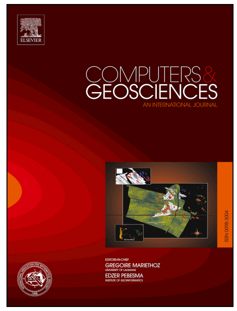

Big geo data surface approximation using radial basis functions: A comparative study

Zuzana Majdisova, Vaclav Skala

PII: S0098-3004(17)30202-9

DOI: 10.1016/j.cageo.2017.08.007

Reference: CAGEO 4004

To appear in: *Computers and Geosciences*

Received Date: 21 February 2017

Revised Date: 31 July 2017

Accepted Date: 7 August 2017

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

Zuzana Majdisova[a,*], Vaclav Skala[a]

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

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.

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

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), Hardy (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 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:

*Corresponding author
Email address: majdisz@kiv.zcu.cz (Zuzana Majdisova)
URL: www.vaclavskala.eu (Vaclav Skala)