On the Efficiency of Interpolation by Radial Basis Functions

Robert Schaback

Abstract. We study the computational complexity, the error behavior, and the numerical stability of interpolation by radial basis functions. It turns out that these issues are intimately connected. For the case of compactly supported radial basis functions, we consider the possibility of getting reasonably good reconstructions of d-variate functions from N data at $\mathcal{O}(Nd)$ computational cost and give some supporting theoretical results and numerical examples.

§1. Optimal Recovery

Given function values $f(x_1), \ldots, f(x_N)$ on a discrete set $X = \{x_1, \ldots, x_N\}$ of scattered locations $x_j \in \mathbb{R}^d$, we want to recover a function f on some given domain $\Omega \subseteq \mathbb{R}^d$ that contains X. Under certain assumptions to be stated below, an optimal reconstruction takes the form of interpolation by another function $s \in S \subset C(\mathbb{R}^d)$ with $s(x_k) = f(x_k)$, $1 \le k \le N$. Due to the Mairhuber–Curtis [3] theorem, the space S of interpolants must depend on X, as is the case for classical splines and finite elements. However, the space S also depends on the continuity requirements that we shall additionally impose, and these have to match those of f. We fix them by picking a (large) function space $\mathcal{F}$ that contains f and consists of real-valued functions on Ω.

Under the assumptions

- $\mathcal{F}$ is a real Hilbert space,
- the point evaluation functionals δ_x for $x \in \Omega$ are continuous on $\mathcal{F}$ (*i.e.*, elements of the dual space $\mathcal{F}^*$),
- the δ_x are linearly independent if they are distinct,

the space S is optimally chosen for the above recovery problem if it takes the form

$$S = S_{X,\Phi} = \mathrm{span}\{\ \Phi(\cdot - x_j)\ :\ 1 \le j \le N,\ x_j \in X\},$$

Proceedings of Chamonix 1996
A. Le Méhauté, C. Rabut, and L. L. Schumaker (eds.), pp. 309–318.

ISBN 1-xxxxx-xxx-x.

with

$$\Phi(x, y) := (\delta_x, \delta_y)_{\mathcal{F}^*} \tag{1}$$

on $\Omega \times \Omega$. This function is positive definite, meaning that the matrix

$$A_{X,\Phi} = (\Phi(x_k - x_j))_{1 \le j,k \le N} \tag{2}$$

is positive definite for any choice of $X \subset \mathbb{R}^d$ and any number N of data points in X. We note in passing that geometric invariance principles of the function space $\mathcal{F}$ and the domain Ω carry over to invariance principles for the function Φ. The most important cases [25] are

- Translation invariance for $\Omega = \mathbb{R}^d: \ \Phi(x, y) = \phi(x - y)$
- Euclidean invariance for $\Omega = \mathbb{R}^d: \ \Phi(x, y) = \phi(\|x - y\|_2)$, the case of *radial* basis functions
- Orthogonal invariance for spheres $\Omega = S^{d-1} \subset \mathbb{R}^d: \ \Phi(x, y) = \phi(x^T y)$, leading to *zonal* basis functions.

Anyway, the interpolant s^f to f can be written as

$$s^f = \sum_{j=1}^{N} \alpha_j^f \Phi(\cdot - x_j),$$

and the interpolation conditions $s^f(x_k) = f(x_k), \ 1 \le k \le N$ lead to the linear system

$$A_{X,\Phi} \alpha^f = (f(x_1), \ldots, f(x_N))^T \tag{3}$$

for $\alpha^f \in \mathbb{R}^N$ with the matrix of (2). The above setting is optimal in the sense that for any given point $x \in \Omega$ the linear error functional $f \mapsto f(x) - s^f(x)$ has minimal norm in $\mathcal{F}^*$ among all other linear recovery processes using the same information, *i.e.*, when compared to all error functionals of the form

$$f \mapsto f(x) - \sum_{j=1}^{N} g_j(x) f(x_j).$$

Among other optimality properties, this feature is the major link of the above reconstruction process to the theories of information–based complexity [28] and optimal recovery [15]. Because we need the notation anyway, let us look at the *power function* $P_{X,\Phi,g}$ of the above linear process, which is defined as the norm of the above functional, and which depends on $g = (g_1, \ldots, g_n)$. Using (1), the square of the power function is the explicitly available quadratic form

