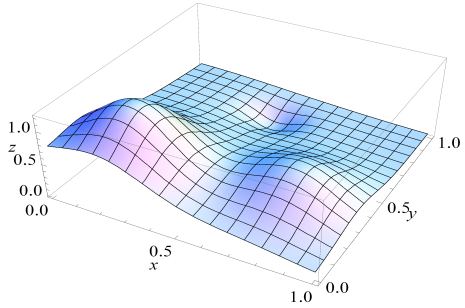
functions are known as F_1, F_7, and F_{10} [27]. The formula for the three test functions are presented as follows:

$$\begin{aligned} F_1(x,y) &= 0.75\exp\left(-\frac{(9x-2)^2+(9y-2)^2}{4}\right) \\ &\quad + 0.75\exp\left(-\frac{(9x+1)^2}{49}-\frac{(9y+1)}{10}\right) \\ &\quad + 0.50\exp\left(-\frac{(9x-7)^2+(9y-3)^2}{4}\right) \\ &\quad - 0.20\exp\left(-(9x-4)^2-(9y-7)^2\right) \\ F_7(x,y) &= 2\cos(10x)\sin(10y)+\sin(10xy) \\ F_{10}(x,y) &= \exp\left(-0.04\sqrt{(80x-40)^2+(90y-45)^2}\right) \\ &\quad \times \cos\left(0.15\sqrt{(80x-40)^2+(90y-45)^2}\right). \end{aligned}$$

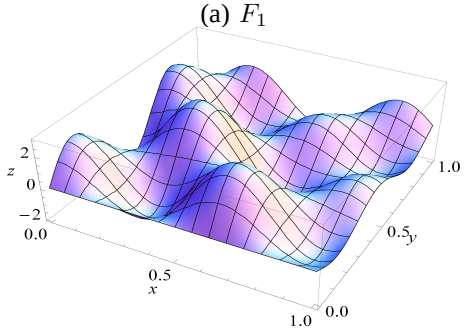
The three test functions have the domain of $[0,1]\times[0,1]$ and their surface can be visualised as in Figure 1.

A total of 4000 2D points are randomly generated using the Mathematica software from the domain $[0,1]\times[0,1]$ in order to sample the 3D data points from the three test function. The similar set of 2D data points is substituted in the three test functions to find its function value in order to generate a set of 3D data points. In this case, the set of 3D data points which is sampled from the three test functions are assumed to be noise-free. The distribution of the 4000 randomly generated 2D data points are shown in Figure 2.

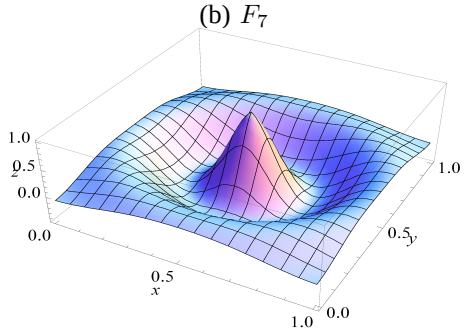
The noise with different noise levels is added to each set of noise-free 3D data points that were sampled from the three test functions. In this case, the noise levels that need to be considered are 0.25, 0.50, 0.75, and 1.00. The $\hat{d}$-noise level implies a simulated noise which is added by normal distribution with the variance of $\hat{d}\cdot h$, whereas h is the average distance between the two nearest points from a set of points. The h values for the set of noise-free data points from F_1, F_7, and F_{10} are 0.01042, 0.02801, and 0.01297, respectively. A total of 3999 stage numbers is generated in the agglomeration schedule using PASW; nevertheless, only



(a) F_1



(b) F_7



(c) F_{10}

Fig. 1: The test functions.

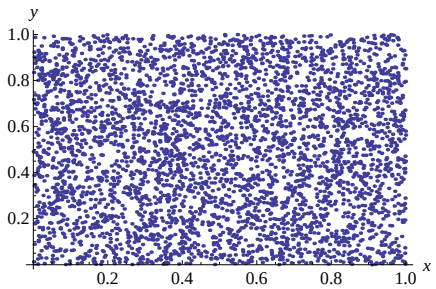


Fig. 2: The randomly generated 2D data points with size 4000 using Mathematica. This set of points is substituted in the test functions in order to generate sets of 3D data point.

the important part is shown and the elbow point is bold together with the optimal K as presented in Tables I–III. The set of data points is presented in green colour whereas the set of reference points is represented in red colour. The following results are based on Algorithm 1.

