



PEMILIHAN NILAI PARAMETER c PADA INTERPOLAN GAUSSIAN

Elin Herlinawati
FMIPA, Universitas Terbuka
e-mail: elin@ecampus.ut.ac.id

ABSTRACT

Interpolation is to find a function $F: X \rightarrow \mathbb{R}$ which passes through a number of given data points. The interpolant that we used is a Gaussian function which has a parameter value of c . Selection of this parameter affects the results of interpolation. This study discusses a method that we used in selecting the optimum parameter of c . This method is inspired by a RMS (Root Means Square) error. Suppose $\{(x_i, y_i)\}$ with $1 \leq i \leq n$ is the set of data points. Define the error vector $g = (g_1, \dots, g_n)^T$ with

$$g_k = y_k - F_k(x_k), \quad k \in \{1, 2, \dots, n\},$$

where y_k is the function value in x_k and F_k is the interpolant which is obtained by deleting a point $\{x_k, y_k\}$ from the given data set. Furthermore, the optimum parameter value of c is selected by minimizing the error vector g . The result of selecting parameter values depends on the amount of data and the distribution of known data.

Keywords: Gaussian function, interpolation, parameter value of c

ABSTRAK

Interpolasi adalah pencarian fungsi $F: X \rightarrow \mathbb{R}$ melalui sejumlah titik data yang diberikan. Interpolan yang digunakan pada artikel ini adalah fungsi Gaussian yang memiliki nilai parameter c . Pemilihan nilai parameter c mempengaruhi hasil interpolasi. Artikel ini membahas metode yang digunakan dalam pemilihan parameter c yang optimum. Metode ini terinspirasi dari galat RMS (*Root Means Square*). Misalkan $\{(x_i, y_i)\}$ dengan $1 \leq i \leq n$ adalah himpunan titik data yang diberikan. Didefinisikan vektor galat $g = (g_1, \dots, g_n)^T$ dengan

$$g_k = y_k - F_k(x_k), \quad k \in \{1, 2, \dots, n\},$$

y_k adalah nilai fungsi di x_k dan F_k adalah interpolan yang diperoleh dengan menghapus satu titik $\{x_k, y_k\}$ dari himpunan data yang diberikan. Selanjutnya, nilai parameter c yang optimum dipilih dengan cara meminimumkan vektor galat g . Hasil dari pemilihan nilai parameter bergantung pada banyaknya data dan sebaran data yang diketahui.

Kata kunci: fungsi Gaussian, interpolasi, nilai parameter c

Misalkan $\{(x_i, y_i)\}$ dengan $1 \leq i \leq n$ adalah himpunan titik data. Interpolasi merupakan suatu cara mencari fungsi F yang memenuhi kondisi $F(x_i) = y_i$ untuk setiap $i \in \{1, \dots, n\}$. Fungsi F yang dicari berbentuk

$$F(x) = \sum_{j=1}^n \beta_j f(\|x - x_j\|), \quad x \in \mathbb{R}^d \quad (1.1)$$

dengan f fungsi bernilai real yang disebut sebagai fungsi basis dari interpolan, dan $\| \cdot \|$ menyatakan norm Euclid. Fungsi F harus memenuhi kondisi interpolasi, yakni

$$F(x_i) = y_i, \quad 1 \leq i \leq n \quad (1.2)$$

Perhatikan bahwa untuk menyelesaikan persamaan (1.1) yang memenuhi kondisi (1.2) sama dengan menyelesaikan sistem persamaan

$$A\beta = y, \quad A = A_{ij} = f(\|x_i - x_j\|) \quad (1.3)$$

dengan $\beta = \{\beta_1, \dots, \beta_n\}$, $\beta_i \in \mathbb{R}$. Masalah interpolasi tersebut mempunyai solusi jika dan hanya jika matriks A adalah matriks nonsingular. Fungsi definit positif memberikan syarat cukup untuk membuat matriks A nonsingular. Selanjutnya, Cheney dan Light (2009) mengemukakan bahwa fungsi Gaussian merupakan fungsi definit positif. Akibatnya, matriks $A = [f(x_i - x_j)]$ dengan f fungsi Gaussian merupakan matriks nonsingular. Oleh karena itu, fungsi Gaussian dipilih sebagai fungsi basis dari interpolan.

