Geometric greedy and greedy points for RBF interpolation

Stefano De Marchi

Department of Computer Science, University of Verona (Italy)

CMMSE09, Gijon July 2, 2009

Motivations

- Stability is very important in numerical analysis: *desirable* in numerical computations; it depends on the *accuracy* of algorithms [3, Higham's book].
- In polynomial interpolation, the stability of the process can be measured by the so-called Lebesgue constant, i.e the norm of the projection operator from $\mathcal{C}(K)$ (equipped with the uniform norm) to $\mathbb{P}_n(K)$ or itselfs ($K \subset \mathbb{R}^d$), which also estimates the interpolation error.
- The Lebesgue constant depends on the *interpolation points* (via the fundamental Lagrange polynomials).
- Are these ideas applicable also in the context of RBF interpolation?

Outline

Motivations

Good interpolation points for RBF

Results

Greedy and geometric greedy algorithms

Numerical examples

Asymptotic points distribution

Lebesgue Constants estimates

References

- DeM: RSMT03;
- DeM, Schaback, Wendland: AiCM05;
- DeM, Schaback: AiCM09;
- DeM: CMMSE09

Notations

- $X = \{x_1, ..., x_N\} \subseteq \Omega \subseteq \mathbb{R}^d$, distinct; *data sites*;
- $\{f_1, ..., f_N\}$, *data values*;
- $\Phi : \Omega \times \Omega \to \mathbb{R}$ *symmetric (strictly) positive definite kernel*

the RBF interpolant

$$s_{f,X} := \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j)\,, \tag{1}$$

Letting $V_X = \text{span}\{\Phi(\cdot, x) : x \in X\}$, $s_{f,X}$ can be written in terms of cardinal functions, $u_j \in V_X$, $u_j(x_k) = \delta_{jk}$, i.e.

$$s_{f,X} = \sum_{j=1}^{N} f(x_j) u_j\,. \tag{2}$$

Error estimates

- Take $V_\Omega = \text{span}\{\Phi(\cdot, x) : x \in \Omega\}$ on which Φ is the reproducing kernel: $\overline{V_\Omega} := \mathcal{N}_\Phi(\Omega)$, the native Hilbert space associated to Φ.
- $f \in \mathcal{N}_\Phi(\Omega)$, using (2) and the reproducing kernel property of Φ on V_Ω, applying the Cauchy-Schwarz inequality, we get the generic pointwise error estimate (cf. e.g., Fasshauer's book, p. 117-118):

$$|f(x) - s_{f,X}(x)| \le P_{\Phi,X}(x)\, \|f\|_{\mathcal{N}_\Phi(\Omega)} \qquad (3)$$

$P_{\Phi,X}$: *power function*.

A power function expression

Letting $\det\left(A_{\Phi,X}(y_1,...,y_N)\right) = \det\left(\Phi(y_i,x_j)\right)_{1\le i,j\le N}$, then

$$u_k(x) = \frac{\det_{\Phi,X}(x_1,\ldots,x_{k-1},x,x_{k+1},\ldots,x_N)}{\det_{\Phi,X}(x_1,\ldots,x_N)}\,, \qquad (4)$$

Letting $u_j(x),\ 0 \le j \le N$ with $u_0(x) := -1$ and $x_0 = x$, then

$$P_{\Phi,X}(x) = \sqrt{\mathbf{u}^T(x) A_{\Phi,Y} \mathbf{u}(x)}\,, \qquad (5)$$

where $\mathbf{u}^T(x) = (-1, u_1(x), \ldots, u_N(x)),\quad Y = X \cup \{x\}$.

The problem

Are there any good points for approximating all functions in the native space?

Our approach

1. **Power function estimates.**
2. **Geometric arguments.**

Overview of the existing literature

- *A. Beyer: Optimale Centerverteilung bei Interpolation mit radialen Basisfunktionen. Diplomarbeit*, Universität Göttingen, 1994.