TABLE I: (a)–(d):Numerical results of elbow rule and the K cluster for F_1 data points with 0.25, 0.50, 0.75, and 1.00 noise level, respectively.

Stage number	Distance
3966	0.027
3967	0.027
3968	0.027
3969	0.028
3970	**0.023**
3971	0.030
3972	0.031
3973	0.032
3974	0.032
3975	0.036
$K = 4000 - 3970 = 30$	

(a)

Stage number	Distance
3964	0.024
3965	0.026
3966	0.026
3967	0.027
3968	**0.027**
3969	0.031
3970	0.031
3971	0.032
3972	0.032
3973	0.033
$K = 4000 - 3968 = 32$	

(b)

Stage number	Distance
3974	0.031
3975	0.032
3976	0.033
3977	0.033
3978	**0.035**
3979	0.039
3980	0.040
3981	0.042
3982	0.042
3983	0.050
$K = 4000 - 3978 = 22$	

(c)

Stage number	Distance
3974	0.034
3975	0.036
3976	0.036
3977	0.037
3978	**0.037**
3979	0.046
3980	0.047
3981	0.051
3982	0.052
3983	0.052
$K = 4000 - 3978 = 22$	

(d)

TABLE II: (a)–(d):Numerical results of elbow rule and the K cluster for F_7 data points with 0.25, 0.50, 0.75, and 1.00 noise level, respectively.

Stage number	Distance
3955	0.117
3956	0.119
3957	0.129
3958	0.131
3959	**0.135**
3960	0.150
3961	0.138
3962	0.150
3963	0.153
3964	0.169
$K = 4000 - 3959 = 41$	

(a)

Stage number	Distance
3947	0.096
3948	0.107
3949	0.108
3950	0.108
3951	**0.110**
3952	0.124
3953	0.133
3954	0.134
3955	0.135
3956	0.140
$K = 4000 - 3951 = 49$	

(b)

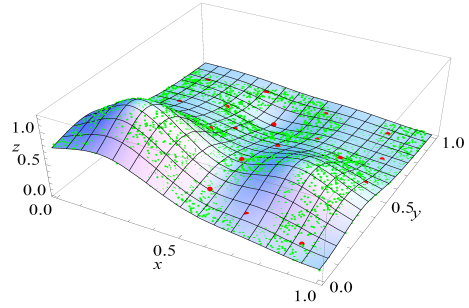
Stage number	Distance
3943	0.100
3944	0.101
3945	0.104
3946	0.102
3947	**0.105**
3948	0.116
3949	0.117
3950	0.120
3951	0.121
3952	0.125
$K = 4000 - 3947 = 53$	

(c)

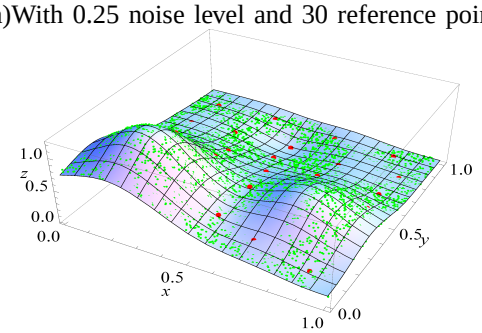
Stage number	Distance
3948	0.105
3949	0.106
3950	0.110
3951	0.110
3952	**0.111**
3953	0.124
3954	0.130
3955	0.134
3956	0.139
3957	0.142
$K = 4000 - 3952 = 48$	

(d)

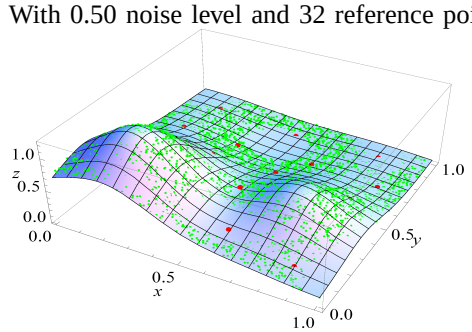
The approximated surfaces of 4000 F_1 data points with noise and the K reference points from Table I using CSRBF are shown in Figure 3.