Telah diketahui fungsi Gaussian memiliki nilai parameter c . Pemilihan nilai parameter tersebut mempengaruhi hasil interpolasi. Sarra & Sturgill (1992) menemukan strategi penentuan nilai parameter c untuk interpolasi dengan fungsi basis radial termasuk fungsi Gaussian. Selain himpunan titik data yang diberikan, Sturgill menggunakan himpunan titik evaluasi untuk menghitung galat interpolasi. Analog dengan Sturgill, Madych (1992) juga menggunakan himpunan titik evaluasi untuk menghitung galat interpolasi. Artikel ini menunjukkan metode lain yang digunakan untuk memilih nilai parameter c dari interpolan Gaussian.

Penjelasan tentang Interpolan Gaussian adalah sebagai berikut. Masalah interpolasi pada persamaan (1.1) memerlukan fungsi basis. Salah satu fungsi yang dapat digunakan adalah fungsi radial. Fungsi radial adalah fungsi yang terdefinisi pada sebuah ruang Euclid \mathbb{R}^d yang nilainya hanya bergantung pada jarak antara titik di \mathbb{R}^d dengan suatu titik c sebagai titik pusat. Dalam Cheney & Light (2009), fungsi radial didefinisikan sebagai berikut.

Definisi 2.1 Misalkan X adalah suatu ruang hasil kali dalam dan f adalah fungsi yang bernilai real. Maka f dikatakan radial apabila $f(x) = f(y)$ untuk setiap $x, y \in X$ dengan $\|x\| = \|y\|$.

Interpolasi dengan fungsi radial telah digunakan di berbagai area. Hardy (1971) mengemukakan bahwa metode ini dapat diterapkan pada bidang geologi, geofisika, survei dan pemetaan, penginderaan jauh, pemrosesan signal, dan lain-lain. Selain itu, Franke (1982) juga membandingkan 30 metode interpolasi 2-D. Dari eksperimen tersebut diperoleh hasil bahwa salah satu metode yang paling baik adalah metode yang menggunakan fungsi basis radial. Salah satu jenis fungsi radial adalah fungsi Gaussian.

Definisi 2.2 Fungsi f dikatakan fungsi Gaussian jika f mempunyai bentuk

$$f(x) = \exp(-c^2 x^2) \quad (2.1)$$

untuk setiap $x \in \mathbb{R}^d$ dan $c \in \mathbb{R}$.

Ada beberapa sifat fungsi Gaussian terkait interpolasi, diantaranya fungsi ini merupakan keluarga fungsi kontinu $C^\infty(\mathbb{R}^d)$. Selain itu, fungsi Gaussian juga merupakan suatu transformasi Fourier dari fungsi tertentu sebagaimana disebutkan dalam lemma berikut.

Lemma 2.3 (Cheney & Light, 2009) Setiap fungsi Gaussian $f(y) = \exp(-c^2\|y\|^2)$, $y \in \mathbb{R}^d$ dan $c > 0$, merupakan sebuah transformasi Fourier dari suatu fungsi, yaitu:

$$\exp(-c^2\|y\|^2) = \int_{\mathbb{R}^d} \left(\frac{\pi}{c^2}\right)^{d/2} \exp\left(-\frac{\pi^2}{2}\|x\|^2\right) \exp(-2\pi ixy) dx \quad (2.2)$$

untuk setiap $x, y \in \mathbb{R}^d$.

Selanjutnya, masalah interpolasi pada persamaan (1.1) erat kaitannya dengan fungsi definit positif dan interpolan yang digunakan dalam artikel ini adalah fungsi Gaussian. Oleh karena itu, perlu ditunjukkan bahwa fungsi Gaussian adalah fungsi definit positif.

Misalkan X adalah sebarang himpunan buka, μ adalah ukuran Borel pada X yakni ukuran yang didefinisikan pada himpunan buka X , dan $\hat{\mu}$ adalah transformasi Fourier dari μ . Lemma berikut menjelaskan kaitan antara ukuran Borel, transformasi Fourier, dan fungsi definit positif.

Lemma 2.4 Misalkan μ adalah suatu ukuran Borel bernilai hingga dan tak negatif pada \mathbb{R}^d dengan penyangga dari μ bukan merupakan himpunan berukuran Lebesgue nol. Maka $\hat{\mu}$ definit positif pada \mathbb{R}^d .