$$\begin{aligned} P^2_{X,\Phi,g}(x) = \|\delta_x - \sum_{j=1}^{N} g_j(x) \delta_{x_j}\|^2_{\mathcal{F}^*} =& \Phi(x, x) - 2 \sum_{j=1}^{N} g_j(x) \Phi(x, x_j) \\ &+ \sum_{j=1}^{N} \sum_{k=1}^{N} g_j(x) g_k(x) \Phi(x_k, x_j) \end{aligned}$$

and its minimum is attained for $g(x) = (g_1(x), \dots, g_n(x))^T$ solving the system

$$A_{X,\Phi} g(x) = (\Phi(x, x_1), \dots, \Phi(x, x_N))^T.$$

But this is the system that is uniquely solved by the functions $u^{opt}(x) = (u_1^{opt}(x), \dots, u_n^{opt}(x))^T$ that make up the Lagrange form of the interpolant

$$s^f = \sum_{j=1}^{N} \alpha_j^f \Phi(\cdot, x_j) = \sum_{j=1}^{N} u_j^{opt}(x) f(x_j).$$

Thus we have the inequality

$$P_{X,\Phi,g}(x) \geq P_{X,\Phi,u^{opt}}(x) =: P_{X,\Phi}(x)$$

betweeen these power functions, and the optimal recovery process is recognized as interpolation by linear combinations of functions $\Phi(\cdot, x_j)$, as required.

§2. Examples

Global radial basis functions $\phi(r)$ with $\phi(\|x-y\|_2) = \Phi(x, y)$ are widely known instances of the above optimal recovery approach, though not fully covered by our simplified presentation here (we omitted *conditional* positive definiteness). If $\mathcal{F}$ is a Beppo-Levi space on $\Omega = \mathbb{R}^d$ induced by a Sobolev-type seminorm, we get the examples [6]

Polyharmonic splines $\phi(r) = r^\beta, \quad \beta \in \mathbb{R}_{>0} \setminus 2\mathbb{N}$

Thin-plate splines $\phi(r) = r^\beta \log r, \quad \beta \in 2\mathbb{N}$

while the spaces $\mathcal{F}$ leading to

Multiquadrics, inverse multiquadrics $\phi(r) = (r^2 + c^2)^{\beta/2}, \quad \beta \in \mathbb{R} \setminus 2\mathbb{N}$

Gaussians $\phi(r) = e^{-\alpha r^2}, \quad \alpha > 0$

are quite small spaces of functions with infinite differentiability. The precise characterization of these spaces requires generalized Fourier transforms and is not included here. Details can be found in the research papers [13,14,32] and in the survey articles [7,8,12,19,20,24]. The most important spaces for applications are the Sobolev spaces $W_2^k(\mathbb{R}^d)$, and these lead to

$$\text{Sobolev splines} \quad \phi(r) = r^{k-d/2} K_{k-d/2}(r), \; k > d/2$$

with the Bessel or Macdonald function K_ν, which has exponential decay towards infinity. Its singularity at zero is compensated by the $r^{k-d/2}$ factor.

In all of the above cases, the matrix $A_{X,\Phi} = (\phi(\|x_j - x_k\|_2))$ is non-sparse in general, which makes the use of the optimal recovery process computationally ineffective. For thin-plate splines, however, there are sophisticated techniques by Beatson, Newsam, and Powell [1,2,21,22] to overcome these problems. A somewhat more direct approach considers compactly supported functions instead of the global ones described above. Examples are

1) Euclidean hat function [29,18,24]
2) Radialized tensor product B-Splines [24]
3) Functions of Wu [31] (plus a related toolbox [27])
4) Functions of Wendland, e.g. $\phi(r) = (1-r)_+^4(4r+1)$ [30]
5) Functions of Buhmann [5]

Case 4 has minimal degree $\lfloor d/2 \rfloor + 3k + 1$ for given smoothness C^{2k} and positive definiteness on $\mathbb{R}^d$, and its related Hilbert space turns out to be norm-equivalent to a Sobolev space (see the contribution of Wendland in this volume). For all of the above functions, the corresponding spaces $\mathcal{F}$ can be formally constructed along the lines of [26] based on [13]. Having compactly supported cases at hand, we devote the rest of the paper to the question of attaining an overall computational complexity of $\mathcal{O}(N \cdot d)$ for a d-variate recovery problem based on N data. Note that this amounts to solving the $N \times N$ system (3) by $\mathcal{O}(N)$ operations. This is the state-of-the-art in preconditioned multilevel techniques for finite elements, and it is a very important open research problem for general recovery processes.