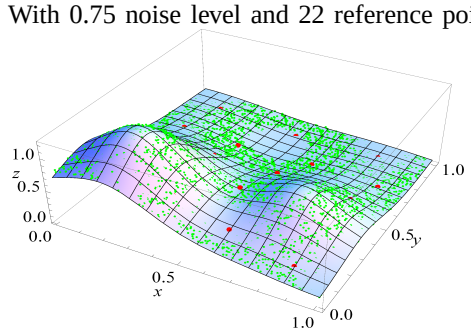
(a)With 0.25 noise level and 30 reference points



(b) With 0.50 noise level and 32 reference points



(c) With 0.75 noise level and 22 reference points



(d) With 1.00 noise level and 22 reference points

Fig. 3: Surface approximation of 4000 F_1 data points with noise and the required reference points using CSRBF.

TABLE III: (a)–(d):Numerical results of elbow rule and the K cluster for F_{10} data points with 0.25, 0.50, 0.75, and 1.00 noise level, respectively.

Stage number	Distance
3966	0.035
3967	0.038
3968	0.041
3969	0.041
3970	**0.043**
3971	0.048
3972	0.049
3973	0.050
3974	0.051
3975	0.052
$K = 4000 - 3970 = 30$	

(a)

Stage number	Distance
3965	0.037
3966	0.038
3967	0.038
3968	0.038
3969	**0.039**
3970	0.044
3971	0.044
3972	0.046
3973	0.047
3974	0.050
$K = 4000 - 3969 = 31$	

(b)

Stage number	Distance
3967	0.035
3968	0.036
3969	0.039
3970	0.040
3971	**0.041**
3972	0.047
3973	0.050
3974	0.050
3975	0.058
3976	0.058
$K = 4000 - 3971 = 29$	

(c)

Stage number	Distance
3972	0.042
3973	0.045
3974	0.046
3975	0.048
3976	**0.050**
3977	0.056
3978	0.061
3979	0.063
3980	0.064
3981	0.065
$K = 4000 - 3976 = 24$	

(d)

The approximated surfaces of 4000 F_7 data points with noise and the K reference points from Table II using CSRBF are shown in Figure 4. Meanwhile, the approximated surfaces of 4000 F_{10} data points with noise and the K reference points from Table III using CSRBF are shown in Figure 5.

In this context, the optimal number of clusters, K is mainly affected by the distance measurement of agglomerative hierarchical method. Therefore, it is safe to say that it is highly dependent on the density and distribution of 3D data points. The presence of different levels of noise is also able to alter the topological distribution. Furthermore, the required number of K reference points for CSRBF approximation was found to be small despite the 4000 data points that were sampled from three test functions. Hence, this scenario leads to the decrease of computational time. The approximated CSRBF surfaces shown in Figures 3–5 are similar to the original noise-free surfaces presented in Figure 1. In relation to this matter, error analysis has to be performed to assess the obtained graphical results. Thus, the mean absolute error (MAE) and root mean square error (RMSE) are unitised to validate the accuracy of the proposed method in this study.

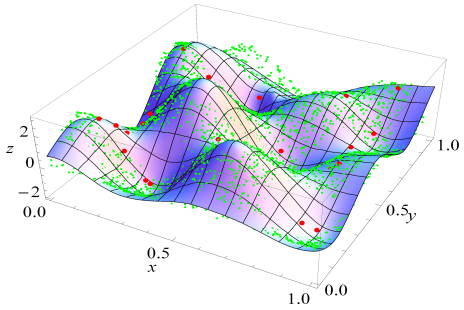
The general formula for MAE and RMSE are given in (9) and (10), respectively.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|v_i - \tilde{v}_i| \tag{9}$$

