View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Strathclyde Institutional Repository

University of Strathclyde Glasgow

Strathprints Institutional Repository

Davydov, Oleg (2007) *Error bound for radial basis interpolation in terms of a growth function.* In: Curve and surface fitting. Nashboro Press, Brentwood, U.S.A., pp. 121-130. ISBN 978-0-9728482-8-2

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (http://strathprints.strath.ac.uk/) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: mailto:strathprints@strath.ac.uk

Error Bound for Radial Basis Interpolation in Terms of a Growth Function*

Oleg Davydov

February 16, 2007

Abstract

We suggest an improvement of Wu-Schaback local error bound for radial basis interpolation by using a polynomial growth function. The new bound is valid without any assumptions about the density of the interpolation centers. It can be useful for the localized methods of scattered data fitting and for the meshless discretization of partial differential equations.

1 Introduction

Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathbb{R}^d$, and let f_j, $j = 1, \dots, N$ be real data values associated with the respective points $\mathbf{x}_j$.

Suppose $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a *radial basis function*, i.e., a positive definite function or a conditionally positive definite function of order $s = 1, 2, \dots$ on $\mathbb{R}^d$, see e.g. [1]. If ϕ is positive definite, we set $s = 0$. Radial basis interpolant has the form

$$r_{\phi,\ell}(\cdot) = \sum_{j=1}^{N} a_j \phi(\|\cdot - \mathbf{x}_j\|_2) + \sum_{j=1}^{m} b_j p_j(\cdot), \qquad \ell \geq s - 1, \tag{1}$$

where $m = 0$ if $\ell = -1$, and $m = \binom{d+\ell}{d}$ otherwise, with $\{p_1, \dots, p_m\}$ in the latter case being a basis for the space Π_ℓ^d of d-variate polynomials of total degree ℓ. The coefficients $\{a_j\}$ and $\{b_j\}$ in (1) are determined from the conditions

$$r_{\phi,\ell}(\mathbf{x}_j) = f_j, \qquad j = 1, \dots, N, \tag{2}$$

and

$$\sum_{j=1}^{N} a_j \, p(\mathbf{x}_j) = 0, \qquad \text{all } p \in \Pi_\ell^d. \tag{3}$$

This is uniquely solvable (see e.g. [1]) under the assumptions that $N \geq m$ and $\mathbf{X}$ is a *norming set* for Π_ℓ^d, i.e. for any $p \in \Pi_\ell^d$, $p|_{\mathbf{X}} = 0$ implies $p \equiv 0$.

---

*University of Strathclyde, Department of Mathematics, 26 Richmond Street, Glasgow G1 1XH, Scotland, UK, `oleg.davydov@strath.ac.uk`, `http://www.maths.strath.ac.uk/~aas04108/`

Radial basis functions are becoming an increasingly popular tool for the approximation of functions, see [1, 17]. The literature studying both their effectiveness in applications and their remarkable mathematical theory is steadily growing.

In this paper we are interested in local error bounds for the radial basis interpolation in the form going back to [18]. We introduce an error bound where a polynomial growth function explicitly enters the estimate, see (4). In contrast to the standard error estimates in terms of fill distance, our bound is valid without any assumptions about the density of the interpolation centers in the domain of interest. The standard bounds can be easily deduced from (4), avoiding complicated techniques of 'local polynomial reproduction.'

The estimate (4) is of special interest for the localized methods of data fitting with radial basis functions, in particular for the design of effective two stage methods using either multivariate splines [15], [3]–[7] or a partition of unity [8, 10, 16].

Another potential application area is the meshless discretization of partial differential equations, where (4) shows how the error bound is influenced by the locations for the centers of radial basis functions, and this can help choosing them in an optimal way.

We present the main result of the paper in Section 2, and its proof in Section 4. In Section 3 we consider various estimates of the growth function. In particular, we show that the error bounds in terms of the fill distance can be obtained by appropriately estimating the growth function in (4). We also discuss a computable estimate of the growth function for small subsets of data, which is related to the error bounds for local discrete least squares polynomials introduced in [2].