§3. Scaling Effects

Let us consider a translation-invariant setting on $\mathbb{R}^d$ with $\Phi(x,y) = \Phi(y,x) = \phi(x-y)$ for an even function $\phi : \mathbb{R}^d \to \mathbb{R}$ which we further assume to be compactly supported inside the unit ball. We introduce a scaling parameter $\delta > 0$, and write

$$\phi_\delta(\cdot) := \phi(\cdot/\delta)$$

such that the support of ϕ_δ fits into a ball of radius δ. For δ sufficiently small, the matrix A_{X,ϕ_δ} will be diagonal, and then the system(3) is trivially solvable using $\mathcal{O}(N)$ operations. However, the recovery will be a superposition of sharp spikes, and the reproduction quality will be hopelessly bad. Thus the question for computational complexity $\mathcal{O}(N \cdot d)$ has to be recast as a question for the achievable reproduction quality while keeping that complexity fixed.

But there is another problem hidden behind the scenery. The stability of reconstruction, as described by the matrix norm $\|A^{-1}_{X,\phi_\delta}\|$, is closely related to the reproduction quality. An *Uncertainty Relation*, as proven in [23], requires

$$P^2_{X,\phi_\delta}(x)\|A^{-1}_{X\cup\{x\},\phi_\delta}\| \geq 1,$$

and makes recovery processes with small power functions notoriously unstable, if no special precautions like preconditioning are taken. The instability has a strong impact on the $\mathcal{O}(N)$ computational complexity for solving $N \times N$ systems, since uniformly bounded condition numbers are an essential hypothesis that cannot be discarded.

The above discussion forces us to study both the error and the stability in terms of scaling. To this end, we have to look at the standard forms of error and stability bounds. We fix X, Ω and ϕ, dropping these symbols in our notation. In case of error bounds, we can proceed as in [23] and bound the optimal power function in terms of the *data distance*

$$h := \sup_{y\in\Omega} \min_{x_j\in X} \|y - x_j\|$$

by

$$\|P^2\|_{\infty,\Omega} \leq F(h/\delta),$$

where F is a monotonic positive function on $[0,\infty)$ with $F(0) = 0$. Note that the overall error is bounded by

$$\|f - s^f\|_{\infty,\Omega} \leq F(h/\delta)\|f\|_{\mathcal{F}_\delta}$$

with an additional factor $\|f\|_{\mathcal{F}_\delta}$ depending on f and δ, but we concentrate on power function bounds in the sequel. A recent improvement of the proof techniques for error bounds [26] yields an additional factor $F(h/\delta)$ in such a bound, but for a more restrictive norm on f and under assumption of additional boundary conditions.

The stability bounds [18,16,23] are in terms of the *separation distance*

$$q := \min_{x_j\neq x_k\in X} \|x_k - x_j\|,$$

and take the form

$$\|A_X^{-1}\|_2 \leq G^{-1}(q/\delta),$$

where G also is a monotonic positive function on $[0,\infty)$ which vanishes at zero.

If the data are not too wildly scattered, they satisfy an *asymptotic uniformity* condition

$$0 < c \leq \frac{q}{h} \leq 1.$$

This can be used to cast the Uncertainty Relation into the inequality

$$G(\cdot) \leq F(\cdot\sqrt{d})$$

between the functions F and G. In cases where ϕ is of limited smoothness, these functions satisfy an asymptotic relation

$$G(t) = \Theta(t^\beta) = F(t) \tag{4}$$

for sufficiently small t, where β increases with the smoothness of ϕ. Details can be found in [23], but the rule of thumb is that for ϕ generating a space equivalent to $W_2^k(\mathbb{R}^d)$ we have $\beta = k - d/2$.

