Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Disertační práce

2019 **Ing. Bc. Zuzana Majdišová**

Západočeská univerzita v Plzni
Fakulta aplikovaných věd

INTERPOLAČNÍ A APROXIMAČNÍ TECHNIKY PRO ROZSÁHLÁ GEOMETRICKÁ DATA

Ing. Bc. Zuzana Majdišová

disertační práce
k získání akademického titulu doktor
v oboru Informatika a výpočetní technika

Školitel: Prof. Ing. Václav Skala, CSc.
Katedra: Katedra informatiky a výpočetní techniky

Plzeň 2019

University of West Bohemia
Faculty of Applied Sciences

INTERPOLATION AND APPROXIMATION METHODS FOR LARGE GEOMETRIC DATASETS

Ing. Bc. Zuzana Majdišová

Doctoral thesis
submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in Computer Science and Engineering

Supervisor: Prof. Ing. Václav Skala, CSc.
Department: Department of Computer Science and Engineering

Pilsen 2019

Prohlášení

Předkládám tímto k posouzení a obhajobě disertační práci zpracovanou na závěr doktorského studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni. Prohlašuji, že tuto práci jsem zpracovala samostatně s použitím odborné literatury a dostupných pramenů uvedených v seznamu, jenž je součástí této práce.

V Plzni dne 18. října 2019 Ing. Bc. Zuzana Majdišová

Abstrakt

Rekonstrukce velkých roztroušených dat pomocí některé z interpolačních nebo aproximačních metod je častým úkolem v mnoha technických aplikacích. Pro interpolaci nebo aproximaci dat bylo sice vyvinuto několik technik, ale obvykle vyžadují, aby byla data v nějakém smyslu uspořádána, např. pravoúhlá síť, strukturovaná síť, nestrukturovaná síť, atd. Bohužel konverze roztroušených dat na polopravidelnou mřížku pomocí některé z teselačních technik je výpočetně velmi náročná. Proto se tato práce zaměřuje na metody využívající radiální bázové funkce (RBF), které jsou vhodné pro zpracování velkých roztroušených dat v n-dimenzionálním prostoru.

RBF metody jsou neseparabilní, jelikož jsou založeny na výpočtu vzdálenosti mezi dvěma body. Tyto metody vedou na řešení systému lineárních rovnic, který je v případě použití aproximační metody přeurčený. Jednou z výhod RBF metod oproti triangulačním metodám je získání analytického popisu daného datasetu. Navíc je v případě RBF aproximace dosaženo významné komprese dat.

Tato práce je vypracována jako soubor komentovaných odborných článků sepsaných autorkou práce a jejích spolupracovníků během autorčina doktorského studia. Články se zaměřují především na výzkum v oblasti RBF aproximace. Příspěvky prezentované v této práci lze rozdělit do tří vzájemně propojených dílčích oblastí: RBF aproximace s polynomiální reprodukcí, RBF aproximace pro velká data a RBF aproximace respektující hlavní rysy dat. Dva z prezentovaných příspěvků byly publikovány v impaktovaných odborných časopisech, jeden článek je přijat k publikovaní v impaktovaném časopisu a čtyři příspěvky byly publikovány ve sbornících mezinárodních konferencí. Důležitou částí této práce je příloha, která obsahuje jednotlivé články v jejich publikované podobě.

Tato disertační práce byla podporována následujícími projekty:

- GA17-05534S – Meshless metody pro vizualizaci velkých časově-prostorových vektorových dat
  Grantová agentura České republiky (GAČR)
- LH12181 – NECPA - Vývoj algoritmů počítačové grafiky a pro CAD/CAM systémy
  Ministerstvo školství, mládeže a tělovýchovy (MŠMT)
- SGS-2013-029 – Pokročilé výpočetní a informační systémy
  Západočeské univerzita v Plzni (ZČU)
- SGS-2016-013 – Pokročilé grafické a výpočetní systémy
  Západočeské univerzita v Plzni (ZČU)
- SGS-2019-016 – Syntéza a analýza geometrických a výpočetních modelů
  Západočeské univerzita v Plzni (ZČU)

Abstract

A reconstruction of large scattered datasets using interpolation or approximation methods is often task in many engineering problems. Several techniques have been developed for data interpolation or approximation, but they usually require an ordered dataset, e.g. rectangular mesh, structured mesh, unstructured mesh, etc. Nevertheless, the conversion of a scattered dataset to a semiregular grid using some tessellation techniques is computationally expensive. Therefore, the thesis is focused to the Radial Basis Function (RBF) methods which are appropriate for large scattered dataset in n-dimensional space.

The RBF methods are non-separable, as it is based on the distance between two points. These methods lead to the solution of a linear system of equation which is overdetermined in the case of the use of some approximation method. Using RBF methods, the analytical description of the data is obtained which is the one of the advantages of such methods over the triangulation methods. Moreover, in the case of RBF approximation methods, the significant compression of the give data is achieved.

The thesis is elaborated as a collection of commented research papers which were written by the author of this thesis and her collaborators during the author's doctoral study. The papers focus mostly on the research in area of the RBF approximation. The presented contributions are from three interconnected subareas: RBF approximation with polynomial reproduction, RBF approximation for big data and RBF approximation respecting features of data. Two of the presented research papers were published in the impacted international journals, one paper is accepted for journal publication and four other papers were published in proceedings of international conferences. Therefore, substantial part forming this thesis is an appendix where the articles are attached.

This dissertation thesis was supported by the following projects:

- GA17-05534S – Meshless Methods for Large Scattered Spatio-temporal Vector Data Visualization
  Czech Science Foundation (GAČR)
- LH12181 – Development of Algorithms for Computer Graphics and CAD/CAM systems
  Ministry of Education, Youth and Sports (MŠMT)
- SGS-2013-029 – Advanced Computing and Information Systems
  University of West Bohemia (UWB)
- SGS-2016-013 – Advanced Graphical and Computing Systems
  University of West Bohemia (UWB)
- SGS-2019-016 – Synthesis and Analysis of Geometric and Computing Models
  University of West Bohemia (UWB)

Acknowledgments

Great thanks belong to my supervisor Prof. Ing. Václav Skala, CSc. for his time, valuable comments, great support and patience. Further, I would like to express my gratitude to my collaborator Michal Šmolík who contributed to the work presented in the thesis. I would like to thank also my other colleagues for their numerous and practical advice. I also want to express my thanks to people from the Department of Computer Science and Engineering from the Faculty of Applied Sciences, University of West Bohemia in Pilsen for great support and pleasant working environment. Last but certainly not least, my thank also belong to my family and my friends whose support was very important to me during my studies.

Contents

1 Introduction **3**
- 1.1 Radial Basis Function (RBF) 5
- 1.2 RBF Approximation 8
- 1.3 Problem Definition 9

2 Overview of Contributions **10**
- 2.1 RBF Approximation with Polynomial Reproduction 10
  - 2.1.1 A New Radial Basis Function Approximation with Reproduction 10
  - 2.1.2 Radial Basis Function Approximations: Comparison and Applications 11
  - 2.1.3 Summary of Achieved Results 11
- 2.2 RBF Approximation for Big Data 12
  - 2.2.1 A Radial Basis Function Approximation for Large Dataset . . 12
  - 2.2.2 Big Geo Data Surface Approximation using Radial Basis Function: A Comparative Study 13
  - 2.2.3 Summary of Achieved Results 14
- 2.3 RBF Approximation Respecting Features of Data 15
  - 2.3.1 Algorithm for Placement of Reference Points and Choice of an Appropriate Variable Shape Parameter for the RBF Approximation 15
  - 2.3.2 Determination of Stationary Points and Their Bindings in Dataset Using RBF Methods 17
  - 2.3.3 Incremental Meshfree Approximation of Real Geographic Data 18
  - 2.3.4 Summary of Achieved Results 19

3 Summary and Future Work **20**

References **22**

Publications of Author **30**

A Project Assignments, Other Activities **34**
- A.1 Conferences and Talks 34

A.2 Participation on Scientific Projects 35
A.3 Teaching Activities 35

B A New Radial Basis Function Approximation with Reproduction 36

C Radial basis function approximations: comparison and applications 45

D A Radial Basis Function Approximation for Large Datasets 62

E Big geo data surface approximation using radial basis functions: A comparative study 69

F Algorithm for Placement of Reference Points and Choice of an Appropriate Variable Shape Parameter for the RBF Approximation 78

G Determination of Stationary Points and Their Bindings in Dataset Using RBF Methods 95

H Incremental Meshfree Approximation of Real Geographic Data 108

Chapter 1

Introduction

Interpolation and approximation are the most frequent operations used in computational techniques. Several techniques have been developed for data interpolation or approximation, but an ordered dataset is mostly expected, e.g. a rectangular mesh, a structured mesh, an unstructured mesh, etc. However, in many engineering problems, data are not ordered, and they are scattered in $n-$dimensional space, in general. In fact, the conversion of a scattered dataset to a semi-regular grid is commonly performed in technical applications using some tessellation techniques. However, this approach is quite prohibitive for the case of $n-$dimensional data due to the computational cost.

Interesting techniques are based on the Radial Basis Function (RBF) method which was originally introduced by [Har71], [Har90] and it is a traditional and powerful tool for the meshless interpolation and approximation of scattered data. These meshless techniques do not require conversion to a semi-regular grid. A good introduction to RBFs is given by [Buh03]. RBF techniques are widely used across many fields solving technical and non-technical problems. RBF techniques are effective tools for solving partial differential equations in engineering and sciences [ECS19], [HSfY15], [LCC13], [Isk04]. RBF applications can be found in fuzzy systems, pattern recognition, data visualization [PRF14], medical applications, surface reconstruction [CLZ+18], [SS18], [IdSPT14], [SPN14], [SPN13], [PS11a], [PS11b], [KHS03], [TO02], [DTS02], [CBC+01], scalar and vector fields visualization [SS17], [SSM18], [SSM19], reconstruction of corrupted images [US05], [ZVS09], etc. The RBF techniques are really meshless and are based on collocation in a set of scattered nodes. These methods are independent with respect to the dimension of the space and lead to the solution of linear system of equations. The computational cost of RBF approximation increases nonlinearly (almost cubic) with the number of points in the given dataset, and linearly with the dimensionality of data. The RBF techniques express the given data using analytical description. Moreover, RBF approximation allows to attain significant compression of the data. It should be noted that the methods for surface reconstruction can be divided into two groups in terms of type of mathematical representation. The first group are RBF methods which lead to the explicit surface representation, i.e. $f(\boldsymbol{x}_i) = h_i$ is solved. The second group of RBF methods leads to implicit representation of surface, i.e. $F(\boldsymbol{x}_i) = 0$ is solved,which

is more difficult because some other constraints have to be added. Therefore, several generalized radial basis function methods which are based on processing of Hermite data, i.e. set of points with their normals, have been developed and were presented in [ZWB19], [BG13], [GTBM13], [MGV11], [MGV09], [PMW09].

Of course, there are other meshless techniques such as discrete smooth interpolation (DSI) [Mal89], kriging [RV84], [MRW+14], [Cre15], which is based on statistical models that include autocorrelation, etc.

The processed point clouds are mostly created by $3D$ scanner, and they contain very large amount of points. Moreover, points in the cloud may not be uniformly distributed, and the holes can be formed. Therefore, it is necessary to develop the methods which are fast and able to reconstruct corrupted datasets [OBS06], [OBS05], [TRS04], [OBA+03], etc. for mathematical representation of surface. It is possible to obtain further acceleration of calculation using high performance computing, such approaches are introduced e.g. in [TGB14] and [CGGS13]. The processing of large amount of points is discussed and solved in [Ska17b] and [Ska17a].

The RBF approximation is generally faster than the RBF interpolation, but the larger errors and inaccuracies are produced. To solve such problems, the modified, robust moving least square method was presented in [JCW+15].

As mentioned above, RBF methods lead to the solution of a linear system of the size equal to the number of data points, further, current $3D$ data scanners allow acquisition of tens of millions points, thus, there is also an important task to find appropriate data structures for storing the point clouds, reconstructed surfaces and representation of matrix of the linear system. There are many publications [SMP+15], [Law13], [LK11], [Šim09], [BG09], [SHK09] [LD08], [BG08], [LZ06], [Mas03] etc. in which introduce a description of several useful data structures.

Another important factor playing a significant role in terms of the quality of approximation and the compression ratio is the good placement of the reference points (i.e. centers of the RBFs) for RBF approximation. The mentioned requirement is fulfilled for placement along significant features of the given dataset. For example, when the geographic data is to be approximated, placement along features such as ridges, peaks, valleys etc. leads to better approximation results.

Further, choice of an appropriate shape parameter of RBFs is extremely important to ensure good approximation. Various authors have focused their research on finding optimal values of the shape parameter for RBFs. Several articles have been dedicated to introducing different algorithms to compute a constant value as an appropriate value for the shape parameter [FZ07], [Fra82], [GIS12], [HLC07], [Rip99], [Sch11]. Many of these focus on finding the minimal error in computations or are based on convergence analysis. Other articles show that variable shape parameters are useful instead of a fixed shape parameter. Sufficient conditions to guarantee a unique solution of the RBF interpolation with variable shape parameters are derived in [ZW15] for CS-RBFs and in [BLRS15], [BLRS04] for global RBFs. The variable shape parameters are possible to determine using a genetic algorithm [AE15] and minimization of the local cost function [SS13], [BLS02] or numerically by minimizing the root-mean-square errors [KC92].

Other approaches generate the variable shape parameters from an estimated range when different distributions of values are used [BH16], [Ran15], [SS09] or use Neural Network RBF approach [XYX+19], [AHJ19c], [Mar19]. However, the approaches mentioned do not reflect features of the given data.

Multiscale RBFs for solving elliptic partial differential equations and distinction between multiscale approximation spaces and multilevel approximation schemes are discussed in [Wen18]. The partition of unity (PU) method which enables to select appropriate size of the different PU subdomains and the suitable shape parameter used to fit the local problems is presented in [CDRP18].

This chapter contains the brief introduce to radial basis functions and the RBF approximation and the summary of thesis goals. Other chapters are organized as follows. Chapter 2 presents an overview of contributions of the papers attached with the thesis. Chapter 3 provides a summary of the research and presents a future work and Appendix A contains the project assignments and other activities of author. Finally, the reprints of the research papers, forming the most important part of the thesis, are provided in appendices.

1.1 Radial Basis Function (RBF)

Radial basis functions (RBFs) are traditional and powerful tools for the meshless interpolation and approximation of scattered data. These functions are real-valued functions which depend only on the distance from the fixed center point. More precisely, let us consider an univariate function:

$$\phi : [0, \infty) \to \mathbb{R} \tag{1.1}$$

then the radial basis function $\Phi_i : \mathbb{R}^d \to \mathbb{R}$ is defined as:

$$\Phi_i(\boldsymbol{x}) = \phi(r_i) = \phi(||\boldsymbol{x} - \boldsymbol{x}_i||), \tag{1.2}$$

where $\{\boldsymbol{x}_i\}_{i=1}^N \subset \mathbb{R}^d$ is a set of N different points which are so-called the centers and $||.||$ is some norm in $\mathbb{R}^d$. The Euclidean norm is usually used.

Name of RBF (specifically word radial) is based on the following property. The value of Φ at any point at a certain fixed distance from the fixed center point is constant, i.e.

$$||\boldsymbol{x}_1|| = ||\boldsymbol{x}_2|| \quad \Rightarrow \quad \Phi(\boldsymbol{x}_1) = \Phi(\boldsymbol{x}_2), \qquad \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d. \tag{1.3}$$

Thus, Φ is radially (or spherically) symmetric around its center. Example of a such function can be seen in Figure 1.1.

Nice property of RBF interpolants is invariance for all Euclidean transformations (i.e. translations, rotations and reflections). This means that it does not matter whether we first compute the RBF interpolant and then apply a Euclidean transformations, or if the transform of the data is performed first and then compute the interpolant. This is a consequence of fact that Euclidean transformations are 2-norm-invariant.

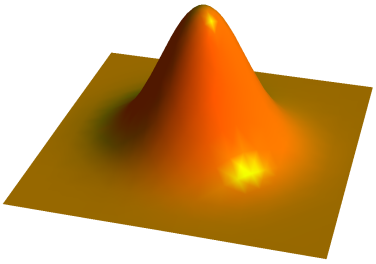


Figure 1.1: Example of RBF

Table 1.1: Typical examples of global RBFs

Global RBF	$\boldsymbol{\phi}(\mathbf{r})$
Gaussian function [Sch79]	$e^{-(\alpha r)^2}$
Inverse Quadric (IQ)	$\frac{1}{1+(\alpha r)^2}$
Inverse Multiquadric (IMQ)	$\frac{1}{\sqrt{1+(\alpha r)^2}}$
Multiquadric (MQ)	$\sqrt{1+(\alpha r)^2}$
Thin-Plate Spline (TPS) [Duc77]	$r^2\log(r)$

There are two main groups of basis functions: global RBFs [Duc77], [Sch79] and "local" Compactly Supported RBFs (CS-RBFs) [Wen06]. Fitting scattered data with CS-RBFs leads to a simpler and faster computation, because a system of linear equations has a sparse matrix. However, an approximation using CS-RBFs is quite sensitive to the density of the approximated scattered data. Global RBFs lead to a linear system of equations with a dense and ill-conditioned matrix and their usage is based on sophisticated techniques such as the fast multipole method [Dar00]. Global RBFs are useful in repairing incomplete datasets and they are significantly less sensitive to the density of approximated data as they cover the whole domain. Typical examples of global RBFs are presented in Table 1.1. It should be noted that Gaussian function, inverse quadric and inverse multiquadric are monotonically decreased with increasing radius r, strictly positive definite, infinitely differentiable and convergent to zero. The multiquadric is monotonically increased with increasing radius r, infinitely differentiable and divergent as radius increases. The last popular global RBF is thin plate spline (TPS) which is shape parameter free and divergent as radius increases. TPS has a singularity at the origin which is removable for the function and its first derivative but this singularity is not removable for the second derivative of TPS. Examples of "local" Wendland's CS-RBFs are presented in Table 1.2. Note that the notation $(1-\alpha r)_+^q$ means:

$$(1-\alpha r)_+^q = \begin{cases} (1-\alpha r)^q & \text{if} \quad 0 \le \alpha r \le 1 \\ 0 & \text{if} \quad \alpha r > 1 \end{cases} \tag{1.4}$$

where q presents some exponent and α is a shape parameter. Figure 1.2 shows behavior of selected Wendland's CS-RBFs for shape parameter $\alpha = 1$; on the x axis is a radius value r (negative part is just for illustration of the symmetry properties).

Table 1.2: Typical examples of "local" Wendland's CS-RBFs $\phi_{d,s}$ [Wen95]. Wendland's functions are univariate polynomial of degree $\lfloor d/2 \rfloor + 3s + 1$, they are always positive definite up to a maximal space dimension d and their smoothness is C^{2s}. For more details see Chapter 11.2 in [Fas07].

ID	**CS-RBF**	$\boldsymbol{\phi}$**(r)**
1	Wendland's $\phi_{1,0}$	$(1-\alpha r)_+$
2	Wendland's $\phi_{1,1}$	$(1-\alpha r)_+^3(3\alpha r+1)$
3	Wendland's $\phi_{1,2}$	$(1-\alpha r)_+^5(8(\alpha r)^2+5\alpha r+1)$
4	Wendland's $\phi_{3,0}$	$(1-\alpha r)_+^2$
5	Wendland's $\phi_{3,1}$	$(1-\alpha r)_+^4(4\alpha r+1)$
6	Wendland's $\phi_{3,2}$	$(1-\alpha r)_+^6(35(\alpha r)^2+18\alpha r+3)$
7	Wendland's $\phi_{3,3}$	$(1-\alpha r)_+^8(32(\alpha r)^3+25(\alpha r)^2+8\alpha r+1)$
8	Wendland's $\phi_{5,0}$	$(1-\alpha r)_+^3$
9	Wendland's $\phi_{5,1}$	$(1-\alpha r)_+^5(5\alpha r+1)$
10	Wendland's $\phi_{5,2}$	$(1-\alpha r)_+^7(16(\alpha r)^2+7\alpha r+1)$

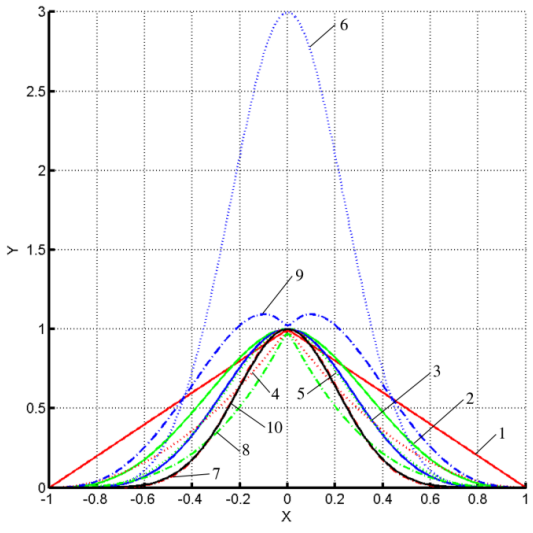


Figure 1.2: Geometrical properties of CS-RBFs [Ska13]

1.2 RBF Approximation

In this section, the RBF approximation method, which was recently introduced in [Ska13], and its properties are decribed.

For simplicity, we assume that we have an unordered dataset $\{\boldsymbol{x}_i\}_1^N$ in $\mathbb{E}^2$. However, note that this approach is generally applicable for d-dimensional space. Further, each point $\boldsymbol{x}_i$ from the dataset is associated with vector $\boldsymbol{h}_i \in \mathbb{E}^p$ of the given values, where p is the dimension of the vector, or scalar value $h_i \in \mathbb{E}^1$. For an explanation of the RBF approximation, let us consider the case when each point $\boldsymbol{x}_i$ is associated with scalar value h_i, e.g. a $2\,1/2\,D$ surface. Let us introduce a set of new reference points $\{\boldsymbol{\xi}_j\}_1^M$, see Figure 1.3.

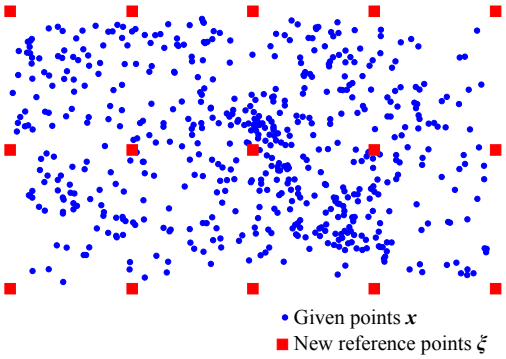


Figure 1.3: RBF approximation and reduction of points.

These reference points may not necessarily be in a uniform grid. It is appropriate that their placement reflects the given surface (e.g. the terrain profile, etc.) as well as possible. The number of added reference points $\boldsymbol{\xi}_j$ is M, where $M \ll N$. The RBF approximation is based on computing the distance of given point $\boldsymbol{x}_i$ and reference point $\boldsymbol{\xi}_j$.

The goal of the RBF approximation is to approximate the given dataset by function:

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \phi(r_j) = \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|), \tag{1.5}$$

where the approximating function $f(\boldsymbol{x})$ is represented as a sum of M RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by an appropriate coefficient c_j.

It can be seen that we get an overdetermined linear system of equations for the given dataset:

$$h_i = f(\boldsymbol{x}_i) = \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|) = \sum_{j=1}^{M} c_j \phi_{i,j} \qquad i = 1, \ldots, N. \tag{1.6}$$

The linear system of equations (1.6) can be represented as the matrix equation:

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{h}, \tag{1.7}$$

where the number of rows is $N \gg M$ and M is the number of unknown weights $[c_1, \ldots, c_M]^T$, i.e. the number of reference points. (1.7) can be expressed in the form:

$$\begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,M} \\ \vdots & \ddots & \vdots \\ \phi_{i,1} & \cdots & \phi_{i,M} \\ \vdots & \ddots & \vdots \\ \phi_{N,1} & \cdots & \phi_{N,M} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_i \\ \vdots \\ h_N \end{pmatrix}. \tag{1.8}$$

Thus, the presented system is overdetermined, i.e. the number of equations N is higher than number of variables M. This linear system of equations can be solved by the least squares method (LSE), QR decomposition etc.

Finally, it should be noted that if the set of reference points is the same as the set of given points, then the problem solving is called the RBF interpolation.

1.3 Problem Definition

This thesis aims to improve the RBF approximation in terms of its stability, solvability, accuracy and compression ratio. For these purposes, the thesis is focused on three interconnected subareas.

The RBF approximation as defined in Section 1.2 can theoretically have problems with stability and solvability. Therefore, the RBF approximant (1.5) is usually extended by polynomial function of degree k. Such RBF approximation is called as the RBF approximation with polynomial reproduction and was introduced by Fasshauer [Fas07]. However, this thesis proves that the mentioned approach causes the inconsistency in the computation and, therefore, the new approach is formulated, clarified and verified. This is the subject of the first subarea that the thesis deals.

The second subarea of this thesis is focused on processing of big scattered data using the RBF approximation. It is very important aspect which has to be considered and solved including the usage of appropriate data structures because as mentioned above, the real data contains tens of millions scattered points.

Finally, the last subarea of the thesis is aimed to improve the quality of the RBF approximation in terms of error and its compression ratio. The main role which affects the results of the RBF approximation is the setting of the shape parameter and the placement of the reference points. Therefore, the last part is focused on these aspects. Moreover, the thesis also deals with the use of the constant shape parameter versus adaptive shape parameter.

Chapter 2

Overview of Contributions

This chapter summarizes main contributions of the work collected in this thesis. The overview is structured into three main sections addressing contributions in the areas of the RBF approximation with polynomial reproduction, the RBF approximation for big data and the RBF approximation respecting features of data.

2.1 RBF Approximation with Polynomial Reproduction

The thesis presents a new approach for the RBF approximation with polynomial reproduction. Moreover, an extensive comparative study of proposed approach with other RBF approximation methods is provided.

2.1.1 A New Radial Basis Function Approximation with Reproduction

Appendix B (paper [MS16a]) presents the new algorithm for the RBF approximation with polynomial reproduction and its mathematical derivation. The classical RBF approximation which was introduced in Section 1.2 can theoretically have problems with stability and solvability. Therefore, the RBF approximant (1.5) is usually extended by polynomial function $P_k(\boldsymbol{x})$ of degree k, i.e.:

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|) + P_k(\boldsymbol{x}). \tag{2.1}$$

The original approach of the RBF approximation with polynomial reproduction was introduced by Fasshauer [Fas07] (Chapter 19.4). However, this original approach applies the additional conditions to the polynomial part and the inconsistency is caused by these conditions. Therefore, the mentioned RBF approximation with the polynomial reproduction is inconveniently formulated as it mixes variables which have a different

physical meaning. Thus, the new approach for the RBF approximation with polynomial reproduction which eliminates the mentioned inconsistency was proposed. The proposed approach is based on the solution of an optimization problem consists of the minimizing the square of error for the approximating function. The experiments prove that the proposed approach is correct and gives significantly better and more stable results than the original method.

2.1.2 Radial Basis Function Approximations: Comparison and Applications

In Appendix C (paper [MS17b]), several existing RBF approximation methods are briefly introduced and their mutually comparison with respect to various criteria is provided. As RBF approximation methods important for comparison, the proposed RBF approximation with reproduction (described in detailed in Appendix B), the classical RBF approximation which was introduced in Section 1.2 and the RBF approximation using Lagrange multipliers introduced by Fasshauer [Fas07] (Chapter 19) have been used. The main emphasis of this comparative study is put on the stability and accuracy of computation. The mentioned methods have been tested on synthetic and real datasets. Moreover, different global radial basis functions, different distributions of reference points and different placement of the dataset in the domain have been used for purposes of comparison.

From the results, it follows that the classical RBF approximation gives the best results due to the smallest error. Nevertheless, the proposed RBF approximation with linear reproduction returns very similar results and the difference in errors compared to the classical RBF approximation in not very significant. Further, the proposed RBF approximation with linear reproduction can be influenced by placement of the given dataset in the space, and therefore, it is appropriate to perform the translation of the estimated center of gravity to the origin of the coordinate system. For all experiments, the worst results according to error were obtained by the RBF approximation using Lagrange multipliers. Moreover, this method has unpredictable behavior and is mostly ill-conditioned. The experiments also proved that the classical RBF approximation and the RBF approximation with reproduction offer a significant data compression. On the other hand, experiments made proved that all methods have problem with the preservation of sharp edges if global RBFs are used.

2.1.3 Summary of Achieved Results

In this section, it will be discussed which significant conclusions can be drawn from our research in the area of the RBF approximation with polynomial reproduction and the related extensive comparative study.

The new algorithm for the RBF approximation with polynomial reproduction and its mathematical derivation was proposed. In comparison with the original approach of

the RBF approximation with polynomial reproduction (see [Fas07] (Chapter 19.4)), the proposed approach has following advantages and properties. The proposed approach does not contain the definition of the additional conditions, and therefore, the inconsistency which causes the mixing variables with a different physical meaning is eliminated. The maximum magnitude of error for the original approach is approximately two times greater that the maximum magnitude of error for the proposed approach. The quality of the RBF approximation in terms of error is better for the proposed approach than for the original approach. When the TPS is used, the quality of the result in terms of error is improved very significant. The proposed approach is easily extendable for general polynomial reproduction and for higher dimensionality.

The comparative study is shown that the numerical stability of the RBF approximation with reproduction can be influenced by placement of the given dataset in space due to large span of the elements in the approximation matrix. Therefore, it is appropriate to apply the translation of the estimated center of gravity to the origin of the coordinate system. Further, it is obvious that the RBF approximation using Lagrange multipliers (see [Fas07] (Chapter 19)) has unpredictable behavior, returns the worst results in terms of error, the matrix for this method is mostly ill-conditioned and its size is high. In terms of global RBFs used, following properties was observed. The Gaussian function has the biggest problems with the preservation of the sharp edges. The Thin-Plate Spline (TPS) is inappropriate for approximation of the synthetic data, on the other hand for real geographic data, it has the advantage that is shape parameter free. In terms of the number and distribution of reference points, it can be concluded that the RBF approximation has problem with the excessive data ordering of the reference points such as points with the uniform distribution. Further, with increasing number of reference points relative to the number of points in the given dataset, the details of the surface are progressively apparent. Moreover, if the number of the reference points is about tenth of the number of approximated points (i.e. $M \approx 0.1 \cdot N$), the mean relative error is smaller than 1%.

2.2 RBF Approximation for Big Data

In practice, real data contain a large number of scattered points which results in high memory requirements for determining their approximation. Therefore, the thesis presents a new method which was proposed for the RBF approximation of big data. Moreover, a modification of this method when the CS-RBFs are used, i.e. the sparse matrix is obtained, have been formulated.

2.2.1 A Radial Basis Function Approximation for Large Dataset

The RBF approximation of big datasets is solved in Appendix D (paper [MS16b]). The RBF approximation leads to the solution of overdetermined linear system of equation (1.8), how it was described in Section 1.2. A large number of points in dataset results

into large matrix of the overdetermined linear system of equation and, therefore, there are the high memory requirements for storing this matrix. Unfortunately, the capacity of RAM memory is limited. Thus, the computation of the RBF approximation would be unbearable computationally expensive due to memory swapping, etc. The new approach for determination of the RBF approximation for such big data was proposed.

The proposed approach is based on the least squares method (LSE) for the solution of the overdetermined linear system of equation, as it was introduced in Section 1.2. For this method a square symmetric matrix, which has a much smaller size than the matrix of linear system, is obtained. Moreover, the determination of each element of such matrix is separable. From the described properties, it follows that the block operations with matrices can be used to save memory requirements and to prevent data bus (PCI) overloading. Futher, due to symmetry of matrix, only upper triangle of the matrix for LSE can be computed, see Figure 2.1. The effectiveness of the proposed procedure is supported by performed experiments.

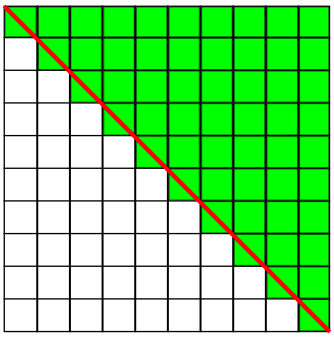


Figure 2.1: $M \times M$ square LSE matrix which is partitioned into $M_B \times M_B$ blocks. The color red is used to denote the main diagonal of the matrix and illustrates the symmetry of the matrix. The color green is used to denote the blocks which have to be computed.

2.2.2 Big Geo Data Surface Approximation using Radial Basis Function: A Comparative Study

Appendix E (paper [MS17a]) focuses on the RBF approximation for big data when the "local" CS-RBFs are used. Moreover, the comparative study is elaborated for this group of basis functions. The proposed approach is the modification of method which is based on partitioning the matrix into blocks and is introduced in Appendix D. It should be noted that the size of block used for the proposed method significantly influences the time performance. If the matrix for LSE is partitioned into small blocks, then the time performance is large due to overhead costs and, moreover, each element of approximation matrix (1.7) has to be calculated more times than for larger size of block. On the other hand, the time performance exceeds the permissible limit for too very large blocks due to memory swapping. Therefore, it is necessary to choose the size of blocks close to optimum.

The modification of previous method, which is described in this contribution, consists in the fact that application of "local" CS-RBFs leads to the sparse matrix for the overdetermined linear system of equations. Therefore, when the special data structures are used for storage of the sparse approximation matrix, then the optimal size of block is much larger than for the classical representation of the matrix. For our purpose, the coordinate (COO) format is most suitable. Moreover, it is evident that we do not want to compute the elements for all pairs of points when CS-RBFs are used, and therefore, the *kd*-tree is used for computing the elements of the approximation matrix. The efficiency of the proposed method is again supported by performed experiments.

2.2.3 Summary of Achieved Results

The significant conclusions of the research in the area of the RBF approximation for big data and advantages and disadvantages of proposed algorithms will be discussed in this section. For purposes of the RBF approximation of the big data, the block-wise approach preventing memory swapping was proposed. The mentioned approach is based on the use of the symmetry of the matrix and partitioning the matrix into blocks, which enables the computation on systems with limited main memory. It should be noted that the efficiency of the proposed approach is strongly dependent on the choice of size of blocks. When the size of block is too small, the time performance is large due to overhead costs. On the other hand, when the size of blocks is too large, the running time begins to rise above the permissible limit due to memory swapping. Therefore, when the "local" CS-RBFs are used (i.e. the sparse matrix is obtained for the overdetermined linear system of equations), the data structures for storage of the sparse matrix can be used and the larger size of blocks can be chosen. This leads as already mentioned to decrease the computational costs. The disadvantage of the proposed approach is dependency on the usage of the least square error method (LSE) which causes worse conditionality of the problem.

Further, during the testing of the proposed approach, following properties and relations have been detected. The RBF methods have problems with the accuracy of computation on the boundary of an object, which is a well-known property. Moreover, the magnitude of error for the RBF approximation is influenced by the presence of a noise in the given dataset. Further, when the input dataset is uniformly distributed within a given area, the "local" CS-RBFs return the better results in terms of the error than the global RBFs. This is caused by fact that global RBFs affect the entire domain of the given datasets, which is usually undesirable behavior. The "local" Wendland's $\phi_{3,0}$ basis function forms sharp undesirable peaks for some types of input data in comparison with other "local" CS-RBFs. The value of constant shape parameter α is depending on the range and number of points of the given dataset.

The results of experiments also proved that the RBF approximation with linear reproduction returns considerably better results in terms of the deviation of error and slightly better results in terms of the mean error than the RBF approximation without polynomial reproduction, particularly if the range of associated values is large.

2.3 RBF Approximation Respecting Features of Data

As already mentioned above, the good placement of the centers of the RBFs and the choice of the appropriate shape parameter for the RBF approximation has an important role to ensure the good quality of approximation and to attain the significant compression of data. For fulfillment of mentioned requirements, the significant features of the given dataset should be given into account. Therefore, the contributions described in this section are focused on use of these significant features for the RBF approximation.

2.3.1 Algorithm for Placement of Reference Points and Choice of an Appropriate Variable Shape Parameter for the RBF Approximation

Two significant aspects which have an influence on the quality of the RBF approximation are determination of the shape parameter and the placement of reference points. Nevertheless, mainly the determination of the optimal shape parameter is difficult problem which is in the majority of cases set up experimentally or using some ad-hoc method. Therefore, a novel algorithm for finding an appropriate set of reference points (i.e. centers of the RBFs) and a variable shape parameter selection for the RBF approximation in $\mathbb{E}^2$ is presented in Appendix F (paper [MSS]). The proposed approach is two-steps and exploits features of the given data such as inflection points and extrema. Determination of the appropriate shape parameter is based on the first curvature of curve.

The briefly summary of the whole proposed algorithm is as follows. As already mentioned, the proposed approach has two steps. The main task of the first step is to perform the primary RBF approximation of the given data so that the associated values will be symmetrically distributed around the x-axis, which leads to the elimination of the problematic course of the sampled function, see Figure 2.2. For these purposes, the inflection points (stationary and non-stationary) and endpoints of the given dataset are TPS (global RBF) interpolated, see Figure 2.3, and an adaptive shift of the given data in terms of associated values is performed. It should be noted that the method for finding significant points of the given dataset (stationary inflection points, non-stationary inflection points, extrema) has to be chosen, because the sampling function for the given

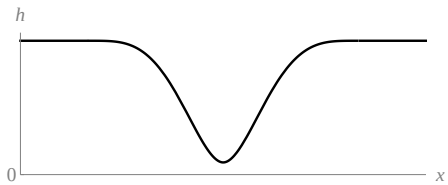


Figure 2.2: The course of the sampled function which is poorly approximated by the RBFs.

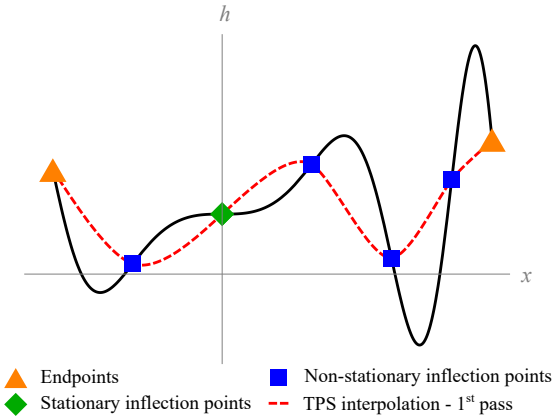
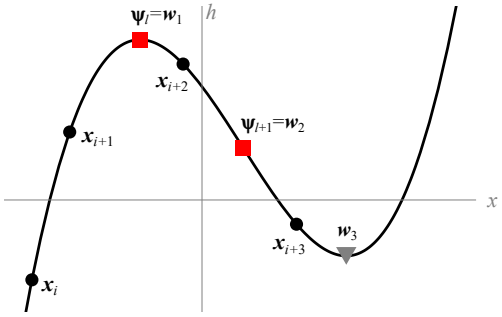
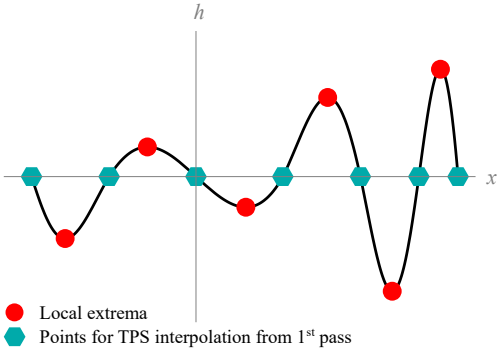


Figure 2.3: Original function and the TPS interpolation for selected significant points (result of the 1st step of our approach).



(a) Four points $\{\boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i+3}\}$ from the dataset interpolated by cubic curve and three significant points $\{\boldsymbol{w}_i\}$ of this curve. The points added to the set of suspicious points (red squares), a point discarded (gray triangle).



(b) Course of the input sampled function for the second step of the proposed approach. The set of reference points for the second step of our approach are marked.

Figure 2.4: Finding significant points of given data.

data is not known. In Appendix F, the method based on piecewise interpolation by a cubic curve was proposed and the four points is used in every step of this piecewise method, see Figure 2.4a.

In the second step of the proposed approach, the RBF approximation with a variable shape parameter and Lagrange multipliers is performed on modified data. The reference points, including the absolute values of the first curvatures in them, are derived using significant points of the shifted data, see Figure 2.4b. For these purposes, the piecewise interpolation by a cubic curve is again uses. After that, the appropriate variable shape parameters are computed according to the the absolute values of the first curvatures at the corresponding reference points. The constraints for the Lagrange multipliers are defined so that the associated values at endpoints are strictly respected. Finally, the approximated value is determined as a sum of the TPS interpolation resulting from the first step of algorithm and the RBF approximation from the second step of the proposed

algorithm:

$$f(\boldsymbol{x}) = \sum_{v=1}^{M_1} c_{Iv}\phi_{TPS}\left(\|\boldsymbol{x} - \hat{\boldsymbol{x}}_v\|\right) + P(\boldsymbol{x}) + \sum_{j=1}^{M} c_j\phi\left(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|, \alpha_j\right), \tag{2.2}$$

where $\boldsymbol{c}_I = (c_{I1}, \ldots, c_{IM_1})$ is the vector of weights for the TPS interpolation, ϕ_{TPS} is the Thin-Plate-Spline, $\{\hat{\boldsymbol{x}}_v\}_1^{M_1}$ are input points for the TPS interpolation, $P(\boldsymbol{x})$ is polynomial function of first order, $\boldsymbol{c} = (c_1, \ldots, c_M)$ is the vector of weights for the RBF approximation, ϕ is the RBF used, $\{\boldsymbol{\xi}_j\}_1^M$ is the set of reference points and $\{\alpha_j\}_1^M$ are appropriate variable shape parameters.

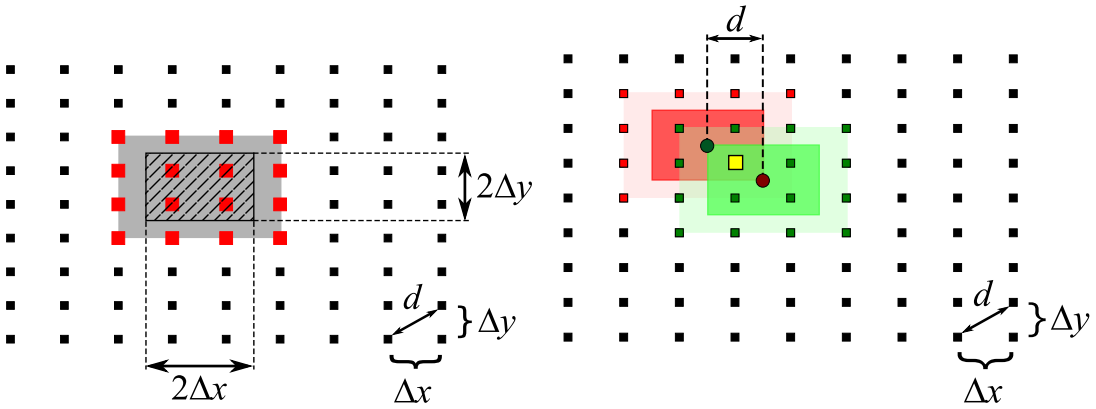
From the experiments, it follows that the proposed approach gives significantly better results over other relevant competing methods. It is also apparent that if the features of the given data are not respected, the RBF approximation is not capable of competing with the RBF approximation respecting features of the data, especially when the number of reference points is too small.

2.3.2 Determination of Stationary Points and Their Bindings in Dataset Using RBF Methods

As already mentioned several times, the placement of reference points with respect to the significant features of the given dataset has an important influence on the result of the RBF approximation. For determination of such significant features of data, the knowledge of stationary points is necessary. However, the given dataset describes the surface for which a sampling function is not known. Therefore, Appendix G (paper [MSS19a]) presents a new algorithm for determining the set of stationary points without knowledge of the sampling function. Moreover, an approach for detecting the bindings between these stationary points (such as stationary points lie on line segment, circle, etc.) is proposed. The knowledge of binding can be appropriate for pruning the subset of related stationary points to the required number of points on the given curve of stationary points. The piecewise RBF interpolation is used during the execution of the algorithm for the determining stationary points of the given dataset.

The main idea of the proposed algorithm is as follows. The all points in the domain are divided into cells, see Figure 2.5a. After that, for the each 3×3 cells (grey area in Figure 2.5a), the RBF interpolation is computed and the stationary points of this interpolating function are determined for the restricted domain (hatched area in Figure 2.5a). Finally, the reduction of the set of stationary points is performed to eliminate identical points or points very close to identical that results from obtaining a stationary point from more RBF interpolations, see Figure 2.5b.

The surface represented by given dataset does not contain only isolated stationary points, but the curves of stationary points such as line segments, circles or some other shapes, can lie on the surface. Therefore, the method of detecting the bindings between stationary points which is based on the *kd*-tree is proposed. Moreover, in the contribution, the derivation of maximal possible distance of two stationary points for



(a) The grey area shows the area of all points from the given dataset which are interpolated by the RBF method during one step of piecewise approach. The hatched area illustrates the domain for which the stationary points of the given dataset are determined from the obtained RBF interpolation.

(b) Visualization of the stationary points reduction which is performed if the two points are identical or very close to identical. The green circle and red circle mark the stationary points which were determined from two different RBF interpolations and which were merged to one stationary point marked by yellow square.

Figure 2.5: Proposed piecewise approach

which these stationary points still lie on the same curve is done. The correctness of the proposed algorithm is proved by performed experiments.

2.3.3 Incremental Meshfree Approximation of Real Geographic Data

In Appendix H (paper [MSS19b]), a new incremental algorithm for the RBF approximation of geographic data that puts the emphasis on good placement of reference points and significantly improves the compression ratio is described. The good placement of reference points for geographic datasets is along features such as ridges, peaks, valleys, etc. Therefore, in the first level, a Gaussian low-pass filter is applied to the given dataset due to elimination of insignificant terrain roughness and the set of stationary points obtained for such filtered data using approach described in Appendix G is used as initial set of reference points. Moreover, this set of reference points is extended by the corners of the bounding box of dataset. It is performed due to avoiding problems on the boundary. After that, the RBF approximation is computed and the residues are determined. In every following level, the set of stationary points for the filtered residues is determined using method described in Appendix G and only local maxima are added to the set of reference points. The RBF approximation is again computed and new residues are determined. The whole process is repeated until the stop conditions are met.

From the experiments, it follows that the quality of the RBF approximation is improving with increasing level of the algorithm. For higher levels, the many details of

the original terrain is already apparent. Moreover, the proposed approach achieves the improvement of results in comparison with other existing methods due to respecting features of the given data.

2.3.4 Summary of Achieved Results

This section is focused on the summary of relevant conclusions obtained from the research in the area of the RBF approximation respecting features of data. As first, the properties and relations of the proposed two-steps RBF approximation with the variable and adaptive shape parameter in $\mathbb{E}^2$ will be described. The proposed approach exploits significant features of the given data to determine reference points and related shape parameters. The experiments proved that the quality of results in terms of error is better for the proposed approach than for other relevant competing methods. Moreover, the RBF approximation for which features of the given data are not respected is not capable of competing with the proposed approach which reflects the features of the given data, especially when the number of reference points is small. This also follows that better compression ratio is obtained for the proposed approach when the mean error of the approximation is fixed. For the real data, the proposed approach can well approximate the global trend of given data for a small set of reference points. The proposed algorithm can be used for the RBF approximation of data describing a curve which is parameterized by one variable in multidimensional space, e.g. a robot path planning, etc. The proposed method also significantly eliminates problems with a shape parameter estimation inherited from the RBF's general properties.

Further, the new incremental approach for the RBF approximation which is based on the determination of stationary points of the input points cloud in the first level and the finding local maxima of residues at each hierarchical level was developed. For purposes of determination such significant points, the new algorithm which is able to determine the stationary points of given sampled surface without knowledge of the sampling function was proposed. This algorithm is based on the piecewise RBF interpolation and, moreover, includes method of detecting the binding between the found points, i.e. associate the points from the same curve. It should be noted that for real data, it is appropriate to filter the given data using a Gaussian low-pass filter before determination of their stationary points. The main reason for filtration of data is a elimination of insignificant stationary points which can cause decline of compression ratio without some benefit for further processing. The proposed incremental RBF approximation achieves the improvement of results in comparison with other existing competing methods when the ratio between number of reference points and number of input points is kept because the features of the given dataset are respected. Moreover, from performed experiments, it was observed that when the approximated data contains some ridges or valleys, the several reference points have to be placement along this data characteristic. Otherwise, such features could be approximated by several peaks or holes.

Chapter 3

Summary and Future Work

This thesis presents issues of the approximation of scattered geometric datasets. For these purposes, the RBF approximation is considered and its improving in terms of different criteria is discussed and solved. The most important part of the thesis is the collection of author's papers. The thesis addresses a variety of different aspects of the RBF approximation: its advantages and disadvantages, stability, solvability, accuracy, compression ratio etc. As shown in the thesis, the RBF approximation provides solution which leads to a significant compression of the given geometric data and the analytical formula is obtained for description of the data. Moreover, the RBF approximation does not require some type of data ordering and, therefore, it is a convenient technique, especially for engineering applications, where the non-linear approximation of data is needed and the large amount of data has to be processed.