Dalam lemma tersebut digunakan istilah *penyangga* dari ukuran Borel μ pada X yang didefinisikan sebagai himpunan $C = X \setminus B$ dengan B merupakan himpunan buka di X dan $\mu(B) = 0$. Kaitannya dengan interpolasi, Lemma 2.4 menjadi salah satu cara untuk menunjukkan suatu fungsi adalah definit positif. Salah satu fungsi definit positif adalah fungsi Gaussian sebagaimana disebutkan pada teorema berikut.

Teorema 2.5 (Cheney & Light, 2009) Jika $c > 0$, maka fungsi $f(x) = \exp(-c^2\|x\|^2)$, $x \in \mathbb{R}^d$, radial dan definit positif pada sebarang ruang hasil kali dalam real.

Bukti. Karena setiap ruang hasil kali dalam real H berdimensi d isomorfik dengan ruang Euclid \mathbb{R}^d , maka titik-titik di H dapat diasumsikan berada di \mathbb{R}^d . Selanjutnya, berdasarkan definisi, fungsi f adalah radial.

Selanjutnya, untuk menunjukkan bahwa f adalah fungsi definit positif, berdasarkan Lemma 2.4, cukup ditunjukkan bahwa f adalah suatu transformasi Fourier dari sebuah ukuran di \mathbb{R}^d dengan penyangga dari ukuran tersebut mempunyai ukuran Lebesgue positif. Definisikan ukuran μ sebagai

$$d\mu = \left(\frac{\pi}{c^2}\right)^{d/2} \exp\left(-\frac{\pi}{c^2}\|x\|^2\right) dx \quad (2.3)$$

Dari Lemma 2.3, untuk setiap $y \in \mathbb{R}^d$,

$$\begin{aligned} \hat{\mu}(y) &= \int_{\mathbb{R}^d} e^{-2\pi ixy} d\mu(x) \\ &= \int_{\mathbb{R}^d} e^{-2\pi ixy} \left(\frac{\pi}{c^2}\right)^{\frac{d}{2}} \exp\left(-\left(\frac{\pi^2}{c^2}\right)\|x\|^2\right) dx \\ &= e^{-c\|y\|^2} \\ &= f(y). \end{aligned} \tag{2.4}$$

Jadi, f adalah transformasi Fourier dari suatu ukuran μ di \mathbb{R}^d .

Karena f adalah fungsi Gaussian, maka f bernilai positif untuk setiap y di \mathbb{R}^d sehingga penyangga dari μ adalah \mathbb{R}^d . Hal ini menunjukkan bahwa ukuran Lebesgue penyangga dari μ tidak nol.

Berdasarkan Lemma 2.4, fungsi f definit positif pada \mathbb{R}^d .

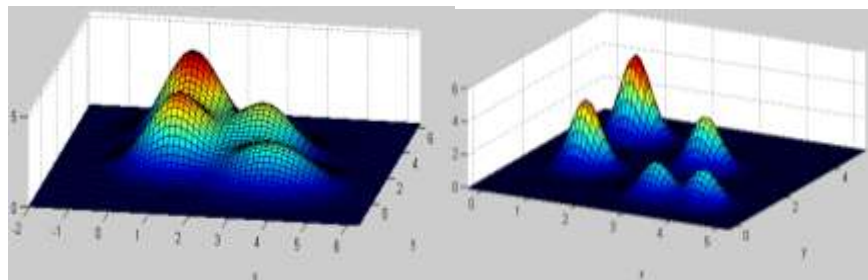
Berdasarkan Teorema 2.5, penggunaan fungsi Gaussian sebagai fungsi basis interpolan mengakibatkan matriks interpolasi taksingular sehingga persamaan (1.1) mempunyai solusi. Hal lain yang perlu diperhatikan adalah keberadaan nilai parameter c pada fungsi Gaussian yang mempengaruhi hasil interpolasi. Perhatikan contoh berikut.

Contoh 2.6 Diberikan himpunan data seperti pada Tabel 1.

Tabel 1. Himpunan Titik Data

Titik pada \mathbb{R}^2	(3,1)	(1,2)	(4,1)	(3,3)	(1,4)
Nilai fungsi	2	4	2	3	5

Masalah interpolasi pada Tabel 1 akan diselesaikan dengan menggunakan interpolan fungsi Gaussian yang berbentuk $F(x) = \sum_{j=1}^5 \beta_j e^{-c\|x-x_j\|^2}$, $c > 0$. Pertama, digunakan nilai $c = 1$, kemudian diproses interpolasi diulang dengan menggunakan $c = 10$. Gambar permukaan dari hasil interpolasi ditunjukkan pada Gambar 1.