2 Error Bound

For any non-empty $\mathbf{Y} \subset \mathbb{R}^d$, we denote by $\rho_q(\mathbf{x}, \mathbf{Y})$ the *growth function* of Π_q^d with respect to $\mathbf{Y}$,

$$\rho_q(\mathbf{x}, \mathbf{Y}) := \max\{|p(\mathbf{x})| \, : \, p \in \Pi_q^d, \, \|p|_{\mathbf{Y}}\|_\infty \leq 1\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Clearly, $\rho_q(\mathbf{x}, \mathbf{Y})$ is finite for all $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{Y}$ is a norming set for Π_q^d. Otherwise, $\rho_q(\mathbf{x}, \mathbf{Y}) = \infty$ for all $\mathbf{x} \notin \mathbf{Y}$. Note that in the case when $\#\mathbf{Y} = \dim \Pi_q^d$, $\rho_q(\mathbf{x}, \mathbf{Y})$ coincides with the standard *Lebesgue function* for polynomial interpolation with knots in $\mathbf{Y}$.

We set

$$\mathcal{F}_\phi = \{f \in L_2(\mathbb{R}^d) \, : \, \|f\|_\phi < \infty\},$$

where

$$\|f\|_\phi := (2\pi)^{-d/2} \left\| \hat{f} / \sqrt{\hat{\Phi}} \right\|_{L_2(\mathbb{R}^d)}, \qquad f \in L_2(\mathbb{R}^d),$$

with $\Phi(\cdot) = \phi(\| \cdot \|_2)$, and $\hat{f}$ denotes the generalized Fourier transform. The latter is given for any $f \in L_1(\mathbb{R}^d)$ by the usual formula

$$\hat{f}(\mathbf{x}) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-i\mathbf{x}\cdot\mathbf{t}} f(\mathbf{t}) \, \mathbf{dt}, \qquad \mathbf{x} \in \mathbb{R}^d,$$

and it is defined in a distributional sense for certain classes of functions non-integrable on $\mathbb{R}^d$, see [17, Section 8.2].

Finally, $E(f,\mathcal{S})_{C(G)}$, where $G \subset \mathbb{R}^d$, denotes the *error of the best uniform approximation* to f from a linear space $\mathcal{S}$ of functions on G,

$$E(f,\mathcal{S})_{C(G)} := \inf_{g\in\mathcal{S}} \|f-g\|_{C(G)}.$$

Here $C(G)$ denotes the space of continuous functions on G, and $\|f\|_{C(G)} := \max_{x\in G} |f(x)|$.

Our main result is the following error bound.

Theorem 1. *Assume that* $\mathbf{f}_j = f(\mathbf{x}_j)$, $j = 1,\ldots,N$, *for a function* $f \in \mathcal{F}_\phi$. *Then for any non-empty* $\mathbf{Y} \subseteq \mathbf{X}$, *and any* $q \geq \max\{\ell, 0\}$, *we have*

$$|f(\mathbf{x}) - r_{\phi,\ell}(\mathbf{x})| \leq \Big(1 + \rho_q(\mathbf{x},\mathbf{Y})\Big) \sqrt{E(\Phi,\Pi_q^d)_{C(B_{\mathbf{x},\mathbf{Y}})}}\, \|f\|_\phi, \quad \mathbf{x} \in \mathbb{R}^d, \tag{4}$$

where $B_{\mathbf{x},\mathbf{Y}}$ *denotes the ball in* $\mathbb{R}^d$ *with center* $\mathbf{0}$ *and radius* $\operatorname{diam}(\{\mathbf{x}\}\cup\mathbf{Y})$.

We postpone the proof of this theorem to Section 4.

Standard error bounds for the radial basis interpolation, e.g. those in terms of fill distance can be obtained from (4) by appropriately estimating the growth function $\rho_q(\mathbf{x},\mathbf{Y})$. We discuss these and other estimates for $\rho_q(\mathbf{x},\mathbf{Y})$ in Section 3.