The thesis presents several new or modified approaches for the RBF approximation which can be divided to three interconnected subareas. The first subarea is focused to the RBF approximation with reproduction which deals the problems with the stability and solvability of the RBF methods. The algorithm presented in this thesis (see Appendix B and Appendix C) eliminates the inconsistency which occurs in the original approach introduced by Fasshauer [Fas07]. In the thesis, it was proved that the inconsistency is caused by adding additional conditions. Therefore, the new approach was proposed and verified.

The second subarea is dealt with the processing of the big data using the RBF approximation. For these purposes, the block-wise algorithm for the RBF approximation (see Appendix D), which uses the symmetry of the matrix, was proposed. Moreover, for the CS-RBFs, its modification (see Appendix E) based on the sparse data structures and $kd-$ tree was defined.

Finally, the last subarea is focused on the improving of the quality of the RBF approximation and the compression ratio. This is closely related to the placement of reference points and the selection of the shape parameter. The novel algorithm for finding an appropriate set of reference points and a variable shape parameter selection for the RBF approximation in $\mathbb{E}^2$ was developed (see Appendix F). The determination of reference points is based on finding the significant features such as extrema, inflection

points etc. The variable shape parameters are derived using the first curvature of curve. Further, the incremental RBF approximation for geographic datasets was proposed (see Appendix H). In this case, the reference points are selected based on the determination of stationary points of filtered residues at each hierarchical level. For these purposes, the algorithm for determination of stationary points of the given dataset without knowledge the sampling function based on the piecewise RBF interpolation was developed (see Appendix G).

Still, there is a lot of space for future work, mainly in terms of improving the quality of the RBF approximation and increasing the compression ratio. Generally, the RBF methods have problem with the accuracy of computation on the boundary of an approximated object. The preservation of sharp edges is also problematic but appropriate placement of the reference points along the significant features of the given data leads to a partial improvement of this problem. Further, the error of the RBF methods is influenced by the presence of a noise in the input data, thus, there is a space for development in terms of lower sensitivity to noise. The question also remains whether optimal shape parameter can be derived empirically and ideally adaptively for the higher dimension of space. Another task possible to solve is additional development in the area of processing of big data so that the least square error method (LSE) will not need to be used because this method leads to worse conditionality of the problem.

References

[AE15] F. Afiatdoust and M. Esmaeilbeigi, "Optimal variable shape parameters using genetic algorithm for radial basis function approximation," *Ain Shams Engineering Journal*, vol. 6, no. 2, pp. 639 – 647, 2015.

[AHJ19] L. Aik, T. Hong, and A. Junoh, "A new formula to determine the optimal dataset size for training neural networks," *ARPN Journal of Engineering and Applied Sciences*, vol. 14, no. 1, pp. 52–61, 2019.

[BG08] N. Bell and M. Garland, "Efficient sparse matrix-vector multiplication on cuda," tech. rep., Nvidia Technical Report NVR-2008-004, Nvidia Corporation, 2008.

[BG09] N. Bell and M. Garland, "Implementing sparse matrix-vector multiplication on throughput-oriented processors," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, p. 18, ACM, 2009.

[BG13] H. C. Batagelo and J. a. P. Gois, "Least-squares hermite radial basis functions implicits with adaptive sampling," in *Proceedings of Graphics Interface 2013*, GI '13, pp. 109–116, Canadian Information Processing Society, 2013.

[BH16] J. Biazar and M. Hosami, "Selection of an interval for variable shape parameter in approximation by radial basis functions," *Advances in Numerical Analysis*, vol. 2016, 2016.

[BLRS04] M. Bozzini, L. Lenarduzzi, M. Rossini, and R. Schaback, "Interpolation by basis functions of different scales and shapes," *CALCOLO*, vol. 41, pp. 77–87, Jun 2004.

[BLRS15] M. Bozzini, L. Lenarduzzi, M. Rossini, and R. Schaback, "Interpolation with variably scaled kernels," *IMA Journal of Numerical Analysis*, vol. 35, no. 1, pp. 199–219, 2015.

[BLS02] M. Bozzini, L. Lenarduzzi, and R. Schaback, "Adaptive interpolation by scaled multiquadrics," *Advances in Computational Mathematics*, vol. 16, pp. 375–387, May 2002.

[Buh03] M. D. Buhmann, *Radial Basis Functions: Theory and Implementations*, vol. 12. Cambridge university press, 2003.

[CBC+01] 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, SIGGRAPH 2001, Los Angeles, California, USA, August 12-17, 2001*, pp. 67–76, 2001.

[CDRP18] R. Cavoretto, A. De Rossi, and E. Perracchione, "Optimal selection of local approximants in rbf-pu interpolation," *Journal of Scientific Computing*, vol. 74, pp. 1–22, Jan 2018.

[CGGS13] S. Cuomo, A. Galletti, G. Giunta, and A. Starace, "Surface reconstruction from scattered point via RBF interpolation on GPU," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, Kraków, Poland, September 8-11, 2013.*, pp. 433–440, 2013.

[CLZ+18] C. Chen, Y. Li, N. Zhao, B. Guo, and N. Mou, "Least squares compactly supported radial basis function for digital terrain model interpolation from airborne lidar point clouds," *Remote Sensing*, vol. 10, no. 4, 2018.

[Cre15] N. Cressie, *Statistics for spatial data*. John Wiley & Sons, 2015.

[Dar00] E. Darve, "The fast multipole method: Numerical implementation," *Journal of Computational Physics*, vol. 160, no. 1, pp. 195–240, 2000.

[DTS02] H. Q. Dinh, G. Turk, and G. G. Slabaugh, "Reconstructing surfaces by volumetric regularization using radial basis functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1358–1371, 2002.

[Duc77] J. Duchon, "Splines minimizing rotation-invariant semi-norms in sobolev spaces," in *Constructive theory of functions of several variables*, pp. 85–100, Springer, 1977.

[ECS19] M. Esmaeilbeigi, O. Chatrabgoun, and M. Shafa, "Numerical solution of time-dependent stochastic partial differential equations using rbf partition of unity collocation method based on finite difference," *Engineering Analysis with Boundary Elements*, vol. 104, pp. 120 – 134, 2019.

[Fas07] G. E. Fasshauer, *Meshfree Approximation Methods with MATLAB*, vol. 6. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 2007.

[Fra82] R. Franke, "Scattered data interpolation: Tests of some methods," *Mathematics of computation*, vol. 38, no. 157, pp. 181–200, 1982.

[FZ07] G. E. Fasshauer and J. G. Zhang, “On choosing "optimal" shape parameters for RBF approximation,” *Numerical Algorithms*, vol. 45, no. 1-4, pp. 345–368, 2007.

[GIS12] M. Gherlone, L. Iurlaro, and M. D. Sciuva, “A novel algorithm for shape parameter selection in radial basis functions collocation method,” *Composite Structures*, vol. 94, no. 2, pp. 453 – 461, 2012.

[GTBM13] J. P. Gois, D. F. Trevisan, H. C. Batagelo, and I. Macêdo, “Generalized hermitian radial basis functions implicits from polygonal mesh constraints,” *The Visual Computer*, vol. 29, no. 6, pp. 651–661, 2013.

[Har71] R. L. Hardy, “Multiquadratic Equations of Topography and Other Irregular Surfaces,” *Journal of Geophysical Research*, vol. 76, pp. 1905–1915, 1971.

[Har90] R. L. Hardy, “Theory and applications of the multiquadric-biharmonic method 20 years of discovery 1968–1988,” *Computers & Mathematics with Applications*, vol. 19, no. 8, pp. 163–208, 1990.

[HLC07] C.-S. Huang, C.-F. Lee, and A.-D. Cheng, “Error estimate, optimal shape factor, and high precision computation of multiquadric collocation method,” *Engineering Analysis with Boundary Elements*, vol. 31, no. 7, pp. 614 – 623, 2007.

[HSfY15] Y.-C. Hon, B. Sarler, and D. fang Yun, “Local radial basis function collocation method for solving thermo-driven fluid-flow problems with free surface,” *Engineering Analysis with Boundary Elements*, vol. 57, pp. 2 – 8, 2015. {RBF} Collocation Methods.

[IdSPT14] D. Izquierdo, M. C. L. de Silanes, M. C. Parra, and J. J. Torrens, “CS-RBF interpolation of surfaces with vertical faults from scattered data,” *Mathematics and Computers in Simulation*, vol. 102, pp. 11–23, 2014.

[Isk04] A. Iske, *Multiresolution Methods in Scattered Data Modelling*, vol. 37 of *Lecture Notes in Computational Science and Engineering*. Springer, 2004.

[JCW+15] G. R. Joldes, H. A. Chowdhury, A. Wittek, B. Doyle, and K. Miller, “Modified moving least squares with polynomial bases for scattered data approximation,” *Applied Mathematics and Computation*, vol. 266, pp. 893–902, 2015.

[KC92] E. Kansa and R. Carlson, “Improved accuracy of multiquadric interpolation using variable shape parameters,” *Computers & Mathematics with Applications*, vol. 24, no. 12, pp. 99 – 120, 1992.

[KHS03] N. Kojekine, I. Hagiwara, and V. V. Savchenko, "Software tools using CSRBFs for processing scattered data," *Computers & Graphics*, vol. 27, no. 2, pp. 311–319, 2003.

[Law13] O. S. Lawlor, "In-memory data compression for sparse matrices," in *Proceedings of the 3rd Workshop on Irregular Applications: Architectures and Algorithms*, p. 6, ACM, 2013.

[LCC13] M. Li, W. Chen, and C. Chen, "The localized RBFs collocation methods for solving high dimensional PDEs," *Engineering Analysis with Boundary Elements*, vol. 37, no. 10, pp. 1300 – 1304, 2013.

[LD08] A. Lagae and P. Dutré, "Compact, fast and robust grids for ray tracing," in *Computer Graphics Forum*, vol. 27, pp. 1235–1244, Wiley Online Library, 2008.

[LK11] S. Laine and T. Karras, "Efficient sparse voxel octrees," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 8, pp. 1048–1059, 2011.

[LZ06] E. Langetepe and G. Zachmann, *Geometric data structures for computer graphics*. A K Peters, 2006.

[Mal89] J.-L. Mallet, "Discrete smooth interpolation," *ACM Trans. Graph.*, vol. 8, pp. 121–144, Apr. 1989.

[Mar19] M. Martynova, "A novel approach of the approximation by patterns using hybrid rbf nn with flexible parameters," in *Computational and Statistical Methods in Intelligent Systems*, pp. 225–235, Springer International Publishing, 2019.

[Mas03] T. Masuda, "Surface curvature estimation from the signed distance field," in *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pp. 361–368, IEEE, 2003.

[MGV09] I. Macêdo, J. P. Gois, and L. Velho, "Hermite interpolation of implicit surfaces with radial basis functions," in *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 1–8, 2009.

[MGV11] I. Macêdo, J. P. Gois, and L. Velho, "Hermite radial basis functions implicits," *Computer Graphics Forum*, vol. 30, no. 1, pp. 27–42, 2011.

[MRW+14] Y. Ma, J.-J. Royer, H. Wang, Y. Wang, and T. Zhang, "Factorial kriging for multiscale modelling," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, pp. 651–659, 08 2014.

[OBA+03] Y. Ohtake, A. G. Belyaev, M. Alexa, G. Turk, and H. Seidel, "Multi-level partition of unity implicits," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 463–470, 2003.

[OBS05] Y. Ohtake, A. G. Belyaev, and H. Seidel, "3d scattered data interpolation and approximation with multilevel compactly supported rbfs," *Graphical Models*, vol. 67, no. 3, pp. 150–165, 2005.

[OBS06] Y. Ohtake, A. G. Belyaev, and H. Seidel, "Sparse surface reconstruction with adaptive partition of unity and radial basis functions," *Graphical Models*, vol. 68, no. 1, pp. 15–24, 2006.

[PMW09] R. Pan, X. Meng, and T. Whangbo, "Hermite variational implicit surface reconstruction," *Science in China Series F: Information Sciences*, vol. 52, no. 2, pp. 308–315, 2009.

[PRF14] D. W. Pepper, C. Rasmussen, and D. Fyda, "A meshless method using global radial basis functions for creating 3-d wind fields from sparse meteorological data," *Computer Assisted Methods in Engineering and Science*, vol. 21, no. 3-4, pp. 233–243, 2014.

[PS11a] R. Pan and V. Skala, "Continuous global optimization in surface reconstruction from an oriented point cloud," *Computer-Aided Design*, vol. 43, no. 8, pp. 896–901, 2011.

[PS11b] R. Pan and V. Skala, "A two-level approach to implicit surface modeling with compactly supported radial basis functions," *Eng. Comput. (Lond.)*, vol. 27, no. 3, pp. 299–307, 2011.

[Ran15] M. Ranjbar, "A new variable shape parameter strategy for gaussian radial basis function approximation methods," *Annals of the University of Craiova-Mathematics and Computer Science Series*, vol. 42, no. 2, pp. 260–272, 2015.

[Rip99] S. Rippa, "An algorithm for selecting a good value for the parameter c in radial basis function interpolation," *Adv. Comput. Math.*, vol. 11, no. 2-3, pp. 193–210, 1999.

[RV84] J.-J. Royer and P. C. Vieira, *Dual Formalism of Kriging*, vol. 2. D. Reidel Publishing Company, g. verly et al. ed., 1984.

[Sch79] I. Schagen, "Interpolation in two dimensions - a new technique," *IMA Journal of Applied Mathematics*, vol. 23, no. 1, pp. 53–59, 1979.

[Sch11] M. Scheuerer, "An alternative procedure for selecting a good value for the parameter c in RBF-interpolation," *Adv. Comput. Math.*, vol. 34, no. 1, pp. 105–126, 2011.

[SHK09] V. Skala, J. Hrádek, and M. Kuchař, "Hash function for triangular mesh reconstruction," pp. 233–238, 2009.

[Šim09] I. Šimecek, "Sparse matrix computations using the quadtree storage format," in *Proceedings of 11th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2009)*, pp. 168–173, 2009.

[Ska13] V. Skala, "Fast Interpolation and Approximation of Scattered Multidimensional and Dynamic Data Using Radial Basis Functions," *WSEAS Transactions on Mathematics*, vol. 12, no. 5, pp. 501–511, 2013.

[SMP+15] N. Sedaghati, T. Mu, L.-N. Pouchet, S. Parthasarathy, and P. Sadayappan, "Automatic selection of sparse matrix representation on gpus," in *Proceedings of the 29th ACM on International Conference on Supercomputing*, pp. 99–108, ACM, 2015.

[SPN13] V. Skala, R. Pan, and O. Nedved, "Simple 3d surface reconstruction using flatbed scanner and 3d print," in *SIGGRAPH Asia 2013, Hong Kong, China, November 19-22, 2013, Poster Proceedings*, p. 7, ACM, 2013.

[SPN14] V. Skala, R. Pan, and O. Nedved, "Making 3d replicas using a flatbed scanner and a 3d printer," in *Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Guimarães, Portugal, June 30 - July 3, 2014, Proceedings, Part VI*, vol. 8584 of *Lecture Notes in Computer Science*, pp. 76–86, Springer, 2014.

[SS09] S. A. Sarra and D. Sturgill, "A random variable shape parameter strategy for radial basis function approximation methods," *Engineering Analysis with Boundary Elements*, vol. 33, no. 11, pp. 1239 – 1245, 2009.

[SS13] Y. Sanyasiraju and C. Satyanarayana, "On optimization of the RBF shape parameter in a grid-free local scheme for convection dominated problems over non-uniform centers," *Applied Mathematical Modelling*, vol. 37, no. 12, pp. 7245 – 7272, 2013.

[SS17] M. Smolik and V. Skala, "Spherical RBF vector field interpolation: Experimental study," in *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 000431–000434, 2017.

[SS18] M. Smolik and V. Skala, "Large scattered data interpolation with radial basis functions and space subdivision," *Integrated Computer-Aided Engineering*, vol. 25, no. 1, pp. 49–62, 2018.

[SSM18] M. Smolik, V. Skala, and Z. Majdisova, "Vector field radial basis function approximation," *Advances in Engineering Software*, vol. 123, pp. 117 – 129, 2018. [IF=3.198, WoS, Scopus].

[SSM19] M. Smolik, V. Skala, and Z. Majdisova, "3D Vector Field Approximation and Critical Points Reduction using Radial Basis Functions," *International journal of mechanics*, vol. 13, no. 1, pp. 100–103, 2019. [Scopus].

[TGB14] D. F. Trevisan, J. P. Gois, and H. C. Batagelo, "A low-cost-memory CUDA implementation of the conjugate gradient method applied to globally supported radial basis functions implicits," *J. Comput. Science*, vol. 5, no. 5, pp. 701–708, 2014.

[TO02] G. Turk and J. F. O'Brien, "Modelling with implicit surfaces that interpolate," *ACM Trans. Graph.*, vol. 21, no. 4, pp. 855–873, 2002.

[TRS04] I. Tobor, P. Reuter, and C. Schlick, "Efficient reconstruction of large scattered geometric datasets using the partition of unity and radial basis functions," in *The 12-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2004, WSCG 2004, University of West Bohemia, Campus Bory, Plzen-Bory, Czech Republic, February 2-6, 2004*, pp. 467–474, 2004.

[US05] K. Uhlir and V. Skala, "Reconstruction of damaged images using radial basis functions," *Proceedings of EUSIPCO 2005*, p. 160, 2005.

[Wen95] H. Wendland, "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree," *Adv. Comput. Math.*, vol. 4, no. 1, pp. 389–396, 1995.

[Wen06] H. Wendland, "Computational aspects of radial basis function approximation," *Studies in Computational Mathematics*, vol. 12, pp. 231–256, 2006.

[Wen18] H. Wendland, *Solving Partial Differential Equations with Multiscale Radial Basis Functions*, pp. 1191–1213. Cham: Springer International Publishing, 2018.

[XYX+19] Y. Xie, J. Yu, S. Xie, T. Huang, and W. Gui, "On-line prediction of ferrous ion concentration in goethite process based on self-adjusting structure rbf neural network," *Neural Networks*, vol. 116, pp. 1 – 10, 2019.

[ZVS09] J. Zapletal, P. Vaněček, and V. Skala, "RBF-based image restoration utilising auxiliary points," in *Proceedings of the 2009 Computer Graphics International Conference*, pp. 39–43, ACM, 2009.

[ZW15] S. Zhu and A. J. Wathen, "Convexity and solvability for compactly supported radial basis functions with different shapes," *Journal of Scientific Computing*, vol. 63, pp. 862–884, Jun 2015.

[ZWB19] D.-Y. Zhong, L.-G. Wang, and L. Bi, "Implicit surface reconstruction based on generalized radial basis functions interpolant with distinct constraints," *Applied Mathematical Modelling*, vol. 71, pp. 408 – 420, 2019.

Publications of Author

Journal publications

- Z. Majdisova and V. Skala, "Big geo data surface approximation using radial basis functions: A comparative study," *Computers & Geosciences*, vol. 109, pp. 51 – 58, 2017. [IF=2.567, WoS, Scopus]

  Cited in:

  - Z. Hoseinzade, A. R. Mokhtari, and H. Zekri, "Application of radial basis function in the analysis of irregular geochemical patterns through spectrum-area method," *Journal of Geochemical Exploration*, vol. 194, pp. 257 – 265, 2018
  - C. Chen, Y. Li, N. Zhao, B. Guo, and N. Mou, "Least squares compactly supported radial basis function for digital terrain model interpolation from airborne lidar point clouds," *Remote Sensing*, vol. 10, no. 4, 2018
  - M. N. Oqielat and O. Ogilat, "Application of gaussion radial basis function with cubic polynomial for modelling leaf surface," *Journal of Mathematical Analysis*, vol. 9, no. 2, pp. 78 – 87, 2018
  - C. Chen, Y. Li, and C. Yan, "A random features-based method for interpolating digital terrain models with high efficiency," *Mathematical Geosciences*, 2019

- Z. Majdisova and V. Skala, "Radial basis function approximations: comparison and applications," *Applied Mathematical Modelling*, vol. 51, pp. 728 – 743, 2017. [IF=2.617, WoS, Scopus]

  Cited in:

  - Y. Xie, J. Yu, S. Xie, T. Huang, and W. Gui, "On-line prediction of ferrous ion concentration in goethite process based on self-adjusting structure rbf neural network," *Neural Networks*, vol. 116, pp. 1 – 10, 2019
  - D.-Y. Zhong, L.-G. Wang, and L. Bi, "Implicit surface reconstruction based on generalized radial basis functions interpolant with distinct constraints," *Applied Mathematical Modelling*, vol. 71, pp. 408 – 420, 2019
  - J. Li, Q. Qin, and Z. Fu, "A dual-level method of fundamental solutions for three-dimensional exterior high frequency acoustic problems," *Applied Mathematical Modelling*, vol. 63, pp. 558 – 576, 2018

  - X. An, B. Song, Z. Mao, and C. Ma, "Layout Optimization Design of Two Vortex Induced Piezoelectric Energy Converters (VIPECs) Using the Combined Kriging Surrogate Model and Particle Swarm Optimization Method," *Energies*, vol. 11, no. 8, 2018
  - M. N. Oqielat, "Scattered data approximation using radial basis function with a cubic polynomial reproduction for modelling leaf surface," *Journal of Taibah University for Science*, vol. 12, no. 3, pp. 331–337, 2018
  - M. N. Oqielat and O. Ogilat, "Application of gaussion radial basis function with cubic polynomial for modelling leaf surface," *Journal of Mathematical Analysis*, vol. 9, no. 2, pp. 78 – 87, 2018
  - L. Aik, T. Hong, and A. Junoh, "A new formula to determine the optimal dataset size for training neural networks," *ARPN Journal of Engineering and Applied Sciences*, vol. 14, no. 1, pp. 52–61, 2019
  - L. Aik, T. Hong, and A. Junoh, "An improved radial basis function networks in networks weights adjustment for training real-world nonlinear datasets," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 63–76, 2019
  - L. Aik, T. Hong, and A. Junoh, "Distance weighted k-means algorithm for center selection in training radial basis function networks," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 54–62, 2019
  - M. M. Nia, S. Shojaee, and S. Hamzehei-Javaran, "A mixed formulation of b-spline and a new class of spherical hankel shape functions for modeling elastostatic problems," *Applied Mathematical Modelling*, vol. 77, pp. 602 – 616, 2020
  - T. Peterka, N. Youssef S. G., I. Grindeanu, V. S. Mahadevan, R. Yeh, and X. Tricoche, "Foundations of multivariate functional approximation for scientific data," in *2018 IEEE 8th Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 61–71, 2018

- M. Smolik, V. Skala, and Z. Majdisova, "Vector field radial basis function approximation," *Advances in Engineering Software*, vol. 123, pp. 117 – 129, 2018. [IF=3.198, WoS, Scopus]

  Cited in:

  - L. Aik, T. Hong, and A. Junoh, "A new formula to determine the optimal dataset size for training neural networks," *ARPN Journal of Engineering and Applied Sciences*, vol. 14, no. 1, pp. 52–61, 2019
  - L. Aik, T. Hong, and A. Junoh, "An improved radial basis function networks in networks weights adjustment for training real-world nonlinear datasets," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 63–76, 2019
  - L. Aik, T. Hong, and A. Junoh, "Distance weighted k-means algorithm for center selection in training radial basis function networks," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 54–62, 2019

  - G. Battineni, N. Chintalapudi, and F. Amenta, "Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (svm)," *Informatics in Medicine Unlocked*, vol. 16, p. 100200, 2019

- Z. Majdisova, V. Skala, and M. Smolik, "Algorithm for Placement of Reference Points and Choice of an Appropriate Variable Shape Parameter for the RBF Approximation," *Integrated Computer-Aided Engineering*. [Accepted 03-09-2019, IF=4.904, WoS, Scopus]

- M. Smolik, V. Skala, and Z. Majdisova, "3D Vector Field Approximation and Critical Points Reduction using Radial Basis Functions," *International journal of mechanics*, vol. 13, no. 1, pp. 100–103, 2019. [Scopus]

- V. Skala, Z. Majdisova, and M. Smolik, "Space Subdivision to Speed-up Convex Hull Construction in E3," *Advances in Engineering Software*, vol. 91, pp. 12–22, 2016. [IF=3.198, WoS, Scopus]

  Cited in:

  - F. Lalem, A. Bounceur, M. Bezoui, M. Saoudi, R. Euler, T. Kechadi, and M. Sevaux, "LPCN: Least polar-angle connected node algorithm to find a polygon hull in a connected euclidean graph," *Journal of Network and Computer Applications*, vol. 93, pp. 38 – 50, 2017
  - G. Mei and S. Guo, "CudaPre2D: A Straightforward Preprocessing Approach for Accelerating 2D Convex Hull Computations on the GPU," in *2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP 2018)* (Merelli, I and Lio, P and Kotenko, I, ed.), Euromicro Conference on Parallel Distributed and Network-Based Processing, pp. 726–732, IEEE, 2018

Conference contributions

- Z. Majdisova, V. Skala, and M. Smolik, "Determination of Stationary Points and Their Bindings in Dataset using RBF Methods," in *Computational and Statistical Methods in Intelligent Systems, 2nd Computational Methods in System and Software 2018 (CoMeSySo 2018)*, vol. 859 of *Advances in Intelligent Systems and Computing series*, pp. 213–224, Springer, 2019. [WoS, Scopus]

- Z. Majdisova, V. Skala, and M. Smolik, "Incremental Meshfree Approximation of Real Geographic Data," in *Applied Physics, System Science and Computers III* (K. Ntalianis, G. Vachtsevanos, P. Borne, and A. Croitoru, eds.), vol. 574 of *Lecture Notes in Electrical Engineering*, (Cham), pp. 222–228, Springer International Publishing, 2019. [WoS, Scopus]

- 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; Game and Entertainment Technologies 2016; and Computer Graphics, Visualization, Computer Vision and Image Processing 2016, pp. 215–222, IADIS Press, July 2016. [Scopus]

Cited in:

- K. Liew, K. Tee, A. Ramli, and W. Ong, "Integrating clustering method in compactly supported radial basis function for surface approximation," *IAENG International Journal of Computer Science*, vol. 46, no. 1, 2019

- Z. Majdisova and V. Skala, "A Radial Basis Function Approximation for Large Datasets," in *Proceedings of SIGRAD 2016*, pp. 9–14, Linköping University Electronic Press, May 2016

  Cited in:

  - K. Liew, K. Tee, A. Ramli, and W. Ong, "Integrating clustering method in compactly supported radial basis function for surface approximation," *IAENG International Journal of Computer Science*, vol. 46, no. 1, 2019
  - J. Zhang, J. Hou, T. Wu, L. Zhong, S. Gong, and Y. Tang, "Rapid surface reconstruction algorithm for 3d scattered point cloud model," *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics*, vol. 30, no. 2, pp. 235–243, 2018

- V. Skala and Z. Majdisova, "Fast Algorithm for Finding Maximum Distance with Space Subdivision in E2," in *Image and Graphics* (Y.-J. Zhang, ed.), vol. 9218 of *Lecture Notes in Computer Science*, pp. 261–274, Springer International Publishing, 2015. [Scopus]

- M. Smolik, V. Skala, and Z. Majdisova, "A New Simple, Fast and Robust Total Least Square Error Computation in E2: Experimental Comparison," in *AETA 2018 - Recent Advances in Electrical Engineering and Related Sciences: Theory and Application*, vol. 554 of *Lecture Notes in Electrical Engineering*, pp. 325–334, Springer, 2020. [WoS, Scopus]

- V. Skala, M. Smolik, and Z. Majdisova, "Reducing the number of points on the convex hull calculation using the polar space subdivision in E2," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 40–47, Oct 2016. [WoS, Scopus]

  Cited in:

  - A. Trabelsi, B. Shi, X. Wei, H. Frigui, X. Zhang, C. McClain, and A. Shahrajooihaghighi, "Molecule Specific Normalization for Protein and Metabolite Biomarker Discovery," in *SAC '19: PROCEEDINGS OF THE 34TH ACM/SIGAPP SYMPOSIUM ON APPLIED COMPUTING*, pp. 25–31, Assoc Comp Machinery Special Interest Grp Appl Comp, 2019

Appendix A

Project Assignments, Other Activities

A.1 Conferences and Talks

- Geographic Point Clouds: RBF Approximation & Compression, Future Forces Forum (Multi-domain Advanced Robotic Systems Conference) 2018, Prague, Czech Republic, October 17-19, 2018
- Incremental Meshfree Approximation of Real Geographic Data, 3rd International Conference on: Applied Physics, System Science and Computers (APSAC) 2018, Croatia, September 26, 2018
- Determination of Stationary Points and Their Bindings in Dataset using RBF Methods, 2nd Computational Methods in Systems and Software (CoMeSySo) 2018, Czech Republic, September 12, 2018
- Determination of Suitable Variable Shape Parameter and Placement of Reference Points for RBF Approximation (in Czech), University of West Bohemia, Plzeň, Czech Republic, June 6, 2017
- Derivation of Radial Basis Function Approximation with Polynomial Reproduction, 25th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG) 2017, Pilsen, Czech Republic, May 31, 2017
- A New Radial Basis Function Approximation with Reproduction, 10th International Conference on Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP), Funchal, Madeira, Portugal, July 4, 2016
- A Radial Basis Function Approximation for Large Datasets, SIGRAD 2016 (the Swedish Chapter of Eurographics) conference, Visby, Sweden, May 24, 2016
- Fast Algorithm for Finding Maximum Distance with Space Subdivision in E^2, 8th International Conference on Image and Graphics, Tianjin, China, August 15, 2015

- Data Structures (in Czech), University of West Bohemia, Plzeň, Czech Republic, May, 19, 2015
- Convex Hull and Space Subdivision in E^3 (in Czech), University of West Bohemia, Plzeň, Czech Republic, January, 30, 2015

A.2 Participation on Scientific Projects

- Meshless Methods for Large Scattered Spatio-temporal Vector Data Visualization, Funded by Czech Science Foundation GAČR, project code GA17-05534S
- Synthesis and Analysis of Geometric and Computing Models. Project code SGS-2019-016.
- Advanced Graphical and Computing Systems. Project code SGS-2016-013.
- Development of Algorithms for Computer Graphics and CAD/CAM Systems. Funded by MŠMT ČR, project code LH12181.
- Advanced Computing and Information Systems. Project code SGS-2013-029.

A.3 Teaching Activities

2014/2015:

- Programming Strategies (KIV/PRO) - tutor
- Introduction to Computer Graphics (KIV/UPG) - tutor

2015/2016:

- Programming Strategies (KIV/PRO) - tutor
- Introduction to Computer Graphics (KIV/UPG) - tutor

2016/2017:

- Programming Techniques (KIV/PT) - tutor

2018/2019:

- Introduction to Computer Graphics (KIV/UPG) - tutor

2019/2020:

- Programming Techniques (KIV/PT) - tutor
- Computers and Programming 1 (KIV/PPA1) - tutor

Appendix B

A New Radial Basis Function Approximation with Reproduction

Majdišová, Z., Skala, V.

Proceedings of the International Conferences on Interfaces and Human Computer Interaction 2016; Game and Entertainment Technologies 2016; and Computer Graphics, Visualization, Computer Vision and Image Processing 2016 (CGVCVIP 2016), pp. 215-222, IADIS Press (2016), ISBN 978-989-8533-52-4

International Conferences Interfaces and Human Computer Interaction 2016,
Game and Entertainment Technologies 2016 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2016

A NEW RADIAL BASIS FUNCTION APPROXIMATION WITH REPRODUCTION

Zuzana Majdisova and Vaclav Skala
Faculty of Applied Sciences, University of West Bohemia
Univerzitni 8, CZ 30614 Plzen, Czech Republic

ABSTRACT

Approximation of scattered geometric data is often a task in many engineering problems. The Radial Basis Function (RBF) approximation is appropriate for large scattered (unordered) datasets in d-dimensional space. This method is useful for a higher dimension $d \geq 2$, because the other methods require a conversion of a scattered dataset to a semi-regular mesh using some tessellation techniques, which is computationally expensive. The RBF approximation is non-separable, as it is based on a distance of two points. It leads to a solution of overdetermined Linear System of Equations (LSE).
In this paper a new RBF approximation method is derived and presented. The presented approach is applicable for d-dimensional cases in general.

KEYWORDS

Radial basis function; RBF; approximation; optimization problem; linear reproduction

1. INTRODUCTION

Radial Basis Functions (RBFs) are widely used across many fields solving technical and non-technical problems. The RBF method was originally introduced by [Hardy, R.L., 1971] and it is an effective tool for solving partial differential equations in engineering and sciences. Moreover, RBF applications can be found in neural networks, fuzzy systems, pattern recognition, data visualization, medical applications, surface reconstruction [Carr, J.C. et al, 2001], [Turk, G. and O'Brien, J.F., 2002], [Pan, R. and Skala, V., 2011a], [Pan, R. and Skala, V., 2011b], [Skala, V. et al, 2013], [Skala, V. et al, 2014], reconstruction of corrupted images [Uhlir, K. and Skala, V., 2005], [Zapletal, J. et al, 2009], etc. The RBF approximation technique is really meshless and is based on collocation in a set of scattered nodes. This method is independent with respect to the dimension of the space. The computational cost of RBF approximation increases nonlinearly with the number of points in the given dataset and linearly with the dimensionality of data.

There are two main groups of basis functions: global RBFs (e.g. [Duchon, J., 1977], [Schagen, I.P, 1979]) and Compactly Supported RBFs (CS-RBFs) [Wendland, H., 2006]. Fitting scattered data with CS-RBFs leads to a simpler and faster computation, because the system of linear equations has a sparse matrix. However, approximation using CS-RBFs is sensitive to the density of approximated scattered data and to the choice of a "shape" parameter. Global RBFs lead to a linear system of equations with a dense matrix and their usage is based on sophisticated techniques such as the fast multipole method [Darve, E., 2000]. Global RBFs are useful in repairing incomplete datasets and they are significantly less sensitive to the density of approximated data.

2. ORIGINAL APPROACH

The original approach of RBF approximation with linear reproduction was introduced by [Fasshauer, G.E., 2007] (Chapter 19.4). Let us briefly summarize the properties of this approach in this section.

The goal of this approach is to approximate a given dataset of N points by a function:

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \, \phi\left(\left\|\boldsymbol{x} - \boldsymbol{\xi}_j\right\|\right) + P_1(\boldsymbol{x}), \tag{1}$$

where the approximating function $f(\boldsymbol{x})$ is represented as a sum of M RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by an appropriate coefficient c_j, and $P_1(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x} + a_0$ is a linear polynomial. This linear polynomial should theoretically solve problems with stability and solvability. Now, it is necessary to determine the vector of weights $\boldsymbol{c} = (c_1, \dots, c_M)^T$and coefficients of the linear polynomial. This is achieved by solving an overdetermined linear system of equations (LSE):

$$\boldsymbol{h}_i = f(\boldsymbol{x}_i) = \sum_{j=1}^{M} c_j \, \phi\left(\left\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\right\|\right) + P_1(\boldsymbol{x}_i) = \sum_{j=1}^{M} c_j \, \phi_{ij} + P_1(\boldsymbol{x}_i), \qquad i = 1, \dots, N, \tag{2}$$

where $\boldsymbol{x}_i$ is point from the given dataset and is associated with scalar value h_i. Moreover, additional conditions are applied:

$$\sum_{i=1}^{M} c_i = 0\,, \qquad \sum_{i=1}^{M} c_i \boldsymbol{\xi}_i = \boldsymbol{0}. \tag{3}$$

It can be seen that for d-dimensional space a linear system of $(N + d + 1)$ equations in $(M + d + 1)$ variables has to be solved, where N is the number of points in the given dataset, M is the number of reference points and d is the dimensionality of the data.

For $d = 2$, vectors $\boldsymbol{x}_i$, $\boldsymbol{\xi}_j$ and $\boldsymbol{a}$ are given as $\boldsymbol{x}_i = (x_i, y_i)^T$, $\boldsymbol{\xi}_j = \left(\xi_j, \eta_j\right)^T$ and $\boldsymbol{a} = \left(a_x, a_y\right)^T$. Thus, for E^2 and the given dataset we can write this LSE in the following matrix form:

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{P} \\ \boldsymbol{\Xi} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{a} \\ a_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{pmatrix} \tag{4}$$

This system is overdetermined ($M \ll N$) and can be solved by the least squares method as:

$$\begin{pmatrix} \boldsymbol{A}^T\boldsymbol{A} + \boldsymbol{\Xi}^T\boldsymbol{\Xi} & \boldsymbol{A}^T\boldsymbol{P} \\ \boldsymbol{P}^T\boldsymbol{A} & \boldsymbol{P}^T\boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{a} \\ a_0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^T\boldsymbol{h} \\ \boldsymbol{P}^T\boldsymbol{h} \end{pmatrix} \tag{5}$$

where

$$\boldsymbol{A}^T\boldsymbol{A} + \boldsymbol{\Xi}^T\boldsymbol{\Xi} = \begin{pmatrix} \sum_{i=1}^{N} \phi_{i1}\phi_{i1} + \xi_1^2 + \eta_1^2 + 1 & \cdots & \sum_{i=1}^{N} \phi_{i1}\phi_{iM} + \xi_1\xi_M + \eta_1\eta_M + 1 \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} \phi_{iM}\phi_{i1} + \xi_M\xi_1 + \eta_M\eta_1 + 1 & \cdots & \sum_{i=1}^{N} \phi_{iM}\phi_{iM} + \xi_M^2 + \eta_M^2 + 1 \end{pmatrix},$$

$$\boldsymbol{P}^T\boldsymbol{A} = (\boldsymbol{A}^T\boldsymbol{P})^T = \begin{pmatrix} \sum_{i=1}^{N} x_i\phi_{i1} & \cdots & \sum_{i=1}^{N} x_i\phi_{iM} \\ \sum_{i=1}^{N} y_i\phi_{i1} & \cdots & \sum_{i=1}^{N} y_i\phi_{iM} \\ \sum_{i=1}^{N} \phi_{i1} & \cdots & \sum_{i=1}^{N} \phi_{iM} \end{pmatrix}, \qquad \boldsymbol{P}^T\boldsymbol{P} = \begin{pmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i y_i & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} y_i x_i & \sum_{i=1}^{N} y_i^2 & \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} y_i & \sum_{i=1}^{N} 1 \end{pmatrix},$$

$$\boldsymbol{A}^T\boldsymbol{h} = \left(\sum_{i=1}^{N} \phi_{i1} h_i \quad \cdots \quad \sum_{i=1}^{N} \phi_{iM} h_i\right)^T, \qquad \boldsymbol{P}^T\boldsymbol{h} = \left(\sum_{i=1}^{N} x_i h_i \quad \sum_{i=1}^{N} y_i h_i \quad \sum_{i=1}^{N} h_i\right)^T.$$

It should be noted that additional conditions (3) introduce inconsistency to the least squares method. Specifically, the inconsistency is caused by adding the term $\boldsymbol{\Xi}^T\boldsymbol{\Xi}$ to $\boldsymbol{A}^T\boldsymbol{A}$. Therefore, the described RBF approximation with linear reproduction is inconveniently formulated, as it mixes variables which have a different physical meaning. Thus, another approach is proposed in the following section.

3. PROPOSED APPROACH

Let us consider that we have an unordered dataset $\{\boldsymbol{x}_i\}_1^N$ in E^2. However, note that this approach is generally applicable for d-dimensional space. Further, each point $\boldsymbol{x}_i$ from the dataset is associated with vector $\boldsymbol{h}_i \in E^p$ of given values, where p is the dimension of the vector, or a scalar value $h_i \in E^1$. For an explanation of the RBF approximation, let us consider the case in which each point $\boldsymbol{x}_i$ is associated with a scalar value h_i, e.g. a 2 ½D surface. Let us introduce a set of new reference points $\left\{\boldsymbol{\xi}_j\right\}_1^M$, see Figure 1.

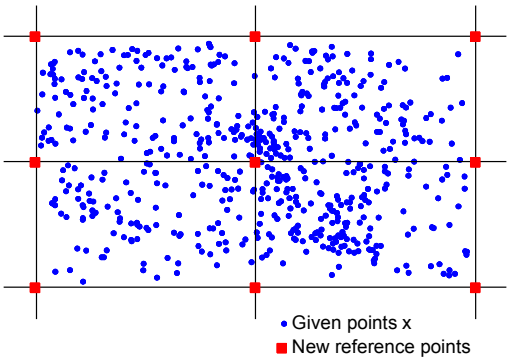


Figure 1. RBF approximation and reduction of points

It should be noted that these reference points may not necessarily be in a uniform grid. It is appropriate that their placements reflect the given surface behavior (e.g. the terrain profile, etc.) as well as possible. The number of added reference points $\boldsymbol{\xi}_j$ is M, where $M \ll N$. The RBF approximation is based on computing the distance of the given point $\boldsymbol{x}_i$ of the given dataset and the reference point $\boldsymbol{\xi}_j$ of the new reference points.

The approximated value can be expressed as:

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \, \phi\left(\left\|\boldsymbol{x} - \boldsymbol{\xi}_j\right\|\right) \; + \; P_1(\boldsymbol{x}), \tag{6}$$

where the approximating function $f(\boldsymbol{x})$ is represented as a sum of M RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by an appropriate coefficient c_j, and $P_1(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x} + a_0$ is a linear polynomial. This linear polynomial should theoretically solve problems with stability and solvability.

It can be seen that for E^2 and the given dataset we get the following overdetermined LSE:

$$\boldsymbol{A}\boldsymbol{c} + \boldsymbol{P}\boldsymbol{k} \; = \; \boldsymbol{h}, \tag{7}$$

where $A_{ij} \; = \phi(\left\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\right\|)$ is the entry of the matrix in the i-th row and j-th column, $\boldsymbol{c} = (c_1, \dots, c_M)^T$ is the vector of weights, $\boldsymbol{P}_i \; = \; (\boldsymbol{x}_i^T, 1)$ is the vector, $\boldsymbol{k} \; = \; (\boldsymbol{a}^T, a_0)^T$ is the vector of coefficients for the linear polynomial and $\boldsymbol{h} = (h_1, \dots, h_N)^T$ is the vector of values in the given points.

The error is then defined as:

$$R = \|\boldsymbol{A}\boldsymbol{c} + \boldsymbol{P}\boldsymbol{k} - \; \boldsymbol{h}\|, \tag{8}$$

then

$$R^2 = (\boldsymbol{A}\boldsymbol{c} + \boldsymbol{P}\boldsymbol{k} - \; \boldsymbol{h})^T(\boldsymbol{A}\boldsymbol{c} + \boldsymbol{P}\boldsymbol{k} - \; \boldsymbol{h}). \tag{9}$$

Our goal is to minimize the square of error, i.e. to find the minimum of R^2 (9). This minimum is obtained by differentiating equation (9) with respect to $\boldsymbol{c}$ and $\boldsymbol{k}$ and finding the zeros of those derivatives. This leads to equations:

$$\begin{aligned} \frac{\partial R^2}{\partial \boldsymbol{c}} \; &= 2(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{k} - \; \boldsymbol{A}^T\boldsymbol{h}) = \boldsymbol{0}, \\ \frac{\partial R^2}{\partial \boldsymbol{k}} \; &= 2(\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{c} + \boldsymbol{P}^T\boldsymbol{P}\boldsymbol{k} - \; \boldsymbol{P}^T\boldsymbol{h}) = \boldsymbol{0}, \end{aligned} \tag{10}$$

which leads to a system of linear equations:

$$\begin{pmatrix} \boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{A}^T\boldsymbol{P} \\ \boldsymbol{P}^T\boldsymbol{A} & \boldsymbol{P}^T\boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{k} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^T\boldsymbol{h} \\ \boldsymbol{P}^T\boldsymbol{h} \end{pmatrix}, \tag{11}$$

i.e.

$$\boldsymbol{B}\boldsymbol{\lambda} = \boldsymbol{f}. \tag{12}$$

The matrix $\boldsymbol{B}$ is a $(M + 3) \times (M + 3)$ symmetric positively semidefinite matrix. Equation (11) can be expressed in the form:

$$\begin{pmatrix} \sum_{i=1}^{N}\phi_{i1}\phi_{i1} & \cdots & \sum_{i=1}^{N}\phi_{i1}\phi_{iM} & \sum_{i=1}^{N}\phi_{i1}x_i & \sum_{i=1}^{N}\phi_{i1}y_i & \sum_{i=1}^{N}\phi_{i1} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{N}\phi_{iM}\phi_{i1} & \cdots & \sum_{i=1}^{N}\phi_{iM}\phi_{iM} & \sum_{i=1}^{N}\phi_{iM}x_i & \sum_{i=1}^{N}\phi_{iM}y_i & \sum_{i=1}^{N}\phi_{iM} \\ \sum_{i=1}^{N}x_i\phi_{i1} & \cdots & \sum_{i=1}^{N}x_i\phi_{iM} & \sum_{i=1}^{N}x_i^2 & \sum_{i=1}^{N}x_iy_i & \sum_{i=1}^{N}x_i \\ \sum_{i=1}^{N}y_i\phi_{i1} & \cdots & \sum_{i=1}^{N}y_i\phi_{iM} & \sum_{i=1}^{N}y_ix_i & \sum_{i=1}^{N}y_i^2 & \sum_{i=1}^{N}y_i \\ \sum_{i=1}^{N}\phi_{i1} & \cdots & \sum_{i=1}^{N}\phi_{iM} & \sum_{i=1}^{N}x_i & \sum_{i=1}^{N}y_i & \sum_{i=1}^{N}1 \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \\ a_x \\ a_y \\ a_0 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N}\phi_{i1}\,h_i \\ \vdots \\ \sum_{i=1}^{N}\phi_{iM}\,h_i \\ \sum_{i=1}^{N}x_i\,h_i \\ \sum_{i=1}^{N}y_i\,h_i \\ \sum_{i=1}^{N}h_i \end{pmatrix}. \quad (13)$$

where $\phi_{ij} = \phi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|)$, point $\boldsymbol{x}_i = (x_i, y_i)^T$ and vector $\boldsymbol{a} = (a_x, a_y)^T$. It can be seen that this approach eliminates the inconsistency introduced in Section 2.

4. EXPERIMENTAL RESULTS

Both presented approaches of the RBF approximation have been compared for a dataset with a Halton distribution of points [Fasshauer, G.E., 2007] (Appendix A.1). Moreover, each point from this dataset is associated with a function value at this point. For this purpose, different functions have been used for experiments. Results for two such functions are presented here. The first is a $2D$ sinc function, see Figure 2 (left), and the second is Franke's function, see Figure 2 (right).

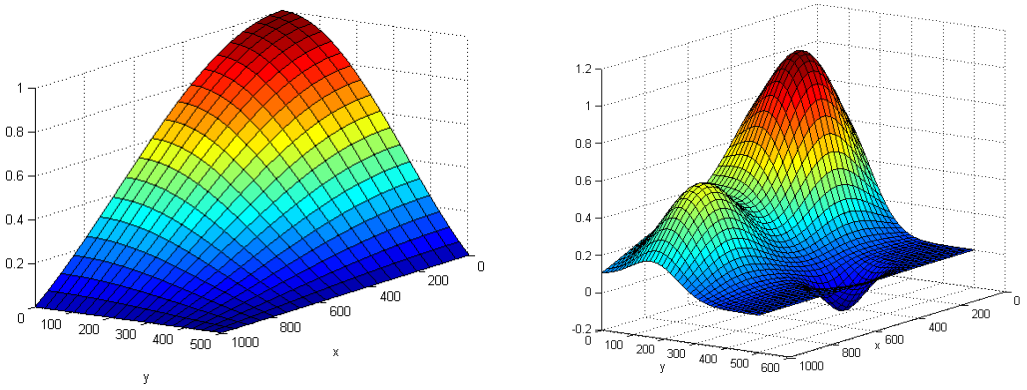


Figure 2. $2D$ sinc function defined as $\text{sinc}\left(\frac{\pi x}{1000}\right)\text{sinc}\left(\frac{\pi y}{500}\right)$, whose domain is restricted to $[0,1000] \times [0,500]$ (left) and Franke's function (right)

In addition, three different global radial basis functions with shape parameter α, see Table 1, have been used for testing. Also different sets of reference points have been used for experiments.

Table 1. Used global RBFs

RBF	$\phi(\boldsymbol{r})$
Gauss function	$e^{-(\alpha r)^2}$
Inverse Quadric (IQ)	$\frac{1}{1+(\alpha r)^2}$
Thin-Plate Spline (TPS)	$(\alpha r)^2 \log(\alpha r)$

These sets of reference points have different types of distributions. The presented types of distribution are the Halton distribution [Fasshauer, G.E., 2007] (Appendix A.1), see Figure 3 (left), an epsilon distribution, which is based on a random drift of points on a regular grid, see Figure 3 (right), and points on a regular grid.

