Some Numerical Experiments for Determining the Optimal Value of the Scale Parameter for Meshfree Interpolation

Msc Alvin ASIMI
Department of Mathematic Engineering, Polytechnic University of Tirana,
Tirana, Albania

Prof.Dr Lulëzim HANELLI
Department of Mathematic Engineering, Polytechnic University of Tirana,
Tirana, Albania

ABSTRACT
This paper consists in the compression of two methods for determining the optimal ε parameter for the Meshfree interpolation of a data set. In the first part we will present two approaches: "Trial and Error" and "Leave one out" cross validation. In the second part are presented some 1D and 2D experiments for comparing this two approaches. For each approach, we will determine the optimal ε scaling parameter for 1D and 2D interpolation with different values of N data points. At the end there will be a comparison between the results issued . We will use MATLAB for all the calculations and the plots.

The data sites used in this paper can be downloaded at http://www.math.unipd.it/~demarchi/TAA2010/.

Keywords
Meshfree Interpolation, Trial and Error, Levae one out, Matlab, optimal ε parameter.

1. INTRODUCTION
Let's consider the interpolation problem of a mesh free data set. This is one of the situations that are shown more often in practice. So, let's assume that we have a data set (x_j, f_j), $j = 1, 2, ..., n$, where $x_j \in \mathbb{R}^s$ (we will focus on s=1 and s=2 on our experiments) and $f_j = f(x_j) \in \mathbb{R}$. The problem

$$P_f(x) = \sum_{j=1}^{N} c_j \varphi(\varepsilon \left\| x - x_j \right\|)$$

which must fulfill the condition $P_f(x_j) = f_j, \ \forall j = 1, ..., n$.

To have the best fit there should be considered the two problems below:

- Which functions φ should we chose
- What parameter ε should we use to scale the basis functions $\varphi_j = \varphi(\left\| x - x_j \right\|)$?

For the first problem we must consider all the information which can be obtained from the data about the functions from which data was generated. In this case the selection of the interpolant must be based on this information . For more of this problem see [1]

In this paper we will be focused on the second problem, the problem of choosing the "best" shape parameter ε. In this porpoise we will perform some experiments with 1D and 2D data.

Gaussian basic function

$$\varphi\left(\left\| x - x_j \right\|_2\right) = e^{-\varepsilon^2 \left\| x - x_j \right\|_2^2} , \ x \in \mathbb{R}^s$$

will be used to fit data generated from the F(x)=sinc(x) , $x \in \mathbb{R}^s$ function in 1D and 2D. We will consider to strategies to find the "better" ε parameter.

2. CHOOSING ε PARAMETER
2.1. Trial and Error approach
This approach consists in performing a series of numerical experiments, in order to choose the "best" parameter. For each experiment performed we will see the average error and the "best" parameter shall be the one with the minimal error. This can be possible if we know the values of the function f that generated the data, which we are interpolating.

Since in real data this is not possible, this approach is for academic results. Also if the function f is not known, it can be chosen as the "best" parameter ε, the minimal value for which the interpolation matrix it is not singular.

We will use this method to make a comparison with the results of the second method that we will use. (Leave-one-out cross validation)

2.2. Cross Validation approach

The second method that we will use for our experiments is Cross Validation. An algorithm that can be used for this porpoise is "*leave-one-out-cross validation*" presented by Rippa (see [2]). In this approach the "best" parameter ε is the one , which minimizes the interpolation error (least square) of the data by leaving one of the centers out . More specifically,

P_f^k is the interpoluesi of data $\{f_1,\ldots,f_{k-1},f_{k+1},\ldots,f_N\}$, so:

$$P_f^{[k]}(x)=\sum_{\substack{j=1\\ j\neq k}}^{N} c_j^{[k]}\varphi(\varepsilon\|x-x_j\|)$$

such that: $P_f^{[k]}(x_j)=f_j,\quad j=1,\ldots,k-1,k+1,\ldots,N$

and the error $E_k=f_k-P_f^{[k]}(x_k)$.

This way a vector with elements the average error is formed for each center left out, $E=[E_1,\ldots,E_N]$.

This procedure is repeated for different values of ε parameter and the optimal parameter is the one with the lower error.