An interesting feature of (4) is that it can be applied in various ways resembling classical notions of h-, p- and spectral convergence. Suppose we consider an approximation process, where f is interpolated by $r_{\phi,\ell}$ with $\mathbf{X}$ becoming denser and denser in Ω. Let M be a positive constant. If we fix q and choose $\mathbf{Y}$ such that $\rho_q(\mathbf{x},\mathbf{Y}) \leq M$ and $h = \operatorname{diam}(\{\mathbf{x}\}\cup\mathbf{Y})$ is as small as possible, then we arrive at *h-convergence.* Indeed, $E(\Phi,\Pi_q^d)_{C(B_{\mathbf{x},\mathbf{Y}})}$ will decay as a certain power of h depending on the smoothness of Φ. To obtain *p-convergence*, we fix a neighborhood U of $\mathbf{x}$, take $\mathbf{Y} = \mathbf{X}\cap U$, and choose q as large as possible with $\rho_q(\mathbf{x},\mathbf{Y}) \leq M$. As $\mathbf{X}$ becomes denser, larger q can be chosen, and $E(\Phi,\Pi_q^d)_{C(B_{\mathbf{x},\mathbf{Y}})}$ can be bounded as a certain power of $1/q$ if Φ has finite smoothness, or it will decay exponentially if Φ is an analytic function.

It is remarkable that both h- and p-convergence happen for the same approximation method (1). Indeed, the estimate (4) is correct for *all* q and $\mathbf{Y}$, and we do not need to know their optimal values in practice. This compares favorably to piecewise polynomial methods, where specific algorithms are needed to design partitions of Ω and choose appropriate polynomial degrees for either h- or p- or, say, hp-convergence.

3 Estimates of Growth Function

In this section we discuss various estimates for $\rho_q(\mathbf{x},\mathbf{Y})$, in particular those that lead to the standard error bounds for radial basis interpolation on bounded domains [1, 17]. We emphasise that the estimates for $\rho_q(\mathbf{x},\mathbf{Y})$ are not only of theoretical interest. Indeed, whenever we can control the placement of the data sites $\mathbf{X}$, the error bounds will be better if we manage to prevent $\rho_q(\mathbf{x},\mathbf{Y})$ from blowing up for largest possible q and smallest possible $\mathbf{Y}$. Such a control is available in localized data fitting, where the algorithms choose appropriate subsets of data to build local approximations. Another important situation where one controls the placement of the data sites is the discretization of partial differential equations.

Clearly, $\rho_q(\mathbf{x},\mathbf{Y}) \geq 1$. This lower bound is achieved when $q = 0$ as we have $\rho_0(\mathbf{x},\mathbf{Y}) = 1$ for any $\mathbf{x} \in \mathbb{R}^d$. This simplest version of (4) has been discussed in [4]. We also have $\rho_1(\mathbf{x},\mathbf{Y}) = 1$ if $\mathbf{x}$ belongs to the convex hull of $\mathbf{Y}$. This fact, together with Theorem 1, can be used to prove error bounds similar to those of [13]. Unfortunately, in general there are no such simple estimates for $\rho_q(\mathbf{x},\mathbf{Y})$, $q \geq 2$.

Assuming that we are only interested in $\mathbf{x}$ in a bounded domain $\Omega \subset \mathbb{R}^d$ such that $\mathbf{X} \subset \Omega$, we observe that the diameter of the set $\{\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_N\}$ is less or equal to the diameter d_Ω of Ω. Since we can always take $\mathbf{Y} = \mathbf{X}$, it follows from Theorem 1 that

$$\|f - r_{\phi,\ell}\|_{C(\Omega)} \leq \Big(1 + \rho_q(\Omega,\mathbf{X})\Big) \sqrt{E(\Phi, \Pi_q^d)_{C(B_{d_\Omega})}}\, \|f\|_\phi, \tag{5}$$

where

$$\rho_q(\Omega,\mathbf{X}) := \max\{\|p\|_{C(\Omega)} \,:\, p \in \Pi_q^d,\ \|p|_\mathbf{X}\|_\infty \leq 1\},$$

and B_r denotes the ball in $\mathbb{R}^d$ with center $\mathbf{0}$ and radius r. Note that $\rho_q(\Omega,\mathbf{X})$ is closely related to the *norming constant* $\nu_q(\Omega,\mathbf{X})$ [9] defined by

$$\nu_q(\Omega,\mathbf{X}) := \min\{\|p|_\mathbf{X}\|_\infty \,:\, p \in \Pi_q^d,\ \|p\|_{C(\Omega)} = 1\}.$$