Gambar 1. Interpolasi menggunakan fungsi Gaussian dengan $c = 1$ (kiri) dan $c = 10$ (kanan)

Contoh 2.6 menunjukkan bahwa perbedaan nilai parameter c yang digunakan mengakibatkan perbedaan pada hasil interpolasi.

METODE

Metode yang digunakan dalam penelitian ini adalah studi literatur, yakni dengan mencari referensi teori yang relevan dengan kasus atau permasalahan yang ditemukan. Kemudian temuan tersebut dikembangkan dan dibuatkan algoritma yang sesuai yang selanjutnya dilakukan uji coba terhadap kebenaran dari algoritma yang telah dibuat.

HASIL DAN PEMBAHASAN

Beberapa metode pemilihan nilai parameter c untuk interpolan multiquadrik dan invers multiquadrik telah dilakukan Hardi(1971) dan Franke (1982). Pada penelitian ini dibahas metode pemilihan nilai parameter c pada interpolan Gaussian dengan memanfaatkan himpunan data yang diketahui. Metode ini terinspirasi dari metode *RMS*.

Misalkan $\{(x_i, y_i)\}$ dengan $1 \leq i \leq n$ adalah himpunan titik data yang diberikan. Dari N buah data tersebut, dipilih sebarang $N - 1$ titik data untuk penentuan interpolan F dan titik data yang lain digunakan sebagai titik evaluasi untuk penentuan galat interpolasi. Definisikan vektor galat

$$g = (g_1, \dots, g_n)^T \text{ dengan } g_k = y_k - F_k(x_k), \quad k \in \{1, 2, \dots, n\},$$

dengan y_k adalah nilai fungsi di x_k dan F_k adalah interpolan yang diperoleh dengan menghapus satu titik $\{x_k, y_k\}$ dari himpunan data yang diberikan. Proses ini dilakukan sebanyak N kali sehingga diperoleh galat g_1, \dots, g_n . Selanjutnya, nilai parameter c yang optimum dipilih dengan cara meminimumkan vektor galat g .

Metode ini dapat dilakukan untuk $N \geq 3$. Hal ini disebabkan himpunan titik data yang diberikan digunakan juga sebagai himpunan titik evaluasi untuk penentuan galat interpolasi. Selanjutnya diberikan lemma dan teorema yang menjelaskan penghitungan galat g .

Lemma 3.1 Misalkan A adalah sebarang matriks, $y = \{y_1, \dots, y_n\}$, dan $z = \{z_1, \dots, z_n\}$. Jika $Ay = z$ dengan $y_k = 0$, maka

$$A^{(k)}y^{(k)} = z^{(k)},$$

dengan $A^{(k)}$ adalah matriks $A_{n \times n}$ yang dihilangkan baris ke- k dan kolom ke- k ,

$$y^{(k)} = \{y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n\}, \text{ dan } z^{(k)} = \{z_1, \dots, z_{k-1}, z_{k+1}, \dots, z_n\}.$$

Teorema berikut analog dengan hasil penelitian Rippa (1999).

Teorema 3.2 Misalkan titik-titik data $\{(x_j, y_j)\}_{j=1}^n$ dan $g = [g_1, \dots, g_n]^T$ adalah vektor galat. Nilai g_k dapat diperoleh dengan menggunakan rumus:

$$g_k = \frac{\beta_k}{A_{kk}^{-1}}, \quad (3.1)$$

Dengan β_k adalah koefisien ke- k dari interpolan F berdasarkan pada semua data yang diketahui, dan A_{kk}^{-1} adalah unsur diagonal ke- k dari invers matriks interpolasi.

Bukti. Misalkan $V = (x_1, \dots, x_n)$, $x_j \in \mathbb{R}^d$ himpunan titik data, $y = (y_1, \dots, y_n)$, $y_j \in \mathbb{R}$, vektor yang bersesuaian dengan V , dan $F(x) = \sum_{j=1}^n \beta_j \varphi(\|x - x_j\|)$ adalah interpolan dengan fungsi Gaussian.