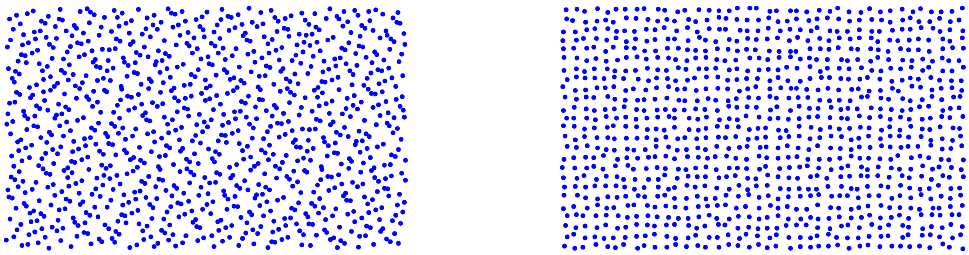
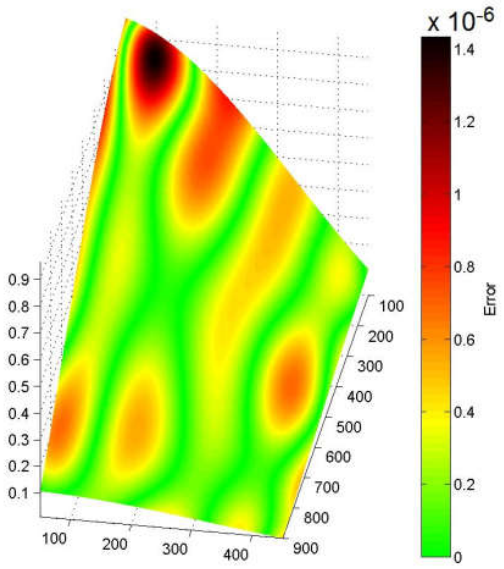


Figure 3. Halton points in E^2 (left) and epsilon points in E^2 (right). Number of points is 10^3 in both cases

4.1 Examples of RBF Approximation Results

An example of RBF approximation of 1089 Halton data points sampled from a $2D$ sinc function, for a Halton set of reference points which consists of 81 points, using both approaches is shown in Figure 4. The graphs are false-colored according to the magnitude of the error.

Original approach of RBF approximation



Proposed approach of RBF approximation

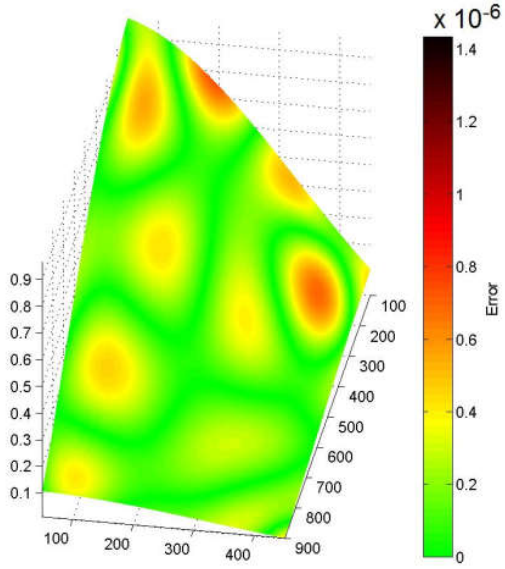


Figure 4. Approximation of 1089 data points sampled from a $2D$ sinc function, i.e. $\text{sinc}\left(\frac{\pi x}{1000}\right)\text{sinc}\left(\frac{\pi y}{500}\right)$, where $(x, y) \in [0,1000] \times [0,500]$, with 81 Halton-spaced Gaussian functions with $\alpha = 0.001$,false-colored by magnitude of error

ISBN: 978-989-8533-52-4 © 2016

A further example of RBF approximation of 4225 Halton data points sampled from a Franke's function and for a set of reference points which consists of 289 points on a regular grid, using both approaches is shown in Figure 5. The graphs are again false-colored by magnitude of error.

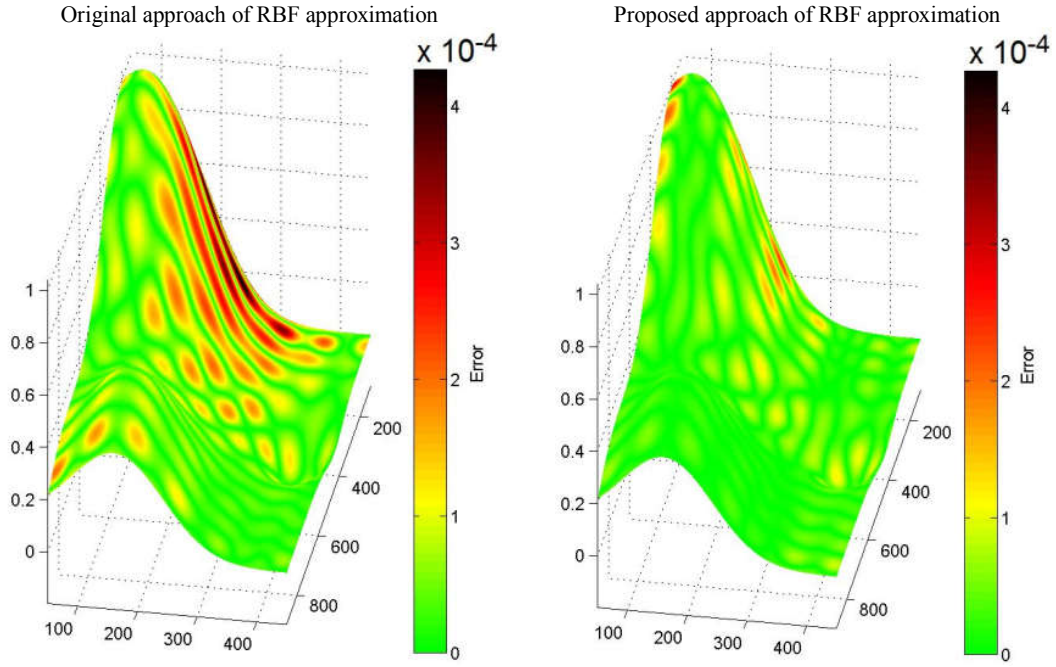


Figure 5. Approximation of 4225 data points sampled from a Franke's function with 289 regularly spaced IQ with $\alpha = 0.005$, false-colored by magnitude of error

It can be seen that the original RBF approximation with a linear reproduction returns a worse result in terms of the error in comparison with the proposed RBF approximation with a linear reproduction. Moreover, we can see from Figure 4 and Figure 5 that for the presented cases the maximum magnitude of error for the original approach is approximately two times greater than the maximum magnitude of error for the proposed approach.

There remains the question of how the RBF approximation depends on the shape parameter α selection. Many papers have been published about choosing optimal shape parameter α, e.g. [Franke, R., 1982], [Rippa, S., 1999], [Fasshauer, G.E. and Zhang, J.G., 2007], [Scheuerer, M., 2011]. In the following section, a comparison depending on the choice of shape parameter α is performed.

4.2 Comparison of Methods

In this section, the original approach and the proposed approach, which were presented in Section 2 and Section 3, are compared. Figure 6 presents the ratio of mean error of the original RBF approximation with the linear reproduction to the mean error of the proposed RBF approximation with the linear reproduction, i.e.:

$$ratio = \frac{mean\ error_{original}}{mean\ error_{proposed}}, \tag{14}$$

for a dataset which consists of 1089 Halton points in the range $[0,1000] \times [0,500]$, sampled from a $2D$ sinc function. The set of reference points contains 81 points with different behavior of the distribution, and for different global RBFs. Graphs in Figure 6 represent the experimentally obtained ratio according to the shape parameter α of the used RBFs.

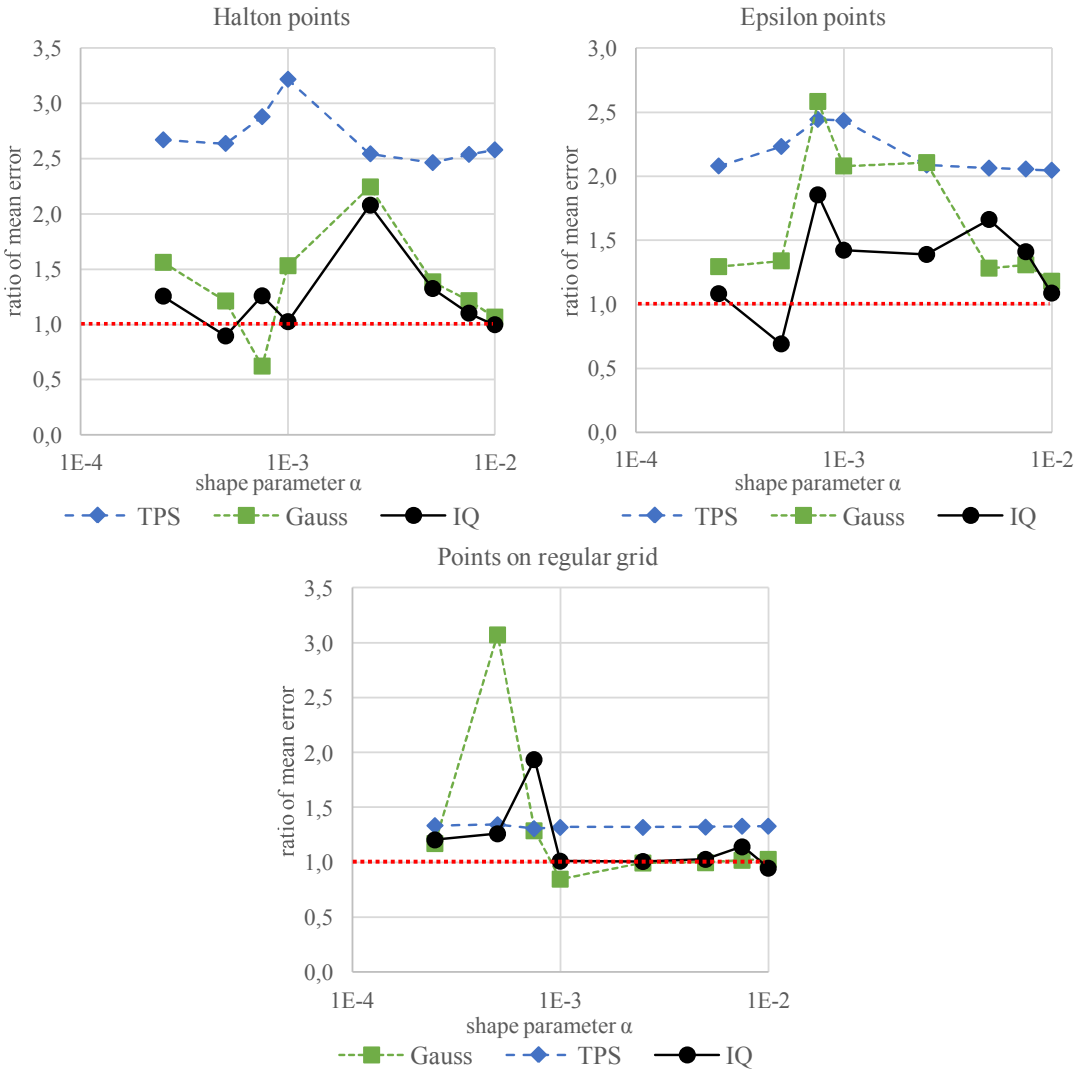


Figure 6. The ratio of mean error of the original approach to the mean error of the proposed approach of RBF approximation of 1089 data points sampled from a $2D$ sinc function with 81 reference points for different RBFs and different shape parameters. The used sets of reference points are: Halton points (top left), Epsilon points (top right) and points on a regular grid (bottom)

We can see that for the TPS, the mean errors of the proposed approach are significantly smaller than those of the original approach (ratio is greater than one). Furthermore, this ratio is not significantly different for the different shape parameters α. For the Gaussian function and epsilon reference points, the proposed RBF approximation gives better results than the original approach in terms of the mean error. In the remaining cases, with five exceptions, the proposed approach is also better.

The experiments prove that the proposed approach to RBF approximation is correct and gives better and more stable results than the original approach [Fasshauer, G.E., 2007].

5. CONCLUSION

This paper presents a new formulation for RBF approximation with a linear reproduction. The proposed approach eliminates inconsistency, which occurs in the original RBF approximation with a linear

reproduction. This inconsistency is caused by adding additional conditions to the polynomial part. The experiments made prove that the proposed approach gives significantly better results than the original method in terms of accuracy. The presented approach is easily extendable for general polynomial reproduction and for higher dimensionality.

In future work, application of the proposed approach is to be performed on large real datasets and the performance can be further measured.

ACKNOWLEDGEMENT

The authors would like to thank their colleagues at the University of West Bohemia, Plzen, for their discussions and suggestions, and also anonymous reviewers for the valuable comments and suggestions they provided. The research was supported by MSMT CR projects LH12181 and SGS 2016-013.

REFERENCES

Carr, J.C. et al, 2001. Reconstruction and representation of 3D objects with radial basis functions. *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001*, Los Angeles, California, USA, pp. 67-76.

Darve, E., 2000. The Fast Multipole Method: Numerical Implementation. *In Journal of Computational Physics*, Vol. 160, No. 1, pp. 195-240.

Duchon, J., 1977. Splines Minimizing Rotation-invariant Semi-norms in Sobolev Spaces. *In Constructive theory of functions of several variables*, pp. 85-100.

Fasshauer, G.E., 2007. *Meshfree Approximation Methods with MATLAB*. World Scientific Publishing Co., River Edge, NJ, USA.

Fasshauer, G.E. and Zhang, J.G., 2007. On choosing optimal shape parameters for RBF approximation. *In Numeric Algorithms*, Vol. 45, pp. 345-368.

Franke, R., 1982. Scattered data interpolation: tests of some methods. *In Mathematical Computing*, Vol. 38, pp. 181-200.

Hardy, R.L., 1971. Multiquadratic Equations of Topography and Other Irregular Surfaces. *In Journal of Geophysical Research*, Vol. 76, No. 8, pp. 1905-1915.

Pan, R. and Skala, V., 2011a. Continuous Global Optimization in Surface Reconstruction from an Oriented Point Cloud. *In Computer-Aided Design*, Vol. 43, No. 8, pp. 896-901.

Pan, R. and Skala, V., 2011b. A Two-level Approach to Implicit Surface Modeling with Compactly Supported Radial Basis Functions. *In Eng. Comput. (Lond.)*, Vol. 27, No. 3, pp. 299-307.

Rippa, S., 1999. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *In Adv. Comput. Math.*, Vol. 11, pp. 193-210.

Schagen, I.P, 1979. Interpolation in Two Dimensions - a New Technique. *In IMA Journal of Applied Mathematics*, Vol. 23, No. 1, pp. 53-59.

Scheuerer, M., 2011. An alternative procedure for selecting a good value for the parameter c in RBF-interpolation. *In Adv. Comput. Math.*, Vol. 34, pp. 105-126.

Skala, V. et al, 2013. Simple 3D Surface Reconstruction Using Flatbed Scanner and 3D Print. *SIGGRAPH Asia 2013, Poster Proceedings.* Hong Kong, China, p. 7.

Skala, V. et al, 2014. Making 3D Replicas Using a Flatbed Scanner and a 3D Printer. *Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Proceedings, Part VI.* Guimarães, Portugal, pp. 76-86.

Turk, G. and O'Brien, J.F., 2002. Modelling with implicit surfaces that interpolate. *In ACM Trans. Graph.*, Vol. 21, No. 4, pp. 855-873.

Uhlir, K. and Skala, V., 2005. Reconstruction of Damaged Images Using Radial Basis Functions. *Proceedings of EUSIPCO.* Antalya, Turkey, p. 160.

Wendland, H., 2006. Computational Aspects of Radial Basis Function Approximation. *In Studies in Computational Mathematics*, Vol. 12, pp. 231-256.

Zapletal, J. et al, 2009. RBF-based Image Restoration Utilizing Auxiliary Points. *Proceedings of the 2009 Computer Graphics International Conference.* Victoria, British Columbia, Canada, pp. 39-43.

Appendix C

Radial basis function approximations: comparison and applications

Majdišová, Z., Skala, V.

Applied Mathematical Modelling, Volume 51, Issue 1, pp. 728-743, Elsevier (2017),
ISSN 0307-904X, IF 2.68

Applied Mathematical Modelling 51 (2017) 728–743

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Radial basis function approximations: comparison and applications

Zuzana Majdisova*, Vaclav Skala

Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitni 8, Plzen CZ 30614, Czechia

ARTICLE INFO

Article history:
Received 5 March 2016
Revised 3 July 2017
Accepted 17 July 2017
Available online 27 July 2017

Keywords:
Radial basis function
RBF
Approximation
Lagrange multipliers

ABSTRACT

Approximation of scattered data is often a task in many engineering problems. The radial basis function (RBF) approximation is appropriate for large scattered (unordered) datasets in d-dimensional space. This approach is useful for a higher dimension $d > 2$, because the other methods require the conversion of a scattered dataset to an ordered dataset (i.e. a semi-regular mesh is obtained by using some tessellation techniques), which is computationally expensive. The RBF approximation is non-separable, as it is based on the distance between two points. This method leads to a solution of linear system of equations (LSE) $\mathbf{Ac} = \mathbf{h}$.

In this paper several RBF approximation methods are briefly introduced and a comparison of those is made with respect to the stability and accuracy of computation. The proposed RBF approximation offers lower memory requirements and better quality of approximation.

1. Introduction

Radial basis functions (RBFs) are widely used across many fields solving technical and non-technical problems. A RBF method was originally introduced by [1] and it is an effective tool for solving partial differential equations in engineering and sciences. Moreover, RBF applications can be found in neural networks, fuzzy systems, pattern recognition, data visualization, medical applications, surface reconstruction [2–5], reconstruction of corrupted images [6,7], etc. The RBF approximation technique is really meshless and is based on collocation in a set of scattered nodes. This method is independent with respect to the dimension of the space. The computational cost of RBF approximation increases nonlinearly with the number of points in the given dataset and linearly with the dimensionality of data.

There are two main groups of basis functions: global RBFs and compactly supported RBFs (CS-RBFs) [8]. Fitting scattered data with CS-RBFs leads to a simpler and faster computation, because a system of linear equations has a sparse matrix. However, approximation using CS-RBFs is quite sensitive to the density of approximated scattered data and to the choice of a shape parameter. Global RBFs lead to a linear system of equations with a dense matrix and their usage is based on sophisticated techniques such as the fast multipole method [9]. Global RBFs are useful in repairing incomplete datasets and they are insensitive to the density of approximated data.

* Corresponding author.
E-mail address: majdisz@kiv.zcu.cz (Z. Majdisova).
URL: http://www.vaclavskala.eu (V. Skala)

http://dx.doi.org/10.1016/j.apm.2017.07.033
0307-904X/

2. RBF approximation using Lagrange multipliers

RBF approximation introduced by Fasshauer [10, Chapter 19] is based on Lagrange multipliers. In this section, the properties of this method will be briefly summarized.

This RBF approximation is formulated as a constrained quadratic optimization problem. The goal of this method is to approximate the given dataset by function:

$$f(\mathbf{x}) = \sum_{j=1}^{M} c_j \phi(\|\mathbf{x} - \boldsymbol{\xi}_j\|), \tag{1}$$

where the approximating function $f(\mathbf{x})$ is represented as a sum of M RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by an appropriate coefficient c_j. Therefore, it is necessary to determine the vector of weights $\mathbf{c} = (c_1, \ldots, c_M)^T$, which leads to the minimization of the quadratic form:

$$\frac{1}{2}\mathbf{c}^T\mathbf{Q}\mathbf{c}, \tag{2}$$

where $\mathbf{Q}$ is some $M \times M$ symmetric positive definite matrix. This quadratic form is minimized subject to the N linear constraints $\mathbf{A}\mathbf{c} = \mathbf{h}$, where $\mathbf{A}$ is an $N \times M$ matrix with full rank, and the right-hand side $\mathbf{h} = (h_1, \ldots, h_N)^T$ is given. Thus the constrained quadratic minimization problem can be described as an LSE:

$$F(\mathbf{c}, \boldsymbol{\lambda}) = \frac{1}{2}\mathbf{c}^T\mathbf{Q}\mathbf{c} - \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{c} - \mathbf{h}), \tag{3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^T$ is the vector of Lagrange multipliers, and we need to find the minimum of (3) with respect to $\mathbf{c}$ and $\boldsymbol{\lambda}$. This leads to solving the following system:

$$\begin{aligned} \frac{\partial F(\mathbf{c}, \boldsymbol{\lambda})}{\partial \mathbf{c}} &= \mathbf{Q}\mathbf{c} - \mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0} \\ \frac{\partial F(\mathbf{c}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= \mathbf{A}\mathbf{c} - \mathbf{h} = \mathbf{0} \end{aligned} \tag{4}$$

or, in matrix form:

$$\begin{pmatrix} \mathbf{Q} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{h} \end{pmatrix}, \tag{5}$$

where $Q_{i,j} = \phi(\|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|)$ and $\mathbf{Q}$ is a symmetric matrix. Eq. (5) is then solved.

It should be noted that we want to minimize M in order to reduce the computational cost of the approximated value $f(\mathbf{x})$ as much as possible.

3. RBF approximation

Another approach is RBF interpolation, which is based on a solution of a linear system of equations (LSE) [11]:

$$\mathbf{A}\mathbf{c} = \mathbf{h}, \tag{6}$$

where $\mathbf{A}$ is a matrix of this system, $\mathbf{c}$ is a column vector of variables and $\mathbf{h}$ is a column vector containing the right sides of equations. In this case, $\mathbf{A}$ is an $N \times N$ matrix, where N is the number of given points, the variables are weights for basis functions and the right sides of equations are values in the given points. The disadvantage of RBF interpolation is the large and usually ill-conditioned matrix of the LSE. Moreover, in the case of an oversampled dataset or intended reduction, we want to reduce the given problem, i.e. reduce the number of weights and used basis functions, and preserve good precision of the approximated solution. The approach, which includes the reduction, is called RBF approximation. In the following, the method recently introduced in [11] is described in detail.

For simplicity, we assume that we have an unordered dataset $\{\mathbf{x}_i\}_1^N$ in E^2. However, note that this approach is generally applicable for d-dimensional space. Further, each point $\mathbf{x}_i$ from the dataset is associated with vector $\mathbf{h}_i \in E^p$ of the given values, where p is the dimension of the vector, or scalar value $h_i \in E^1$. For an explanation of the RBF approximation, let us consider the case when each point $\mathbf{x}_i$ is associated with scalar value h_i. Now we extend the given dataset by a set of new reference points $\{\boldsymbol{\xi}_j\}_1^M$, see Fig. 1.

These reference points may not necessarily be in a uniform grid. It is appropriate, that their placement reflects the given surface as well as possible. A good placement of the reference points improves the approximation of the underlying data. For example, when a terrain is to be approximated, placement along features such as break lines leads to better approximation results. The number of added reference points $\boldsymbol{\xi}_j$ is M, where $M \ll N$. The RBF approximation is based on computing the distance of given point $\mathbf{x}_i$ and reference point $\boldsymbol{\xi}_j$ from the extended dataset.

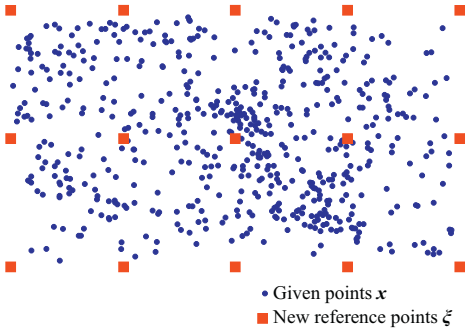


Fig. 1. RBF approximation and reduction of points.

The approximated value can be determined similarly as for interpolation (see [11]):

$$f(\mathbf{x}) = \sum_{j=1}^{M} c_j \phi(r_j) = \sum_{j=1}^{M} c_j \phi(\|\mathbf{x} - \boldsymbol{\xi}_j\|), \tag{7}$$

where the approximating function $f(\mathbf{x})$ is represented as a sum of M RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by an appropriate coefficient c_j.

It can be seen that we get an overdetermined LSE for the given dataset:

$$h_i = f(\mathbf{x}_i) = \sum_{j=1}^{M} c_j \phi(\|\mathbf{x}_i - \boldsymbol{\xi}_j\|) = \sum_{j=1}^{M} c_j \phi_{i,j}, \qquad i = 1, \ldots, N. \tag{8}$$

The linear system of Eq. (8) can be represented as the matrix equation:

$$\mathbf{Ac} = \mathbf{h}, \tag{9}$$

where the number of rows is $N \gg M$ and M is the number of unknown weights $[c_1, \ldots, c_M]^T$, i.e. the number of reference points. Eq. (9) can be expressed in the form:

$$\begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,M} \\ \vdots & \ddots & \vdots \\ \phi_{i,1} & \cdots & \phi_{i,M} \\ \vdots & \ddots & \vdots \\ \phi_{N,1} & \cdots & \phi_{N,M} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_i \\ \vdots \\ h_N \end{pmatrix}. \tag{10}$$

Thus the presented system is overdetermined, i.e. the number of equations N is higher than number of variables M. This LSE can be solved by the least squares method as $\mathbf{A}^T\mathbf{Ac} = \mathbf{A}^T\mathbf{h}$ or singular value decomposition, etc.

4. RBF approximation with polynomial reproduction

The method which was introduced in Section 3 can theoretically have problems with stability and solvability. Therefore, the RBF approximant (7) is usually extended by polynomial function $P_k(\mathbf{x})$ of degree k. Now, the approximated value can be expressed in the form:

$$f(\mathbf{x}) = \sum_{j=1}^{M} c_j \phi(\|\mathbf{x} - \boldsymbol{\xi}_j\|) + P_k(\mathbf{x}). \tag{11}$$

where $\boldsymbol{\xi}_j$ are reference points specified by a user. This leads to solving the LSE:

$$\begin{aligned} h_i = f(\mathbf{x}_i) &= \sum_{j=1}^{M} c_j \phi(\|\mathbf{x}_i - \boldsymbol{\xi}_j\|) + P_k(\mathbf{x}_i) \\ &= \sum_{j=1}^{M} c_j \phi_{i,j} + P_k(\mathbf{x}_i), \qquad i = 1, \ldots, N. \end{aligned} \tag{12}$$

In practice, a linear polynomial

$$P_1(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + a_0 \tag{13}$$

Table 1
Used global RBFs (α is a shape parameter).

RBF	$\phi(r)$
Gauss function [12]	$e^{-(\alpha r)^2}$
Inverse quadric (IQ)	$\frac{1}{1+(\alpha r)^2}$
Thin-plate spline (TPS) [13]	$(\alpha r)^2 \log(\alpha r)$

is used. Geometrically, the coefficient a_0 determines the placement of the hyperplane and the expression $\mathbf{a}^T\mathbf{x}$ represents the inclination of the hyperplane.

It can be seen that for d-dimensional space a linear system of N equations in $(M+d+1)$ variables has to be solved, where N is the number of points in the given dataset, M is the number of reference points and d is the dimensionality of space, e.g. for $d=2$ vectors $\mathbf{x}_i$ and $\mathbf{a}$ are given as $\mathbf{x}_i = (x_i, y_i)^T$ and $\mathbf{a} = (a_x, a_y)^T$. Using the matrix notation, we can write for E^2:

$$\begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,M} & x_1 & y_1 & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \phi_{i,1} & \cdots & \phi_{i,M} & x_i & y_i & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \phi_{N,1} & \cdots & \phi_{N,M} & x_N & y_N & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \\ a_x \\ a_y \\ a_0 \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_i \\ \vdots \\ h_N \end{pmatrix}. \tag{14}$$

Eq. (14) can also be expressed in the form:

$$\begin{pmatrix} \mathbf{A} & \mathbf{P} \end{pmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{a} \\ a_0 \end{pmatrix} = \mathbf{h}. \tag{15}$$

It can be seen that for E^2 we have a linear system of N equations in $(M+3)$ variables, where $M \ll N$. Thus the presented system is overdetermined again and can also be solved by the method of least squares or singular value decomposition.

5. Experimental results

The above presented methods of the RBF approximation have been tested on synthetic and real datasets. Moreover, different global radial basis functions with shape parameter α, see Table 1, and different sets of reference points have been used for testing. These sets of reference points have different types of distributions described in Section 5.1.

5.1. Distribution of reference points

For these experiments, the following sets of reference points were used:

Points on regular grid: This set contains the points on a regular grid in E^2.
Epsilon points: This distribution of reference points is described in the following text.
Epsilon points + AABB corners: This set of points is determined in the same manner as the previous case. Moreover, the corners of axis aligned bounding box (AABB) of Epsilon points are added to the set of reference points.
Halton points: This distribution of points is described in the following text in detail. However, note that this set of reference points equals the subset of the given dataset, for which we determine the RBF approximation.
Halton points + AABB corners: This set of reference points is determined in the same manner as Halton points. Moreover, the corners of AABB are added to this set.

5.1.1. Halton points

Construction of a Halton sequence is based on a deterministic method. This sequence generates well-spaced "draws" points from the interval [0, 1]. The sequence uses a prime number as its base and is constructed based on finer and finer prime-based divisions of sub-intervals of the unit interval. The Halton sequence [10] can be described by the following recurrence formula:

$$\text{Halton}(p)_k = \sum_{i=0}^{\lfloor \log_p k \rfloor} \frac{1}{p^{i+1}} \left(\left\lfloor \frac{k}{p^i} \right\rfloor \bmod p \right), \tag{16}$$

where p is the prime number and k is the index of the calculated element.

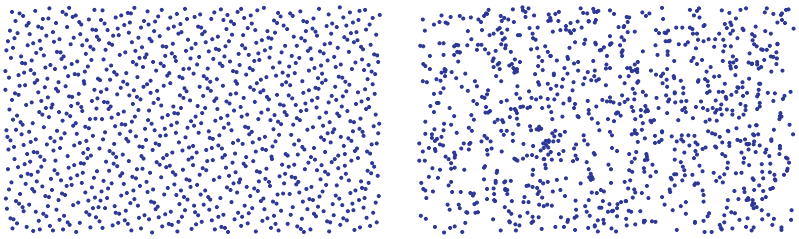


Fig. 2. Halton points in E^2 generated by Halton(2,3) (left) and random points in a rectangle with uniform distribution (right). The number of points is 10^3 in both cases.

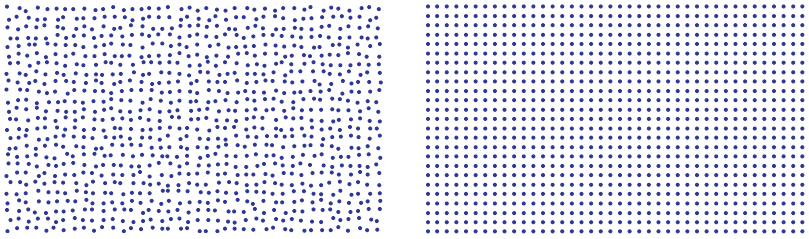


Fig. 3. Epsilon points (left) and points on a 2D regular grid (right). The number of points is $40 \times 25 = 10^3$ in both cases.

For the E^2 space, subsequent prime numbers are used as a base. In this test, {2, 3} were used for the Halton sequence and the following sequence of points in a rectangle (a, b) was derived:

$$\begin{aligned}\text{Halton}(2,3) = \Big\{&\Big(\frac{1}{2}a, \frac{1}{3}b\Big), \Big(\frac{1}{4}a, \frac{2}{3}b\Big), \Big(\frac{3}{4}a, \frac{1}{9}b\Big), \Big(\frac{1}{8}a, \frac{4}{9}b\Big), \Big(\frac{5}{8}a, \frac{7}{9}b\Big),\\ &\Big(\frac{3}{8}a, \frac{2}{9}b\Big), \Big(\frac{7}{8}a, \frac{5}{9}b\Big), \Big(\frac{1}{16}a, \frac{8}{9}b\Big), \Big(\frac{9}{16}a, \frac{1}{27}b\Big), \dots\Big\},\end{aligned} \tag{17}$$

where a is the width of the rectangle and b is the height of the rectangle.

Visualization of the dataset with 10^3 points of the Halton sequence from (17) can be seen in Fig. 2. We can see that the Halton sequence in E^2 space covers this space more evenly than randomly distributed uniform points in the same rectangle.

5.1.2. Epsilon points

This is a special distribution of points in E^2, which is based on a regular grid. Each point is determined as follows:

$$\begin{aligned}P_{ij} = \Big[&i \cdot \Delta x + \text{rand}(-\varepsilon_x, \varepsilon_x),\ j \cdot \Delta y + \text{rand}(-\varepsilon_y, \varepsilon_y)\Big],\\ &\varepsilon_x \approx 0.25 \cdot \Delta x, \qquad i = 0, \dots, N_x,\\ &\varepsilon_y \approx 0.25 \cdot \Delta y, \qquad j = 0, \dots, N_y,\end{aligned} \tag{18}$$

where Δx and Δy are real numbers representing the grid spacing, N_x indicates the number of grid columns, N_y is the number of grid rows and $\text{rand}(-\varepsilon_x, \varepsilon_x)$ or $\text{rand}(-\varepsilon_y, \varepsilon_y)$ is a random drift with a uniform distribution from $-\varepsilon_x$ to ε_x or from $-\varepsilon_y$ to ε_y.

Fig. 3 presents the dataset with 40×25, (i.e. 10^3) epsilon points. Moreover, we can see the comparison of this distribution of points with points on a regular grid.

5.2. Synthetic datasets

The Halton distribution of points was used for synthetic data. Moreover, each point from this dataset is associated with a function value at this point. For this purpose, different functions have been used for experiments. Results for a 2D sinc function, see Fig. 4, are presented in this paper.

5.2.1. Examples of RBF approximation results

Some examples of RBF approximation to 1089 Halton data points sampled from a 2D sinc function, for a Halton set of reference points, which consists of 81 points, and different RBFs are shown in Figs. 5 and 6.

It can be seen that the RBF approximation using Lagrange multiplies (Fasshauer [10]) returns the worst result in terms of the error in comparison with the proposed methods. Further, in Fig. 6, it can be seen that the errors for all RBF approximation methods are much higher when the TPS is used.

There is a question of how the RBF approximation depends on the shape parameter α. This is described in the following section.

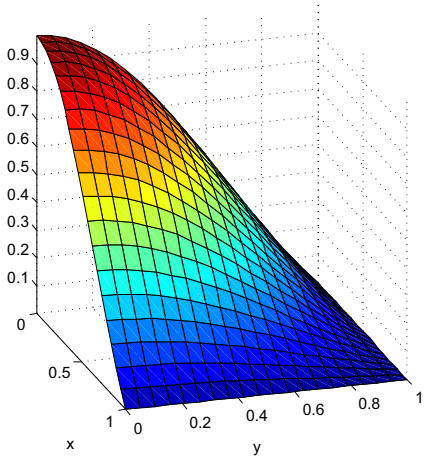


Fig. 4. 2D sinc function $\mathrm{sinc}(\pi x)\cdot\mathrm{sinc}(\pi y)$ whose domain is restricted to $[0, 1]\times[0, 1]$.

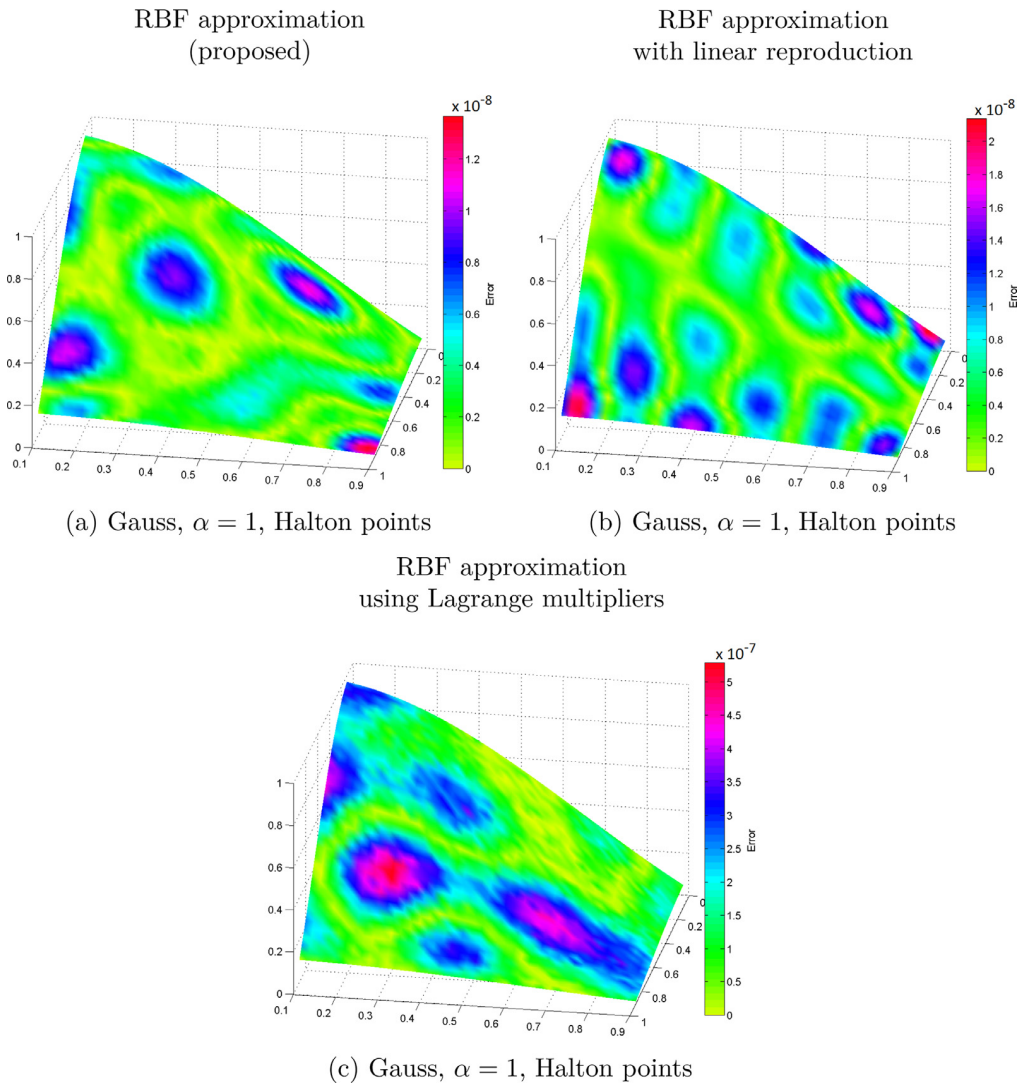


Fig. 5. Approximation to 1089 data points sampled from a 2D sinc function with 81 Halton-spaced Gaussian basis functions false-colored by magnitude of absolute error.

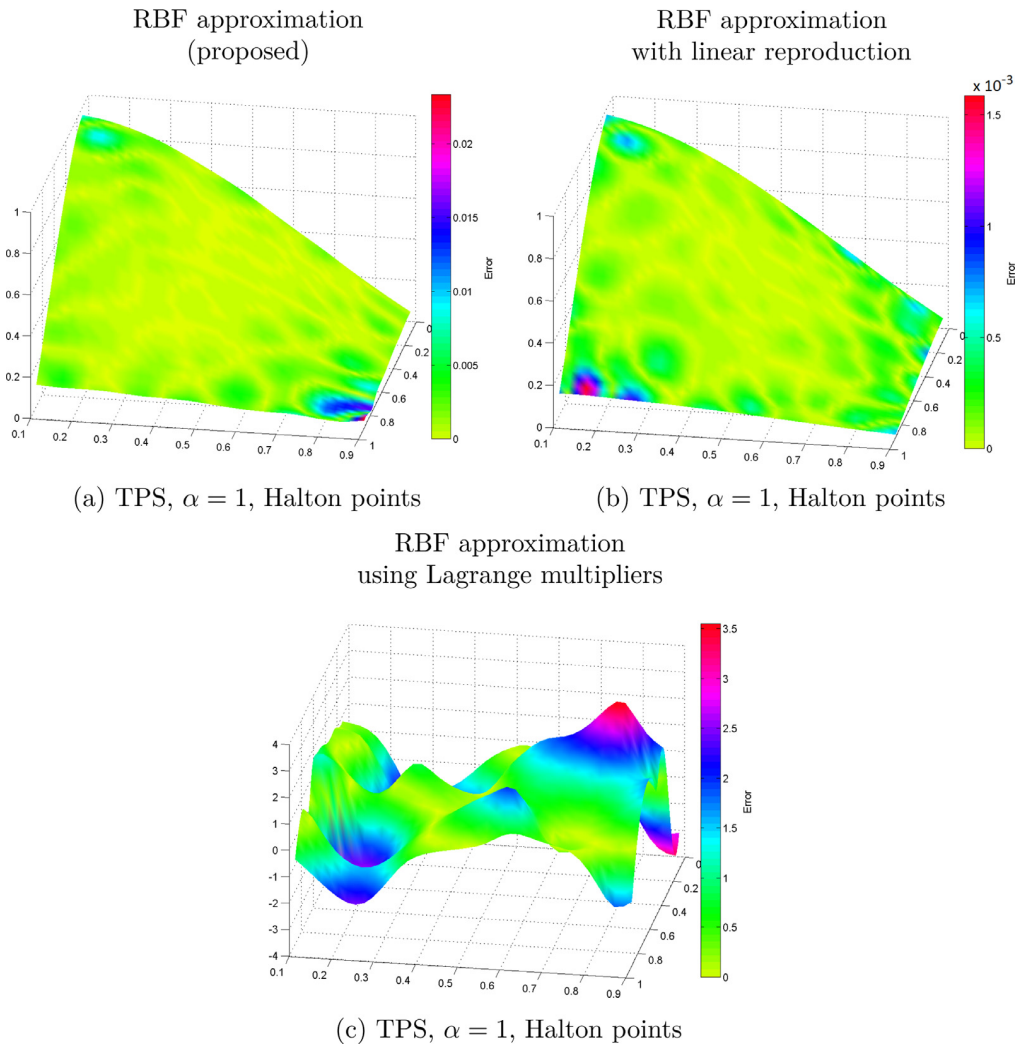


Fig. 6. Approximation to 1089 data points sampled from a 2D sinc function with 81 Halton-spaced TPS false-colored by magnitude of absolute error.

5.2.2. Comparison of methods

In this section, the different versions of RBF approximation which were presented in Sections 2–4 are compared. Fig. 7 presents the mean absolute error of RBF approximation for the dataset, which consists of 1089 Halton points in the range $[0, 1] \times [0, 1]$, sampled from a 2D sinc function, while the set of reference points contains 81 points with Halton behavior of the distribution, and for different global radial basis functions. The graphs represent the mean absolute error according to a shape parameter α of used RBFs. We can see that for RBF approximation using Lagrange multipliers (Fasshauer [10]) we obtain a higher mean absolute error. Mean absolute errors for RBF approximation and RBF approximation with linear reproduction are almost the same. Moreover, the Gaussian RBF gives the best result for shape parameter $\alpha = 1$ and the inverse quadric for $\alpha = 0.5$. Further, the TPS function is not appropriate to solve the given problem, see Fig. 7c. Note, the standard deviation of errors was also measured and the same behavior and order of magnitude was obtained as for the mean absolute errors.

5.2.3. Comparison of different distributions of reference points

In this section, we focus on a comparison of the presented RBF approximation methods due to used distribution of reference points. Measurements of errors were performed for different type of RBFs with different shape parameters. Mean absolute error according to shape parameter α for Gaussian RBF is presented in Fig. 8.

We can see that for all versions of RBF approximation the worst result is obtained for reference points on a regular grid (u). For the proposed RBF approximation, the remaining sets of reference points give almost the same results. Reference points corresponding to epsilon points + AABB (epsaabb) almost always give the best result for RBF approximation with linear reproduction. For RBF approximation using Lagrange multipliers, the best results are for the reference points which have a Halton distribution.

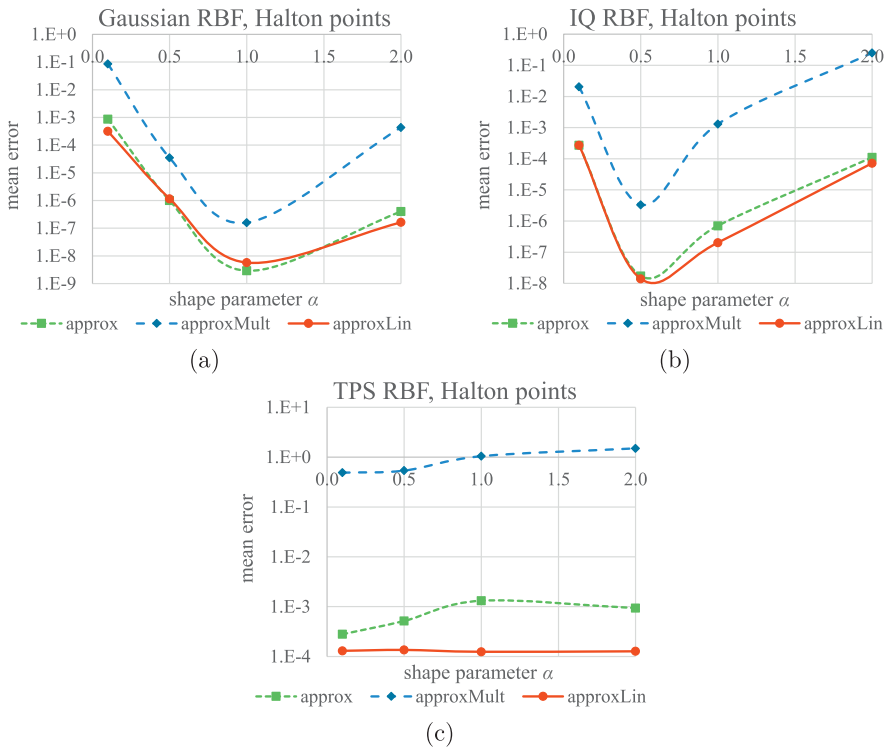


Fig. 7. The mean absolute error of approximation to 1089 data points sampled from a 2D sinc function with 81 reference Halton points for different RBF approximation methods, different RBFs and different shape parameters. The used approximation methods are: proposed RBF approximation (approx), RBF approximation using Lagrange multipliers (approxMult) and RBF approximation with linear reproduction (approxLin). RBFs are: (a) Gauss function, (b) IQ, (c) TPS.

5.2.4. Comparison by placement of the dataset in E^2

This section is focused on placement of the actual dataset in the domain space and the used function generating associated scalar values in E^2. The given dataset has a range of one in both axes and the function generating associated scalar values is a 2D sinc function. Two configurations for placement of the origin of the dataset and the maximum of the 2D sinc function were used. The first configuration is at point (0; 0); the second is moved to point (3, 951, 753; 2, 785, 412).

Fig. 9 presents the mean absolute error for these configurations, when the Gaussian basis functions and Halton set of reference points were chosen. We can see that RBF approximation with linear reproduction gives a higher error for the second configuration, i.e. placement at point (3, 951, 753; 2, 785, 412). For RBF approximation using Lagrange multiplier the decision is not ambiguous. Note that a graph for the proposed RBF approximation is not presented, because both configurations give the same results.

5.2.5. Optimal number of reference points

This section focuses on the influence of the number of reference points. The number of reference points is determined relative to the number of points in the given dataset. Measurements for different shape parameters were performed many times and average mean absolute errors were computed, see Figs. 10–12. Note that the reference points were distributed by Halton distribution. Fig. 10 presents the mean absolute error for the Gaussian RBF approximation. Experimental results for the IQ are shown in Fig. 11. We can see that for the small shape parameter α the mean absolute errors are almost constant. However, for greater shape parameters the mean absolute error decreases with the increasing number of reference points.

Fig. 12 presents experimental results obtained for the TPS function. We can see that the mean absolute error decreases with the increasing number of reference points as would be expected.

Finally, note that the results for RBF approximation with reproduction are very similar to the proposed RBF approximation. RBF using Lagrange multipliers has unpredictable behavior and no trend can be established.