Clearly,

$$\rho_q(\Omega,\mathbf{X}) = \nu_q^{-1}(\Omega,\mathbf{X}). \tag{6}$$

Upper bounds for $\rho_q(\Omega,\mathbf{X})$ can be obtained under certain assumptions on the *fill distance* $h(\Omega,\mathbf{X})$ of $\mathbf{X}$ with respect to Ω, where

$$h(\Omega,\mathbf{X}) := \sup_{\mathbf{y}\in\Omega} \inf_{\mathbf{x}\in\mathbf{X}} \|\mathbf{y} - \mathbf{x}\|_2.$$

In the case when Ω is a cube in $\mathbb{R}^d$, a result from [11] can be used. Its consequence is that $h(\Omega,\mathbf{X}) < \frac{a}{2\gamma_d(q+1)}$ implies $\rho_q(\Omega,\mathbf{X}) \leq e^{2d\gamma_d(q+1)}$, where a is the sidelength of the cube, and γ_d is defined recursively as $\gamma_1 = 2$, $\gamma_d = 2d(1 + \gamma_{d-1})$, $d \geq 2$, see [11, Lemma 1]. Another approach originated in [9] is based on an elegant application of Markov inequality for polynomials. It can be used on any domains satisfying the interior cone condition. For example, in the case when Ω is a ball of radius r in $\mathbb{R}^d$, it can be shown that $h(\Omega,\mathbf{X}) \leq \frac{\sqrt{3}r}{4(2+\sqrt{3})q^2}$ implies $\rho_q(\Omega,\mathbf{X}) \leq 2$, see [17, Corollary 3.11].

These upper bounds on $\rho_q(\Omega,\mathbf{X})$, combined with (4), can be used to obtain the standard error bounds in terms of fill distance, such as those in [1, Theorem 5.5], [17, Section 11.3] or the spectral convergence orders for multiquadric and Gaussian in [17, Section 11.4]. Indeed, the second factor in (4) can be estimated following the same argumentation as e.g. in the proofs in [17, Chapter 11]. Comparing to the standard method of proof, the approach based on (4) is simpler because it does not require the complicated techniques of local polynomial reproduction [17, Chapter 3].

Despite their theoretical significance, the bounds on $\rho_q(\Omega,\mathbf{X})$ in terms of fill distance do not seem to be of much practical use. Consider, for example, the case $q = 2$ in $\mathbb{R}^2$. For the unit square the assumption $h(\Omega,\mathbf{X}) < \frac{a}{2\gamma_d(q+1)}$ of [11] becomes $h(\Omega,\mathbf{X}) < 1/72$. Obviously, one can place at least $36^2 = 1296$ disjoint circles of radius $1/36$ inside the unit square. Since $h(\Omega,\mathbf{X}) < 1/72$, each of these circles must contain a point in $\mathbf{X}$, which implies that $\#\mathbf{X} \geq 1296$. Similarly, if Ω is a circle of radius 1, then $h(\Omega,\mathbf{X}) \leq \frac{\sqrt{3}r}{4(2+\sqrt{3})q^2}$

becomes $h(\Omega,\mathbf{X})\le 0.029$ for $q=2$, and a crude lower bound is $\#\mathbf{X}\ge 594$. Recall that $\dim\Pi_2^2$ is only 6, which suggests that a set $\mathbf{X}$ consisting of 6 points in good position should have a reasonably bounded $\rho_2(\Omega,\mathbf{X})$. In practice the growth function $\rho_2(\mathbf{x},\mathbf{X})$ with respect to great many datasets consisting of, say, 10 points spread out near $\mathbf{x}$ will be reasonably small. Of course, $\rho_2(\mathbf{x},\mathbf{X})$ may be large if the point constellation is unfortunate, i.e. if $\mathbf{X}$ is close to a subset of a conic section. However, it would hardly be practical, especially for a localized method of data fitting, to use hundreds of points instead of just several in order to avoid these bad constellations. Instead, one needs to know the exact value of $\rho_q(\Omega,\mathbf{X})$ or a good estimate for it, as an indicator of whether the points in $\mathbf{X}$ are in good position.