  He considered numerical aspects of the problem.

- *L. P. Bos and U. Maier: On the asymptotics of points which maximize determinants of the form* $\det(g(|x_i - x_j|))$, in *Advances in Multivariate Approximation* (Berlin, 1999),

  They investigated on *Fekete-type* points for univariate RBFs, proving that if g is s.t. $g'(0) \neq 0$ then points that maximize the Vandermonde determinant are the ones asymptotically equidistributed.

Literature

- *A. Iske:*, *Optimal distribution of centers for radial basis function methods.* Tech. Rep. M0004, Technische Universität München, 2000.
  He studied admissible sets of points by varying the centers for stability and quality of approximation by RBF, proving that uniformly distributed points gives better results. He also provided a bound for the so-called uniformity:
  $\rho_{X,\Omega} \leq \sqrt{2(d+1)/d}$, $d=$ space dimension.
- *R. Platte and T. A. Driscoll:*, *Polynomials and potential theory for GRBF interpolation*, SINUM (2005), they used potential theory for finding near-optimal points for gaussians in 1d.

Main result

Idea: data sets, for good approximation for all $f \in \mathcal{N}_\Phi(\Omega)$, should have regions in Ω without large holes.
Assume Φ, translation invariant, integrable and its Fourier transform decays at infinity with $\beta > d/2$

Theorem

[DeM., Schaback&Wendland, AiCM05]. *For every $\alpha > \beta$ there exists a constant $M_\alpha > 0$ with the following property: if $\epsilon > 0$ and $X = \{x_1, \ldots, x_N\} \subseteq \Omega$ are given such that*

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \leq \epsilon \|f\|_\Phi, \qquad \text{for all } f \in W_2^\beta(\mathbb{R}^d), \tag{6}$$

then the fill distance of X satisfies

$$h_{X,\Omega} \leq M_\alpha \epsilon^{\frac{1}{\alpha - d/2}}. \tag{7}$$

Remarks

1. The interpolation error can be bounded in terms of the fill-distance (cf. e.g., Fasshauer's book, p. 121):

$$\|f - s_{f,X}\|_{L_\infty(\Omega)} \le C\, h_{X,\Omega}^{\beta - d/2} \|f\|_{W_2^\beta(\mathbb{R}^d)}. \qquad (8)$$

   provided $h_{X,\Omega} \le h_0$, for some h_0

2. $M_\alpha \to \infty$ when $\alpha \to \beta$, so from (8) we cannot get $h_{X,\Omega}^{\beta - d/2} \le C\,\epsilon$ but as close as possible.
3. *The proof does not work for gaussians* (no compactly supported functions in the native space of the gaussians).

To remedy, we made the additional assumption that X is already quasi-uniform, i.e. $h_{X,\Omega} \approx q_X$.

- As a consequence, $P_{\Phi,X}(x) \le \epsilon$. The result follows from the lower bounds of $P_{\Phi,X}$ (cf. [Schaback AiCM95] where they are given in terms of q_X).
- Quasi-uniformity brings back to bounds in term of $h_{X,\Omega}$.

Observation: optimally distributed data sites are sets that cannot have a large region in Ω without centers, i.e. $h_{X,\Omega}$ is sufficiently small.

On computing near-optimal points

We studied two algorithms.

1. Greedy Algorithm (GA)
2. Geometric Greedy Algorithm (GGA)

The Greedy Algorithm (GA)

At each step we determine a point where the power function attains its maxima w.r.t. the preceding set. That is,

- **starting step:** $X_1 = \{x_1\}, \ \ x_1 \in \Omega$, *arbitrary*.
- **iteration step:** $X_j = X_{j-1} \cup \{x_j\}$ with $P_{\Phi, X_{j-1}}(x_j) = \|P_{\Phi, X_{j-1}}\|_{L_\infty(\Omega)}$.

Remark: practically we maximize over some very large discrete set $X \subset \Omega$ instead of Ω.