§4. Efficiency

We now look at special choices for h, q, and δ and their connections. In the *stationary* case, the parameters h and δ are proportional, and then the Uncertainty Relation in the form

$$P^2(x) \geq \|A^{-1}_{X\cup\{x\}}\|^{-1} \geq G(q/\delta)$$

shows us that there will be no convergence of the power function to zero for $h \to 0$. This concides with M. Buhmann's findings [4] on the grid $\mathbb{Z}^d$. But the error can be made small by making the ratio h/δ small, and this amounts to working with increasing bandwidth of the matrix A_X. We can roughly account for the bandwidth by

$$B := (\delta/h)^d$$

for asymptotically uniformly scattered data, and we see that the power function bound can be rephrased in terms of bandwidth B as

$$\|P^2\|_{\infty,\Omega} \leq F(B^{-1/d}),$$

yielding an error factor that decreases with increasing bandwidth. This will be confirmed by numerical evidence in the final section. Under the assumption (4) and ignoring constant factors from now on, we find

$$\|P^2\|_{\infty,\Omega} \leq B^{-\beta/d},$$

such that high-order methods have a strongly positive influence.

Now we look at computational complexity. The condition of the matrix can be roughly estimated by

$$C = B \cdot S = (\delta/h)^d \cdot G^{-1}(q/\delta),$$

where S is the stability, *i.e.*, an upper bound for $\|A_X^{-1}\|$. The computational complexity of, say, the conjugate gradient method then is $\mathcal{O}(N \cdot B \cdot \sqrt{C})$, because each matrix-vector multiplication takes $\mathcal{O}(N \cdot B)$ operations and the number of iterations is proportional to the square root of the condition. Since we assert an overall $\mathcal{O}(N)$ complexity, we look at the computational cost per data item. This then is

$$B \cdot \sqrt{C} = (\delta/h)^{3d/2} \cdot G^{-1/2}(q/\delta),$$

but we still have to relate it to the error behavior. Since this is (at least partially) ruled by P and its bound

$$\epsilon = F^{1/2}(h/\delta),$$

we can combine the above identities under the assumptions (4) and asymptotic uniformity of the data. If we express everything by bandwidth B, we have to keep in mind that large bandwidth B and small error ϵ are intimately related by

$$\epsilon = F^{1/2}(B^{-1/d}).$$

Then the computational complexity per data point and related to the error is

$$\begin{aligned} B \cdot \sqrt{C}/\epsilon &= (\delta/h)^{3d/2} \cdot G^{-1/2}(q/\delta) F^{-1/2}(B^{-1/d}) \\ &= B^{3/2} B^{\beta/d}. \end{aligned}$$

If we express everything by the bound ϵ for the power function, we get

$$B \cdot \sqrt{C}/\epsilon = \epsilon^{-2-3d/\beta}$$

for the computational complexity per data point and related to the error. The consequence of this bound is that higher-order methods should be used for increasing space dimension d, because then β/d can be kept in a feasible range.

§5. Numerical Evidence

The following tables were obtained by interpolation of Franke's function [11] on $N = (2n+1)^2$ regularly distributed data points by Wendland's C^2 function $\phi(r) = (1-r)_+^4(1+4r)$ with varying support scale δ and bandwidth B, the maximum number of nonzero entries in each row/column of the interpolation matrix.

N/B	5	9	13	21	25	29	37	45
9	8.0509	5.4807	5.3385	5.3521	5.3806	5.4089	5.4845	5.5442
25	4.4284	1.2000	0.9532	0.9210	0.8963	0.8719	0.8399	0.8187
81	4.6422	1.1014	0.8827	0.7237	0.6762	0.6310	0.5362	0.4666
289	4.3543	0.8139	0.7562	0.7672	0.7110	0.6605	0.5441	0.4605
1089	4.2275	0.7813	0.7898	0.7921	0.7363	0.6830	0.5643	0.4786
4225	4.2164	0.7791	0.8008	0.8046	0.7486	0.6950	0.5757	0.4883
16641	4.2132	0.7786	0.7791	0.7630	0.7096	0.6580	0.5451	0.4608
66049	4.2123	0.7786	0.7179	0.7658	0.7130	0.6611	0.5477	0.4633

Table 1. Stationary case, maximum error on domain.