In a localized method, the estimate (5) is applied to a number of small subdomains ω_i, $i=1,\dots,I$, of the original domain Ω. Knowing $\rho_q(\omega_i,\mathbf{X})$, one can sensibly decide whether more distant points should be invoked to improve a local approximation in ω_i, or some points can be discarded to save computation cost without causing significant damage to the approximation quality.

A computable estimate for $\rho_q(\Omega,\mathbf{X})$ in the case when $\#\mathbf{X}$ is small is provided by the reciprocal of the minimum singular value of the polynomial collocation matrix

$$P_{\mathbf{X}} := \begin{bmatrix} p_1(\mathbf{x}_1) & \dots & p_m(\mathbf{x}_1) \\ \vdots & & \vdots \\ p_1(\mathbf{x}_N) & \dots & p_m(\mathbf{x}_N) \end{bmatrix}.$$

More precisely, in view of (6) it follows from [2, Eq. (2.7)] that

$$K_1\sigma_{\min}^{-1}(P_{\mathbf{X}}) \le \rho_q(\Omega,\mathbf{X}) \le K_2\sqrt{\#\mathbf{X}}\,\sigma_{\min}^{-1}(P_{\mathbf{X}}), \tag{7}$$

where $\sigma_{\min}(P_{\mathbf{X}})$ is the minimum singular value of the matrix $P_{\mathbf{X}}$, and K_1, K_2 are positive constants such that

$$K_1\|a\|_2 \le \|\sum_{j=1}^m a_j p_j\|_{C(\Omega)} \le K_2\|a\|_2$$

for all coefficient vectors $a=(a_1,\dots,a_m)^T\in\mathbb{R}^m$.

Note that (7) remains valid if $\rho_q(\Omega,\mathbf{X})$ is replaced with the norm of the polynomial least squares operator, see [2]. Therefore, the quality of local least squares polynomial approximations can be judged on the basis of the size of $\sigma_{\min}(P_{\mathbf{X}})$, which has been exploited in the data fitting method of [6].

4 Proof of Theorem 1

We first prove the following two lemmas, and then proceed to prove the theorem.

Lemma 2. *Let X be a finite dimensional vector space and X^* its dual. Suppose that $X^*=\operatorname{span}\{\lambda_1,\dots,\lambda_k\}$ for some $\lambda_1,\dots,\lambda_k\in X^*$. Then for any functional $\lambda\in X^*$ we have*

$$\max_{\substack{x\in X\\ |\lambda_i(x)|\le 1,\, i=1,\dots,k}} |\lambda(x)| = \min_{\substack{c\in\mathbb{R}^k\\ \lambda=\sum_{i=1}^k c_i\lambda_i}} \sum_{i=1}^k |c_i|. \tag{8}$$

Proof. Let us turn X into a normed space by introducing the norm $\|x\|_X = \max_i |\lambda_i(x)|$. Then the left hand side of (8) is $\|\lambda\|_{X^*}$. Obviously, $\|\lambda_i\|_{X^*} \le 1$, $i = 1, \ldots, k$, which implies $\|\lambda\|_{X^*} \le \sum_{i=1}^k |c_i|$ for any $c = (c_1, \ldots, c_k) \in \mathbb{R}^k$ such that $\lambda = \sum_{i=1}^k c_i\lambda_i$. If $\dim X = k$, then $\{\lambda_1, \ldots, \lambda_k\}$ is a basis for X^*, and (8) is easily recognized as the duality relation for $\|\lambda\|_{X^*}$. In general, similar to [9, Proposition 2] and [12, Proposition 4.1], we consider the linear "sampling operator" $T : X \to \mathbb{R}^k$, where $T(x) = (\lambda_1(x), \ldots, \lambda_k(x))$. Let $\tilde{X} = T(X) \subset \mathbb{R}^k$. Since $\{\lambda_1, \ldots, \lambda_k\}$ is a spanning set for X^*, T is injective and hence $T^{-1} : \tilde{X} \to X$ exists. Moreover, the functionals $\lambda_i \circ T^{-1}$, $i = 1, \ldots, k$, on $\tilde{X}$ are the restrictions of coordinate projections Λ_i, $i = 1, \ldots, k$ on $\mathbb{R}^k$, and the norm $\|\cdot\|_{\tilde{X}}$ on $\tilde{X}$ defined by $\|\tilde{x}\|_{\tilde{X}} = \|T^{-1}(\tilde{x})\|_X$ coincides with the norm $\|\cdot\|_\infty$ induced from $\mathbb{R}^k$. Therefore, $\tilde{\lambda} := \lambda \circ T^{-1}$, as a linear functional on $\tilde{X}$ with $\|\tilde{\lambda}\|_{\tilde{X}^*} = \|\lambda\|_{X^*}$, is extendible by Hahn-Banach theorem to a linear functional Λ on $\mathbb{R}^k$ with norm $\|\Lambda\|_1 = \|\lambda\|_{X^*}$. Now, there exists a unique representation $\Lambda = \sum_{i=1}^k \tilde{c}_i\Lambda_i$, $\tilde{c}_i \in \mathbb{R}$, with $\|\Lambda\|_1 = \sum_{i=1}^k |\tilde{c}_i|$, and, hence, $\lambda = \sum_{i=1}^k \tilde{c}_i\lambda_i$ and $\|\lambda\|_{X^*} = \sum_{i=1}^k |\tilde{c}_i|$. ■