The Geometric Greedy Algorithm (GGA)

The points are computed independently of the kernel Φ.

- **starting step:** $X_0 = \emptyset$ and define $\text{dist}(x, \emptyset) := A,\ A > \text{diam}(\Omega)$.
- **iteration step:** given $X_n \in \Omega, |X_n| = n$ pick $x_{n+1} \in \Omega \setminus X_n$ s.t. $x_{n+1} = \max_{x \in \Omega \setminus X_n} \text{dist}(x, X_n)$. Then, form $X_{n+1} := X_n \cup \{x_{n+1}\}$.

Notice: this algorithm works quite well for subset Ω of cardinality n with small $h_{X,\Omega}$ and large q_X.

On the convergence of GA e GGA

- Experiments showed that the GA fills the currently largest hole in the data set close to the center of the hole and converges at least like

$$\|P_j\|_{L_\infty(\Omega)} \leq C\, j^{-1/d}\,, \quad C > 0.$$

- Defining the *separation distance* for X_j as $q_j := q_{X_j} = \frac{1}{2}\min_{x \neq y \in X_j} \|x - y\|_2$ and the *fill distance* as $h_j := h_{X_j,\Omega} = \max_{x \in \Omega} \min_{y \in X_j} \|x - y\|_2$ then, we proved that

$$h_j \geq q_j \geq \frac{1}{2} h_{j-1} \geq \frac{1}{2} h_j, \quad \forall\, j \geq 2$$

i.e. the GGA produces point sets quasi-uniformly distributed in the euclidean metric.

Connections with (discrete) Leja-like sequences

- Let Ω_N be a discretization of a compact domain of $\Omega \subset \mathbb{R}^d$ and let x_0 arbitrarily chosen in Ω. The sequence of points

$$\min_{0 \le k \le n-1} \|x_n - x_k\|_2 = \max_{x \in \Omega_N \setminus \{x_0, \ldots, x_{n-1}\}} \left(\min_{0 \le k \le n-1} \|x - x_k\|_2 \right) \tag{9}$$

  is known as *Leja-Bos sequence* on Ω_N or *Greedy Best Packing sequence* (cf. López-G.Saff09).

- Hence, the construction technique of GGA is conceptually similar to finding Leja-like sequences : *both maximize a function of distances*.
- The GGA can be generalized to any metric. Indeed, if μ is any metric on Ω, the GGA produces points asymptotically equidistributed in that metric (cf. CDeM.V AMC2005).

How good are the point sets computed by GA and GAA?

We could check these quantities:

- Interpolation error
- Uniformity

$$\rho_{X,\Omega} := \frac{q_X}{h_{X,\Omega}},$$

Notice: *GGA maximizes the uniformity* (since it works well with subsets $\Omega_n \subset \Omega$ with large q_X and small $h_{X,\Omega}$).

- Lebesgue constant

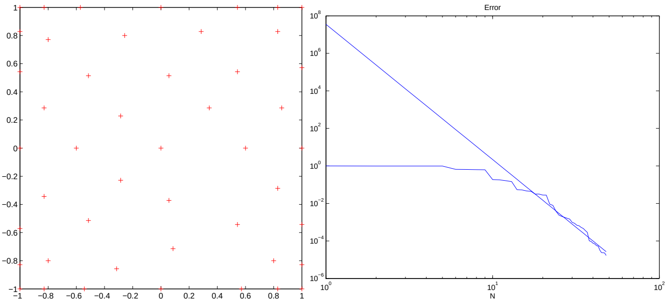
$$\Lambda_N := \max_{x \in \Omega} \lambda_N(x) = \max_{x \in \Omega} \sum_{k=1}^{N} |u_k(x)|\,.$$

Numerical examples in $\mathbb{R}^2$

1. We considered a discretization of $\Omega = [-1, 1]^2$ with 10000 random points.
2. The GA run until $\|P_{X,\Omega}\|_\infty \leq \eta$, η a chosen threshold.
3. The GGA, thanks to the connection with the Leja-like sequences, has been run once and for all. We extracted 406 points from 406^3 random on $\Omega = [-1, 1]^2$, $406 = \dim(\mathbb{P}_{27}(\mathbb{R}^2))$.