5.3. Real datasets

The presented methods of the RBF approximation have been also tested on real data. Let us introduce results for real dataset which was obtained from GPS data of mount Veľký Rozsutec in the Malá Fatra, Slovakia (Fig. 13).[1] Each point of this

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

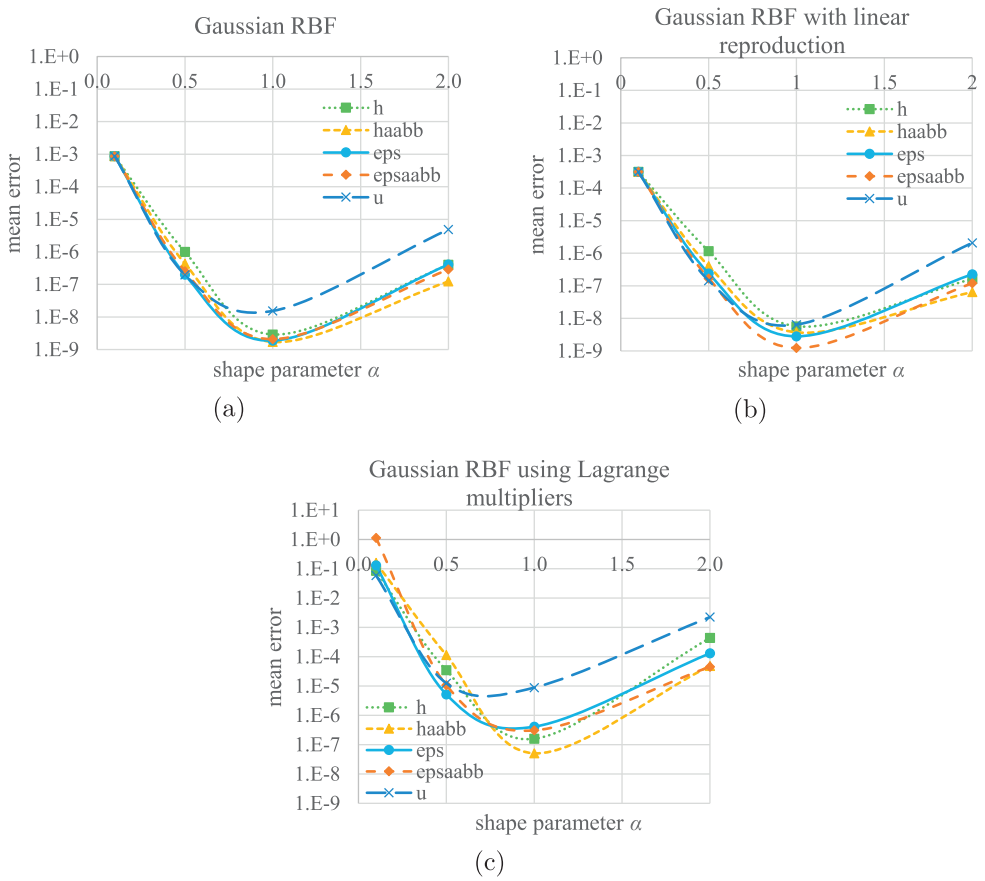


Fig. 8. The mean absolute error of approximation to 1089 data points sampled from a 2D sinc function with 81 spaced Gaussian basis functions for different RBF approximation methods, different shape parameters and different sets of reference points. The sets of reference points are: Halton points (h), Halton points + AABB (haabb), epsilon points (eps), epsilon points + AABB (epsaabb), points on a regular grid (u). Their description is in Section 5.1. Versions of approximation are: (a) RBF approximation, (b) RBF approximation with linear reproduction, (c) RBF approximation using Lagrange multipliers.

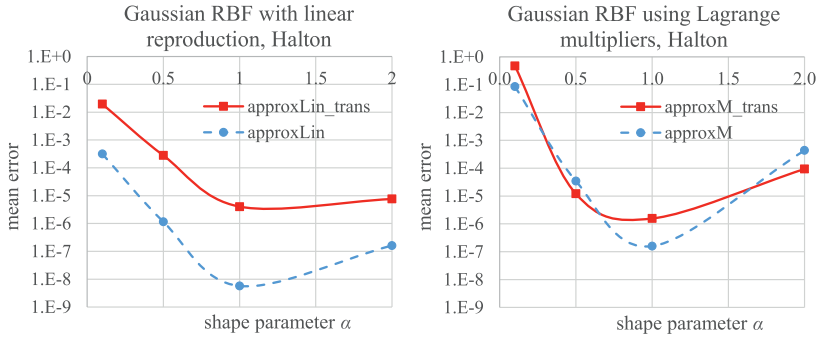


Fig. 9. The mean absolute error of approximation to 1089 data points sampled from a 2D sinc function with 81 spaced Gaussian basis functions for a Halton set of reference points, different RBF approximation methods and different shape parameters. The placement of the given dataset and the maximum of the 2D sinc function are at point (0; 0) (circles) or at point (3, 951, 753; 2, 785, 412) (squares). Versions of approximation are RBF approximation with linear reproduction (left) and RBF approximation using Lagrange multipliers (right).

dataset is associated with its elevation. Moreover, as a first step, the real dataset is translated so that its estimated center of gravity corresponds to the origin of the coordinate system. This step is used due to the limitation of the influence of dataset placement in space and it was chosen based on the results of experiments described in Section 5.2.4. Table 2 gives an overview of the used dataset.

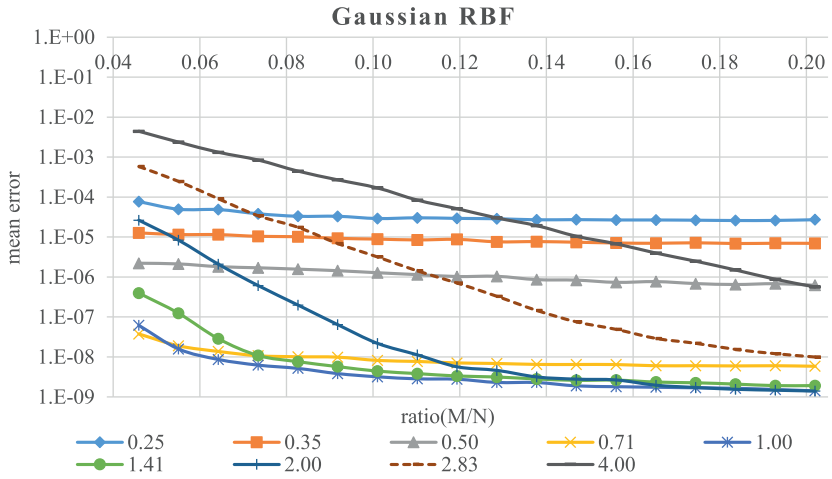


Fig. 10. The mean absolute error of the proposed RBF approximation to 1089 data points sampled from a 2D sinc function for different numbers of reference points, Gaussian RBF with different shape parameters α.

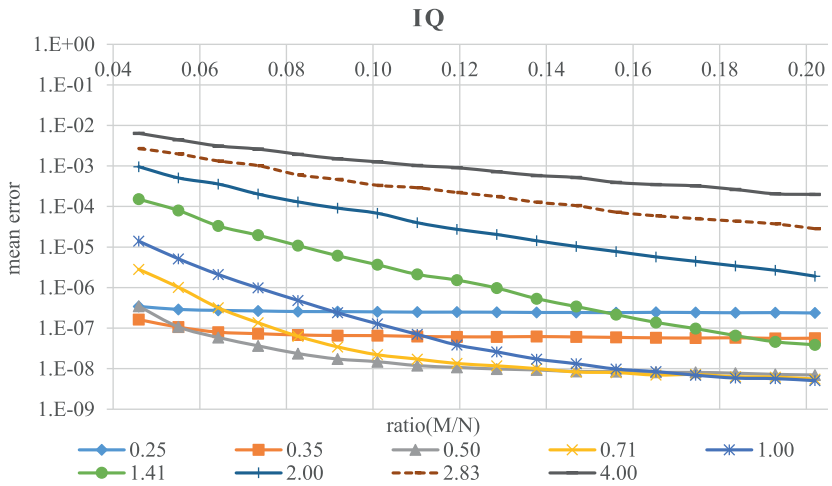


Fig. 11. The mean absolute error of the proposed RBF approximation to 1089 data points sampled from a 2D sinc function for different numbers of reference points, IQ RBF with different shape parameters α.

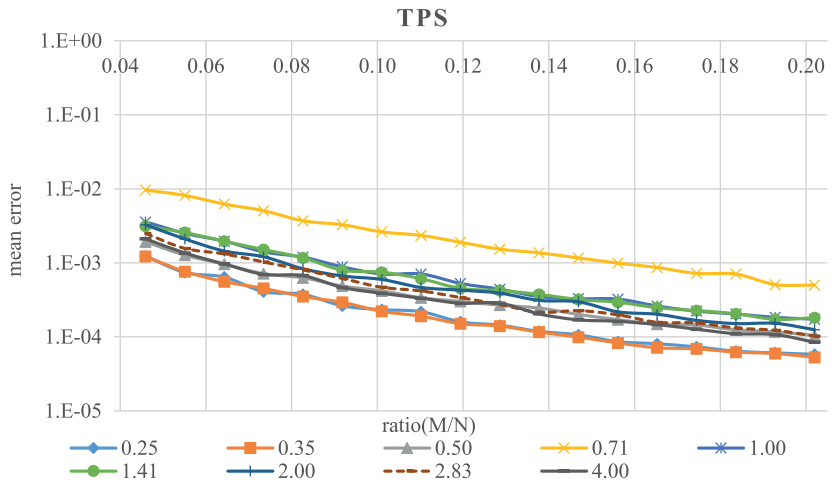


Fig. 12. The mean absolute error of the proposed RBF approximation to 1089 data points sampled from a 2D sinc function for different numbers of reference points, TPS RBF with different shape parameters α.

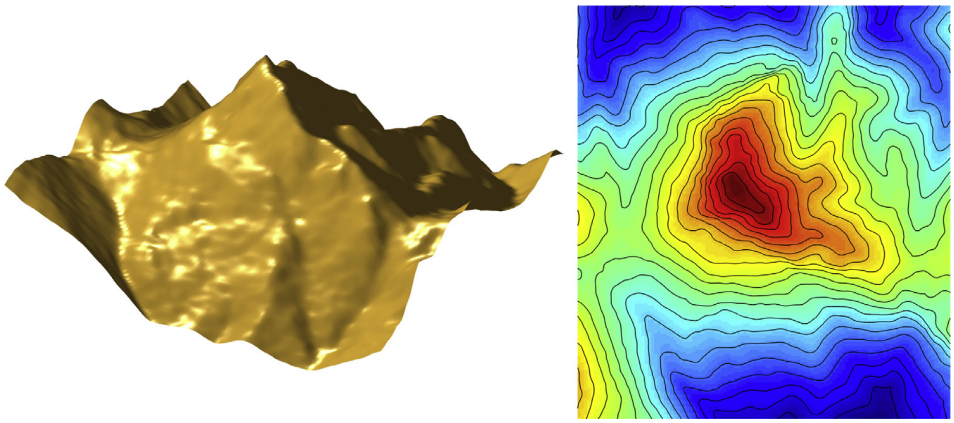


Fig. 13. Mount Veľký Rozsutec, Slovakia (left) and its contour map (right).

Table 2
Overview information for the tested real dataset. The axis-aligned bounding box (AABB) of the tested dataset has a size width × length × relief, i.e. $x_{range} \times y_{range} \times z_{range}$.

Veľký Rozsutec	
Number of pts.	24,190
Relief (m)	818.8000
Width (m)	2608.5927
Length (m)	2884.1169

5.3.1. Examples of RBF approximation results

Results for RBF approximation of mount Veľký Rozsutec dataset using Halton set of reference points, which consists of 484 points, and Gaussian RBF with shape parameter $\alpha = 0.0025$ are shown in Fig. 14 and histograms of errors for these results are shown in Fig. 15.

Note, that the results of RBF approximation using Lagrange multipliers are not presented for real data because this method has unpredictable behavior and is unusable for real dataset, which was already evident from results for synthetic datasets. From presented results, it can be seen that the RBF approximation with linear reproduction returns the worst result in terms of the error in comparison with the proposed method. Moreover, if the results of approximation are compared with the original, it can be seen that the RBF approximation with the global Gaussian RBFs cannot preserve the sharp ridge.

Results for RBF approximation of mount Veľký Rozsutec dataset using Halton set of reference points, which contains different number of points, and TPS with shape parameter $\alpha = 0.005$ are shown in Fig. 16. The histograms of errors for these results are shown in Fig. 17. From these results, it can be seen that with an increasing number of reference points, approximation error is improved and some surface details also begin to appear. However, it can be again seen that the RBF approximation with the global TPS cannot preserve the sharp ridge.

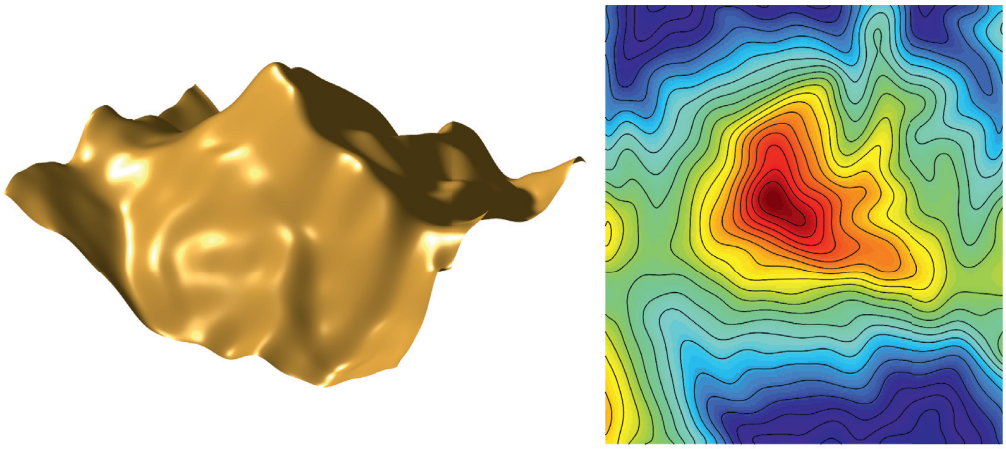
There is a question of how the RBF approximation of real dataset depends on the shape parameter α and distribution of reference points. This is described in the following sections.

5.3.2. Comparison of different distributions of reference points

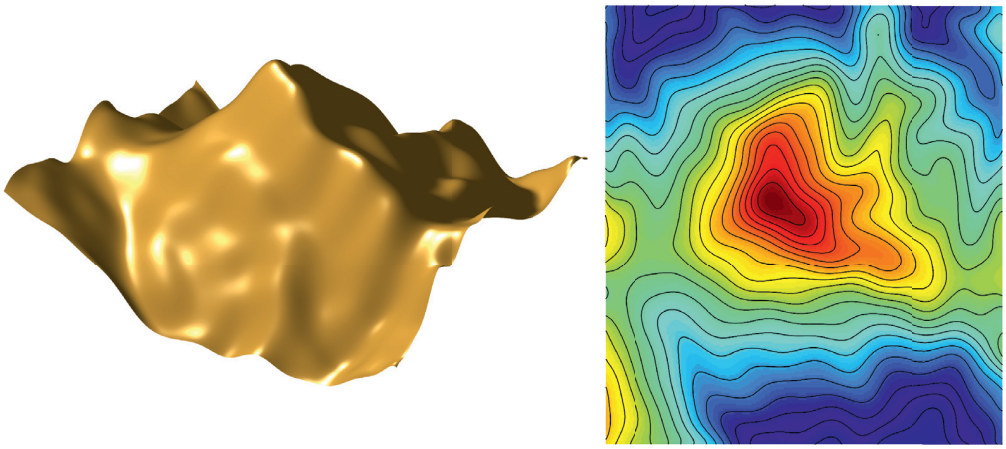
In this section, we focus on a comparison of the presented RBF approximation methods due to used distribution of reference points when the real data are approximated. Measurements of errors were performed for different type of RBFs with different shape parameters. Mean relative error according to shape parameter α for the Gaussian RBF is presented in Fig. 18 and for the IQ is shown in Fig. 19. Note that the mean *relative* error is presented for real data. The reason for this choice is that the function values of the real dataset are not normalized to the interval [0, 1].

We can see that for all versions of RBF approximation, if the shape parameter α is not close to the optimum, the worst results are obtained for reference points with Halton distribution ((h) and (haabb)). The best results are obtained for reference points on a regular grid (u). If the shape parameter is chosen close to the optimum (for the presented configuration $\alpha \approx 0.0025$) then the mean relative error has only minor differences for different distribution of reference points. These results are different in comparison with results obtained for synthetic data.

Finally, note that the mean relative error for approximation of mount Veľký Rozsutec dataset according to shape parameter is constant for the TPS and deviation of mean relative error for different distribution of reference points is almost negligible.



(a) RBF approximation: Gauss, $N = 24,190$, $M = 484$, $\alpha = 0.0025$, Halton points (left) and its contour map (right)



(b) RBF approximation with reproduction: Gauss, $N = 24,190$, $M = 484$, $\alpha = 0.0025$, Halton points (left) and its contour map (right)

Fig. 14. Results for mount Veľký Rozsutec approximated by 484 Halton-spaced Gaussian basis functions with shape parameter $\alpha = 0.0025$.

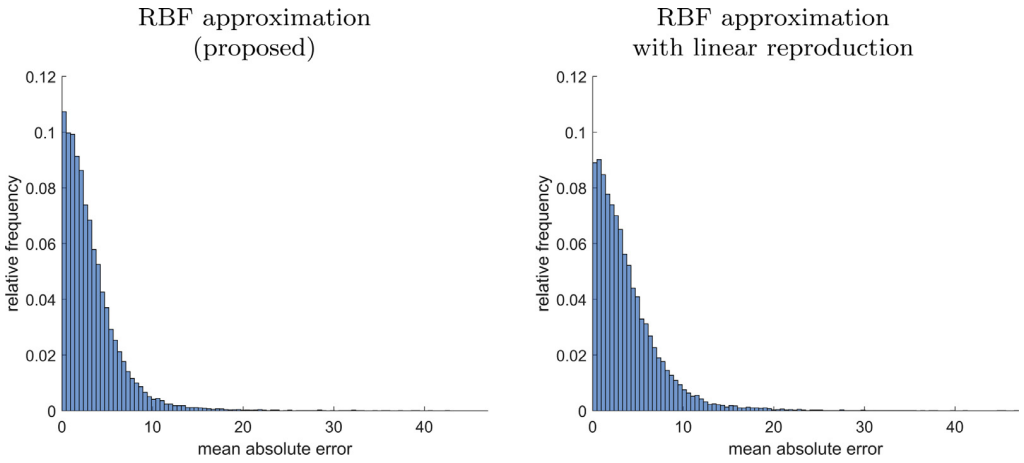
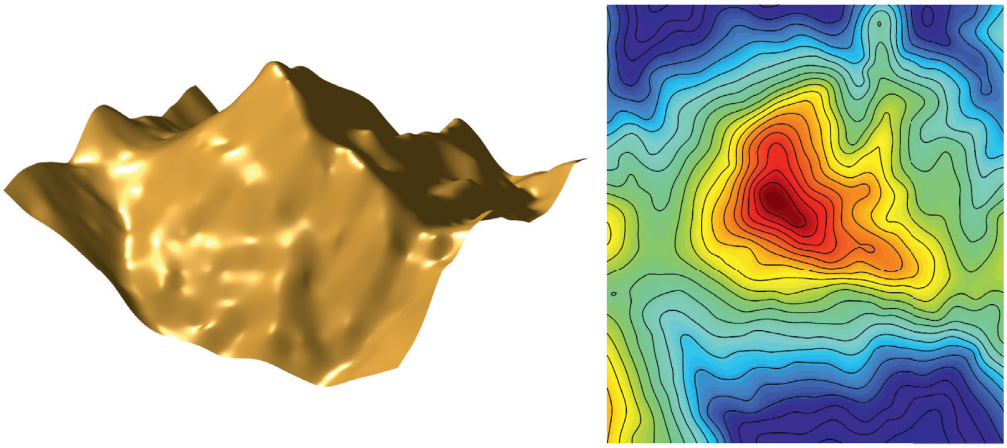
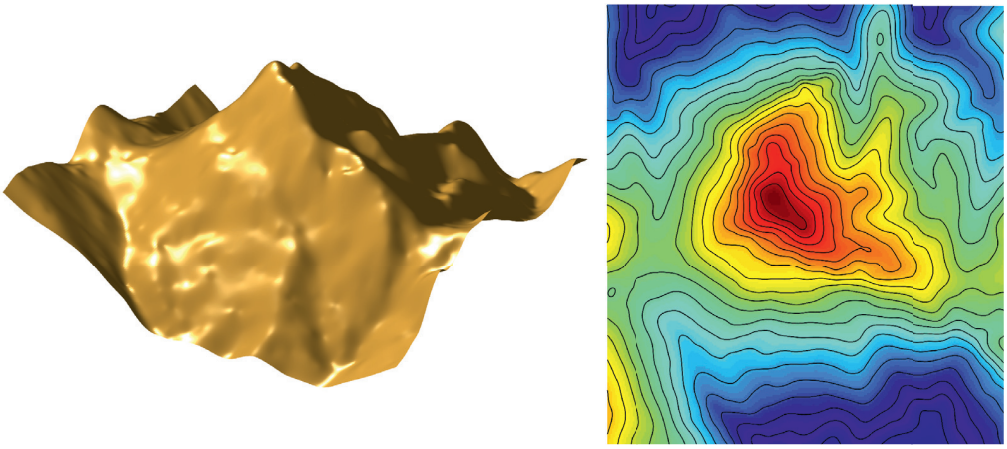


Fig. 15. Histograms of errors for mount Veľký Rozsutec approximated by 484 Halton-spaced Gaussian basis functions with shape parameter $\alpha = 0.0025$.

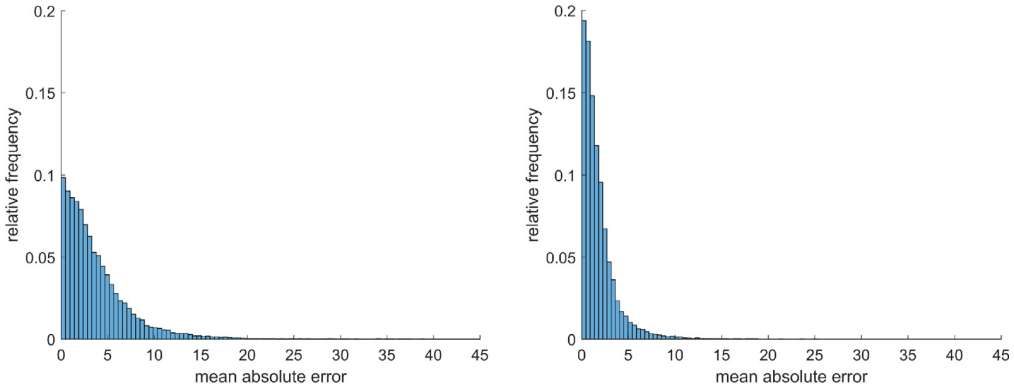


(a) RBF approximation: TPS, $N = 24,190$, $M = 484$, $\alpha = 0.005$, Halton points (left) and its contour map (right)



(b) RBF approximation: TPS, $N = 24,190$, $M = 1089$, $\alpha = 0.005$, Halton points (left) and its contour map (right)

Fig. 16. Results for mount Veľký Rozsutec approximated by Halton-spaced TPS with shape parameter $\alpha = 0.005$.



(a) $M = 484$ (b) $M = 1089$

Fig. 17. Histograms of errors for mount Veľký Rozsutec approximated by Halton-spaced TPS with shape parameter $\alpha = 0.005$.

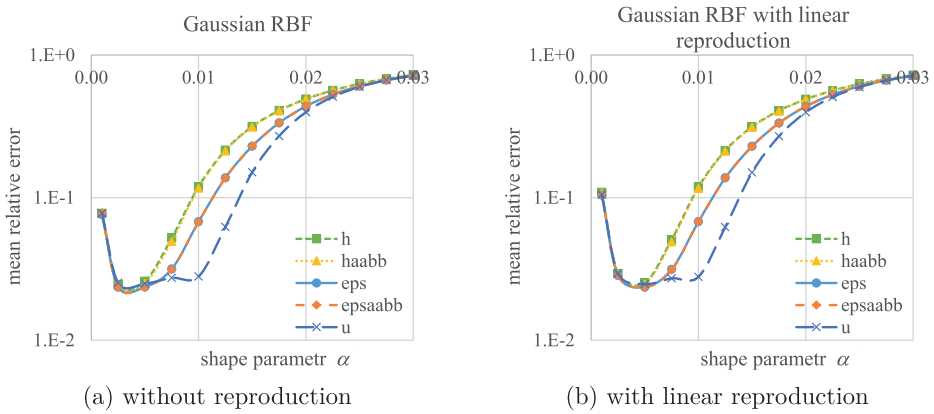


Fig. 18. The mean relative error of approximation for mount Veľký Rozsutec with 484 spaced Gaussian basis functions for different RBF approximation methods, different shape parameters and different sets of reference points. The sets of reference points are: Halton points (h), Halton points + AABB (haabb), epsilon points (eps), epsilon points + AABB (epsaabb), points on a regular grid (u), described in Section 5.1.

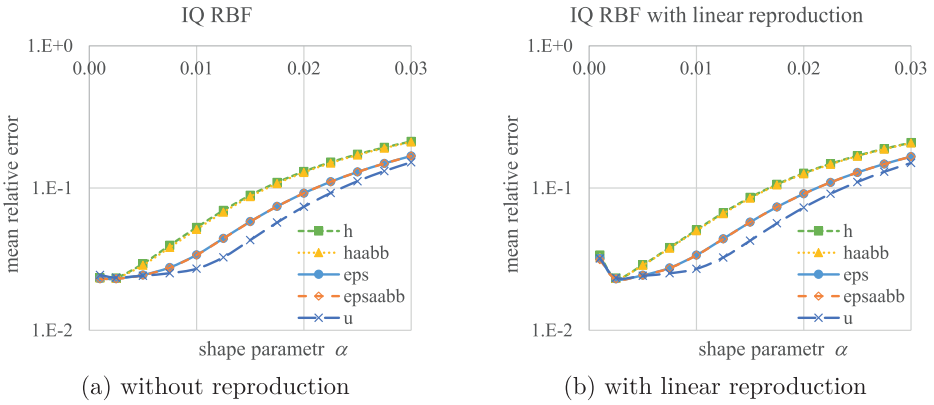


Fig. 19. The mean relative error of approximation for mount Veľký Rozsutec with 484 spaced IQ for different RBF approximation methods, different shape parameters and different sets of reference points. The sets of reference points are: Halton points (h), Halton points + AABB (haabb), epsilon points (eps), epsilon points + AABB (epsaabb), points on a regular grid (u), described in Section 5.1.

5.3.3. Comparison of different radial basis functions

In this section, we focus on a comparison of the results of RBF approximations using different types of RBFs. Real datasets were used for experiments and results of the mount Veľký Rozsutec are presented. Measurements of errors were performed for Halton set with 484 reference points. The shape parameter $\alpha = 0.0025$ was chosen for all types of RBFs. The differences of frequencies of errors are shown in Fig. 20. It can be seen that the best error is obtained for the RBF approximation using the IQ function. On the contrary, the worst error returns the RBF approximation using the TPS function.

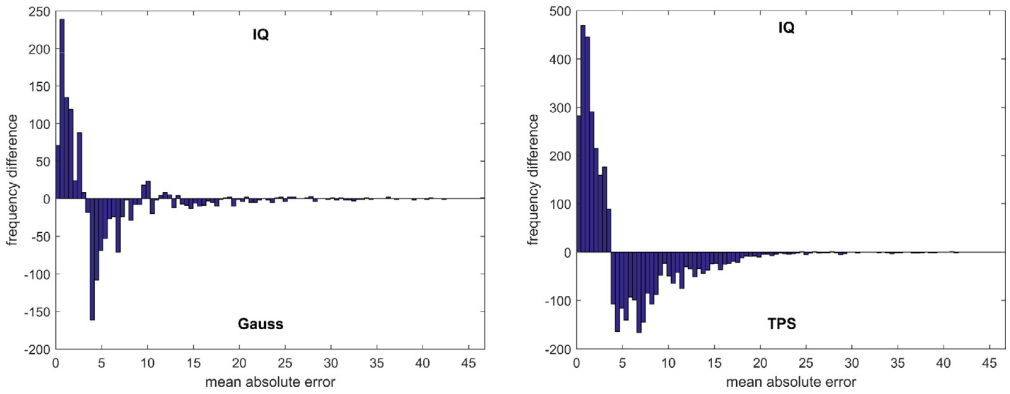
Finally, note that the results for the RBF approximation with linear reproduction are similar to the proposed RBF approximation.

5.3.4. Optimal number of reference points

This section focuses on the influence of the number of reference points for RBF approximation of mount Veľký Rozsutec dataset. The number of reference points is determined relative to the number of points in the given dataset. Measurements for different shape parameters were performed many times and average mean relative errors were computed, see Figs. 21–23. Note that the reference points were distributed by Halton distribution. Fig. 21 presents the mean relative error for the Gaussian RBF approximation and Fig. 22 presents results for the IQ. It can be seen that for small shape parameter α the mean relative errors are almost constant. However, for the greater shape parameters the mean relative error decreases with the increasing number of reference points. These results are consistent with results for synthetic dataset.

Fig. 23 presents experimental results obtained for TPS. We can see that the mean relative error is independent on the shape parameter α and decreases with the increasing number of reference points.

Finally, note that the results for RBF approximation with linear reproduction are very similar to the proposed RBF approximation.



(a) IQ vs. Gaussian RBF (b) IQ vs. TPS

Fig. 20. Difference of frequencies of error for mount Veľký Rozsutec approximated by 484 Halton-spaced RBFs with shape parameter $\alpha = 0.0025$.

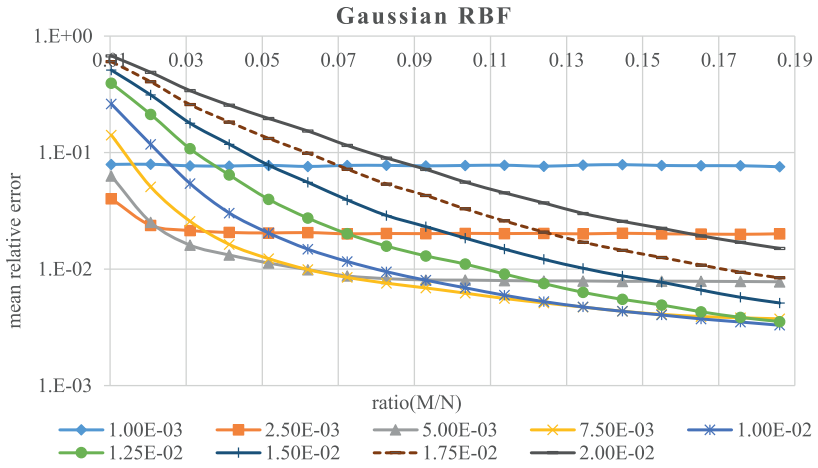


Fig. 21. The mean relative error of the proposed RBF approximation of mount Veľký Rozsutec dataset for different numbers of reference points, Gaussian RBF with different shape parameters α.

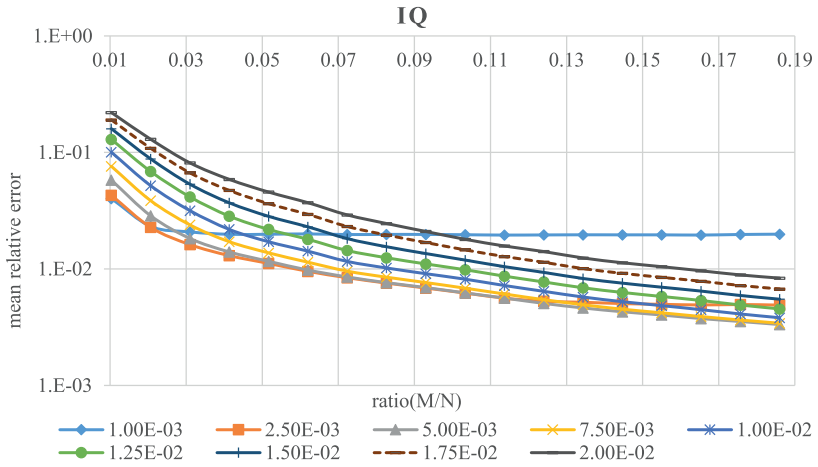


Fig. 22. The mean relative error of the proposed RBF approximation of mount Veľký Rozsutec dataset for different numbers of reference points, IQ RBF with different shape parameters α.

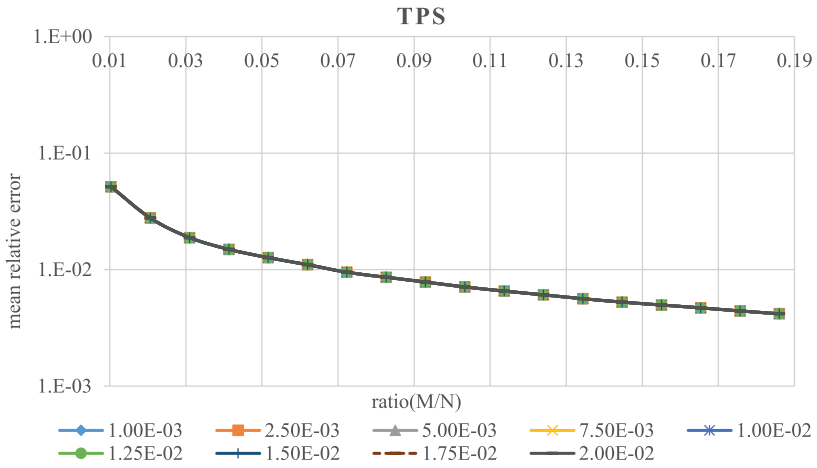


Fig. 23. The mean relative error of the proposed RBF approximation of mount Veľký Rozsutec dataset for different numbers of reference points, TPS RBF with different shape parameters α.

6. Conclusion

Comparisons of different methods of RBF approximation with respect to various criteria were presented. The proposed RBF approximation introduced in Section 3 gives the best results due to the smallest error. The RBF approximation with a linear reproduction can be influenced by placement of the given dataset in space. Therefore, it is appropriate that the translation of the estimated center of gravity to the origin of the coordinate system is made as the first step. The worst results according to error were obtained using the RBF approximation using Lagrange multipliers. Moreover, this method of approximation has unpredictable behavior, the matrix for RBF approximation using Lagrange multipliers is mostly ill-conditioned and its size is high, i.e. it is of the $(M+N) \times (M+N)$ size.

The experiments proved that the proposed RBF approximation gives significantly better result over other methods used in the experiments described above. It also offers a possible data compression as the matrix is only $M \times M$, where $M \ll N$, which is a significant factor for large datasets processing. On the other hand, experiments made also proved that all methods have problems with the preservation of sharp edges if global functions are used.

Future work will be devoted to evaluation of compactly-supported RBFs (CS-RBFs) which will lead to sparse matrices, decrease of memory requirements and significant increase of speed of computation. A special attention will be given to finding optimal shape parameters which are critical for the RBF approximation quality.

Acknowledgments

The authors would like to thank their colleagues at the University of West Bohemia, Plzen, for their discussions and suggestions, and anonymous reviewers for their valuable comments and hints provided. The research was supported by the National Science Foundation GAČR project GA17-05534S and partially supported by SGS-2016-013.

References

[1] R.L. Hardy, Multiquadratic equations of topography and other irregular surfaces, J. Geophys. Res. 76 (1971) 1905–1915.
[2] R. Pan, V. Skala, Continuous global optimization in surface reconstruction from an oriented point cloud, Comput. Aided Des. 43 (8) (2011) 896–901.
[3] R. Pan, V. Skala, A two-level approach to implicit surface modeling with compactly supported radial basis functions, Eng. Comput. 27 (3) (2011) 299–307.
[4] V. Skala, R. Pan, O. Nedved, Simple 3D surface reconstruction using flatbed scanner and 3D print, in: SIGGRAPH Asia 2013, Hong Kong, China, November 19–22, 2013, Poster Proceedings, ACM, 2013, p. 7.
[5] V. Skala, R. Pan, O. Nedved, Making 3D replicas using a flatbed scanner and a 3D printer, in: Computational Science and Its Applications – ICCSA 2014 – 14th International Conference, Guimarães, Portugal, June 30–July 3, 2014, Proceedings, Part VI, in: Lecture Notes in Computer Science, 8584, Springer, 2014, pp. 76–86.
[6] K. Uhlir, V. Skala, Reconstruction of damaged images using radial basis functions, in: Proceedings of EUSIPCO 2005, 2005, p. 160.
[7] J. Zapletal, P. Vaněček, V. Skala, RBF-based image restoration utilising auxiliary points, in: Proceedings of the 2009 Computer Graphics International Conference, ACM, 2009, pp. 39–43.
[8] H. Wendland, Computational aspects of radial basis function approximation, Stud. Comput. Math. 12 (2006) 231–256.
[9] E. Darve, The fast multipole method: numerical implementation, J. Comput. Phys. 160 (1) (2000) 195–240.
[10] G.E. Fasshauer, Meshfree Approximation Methods with MATLAB, 6, World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2007.
[11] V. Skala, Fast interpolation and approximation of scattered multidimensional and dynamic data using radial basis functions, WSEAS Trans. Math. 12 (5) (2013) 501–511.
[12] I. Schagen, Interpolation in two dimensions - a new technique, IMA J. Appl. Math. 23 (1) (1979) 53–59.
[13] J. Duchon, Splines minimizing rotation-invariant semi-norms in Sobolev spaces, Constructive Theory of Functions of Several Variables, Springer, 1977, pp. 85–100.

Appendix D

A Radial Basis Function Approximation for Large Datasets

Majdišová, Z., Skala, V.

Proceedings of SIGRAD 2016, pp. 9-14, Linköping University Electronic Press (2016),
ISSN 1650-3686, ISBN 978-91-7685-731-1

SIGRAD 2016
M. Hayashi (Editors)

A Radial Basis Function Approximation for Large Datasets

Z. Majdisova[1] and V. Skala[1]

[1]Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitni 8, CZ 30614 Plzen, Czech Republic

Abstract

Approximation of scattered data is often a task in many engineering problems. The Radial Basis Function (RBF) approximation is appropriate for large scattered datasets in d-dimensional space. It is non-separable approximation, as it is based on a distance between two points. This method leads to a solution of overdetermined linear system of equations.

In this paper a new approach to the RBF approximation of large datasets is introduced and experimental results for different real datasets and different RBFs are presented with respect to the accuracy of computation. The proposed approach uses symmetry of matrix and partitioning matrix into blocks.

Categories and Subject Descriptors (according to ACM CCS): G.1.2 [Numerical Analysis]: Approximation—Approximation of Surfaces and Contours

1. Introduction

Interpolation and approximation are the most frequent operations used in computational techniques. Several techniques have been developed for data interpolation or approximation, but they mostly expect an ordered dataset, e.g. rectangular mesh, structured mesh, unstructured mesh etc. However, in many engineering problems, data are not ordered and they are scattered in $d-$dimensional space, in general. Usually, in technical applications the conversion of a scattered dataset to a semi-regular grid is performed using some tessellation techniques. However, this approach is quite prohibitive for the case of $d-$dimensional data due to the computational cost.

Interesting techniques are based on the Radial Basis Function (RBF) method which was originally introduced by [Har71]. They are widely used across of many fields solving technical and non-technical problems. The RBF applications can be found in neural networks, data visualization [PRF14], surface reconstruction [CBC*01], [TO02], [PS11], [SPN13], [SPN14], solving partial differential equations [LCC13], [HSfY15], etc. The RBF techniques are really meshless and are based on collocation in a set of scattered nodes. These methods are independent with respect to the dimension of the space. The computational cost of this techniques increase nonlinearly with the number of points in the given dataset and linearly with the dimensionality of data.

There are two main groups of basis functions: global RBFs and Compactly Supported RBFs (CS-RBFs) [Wen06]. Fitting scattered data with CS-RBFs leads to a simpler and faster computation, but techniques using CS-RBFs are sensitive to the density of scattered data. Global RBFs lead to a linear system of equations with a dense matrix and their usage is based on sophisticated techniques such as the fast multipole method [Dar00]. Global RBFs are useful in repairing incomplete datasets and they are insensitive to the density of scattered data.

For the processing of scattered data we can use the RBF interpolation or the RBF approximation. The RBF interpolation, e.g. presented by [Ska15], is based on a solution of a linear system of equations:

$$\mathbf{Ac} = \mathbf{h}, \tag{1}$$

where $\mathbf{A}$ is a matrix of this system, $\mathbf{c}$ is a column vector of variables and $\mathbf{h}$ is a column vector containing the right sides of equations. In this case, $\mathbf{A}$ is an $N \times N$ matrix, where N is the number of points in the given scattered dataset, the variables are weights for basis functions and the right sides of equations are values in the given points. The disadvantage of RBF interpolation is the large and usually ill-conditioned matrix of the linear system of equations. Moreover, in the case of an oversampled dataset or intended reduction, we want to reduce the given problem, i.e. reduce the number of weights and used basis functions, and preserve good preci-

sion of the approximated solution. The approach which includes the reduction is called the RBF approximation. In the following section, the method recently introduced in [Ska13] is described in detail. This approach requires less memory and offer higher speed of computation than the method using Lagrange multipliers [Fas07]. Further, a new approach to RBF approximation of large datasets is presented in the Section 3. These approach uses symmetry of matrix and partitioning matrix into blocks.

2. RBF Approximation

For simplicity, we assume that we have an unordered dataset $\{\mathbf{x}_i\}_1^N \in E^2$. However, this approach is generally applicable for d-dimensional space. Further, each point $\mathbf{x}_i$ from the dataset is associated with a vector $\mathbf{h}_i \in E^p$ of the given values, where p is the dimension of the vector, or scalar value, i.e. $h_i \in E^1$. For an explanation of the RBF approximation, let us consider the case when each point $\mathbf{x}_i$ is associated with a scalar value h_i, e.g. a $2^1/_2D$ surface. Let us introduce a set of new reference points $\{\boldsymbol{\xi}_j\}_1^M$, see Figure 1.

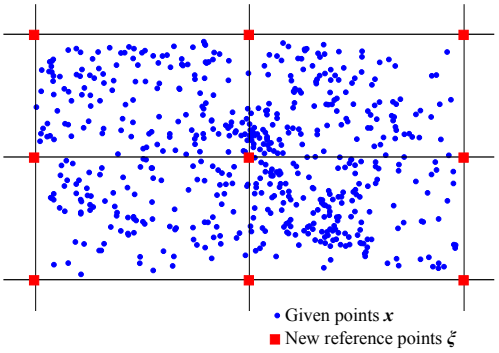


Figure 1: *The RBF approximation and reduction of points.*

These reference points may not necessarily be in a uniform grid. It is appropriate that their placement reflects the given surface (e.g. the terrain profile, etc.) as well as possible. The number of reference points $\boldsymbol{\xi}_j$ is M, where $M \ll N$. Now, the RBF approximation is based on the distance computation of the given point $\mathbf{x}_i$ and the reference point $\boldsymbol{\xi}_j$.

The approximated value is determined similarly as for interpolation (see [Ska15]):

$$f(\mathbf{x}) = \sum_{j=1}^{M} c_j \phi(r_j) = \sum_{j=1}^{M} c_j \phi(\|\mathbf{x} - \boldsymbol{\xi}_j\|), \tag{2}$$

where $\phi(r_j)$ is a used RBF centered at point $\boldsymbol{\xi}_j$ and the approximating function $f(\mathbf{x})$ is represented as a sum of these RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by a coefficient c_j which has to be determined.

It can be seen that we get an overdetermined linear system of equations for the given dataset:

$$\begin{aligned} h_i = f(\mathbf{x}_i) &= \sum_{j=1}^{M} c_j \phi(\|\mathbf{x}_i - \boldsymbol{\xi}_j\|) \\ &= \sum_{j=1}^{M} c_j \phi_{i,j} \qquad i = 1, \ldots, N. \end{aligned} \tag{3}$$

The linear system of equations (3) can be represented in a matrix form as:

$$\mathbf{A}\mathbf{c} = \mathbf{h}, \tag{4}$$

where the number of rows is $N \gg M$ and M is the number of unknown weights $[c_1, \ldots, c_M]^T$, i.e. the number of reference points. Equation (4) represents system of linear equations:

$$\begin{pmatrix} \phi_{1,1} & \cdots & \phi_{1,M} \\ \vdots & \ddots & \vdots \\ \phi_{i,1} & \cdots & \phi_{i,M} \\ \vdots & \ddots & \vdots \\ \phi_{N,1} & \cdots & \phi_{N,M} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_i \\ \vdots \\ h_N \end{pmatrix}. \tag{5}$$

The presented system is overdetermined, i.e. the number of equations N is higher than the number of variables M. This linear system of equations can be solved by the least squares method as $\mathbf{A}^T\mathbf{A}\mathbf{c} = \mathbf{A}^T\mathbf{h}$ or singular value decomposition, etc.

3. RBF Approximation for Large Data

In practice, the real datasets contain a large number of points which results into high memory requirements for storing the matrix $\mathbf{A}$ of the overdetermined linear system of equations (5). For example when we have dataset contains $3,000,000$ points, number of reference points is $10,000$ and double precision floating point is used then we need 223.5 GB memory for storing the matrix $\mathbf{A}$ of the overdetermined linear system of equations (5). Unfortunately, we do not have an unlimited capacity of RAM memory and therefore calculation of unknown weights c_j for RBF approximation would be prohibitively computationally expensive due to memory swapping, etc. In this section, a proposed solution to this problem is described.

In Section 2, it was introduced that overdetermined system of equations can be solved by the least squares method. For this method the $M \times M$ square matrix:

$$\mathbf{B} = \mathbf{A}^T\mathbf{A} \tag{6}$$

is to be determined. Advantages of matrix $\mathbf{B}$ are that it is a symmetric matrix and moreover only two vectors of length N are needed te determine of one entry, i.e.:

$$b_{ij} = \sum_{k=1}^{N} \phi_{ki} \cdot \phi_{kj}, \tag{7}$$

where b_{ij} is the entry of the matrix $\mathbf{B}$ in the $i-$th row and $j-$th column.

To save memory requirements and data bus (PCI) load block operations with matrices are used. Based on the above properties of the matrix $\mathbf{B}$, only the upper triangle of this matrix is computed. Moreover the matrix is partitioned into $M_B \times M_B$ blocks, see Figure 2, and the calculation is performed sequentially for each block:

$$\begin{gathered}\mathbf{B}_{kl} = (\mathbf{A}_{*,k})^T (\mathbf{A}_{*,l}) \\ k = 1,\ldots,\frac{M}{M_B}, \qquad l = k,\ldots,\frac{M}{M_B},\end{gathered} \tag{8}$$

where $\mathbf{B}_{kl}$ is sub-matrix in the $k-$th row and $l-$th column and $\mathbf{A}_{*,k}$ is defined as:

$$\mathbf{A}_{*,k} = \begin{pmatrix} \phi_{1,(k-1)\cdot M_B+1} & \cdots & \phi_{1,k\cdot M_B} \\ \vdots & \ddots & \vdots \\ \phi_{i,(k-1)\cdot M_B+1} & \cdots & \phi_{i,k\cdot M_B} \\ \vdots & \ddots & \vdots \\ \phi_{N,(k-1)\cdot M_B+1} & \cdots & \phi_{N,k\cdot M_B} \end{pmatrix}. \tag{9}$$

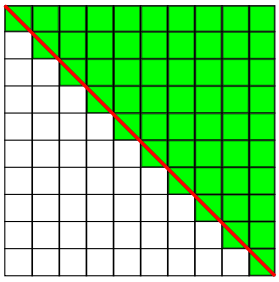


Figure 2: *$M \times M$ square matrix which is partitioned into $M_B \times M_B$ blocks. Main diagonal of matrix is represented by red color and illustrates the symmetry of matrix. Blocks, which must be computed, are represented by green color.*

The size of block M_B is chosen so that M_B is multiple of M and there is no swapping, i.e.:

$$M_B \cdot (M_B + 2 \cdot N) \cdot prec < \text{size of RAM [B]}, \tag{10}$$

where *prec* is size of data type in bytes.

4. Experimental results

The presented modification of the RBF approximation method has been tested on synthetic and real data. Let us introduce results for two real datasets.

The first dataset was obtained from LiDAR data of the Serpent Mound in Adams Country, Ohio[†]. The second dataset is LiDAR data of the Mount Saint Helens in Skamania Country, Washington[†]. Each point of these datasets is associated with its elevation. Summary of the dimensions of terrain for the given datasets is in Table 1.

Table 1: *Summary of the dimensions of terrain for tested datasets. Note that one feet [ft] corresponds to 0.3048 meter [m].*

Dimensions	Serpent Mound	St. Helens
number of points	3,265,110	6,743,176
lowest point [ft]	166.7800	3,191.5269
highest point [ft]	215.4800	8,330.2219
width [ft]	1,085.1199	26,232.3696
length [ft]	2,698.9601	35,992.6861

For experiments, two different radial basis functions have been used, see Table 2. Shape parameters α for used RBFs were determined experimentally with regard to the quality of approximation and they are presented in Table 3. Note that value of shape parameter α is inversely proportional to range of datasets.