Based on [2] we can calculate errors E_k with the formula:

$$E_k=\frac{c_k}{A_{kk}^{-1}}$$

where c_k is the k-th coefficient of the interpolant P_f based on all the data and A_{kk}^{-1} is the k-th diagonal element of the inverse interpolant matrix. This formula reduces the operations number for the calculation of errors for each parameter ε , becouse only one interpolation matrix is calculated for each parameter.

3. NUMERICAL EXPERIMENTS

In this part we are going to perform some experiments to implement the two methods above. As testing function we will take function F(x)=sinc(x). Based on 1D and 2D data we are going to perform the results comparison. Firstly, we will apply the "Trial and Error" approach by using as centers N- halton points in 1D and 2D.

In the figure below are shown the errors curves for some values of N, for 1D data interpolation of the F(x)= sinc (x) function.

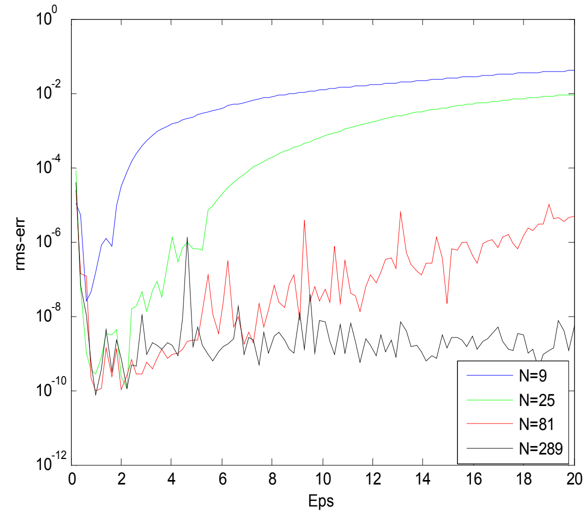

Fig 1. Optimal ε curves for 1D Gaussian interpolation of F(x)=sinc(x) function for different N halton points

As we can see from Fig.1 for the 1D interpolation, the minimal error is of the same range for N=25, N=61, N=289 and is reached approximately at the same value of ε

Let's perform the experiment for the 2D case. In figure 2 are shown graphically the errors curve for different values of N.

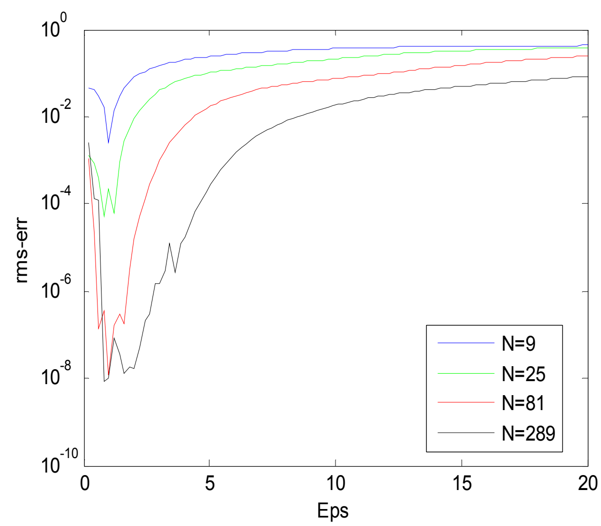

Fig 2. Optimal ε curves for 2D Gaussian interpolation of F(x)=sinc(x) function for different N halton points

The minimal error is of the same range for N=81 and N=289 and is reached for the same value ε =1.01.

Now let's perform the algorithm "Leave one Out" for N halton points as Centers in 1D. Optimal ε curves are shown in figure 3.

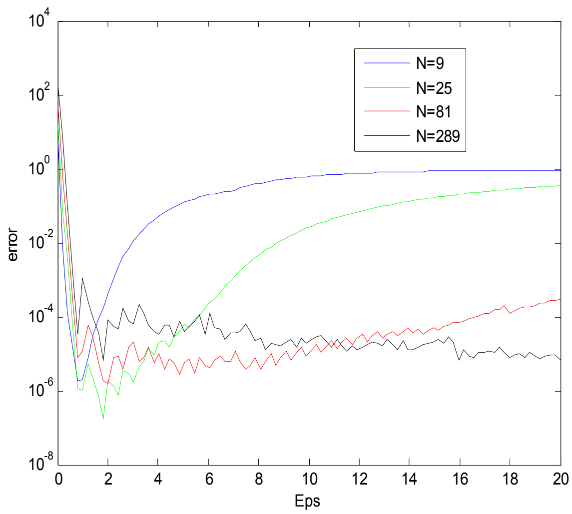

Fig 3. Optimal ε curves for 1D Gaussian interpolation of F(x)=sinc(x) function for different N halton points based on "Leave one out" algorithm

In the figure above we can see that the error calculated for N=81, N=289, for each value ε is of the range of ($10^{-4} - 10^{-6}$)
In the figure below are shown Optimal ε curves for the 2D case.