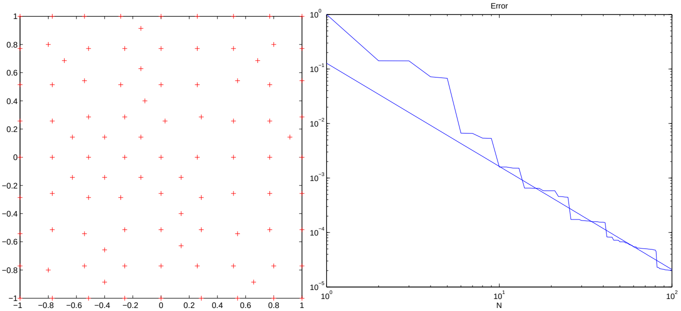
GA: Gaussian

Gaussian kernel with scale 1, 48 points, $\eta = 2 \cdot 10^{-5}$. The "error", in the right–hand figure, is $\|P_N\|^2_{L_\infty(\Omega)}$ which decays as a function of N, the number of data points. The decay, which has been determined by the regression line, behaves like $N^{-7.2}$.



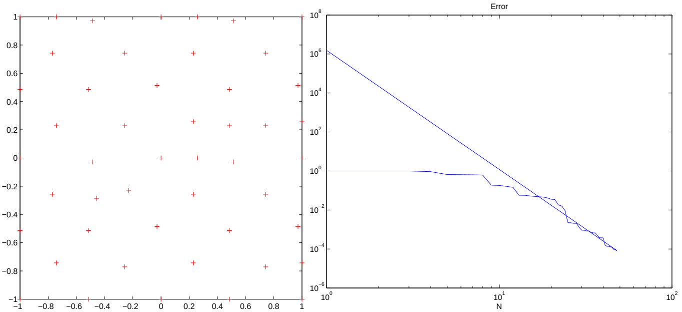
GA: Wendland

C^2 Wendland function scale 15, $N = 100$ points to depress the power function down to $2 \cdot 10^{-5}$. The error decays like $N^{-1.9}$ as determined by the regression line depicted in the right figure.



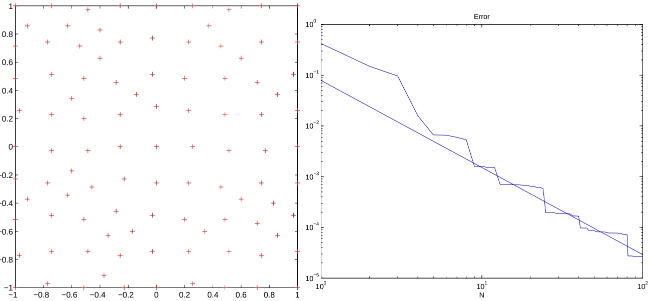
GGA: Gaussian

Error decay when the Gaussian power function is evaluated on the data supplied by the GGA up to X_{48}. The final error is larger by a factor of 4, and the estimated decrease of the error is only like $N^{-6.1}$.



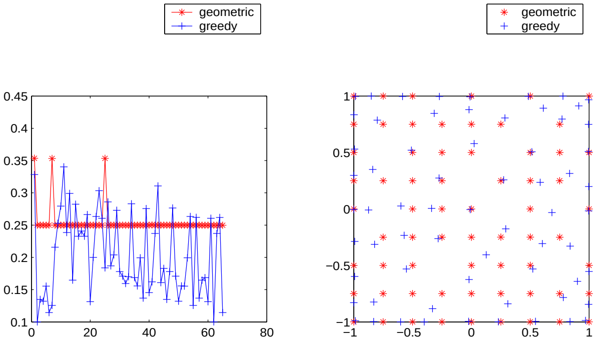
GGA: Wendland

The error factor is *only* 1.4 bigger, while the estimated decay order is -1.72.