Perhatikan bahwa F memenuhi kondisi $F(x_l) = y_l$ untuk setiap $l = 1, 2, \dots, n$. Masalah interpolasi tersebut dapat dituliskan dalam bentuk sistem persamaan linear

$$A\beta = y, \text{ dengan } A = [A_{lj}] = [\varphi(\|x_l - x_j\|)] \quad (3.2)$$

Misalkan $V^{(k)}, y^{(k)}, \beta^{(k)}$ berturut-turut diperoleh dari penghapusan entri ke- k dari vektor-vektor V, y , dan β . Kemudian misalkan $p^{[k]}$ adalah solusi dari

$$Ap^{[k]} = e^{[k]}, \quad (3.3)$$

dengan $e^{[k]}$ adalah kolom ke- k dari matriks identitas $N \times N$. Dalam hal ini $p_k^{[k]} \neq 0$ karena andaikan $p_k^{[k]} = 0$, dari Lemma 3.1 dan persamaan (3.3), diperoleh

$$A^{(k)} \left(p_1^{[k]}, \dots, p_{k-1}^{[k]}, p_{k+1}^{[k]}, \dots, p_n^{[k]} \right)^T = 0.$$

Karena $A^{(k)}$ taksingular, maka $\left(p_1^{[k]}, \dots, p_{k-1}^{[k]}, p_{k+1}^{[k]}, \dots, p_n^{[k]} \right)^T = 0$. Akibatnya $p^{[k]} = 0$, kontradiksi dengan $p^{[k]}$ adalah solusi dari persamaan (3.3). Jadi, haruslah $p_k^{[k]} \neq 0$.

Misalkan $b^{[k]} \in \mathbb{R}^n$ dengan $b^{[k]} = \beta - \frac{\beta_k}{p_k^{[k]}} p^{[k]}$. Maka diperoleh

$$\begin{aligned} Ab^{[k]} &= A \left(\beta - \frac{\beta_k}{p_k^{[k]}} p^{[k]} \right) = A\beta - \frac{\beta_k}{p_k^{[k]}} Ap^{[k]} = y - \frac{\beta_k}{p_k^{[k]}} e^{[k]} \\ &= \left(y_1, \dots, y_{k-1}, y_k - \frac{\beta_k}{p_k^{[k]}}, y_{k+1}, \dots, y_n \right)^T. \end{aligned} \quad (3.4)$$

Karena $b_k^{[k]} = 0$, berdasarkan Lemma 3.1, maka $\beta^k = (b_1^k, \dots, b_{k-1}^k, b_{k+1}^k, \dots, b_k^k)^T$. Akibatnya adalah,

$$\begin{aligned} F^{(k)}(x_k) &= \sum_{\substack{j=1 \\ j \neq k}}^n \beta_j^k \varphi(\|x_k - x_j\|) = \sum_{\substack{j=1 \\ j \neq k}}^n b_j^{[k]} \varphi(\|x_k - x_j\|) \\ &= \sum_{j=1}^n b_j^{[k]} \varphi(\|x_k - x_j\|) = [Ab^{[k]}]_k \\ &= y_k - \frac{\beta_k}{p_k^{[k]}}. \end{aligned} \quad (3.5)$$

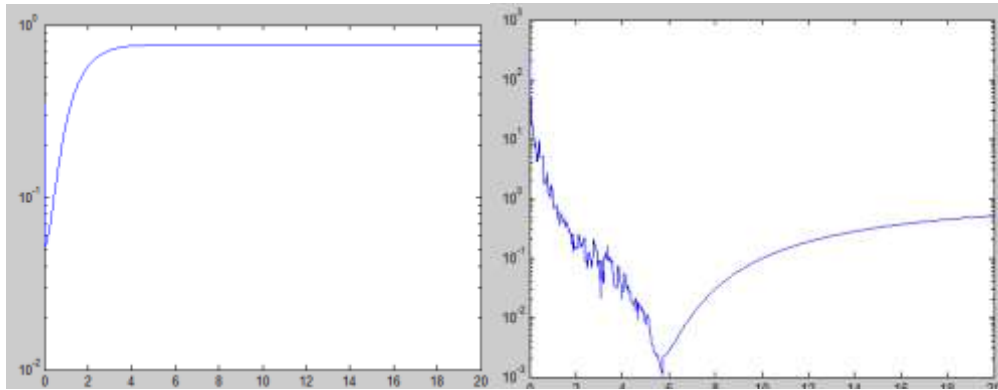
Karena $p^{[k]}$ adalah solusi dari persamaan (3.3), maka $p_k^{[k]} = A_{kk}^{-1}$. Sehingga rumus sederhana untuk komponen galat adalah

$$g_k = y_k - F^{(k)}(x_k) = y_k - \left(y_k - \frac{\beta_k}{p_k^{[k]}} \right) = \frac{\beta_k}{p_k^{[k]}} = \frac{\beta_k}{A_{kk}^{-1}}. \quad (3.6)$$