The first table contains the stationary case h/δ =const. where the bandwidth B is the relevant parameter. Note that for B fixed there is no convergence for $N \to \infty$, while an increase of B decreases the error gradually. Each entry can be calculated by a $\mathcal{O}(N)$ computational complexity, the factor increasing with bandwidth B.

N/δ	0.0625	0.125	0.25	0.5	1	2	4	8
9	12.9202	12.9202	12.9202	12.0378	5.4724	5.5349	5.8105	5.9030
25	12.8825	12.4130	10.0098	1.1960	0.8177	0.6902	0.6996	0.7186
81	12.8825	11.3262	1.1004	0.4664	0.3616	0.3692	0.3713	0.3718
289	11.4934	0.8139	0.4300	0.1180	0.0433	0.0273	0.0244	0.0255
1089	0.7813	0.4474	0.1175	0.0277	0.0090	0.0070	0.0065	0.0041
4225	0.4562	0.1182	0.0273	-	-	-	-	-
16641	0.1228	-	-	-	-	-	-	-

Table 2. Nonstationary case, maximum error on domain.

The columns of the second table contain the fully nonstationary case δ = const. where the interpolation matrix usually is non-sparse for large δ and cannot be treated with $\mathcal{O}(N)$ computational complexity. The error behavior for $N \to \infty$ is good, but bad condition does not allow large values of N for large δ.

N/B	9	13	21	25	29	37	45
9	5.48067	5.33852	5.35213	5.38056	5.40887	5.48450	5.54424
25	0.93277	0.78306	0.70164	0.69314	0.68419	0.68081	0.68469
81	0.42254	0.38226	0.35683	0.35762	0.36036	0.36862	0.37383
289	0.06797	0.05173	0.03522	0.03304	0.03148	0.03041	0.02889
1089	0.08255	0.02560	0.01866	0.01195	0.01106	0.01054	0.00926
4225	0.03001	0.00974	0.00682	0.00411	0.00375	0.00356	0.00298
16641	0.01142	0.00346	0.00236	0.00136	0.00123	0.00117	0.00094
66049	0.00516	0.00130	0.00083	0.00043	0.00038	0.00036	0.00027

Table 3. Multilevel case, maximum error on domain

The third table uses a stationary scheme again, but proceeds vertically through the columns by interpolating residuals of the previous step at the finer data set. This is the multilevel technique that also has been used by Floater and Iske [9,10], and it shows linear convergence at a $\mathcal{O}(N)$ computational complexity. A proof for this behavior still is missing. First results for a similar technique are in [17].

References

1. Beatson, R. K., and G. N. Newsam, Fast evaluation of radial basis functions: I. Advances in the theory and applications of radial basis functions, Comput. Math. Appl. **24** (1992), 7–19.
2. Beatson, R. K., and G. N. Newsam, Fast evaluation of radial basis functions: Moment based methods, preprint, 1995.