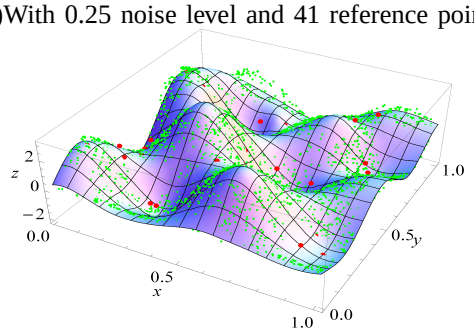
$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i - \tilde{v}_i)^2} \tag{10}$$

where v_i is the actual observation, $\tilde{v}_i$ represents the approximated observation, and N refers to the number of observations.

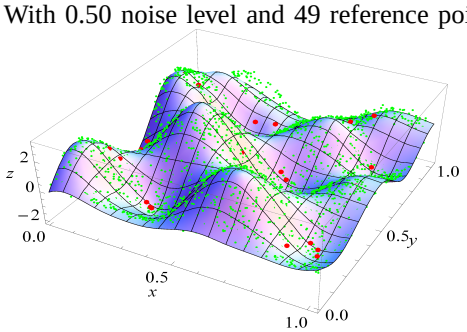
The error analysis performed for CSRBF surface approximation from a set of data points is generated by F_1, F_7, and F_{10} functions in the presence of noise are presented in Table IV.



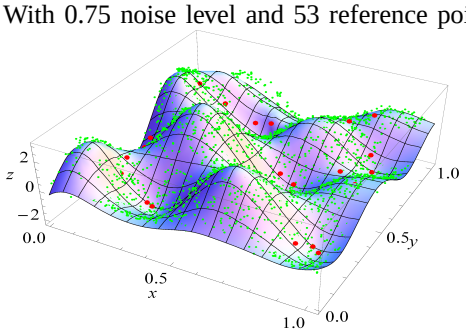
(a)With 0.25 noise level and 41 reference points



(b) With 0.50 noise level and 49 reference points



(c) With 0.75 noise level and 53 reference points

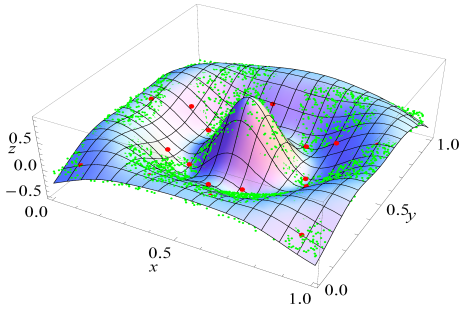


(d) With 1.00 noise level and 48 reference points

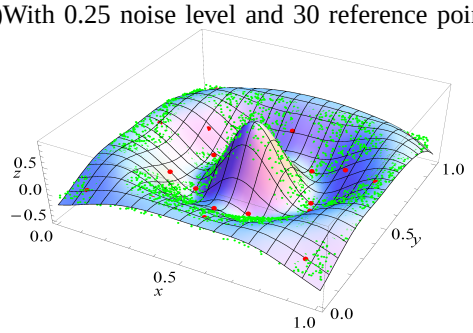
Fig. 4: Surface approximation of 4000 F_7 data points with noise and the required reference points using CSRBF.

TABLE IV: The MAE and RMSE of the CSRBF surface approximation for F_1, F_7, and F_{10} functions with different noise levels.

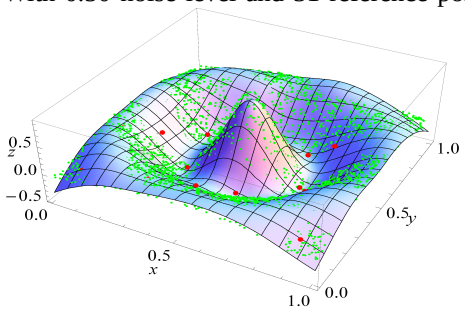
Noise level	Error	CSRBF surface approximation for :		
		F_1	F_7	F_{10}
0.25	MAE	0.007911	0.144920	0.034877
	RMSE	0.010063	0.214644	0.042705
0.50	MAE	0.009011	0.140880	0.033608
	RMSE	0.010672	0.175983	0.042268
0.75	MAE	0.012925	0.154955	0.035601
	RMSE	0.016995	0.161280	0.043450
1.00	MAE	0.014360	0.180871	0.044854
	RMSE	0.017439	0.184489	0.057994