Lemma 3. *Let* $\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ *and* $c_1, \ldots, c_n \in \mathbb{R}$. *Suppose that*

$$p(\mathbf{x}) = \sum_{j=1}^n c_j p(\mathbf{x}_j) \qquad \text{for all } p \in \Pi_q^d. \tag{9}$$

Then for all $p \in \Pi_q^d$ *we have*

$$p(\mathbf{0}) - 2\sum_{j=1}^n c_j\, p(\mathbf{x} - \mathbf{x}_j) + \sum_{j,k=1}^n c_j c_k\, p(\mathbf{x}_j - \mathbf{x}_k) = 0. \tag{10}$$

Proof. If $p \in \Pi_q^d$ and $\mathbf{y} \in \mathbb{R}^d$, then both $p(\cdot - \mathbf{y})$ and $p(\mathbf{y} - \cdot)$ belong to Π_q^d, and it follows from (9) that

$$p(\mathbf{x} - \mathbf{y}) = \sum_{j=1}^n c_j p(\mathbf{x}_j - \mathbf{y}), \qquad p(\mathbf{y} - \mathbf{x}) = \sum_{j=1}^n c_j p(\mathbf{y} - \mathbf{x}_j).$$

By taking $\mathbf{y} = \mathbf{x}$ in the second identity and $\mathbf{y} = \mathbf{x}_k$ in the first, we get

$$p(\mathbf{0}) = \sum_{j=1}^n c_j p(\mathbf{x} - \mathbf{x}_j), \quad p(\mathbf{x} - \mathbf{x}_k) = \sum_{j=1}^n c_j p(\mathbf{x}_j - \mathbf{x}_k), \quad k = 1, \ldots, n,$$

and (10) is easily verified. ■

Proof of Theorem 1. Let $\mathbf{x} \in \mathbb{R}^d \setminus \mathbf{X}$. As shown in [18] (see also [17]),

$$|f(\mathbf{x}) - r_{\phi,\ell}(\mathbf{x})| \le P(\mathbf{x})\, \|f\|_\phi, \tag{11}$$

with $P(\mathbf{x})$ being the *power function* that satisfies

$$P(\mathbf{x}) = \min\{\sqrt{F(c)} \;:\; c \in \mathbb{R}^N,\, p(\mathbf{x}) = \sum_{j=1}^N c_j\, p(\mathbf{x}_j) \text{ for all } p \in \Pi_\ell^d\}, \tag{12}$$

where for $c = (c_1, \dots, c_N) \in \mathbb{R}^N$,

$$F(c) := \Phi(\mathbf{0}) - 2\sum_{j=1}^{N} c_j \Phi(\mathbf{x} - \mathbf{x}_j) + \sum_{j,k=1}^{N} c_j c_k \Phi(\mathbf{x}_j - \mathbf{x}_k).$$