3. Braess, D., *Nonlinear Approximation Theory*, Springer, Berlin–Heidelberg–New York, 1986.
4. Buhmann, M. D., Multivariable Interpolation Using Radial Basis Functions, dissertation, Univ. of Cambridge, 1989.
5. Buhmann, M. D., Radial functions on compact support, Seminar für Angewandte Mathematik, ETH Zürich, Research Report 95–13, 1995.
6. Duchon, J., Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, Rev. Française Automat. Informat. Rech. Opér., Anal. Numer. **10** (1976), 5–12.
7. Dyn, N., Interpolation of scattered data by radial functions, in *Topics in Multivariate Approximation*, C.K. Chui and L.L. Schumaker and Florencio Utreras (eds.), Academic Press, Boston, 1987, 47–61.
8. Dyn, N., Interpolation and approximation by radial and related functions, in *Approximation Theory VI, Vol. 1*, C.K. Chui and L.L. Schumaker and J.D. Ward (eds.) Academic Press, Boston, 1989, 211–234.
9. Floater, M. S., and A. Iske, Multistep scattered data interpolation using compactly supported radial basis functions, to appear in J. Comput. Appl. Math., 1996.
10. Floater, M. S. and A. Iske, Scattered data analysis and industrial applications, in *Advanced Topics in Multivariate Approximation*, F. Fontanella, K. Jetter and P.-J. Laurent (eds.), World Scientific Publishing, 1996, 1–10.
11. Franke, R., Scattered data interpolation. test of some methods, Math. Comp. **38** (1982), 181–200.
12. Light, W., Using radial functions on compact domains, in *Wavelets, Images, and Surface Fitting*, P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker (eds.), A. K. Peters, Wellesley MA, 1994, 351–370.
13. Madych, W. R., and S. A. Nelson, Multivariate interpolation: a variational theory, manuscript, 1983.
14. Madych, W. R., and S. A. Nelson, Multivariate interpolation and conditionally positive definite functions II, Math. Comp. **54** (1990), 211–230.
15. Micchelli, C. A., and T. J. Rivlin, Lectures on optimal recovery, in *Numerical Analysis*, Lancaster 1984, P.R. Turner (ed.), Springer Verlag, Lecture Notes in Mathematics 1129, 1984, 12–93.
16. Narcowich, F. J., Sivakumar, N., and J. D. Ward, Norm of inverses and condition numbers for matrices associated with scattered data, J. Approx. Theory **64**, 1991, 69–94.
17. Narcowich, F. J., Schaback, R., and J .D. Ward, Multilevel approximation, preprint, Texas A&M University, 1996.
18. Narcowich, F. J., and J. D. Ward, Norm estimates for the inverses of a general class of scattered–data radial–function interpolation matrices, J. Approx. Theory **69** (1992), 84–109.

19. Powell, M. J. D., Radial basis functions for multivariable interpolation: a review, in *Numerical Analysis*, D. F. Griffiths and G. A. Watson (eds.), Longman Scientific & Technical (Harlow), 1987, 223–241.

20. Powell, M. J. D., The theory of radial basis function approximation in 1990, in *Advances in Numerical Analysis II: Wavelets, Subdivision Algorithms, and Radial Basis Functions*, W.A. Light (ed.), Oxford Univ. Press, Oxford, 1992, 105–210.

21. Powell, M. J. D., Tabulation of thin plate splines on a very fine two–dimensional grid, in *Numerical Methods of Approximation Theory*, D. Braess and L.L. Schumaker (eds.), Birkhäuser, Basel 1992, 221–244.

22. Powell, M. J. D., Truncated Laurent expansions for the fast evaluation of thin plate splines, Numer. Algorithms **5** (1993), 99–120.

23. Schaback, R., Error Estimates and Condition Numbers for Radial Basis Function Interpolation, Advances in Comp. Math. **3** (1995), 251–264.

24. Schaback, R., Creating surfaces from scattered data using radial basis functions, in *Mathematical Methods for Curves and Surfaces*, Morten Dæhlen, Tom Lyche, Larry L. Schumaker (eds.), Vanderbilt University Press, Nashville & London, 1995, 477–496.

25. Schaback, R., Optimal recovery in translation–invariant spaces of functions, preprint 1996, to appear in a Festschrift on the occasion of T. Rivlin's 70th birthday.

26. Schaback, R., Improved error bounds for scattered data interpolation by radial basis functions, Preprint, Göttingen, 1996.

27. Schaback, R. and Z. Wu, Operators on radial functions, J. Comput. Appl. Math. **73** (1996), 257–270.

28. Traub, J. F., and H. Woźniakowski, *A General Theory of Optimal Algorithms*, Academic Press, New York, London, Toronto, Sydney, San Francisco, Elsevier, New York, 1980.

29. Wendland, H., Ein Beitrag zur Interpolation mit radialen Basisfunktionen, Diplomarbeit, Universität Göttingen, 1994.

30. Wendland, H., Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree, Advances in Comp. Math. **4** (1995), 389–396.

31. Wu, Z., Multivariate compactly supported positive definite radial functions, Advances in Comp. Math. **4** (1995), 283–292.

32. Wu, Z., and R. Schaback, Local error estimates for radial basis function interpolation of scattered data, IMA J. Numer. Anal. **13** (1993), 13–27.

Robert Schaback
Universität Göttingen, Lotzestraße 16-18
D-37083 Göttingen, GERMANY
`schaback@math.uni-goettingen.de`