Table 2: *Used RBFs*

RBF	type	$\phi(\boldsymbol{r})$
Gaussian RBF	global	$e^{-(\alpha r)^2}$
Wendland's $\phi_{3,1}$	local	$(1-\alpha r)_+^4(4\alpha r+1)$

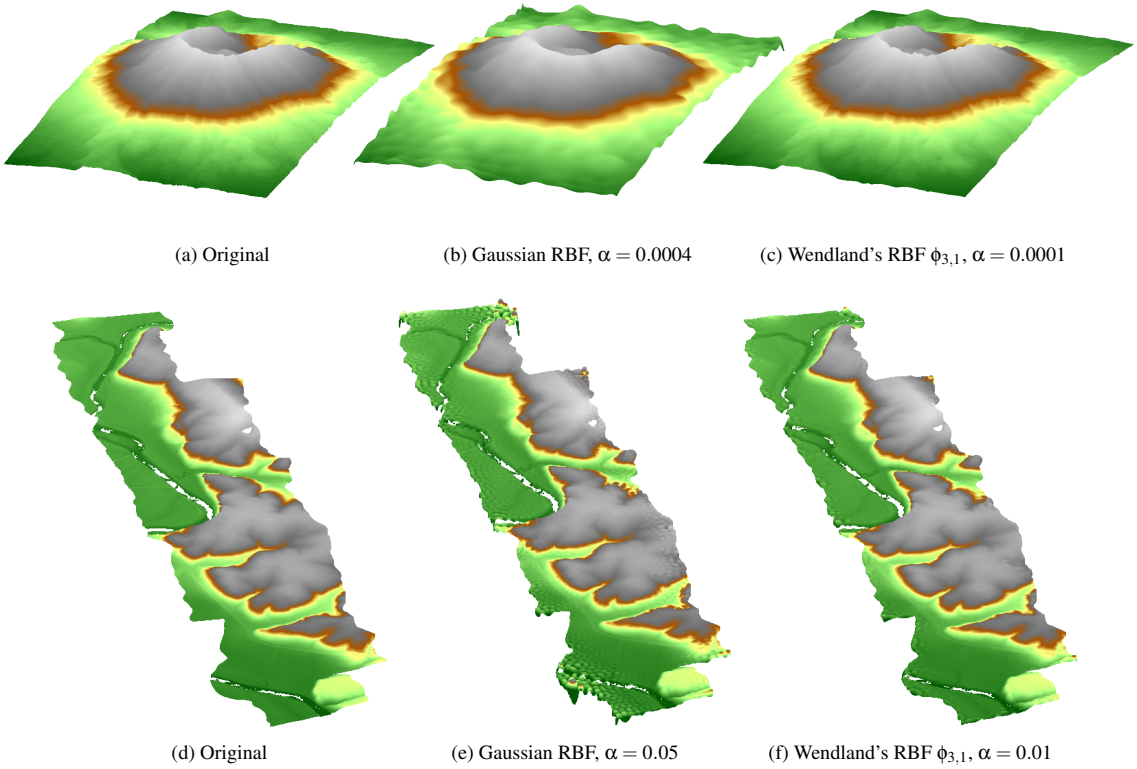
Table 3: *Experimentally determined shape parameters α for used RBFs*

RBF	shape parameter	
	Serpent Mound	**St. Helens**
Gaussian RBF	$\alpha = 0.05$	$\alpha = 0.0004$
Wendland's $\phi_{3,1}$	$\alpha = 0.01$	$\alpha = 0.0001$

The set of reference points equals the subset of the given dataset for which we determine the RBF approximation. Moreover, the distribution of reference points is uniform and the set of reference points has a cardinality 10,000 in both experiments.

Approximation of Mount Saint Helens for both BRFs and its original are shown in Figure 3a-3c. In Figure 3b can be seen that the RBF approximation with the global Gaussian RBFs cannot preserve the sharp rim of a crater. Further, visualization of magnitude of error at each point of the original points cloud is presented in Figure 4 and Figure 5. It can be seen that the RBF approximation with the global Gaussian

[†] http://www.liblas.org/samples/



(a) Original (b) Gaussian RBF, $\alpha = 0.0004$ (c) Wendland's RBF $\phi_{3,1}$, $\alpha = 0.0001$

(d) Original (e) Gaussian RBF, $\alpha = 0.05$ (f) Wendland's RBF $\phi_{3,1}$, $\alpha = 0.01$

Figure 3: *Serpent Mound in Adams Country, Ohio (top) and Mount Saint Helens is Skamania Country, Washington (bottom)*

RBFs returns worse result than RBF approximation with local Wendland's $\phi_{3,1}$ basis functions in terms of the error. In Table 4 can be seen the value of mean absolute error, its deviation and mean relative error for both approximations.

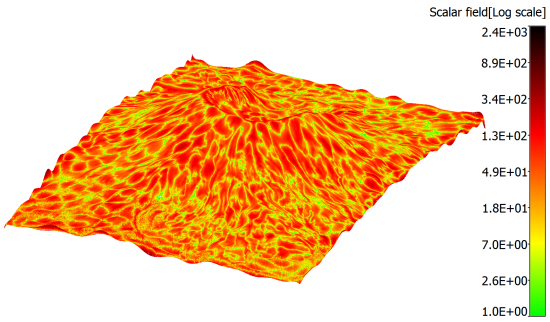


Figure 4: *Approximation of Mount Saint Helens with* $10,000$ *global Gaussian basis functions with shape parameter* $\alpha = 0.0004$ *false-colored by magnitude of error.*

Results of the RBF approximation for Serpent Mound and its original are shown in Figure 3d-3f. It can be seen that

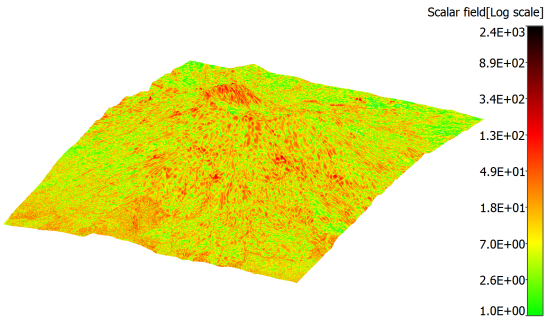


Figure 5: *Approximation of Mount Saint Helens with* $10,000$ *local Wendland's* $\phi_{3,1}$ *basis functions with shape parameter* $\alpha = 0.0001$ *false-colored by magnitude of error.*

the approximation using local Wendland's $\phi_{3,1}$ basis function (Figure 3f) returns again better result than approximation using the global Gaussian RBF (Figure 3e) in terms of the error. It is also seen in Figure 6 and Figure 7 where magnitude of error at each point of original points cloud is visualized. Moreover, we can see that the highest errors occur

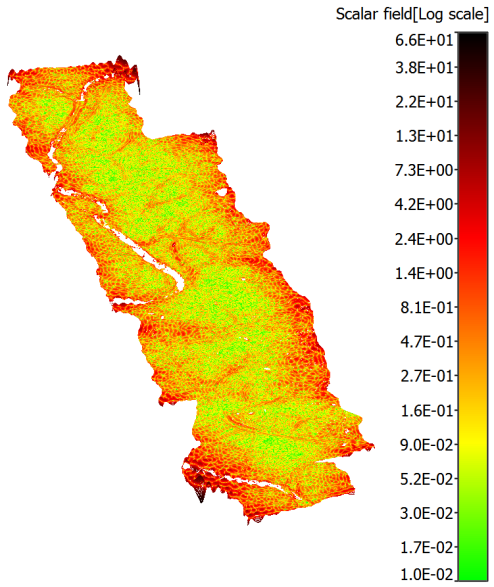


Figure 6: *Approximation of the Serpent Mound with* $10,000$ *global Gaussian basis functions with shape parameter* $\alpha = 0.05$ *false-colored by magnitude of error.*

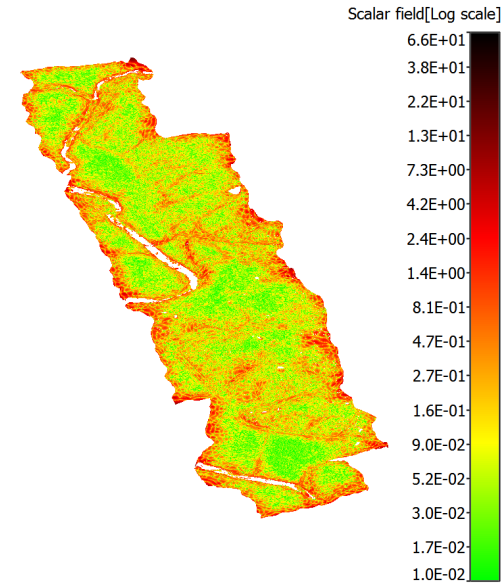


Figure 7: *Approximation of the Serpent Mound with* $10,000$ *local Wendland's* $\phi_{3,1}$ *basis functions with shape parameter* $\alpha = 0.01$ *false-colored by magnitude of error.*

on the boundary of terrain, which is a general problem of RBF methods. Value of mean absolute error, its deviation and mean relative error due to elevation for both used RBFs are again mentioned in Table 4.

Mutual comparison both datasets in terms of the mean relative error (Table 4) indicates that mean relative error for Serpent Mount is smaller than for Mount Saint Helens. It is caused by the presence of vegetation, namely forest, in LiDAR data of the Mount Saint Helens. This vegetation operates in our RBF approximation as noise and therefore the resulting mean relative error is higher.

The implementation of the RBF approximation has been performed in Matlab and tested on PC with the following configuration:

- CPU: Intel® Core™ i7-4770 (4× 3.40GHz + hyper-threading),
- memory: 32 GB RAM,
- operating system Microsoft Windows 7 64bits.

For the approximation of the Serpent Mound with $10,000$ local Wendland's $\phi_{3,1}$ basis function with shape parameter $\alpha = 0.01$ the running times for different sizes of blocks were measured. These times were converted relative to the time for 100×100 blocks and are presented in Figure 8. We can see that for the approximation matrix which is partitioned into small blocks (i.e. smaller than 25×25 blocks) the time performance is large. This is caused by overhead costs. On the other hand, for the approximation matrix which is partitioned into large blocks (i.e. larger than 125×125 blocks) the running time begins to grow above the permissible limit due to memory swapping.

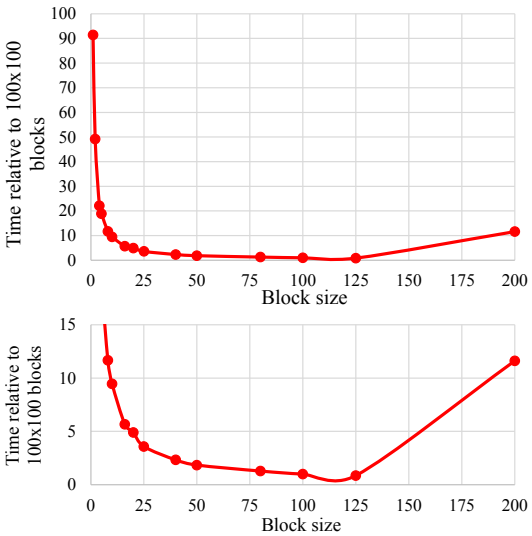


Figure 8: *Time performance for approximation of the Serpent Mound depending on the block size. The times are presented relative to the time for* 100×100 *blocks.*

Table 4: *The RBF approximation error for testing datasets and different radial basis functions. Note that one feet [ft] corresponds to 0.3048 meter [m].*

Error	Serpent Mound		St. Helens	
	Gaussian RBF	Wendland's $\phi_{3,1}$	Gaussian RBF	Wendland's $\phi_{3,1}$
mean absolute error [ft]	0.4477	0.2289	44.4956	12.1834
deviation of error [ft]	1.4670	0.1943	680.3659	169.2800
mean relative error [%]	0.0024	0.0012	0.0087	0.0023

5. Conclusions

This paper presents a new approach to the RBF approximation of large datasets. The proposed approach uses symmetry of matrix and partitioning matrix into blocks, thus preventing memory swapping. The experiments made proved that the proposed approach is able to determine the RBF approximation for large dataset. Moreover, from the experimental results we can see that use of a local RBFs is better than global RBFs, if data are sufficiently sampled. Futher, it is obvious that approximation using the global Gaussian RBFs has problems with the preservation of sharp edges. The experiments made also proved that RBF methods have problems with the accuracy of calculation on the boundary of an object, which is a well known property, and the magnitude of the RBF approximation error is influenced by the presence of a noise.

For the future work, the RBF approximation method can be explored in terms of lower sensitivity to noise, more accurate calculation on the boundary or better approximation of sharp edges and improvements of the computational cost without loss of approximation accuracy.

Acknowledgments

The authors would like to thank their colleagues at the University of West Bohemia, Plzen, for their discussions and suggestions, and also anonymous reviewers for the valuable comments and suggestions they provided. The research was supported by MSMT CR projects LH12181 and SGS 2016-013.

References

[CBC*01] CARR J. C., BEATSON R. K., CHERRIE J. B., MITCHELL T. J., FRIGHT W. R., MCCALLUM B. C., EVANS T. R.: Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, Los Angeles, California, USA, August 12-17, 2001* (2001), pp. 67–76. 1

[Dar00] DARVE E.: The fast multipole method: Numerical implementation. *Journal of Computational Physics 160*, 1 (2000), 195–240. 1

[Fas07] FASSHAUER G. E.: *Meshfree Approximation Methods with MATLAB*, vol. 6. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2007. 2

[Har71] HARDY R. L.: Multiquadratic Equations of Topography and Other Irregular Surfaces. *Journal of Geophysical Research 76* (1971), 1905–1915. 1

[HSfY15] HON Y.-C., SARLER B., FANG YUN D.: Local radial basis function collocation method for solving thermo-driven fluid-flow problems with free surface. *Engineering Analysis with Boundary Elements 57* (2015), 2 – 8. {RBF} Collocation Methods. 1

[LCC13] LI M., CHEN W., CHEN C.: The localized {RBFs} collocation methods for solving high dimensional {PDEs}. *Engineering Analysis with Boundary Elements 37*, 10 (2013), 1300 – 1304. 1

[PRF14] PEPPER D. W., RASMUSSEN C., FYDA D.: A meshless method using global radial basis functions for creating 3-d wind fields from sparse meteorological data. *Computer Assisted Methods in Engineering and Science 21*, 3-4 (2014), 233–243. 1

[PS11] PAN R., SKALA V.: A two-level approach to implicit surface modeling with compactly supported radial basis functions. *Eng. Comput. (Lond.) 27*, 3 (2011), 299–307. 1

[Ska13] SKALA V.: Fast Interpolation and Approximation of Scattered Multidimensional and Dynamic Data Using Radial Basis Functions. *WSEAS Transactions on Mathematics 12*, 5 (2013), 501–511. 2

[Ska15] SKALA V.: Meshless interpolations for computer graphics, visualization and games. In *Eurographics 2015 - Tutorials, Zurich, Switzerland, May 4-8, 2015* (2015), Zwicker M., Soler C., (Eds.), Eurographics Association. 1, 2

[SPN13] SKALA V., PAN R., NEDVED O.: Simple 3d surface reconstruction using flatbed scanner and 3d print. In *SIGGRAPH Asia 2013, Hong Kong, China, November 19-22, 2013, Poster Proceedings* (2013), ACM, p. 7. 1

[SPN14] SKALA V., PAN R., NEDVED O.: Making 3d replicas using a flatbed scanner and a 3d printer. In *Computational Science and Its Applications - ICCSA 2014 - 14th International Conference, Guimarães, Portugal, June 30 - July 3, 2014, Proceedings, Part VI* (2014), vol. 8584 of *Lecture Notes in Computer Science*, Springer, pp. 76–86. 1

[TO02] TURK G., O'BRIEN J. F.: Modelling with implicit surfaces that interpolate. *ACM Trans. Graph. 21*, 4 (2002), 855–873. 1

[Wen06] WENDLAND H.: Computational aspects of radial basis function approximation. *Studies in Computational Mathematics 12* (2006), 231–256. 1

Appendix E

Big geo data surface approximation using radial basis functions: A comparative study

Majdišová, Z., Skala, V.

Computers & Geosciences, Volume 109, Issue 1, pp. 51-58, Elsevier (2017), ISSN 0098-3004, IF 2.567

Computers and Geosciences 109 (2017) 51–58

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

Big geo data surface approximation using radial basis functions: A comparative study

Zuzana Majdisova *, Vaclav Skala

Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, CZ 30614 Plzeň, Czech Republic

ARTICLE INFO

Keywords:
Radial basis functions
CS-RBF
Approximation
Wendland's RBF
Big data
Point clouds

ABSTRACT

Approximation of scattered data is often a task in many engineering problems. The Radial Basis Function (RBF) approximation is appropriate for big scattered datasets in *n*–dimensional space. It is a non-separable approximation, as it is based on the distance between two points. This method leads to the solution of an overdetermined linear system of equations.

In this paper the RBF approximation methods are briefly described, a new approach to the RBF approximation of big datasets is presented, and a comparison for different Compactly Supported RBFs (CS-RBFs) is made with respect to the accuracy of the computation. The proposed approach uses symmetry of a matrix, partitioning the matrix into blocks and data structures for storage of the sparse matrix. The experiments are performed for synthetic and real datasets.

1. Introduction

Interpolation and approximation are the most frequent operations used in computational techniques. Several techniques have been developed for data interpolation or approximation, but they usually require an ordered dataset, e.g. rectangular mesh, structured mesh, unstructured mesh, etc. However, in many engineering problems, data are not ordered and they are scattered in *n*–dimensional space, in general. Usually, in technical applications the conversion of a scattered dataset to a semi-regular grid is performed using some tessellation techniques. However, this approach is quite prohibitive for the case of *n*–dimensional data due to the computational cost.

Interesting techniques are based on the Radial Basis Function (RBF) method, which was originally introduced by Hardy (1971, 1990). A good introduction to RBFs is given by Buhmann (2003). RBF techniques are widely used across many fields solving technical and non-technical problems, e.g. surface reconstruction (Carr et al. (2001), Turk and O'Brien (2002)), data visualization (Pepper et al. (2014)) and pattern recognition. It is an effective tool for solving partial differential equations (Hon et al. (2015), Li et al. (2013)). The RBF techniques are really meshless and are based on collocation in a set of scattered nodes. These methods are independent with respect to the dimension of the space. The computational cost of the RBF approximation increases nonlinearly (almost cubic) with the number of points in the given dataset and linearly with the dimensionality of the data. Of course, there are other meshless techniques such as discrete smooth interpolation (DSI) (Mallet (1989)), kriging (Royer and Vieira (1984), Ma et al. (2014), Cressie (2015)), which is based on statistical models that include autocorrelation, etc.

The radial basis functions are divided into two main groups of basis functions: global RBFs and Compactly Supported RBFs (CS–RBFs) (Wendland (2006)). In this paper, we will mainly focus on CS-RBFs. Fitting scattered data with CS–RBFs leads to a simpler and faster computation, because the system of linear equations has a sparse matrix. However, an approximation using CS–RBFs is sensitive to the density of the approximated scattered data and to the choice of a shape parameter. Global RBFs are useful in repairing incomplete datasets and they are insensitive to the density of scattered data. However, global RBFs lead to a linear system of equations with a dense matrix and therefore they have high computational and memory costs. Typical global RBFs are Gauss $\phi(r) = e^{-(\alpha r)^2}$, inverse quadratic $(1 + (\alpha r)^2)^{-1}$ and inverse multiquadric $(1 + (\alpha r)^2)^{-1/2}$, where α is shape parameter which defines behavior of function. These RBFs are monotonically decreased with increasing radius r, strictly positive definite, infinitely differentiable and convergent to zero. Other global RBF is multiquadric $\sqrt{1 + (\alpha r)^2}$ which is monotonically increased with increasing radius r, infinitely differentiable and

* Corresponding author.
E-mail address: majdisz@kiv.zcu.cz (Z. Majdisova).
URL: http://www.vaclavskala.eu

http://dx.doi.org/10.1016/j.cageo.2017.08.007
Received 21 February 2017; Received in revised form 31 July 2017; Accepted 7 August 2017
Available online 10 August 2017

divergent as radius increases. The last popular global RBF is thin plate spline (TPS) $r^2\log(r)$ which is shape parameter free and divergent as radius increases. TPS has a singularity at the origin which is removable for the function and its first derivative but this singularity is not removable for the second derivative of TPS.

For the processing of scattered data we can use the RBF interpolation or the RBF approximation. The unknown function sampled at given points $\{\boldsymbol{x}_i\}_1^N$ by values $\{h_i\}_1^N$ can be determined using the RBF interpolation, e.g. presented by Skala (2015), as:

$$f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j \phi(r_j) = \sum_{j=1}^{N} c_j \phi(\|\boldsymbol{x} - \boldsymbol{x}_j\|), \tag{1}$$

where the interpolating function $f(\boldsymbol{x})$ is represented as a sum of N RBFs, each centered at a different data point $\boldsymbol{x}_j$ and weighted by an appropriate weight c_j which has to be determined. This leads to a solution of linear system of equations:

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{h}, \tag{2}$$

where the matrix $\boldsymbol{A} = \{A_{ij}\} = \{\phi(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|)\}$ is $N \times N$ symmetric square matrix, the vector $\boldsymbol{c} = (c_1, \ldots, c_N)^T$ is the vector of unknown weights and $\boldsymbol{h} = (h_1, \ldots, h_N)^T$ is a vector of values in the given points. The disadvantage of RBF interpolation is the large and usually ill-conditioned matrix of the linear system of equations. Note that the one of the possible solution of ill-condition problems based on modified orthogonal least squares is described in Chen and Li (2012). Moreover, in the case of an oversampled dataset or intended reduction, we want to reduce the given problem, i.e. reduce the number of weights and used basis functions, and preserve good precision of the approximated solution. The approach which includes such a reduction is called the RBF approximation. In the following section, the approach recently introduced in Skala (2013) will be described in detail. This approach requires less memory and offers higher speed of computation than the method using Lagrange multipliers (Fasshauer (2007)). Further, a new approach to RBF approximation of large datasets is presented in Section 5. This approach uses symmetry of a matrix, partitioning the matrix into blocks and data structures for storage of the sparse matrix (see Section 4).

2. RBF approximation

For simplicity, we assume that we have an unordered dataset $\{\boldsymbol{x}_i\}_1^N \in E^2$. However, this approach is generally applicable for n-dimensional space. Further, each point $\boldsymbol{x}_i$ from the dataset is associated with a vector $\boldsymbol{h}_i \in E^p$ of the given values, where p is the dimension of the vector, or a scalar value, i.e. $h_i \in E^1$. For an explanation of the RBF approximation, let us consider the case when each point $\boldsymbol{x}_i$ is associated with a scalar value h_i, e.g. a $21/2D$ surface. Let us introduce a set of new reference points (knots of RBF) $\{\boldsymbol{\xi}_j\}_1^M$, see Fig. 1.

These reference points may not necessarily be in a uniform grid. A good placement of the reference points improves the approximation of the underlying data. For example, when a terrain is approximated, placement along features such as break lines leads to better approximation results. The number of reference points $\boldsymbol{\xi}_j$ is M, where $M \ll N$. The RBF approximation is based on the distance computation between the given point $\boldsymbol{x}_i$ and the reference point $\boldsymbol{\xi}_j$.

The approximated value is determined as (see Skala (2013)):

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \phi(r_j) = \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|), \tag{3}$$

where $\phi(r_j)$ is an RBF centered at point $\boldsymbol{\xi}_j$ and the approximating function $f(\boldsymbol{x})$ is represented as a sum of these RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by a coefficient c_j which has to be determined.

When inserting all data points $\boldsymbol{x}_i$, with $i = 1, \ldots, N$, into (3), we get an overdetermined linear system of equations.

$$h_i = f(\boldsymbol{x}_i) = \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|) = \sum_{j=1}^{M} c_j \phi_{i,j} \quad i = 1, \ldots, N \tag{4}$$

The linear system of equation (4) can be represented in a matrix form as:

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{h}, \tag{5}$$

where $A_{ij} = \phi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|)$ is the entry of the matrix in the i-th row and j-th column, the number of rows is $N \gg M$, M is the number of unknown weights $\boldsymbol{c} = (c_1, \ldots, c_M)^T$, i.e. a number of reference points, and $\boldsymbol{h} = (h_1, \ldots, h_N)^T$ is a vector of values in the given points. The presented system is overdetermined, i.e. the number of equations N is higher than the number of variables M. This linear system of equations can be solved by the least squares method (LSE) as $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{h}$.

3. RBF approximation with polynomial reproduction

The method which was described in Section 2 can have problems with stability and solvability. Therefore, the RBF approximant (3) is usually extended by a polynomial function $P_k(\boldsymbol{x})$ of the degree k. This approach was introduced in Majdisova and Skala (2016).

The approximated value $f(\boldsymbol{x})$ is determined as:

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|) + P_k(\boldsymbol{x}), \tag{6}$$

where $\boldsymbol{\xi}_j$ are reference points specified by a user. The approximating function $f(\boldsymbol{x})$ is represented as a sum of M RBFs, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by an appropriate coefficient c_j, and $P_k(\boldsymbol{x})$ is a polynomial function of degree k. It should be noted that the polynomial function affects only global behavior of the approximated dataset. In practice, a linear polynomial $P_1(\boldsymbol{x})$:

$$P_1(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + a_0 \tag{7}$$

is used (e.g. $P_1(\boldsymbol{x}) = a_1x + a_2y + a_0$ for $\boldsymbol{x} \in E^2$). Geometrically, the coefficient a_0 determines the "vertical" placement of the hyperplane and the expression $\boldsymbol{a}^T\boldsymbol{x}$ represents the inclination of the hyperplane.

Thus, the following overdetermined linear system of equations is obtained:

$$\begin{aligned} h_i = f(\boldsymbol{x}_i) &= \sum_{j=1}^{M} c_j \phi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|) + \boldsymbol{a}^T\boldsymbol{x} + a_0 \\ &= \sum_{j=1}^{M} c_j \phi_{i,j} + \boldsymbol{a}^T\boldsymbol{x} + a_0 \quad i = 1, \ldots, N. \end{aligned} \tag{8}$$

The linear system of equation (8) can be represented in a matrix form as:

$$\boldsymbol{A}\boldsymbol{c} + \boldsymbol{P}\boldsymbol{k} = \boldsymbol{h}, \tag{9}$$

where $A_{ij} = \phi(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|)$ is the entry of the matrix in the i-th row and j-th column, $\boldsymbol{c} = (c_1, \ldots, c_M)^T$ is the vector of unknown weights, $\boldsymbol{P}_i = (\boldsymbol{x}_i^T, 1)$ is the vector of basis functions of linear polynomial at point $\boldsymbol{x}_i$, $\boldsymbol{k} = (\boldsymbol{a}^T, a_0)^T$ is the vector of the coefficient for the linear polynomial and $\boldsymbol{h} = (h_1, \ldots, h_N)^T$ is the vector of values in the given points. The presented linear system of equations can be solved by the minimization of the square of error, which leads to a system of linear equations:

$$\begin{pmatrix} \boldsymbol{A}^T\boldsymbol{A} & \boldsymbol{A}^T\boldsymbol{P} \\ \boldsymbol{P}^T\boldsymbol{A} & \boldsymbol{P}^T\boldsymbol{P} \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{k} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}^T\boldsymbol{h} \\ \boldsymbol{P}^T\boldsymbol{h} \end{pmatrix}. \tag{10}$$

Finally, it should be noted that the polynomial of degree $k>1$ can be used in general. However, in this case, it is necessary be careful because the polynomial of higher degree in combination with a large range of data might cause numerical problems. This is due to the fact that the elements of sub-matrix $\boldsymbol{P}^T\boldsymbol{P}$ in relation (10) contain much larger values than elements of sub-matrix $\boldsymbol{A}^T\boldsymbol{A}$ in the same relation.

4. Data structures for storage of the sparse matrix

If the CS-RBFs are used, the matrix of the linear system of equations is sparse. Therefore, the most important part of each approximation using CS-RBFs is a data structure used to store the approximation matrix. There are a number of existing sparse matrix representations, e.g. Bell and Garland (2009), Šimecek (2009), each with different computational characteristics, storage requirements and methods of accessing and manipulating entries of the matrix. The main difference among existing storage formats is the sparsity pattern, or the structure of the nonzero elements, for which they are best suited. For our purpose, the coordinate format is used, which is briefly described in the following.

The coordinate (COO) format is the simplest storage scheme. The sparse matrix is represented by three arrays: `data`, where the N_{NZ} nonzero values are stored, `row`, where the row index of each nonzero element is kept, and `col`, where the column indices of the nonzero values are stored.

Example of the COO format for matrix $\boldsymbol{Q}$:

$$\boldsymbol{Q} = \begin{pmatrix} 1 & 0 & 6 & 0 & 0 \\ 9 & 2 & 0 & 7 & 0 \\ 0 & 1 & 3 & 0 & 8 \\ 4 & 0 & 2 & 4 & 0 \\ 0 & 5 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} \texttt{row} &= [0\ 0\ 1\ 1\ 1\ 2\ 2\ 2\ 3\ 3\ 3\ 4] \\ \texttt{col} &= [0\ 2\ 0\ 1\ 3\ 1\ 2\ 4\ 0\ 2\ 3\ 1] \\ \texttt{data} &= [1\ 6\ 9\ 2\ 7\ 1\ 3\ 8\ 4\ 2\ 4\ 5] \end{aligned}$$

So, if the COO format is used for representation of matrix $\boldsymbol{Q}$ (in form as described above) and the equation $\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{x}$, where $\boldsymbol{x}$ is vector of the given values, has been solved, the following pseudocode is used for calculation:

$\forall i = 0, \ldots, N : \boldsymbol{y}_{[i]} = 0$

for $i = 0, \ldots, N_{NZ} - 1$ **do**

$\quad \mathbf{y}_{\mathbf{row[i]}} = \mathbf{y}_{\mathbf{row[i]}} + \mathbf{data}_{\mathbf{[i]}} \cdot \mathbf{x}_{\mathbf{col[i]}}$

Note that vector of given values has form $\boldsymbol{x} = [x_0, x_1, \ldots, x_M]$, where M is number of columns of matrix $\boldsymbol{Q}$, and the resulting vector is $\boldsymbol{y} = [y_0, y_1, \ldots, y_N]$, where N is number of rows of matrix $\boldsymbol{Q}$.

The benefit of the COO format is its generality, i.e. an arbitrary sparse matrix can be represented by the COO format and the required storage is always proportional to the number of nonzero values.

The disadvantage of the COO format is that both row and column indices are stored explicitly, which reduces the efficiency of memory transactions (e.g. read operations).

5. RBF approximation for large data

In practice, real datasets contain a large number of points, which results in high memory requirements for storing the matrix $\boldsymbol{A}$ of the overdetermined linear system of equation (5). Unfortunately, we do not have an unlimited capacity of RAM memory; therefore, calculation of unknown weights c_j for RBF approximation would be prohibitively computationally expensive due to memory swapping, etc. In this section, a proposed solution to this problem is described.

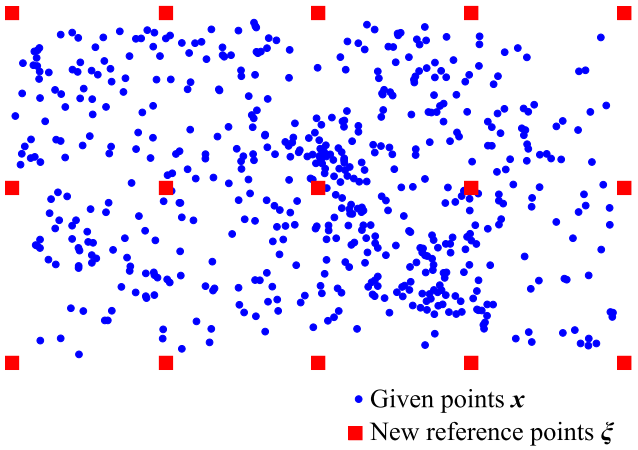


Fig. 1. The RBF approximation and reduction of points. Note that the reference points (knots) can be distributed arbitrarily.

In Section 2, it was mentioned that an overdetermined system of equations can be solved by the least squares method. For this method the square $M \times M$ matrix:

$$\boldsymbol{B} = \boldsymbol{A}^T\boldsymbol{A} \tag{11}$$

is to be determined. Advantages for computation of the matrix $\boldsymbol{B}$ are that it is a symmetric matrix and, moreover, only two vectors of length N are needed for determination of one entry, i.e.:

$$b_{ij} = \sum_{k=1}^{N} \phi_{ki} \cdot \phi_{kj}, \tag{12}$$

where b_{ij} is the entry of the matrix $\boldsymbol{B}$ in the $i-$th row and $j-$th column.

To save memory requirements and to prevent data bus (PCI) overloading, block operations with matrices are used. Based on the above properties of the matrix $\boldsymbol{B}$, only the upper triangle of this matrix is computed. Moreover, the matrix $\boldsymbol{B}$ is partitioned into $M_B \times M_B$ blocks, see Fig. 2, and the calculation is performed sequentially for each block:

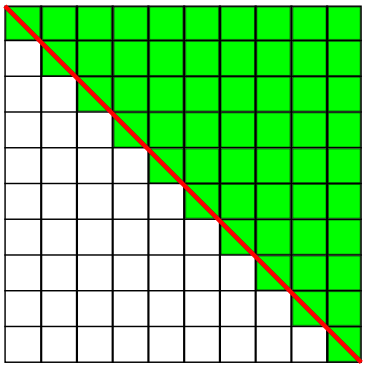


Fig. 2. $M \times M$ square matrix which is partitioned into $M_B \times M_B$ blocks. The color red is used to denote the main diagonal of the matrix and illustrates the symmetry of the matrix. The color green is used to denote the blocks which must be computed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\boldsymbol{B}_{kl} = (\boldsymbol{A}_{*,k})^T(\boldsymbol{A}_{*,l})$$
$$k = 1, \ldots \left\lceil \frac{M}{M_B} \right\rceil, \quad l = k, \ldots, \left\lceil \frac{M}{M_B} \right\rceil, \tag{13}$$

where $\boldsymbol{B}_{kl}$ is a sub-matrix in the $k-$th row and $l-$th column, the index $*$ denotes that the sub-matrix $\boldsymbol{A}_{*,k}$ contains all values in the appropriate block of columns (given by the index k) of the original matrix $\boldsymbol{A}$, i.e. $\boldsymbol{A}_{*,k}$ is defined as:

$$\boldsymbol{A}_{*,k} = \begin{pmatrix} \phi_{1,(k-1)\cdot M_B+1} & \cdots & \phi_{1,\min\{k\cdot M_B,M\}} \\ \vdots & \ddots & \vdots \\ \phi_{i,(k-1)\cdot M_B+1} & \cdots & \phi_{i,\min\{k\cdot M_B,M\}} \\ \vdots & \ddots & \vdots \\ \phi_{N,(k-1)\cdot M_B+1} & \cdots & \phi_{N,\min\{k\cdot M_B,M\}} \end{pmatrix}, \tag{14}$$

where the size of this matrix is $N \times M_B$ except of the last block and the index k denotes the $k-$th block of columns. This enables the computation of big datasets on hardware systems with limited main memory.

The size of block M_B is chosen so that swapping of memory (RAM) does not occur during the computation, i.e.:

$$\left(M^2 + 2\cdot M_B \cdot N\right)\cdot prec \; < \; \text{size of RAM [B]}, \tag{15}$$

where *prec* is the size of the data type in bytes. Note that this relation is valid when the matrix $\boldsymbol{A}$ of the overdetermined linear system of equations is dense. If CS-RBFs are used for RBF approximation and the matrix $\boldsymbol{A}$ of the overdetermined linear system of the equation is stored using special data structures, see Section 4, then the optimal size of block M_B is much larger than given in relation (15). For this case, the optimal size of block M_B should satisfy:

$$\left(M^2 + 2\cdot N_{NZ}\right)\cdot prec \; < \; \text{size of RAM [B]}, \tag{16}$$

where N_{NZ} is the maximum number of non-zero elements in sub-matrices $\boldsymbol{A}_{*,k}, k = 1, \ldots, \left\lceil \frac{M}{M_B} \right\rceil$. Naturally, it is obvious that the size of the block should be selected as the largest possible value satisfying (16).

Moreover, note that the elements in sub-matrices $\boldsymbol{A}_{*,k}$ are zero for far away points, when CS-RBFs are used. Therefore, we do not want to compute the elements for all pairs of points, so the *kd*-tree (A.2 in Fasshauer (2007)) is used for computing the sub-matrices $\boldsymbol{A}_{*,k}$. Algorithm for determination of the sparse sub-matrix $\boldsymbol{A}_{*,k}$ is described in Algorithm 1.

Algorithm 1 Determination of the sub-matrix $\boldsymbol{A}_{*,k}$ when CS-RBFs are used. Note that the order of the elements in the triplet (row 5) is {row index, col index, value}

Input: given points $\{\boldsymbol{x}_i\}_1^N$, reference points $\{\boldsymbol{\xi}_i\}_{(k-1)\cdot M_B}^{\min\{k\cdot M_B,M\}}$, shape parameter α, CS-RBF ϕ

Output: sub-matrix $\boldsymbol{A}_{*,k}$ in COO format, i.e. return three arrays `row`, `col`, `data`

1: Build a *kd*-tree for the given points $\{\boldsymbol{x}_i\}_1^N$
2: **for each** reference point $\boldsymbol{\xi}_j$ **do**
3: Query the *kd*-tree for points $\{\boldsymbol{x}_q\}$ such that $\|\boldsymbol{x}_q - \boldsymbol{\xi}_j\| < \frac{1}{\alpha}$
4: **for each** point in a support radius $\boldsymbol{x}_q$ **do**
5: Add triplet $\{q, j, \phi(\|\boldsymbol{x}_q - \boldsymbol{\xi}_j\|)\}$ to COO format

In general, the mentioned approach could be used in combination with massive parallel computing on GPU, but the calculation would have to be done in single precision to exploit the full potential of GPU. However, in this case, problems with numerical stability and solvability of the RBF approximation can be expected.

Finally, note that it is possible to modify this approach easily for the RBF approximation with a polynomial reproduction, see Section 3.

6. Experimental results

The presented RBF approximation method was tested on synthetic and real data. The implementation was performed in Matlab. Experimental results for one synthetic and two real datasets follow.

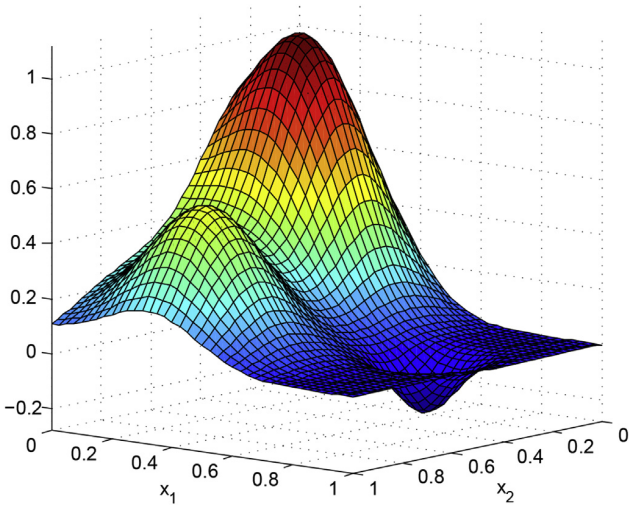


Fig. 3. Franke's function defined as (17).

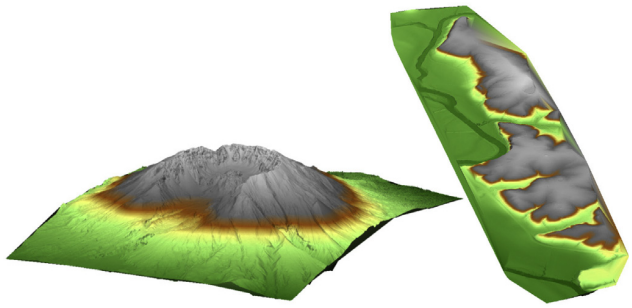


Fig. 4. Original datasets: Mount Saint Helens (left); Serpent Mound (right).

Table 1
Overview information for the tested datasets. The Axis-Aligned Bounding Boxes (AABBs) of the tested datasets have a size width × length × relief, i.e. $x_{range} \times y_{range} \times z_{range}$. Note that one foot [ft] corresponds to 0.3048 m [m].

	Synth.	Serpent Mound	St. Helens
Number of pts.	1089	3,265,110	6,743,176
Number of ref. pts.	81	10,000	10,000
Relief [ft]	1.238	48.70	5138.69
Width [ft]	1.000	1,085.12	26,232.37
Length [ft]	1.000	2,698.96	35,992.69

Table 2
Used Wendland's CS-RBFs $\phi_{d,s}$. Wendland's functions are univariate polynomial of degree $\lfloor d/2 \rfloor + 3s + 1$, they are always positive definite up to a maximal space dimension d and their smoothness is C^{2s}. For more details see Chapter 11.2 in Fasshauer (2007).

CS-RBF	$\phi(r)$
$\phi_{3,0}$	$(1-\alpha r)_+^2$
$\phi_{3,1}$	$(1-\alpha r)_+^4(4\alpha r+1)$
$\phi_{3,3}$	$(1-\alpha r)_+^8(32(\alpha r)^3+25(\alpha r)^2+8\alpha r+1)$

Table 3
Experimentally determined shape parameters α for the used CS-RBFs.

CS-RBF	Shape parameter		
	Synthetic	Serpent Mound	St. Helens
Wendland's $\phi_{3,0}$	$\alpha = 0.707$	$\alpha = 0.01$	$\alpha = 0.0005$
Wendland's $\phi_{3,1}$	$\alpha = 0.500$	$\alpha = 0.01$	$\alpha = 0.0007$
Wendland's $\phi_{3,3}$	$\alpha = 0.250$	$\alpha = 0.01$	$\alpha = 0.0005$

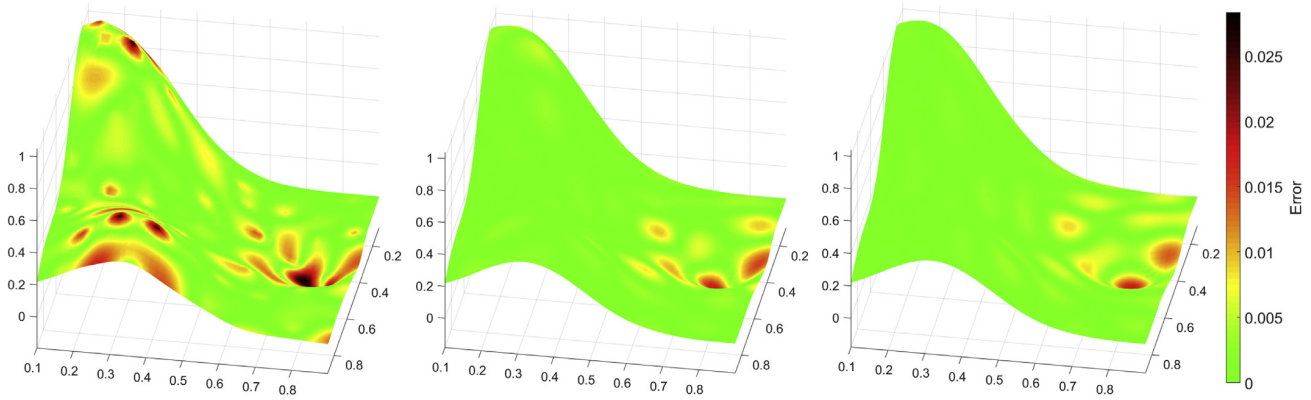


Fig. 5. Results for synthetic dataset false-colored by magnitude of absolute error: Wendland's RBF $\phi_{3,0}$, $\alpha = 0.707$ (left); Wendland's RBF $\phi_{3,1}$, $\alpha = 0.500$ (center) and Wendland's RBF $\phi_{3,3}$, $\alpha = 0.250$ (right).

The synthetic dataset has a Halton distribution (A.1 in Fasshauer (2007)) of points and each point is associated with a value from Franke's function (Franke (1979)):

$$
\begin{aligned}
f(\boldsymbol{x}) &= f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + f_3(\boldsymbol{x}) - f_4(\boldsymbol{x}),\\
f_1(\boldsymbol{x}) &= 0.75\cdot\exp\left(-\frac{(9x_1-2)^2}{4} - \frac{(9x_2-2)^2}{4}\right),\\
f_2(\boldsymbol{x}) &= 0.75\cdot\exp\left(-\frac{(9x_1+1)^2}{49} - \frac{(9x_2+1)^2}{10}\right),\\
f_3(\boldsymbol{x}) &= 0.50\cdot\exp\left(-\frac{(9x_1-7)^2}{4} - \frac{(9x_2-3)^2}{4}\right),\\
f_4(\boldsymbol{x}) &= 0.20\cdot\exp\left(-(9x_1-4)^2 - (9x_2-7)^2\right),
\end{aligned}
\tag{17}
$$

where $\boldsymbol{x} = (x_1, x_2)$ is a point for which the associated value has been computed. This function is shown in Fig. 3.

The first real dataset was obtained from LiDAR data of Mount Saint Helens in Skamania County, Washington,[1] see Fig. 4 (left). The second real dataset is LiDAR data of the Serpent Mound in Adams County, Ohio[21], see Fig. 4 (right).

Each point of these datasets is associated with its elevation. Moreover, as a first step, the real datasets are translated so that their estimated center of gravity corresponds to the origin of the coordinate system. This step is used due to the limitation of the influence of dataset placement in space. The set of reference points is a subset of the given dataset, for which we determine the RBF approximation. In addition, reference points are uniformly distributed within a given area. Table 1 gives an overview of the used datasets.

Because the global RBFs affect the entire domain of given datasets, which is usually undesirable behavior, the CS-RBFs have been used for the presented experiments. All CS-RBFs from the catalog of RBFs in Fasshauer (2007) (see D.2.7) have been used for the experiments. Depending on the quality, the obtained results are divided into three groups. The results are presented for a representative of each group, see Table 2.

Note that the notation $(1-\alpha r)_+^q$ means:

$$
(1-\alpha r)_+^q = \begin{cases} (1-\alpha r)^q & \text{if } 0 \le \alpha r \le 1 \\ 0 & \text{if } \alpha r > 1 \end{cases},
\tag{18}
$$

where r is the variable which denotes the distance of the given point from the appropriate reference point and α is a shape parameter. The shape parameters α for the used CS-RBFs were determined experimentally with regard to the quality of approximation and they are presented in Table 3. Some papers have also been published on choosing the optimal shape parameter α, e.g. Franke (1982), Rippa (1999), Fasshauer and Zhang (2007), Scheuerer (2011). Note that the value of the shape parameter α is inversely proportional to the width, length, and number of points of the datasets.

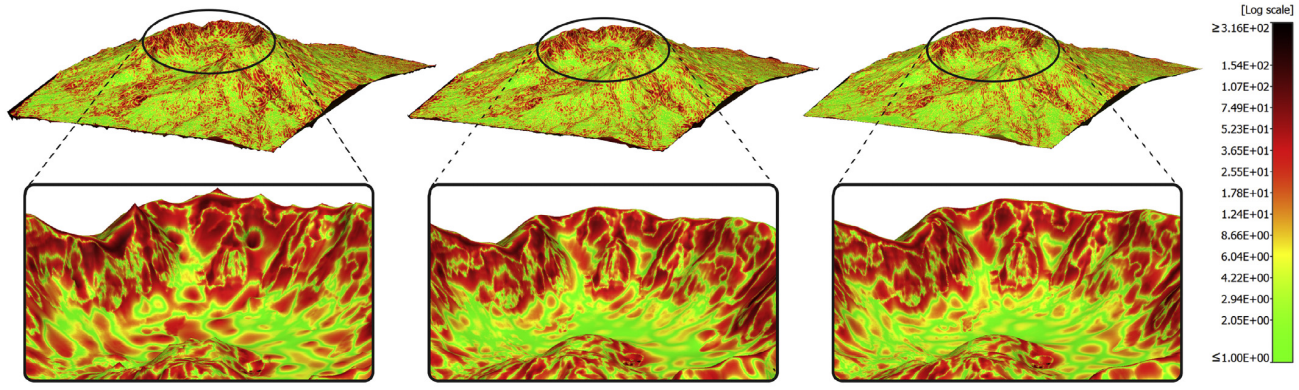
Fig. 5 presents the approximations of the synthetic dataset without polynomial reproduction for all CS−RBFs.