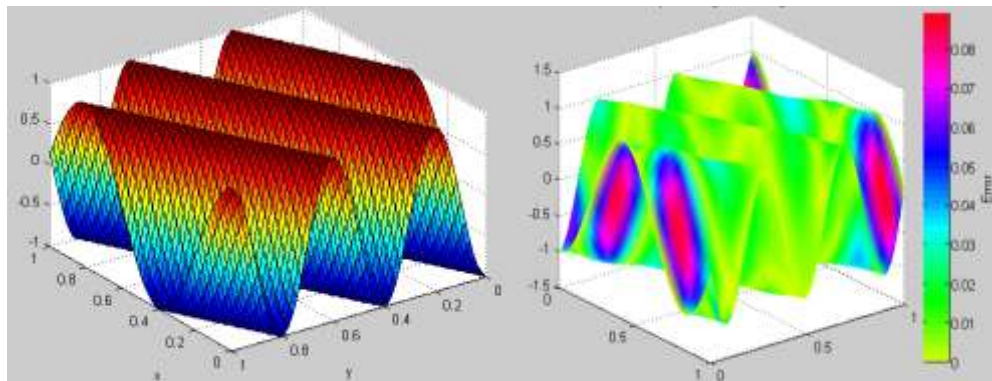
Selanjutnya, dilakukan percobaan numerik dengan N yang berbeda-beda, yakni $N = 9, N = 25, N = 49, N = 81$, dan $N = 289$ pada $[0,1]^2 \subset \mathbb{R}^2$. Kemudian menetapkan interval nilai c yang akan dipilih. Dalam hal ini, nilai c yang dipilih dari interval $(0,20)$, yang didiskritisasi sebanyak 1000 buah. Selanjutnya, pilih nilai c yang meminimumkan galat g . Tabel berikut menunjukkan nilai c yang meminimumkan galat g untuk setiap banyaknya data (N) yang diberikan.

Tabel 2. Nilai Parameter Optimum untuk N Berbeda

N	c opt	Galat g
9	0,04	5.217222e-02
25	2,94	3.933688e-01
49	3,32	8.201503e-02
81	4,20	5.029745e-02
289	5,74	1.056081e-03



Gambar 2. Kurva galat untuk $N = 9$ (kiri) dan $N = 289$ (kanan)



Gambar 3. Gambar fungsi sumber (kiri) dan gambar permukaan hasil interpolasi (kanan)

SIMPULAN

Pemilihan nilai parameter c pada interpolan Gaussian dapat dilakukan dengan cara meminimumkan galat interpolasi yang diperoleh dengan menghilangkan satu titik data yang diberikan untuk dijadikan sebagai titik evaluasi. Hasil dari pemilihan nilai parameter bergantung pada banyaknya data dan sebaran data yang diketahui. Kelebihan dari metode ini adalah tidak dibutuhkannya sejumlah titik evaluasi untuk menghitung galat interpolasi. Akan tetapi, metode ini juga memiliki kelemahan yakni galat bergantung pada sebaran titik data yang diberikan.

REFERENSI

- Cheney, W & Light W. (2009). *A Course in approximation theory*. American Mathematical Society.
- Franke, R. (1982). Scattered data interpolation: Tests of some methods. *Math. Comp.*, 38, 181–200.
- Hardy, Rolland L. (1971). Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.*, 76, 1905–1915.
- Madych, W.R. (1992). Miscellaneous error bounds for multiquadric and related interpolants. *Comput. Math. Appl.*, 24, 121-138.
- Rippa, S. (1999). An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Journal of Advance in Computational Mathematics*, 11, 193-210.
- Sarra, S.A. & Sturgill, D. (1992). A random variable shape parameter strategy for radial basis function approximation methods, *Eng. Anal. Bound. Elem*, 33(1), 99-120.