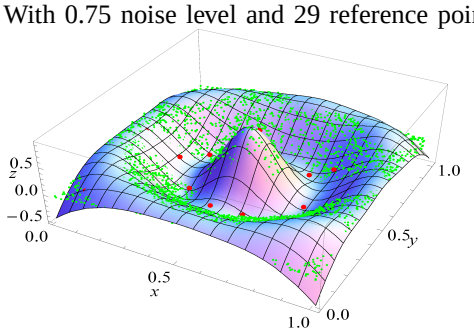
(a)With 0.25 noise level and 30 reference points



(b) With 0.50 noise level and 31 reference points



(c) With 0.75 noise level and 29 reference points



(d) With 1.00 noise level and 24 reference points

Fig. 5: Surface approximation of 4000 F_{10} data points with noise and the required reference points using CSRBF.

From Table IV, the recommended K obtained from Algorithm 1 contributes to the small MAE and RMSE values. Furthermore, it is possible to explain that small MAE value indicates that the approximated data points for the three test functions using CSRBF are closer to the noise-free original data points though there is an increment of the noise levels. For the purpose of approximating the surface from a set of noisy data points with different noise levels, the small value of RMSE represents high accuracy of CSRBF approximation model. On top of that, it has been noted the approximation of the complicated surface, such as from a set of F_7 data points requires larger number of K compared to the F_1 and F_{10} data points. Apart from that, the increase of number of K will lead to the decrease of RMSE values, which means the increment in accuracy. However, this leads to the increment of computational time when solving the system in (6). In addition, the MAE values will decrease as well. But, the noise pollution is still the main contribution for these MAE values. Nonetheless, the computational times tend to decrease when the number of K is decreased, which allows both MAE and RMSE values to increase. Therefore, the recommended number of reference points is important in the CSRBF approximation due to the trade-off between the accuracy and the computational time.

IV. Conclusion

In this paper, an integration of clustering method was proposed to determine the optimal number of reference points in the surface approximation. It is also considered as a simple approach because it requires less memory, high speed computation, and stability. Overall, it can be concluded that this process can be accomplished using CSRBF, clustering method, and elbow rule. Meanwhile, the function of QR decomposition is to ensure that the linear system is solvable. The experiment which involves large numbers of points will be the future research due to the limitation in computation capacity. Finally, the proposed method is hoped to contribute to the growing body of literature on CSRBF surface approximation.

Acknowledgment

The authors like to extend their gratitude to the School of Mathematical Sciences, Universiti Sains Malaysia. Finally, the first and the second authors of this manuscript also gratefully acknowledge the generous financial support contributed by the Ministry of Higher Education Malaysia under postdoctoral fellowship and MyMaster scholarship, respectively.

References

[1] R. L. Hardy, “Multiquadric equations of topography and other irregular surfaces,” *Journal of Geophysical Research*, vol. 76, no. 8, pp. 1905–1915, March 1971.

[2] S. M. Wong, Y. C. Hon, and T. S. Li, “Radial basis functions with compactly supported and multizone decomposition: applications to environmental modelling,” *Boundary Element Technology*, vol. 13, pp. 355–364, 1999.

[3] T. Hacib, M. Mekideche, and N. Ferkha, “Computational investigation on the use of fem and rbf neural network in the inverse electromagnetic problem of parameter identification,” *IAENG International Journal of Computer Science*, vol. 33, no. 2, pp. 18–24, 2007.

[4] A. Rubio-Solis and G. Panoutsos, “Interval type-2 radial basis function neural network: a modeling framework,” *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 2, pp. 457–473, April 2015.

[5] O. Haddadi, Z. Abbasi, and H. TooToonchy, “The hamming code performance analysis using rbf neural network,” in *Proceedings of the World Congress on Engineering and Computer Science*, vol. 2, 2014.