In this figure, the surfaces are false-colored by the magnitude of the

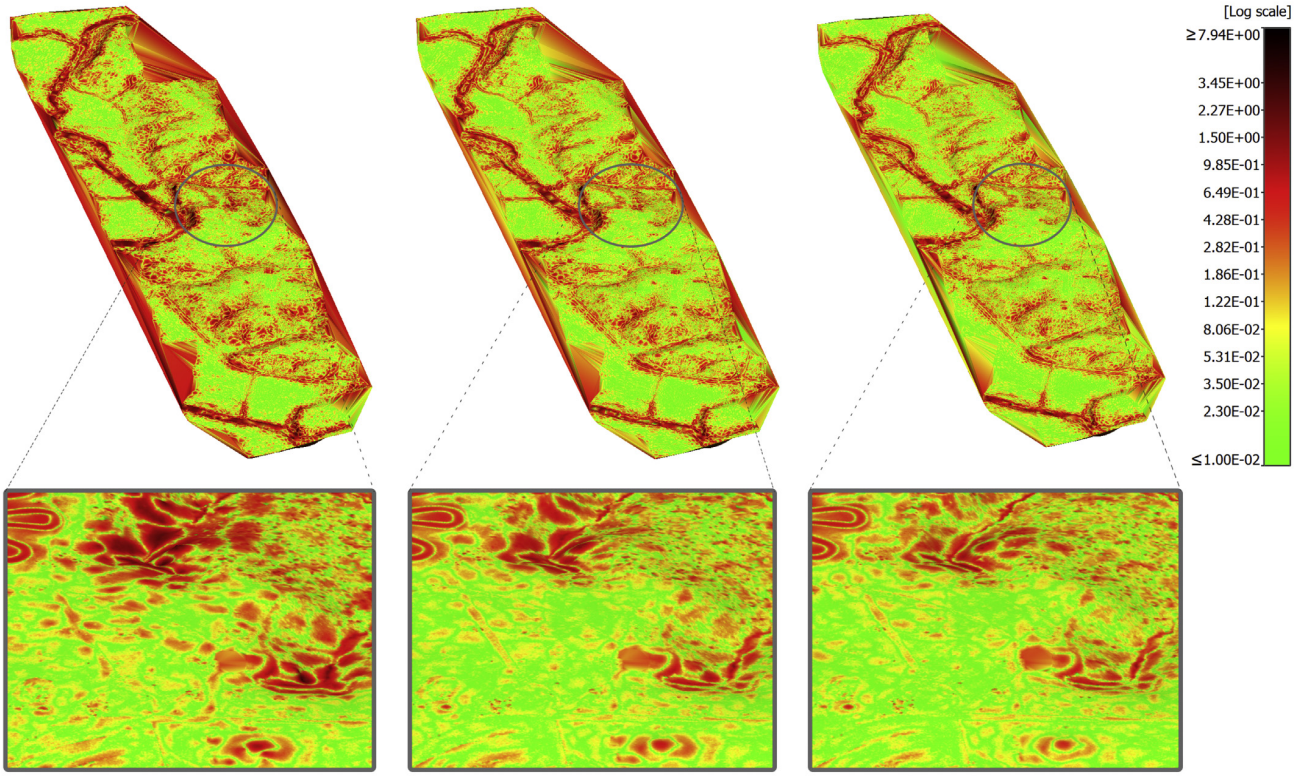
Table 4
The RBF approximation error and density of least square matrix for the tested datasets and different radial basis functions. Note that density of least square matrix expresses percentage of non-zero elements in matrix and that one foot [ft] corresponds to 0.3048 m [m].

Phenomenon	without polynomial Wendland's			with linear polynomial Wendland's		
	$\phi_{3,0}$	$\phi_{3,1}$	$\phi_{3,3}$	$\phi_{3,0}$	$\phi_{3,1}$	$\phi_{3,3}$
Synthetic data						
Mean absolute error [ft]	0.0041	0.0021	0.0019	0.0040	0.0019	0.0019
Deviation of error [ft]	1.92E-5	6.06E-6	5.25E-6	1.90E-5	5.45E-6	5.12E-6
Mean relative error [%]	0.0151	0.0076	0.0072	0.0150	0.0070	0.0072
Serpent Mound						
Mean absolute error [ft]	0.173	0.141	0.130	0.164	0.139	0.129
Deviation of error [ft]	0.072	0.047	0.037	0.068	0.047	0.037
Mean relative error [%]	0.015	0.012	0.011	0.014	0.012	0.011
Density of LSE matrix [%]	8.413	8.413	8.413	8.468	8.468	8.468
Mount St. Helens						
Mean absolute error [ft]	12.568	11.589	9.881	12.129	10.935	9.773
Deviation of error [ft]	188.595	165.574	100.738	159.139	122.659	98.993
Mean relative error [%]	0.013	0.012	0.010	0.012	0.011	0.010
Density of LSE matrix [%]	6.470	3.452	6.470	6.536	3.510	6.536

[1] http://www.liblas.org/samples/.



(a) Mount Saint Helens dataset: Wendland's RBF $\phi_{3,0}$, $\alpha = 0.0005$ (left), Wendland's RBF $\phi_{3,1}$, $\alpha = 0.0007$ (center) and Wendland's RBF $\phi_{3,3}$, $\alpha = 0.0005$ (right)



(b) Serpent Mound dataset: Wendland's RBF $\phi_{3,0}$, $\alpha = 0.01$ (left), Wendland's RBF $\phi_{3,1}$, $\alpha = 0.01$ (center) and Wendland's RBF $\phi_{3,3}$, $\alpha = 0.01$ (right)

Fig. 6. Results for the tested real datasets false-colored by magnitude of absolute error.

error. The error is defined as the absolute value of the difference between Franke's function (17) and approximated function. It can be seen that for the synthetic dataset, the RBF approximation with Wendland's $\phi_{3,3}$ basis function returns the best result in terms of the error. On the contrary, the worst result is obtained for the RBF approximation with Wendland's $\phi_{3,0}$ basis function. Table 4 shows three different error measures of the datasets depending on the chosen basis functions: mean absolute error, deviation and mean relative error.

These error measures are performed for approximation without polynomial reproduction and for approximation with linear polynomial reproduction. It can be seen that the RBF approximation with linear polynomial reproduction produces slightly better results than the RBF approximation without reproduction in terms of the error, but this improvement seems to be insignificant.

The RBF approximation for the real datasets was solved using "block-wise" approach described above. Approximations of Mount Saint Helens dataset without polynomial reproduction for all CS-RBFs are shown in Fig. 6a.

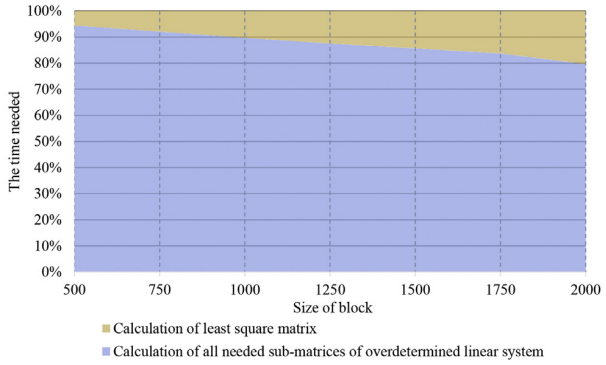
It illustrates the magnitude of error at each point of the original point cloud. Moreover, the detail of a crater is shown for each approximation. It can be seen that the RBF approximation with Wendland's $\phi_{3,3}$ basis function returns the best results in terms of the error. On the contrary, the worst result is obtained for the RBF approximation with Wendland's $\phi_{3,0}$ basis function again. For this approximation, sharp peaks are formed. It is most evident around the rim of a crater. Also for the Mount Saint Helens dataset, the three error measures of the computed elevation for all used CS-RBFs and for both types of RBF approximation (i.e. approximation without polynomial reproduction and approximation with linear polynomial

Fig. 7. The signed errors for the Serpent Mound dataset and Wendland's RBF $\phi_{3,1}$ with $\alpha = 0.01$: the positive error is colored white and the negative error is colored black.

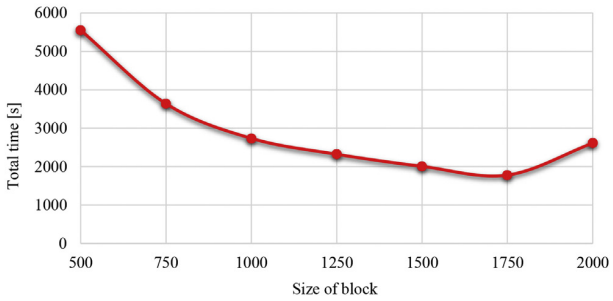
reproduction) are presented in Table 4. These results confirm the statements above. Further, it can be seen that the RBF approximation with linear reproduction again produces better results than the RBF approximation without reproduction, especially in terms of deviation of error.

The last presented experimental results are for the RBF approximation of Serpent Mound without polynomial reproduction and are shown in Fig. 6b. It illustrates the magnitude of error at each point of the original point cloud. Moreover, the detail of Serpent Mound is shown for each approximation. It can be seen that the RBF approximation with Wendland's $\phi_{3,3}$ basis function returns a slightly better result than RBF approximation with Wendland's $\phi_{3,1}$ basis function in terms of the error for the Serpent Mound dataset. The RBF approximation with Wendland's $\phi_{3,0}$ basis function returns the worst results. These facts are mainly evident in the details. Further, we can see that the highest errors occur on the boundary of the terrain for all cases. The three error measures of the elevation for all used CS−RBFs and for both types of RBF approximation (i.e. approximation without polynomial reproduction and approximation with linear polynomial reproduction) are presented in Table 4. These results again confirm the statements above. Further, it can be seen that the RBF approximation with linear polynomial reproduction produces slightly better results than the RBF approximation without reproduction in terms of the error, but this improvement is not significant. The mutual comparison of both real datasets in terms of the deviation of error (Table 4) indicates that RBF approximation with linear reproduction returns considerably better results than RBF approximation without polynomial reproduction if the range of associated values is large. Moreover, it should be noted that the degree of smoothness for the tested type of real datasets is lower than degree of smoothness for Wendland's $\phi_{3,1}$ and Wendland's $\phi_{3,3}$ basis functions and, therefore, the comparison of RBF approximation with Wendland's $\phi_{3,1}$ basis function and RBF approximation with Wendland's $\phi_{3,3}$ basis function returns less significant results. The situation is different for comparison of RBF approximation with Wendland's $\phi_{3,0}$ basis function and RBF approximation with Wendland's $\phi_{3,1}$ basis function where the difference is significant. The signed errors for the Serpent Mound dataset and Wendland's $\phi_{3,1}$ basis function are shown in Fig. 7. We can see that the signs are different at various locations. Similar results are obtained for the rest of the experiments.

The implementation of the RBF approximation was performed in MATLAB and tested on a PC with the following configuration:



(a) The time needed for calculation of all sub-matrices of the matrix $\boldsymbol{A}$ (blue color) and for determination of least square matrix $\boldsymbol{A}^T\boldsymbol{A}$ (orange color) for the Serpent Mound depending on block size. Note that 100% corresponds to total time of computation.



(b) The total time needed for approximation of the Serpent Mound depending on block size.

Fig. 8. Time performance for approximation of the Serpent Mound depending on the block size.

- CPU: Intel® Core™ i7-4770 (4 × 3.40 GHz + hyper-threading),
- memory: 32 GB RAM,
- operation system: Microsoft Windows 7 64 bits.

For the approximation of the Serpent Mound dataset with 10,000 local Wendland's $\phi_{3,1}$ basis functions with shape parameter $\alpha = 0.01$, the running times for different sizes of blocks were measured. These computational times are presented in Fig. 8b. We can see that the time performance is large for the approximation matrix which is partitioned into small blocks (i.e. smaller than 500 × 500 blocks). This is caused by overhead costs and, moreover, each entry in the matrix $\boldsymbol{A}$ of the overdetermined linear system has to be calculated more times than for larger sizes of block. On the other hand, the running time begins to rise above the permissible limit due to memory swapping for the approximation matrix which is partitioned into larger blocks (i.e. larger than 2500 × 2500 blocks).

The running time for determination of RBF approximation with the mentioned parameters was divided into two steps. The running time needed for calculation of all sub-matrices formed from the matrix $\boldsymbol{A}$ of the original overdetermined linear system of equations by the block-wise approach is determined in the first step. The running time needed for calculation of the least square matrix $\boldsymbol{A}^T\boldsymbol{A}$ and for calculation of the vector of unknown weights is measured in the second step. The comparison of the perceptual time performance of these two steps can be seen in Fig. 8a. It can be seen that the most time-consuming part is the first step, in which all needed sub-matrices are calculated (lower part in the graph).

7. Conclusion

In this paper two different RBF approximation methods are experimentally verified using one synthetic and two real datasets. The first method is an RBF approximation without polynomial reproduction and the second method is an RBF approximation with linear reproduction. Moreover, a new approach to the RBF approximation of large datasets is presented. The proposed approach uses symmetry of the matrix, partitioning the matrix into blocks and block-wise solving which enables the computation on systems with limited main memory. Because CS-RBFs are used for approximation, data structures for storage of the sparse matrix can be employed; thereby a larger size of blocks can be chosen and the computational costs decrease. The experiments proved that the proposed approach is fully applicable for the RBF approximation for large datasets.

The experiments also showed that, depending on the quality of the results, it is possible to divide the CS-RBFs from the catalog of RBFs (D.2.7 in Fasshauer (2007)) into three groups. The results of the experiments proved that RBF approximation with linear reproduction returns better result than RBF approximation without polynomial reproduction, particularly if the range of associated values is large. The experiments also proved that the RBF methods have problems with the accuracy of calculation on the boundary of an object, which is a well-known property. The presented approach is directly applicable in GIS and geoscience fields.

Future work will be aimed at improving the accuracy at the boundaries, on the computational performance without loss of approximation accuracy and computation of optimal shape parameters. Also, the "moving window" technique will be explored to increase speed of computation.

Acknowledgments

The authors would like to thank their colleagues at the University of West Bohemia, Plzeň, for their discussions and suggestions, and the anonymous reviewers for their valuable comments. The research was supported by the National Science Foundation GAČR project GA17-05534S and partially supported by SGS 2016-013 project.

References

Bell, N., Garland, M., 2009. Implementing sparse matrix-vector multiplication on throughput-oriented processors. In: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. ACM, p. 18.

Buhmann, M.D., 2003. Radial Basis Functions: Theory and Implementations, vol. 12. Cambridge university press.

Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R., 2001. Reconstruction and representation of 3d objects with radial basis functions. August 12-17, 2001. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, Los Angeles, California, USA, pp. 67–76.

Chen, C., Li, Y., 2012. A robust method of thin plate spline and its application to DEM construction. Comput. Geosci. 48, 9–16.

Cressie, N., 2015. Statistics for Spatial Data. John Wiley & Sons.

Fasshauer, G.E., 2007. Meshfree Approximation Methods with MATLAB, vol. 6. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

Fasshauer, G.E., Zhang, J.G., 2007. On choosing "optimal" shape parameters for RBF approximation. Numer. Algorithms 45 (1–4), 345–368.

Franke, R., 1979. A Critical Comparison of Some Methods for Interpolation of Scattered Data. Tech. Rep. NPS53-79-003. NAVAL POSTGRADUATE SCHOOL MONTEREY CA.

Franke, R., 1982. Scattered data interpolation: tests of some methods. Math. Comput. 38 (157), 181–200.

Hardy, R.L., 1971. Multiquadratic equations of topography and other irregular surfaces. J. Geophys. Res. 76, 1905–1915.

Hardy, R.L., 1990. Theory and applications of the multiquadric-biharmonic method 20 years of discovery 19681988. Comput. Math. Appl. 19 (8), 163–208.

Hon, Y.-C., Šarler, B., Yun, D.-F., 2015. Local radial basis function collocation method for solving thermo-driven fluid-flow problems with free surface. Eng. Anal. Bound. Elem. 57, 2–8.

Li, M., Chen, W., Chen, C., 2013. The localized RBFs collocation methods for solving high dimensional PDEs. Eng. Anal. Bound. Elem. 37 (10), 1300–1304.

Ma, Y., Royer, J.-J., Wang, H., Wang, Y., Zhang, T., 08 2014. Factorial kriging for multiscale modelling. J. South. Afr. Inst. Min. Metall. 114 (8), 651–659.

Majdisova, Z., Skala, V., 2016. A new radial basis function approximation with reproduction. In: Blashki, K., Xiao, Y. (Eds.), Proceedings of the International Conferences on Interfaces and Human Computer Interaction 2016, Game and Entertainment Technologies 2016 and Computer Graphics, Visualization, Computer Vision and Image Processing 2016. IADIS Press, pp. 215–222.

Mallet, J.-L., Apr. 1989. Discrete smooth interpolation. ACM Trans. Graph 8 (2), 121–144.

Pepper, D.W., Rasmussen, C., Fyda, D., 2014. A meshless method using global radial basis functions for creating 3-d wind fields from sparse meteorological data. Comput. Assisted Methods Eng. Sci. 21 (3–4), 233–243.

Rippa, S., 1999. An algorithm for selecting a good value for the parameter *c* in radial basis function interpolation. Adv. Comput. Math. 11 (2–3), 193–210.

Royer, J.-J., Vieira, P.C., 1984. Dual Formalism of Kriging, g. verly et al. Edition, vol. 2. D. Reidel Publishing Company.

Scheuerer, M., 2011. An alternative procedure for selecting a good value for the parameter *c* in rbf-interpolation. Adv. Comput. Math. 34 (1), 105–126.

Šimecek, I., 2009. Sparse matrix computations using the quadtree storage format. In: Proceedings of 11th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2009), pp. 168–173.

Skala, V., 2013. Fast interpolation and approximation of scattered multidimensional and dynamic data using radial basis functions. WSEAS Trans. Math. 12 (5), 501–511.

Skala, V., 2015. Meshless interpolations for computer graphics, visualization and games. May 4-8, 2015. In: Zwicker, M., Soler, C. (Eds.), Eurographics 2015-Tutorials. Eurographics Association, Zurich, Switzerland.

Turk, G., O'Brien, J.F., 2002. Modelling with implicit surfaces that interpolate. ACM Trans. Graph 21 (4), 855–873.

Wendland, H., 2006. Computational aspects of radial basis function approximation. Stud. Comput. Math. 12, 231–256.

Appendix F

Algorithm for Placement of Reference Points and Choice of an Appropriate Variable Shape Parameter for the RBF Approximation

Majdišová, Z., Skala, V., Šmolík, M.

Accepted to:
Integrated Computer-Aided Engineering, IOS Press, ISSN 1069-2509, IF 4.904

Undefined 0 (2019) 1–0
IOS Press

Algorithm for Placement of Reference Points and Choice of an Appropriate Variable Shape Parameter for the RBF Approximation

Zuzana Majdisova [a,*], Vaclav Skala [a] and Michal Smolik [a]
[a] *Department of Computer Science and Engineering, Faculty of Applied Sciences*
University of West Bohemia, Plzeň, Czech Republic
E-mail: majdisz@kiv.zcu.cz, smolik@kiv.zcu.cz
URL: http://www.VaclavSkala.eu

Abstract. Many Radial Basis Functions (RBFs) contain a shape parameter which has an important role to ensure good quality of the RBF approximation. Determination of the optimal shape parameter is a difficult problem. In the majority of papers dealing with the RBF approximation, the shape parameter is set up experimentally or using some ad-hoc method. Moreover, the constant shape parameter is almost always used for the RBF approximation, but the variable shape parameter produces more accurate results. Several variable shape parameter methods, which are based on random strategy or on an evolutionary algorithm, have been developed. Another aspect which has an influence on the quality of the RBF approximation is the placement of reference points.

A novel algorithm for finding an appropriate set of reference points and a variable shape parameter selection for the RBF approximation of functions $y = f(x)$ (i.e. the case when a one-dimensional dataset is given and each point from this dataset is associated with a scalar value) is presented. Our approach has two steps and is based on exploiting features of the given dataset, such as extreme points or inflection points, and on comparison of the first curvature of a curve. The proposed algorithm can be used for the approximation of data describing a curve parameterized by one variable in multidimensional space, e.g. a robot path planning, etc.

Keywords: Radial basis functions, Approximation, Variable shape parameter, Curvature, Lagrange multipliers

1. Introduction

Radial basis functions (RBFs) are used to solve many technical and non-technical problems. RBFs are real-valued functions which depend only on the distance from the fixed center point. A RBF method was originally introduced by [1], [2]. It is a powerful tool for the meshless interpolation and approximation of scattered data, as space tessellation is not required. Moreover, this method is independent with respect to the dimension of the space. RBF applications can be found in data visualization [3], surface reconstruction [4], [5], [6], vector fields visualization [7], solving partial differential equations [8], [9], etc.

RBFs can be divided into two main groups of basis functions: global RBFs and Compactly Supported RBFs (CS-RBFs) [10]. The use of CS-RBFs leads to a simpler and faster computation, because a system of linear equations has a sparse matrix. However, approximation using CS-RBFs is quite sensitive to the density of given scattered data. Global RBFs are useful in repairing incomplete datasets and they are significantly less sensitive to the density of data as they cover the whole domain. However, they lead to a linear system of equations with a dense and ill-conditioned matrix.

*Corresponding author. E-mail: majdisz@kiv.zcu.cz

Choice of an appropriate shape parameter of RBFs is extremely important to ensure good approximation. Several articles have been dedicated to introducing different algorithms to compute a constant value as an appropriate value for the shape parameter [11], [12], [13], [14], etc. Many of these focus on finding the minimal error in computations or are based on convergence analysis. Other articles show that variable shape parameters are useful instead of a fixed shape parameter. Sufficient conditions to guarantee a unique solution of the RBF interpolation with variable shape parameters are derived in [15] for CS-RBFs and in [16] for global RBFs. The variable shape parameters are determined by used of genetic algorithm [17] and minimization of the local cost function [18], [19] or numerically by minimizing the root-mean-square errors [20]. For these purposes, there are many other papers which are dealing with the general global optimization such as [21], [22], [23]. Other approaches generate the variable shape parameters from an estimated range when different distributions of values are used [24], [25], use Neural Network RBF approach [26], [27], [28], [29], [30] or ortogonal least square [31]. However, the approaches mentioned do not reflect features of the given data.

Our approach for 1.5D case eliminating above mentioned drawbacks will be described in this paper. The proposed method consists of two steps. In the first step, the Thin-Plate Spline (TPS) function is used. The second step is focused on RBFs which have smoothness at least C^3 at the origin (e.g. Gaussian RBF, Wendland's $\phi_{3,2}$, etc.). This condition follows from the requirement that the first curvature of curves is smooth, which is a direct result of the algorithm described below. The proposed approach leads to a significant compression of the given data and obtaining their analytical form. Our approach can be applied to many real data in the different areas of interest, e.g. data obtained from GPS navigation describing the terrain profile [32], data for recovering smooth robot trajectory [33], total electron content data [34], etc.

In the following section, the fundamental theoretical background needed for description of the proposed algorithm will be mentioned. The proposed two steps algorithm, including the derivation of the appropriate variable shape parameter, will be described in Section 3. The results of the proposed algorithm for synthetic and real data will be presented in Section 4.

2. Theoretical Background

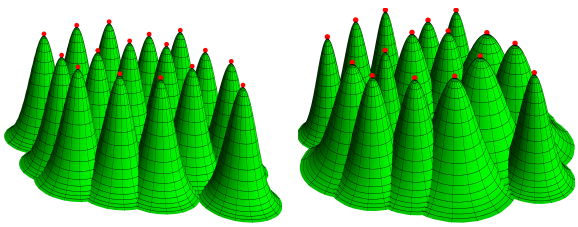
In this section, some theoretical aspects needed for description of the proposed algorithm for placement of reference points and choice of an appropriate variable shape parameter for the RBF approximation will be introduced.

2.1. RBF Approximation with a Variable Shape Parameter

We assume that we have an unordered dataset $\{\boldsymbol{x}_i\}_1^N \in \mathbb{E}^n$, where n denotes the dimension of space. Further, each point $\boldsymbol{x}_i$ from the dataset is associated with a vector $\boldsymbol{h}_i \in \mathbb{E}^p$ of the given values, where p is the dimension of the vector, or a scalar value, i.e. $h_i \in \mathbb{E}^1$. In the following, we will deal with scalar data approximation, i.e. the case when each point $\boldsymbol{x}_i$ is associated with a scalar value h_i is considered. Let us introduce a set of new reference points (knots of RBF) $\{\boldsymbol{\xi}_j\}_1^M \in \mathbb{E}^n$, where n denotes the dimension of space, M is the number of reference points and $M \ll N$.

These reference points may not necessarily have any special distribution as uniform distribution, etc. However, a good placement of the reference points improves the approximation of the underlying data. The RBF approximation is based on the distance computation between the given point $\boldsymbol{x}_i$ and the reference point $\boldsymbol{\xi}_j$.

As generally known, most RBFs are dependent on the shape parameter α, which influences the radius of support. In the case of the fundamental RBF approximation (see [35], [36], [37]), the shape parameter of the RBF used is set to a constant value for all M RBFs, see Fig. 1a. Nevertheless, it is possible to set a different shape parameter for each of the M RBFs, where the shape parameter can be determined depending on features of



(a) Constant shape parameter α (b) Variable shape parameters $\{\alpha_j\}$

Fig. 1. RBF collocation functions centered at reference points $\{\boldsymbol{\xi}_j\} \in \mathbb{E}^2$.

the neighborhood of the reference point at which a given RBF is centered or some other criterion can be used. In such case, it is the RBF approximation with a variable shape parameter and the approximated value is determined as:

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} c_j \phi\left(r_j, \alpha_j\right) = \sum_{j=1}^{M} c_j \phi\left(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|, \alpha_j\right), \quad (1)$$

where $\phi(r_j, \alpha_j)$ is an RBF with shape parameter α_j centered at point $\boldsymbol{\xi}_j$, see Fig. 1b. The approximating function $f(\boldsymbol{x})$ is represented as a sum of the M RBFs with a variable shape parameter, each associated with a different reference point $\boldsymbol{\xi}_j$, and weighted by a coefficient c_j which has to be determined.

It can be seen that the overdetermined linear system of equations is obtained when inserting all points $\boldsymbol{x}_i$, with $i = 1, \ldots, N$, into (1):

$$h_i = f(\boldsymbol{x}_i) = \sum_{j=1}^{M} c_j \phi\left(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|, \alpha_j\right) \; i = 1, \ldots, N. \quad (2)$$

Using the matrix notation, the linear system of equations (2) can be expressed:

$$\begin{pmatrix} \phi\left(r_{11}, \alpha_1\right) \cdots \phi\left(r_{1M}, \alpha_M\right) \\ \vdots \quad \ddots \quad \vdots \\ \phi\left(r_{i1}, \alpha_1\right) \cdots \phi\left(r_{iM}, \alpha_M\right) \\ \vdots \quad \ddots \quad \vdots \\ \phi\left(r_{N1}, \alpha_1\right) \cdots \phi\left(r_{NM}, \alpha_M\right) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_M \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_i \\ \vdots \\ h_N \end{pmatrix}, \quad (3)$$

where $r_{ij} = \|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|$ is the distance between the given point $\boldsymbol{x}_i$ and the reference point $\boldsymbol{\xi}_j$.

Equation (3) can also be expressed in the form:

$$\boldsymbol{A}_{Var}\, \boldsymbol{c} = \boldsymbol{h}. \quad (4)$$

The presented system is again overdetermined, $M \ll N$, and can be solved by the least squares method, QR decomposition, etc.

The use of variable shape parameter α_j disrupts the proof of non-singularity of approximation matrix $\boldsymbol{A}_{Var}$. In practice, however, the constant shape parameter does not prevent approximation matrix becoming so ill-conditioned as to be essentially singular [37], and the benefits of variable shape parameter are considered substantial. Moreover, in [38], it is shown that the variable shape parameter is improving the conditionality.

2.2. RBF Approximation with a Variable Shape Parameter and Lagrange Multipliers

In many cases, it is required that the approximate function must have exactly the given values $\{s_k\}_1^K \in \mathbb{E}^1$ at some set of points $\{\boldsymbol{\eta}_k\}_1^K \in \mathbb{E}^n$, where $K \ll N$. It follows that the aim is finding the RBF approximation of the dataset in the form (1) subject to K constraints:

$$\{f(\boldsymbol{\eta}_k) = s_k\}_1^K. \quad (5)$$

This problem can be solved as minimization of the square error of the RBF approximation subject to K constraints and the method of Lagrange multipliers can be used for this purpose. Specifically, our goal is to minimize the following function:

$$\begin{aligned} F(\boldsymbol{c}, \boldsymbol{\lambda}) &= \sum_{i=1}^{N} \left(\sum_{j=1}^{M} c_j \phi\left(\|\boldsymbol{x}_i - \boldsymbol{\xi}_j\|, \alpha_j\right) - h_i \right)^2 \\ &\quad + \sum_{k=1}^{K} \lambda_k \left(\sum_{j=1}^{M} c_j \phi\left(\|\boldsymbol{\eta}_k - \boldsymbol{\xi}_j\|, \alpha_j\right) - s_k \right) \\ &= (\boldsymbol{A}_{Var}\boldsymbol{c} - \boldsymbol{h})^2 + (\boldsymbol{c}^T \boldsymbol{R}^T - \boldsymbol{s}^T) \cdot \boldsymbol{\lambda}. \end{aligned} \quad (6)$$

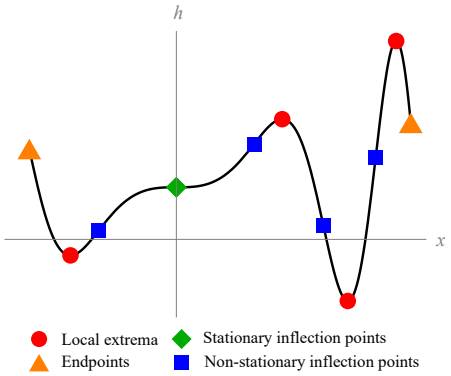
This minimum is obtained by differentiating equation (6) with respect to $\boldsymbol{c}$ and $\boldsymbol{\lambda}$ and finding the zeros of those derivatives. This leads to equations:

$$\begin{aligned} \frac{\partial F}{\partial \boldsymbol{c}} &= 2\boldsymbol{A}_{Var}^T \boldsymbol{A}_{Var} \boldsymbol{c} - 2\boldsymbol{A}_{Var}^T \boldsymbol{h} + \boldsymbol{R}^T \boldsymbol{\lambda} = \boldsymbol{0} \\ \frac{\partial F}{\partial \boldsymbol{\lambda}} &= \boldsymbol{R}\boldsymbol{c} - \boldsymbol{s} = \boldsymbol{0}, \end{aligned} \quad (7)$$

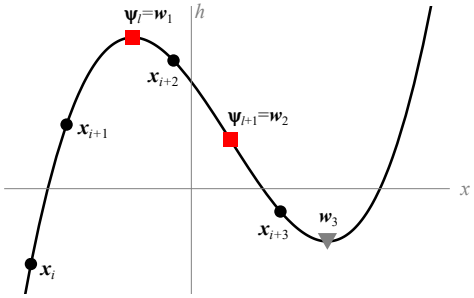
which leads to a system of linear equations:

$$\begin{pmatrix} 2\boldsymbol{A}_{Var}^T \boldsymbol{A}_{Var} & \boldsymbol{R}^T \\ \boldsymbol{R} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} 2\boldsymbol{A}_{Var}^T \boldsymbol{h} \\ \boldsymbol{s} \end{pmatrix}. \quad (8)$$

The presented system has an $(M + K) \times (M + K)$ symmetric matrix, where $K \ll M$, and can be solved by the LU decomposition, QR decomposition, etc. Then the RBF approximation of the given dataset can be expressed using equation (1) and the vector $\boldsymbol{c}$ which was computed from the linear system (8). However, the matrix $\boldsymbol{A}_{Var}$ depends on shape parameters and their estimation will be explained in the following sections.



(a) The significant points of the function are local extrema, stationary and non-stationary inflection points and endpoints.



(b) Four points $\{\boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i+3}\}$ from the dataset are interpolated by cubic curve and three significant points $\{\boldsymbol{w}_t\}$ of this curve are shown: (red squares) the points added to the set of suspicious points, (gray triangle) a point discarded because the condition (12) is not valid.

Fig. 2. Finding significant points of given data

2.3. Determination of Significant Points and Their Properties

In this section, the proposed approach for determination of the significant points of the given dataset will be described. These points have a large influence on the quality of the RBF approximation.

The paper is focused on a 1.5D case, i.e. we have given a dataset $\{\boldsymbol{x}_i\}_1^N \in \mathbb{E}^1$ and each point $\boldsymbol{x}_i$ from this dataset is associated with a value $h_i \in \mathbb{E}^1$.

The local extrema, stationary and non-stationary inflection points and endpoints of the dataset are included among significant points, see Fig. 2a. The points which will be used for the determination of the set of significant points of the given data will be called the set of suspicious points $\{\boldsymbol{\psi}_l\}_{l=1}^{N_S}$, i.e. a set of significant points is a reduced set of suspicious points.

First, the ordering of the given dataset is performed. After that, the set of suspicious points is determined. For these purposes, every four points $\{\boldsymbol{x}_i, \ldots, \boldsymbol{x}_{i+3}\}$ from the given data are interpolated by a cubic curve in the form:

$$g_q(\boldsymbol{x}) = \beta_{1q} \cdot \boldsymbol{x}^3 + \beta_{2q} \cdot \boldsymbol{x}^2 + \beta_{3q} \cdot \boldsymbol{x} + \beta_{4q} \quad q = 1, \ldots, N-3, \tag{9}$$

which leads to the solution of the $(N-3)$ linear systems of the size 4×4. Then, the significant points $\{\boldsymbol{w}_t\}$ of this cubic are determined, see Fig. 2b. The significant points for the cubic curve (9) meet at least one of the following conditions:

$$\frac{\partial g_q}{\partial \boldsymbol{x}} = 0 \quad \text{or} \quad \frac{\partial^2 g_q}{\partial \boldsymbol{x}^2} = 0, \tag{10}$$

which leads in our case to the following set:

$$\{\boldsymbol{w}_t\} = \left\{-\tfrac{\beta_{2q}}{3\beta_{1q}}\right\} \cup \left\{(\beta_{2q}^2 - 3\beta_{1q}\beta_{3q}) > 0 : \tfrac{-\beta_{2q} \pm \sqrt{\beta_{2q}^2 - 3\beta_{1q}\beta_{3q}}}{3\beta_{1q}}\right\} \tag{11}$$

The significant point $\boldsymbol{w}_t$ is further added to the set of suspicious points $\{\boldsymbol{\psi}_l\}_{l=1}^{N_S}$, if the necessary condition:

$$\boldsymbol{x}_i \leq \boldsymbol{w}_t \leq \boldsymbol{x}_{i+3} \tag{12}$$

is valid. Moreover, the functional value $g_q(\boldsymbol{\psi}_l)$ of the associated cubic is calculated at such a suspicious point $\boldsymbol{\psi}_l$:

$$g_q(\boldsymbol{\psi}_l) = \beta_{1q} \cdot \boldsymbol{\psi}_l^3 + \beta_{2q} \cdot \boldsymbol{\psi}_l^2 + \beta_{3q} \cdot \boldsymbol{\psi}_l + \beta_{4q} \tag{13}$$

and the absolute value of the first curvature $\left|{}^1k_q(\boldsymbol{\psi}_l)\right|$ of the associated cubic is determined (using the symbolic manipulation):

$$\left|{}^1k_q(\boldsymbol{\psi}_l)\right| = \left|{}^1k\left(g_q(\boldsymbol{\psi}_l)\right)\right| = = \left|\frac{6\beta_{1q}\boldsymbol{\psi}_l + 2\beta_{2q}}{\left(1 + \left(3\beta_{1q}\boldsymbol{\psi}_l^2 + 2\beta_{2q}\boldsymbol{\psi}_l + \beta_{3q}\right)^2\right)^{3/2}}\right|. \tag{14}$$

It should be noted that absolute values of the first curvature for the endpoints are calculated using

the cubic curve interpolating points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4\}$, or $\{\boldsymbol{x}_{N-3}, \ldots, \boldsymbol{x}_N\}$.

The set of suspicious points may contain two or more identical points or points very close to identical. This problem is caused by the fact that one significant point can be obtained from up to three cubics. Therefore, the reduction of the set of suspicious points is performed. The resulting set of significant points of the given data $\{\boldsymbol{\chi}_u\}$ is determined as follows. First, the endpoints are added to the set of significant points $\{\boldsymbol{\chi}_u\}$, their associated functional values are added to set $\{g(\boldsymbol{\chi}_u)\}$ and their associated absolute values of the first curvatures are added to set $\left\{\left|{}^1k(\boldsymbol{\chi}_u)\right|\right\}$. Now, let δ is the average step between the given sorted points, then the reduction of the set of suspicious points can be performed as follows. The suspicious points which meet the condition:

$$\left(\|\boldsymbol{\psi}_l - \boldsymbol{x}_1\| \leq \delta\right) \text{or} \left(\|\boldsymbol{\psi}_l - \boldsymbol{x}_N\| \leq \delta\right), \tag{15}$$

are deleted. Further, the subset Ψ_u of suspicious points, where each point meets the relation:

$$\Psi_u = \left\{\boldsymbol{\psi}_{\hat{l}} : \|\boldsymbol{\psi}_{\hat{l}} - \boldsymbol{\psi}_1\| \leq \delta\right\}, \tag{16}$$

is removed from the set of suspicious points $\{\boldsymbol{\psi}_l\}$ and the new significant point is determined from them by averaging:

$$\boldsymbol{\chi}_u = \frac{\sum \boldsymbol{\psi}_{\hat{l}}}{|\Psi_u|}, \tag{17}$$

where $|\Psi_u|$ is a size (cardinality) of the subset Ψ_u. Moreover, the associated functional value $g(\boldsymbol{\chi}_u)$ and the associated absolute value of the first curvature $\left|{}^1k(\boldsymbol{\chi}_u)\right|$ are determined in the same way. The process is repeated until the set of suspicious points is not empty.

The whole algorithm for finding the set of significant points of the given data, the calculation of the first curvatures and associated functional values in them is summarized in Algorithm 1.

Algorithm 1: Determination of the set of significant points $\{\boldsymbol{\chi}_u\}_1^S$, the absolute values of the first curvatures $\left\{\left|{}^1k(\boldsymbol{\chi}_u)\right|\right\}_1^S$ and the associated functional values $\{g(\boldsymbol{\chi}_u)\}_1^S$.

Input: given points $\{\boldsymbol{x}_i\}_1^N \in \mathbb{E}^1$ and their associated scalar values $\{h_i\}_1^N \in \mathbb{E}^1$, the average step between given sorted points δ

Output: significant points, their associated first curvatures and their associated functional values $\left\{\boldsymbol{\chi}_u, \left|{}^1k(\boldsymbol{\chi}_u)\right|, g(\boldsymbol{\chi}_u)\right\}_1^S$

```
Sort the given points {x_i}_1^N in ascending order.
for i = 1, ..., N − 3 do
    Determine the significant points {w_t} for cubic curve defined by {x_i, ..., x_{i+3}}, (eq. (9), eq. (11)).
    if i = 1 then
        Add the triplet {x_1, |1k_1(x_1)|, h_1} to the output (using eq. (14)).
    foreach w_t do
        if x_i ≤ w_t ≤ x_{i+3} then
            Add the point w_t to the set of suspicious points {ψ_l}.
            Compute the first curvature |1k_i(w_t)|, eq. (14), and the functional value g_i(w_t), eq. (13), and
              add these values to appropriate sets
    if i = (N − 3) then
        Add the triplet {x_N, |1k_{N−3}(x_N)|, h_N} to the output (using eq. (14)).
From the set of suspicious points {ψ_l}, delete all points such that (15) is valid.
while the set of suspicious points is not empty do
    Find Ψ_u = {ψ_î : ||ψ_î − ψ_1|| ≤ δ} in the set of suspicious points {ψ_l}.
    Add the triplet { Σψ_î/|Ψ_u| , Σ|1k_q(ψ_î)|/|Ψ_u| , Σg_q(ψ_î)/|Ψ_u| } to the output (|Ψ_u| is cardinality of Ψ_u).
    Delete all points ψ_î ∈ Ψ_u and their associated values from the sets {ψ_l}, {|1k_q(ψ_l)|} and {g_q(ψ_l)}.
```

3. Proposed Two Steps Algorithm

In this section, the proposed two steps algorithm for the RBF approximation of the given data (in form $y = f(x)$) including the determination of placement of reference points and the derivation of an appropriate variable shape parameter will be described.

3.1. First Step of the Proposed Approach

In this section, the first step of our approach will be described. The main goal of this step is to perform the primary RBF approximation of the given dataset such that the input data for the second step of our method will be symmetrically distributed around the x-axis. This will be done using inflection points (stationary and non-stationary) and endpoints of the given data. Moreover, this step executes the shift of associated values $\{h_i\}_1^N$ so that the newly obtained associated values are better approximated using the RBF, i.e. the problematic course of the sampled function as in Fig. 3 will be eliminated.

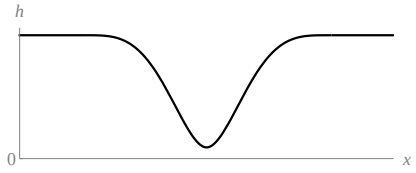


Fig. 3. The course of the sampled function which is poorly approximated using the RBF.

In the first step of our algorithm, the significant points of the given data $\{\boldsymbol{\chi}_u\}$ are firstly found using the method introduced in Section 2.3. From the set of significant points, only inflection points (stationary or non-stationary) and endpoints are used, see Fig. 4a.

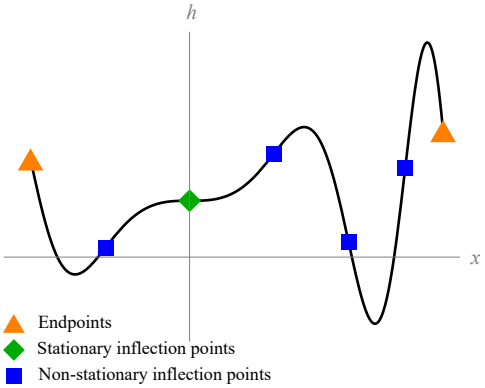
Such a set of points is marked as $\{\hat{\boldsymbol{x}}_v\}_1^{M_1}$, where M_1 is number of points of interest, and the set of their associated functional values is marked as $\{\hat{g}_v\}_1^{M_1}$. The significant point $\boldsymbol{\chi}_u$ is an inflection point (stationary or non-stationary) if its associated absolute value of the first curvature $\left|{}^1k(\boldsymbol{\chi}_u)\right|$ is zero (or close to zero).

Now, the RBF interpolation with polynomial reproduction [39] is performed for the set $\{\hat{\boldsymbol{x}}_v\}_1^{M_1}$, where each point $\hat{\boldsymbol{x}}_v$ is associated with a value $\hat{g}_v$, see Fig. 4b. This means that the vector of unknown weights $\boldsymbol{c}_I = (c_{I1}, \ldots, c_{IM_1})$ and vector of coefficients for the polynomial $\boldsymbol{a}_I$ are computed from the linear system of equations:

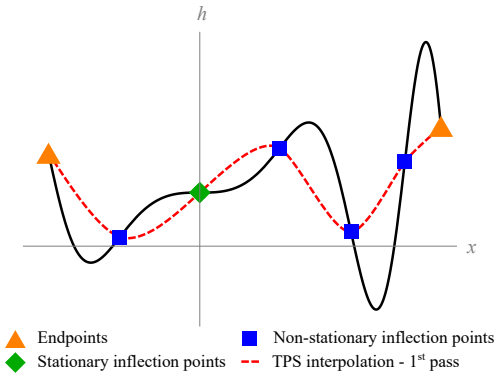
$$\begin{pmatrix} \boldsymbol{A}_I & \boldsymbol{P}_I \\ \boldsymbol{P}_I^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{c}_I \\ \boldsymbol{a}_I \end{pmatrix} = \begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{pmatrix}, \tag{18}$$

where the index I means the interpolation case, the vector of associated values is $\boldsymbol{h} = (\hat{g}_1, \ldots, \hat{g}_{M_1})^T$, the matrix $\boldsymbol{A}_I = \{A_{ij}\} = \{\phi_{TPS}\left(\|\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_j\|\right)\}$ and the matrix $\boldsymbol{P}_I = \{P_i\} = \left\{\left(\hat{\boldsymbol{x}}_i^T, 1\right)\right\}$.
The Thin-Plate Spline (TPS) is used as ϕ_{TPS}, i.e.:

$$\phi_{TPS}(r) = r^2 \log(r) = \frac{1}{2} \cdot r^2 \log(r^2). \tag{19}$$



(a) In the first step, stationary inflection points, non-stationary inflection points and endpoints are used as the significant points of the function.



(b) Original function and the RBF interpolation for selected significant points (result of the first step of our approach).

Fig. 4. The first step of our proposed method.

Algorithm 2: Determination of shifted associated values $\{\hat{h}_i\}_1^N$, i.e. the first step of the proposed approach.

Input: given points $\{\boldsymbol{x}_i\}_1^N$ and their associated scalar values $\{h_i\}_1^N$
Output: interpolation points for the TPS interpolation $\{\hat{\boldsymbol{x}}_v\}_1^{M_1}$, vector of weights for the TPS interpolation $\boldsymbol{c}_I$, vector of coefficients for polynomial $\boldsymbol{a}_I$ and shifted associated values $\{\hat{h}_i\}_1^N$

Determine the significant points of the given data $\{\boldsymbol{\chi}_u\}_1^S$ and calculate the functional values and absolute values of the first curvatures in them, **Algorithm 1**.
Add $\boldsymbol{\chi}_1$ to set $\{\hat{\boldsymbol{x}}_v\}$ and $g(\boldsymbol{\chi}_1)$ to set $\{\hat{g}_v\}$
for $u = 2, \ldots, S-1$ **do**
    if $\left|{}^1k(\boldsymbol{\chi}_u)\right|$ *is zero* **then**
        Add $\boldsymbol{\chi}_u$ to set $\{\hat{\boldsymbol{x}}_v\}$ and $g(\boldsymbol{\chi}_u)$ to set $\{\hat{g}_v\}$
Add $\boldsymbol{\chi}_S$ to set $\{\hat{\boldsymbol{x}}_v\}$ and $g(\boldsymbol{\chi}_S)$ to set $\{\hat{g}_v\}$
Compute the vector of weights $\boldsymbol{c}_I = (c_{I1}, \ldots, c_{IM_1})$ and vector of polynomial coefficients $\boldsymbol{a}_I$, eq. (18)
Compute shifted values $\hat{h}_i$, where $i = 1, \ldots, N$, eq. (20)

This global RBF is chosen because it is not dependent on a shape parameter.

When the vector of weights $\boldsymbol{c}_I$ and vector of coefficients $\boldsymbol{a}_I$ are determined, the new associated values (i.e. shifted associated values) for original dataset $\{\boldsymbol{x}_i\}_1^N$ can be computed:

$$\hat{h}_i = h_i - \sum_{v=1}^{M_1} c_{Iv}\phi_{TPS}\left(\|\boldsymbol{x_i} - \hat{\boldsymbol{x}}_v\|\right) - P(\boldsymbol{x}_i) \qquad (20)$$
$$i = 1, \ldots, N,$$

where ϕ_{TPS} is given by (19).

The whole algorithm for the first step of the proposed approach, i.e. determination of shifted associated values $\{\hat{h}_i\}_1^N$, is summarized in Algorithm 2.

3.2. Second Step of the Proposed Approach

This section will be focused on the second step of our approach. In the input of this step, we assume that we have given the unordered dataset $\{\boldsymbol{x}_i\}_1^N \in \mathbb{E}^1$ and each point $\boldsymbol{x}_i$ from this dataset is associated with a shifted scalar value $\hat{h}_i$ (these values were calculated in the first step), see Fig. 5. Our goal is to determine the RBF approximation with a variable shape parameter and Lagrange multipliers for the described data. Therefore, the set of reference points has to be determined. It should be noted that the placement of reference points has a significant influence on the quality of the approximation. If the reference points are located at the significant points of the given data, then better approximation results are obtained.

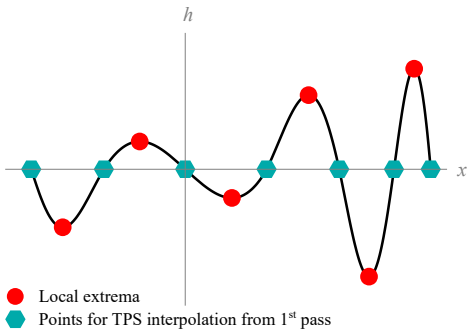


Fig. 5. Course of the input sampled function for the second step of the proposed approach. The set of reference points for the second step of our approach are marked. Hexagons indicate input points for the TPS interpolation from the first step (i.e. stationary and non-stationary inflection points and endpoints) and circles denote local extrema of shifted functional values.

Therefore, the significant points $\{\hat{\boldsymbol{\chi}}_u\}$ of the input data for the second step are found using the method which was introduced in Section 2.3. This set of significant points is further used as the reference points of the RBF approximation $\{\boldsymbol{\xi}_j\}_1^M$. It should be noted that the inflection points (stationary and non-stationary) and endpoints from the second step correspond to the input points for the TPS interpolation $\{\hat{\boldsymbol{x}}_v\}_1^{M_1}$ from the first step, i.e. only local extrema are newly added, see Fig. 5.