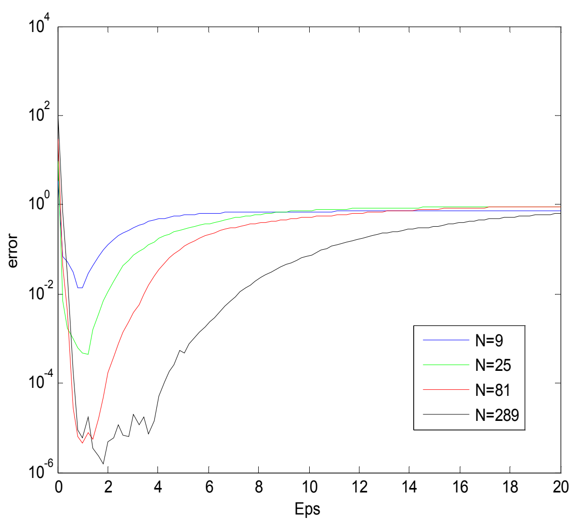

Fig 4. Optimal ε curves for 2D Gaussian interpolation of F(x)=sinc(x) function for different N halton points based on "Leave one out" algorithm

The optimal values of ε together with the respective errors for the 1D Gaussian Interpolation of the F(x)=sinc(x) function are shown in table 1, and the 2D interpolation case is shown in table 2.

Table 1. Optimal ε and errors for 1D Gaussian interpolation of F(x)=sinc(x) function for different N halton points

	Trial and Error		Leave one out	
N	Eps	error	Eps	error
9	0.6	2.5159E-08	0.81	1.8399E-06
25	2.22	1.1627E-10	1.82	1.7462E-07
81	1.01	9.8934E-11	2.02	1.5463E-06
289	1.01	7.5343E-11	15.96	6.8118E-06

From the interpolation of N=9 and N=25 halton points we get approximate values of optimal ε for both approaches.
The maximum difference of ε value for both approaches is reached for N=289 data points.

In the 2D case (table 2) we can see that the optimal values of ε are the same for N=9 and N=81 halton points, they vary a little for N=25 and they reach the maximum difference for N=289.

Table 2. Optimal ε and errors for 2D Gaussian interpolation of F(x)=sinc(x) function for different N halton points

	Trial and Error		Leave one out	
N	Eps	error	Eps	error
9	1.01	2.5000E-03	1.01	1.3787E-02
25	0.81	5.0979E-05	1.21	4.3946E-04
81	1.01	1.1968E-08	1.01	4.5232E-06
289	0.81	8.3476E-09	1.81	1.5268E-06

Based on Tab.2 we can see that for 2D interpolation we get approximate values of ε parameter for each N.

4. CONCLUSIONS

The "Trial and Error" method used for determining ε parameter, is an approach performed for academic results, because it needs to know the function from which the data are generated. But, it can be useful as a comparative approach with another method for determining the optimal ε parameter.

Based on the experiments developed above results that a good approach for determining ε parameter is the "Leave one out" cross validation algorithm. The experiments were performed for N=9,N=25, N=81, N=289 halton points for the 1D and 2D interpolation case.

For the 1D interpolation case the optimal values of ε parameter results approximately the same to each other for both methods, except for N=289 data points .

In the 2D case of data interpolation, from all the data analyzed, optimal values of ε parameter results approximately the same to each other for both methods.

The "Leave one out" algorithm in almost all cases gives approximate results with the "Trial and Error" algorithm for the optimal value of ε and also the respective range of error. This is a good approach for scaling the data with the best parameter, since it gives approximates results with the "Trial and error approach".

REFERENCES

[1] Schaback, R and Welland, H (2006). Kernel techniques: From machine learning to meshless methods, Acta Numerica, 15, pp. 543-639.

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

[3] used data: http://www.math.unipd.it/~demarchi/TAA2010/