[6] Y. Yang, Z. Wang, X. Li, and M. Fu, “Real-time terrain estimation based on multi-scale radial basis function for unmanned ground vehicle,” in *2016 IEEE International Conference on Information and Automation (ICIA)*, August 2016, pp. 659–664.

[7] S. E. Hubera and M. R. Trummerb, “Radial basis functions for solving differential equations: Ill-conditioned matrices and numerical stability,” *Computers & Mathematics with Applications*, vol. 71, no. 1, pp. 319–327, January 2016.

[8] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans, “Reconstruction and representation of 3d objects with radial basis functions,” in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 2001, pp. 67–76.

[9] L. Wang, B. Yuan, and X. Tang, "Implicit surface reconstruction from noisy 3d scattered data," in *2006 8th International Conference on Signal Processing*, vol. 2, 2006.

[10] K. J. Liew, A. Ramli, and A. Abd Majid, "Point set denoising using bootstrap-based radial basis function," *PLOS ONE*, vol. 11, no. 6, pp. 1–18, 2016.

[11] J. Duchon, *Constructive Theory of Functions of Several Variables*. Springer Berlin Heidelberg, 1977, ch. Splines Minimizing Rotation-invariant Semi-norms in Sobolev Spaces, pp. 85–100.

[12] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree," *Advances in Computational Mathematics*, vol. 4, no. 1, pp. 389–396, 1995.

[13] L. Ling, "An adaptive-hybrid meshfree approximation method," *International Journal for Numerical Methods in Engineering*, vol. 89, no. 5, pp. 637–657, 2012.

[14] A. Gelas, Y. Ohtake, T. Kanai, and R. Prost, "Approximation of unorganized point set with composite implicit surface," in *2006 International Conference on Image Processing*, Oct 2006, pp. 1217–1220.

[15] A. Gelas and R. Prost, "Multi-resolution reconstruction of irregularly sampled signals with compactly supported radial basis functions," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 3, May 2006, pp. 388–391.

[16] Z. Majdisova and V. Skala, "A radial basis function approximation for large datasets," in *Proceedings of SIGRAD 2016, Visby, Sweden*, no. 127. Linköping University Electronic Press, 2016, pp. 9–14.

[17] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statiscal Analysis*. Pearson Education, Inc., 2007, ch. Clustering, Distance Methods, and Ordination, pp. 671–701.

[18] Z. Wu, "Compactly supported positive definite radial functions," *Advances in Computational Mathematics*, vol. 4, no. 1, pp. 283–292, 1995.

[19] S. Zhu, "Compactly supported radial basis functions: how and why?" *SIAM Review*, 2012.

[20] G. Fasshauer, "On smoothing for multilevel approximation with radial basis functions," in *Vanderbilt University Press, Nashville TN*. University Press, 1999, pp. 55–62.

[21] Z. Majdisova and V. Skala, "A new radial basis function approximation with reproduction," in *Proceedings of the International Conferences on Interfaces and Human Computer Interaction 2016*, 2016, pp. 1–8.

[22] C. Hennig, M. Meila, F. Murtagh, and R. Rocci, Eds., *Handbook of Cluster Analysis*. Taylor & Francis Group, LLC, 2016, ch. Cluster Analayis: An Overview, pp. 1–19.

[23] M. J. Norušis, *SPSS Advanced Statistics: Student Guide*. SPSS Incorporated, 1990, ch. Cluster Analysis, pp. 200–201.

[24] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif: University of California Press, 1967, pp. 281–297.

[25] P. Bholowalia and A. Kumar, "Ebk-means: A clustering technique based on elbow method and k-means in wsn," *International Journal of Computer Applications*, vol. 105, no. 9, pp. 17–24, November 2014.

[26] D. L. Hahs-Vaughn, *Applied Multivariate Statistical Concepts*. Taylor & Francis, 2016, ch. Cluster Analysis, pp. 342–343.

[27] R. J. Renka and R. Brown, "Algorithm 792: accuracy test of acm algorithms for interpolation of scattered data in the plane," *ACM Transactions on Mathematical Software (TOMS)*, vol. 25, no. 1, pp. 78–94, 1999.