3.3. Determination of Appropriate Variable Shape Parameter

The variable shape parameters of RBF α_j at appropriate reference points $\boldsymbol{\xi}_j$ can be determined when the reference points and absolute values of the first curvatures in them are known. Our approach for determining the variable shape parameters is based on a requirement that the validity of the following equality is required:

$$\left(\hat{h}_{max} - \hat{h}_{min}\right) \cdot \left|{}^1k_\phi(0, \alpha_j)\right| = |\phi(0)| \cdot \left|{}^1k(\boldsymbol{\xi}_j)\right|, \quad (21)$$

where $\hat{h}_{max} = \max_{i=1,\dots,N}(\hat{h}_i)$ is the maximum of shifted associated values and $\hat{h}_{min} = \min_{i=1,\dots,N}(\hat{h}_i)$ is the minimum of shifted associated values, $\phi(0)$ is the value of the radial basis function at the center, $\left|{}^1k_\phi(0, \alpha_j)\right|$ is the absolute value of the first curvature for radial basis function $\phi(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|)$ at point $\boldsymbol{x} = \boldsymbol{\xi}_j$. The above mentioned equality is derived based on the consideration that the RBF curve centered at reference point $\boldsymbol{\xi}_j$ has the greatest influence from all used RBFs on the shape of the approximating function at this point, and therefore, the match of the absolute value of first curvature is required. Moreover, the normalization of both function is taken to account. The absolute value of the first curvature for the RBF curve centered at reference point $\boldsymbol{\xi}_j$ is obtained as:

$$\left|{}^1k_\phi(r, \alpha_j)\right| = \left|\frac{\phi_{rr}(r, \alpha_j)}{(1 + \phi_r^2(r, \alpha_j))^{3/2}}\right|, \quad (22)$$

where $\phi_{rr}(r, \alpha_j)$ denotes the second derivative and $\phi_r(r, \alpha_j)$ denotes the first derivative of the RBF.

From equality (21), the following equation can be derived for variable shape parameter α_j of the RBF:

$$\alpha_j = \frac{1}{2}\sqrt{\frac{|{}^1k(\boldsymbol{\xi}_j)| \cdot |\phi(0)|}{\omega \cdot \left(\hat{h}_{max} - \hat{h}_{min}\right)}}, \quad (23)$$

where $\hat{h}_{max}$ is the maximum of shifted values, $\hat{h}_{min}$ is the minimum of shifted values, $\left|{}^1k(\boldsymbol{\xi}_j)\right|$ is the associated absolute value of the first curvature, which was determined by the algorithm described in Section 2.3, and ω is a constant parameter depending on the type of RBF used, see Table 1.

Table 1
Different RBFs and their parameter ω, eg. (23).

RBF	$\boldsymbol{\phi(r)}$	$\boldsymbol{\omega}$
Gaussian RBF	$e^{-(\alpha r)^2}$	2
Inverse quadric	$\frac{1}{1+(\alpha r)^2}$	2
Wendland's $\phi_{3,2}$	$(1-\alpha r)_+^6(35(\alpha r)^2 + 18\alpha r + 3)$	56

It should be noted that this approach can be used only for RBFs which have smoothness of at least C^3 at the origin, because the first curvature of the RBF curve should be smooth.

The last problem which has to be solved is the case when shape parameter α_j associated with reference point $\boldsymbol{\xi}_j$ is zero (i.e. the reference point is an inflection point of the given data), because for such shape parameter the constant function would be obtained. In these cases, correction of the shape parameter is made. Specifically, the weighted average of shape parameters associated with the neighboring points is established and is used as the value of shape parameter α_j.

The whole algorithm for determination of appropriate variable shape parameters is summarized in Algorithm 3.

3.4. Algorithm summary

In this section, a summary of the whole proposed algorithm is provided. First, the first step of the proposed approach is performed, see Algorithm 2, i.e. the TPS interpolation and shift of associated values $\left\{\hat{h}_i\right\}_1^N$ are determined, which leads to the elimination of the problematic course of the sampled function. Next, the reference points $\{\boldsymbol{\xi}_j\}_1^M$ are found and their appropriate absolute values of the first curvatures are calculated for the newly determined data, see Algorithm 1. Then, the appropriate variable shape parameters $\{\alpha_j\}_1^M$ are computed, see Algorithm 3. After that, the RBF approximation can be performed. For these purposes, the RBF approximation with a variable shape parameter and Lagrange multipliers, see Section 2.2, is used. Therefore, the constraints (see (5)) have to be defined. For the proposed approach, the following constraints are used:

$$\left\{f(\boldsymbol{x}_1) = \hat{h}_1 = 0, \;\; f(\boldsymbol{x}_N) = \hat{h}_N = 0\right\}, \quad (24)$$

Algorithm 3: Determination of the variable shape parameters α_j at appropriate reference points $\boldsymbol{\xi}_j$.

Input: reference points and their associated absolute values of the first curvatures $\left\{\boldsymbol{\xi}_j, \left|{}^1k(\boldsymbol{\xi}_j)\right|\right\}_1^M$, the minimum of shifted associated values $\hat{h}_{min} = \min_{i=1,\dots,N}(\hat{h}_i)$, the maximum of shifted associated values $\hat{h}_{max} = \max_{i=1,\dots,N}(\hat{h}_i)$ and the coefficient ω for used RBF (see Table 1)

Output: variable shape parameters associated with appropriate reference points $\{\alpha_j\}_1^M$

Sort the given pairs $\left\{\boldsymbol{\xi}_j, \left|{}^1k(\boldsymbol{\xi}_j)\right|\right\}_1^M$ in ascending order with respect to coordinates of reference points.

for $j = 1, \dots, M$ **do**

$$\alpha_j = \frac{1}{2}\sqrt{\frac{|{}^1k(\boldsymbol{\xi}_j)| \cdot |\phi(0)|}{(h_{max} - h_{min}) \cdot \omega}}$$

for $j = 1, \dots, M$ **do**

  if α_j *is zero* **then**

    if $j = 1$ **then**

      $\alpha_j = \alpha_{j+1}$

    else if $j = M$ **then**

      $\alpha_j = \alpha_{j-1}$

    else

$$\alpha_j = \frac{\alpha_{j-1} \cdot \|\boldsymbol{\xi}_j - \boldsymbol{\xi}_{j-1}\| + \alpha_{j+1} \cdot \|\boldsymbol{\xi}_{j+1} - \boldsymbol{\xi}_j\|}{\|\boldsymbol{\xi}_{j+1} - \boldsymbol{\xi}_{j-1}\|}$$

i.e. the given values $\{s_k\} = \{0, 0\}$ have to be strictly respected at endpoints $\{\boldsymbol{\eta}_k\} = \{\boldsymbol{x}_1, \boldsymbol{x}_N\}$. Now, using eq. (8), the vector of unknown weights $\boldsymbol{c} = (c_1, \dots, c_M)^T$ can be determined.

Finally, the approximated value is determined as:

$$\begin{aligned} f(\boldsymbol{x}) = &\sum_{v=1}^{M_1} c_{Iv}\phi_{TPS}\left(\|\boldsymbol{x} - \hat{\boldsymbol{x}}_v\|\right) + P(\boldsymbol{x}) \\ &+ \sum_{j=1}^{M} c_j \phi\left(\|\boldsymbol{x} - \boldsymbol{\xi}_j\|, \alpha_j\right), \end{aligned} \tag{25}$$

where $\boldsymbol{c}_I = (c_{I1}, \dots, c_{IM_1})$ is the vector of weights for the TPS interpolation, ϕ_{TPS} is the Thin-Plate-Spline, $\{\hat{\boldsymbol{x}}_v\}_1^{M_1}$ are input points for the TPS interpolation, M_1 is number of interpolation points for the first step, $P(\boldsymbol{x})$ is polynomial function of first order, $\boldsymbol{c} = (c_1, \dots, c_M)$ is the vector of weights for the RBF approximation, ϕ is the RBF used (see Table 1), $\{\boldsymbol{\xi}_j\}_1^M$ is the set of reference points, M is number of reference points for the second step and $\{\alpha_j\}_1^M$ are appropriate variable shape parameters.

4. Experimental Results

The above-proposed method of the RBF approximation has been tested on different datasets using Matlab. Moreover, a comparison with the RBF approximation using the constant shape parameter for different distributions of the set of reference points has been made using different radial basis functions, see Table 1.

4.1. Distribution of Reference Points

For the comparison of our approach, the following sets of reference points were used:

Uniform points: This set contains the points distributed uniformly at a given interval.

Epsilon points: This is a special distribution of points which is based on uniform points. Specifically, the points from uniform distribution are randomly drift about value from a range $(-\varepsilon_x, \varepsilon_x)$, where ε_x responds to a quarter of the step of uniform points.

Optimal points: The set of reference points from the second step of the proposed approach is used.

4.2. Testing datasets

A uniform distribution of points was used for the testing data. The given dataset contains 200 points uniformly distributed in the interval $[0, 1]$. Moreover, each point from this dataset is associated with a function value at this point. For this purpose, many different functions have been used. Results for some representative functions are presented below.

$$f_1(x) = \sin\left(15x^2\right) + 5x \tag{26}$$

$$f_2(x) = (4.88x - 1.88) \cdot \sin\left((4.88x - 1.88)^2\right) + 1 \tag{27}$$

$$f_3(x) = e^{10x-6} \cdot \sin\left((5x-2)^2\right) + (3x-1)^3 \tag{28}$$

4.3. Experimental Results and Comparisons

The experimental results for the proposed approach will be presented and their comparison with results for another RBF approximation using the constant shape parameter for different distributions of reference points will be made. The shape parameters α for the RBFs used, in the case of approximation with the constant shape parameter, were determined experimentally with regard to the quality of the approximation, i.e. they were selected the shape parameters α for which the lowest mean absolute error of the approximation was obtained. Moreover, the RBF approximation using the constant shape parameter was applied in two ways. The first one is that the RBF approximation with the constant shape parameter was performed for the original input data. The second one is that the original input data was preprocessed and then the RBF approximation with the constant shape parameter was applied. This preprocessing consists of the application of the first step from the proposed approach to the original input data, i.e. the RBF approximation using the constant shape parameter is applied to shifted data. The setups for presented experiments are presented in Table 2.

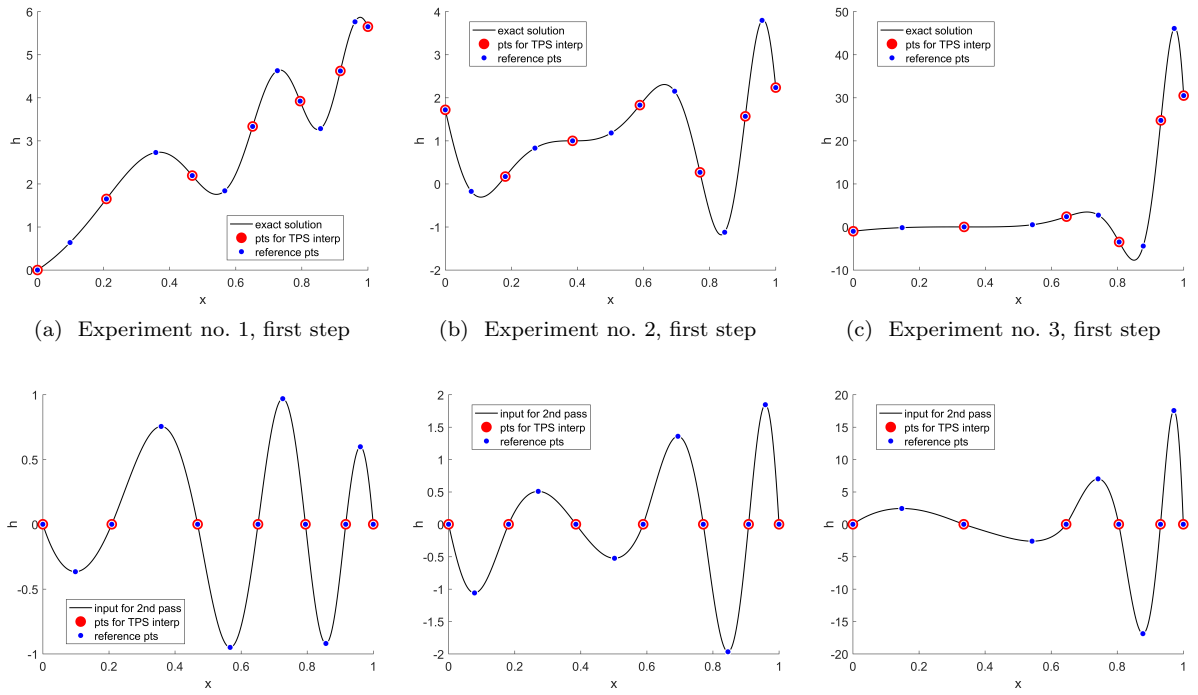
The trends of the original data for the different experiments from Table 2 are shown in Fig. 6 (top). Figure 7 presents the results for the different experiments in which each point is associated with a value from some sampling function (26) - (28) and some RBF from Table 1 is used. The specific choice of the sampling function and RBF for each experiment is mentioned in Table 2. Using the proposed approach, M significant points (see Table 2) were found for the chosen datasets and M_1 of them (see Table 2) were classified as inflection points or endpoints. These M_1 significant points were used for the TPS interpolation in the first step. The trends of the data after the first step, i.e. after performing the shift of associated values, can be seen in Fig. 6 (bottom) for the different experiments from Table 2. The points for the TPS interpolation from the first step and reference points used for the second step for the different experiments are visualized in Fig. 6 (bottom) on the shifted data and in Fig. 6 (top) on the original data.

Figure 7 (left), in addition to the RBF approximation using the proposed approach, also presents the results of the RBF approximation using the constant shape parameter, where the set of reference points has uniform, epsilon or optimal distribution and contains M points (see Table 2). In this case, the RBF approximation using the constant shape parameter was applied to the original input data, i.e. the preprocessing is not included. The magnitudes of error for these approximations can be seen in Fig. 7 (right). The error is defined as the absolute value of the difference between the sampling function, some of the equations (26) - (28), and the approximated function. The differences of frequencies of errors for different experiments are shown in Fig. 8. Moreover, the three basic error measures (mean absolute error, deviation of error and mean relative error) for all experiments men-

Table 2

Experimental setups - $\{h_i\}_{i=1}^{N}$ indicates the sampling function of the associated values, N is size of input dataset, M_1 is number of interpolation points for the first step, M is number of reference points for the second step and $\phi(r)$ is RBF used.

	$\{h_i\}_{i=1}^{N}$	N	M_1	M	$\phi(r)$
Experiment no. 1	f_1, eq. (26)	200	7	13	Gaussian RBF
Experiment no. 2	f_2, eq. (27)	200	7	13	Inverse quadric
Experiment no. 3	f_3, eq. (28)	200	6	11	Wendland's $\phi_{3,2}$



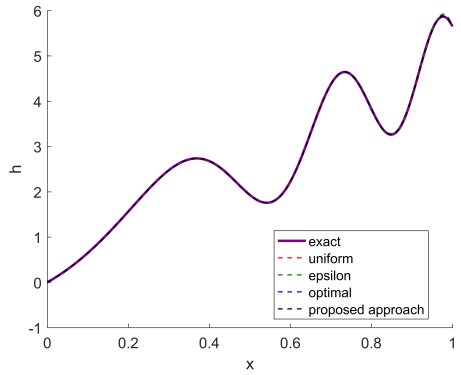
(a) Experiment no. 1, first step (b) Experiment no. 2, first step (c) Experiment no. 3, first step

(d) Experiment no. 1, second step (e) Experiment no. 2, second step (f) Experiment no. 3, second step

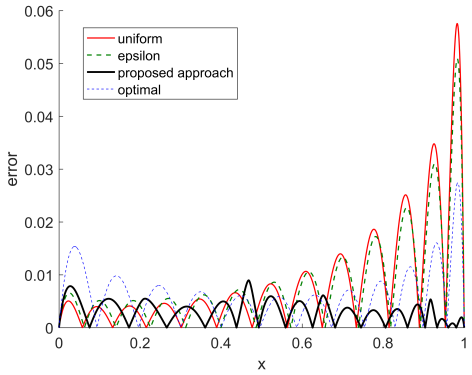
Fig. 6. Trends of input data for the first step (top) and trends of input data for the second step (bottom) for different experiments, see Table 2. The sets of reference points for both steps are marked.

Table 3
The RBF approximation error for the tested datasets and different initial configurations.

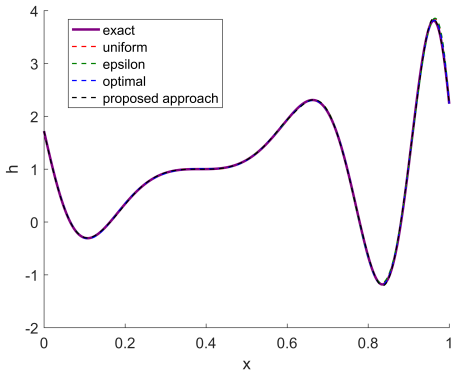
Phenomenon	**proposed approach**	**original data**			**shifted data**		
		uniform ref. pts.	**epsilon ref. pts.**	**optimal ref. pts.**	**uniform ref. pts.**	**epsilon ref. pts.**	**optimal ref. pts.**
Experiment no. 1							
mean absolute error	3.13E-03	8.55E-03	8.20E-03	6.21E-03	1.23E-02	1.22E-02	9.09E-03
deviation of error	3.71E-06	1.01E-04	7.56E-05	2.11E-05	1.74E-04	1.62E-04	1.48E-04
mean relative error	1.17E-03	3.18E-03	3.05E-03	2.31E-03	4.59E-03	4.54E-03	3.38E-03
Experiment no. 2							
mean absolute error	9.49E-03	1.80E-02	1.77E-02	1.18E-02	2.23E-02	2.23E-02	1.52E-02
deviation of error	9.94E-05	5.24E-04	6.09E-04	1.37E-04	5.51E-04	6.06E-04	1.21E-04
mean relative error	7.94E-03	1.51E-02	1.48E-02	9.83E-03	1.86E-02	1.86E-02	1.27E-02
Experiment no. 3							
mean absolute error	7.52E-02	2.04E+00	1.98E+00	8.12E-02	2.14E+00	2.37E+00	1.45E-01
deviation of error	8.62E-03	1.37E+01	1.48E+01	2.18E-02	1.67E+01	1.58E+01	4.79E-02
mean relative error	1.75E-02	4.74E-01	4.61E-01	1.89E-02	4.98E-01	5.52E-01	3.37E-02



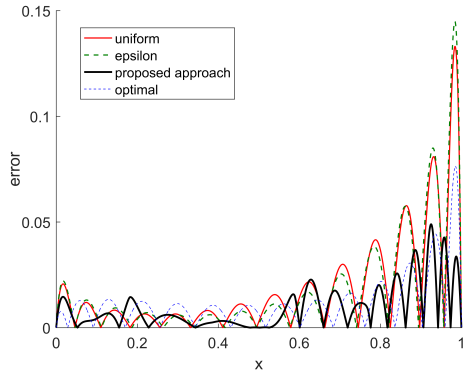
(a) Results for experiment no. 1



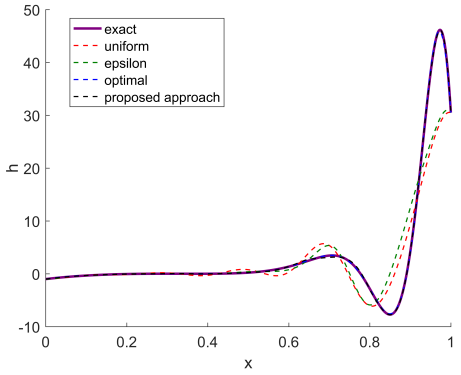
(b) Absolute errors for experiment no. 1



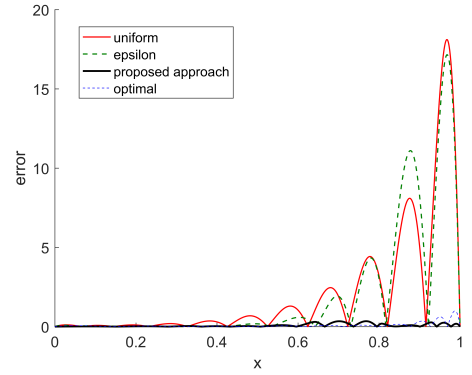
(c) Results for experiment no. 2



(d) Absolute errors for experiment no. 2



(e) Results for experiment no. 3

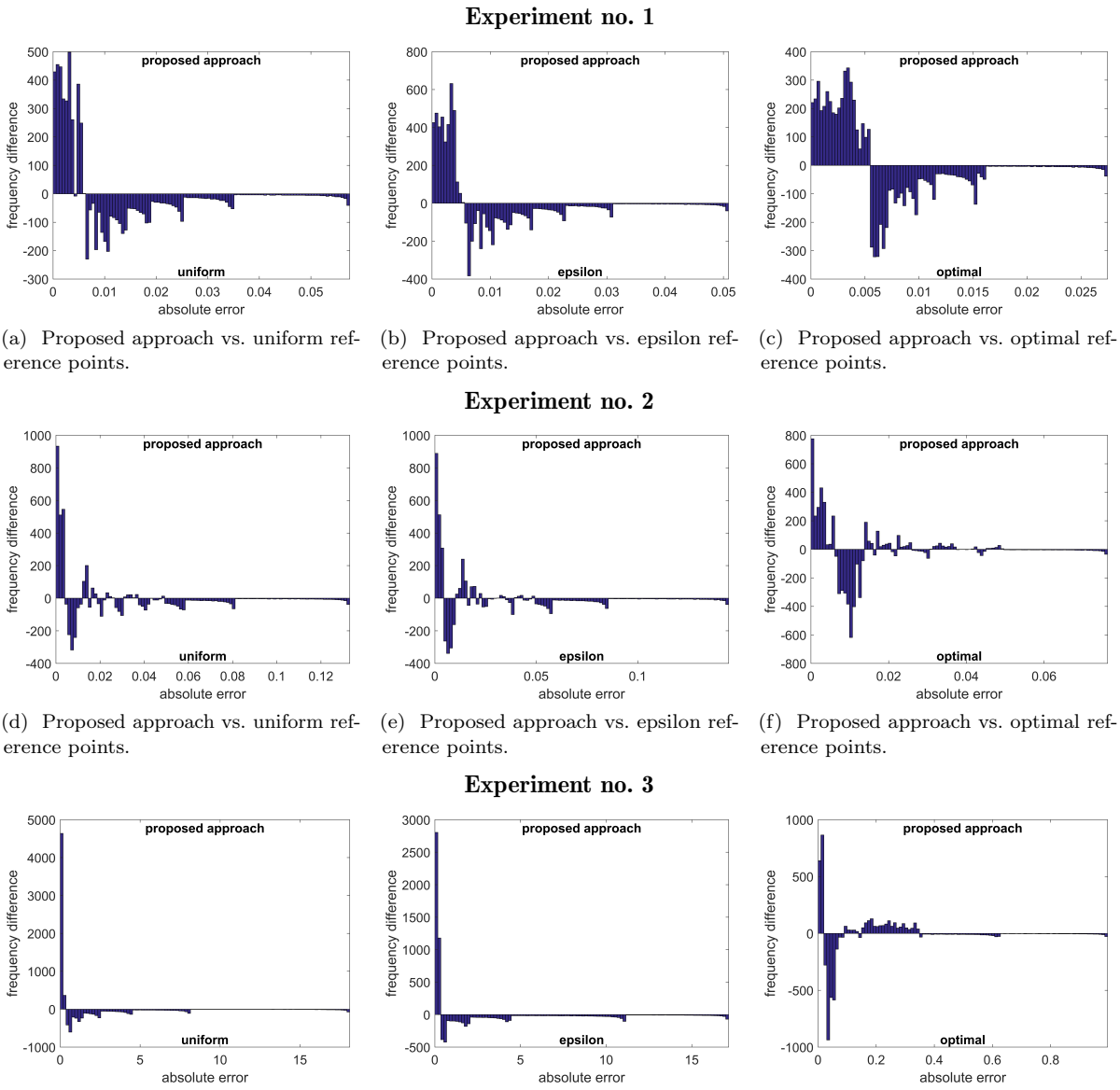


(f) Absolute errors for experiment no. 3

Fig. 7. Results of the RBF approximation (left) and their errors (right) for different initial configurations and different experiments, see Table 2. The RBF approximation with the constant shape parameter is applied to the original data, i.e. the preprocessing is not included. The initial configurations are: constant shape parameter and uniform reference points (uniform), constant shape parameter and epsilon reference points (epsilon), constant shape parameter and optimal reference points (optimal) and the proposed approach (note that values of the proposed method are nearly equal to the exact ones).

tioned are shown in Table 3. It can be observe that the proposed approach returns better results than the other methods in terms of the error.

The results of comparison of the proposed approach with the RBF approximation using the constant shape parameter for different distributions of the set of reference points and different exper-



(a) Proposed approach vs. uniform reference points.

(b) Proposed approach vs. epsilon reference points.

(c) Proposed approach vs. optimal reference points.

(d) Proposed approach vs. uniform reference points.

(e) Proposed approach vs. epsilon reference points.

(f) Proposed approach vs. optimal reference points.

(g) Proposed approach vs. uniform reference points.

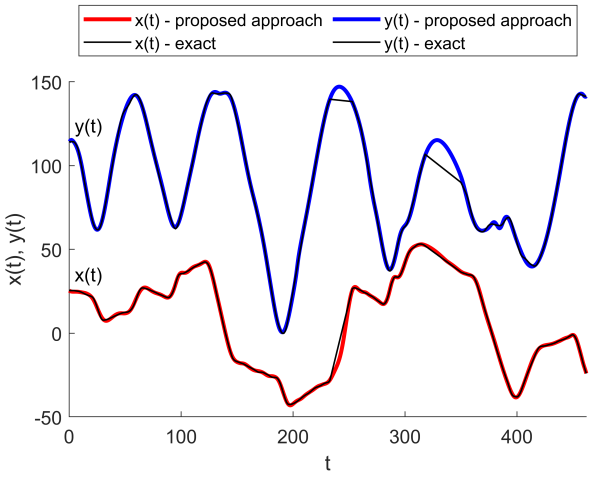
(h) Proposed approach vs. epsilon reference points.

(i) Proposed approach vs. optimal reference points.

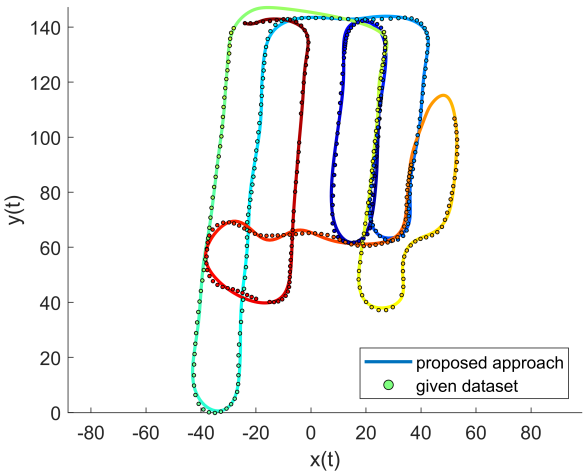
Fig. 8. Difference histograms of approximation errors for different experiments, see Table 2. The RBF approximation with the constant shape parameter is applied to the original data, i.e. the preprocessing is not included.

iments from Table 2 when the preprocessing was used, i.e. the approximation was applied to shifted data, see Fig. 6 (bottom), have the similar visual results as when the preprocessing is not included, and therefore, these experiments are presented only by the three basic error measures, see Table 3.

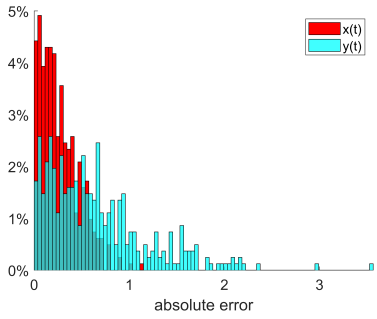
The proposed algorithm was applied on data for recovering smooth robot trajectory in the space which can be computed as the curve parameterized by time. Description of results for this experiment follows. The two separate functions, $x(t)$ and $y(t)$, each representing its respective coordinate and depending on time t on which the proposed algorithm was used, are shown in Fig. 9 (a). These two functions are represented the parametric curve which is presented together with the original data in Fig. 9 (b). The histograms of absolute errors for both functions $x(t)$ and $y(t)$ when the proposed ap-



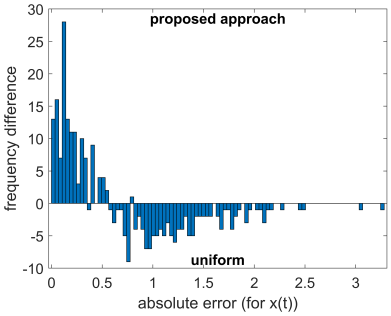
(a) Results of proposed approach - parametric representation



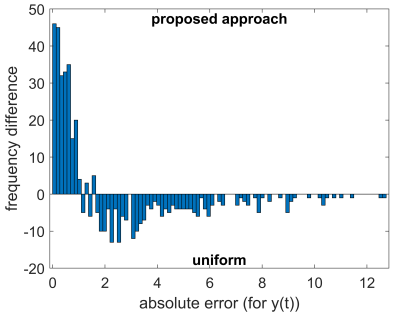
(b) Results of proposed approach



(c) Histogram of absolute errors for proposed approach



(d) Curve x(t) - proposed approach vs. uniform reference points



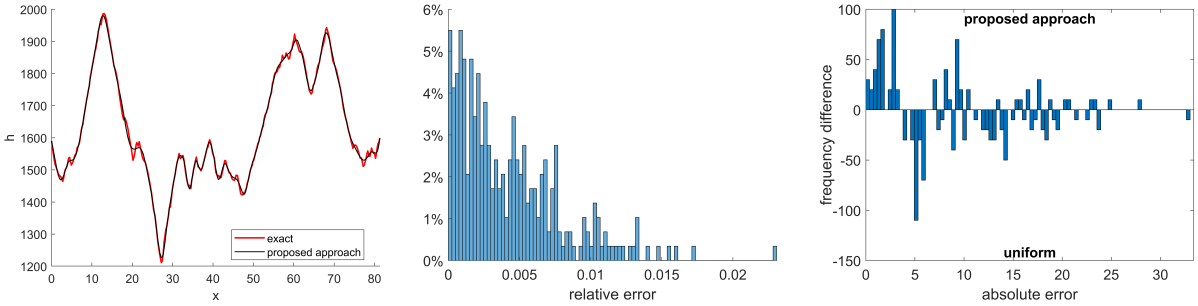
(e) Curve y(t) - proposed approach vs. uniform reference points

Fig. 9. Recovering the smooth robot trajectory from real data [40] using the proposed RBF approximation ($N = 407$ - number of given points for both curves $x(t)$ and $y(t)$, $M_{1x} = 38$ and $M_{1y} = 28$ - number of points for the TPS interpolation, $M_x = 63$ and $M_y = 50$ - number of reference points for the second step of the proposed algorithm.

proach is used can be seen in Fig. 9 (c). Finally, the differences of frequencies of errors for comparison with the RBF approximation using the constant shape parameter for uniform distribution of set of reference points is presented in Fig. 9 (d) - (e). It can be seen that the proposed approach returns better results than the other methods in terms of the error. Moreover, the proposed algorithm is able to reconstruct and smoothly connect a path even if data is missing for a certain period of time. It should be noted that the smooth connection of path is the key property for robot path planning.

Further, the application of the proposed approach on real dataset which represents the terrain profile (2711 points) was performed and the comparison with the RBF approximation using the constant shape parameter for uniform distribution of the set of reference points is presented, see Fig. 10. It can be seen that the proposed approach can well approximate the global trend of terrain profile for a small set of reference points. The further improving of result could be obtained e.g. by application of some incremental method. Moreover, it can be seen that the proposed approach returns again better results than the other methods in terms of the error.

It can be concluded that the RBF approximation for which the distribution of the set of reference points does not reflect the features of the approximated data and the number of reference points is minimal in terms of usability, i. e. uniform and epsilon distribution in our experiments, returns much worse results than the approximation for which features of the given data are reflected.



(a) Results of proposed approach. (b) Histogram of relative errors for proposed approach. (c) Proposed approach vs. uniform reference points.

Fig. 10. Result of the proposed RBF approximation for real dataset which represents the terrain profile ($N = 2711$ - number of given points, $M_1 = 30$ - number of points for the TPS interpolation and $M = 55$ number of reference points for the second step of the proposed algorithm).

5. Conclusion

In this paper, a new algorithm for the radial basis function (RBF) approximation of functions $y = f(x)$ with the variable and adaptive shape parameter based on curve curvature behavior is presented. The proposed method has two steps based on exploiting features of the given dataset, such as extreme points and inflection points. The first step of the proposed approach is applying the global RBF interpolation of the selected subset of significant points, which leads to an adaptive shift of the given data in terms of associated values. After that, the RBF approximation with the variable shape parameter is performed on modified data. The set of reference points is derived using significant points of the shifted data and the variable shape parameters are determined according to the first curvature in them.

The experiments proved that the proposed method gives significantly better results over other relevant competing methods. Moreover, it can be observed that the RBF approximation for which features of the given data are not respected is not capable of competing with the RBF approximation respecting data features, especially when the number of reference points is small.

The proposed method significantly eliminates problems with a shape parameter estimation inherited from the RBF's general properties. The proposed algorithm can be used for the RBF approximation of a curve which is parameterized by one variable in multidimensional space. In future, the proposed approach is to be extended for explicit functions of two variables, i.e. to higher dimensions.

Acknowledgments

The authors would like to thank their colleagues at the University of West Bohemia, Plzeň, for their discussions and suggestions. The research was supported by the Czech Science Foundation GAČR project GA17-05534S and partially supported by the SGS 2019-016 project.

References

[1] Hardy RL. Multiquadratic Equations of Topography and Other Irregular Surfaces. Journal of Geophysical Research. 1971;76:1905–1915.

[2] Hardy RL. Theory and applications of the multiquadric-biharmonic method 20 years of discovery 19681988. Computers & Mathematics with Applications. 1990;19(8):163–208.

[3] Pepper DW, Rasmussen C, Fyda D. A meshless method using global radial basis functions for creating 3-D wind fields from sparse meteorological data. Computer Assisted Methods in Eng & Science. 2014;21(3-4):233–243.

[4] Carr JC, Beatson RK, Cherrie JB, Mitchell TJ, Fright WR, McCallum BC, et al. Reconstruction and representation of 3D objects with radial basis functions. In: SIGGRAPH 2001. ACM; 2001. p. 67–76.

[5] Majdisova Z, Skala V. Big geo data surface approximation using radial basis functions: A comparative study. Computers & Geosciences. 2017;109:51–58.

[6] Smolik M, Skala V. Large scattered data interpolation with radial basis functions and space subdivision. Integrated Computer-Aided Engineering. 2018;25(1):49–62.

[7] Smolik M, Skala V, Majdisova Z. Vector field radial basis function approximation. Advances in Engineering Software. 2018;123:117 – 129.

[8] Hon YC, Šarler B, Yun DF. Local radial basis function collocation method for solving thermo-driven fluid-flow

problems with free surface. Engineering Analysis with Boundary Elements. 2015;57:2–8.

[9] Li M, Chen W, Chen CS. The localized RBFs collocation methods for solving high dimensional PDEs. Engineering Analysis with Boundary Elements. 2013;37(10):1300–1304.

[10] Wendland H. Computational aspects of radial basis function approximation. Studies in Computational Mathematics. 2006;12:231–256.

[11] Fasshauer GE, Zhang JG. On choosing "optimal" shape parameters for RBF approximation. Numerical Algorithms. 2007;45(1-4):345–368.

[12] Gherlone M, Iurlaro L, Sciuva MD. A novel algorithm for shape parameter selection in radial basis functions collocation method. Composite Structures. 2012;94(2):453–461.

[13] Huang CS, Lee CF, Cheng AHD. Error estimate, optimal shape factor, and high precision computation of multiquadric collocation method. Engineering Analysis with Boundary Elements. 2007;31(7):614 – 623.

[14] Scheuerer M. An alternative procedure for selecting a good value for the parameter c in RBF-interpolation. Adv Comput Math. 2011;34(1):105–126.

[15] Zhu S, Wathen AJ. Convexity and Solvability for Compactly Supported Radial Basis Functions with Different Shapes. Journal of Scientific Computing. 2015;63(3):862–884.

[16] Bozzini M, Lenarduzzi L, Rossini M, Schaback R. Interpolation with variably scaled kernels. IMA Journal of Numerical Analysis. 2015;35(1):199–219.

[17] Afiatdoust F, Esmaeilbeigi M. Optimal variable shape parameters using genetic algorithm for radial basis function approximation. Ain Shams Engineering Journal. 2015;6(2):639 – 647.

[18] Sanyasiraju YVSS, Satyanarayana C. On optimization of the RBF shape parameter in a grid-free local scheme for convection dominated problems over non-uniform centers. Applied Mathematical Modelling. 2013;37(12):7245 – 7272.

[19] Bozzini M, Lenarduzzi L, Schaback R. Adaptive Interpolation by Scaled Multiquadrics. Advances in Computational Mathematics. 2002;16(4):375–387.

[20] Kansa EJ, Carlson RE. Improved accuracy of multiquadric interpolation using variable shape parameters. Computers & Mathematics with Applications. 1992;24(12):99 – 120.

[21] Cheng J, Zhang G, Caraffini F, Neri F. Multicriteria adaptive differential evolution for global numerical optimization. Integrated Computer-Aided Engineering. 2015;22(2):103–107.

[22] Rostami S, Neri F. Covariance matrix adaptation pareto archived evolution strategy with hypervolume-sorted adaptive grid algorithm. Integrated Computer-Aided Engineering. 2016;23(4):313–329.

[23] Rostami S, Neri F, Epitropakis M. Progressive preference articulation for decision making in multi-objective optimisation problems. Integrated Computer-Aided Engineering. 2017;24(4):315–335.

[24] Biazar J, Hosami M. Selection of an interval for variable shape parameter in approximation by radial basis functions. Advances in Numerical Analysis. 2016;2016.

[25] Ranjbar M. A new variable shape parameter strategy for Gaussian radial basis function approximation methods. Annals of the University of Craiova-Mathematics and Computer Science Series. 2015;42(2):260–272.

[26] Martynova M. A Novel Approach of the Approximation by Patterns Using Hybrid RBF NN with Flexible Parameters. In: CoMeSySo 2018. Springer; 2019. p. 225–235.

[27] Adeli H, Karim A. Fuzzy-wavelet RBFNN model for freeway incident detection. Journal of Transportation Engineering. 2000;126(6):464–471.

[28] Karim A, Adeli H. Comparison of Fuzzy-Wavelet Radial Basis Function Neural Network Freeway Incident Detection Model with California Algorithm. Journal of Transportation Engineering. 2002;128(1):21–30.

[29] Karim A, Adeli H. Radial basis function neural network for work zone capacity and queue estimation. Journal of Transportation Engineering. 2003;129(5):494–503.

[30] Ghosh-Dastidar S, Adeli H, Dadmehr N. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. IEEE Tran on Biomedical Eng. 2008;55(2):512–518.

[31] Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Transactions on Neural Networks. 1991;2(2):302–309.

[32] Chen M, Xu M, Franti P. Compression of GPS Trajectories. In: DCC 2012. IEEE; 2012. p. 62–71.

[33] Aleotti J, Caselli S, Maccherozzi G. Trajectory reconstruction with NURBS curves for robot programming by demonstration. In: 2005 International Symposium on Computational Intelligence in Robotics and Automation. IEEE; 2005. p. 73–78.

[34] Tian Y, Hu C, Dong X, Zeng T, Long T, Lin K, et al. Theoretical Analysis and Verification of Time Variation of Background Ionosphere on Geosynchronous SAR Imaging. IEEE Geoscience and Remote Sensing Letters. 2015;12(4):721–725.

[35] Majdisova Z, Skala V. Radial basis function approximations: comparison and applications. Applied Mathematical Modelling. 2017;51:728–743.

[36] Skala V. Fast Interpolation and Approximation of Scattered Multidimensional and Dynamic Data Using Radial Basis Functions. WSEAS Transactions on Mathematics. 2013;12(5):501–511.

[37] Skala V. RBF Approximation of Big Data Sets with Large Span of Data. In: MCSI 2017. IEEE; 2017. p. 212–218.

[38] Fornberg B, Zuev J. The Runge phenomenon and spatially variable shape parameters in RBF interpolation. Computers & Mathematics with Applications. 2007;54(3):379 – 398.

[39] Skala V. RBF Interpolation with CSRBF of Large Data Sets. Procedia Computer Science. 2017;108:2433 – 2437. ICCS 2017.

[40] Blanco JL, Moreno FA, González J. A Collection of Outdoor Robotic Datasets with centimeter-accuracy Ground Truth. Autonomous Robots. 2009;27(4):327–351.

Appendix G

Determination of Stationary Points and Their Bindings in Dataset Using RBF Methods

Majdišová, Z., Skala, V., Šmolík, M.

Computational and Statistical Methods in Intelligent Systems, 2nd Computational Methods in System and Software 2018, Volume 859 of Advances in Intelligent Systems and Computing series, pp. 213-224, Springer (2019), ISSN 2194-5357, ISBN 978-3-030-00211-4

Determination of Stationary Points and Their Bindings in Dataset Using RBF Methods

Zuzana Majdisova(✉), Vaclav Skala, and Michal Smolik

Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 30614 Plzeň, Czech Republic
{majdisz,smolik}@kiv.zcu.cz
http://www.vaclavskala.eu

Abstract. Stationary points of multivariable function which represents some surface have an important role in many application such as computer vision, chemical physics, etc. Nevertheless, the dataset describing the surface for which a sampling function is not known is often given. Therefore, it is necessary to propose an approach for finding the stationary points without knowledge of the sampling function.

In this paper, an algorithm for determining a set of stationary points of given sampled surface and detecting the bindings between these stationary points (such as stationary points lie on line segment, circle, etc.) is presented. Our approach is based on the piecewise RBF interpolation of the given dataset.

Keywords: Stationary points · RBF interpolation
Shape parameter · Shape detection · Nearest neighbor

1 Introduction

Stationary points of the given explicit function $f(\boldsymbol{x})$ are points where the gradient of the function $f(\boldsymbol{x})$ is zero in all directions, i.e. all partial derivatives are zero:

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= \mathbf{0} \qquad \boldsymbol{x} \in \mathbb{E}^n, \text{ i.e.} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_k} &= 0 \qquad k = 1, \ldots, n, \end{aligned} \tag{1}$$

where n denotes the dimension of space. The knowledge of stationary points is required in many areas that are used a multidimensional data analysis, e.g. [1–5]. The significant features of the given dataset can be determined using the set of stationary points. This properties can be further used for improving the quality of the RBF approximation [6,7], etc. In the technical applications, the sampling function is not often known and only the dataset describing the given surface is specified. Therefore, it is necessary determining the stationary points without knowledge of the sampling function. Moreover, for a higher dimension

R. Silhavy et al. (Eds.): CoMeSySo 2018, AISC 859, pp. 213–224, 2019.
https://doi.org/10.1007/978-3-030-00211-4_20

of space $n \geq 2$, it is possible that the stationary points of given surface are not only isolated but they can be formed into line segments, circles or some other shapes. A new approach for searching of bindings between stationary points will be described in this paper. Knowledge of these bindings is suitable, for example, for pruning purposes.

In the following sections, the fundamental the RBF interpolation will be described. The finding of stationary points of surface using the RBF interpolation will be described in Sect. 3. Moreover, the method, how the bindings between stationary points are searching, is introduced in this section. In the section Sect. 4, the results of our proposed algorithm will be presented. Finally, a final discussion of results will be performed.

2 RBF Interpolation

In this section, the RBF interpolation method, recently introduced, e.g. in [8,9], and its properties are described.

We assume that we have an unordered dataset $\{\boldsymbol{x}_i\}_1^N \in \mathbb{E}^n$, where n denotes the dimension of space and N is the number of given points. Further, each point $\boldsymbol{x}_i$ from the dataset is associated with a vector $\boldsymbol{h}_i \in \mathbb{E}^p$ of the given values, where p is the dimension of the vector, or a scalar value, i.e. $h_i \in \mathbb{E}^1$. In the following, we will deal with scalar data interpolation, i.e. the case when each point $\boldsymbol{x_i}$ is associated with a scalar value h_i is considered. Our goal is determined the unknown function which is sampled at given points $\{\boldsymbol{x}_i\}_1^N$ by values $\{h_i\}_1^N$. For these purposes, it can be used the RBF interpolation which is based on the distance computation between two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ from the given dataset.

The interpolated value can be determined as:

$$f(\boldsymbol{x}) = \sum_{j=1}^{N} c_j \phi(r_j) = \sum_{j=1}^{N} c_j \phi\left(\|\boldsymbol{x} - \boldsymbol{x}_j\|_2\right), \tag{2}$$

where the interpolating function $f(\boldsymbol{x})$ is represented as a sum of N RBFs, each centered at a different data points $\boldsymbol{x}_j$ and weighted by an appropriate weight c_j which has to be determined, see Fig. 1.

Applying (2) for all data points $\boldsymbol{x}_i, i = 1, \ldots, N$, we get a linear system of equations:

$$h_i = f(\boldsymbol{x}_i) = \sum_{j=1}^{N} c_j \phi\left(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2\right) \quad i = 1, \ldots, N. \tag{3}$$

The linear system of equations can be represented in a matrix form as:

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{h}, \tag{4}$$

where the matrix $\boldsymbol{A} = \{A_{ij}\} = \{\phi\left(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2\right)\}$ is $N \times N$ symmetric square interpolation matrix, the vector $\boldsymbol{c} = (c_1, \ldots, c_N)^T$ is the vector of unknown weights and $\boldsymbol{h} = (h_1, \ldots, h_N)^T$ is a vector of values in the given points. This

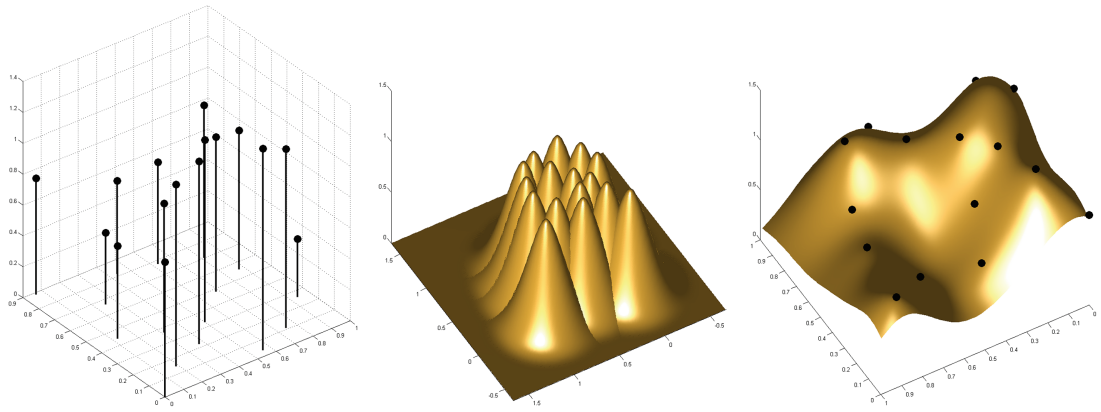


Fig. 1. Data values, the RBF collocation functions, the resulting interpolant.

linear system of equations can be solved by the Gauss elimination method, the LU decomposition, etc.

From the above, it can be seen that, in order to solve the interpolation problem, the distance matrix and a radial basis expansion are used.

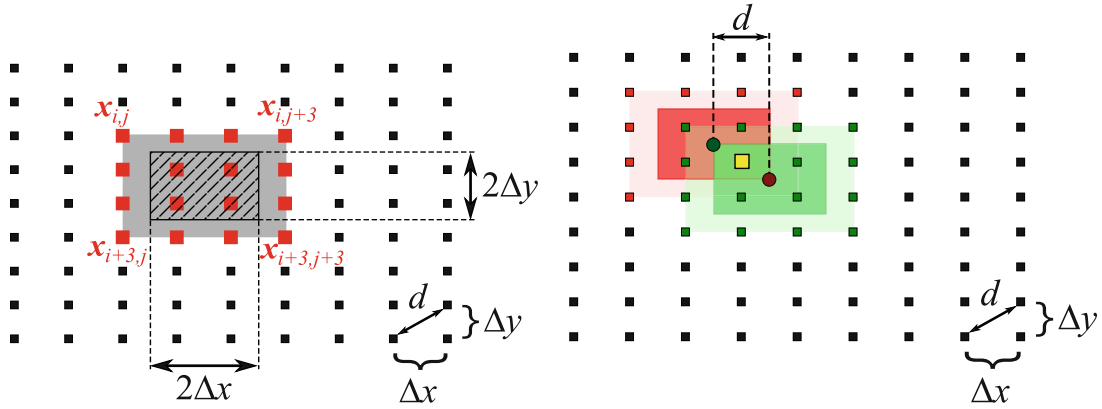
3 Proposed Approach

In this section, determination of stationary points of the given dataset is described. Moreover, the approach includes the method for searching of bindings between stationary points because whole shape of stationary points may lie on the sampled surface.