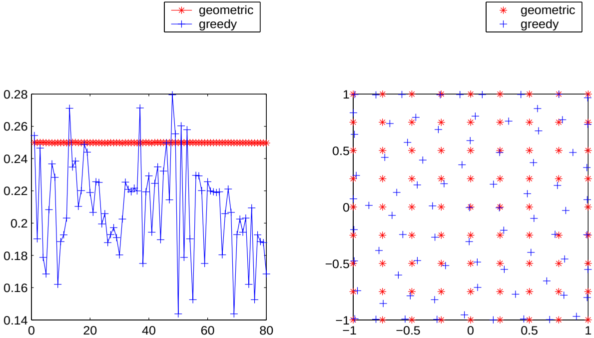
Gaussian

Below: 65 points for the gaussian with scale 1. *Left:* their separation distances; *Right:* the points ($+$) are the one computed with the GA with $\eta = 2.0e-7$, while the ($*$) the one computed with the GGA.



$\mathcal{C}^2$ *Wendland*

Below: 80 points for the Wendland's RBF with scale 1. *Left:* their separation distances; *Right:* the points ($+$) are the one computed with the GA with $\eta = 1.0e - 1$, while the (*) the one computed with the GGA.



For points quasi-uniformly distributed, i.e. points for which $\exists M_1, M_2 \in \mathbb{R}_+$ such that $M_1 \le \frac{h_n}{q_n} \le M_2$, $\forall n \in \mathbb{N}$, holds the following.

Proposition

(cf. [7, Prop. 14.1]) There exists constants c_1 , $c_2 \in \mathbb{R}$, $n_0 \in \mathbb{N}$ such that

$$c_1\, n^{-1/d} \le h_n \le C_2\, n^{-1/d}, \quad \forall n \ge n_0. \tag{10}$$

The GGA algorithm

Defining C_Ω by

$$C_\Omega := \frac{\mathrm{vol}(\Omega) 2^{d+1} \pi \Gamma(d/2+1)}{\alpha \pi^{d/2}},$$

we get

$$C_\Omega \geq n(q_n)^d \geq n(h_n/M_2)^d.$$

Hence,

$$h_n \leq M_2 (C_\Omega/n)^{1/d} = \underbrace{C_\Omega^{1/d} M_2}_{=: C_{\Omega,M_2}} n^{-1/d}\ . \qquad (11)$$

The GGA algorithm

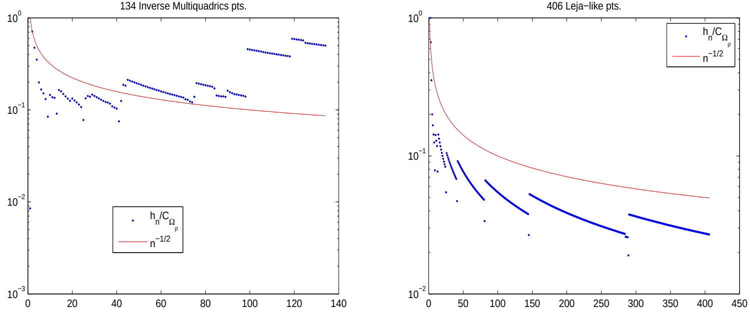


Figure: Plots of $h_n/C_{\Omega,M}$ and $1/\sqrt{n}$ for IM134 and 406Leja-like pts