3.1 Piecewise Approach for Determination of Stationary Points

For simplicity we assume that we have given dataset $\{\boldsymbol{x}_i\}_1^N \in \mathbb{E}^2$ and each point $\boldsymbol{x}_i$ from this dataset is associated with a scalar value $h_i \in \mathbb{E}^1$. Further, for purposes of determination of stationary points, we assume that the given dataset contains the points on a $N_x \times N_y$ regular grid, where Δx and Δy are real numbers representing its grid spacing. Moreover, the row-major ordering of the given data is performed at first. After that, the piecewise approach is applied on the given data.

The process which is performed at each step of the piecewise approach is following. Every sixteen points $\{\boldsymbol{x}_m\}_1^{16} = \{\boldsymbol{x}_{i,j}, \ldots, \boldsymbol{x}_{i,j+3}, \ldots, \boldsymbol{x}_{i+3,j}, \ldots, \boldsymbol{x}_{i+3,j+3}\}$ from the given dataset, where $i \in \{1, \ldots, N_y - 3\}$ denotes the row index and $j \in \{1, \ldots, N_x - 3\}$ denotes the column index, are interpolated by the RBF interpolation (2), i.e. the linear system (4) has to be solved and the vector of weights $\hat{\boldsymbol{c}} = (c_1, \ldots, c_{16})$ is computed. It mean that during one step of proposed approach, the RBF interpolation for $3\Delta x \times 3\Delta y$ area, where Δx and Δy are real numbers representing the input grid spacing, is performed, see Fig. 2a.



(a) The grey area shows the all points from the given dataset which are interpolated by the RBF method during one step of piecewise approach. The hatched area illustrates the domain for which the stationary points of the given dataset are determined from the obtained RBF interpolation.

(b) Visualization of the stationary points reduction which is performed if the two points are identical or very close to identical. The green circle and red circle mark the stationary points which were determined from two different RBF interpolations and which were merged to one stationary point marked by yellow square.

Fig. 2. Proposed piecewise approach

Then, the stationary points $\{\boldsymbol{s}_q\}$ of this interpolation function are determined using (1). Specifically, for stationary points of the RBF interpolation function the nonlinear system of equations:

$$\mathbf{0} = \sum_{m=1}^{16} c_m \frac{\phi'\left(\|\boldsymbol{x} - \boldsymbol{x}_m\|_2\right)}{\|\boldsymbol{x} - \boldsymbol{x}_m\|_2} * (\boldsymbol{x} - \boldsymbol{x}_m)\,, \tag{5}$$

where $\phi'(r)$ is the derivation of RBF function ϕ with respect to variable r, $*$ denotes the element-wise multiplication and $\hat{\boldsymbol{c}} = (c_1, \ldots, c_{16})$ is the vector of weights, has to be solved. The solution of (5), i.e. the stationary points $\{\boldsymbol{s}_q\}$ of the RBF interpolation, is searched for the domain defined as:

$$\boldsymbol{x}_{i,j} + \boldsymbol{\varepsilon}_{min} \leq \boldsymbol{s}_q \leq \boldsymbol{x}_{i+3,j+3} - \boldsymbol{\varepsilon}_{max},$$

$$\boldsymbol{\varepsilon}_{min} = \begin{cases} \left[\frac{\Delta x}{2}, 0\right] & \text{if } i = 1 \\ \left[0, \frac{\Delta y}{2}\right] & \text{if } j = 1 \\ \left[\frac{\Delta x}{2}, \frac{\Delta y}{2}\right] & \text{otherwise} \end{cases} \quad \boldsymbol{\varepsilon}_{max} = \begin{cases} \left[\frac{\Delta x}{2}, 0\right] & \text{if } i = N_y - 3 \\ \left[0, \frac{\Delta y}{2}\right] & \text{if } j = N_x - 3 \\ \left[\frac{\Delta x}{2}, \frac{\Delta y}{2}\right] & \text{otherwise} \end{cases} \tag{6}$$

where Δx and Δy are real numbers representing the input grid spacing, N_x indicates the number of grid column and N_y is the number of grid rows, see Fig. 2a, and the resulting set is added to the set of stationary points $\{\boldsymbol{s}_l\}$. It should be noted, that the values $\boldsymbol{\varepsilon}_{min}$ and $\boldsymbol{\varepsilon}_{max}$ include the correction for the boundary areas.

The determination of stationary points of a function corresponds to the problem of finding critical points of the vector field, where the vector field is defined

by Eq. (5) for our purposes, and, therefore the method for determining critical points [10,11] may be used for obtaining the result.

The advantage of the above mentioned process is that the matrix $\boldsymbol{A}$ of the linear system (4) for the RBF interpolation, is not dependent on the position of the given points (the matrix is dependent only on the distances between given points) and, therefore, this matrix is constant for all steps of piecewise approach. It should be noted that the approximation by a quadric surface could be used instead of the RBF interpolation, but the experimental results proved that this variant returns worse results in terms of stationary point locations.

The set of stationary points in the current form $\{\boldsymbol{s}_l\}$ may contain two identical points or points very close to identical. This problem is caused by the fact that one stationary point can be obtained from more RBF interpolations. The situation is illustrated in Fig. 2b. However, this problem can be solved by reduction of the set of stationary points. Then, the final set of stationary points $\{\boldsymbol{\sigma}_u\}$ of the given data is determined as follows. The subset S_u of stationary points is removed from the unreduced set of stationary points $\{\boldsymbol{s}_l\}$. The points in the subset S_u meet relation:

$$S_u = \{\boldsymbol{s}_k : \|\boldsymbol{s}_k - \boldsymbol{s}_1\| \leq d\}\,, \tag{7}$$

where $d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ is the diagonal step in the regular grid, and the new stationary point is determined as a centroid of points from subset S_u:

$$\boldsymbol{\sigma}_u = \frac{\sum \boldsymbol{s}_k}{|S_u|}, \tag{8}$$

where $|S_u|$ is a number of points in the subset S_u. The process is repeated until the unreduced set of the stationary points is not empty.

The whole algorithm for determining the stationary points of the given dataset is summarized in Algorithm 1.

3.2 Estimation of Shape Parameter for RBF Interpolation

The piecewise RBF interpolation is used during the process of the determining the stationary points of the given dataset. Nevertheless, the quality of the resulting RBF interpolation strongly depends on the choice of the shape parameter α. Therefore, in this section, the determination of suitable shape parameter α will be performed.

For the above mentioned process, the surface with the least possible tension is required, i.e. the surface must contain as little wavy as possible if the interpolated points allow it. It means that the shape parameter α has to be sufficiently large.

Therefore, for these purposes, we proposed and experimentally verified that shape parameter α is chosen so that the radius of circle of non-stationary inflection points of used RBF function $\phi(r)$ corresponds to the maximum distance of the interpolated points, within one step of proposed piecewise approach, which is $r = 3d$, where $d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ is the diagonal step in the regular grid.

Algorithm 1. Determination of the stationary points $\{\boldsymbol{\sigma}_u\}_1^{N_S}$.

Input: given points $\{\boldsymbol{x}_i\}_1^N$ and their associated scalar values $\{h_i\}_1^N$, size of grid $N_x \times N_y$, grid spacing Δx and Δy, used RBF ϕ and its shape parameter α.
Output: stationary points $\{\boldsymbol{\sigma}_u\}_1^{N_S}$

Row-major ordering the given points $\{\boldsymbol{x}_i\}_1^N$.
$d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$.
Compute matrix $\boldsymbol{A}$ of linear system (4) for the set of points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4, \boldsymbol{x}_{N_x+1}, \ldots, \boldsymbol{x}_{N_x+4}, \boldsymbol{x}_{2N_x+1}, \ldots, \boldsymbol{x}_{2N_x+4}, \boldsymbol{x}_{3N_x+1}, \ldots, \boldsymbol{x}_{3N_x+4}\}$.
for $i = 1, \ldots, N_y - 3$ **do**
    for $j = 1, \ldots, N_x - 3$ **do**
        $\hat{\boldsymbol{x}} = \{\boldsymbol{x}_{(i-1)N_x+j}, \ldots, \boldsymbol{x}_{(i-1)N_x+j+3}, \ldots, \boldsymbol{x}_{(i+2)N_x+j}, \ldots, \boldsymbol{x}_{(i+2)N_x+j+3}\}$
        $\hat{\boldsymbol{h}} = \{h_{(i-1)N_x+j}, \ldots, h_{(i-1)N_x+j+3}, \ldots, h_{(i+2)N_x+j}, \ldots, h_{(i+2)N_x+j+3}\}$
        Compute the vector of unknown weights $\hat{\boldsymbol{c}}$, eq. (4), where $\boldsymbol{h} = \hat{\boldsymbol{h}}$.
        Compute the coefficients $\boldsymbol{\varepsilon}_{min}$ and $\boldsymbol{\varepsilon}_{max}$, eq. (6).
        Determine the stationary points $\{\boldsymbol{s}_q\}$ from eq. (5) in the domain (6).
        $\{\boldsymbol{s}_l\} = \{\boldsymbol{s}_l\} \cup \{\boldsymbol{s}_q\}$
while the set $\{\boldsymbol{s}_k\}$ is not empty **do**
    Find $S_u = \{\boldsymbol{s}_k : \|\boldsymbol{s}_k - \boldsymbol{s}_1\|_2 \leq d\}$ in the set $\{\boldsymbol{s}_l\}$.
    Add the stationary point $\frac{\sum \boldsymbol{s}_k}{|S_u|}$ to the final set of stationary points $\{\boldsymbol{\sigma}_u\}$.
    Delete all points $\boldsymbol{s}_k \in S_u$ from the set $\{\boldsymbol{s}_l\}$.

From this assumption, the following expression for shape parameter α was derived:

$$\alpha = \frac{\omega}{3d}, \tag{9}$$

where $d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ is the diagonal step in the regular grid and ω is a constant parameter depending on the type of used RBF, see Table 1.

Table 1. Different RBFs, their derivation $\phi'(r)$ and their parameter ω, Eq. (9).

RBF	$\boldsymbol{\phi(r)}$	$\boldsymbol{\phi'(r)}$	$\boldsymbol{\omega}$
Gaussian RBF	$e^{-(\alpha r)^2}$	$-2\alpha^2 r e^{-(\alpha r)^2}$	$1/\sqrt{2}$
Inverse quadric	$\left(1 + (\alpha r)^2\right)^{-1}$	$-2\alpha^2 r \left(1 + (\alpha r)^2\right)^{-2}$	$1/\sqrt{3}$
Wendland's $\phi_{3,1}$	$(1 - \alpha r)_+^4 (4\alpha r + 1)$	$-20\alpha^2 r (1 - \alpha r)_+^3$	$1/4$

3.3 Searching of Bindings Between Stationary Points

It is possible that the given surface does not contain only isolated stationary points, but the curves of stationary points, such as line segments, circles, parabolas or some other shapes, can lie on the given surface. Therefore, the method for searching of bindings between stationary points will be described.

At the beginning, the maximal possible distance δ_{max} of two stationary points for which these stationary points still lie on the same curve has to be established. The situation of the worst case is illustrated in Fig. 3. In this figure, it can be

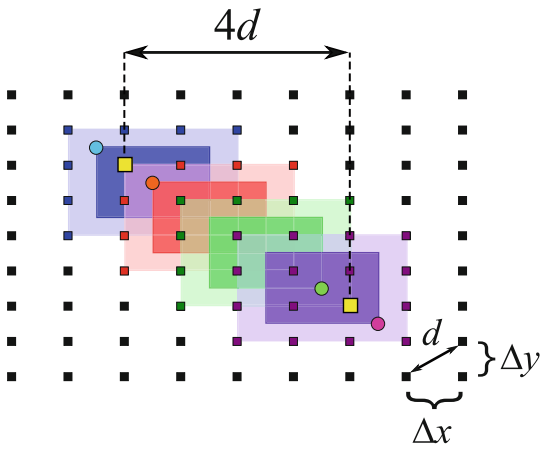


Fig. 3. The figure shows the worst case in which two stationary points (yellow squares) still lie on the same curve of stationary points, i.e. the distance between two stationary points is maximal possible distance. Moreover, the reduction of stationary points is again shown.

seen four subdomains of the piecewise approach and for each of them, the one stationary point is indicated using circle mark. Based on Eq. (7), the stationary points of blue and red subdomains are reduced and are replaced by their centroid. The same case occurs for the green and purple subdomain. New stationary points which are obtained after the reduction are indicates by yellow squares in the figure. It is also obvious that the distance of these two stationary points, which is also the maximum possible distance δ_{max}, is:

$$\delta_{max} = 4d,$$

where $d = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ is the diagonal step in the regular grid.

Now, the stationary points $\{\boldsymbol{\sigma}_u\}$ of the given dataset are sequentially processed by following. For the current stationary point $\boldsymbol{\sigma}_u$, the all stationary points $\{\boldsymbol{\sigma}_w\}$ which lying in the distance δ_{max} are determined:

$$\{\boldsymbol{\sigma}_w\} = \{\boldsymbol{\sigma}_w : \|\boldsymbol{\sigma}_w - \boldsymbol{\sigma}_u\| \leq \delta_{max}\}\,. \tag{10}$$

If no stationary point is found, then the stationary point $\boldsymbol{\sigma}_u$ is isolated. Otherwise, the binding $f_v = \{\boldsymbol{\sigma}_w\} \cup \{\boldsymbol{\sigma}_u\}$ is obtained and newly added stationary points are processed in the same way. Finally, the result of this approach is the set of points described the curve of stationary points. This procedure is repeated until the all stationary points $\{\boldsymbol{\sigma}_u\}$ are processed.

One of the possible solution of this problem is the $kd-$tree which can be simply applied for purposes of searching of bindings between stationary points.

4 Experimental Results

In this section, the experimental results for our proposed approach will be presented and their comparison with the exact stationary points which were determined analytically from the sampling function will be made. The implementation was performed in Matlab. In addition, different radial basis functions have been used, see Table 1.

For purposes of our experiments, a uniform distribution of points on a rectangular domain was used for the testing data. Thus, the given dataset contains 120×120 points uniformly distributed in the interval $[x_{min}, x_{max}] \times [y_{min}, y_{max}]$, where the values x_{min}, x_{max}, y_{min} and y_{max} are chosen based on the used sampling function, see (11a) – (11b) and (12a) – (12d). Moreover, each point from this dataset is associated with a function value of the selected sampling function at this point.

4.1 Comparison of Determined Stationary Points with Exact Stationary Points

In this section, the results for datasets whose stationary points do not contain mutual bindings, i.e. all stationary points are isolated, will be presented. The sampling functions f_1 (11a) and f_2 (11b), which were defined in [12], fulfill these properties.

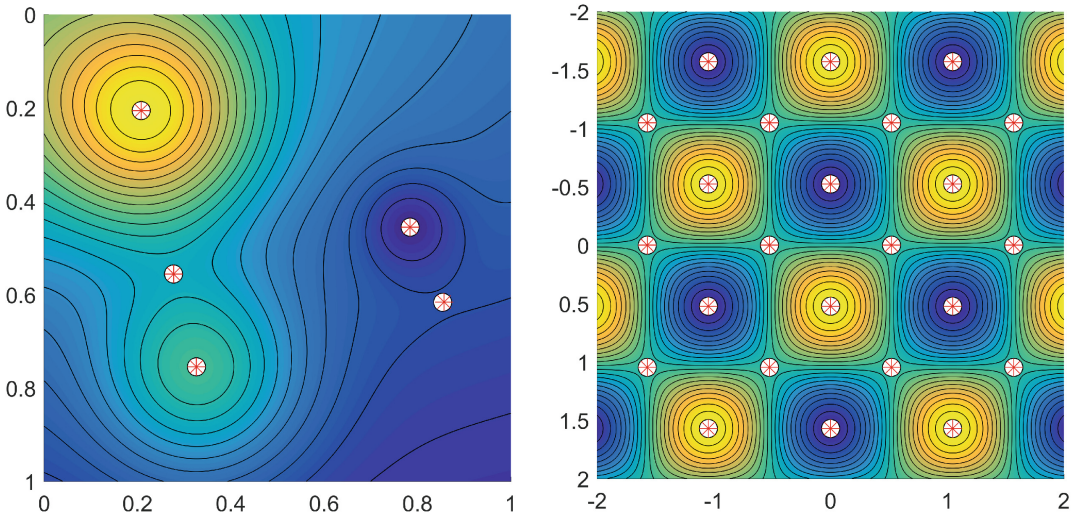
$$f_1(\boldsymbol{x}) = \frac{3}{4}e^{-\frac{(9x_1-2)^2}{4}-\frac{(9x_2-2)^2}{4}} + \frac{3}{4}e^{-\frac{(9x_1+1)^2}{49}-\frac{(9x_2+1)}{10}} + \frac{1}{2}e^{-\frac{(9x_1-7)^2}{4}-\frac{(9x_2-3)^2}{4}} - \frac{1}{5}e^{-(9x_1-4)^2-(9x_2-7)^2} \qquad \boldsymbol{x} \in [0,1] \times [0,1] \quad (11a)$$

$$f_2(\boldsymbol{x}) = \sin(3 \cdot x_1) \cdot \cos(3 \cdot x_2) \qquad \boldsymbol{x} \in [-2,2] \times [-2,2] \quad (11b)$$

Figure 4a presents the results for the dataset in which each point is associated with a value from the f_1 function (11a) when the Gaussian RBF has been used for the piecewise RBF interpolation. Using our proposed approach, five isolated stationary points which are marked by white circles were found for this dataset. The exact stationary points of f_1 function (11a) are shown using the red asterisks $(*)$.

The results for the dataset in which each point is associated with a value from the f_2 function (11b), when the Gaussian RBF has been used for the piecewise RBF interpolation, are presented in Fig. 4b. Twenty four isolated stationary points which are represented by white circles were found for this dataset using our proposed approach. The exact stationary points of f_2 function (11b) are again shown using the red asterisks $(*)$.

It can be seen that obtained results for both datasets correspond to the stationary points calculated analytically from the sampling functions. Moreover, it should be noted, that the same results were obtained even when other RBF function, see Table 1, were used for the piecewise RBF interpolation.



(a) Dataset sampled from f_1, eq.(11a) (b) Dataset sampled from f_2, eq.(11b)

Fig. 4. The white circles indicate the stationary points of the given dataset that are obtained using the proposed approach when the RBF interpolation used the Gaussian RBF. The tested dataset contains 120×120 points. The red asterisks $(*)$ denote the exact positions of the stationary points of the appropriate function. Furthermore, the contour map of given dataset is shown.

4.2 Bindings Between Stationary Points

In this section, the results for datasets whose stationary points contains mutual bindings will be presented. The four following sampling functions (12a) – (12d) fulfill these properties.

$$f_{11}(\boldsymbol{x}) = -\left(x_1 - x_2\right)^2, \qquad \boldsymbol{x} \in [-1,1] \times [-1,1] \tag{12a}$$

$$f_{12}(\boldsymbol{x}) = \sin\left(x_1 + x_2^2\right), \qquad \boldsymbol{x} \in [-3,3] \times [-2,2] \tag{12b}$$

$$f_{13}(\boldsymbol{x}) = \sin\left(3\pi\left(\sqrt{x_1^2 + x_2^2} + 0.25\right)\right), \qquad \boldsymbol{x} \in [-1,1] \times [-1,1] \tag{12c}$$

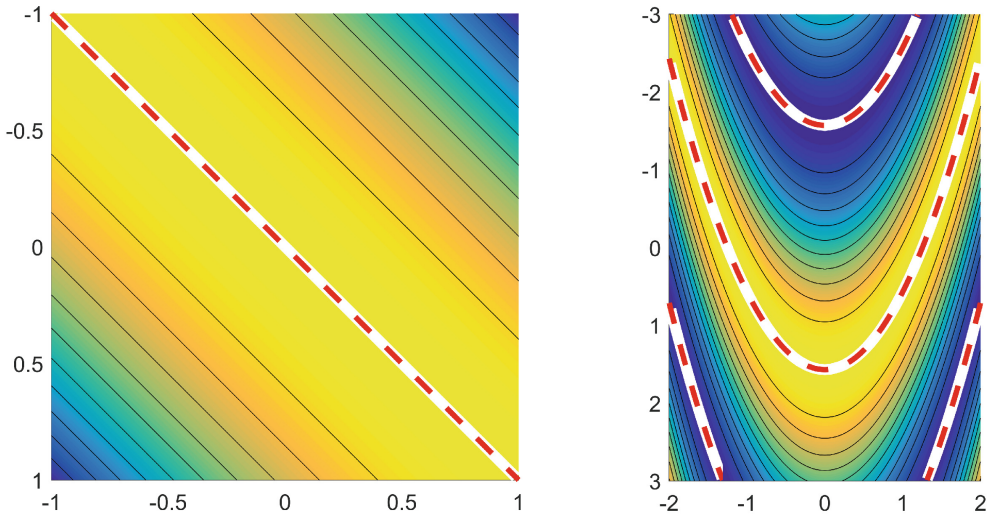
$$f_{14}(\boldsymbol{x}) = -2 \cdot \left(x_1^2 - x_2^2\right)^2 + 1, \qquad \boldsymbol{x} \in [-1,1] \times [-1,1] \tag{12d}$$

At the beginning, it should be noted that the white solid line indicates the curve of stationary points obtained for the given dataset using our proposed approach and the isolated stationary point obtained using our proposed approach is marked by the white circle. The red dashed line indicates the curve of stationary points calculated analytically from the given sampling function and the isolated stationary point calculated analytically from the given sampling function is represented by the red asterisk $(*)$.

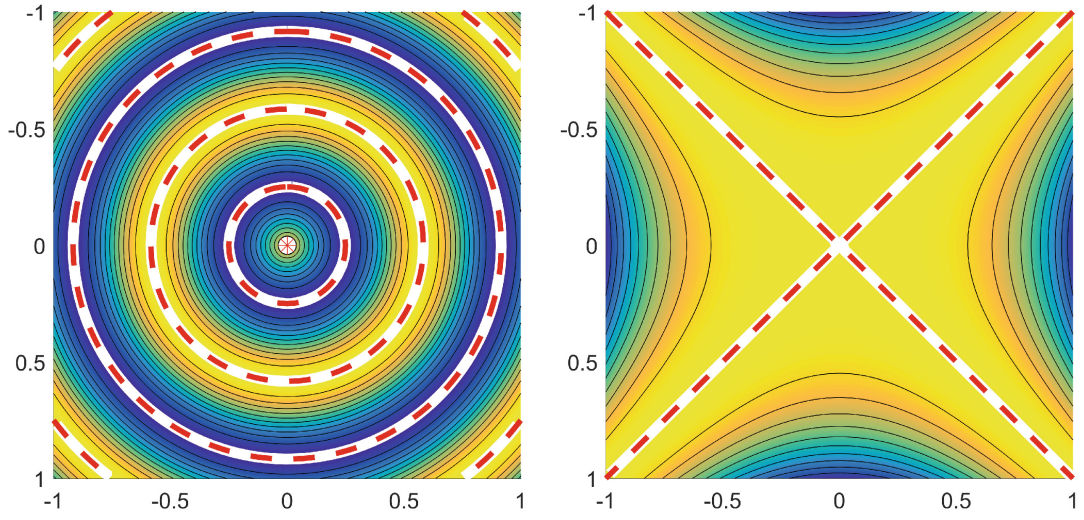
Figure 5a presents the results for the dataset in which each point is associated with a value from the f_{11} function (12a) when the Gaussian RBF has been used for the piecewise RBF interpolation. Using our proposed approach, one curve of stationary points, specifically the line segment, were found for this dataset. This result coincides with the result obtained using analytically approach.

The results for the dataset in which each point is associated with a value from the f_{12} function (12b), when the Gaussian RBF has been used for the piecewise RBF interpolation, are presented in Fig. 5b. For this dataset, the four curves of stationary points, specifically two parabolas and two segments of parabola, were found using our proposed approach.

Figure 5c presents the results for the dataset in which each point is associated with a value from the f_{13} function (12c) when the Gaussian RBF has been used



(a) Dataset sampled from f_{11}, eq.(12a) (b) Dataset sampled from f_{12}, eq.(12b)



(c) Dataset sampled from f_{13}, eq.(12c) (d) Dataset sampled from f_{14}, eq.(12d)

Fig. 5. The white solid lines indicate the curves of stationary points of the given dataset that are obtained using the proposed approach when the RBF interpolation used the Gaussian RBF. The tested dataset contains 120×120 points. The red dashed lines denote the exact curves of the stationary points of the appropriate function. Furthermore, the contour map of given dataset is shown.

for the piecewise RBF interpolation. Using our proposed approach, seven curves of stationary points, specifically three circles and four arcs, and one isolated stationary point were found for this dataset.

The results for the dataset in which each point is associated with a value from the f_{14} function (12d), when the Gaussian RBF has been used for the piecewise RBF interpolation, are presented in Fig. 5d. For this dataset, the two curves of stationary points, specifically two line segments, were found using our proposed approach.

For all mentioned experiments, it can be seen that the results obtained using our proposed approach correspond to results obtained using analytically approach. Moreover, it should be again noted that the same results were obtained even when other RBF function, see Table 1, were used for the piecewise RBF interpolation.

5 Conclusion

In this paper, a new approach for determination of stationary points of given sampled surface without knowledge of the sampling function is presented. The proposed method is based on the piecewise RBF interpolation of the given dataset. Moreover, the proposed approach includes the method of detecting the bindings between the found stationary points, i.e. the approach is able to associate the points from the same curve of stationary points.

The experiments proved that the stationary points determined by our proposed approach coincide with the exact stationary points which were determined analytically form the sampling function.

The results of the proposed approach can, for example, be used for determination of the set of reference points for the RBF approximation which enable appropriate compression of given dataset. The knowledge of the bindings between stationary points is possible to use for pruning the subset of related stationary points to the required number of points on the appropriate curve of stationary points.

In the future work, the proposed approach can be generalized for scattered data using the k-nearest neighbors algorithm.

Acknowledgments. The authors would like to thank their colleagues at the University of West Bohemia, Plzeň, for their discussions and suggestions, and the anonymous reviewers for their valuable comments. Special thanks belong to Jan Dvorak, Lukas Hruda and Martin Červenka for their independent experiments and valuable comments. The research was supported by the Czech Science Foundation GAČR project GA17-05534S and partially supported by the SGS 2016-013 project.

References

1. Banerjee, A., Adams, N., Simons, J., Shepard, R.: Search for stationary points on surfaces. J. Phys. Chem. **89**(1), 52–57 (1985)
2. Tsai, C.J., Jordan, K.D.: Use of an eigenmode method to locate the stationary points on the potential energy surfaces of selected argon and water clusters. J. Phys. Chem. **97**(43), 11227–11237 (1993)
3. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 603–619 (2002)
4. Strodel, B., Wales, D.J.: Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. Chem. Phys. Lett. **466**(4), 105–115 (2008)
5. Liu, Y., Burger, S.K., Ayers, P.W.: Newton trajectories for finding stationary points on molecular potential energy surfaces. J. Math. Chem. **49**(9), 1915–1927 (2011)
6. Majdisova, Z., Skala, V.: Radial basis function approximations: comparison and applications. Appl. Math. Model. **51**, 728–743 (2017)
7. Majdisova, Z., Skala, V.: Big geo data surface approximation using radial basis functions: a comparative study. Comput. Geosci. **109**, 51–58 (2017)
8. Skala, V.: RBF interpolation with CSRBF of large data sets. Procedia Comput. Sci. **108**, 2433–2437 (2017). International Conference on Computational Science, ICCS 2017, 12–14 June 2017, Zurich, Switzerland
9. Smolik, M., Skala, V.: Large scattered data interpolation with radial basis functions and space subdivision. Integr. Comput.-Aided Eng. **25**(1), 49–62 (2018)
10. Bhatia, H., Gyulassy, A., Wang, H., Bremer, P.T., Pascucci, V.: Robust detection of singularities in vector fields. In: Topological Methods in Data Analysis and Visualization III, pp. 3–18. Springer (2014)
11. Wang, W., Wang, W., Li, S.: Detection and classification of critical points in piecewise linear vector fields. J. Vis. **21**, 147–161 (2018)
12. Franke, R.: A critical comparison of some methods for interpolation of scattered data. Technical report NPS53-79-003, Naval Postgraduate School, Monterey, CA (1979)

Appendix H

Incremental Meshfree Approximation of Real Geographic Data

Majdišová, Z., Skala, V., Šmolík, M.

Applied Physics, System Science and Computers III., 3rd International Conference on Applied Physics, System Science and Computers (APSAC2018), Volume 574 of Lecture Notes in Electrical Engineering series, pp.222-228, Springer (2019), ISSN 1876-1100, ISBN 978-3-030-21506-4

Incremental Meshfree Approximation of Real Geographic Data

Zuzana Majdisova(✉), Vaclav Skala, and Michal Smolik

Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 30614 Plzeň, Czech Republic
{majdisz,skala,smolik}@kiv.zcu.cz
http://meshfree.zcu.cz/

Abstract. In many technical applications, reconstruction of the scattered data is often task. For big scattered dataset in n-dimensional space, the using some meshless method such as the radial basis function (RBF) approximation is appropriate. RBF approximation is based on the distance computation, and therefore, it is dimensionally non-separable. This approximation can be converted to an overdetermined linear system of equations which has to be solved.

A new incremental approach for meshless RBF approximation which respects the significant features of the given terrain data such as break lines is proposed in this paper. Using this approach, the improving approximation of the underlying data is achieved. Moreover, the proposed approach leads to a significant compression of the given dataset and the analytical description of the data is obtained. In comparison with other existing methods, the proposed approach achieves the better results due to respecting the features of the given data.

Keywords: RBF approximation · Stationary points · Extrema · Incremental algorithm · TPS · Point clouds

1 Introduction

The most frequent task for many engineering problems is the reconstruction of the given data. There have been developed several algorithms for interpolation or approximation of the given data. Nevertheless, they mostly expect some kind of data ordering, e.g. rectangular mesh, structured mesh, unstructured mesh, etc. This requirement is not necessary when the meshless techniques such as the Radial Basis Function (RBF) methods originally introduced in [1,2] are used. RBF techniques can be used in many fields of the technical or non-technical problems, e.g. reconstruction of surfaces [3–5], visualization of data [6], solving partial differential equations [7,8]. The RBF methods are based on the distance computation between two points and they are independent of the dimension of the space. When the RBF techniques are applied, the given data can be described using analytical formula. The significant compression of the given data is also achieved by using RBF approximation.

K. Ntalianis et al. (Eds.): APSAC 2018, LNEE 574, pp. 222–228, 2019.
https://doi.org/10.1007/978-3-030-21507-1_32

A significant role in terms of the quality of RBF approximation and the compression ratio plays the appropriate placement of the reference points for RBF approximation. In the case of geographic data, this requirement is met for placement along significant features such as ridges, peaks, valleys, etc. A new incremental approach for RBF approximation that puts the emphasis on good placement of reference points and significantly improves the compression ratio will be described in this paper.

In the following sections, the fundamental theoretical background needed for description of the proposed approach will be mentioned. The proposed incremental RBF approximation will be described in Sect. 3. In Sect. 4, the results of our proposed algorithm will be presented. Finally, a final discussion of results will be performed.

2 Theoretical Background

In this section, some theoretical aspects needed for description of the proposed incremental approach will be introduced.

2.1 RBF Approximation

For scattered data processing, the RBF approximation can be used. This technique is based on computing the distance between two points and leads to a solution of linear system of equations which can be solved by singular value decomposition, QR decomposition etc. The RBF approximation is described in [9] or [4] in detail.

2.2 Determination of Stationary Points

Stationary points of an explicit function $f(\boldsymbol{x})$ are points where the gradient of the function $f(\boldsymbol{x})$ is zero vector, i.e. all partial derivatives are zero. In the case, when an analytical explicit expression is not known for the given dataset, the piecewise approach [10] based on RBF interpolation can be used for determination of stationary points in the given dataset.

3 Proposed Approach

In this section, the proposed incremental approach for approximation geographic data using radial basis functions is described.

The main influence on the quality of approximation and the compression ratio has a good placement of the reference points. In case of geographic data, placement along features such as break lines leads to better results. Therefore, in the first level, the set of stationary points obtained for the filtered data using algorithm in [10] is used as set of reference points. The filtered data are determined by applying a Gaussian low-pass filter to the given dataset. The main

reason for filtration of data is a elimination of insignificant stationary points. Moreover, the set of reference points is extended by corners of dataset bounding box due to avoiding problems on the boundary. Now, the RBF approximation (described in Sect. 2.1) is computed and residues $\boldsymbol{r}_1$ are determined. For this purpose, the following equation is used:

$$\boldsymbol{r}_k = |\boldsymbol{h} - f_k(\boldsymbol{X})| \qquad k = 1, \ldots, L, \tag{1}$$

where $\{\boldsymbol{X}, \boldsymbol{h}\} = \{\boldsymbol{x}_i, h_i\}_1^N$ represents the given dataset, $f_k(\boldsymbol{x})$ is approximating function in the k^{th} level and L is number of levels.

In every following level $k > 2$, the residues $\boldsymbol{r}_{k-1}$ are filtered by applying the Gaussian low-pass filter due to eliminating insignificant local maxima. Then, the set of stationary points for filtered residues are determined using algorithm in [10] and only local maxima are added to the set of reference points. Moreover, the uniqueness of the added reference points is checked. When the new set of reference points is obtained, the RBF approximation (described in Sect. 2.1) is again computed and residues $\boldsymbol{r}_k$ are calculated using Eq. (1). The whole process is repeated until the required accuracy of approximation is achieved or the maximum permissible compression ratio is exceeded.

Finally, it should be noted that the value of standard deviation σ_k of Gaussian low-pass filter in k^{th} level is set as:

$$\sigma_k = \begin{cases} \sigma & k = 1, 2 \\ \dfrac{\sigma_{k-1}}{2} & k = 3, \ldots, L \end{cases} \tag{2}$$

where σ is initial value. The whole pseudocode is in Algorithm 1.

Algorithm 1. The incremental RBF approximation of geographic data

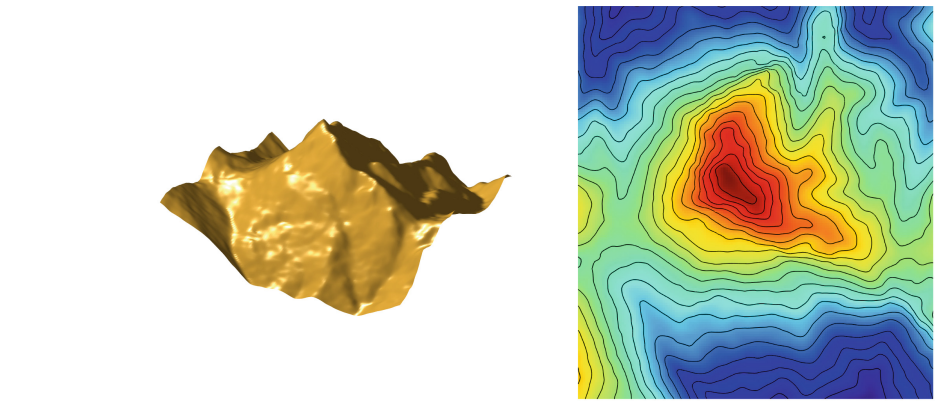
Input: given dataset $\{\boldsymbol{X}, \boldsymbol{h}\} = \{\boldsymbol{x}_i, h_i\}_1^N$, initial value of standard deviation σ for Gaussian low-pass filter, stop conditions c_1 and c_2
Output: approximating function $f_k(\boldsymbol{x})$

$\boldsymbol{h}_f = Gauss(\boldsymbol{h}, \sigma)$ `// Gaussian low-pass filter`
$\boldsymbol{\Xi}$ = Compute stationary points of $\{\boldsymbol{X}, \boldsymbol{h}_f\}$ (using algorithm in [10])
$\boldsymbol{\Xi} = \boldsymbol{\Xi} \cup$ (corners of dataset bounding box)
$f_k(\boldsymbol{x})$ = RBF approximation $(\{\boldsymbol{X}, \boldsymbol{h}\}, \boldsymbol{\Xi})$
$\boldsymbol{r}_k = |\boldsymbol{h} - f_k(\boldsymbol{X})|$
$\sigma_k = \sigma$
while $c_1 || c_2$ **do**
  $\boldsymbol{r}_{kf} = Gauss(\boldsymbol{r}_k, \sigma_k)$ `// Gaussian low-pass filter`
  $\boldsymbol{\Xi}_k$ = Compute stationary points of $\{\boldsymbol{X}, \boldsymbol{r}_{kf}\}$ (using algorithm in [10])
  $\boldsymbol{\Xi} = \boldsymbol{\Xi} \cup$ (only local maxima from $\boldsymbol{\Xi}_k$)
  $f_k(\boldsymbol{x})$ = RBF approximation $(\{\boldsymbol{X}, \boldsymbol{h}\}, \boldsymbol{\Xi})$
  $\boldsymbol{r}_k = |\boldsymbol{h} - f_k(\boldsymbol{X})|$
  $\sigma_k = \sigma_k / 2$

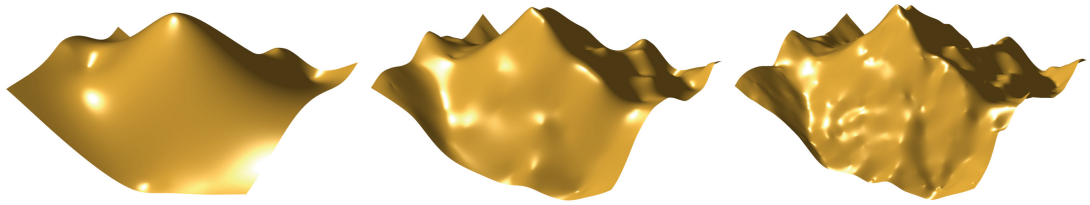
4 Experimental Results

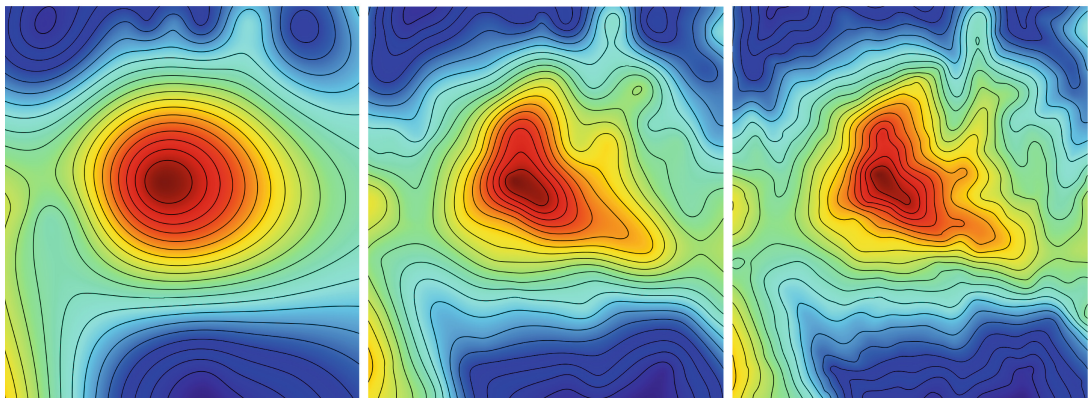
In this section, the experimental results for our proposed approach will be presented. The implementation was performed in Matlab. The thin plate spline (TPS) function $r^2 \log(r^2)$ which is shape parameter free and divergent as radius increases has been used for RBF approximation.

For the purposes of below mentioned experiments, two geographic point clouds were used. The first dataset was obtained from GPS data of the mount Veľký Rozsutec in the Malá Fatra, Slovakia (Fig. 1a) and contains 24,190 points. The second dataset is GPS data of the part of Pennine Alps, Switzerland (Fig. 2a) and contains 131,044 points.



(a) original data, $N = 24,190$



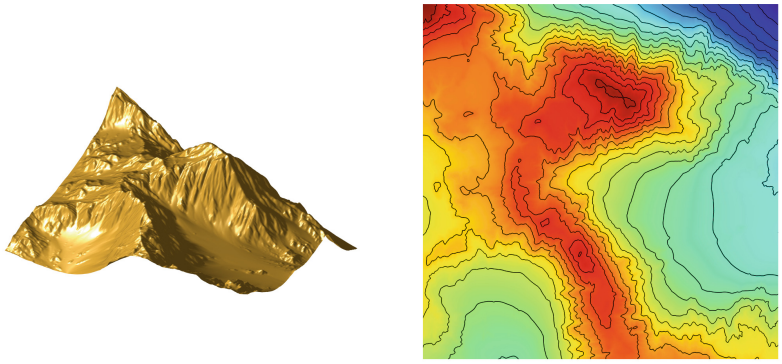


(b) 1^{st} level, $M = 19$ (c) 3^{rd} level, $M = 115$ (d) 8^{th} level, $M = 595$

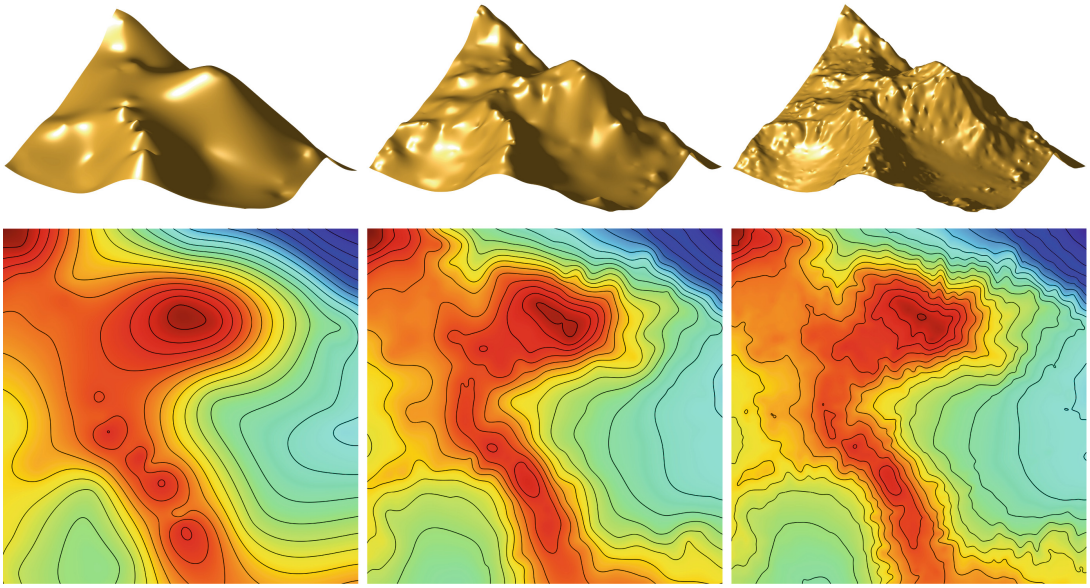
Fig. 1. The mount Veľký Rozsutec, Slovakia and its contour map: original data and different levels of proposed incremental RBF approximation when TPS is used.

Results for different levels of RBF approximation of the mount Veľký Rozsutec are shown in Figs. 1b–d. We can see that the quality of approximation in terms of error is improving with increasing level of the incremental RBF approximation. For 8^{th} level (see Fig. 1d), the many details of the original terrain are already apparent.

In Fig. 2b–d, the results for different levels of incremental RBF approximation of the part of Pennine Alps are shown. It can be again seen that the quality of approximation is improving with increasing level of the incremental approach. For the first level (see Fig. 2b), it is evident, that the small number of reference points is defined for the ridge in the foreground, and therefore, this ridge is approximated by several peaks in the first level. This problem is eliminated with increasing level of the incremental RBF approximation. For 7^{th} level (see Fig. 2d), the many details of the original terrain are again apparent.

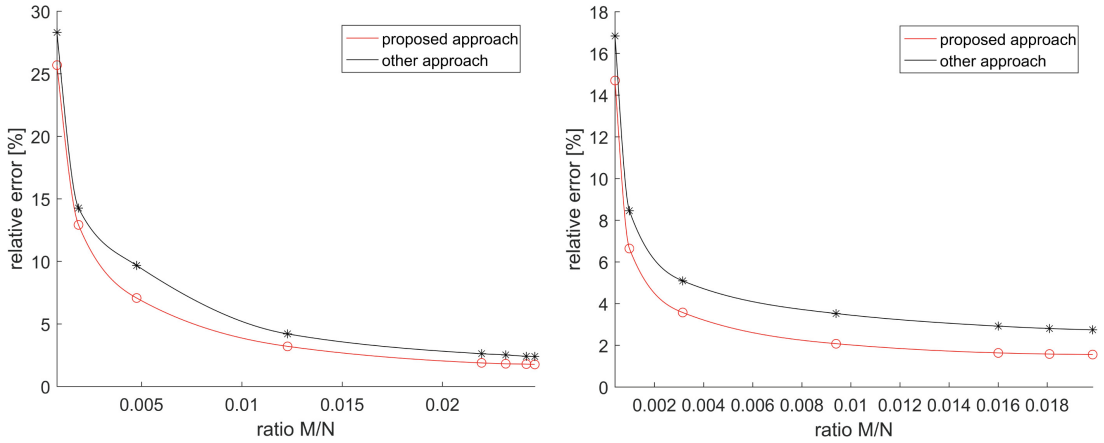


(a) original data, $N = 131,044$



(b) 1^{st} level, $M = 51$ (c) 3^{rd} level, $M = 413$ (d) 7^{th} level, $M = 2602$

Fig. 2. The part of Pennine Alps, Switzerland and its contour map: original data and different levels of proposed incremental RBF approximation when TPS is used.



(a) The mount Veľký Rozsutec (b) The part of Pennine Alps

Fig. 3. The mean relative error of the proposed incremental RBF approximation in comparison with classical RBF approximation [9] for different compression ratio.

The mean relative error in dependency on compression ratio is presented for both geographic datasets in Fig. 3. Moreover, the comparison of the proposed incremental approach with the classical RBF approximation [9] is performed. From the results, it can be seen that the proposed approach achieves the better quality of results in terms of error.

5 Conclusion

In this paper, a new incremental approach for RBF approximation for geographic data is presented. Selection of the set of reference points for proposed incremental approximation is based on the determination of stationary points of the input point cloud in the first level and the finding local maxima of residues at each hierarchical level. In addition, the Gaussian low-pass filter is used to smooth the trend of the input points, resp. the residues before finding significant points.

The proposed approach achieves the improvement of results in comparison with other existing methods because the features of the given dataset are respected.

In the future work, the proposed approach can be extended to higher dimensions, as the extension should be straightforward. Also, the improving the computational performance without loss of accuracy can be explored.

Acknowledgments. The authors would like to thank their colleagues at the University of West Bohemia, Plzeň, for their discussions and suggestions, and the anonymous reviewers for their valuable comments. The research was supported by the Czech Science Foundation GAČR project GA17-05534S and partially supported by the SGS 2016-013 project.

References

1. Hardy, R.L.: Multiquadratic equations of topography and other irregular surfaces. J. Geophys. Res. **76**, 1905–1915 (1971)
2. Hardy, R.L.: Theory and applications of the multiquadric-biharmonic method 20 years of discovery 1968–1988. Comput. Math. Appl. **19**(8), 163–208 (1990)
3. Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R.: Reconstruction and representation of 3D objects with radial basis functions. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, Los Angeles, California, USA, 12-17 August 2001, pp. 67–76 (2001)
4. Majdisova, Z., Skala, V.: Big geo data surface approximation using radial basis functions: a comparative study. Comput. Geosci. **109**, 51–58 (2017)
5. Smolik, M., Skala, V.: Large scattered data interpolation with radial basis functions and space subdivision. Integr. Comput.-Aided Eng. **25**(1), 49–62 (2018)
6. Pepper, D.W., Rasmussen, C., Fyda, D.: A meshless method using global radial basis functions for creating 3-D wind fields from sparse meteorological data. Comput. Assisted Methods Eng. Sci. **21**(3–4), 233–243 (2014)
7. Hon, Y.C., Sarler, B., Yun, D.F.: Local radial basis function collocation method for solving thermo-driven fluid-flow problems with free surface. Eng. Anal. Boundary Elem. **57**, 2–8 (2015)
8. Li, M., Chen, W., Chen, C.: The localized RBFs collocation methods for solving high dimensional PDEs. Eng. Anal. Boundary Elem. **37**(10), 1300–1304 (2013)
9. Majdisova, Z., Skala, V.: Radial basis function approximations: comparison and applications. Appl. Math. Model. **51**, 728–743 (2017)
10. Majdisova, Z., Skala, V., Smolik, M.: Determination of stationary points and their bindings in dataset using RBF methods. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds.) Computational and Statistical Methods in Intelligent Systems. Advances in Intelligent Systems and Computing series, vol. 859, pp. 213–224. Springer, Cham (2019)