The GGA algorithm

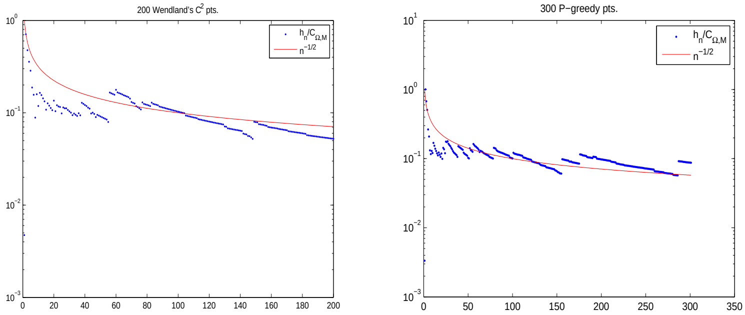


Figure: Plots of $h_n/C_{\Omega,M}$ and $1/\sqrt{n}$ for W2_200 and 300PGreedy pts

Remarks

1. The GGA is independent on the kernel and generates asymptotically equidistributed optimal sequences. It still inferior to the GA that considers the power function.
2. The points generated by the GGA are such that $h_{X_n,\Omega} = \max_{x\in\Omega} \min_{y\in X_n} \|x-y\|_2$.
3. GGA generates sequences with $h_n \le Cn^{-1/d}$, as required by the asymptotic optimality.
4. So far, we have no theoretical results on the asymptotic distribution of points generated by the GA. We are convinced that the use of disk covering strategies could help.

Theorem

(cf. DeM.S08) The classical Lebesgue constant for interpolation with Φ *on* $N = |X|$ *data locations in a bounded* $\Omega \subseteq \mathbb{R}^d$ *has a bound of the form*

$$\Lambda_X \leq C\sqrt{N}\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2}. \qquad (12)$$

For quasi-uniform sets, with uniformity bounded by $\gamma < 1$*, this simplifies to* $\Lambda_X \leq C\sqrt{N}$*.*
Each single cardinal function is bounded by

$$\|u_j\|_{L_\infty(\Omega)} \leq C\left(\frac{h_{X,\Omega}}{q_X}\right)^{\tau-d/2}, \qquad (13)$$

which, in the quasi-uniform case, simplifies to $\|u_j\|_{L_\infty(\Omega)} \leq C$*.*

Corollary

Interpolation on sufficiently many quasi–uniformly distributed data is stable in the sense of

$$\|s_{f,X}\|_{L_\infty(\Omega)} \leq C\left(\|f\|_{\ell_\infty(X)} + \|f\|_{\ell_2(X)}\right) \tag{14}$$

and

$$\|s_{f,X}\|_{L_2(\Omega)} \leq C h_{X,\Omega}^{d/2} \|f\|_{\ell_2(X)} \tag{15}$$

with a constant C independent of X.

- In the right-hand side of (15), ℓ_2 is a properly scaled discrete version of the L_2 norm.
- Proofs have been done by resorting to classical error estimates. An alternative proof based on sampling inequality [Rieger, Wendland NM05], has been proposed in [DeM.Schaback,RR59-08,UniVR].

Lebesgue constants: kernels

1. **Matérn/Sobolev kernel** (*finite smoothness*, definite positive)

$$\Phi(r) = (r/c)^{\nu} K_{\nu}(r/c), \text{ of order } \nu\,.$$

K_{ν} is the *modified Bessel function of second kind*. *Schaback* call them *Sobolev splines*. Examples were done with $\nu = 1.5$ at scale $c = 20, 320$.

2. **Gauss kernel** (*infinite smoothness*, definite positive)

$$\Phi(r) = \mathrm{e}^{-\nu r}, \; \nu > 0\,.$$

Examples with $\nu = 1$ at scale $c = 0.1,\ 0.2,\ 0.4$.

Lebesgue constants

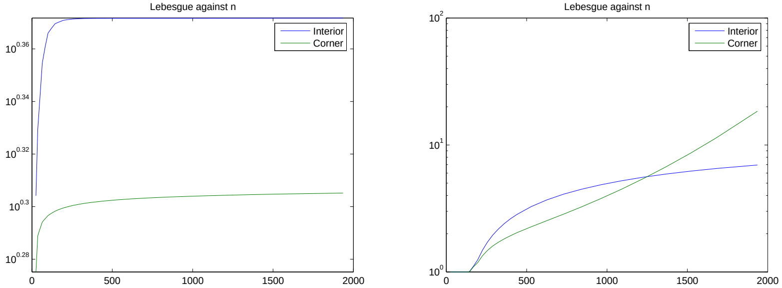


Figure: Lebesgue constants for the Matérn/Sobolev kernel (left) and Gauss kernel (right)

Lebesgue constants

Here we collect some computed Lebesgue constants on a grid of centers consisting of 225 pts on $[-1,1]^2$. The constants were computed on a finer grid made of 7225 pts. Matérn and Wendland had scaled by 10, IMQ and GA scaled by 0.2.

Matern	W2	IMQ	GA
2.3	2.3	2.7	4.3
1.3	1.3	1.3	1.7

First line contains the max of Lebesgue functions. The second are the *estimated* constants, by the Lebesgue function computed by the formula [Wendland's book, p. 208]

$$1 + \sum_{i=1}^{N} (u_j^*(x))^2 \le \frac{P^2_{\Phi,X}(x)}{\lambda_{min}(A_{\Phi,X\cup\{x\}})}, \quad x \notin X\,.$$

in a neighborhood of the point that maximizes the "classical" Lebesgue constant.

Remarks on the finite smooth case

1. In all examples, our bounds on the Lebesgue constants, are confirmed.
2. In all experiments, the Lebesgue constants seem to be uniformly bounded.
3. The maximum of the Lebesgue function is attained in the *interior points*.

Remarks on the infinite smoothness

... things are moreless specular ...

1. The Lebesgue constants do not seem to be uniformly bounded.
2. In all experiments, the Lebesgue function attains its maximum near the *corners* (for large scales).
3. The limit for large scales is called flat limit which corresponds to the Lagrange basis function for polynomial interpolation (see Larsson and Fornberg talks, [Driscoll, Fornberg 2002], [Schaback 2005],...).

A possible solution

Schaback, in a recent paper with S. Müller [Müeller, Scahaback JAT08], studied a Newton's basis for overcoming the ill-conditioning of linear systems in RBF interpolation. The basis is orthogonal in the native space in which the kernel is reproducing and more stable.

Other references

- Driscoll T. and Fornberg B, *Interpolation in the limit of increasingly flat radial basis functions*, Computers Math. Appl. 43 2002, 413-422.
- G. E. Fasshauer, Meshfree Approximation Methods with MATLAB, *Interdisciplinary Mathematical Sciences*, Vol. 6, 2007.
- Nicholas J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996.
- S. Jokar and B. Meheri,*Lebesgue function for multivariate interpolation by RBF*, Appl. Math. Comp. 187(1) 2007, 306–314.
- López García A. and Saff. E. B., *Asymptotics of greedy energy points*, Preprint 2009.
- S. Müller and R. Schaback, *A Newton basis for Kernel Spaces*, J. Approx. Theory, 2009.
- H. Wendland, Scattered Data Approximation, *Cambridge Monographs on Applied and Computational Mathematics*, Vol. 17, 2005.
- H. Wendland, C. Rieger, *Approximate interpolation with applications to selecting smoothing parameters,*, Num. Math. **101** 2005, 643-662.

2nd Dolomites Workshop on Constructive Approximation and Applications

Alba di Canazei (Italy), 4-9 Sept. 2009.

- Keynote speakers: C. de Boor, N. Dyn, G. Meurant, R. Schaback, I. Sloan, N. Trefethen, H. Wendland, Y. Xu
- Sessions on: Polynomial and rational approximation (Org.: *J. Carnicer, A. Cuyt*), Approximation by radial bases (Org.: *A. Iske, J. Levesley*), Quadrature and cubature (Org. *B. Bojanov*†*, E. Venturino*, Approximation in linear algebra (Org. *C. Brezinski, M. Eiermann*).

Happy Birthday Gianpietro!

... and thank you for your attention!