We now take a $q \geq \ell$ and choose any subset $\mathbf{Y} \subseteq \mathbf{X}$ such that $\rho_q(\mathbf{x}, \mathbf{Y}) < \infty$. Assume without loss of generality that $\mathbf{Y} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, where $n \leq N$. Clearly, the condition $\rho_q(\mathbf{x}, \mathbf{Y}) < \infty$ holds if and only if $\mathbf{Y}$ is a norming set for Π_q^d. Therefore the mapping $\delta_{\mathbf{Y}} : \Pi_q^d \to \mathbb{R}^n$ defined by $\delta_{\mathbf{Y}}(p) = p|_{\mathbf{Y}}$ is injective, and its image has dimension $\binom{d+q}{d} = \dim \Pi_q^d$. This implies that among the point evaluation functionals $\delta_{\mathbf{x}_j} : \Pi_q^d \to \mathbb{R}$, $j = 1, \dots, n$, that form the components of $\delta_{\mathbf{Y}}$, there are $\binom{d+q}{d}$ that are linearly independent over Π_q^d. Therefore, $\{\delta_{\mathbf{x}_j}\}_{j=1}^n$ span the dual space $(\Pi_q^d)^*$. Now, the linear functional $\delta_{\mathbf{x}}$ defined by $\delta_{\mathbf{x}}(p) = p(\mathbf{x})$ is also in $(\Pi_q^d)^*$, and hence it can be written as a linear combination of $\delta_{\mathbf{x}_j}$, $j = 1, \dots, n$. We conclude that there exist vectors $c \in \mathbb{R}^N$ satisfying

$$p(\mathbf{x}) = \sum_{j=1}^{n} c_j \, p(\mathbf{x}_j) \qquad \text{for all } \ p \in \Pi_q^d \tag{13}$$

and

$$c_j = 0, \qquad \text{for all } j = n+1, \dots, N. \tag{14}$$

Let us fix for a moment one such $c \in \mathbb{R}^N$. In view of (13), Lemma 3 implies that for any $p \in \Pi_q^d$,

$$p(\mathbf{0}) - 2\sum_{j=1}^{n} c_j \, p(\mathbf{x} - \mathbf{x}_j) + \sum_{j,k=1}^{n} c_j c_k \, p(\mathbf{x}_j - \mathbf{x}_k) = 0.$$

Since $\Pi_\ell^d \subset \Pi_q^d$, we obtain by taking into account (14),

$$\begin{aligned} F(c) &= [\Phi(\mathbf{0}) - p(\mathbf{0})] - 2\sum_{j=1}^{n} c_j \, [\Phi(\mathbf{x} - \mathbf{x}_j) - p(\mathbf{x} - \mathbf{x}_j)] \\ &\quad + \sum_{j,k=1}^{n} c_j c_k \, [\Phi(\mathbf{x}_j - \mathbf{x}_k) - p(\mathbf{x}_j - \mathbf{x}_k)] \\ &\leq \Big(1 + \sum_{j=1}^{n} |c_j|\Big)^2 \|\Phi - p\|_{C(B_{\mathbf{x},\mathbf{Y}})}. \end{aligned}$$

Since $p \in \Pi_q^d$ is arbitrary, it follows that

$$F(c) \leq \Big(1 + \sum_{j=1}^{n} |c_j|\Big)^2 E(\Phi, \Pi_q^d)_{C(B_{\mathbf{x},\mathbf{Y}})}$$

for any $c \in \mathbb{R}^N$ such that (13) and (14) hold.

By Lemma 2, where we take $X = \Pi_q^d$, $\lambda = \delta_{\mathbf{x}}$ (point evaluation at $\mathbf{x}$), $\lambda_j = \delta_{\mathbf{x}_j}$, $j = 1, \dots, n$, there exist $\tilde{c}_1, \dots, \tilde{c}_n \in \mathbb{R}$ such that $p(\mathbf{x}) = \sum_{j=1}^{n} \tilde{c}_j \, p(\mathbf{x}_j)$ for all $p \in \Pi_q^d$, and

$$\rho_q(\mathbf{x}, \mathbf{Y}) = \max\{|p(\mathbf{x})| \, : \, p \in \Pi_q^d, \ \|p|_{\mathbf{Y}}\|_\infty \leq 1\} = \sum_{j=1}^{n} |\tilde{c}_j|.$$

Thus, by setting $\tilde{c} = (\tilde{c}_1, \ldots, \tilde{c}_n, 0, \ldots, 0) \in \mathbb{R}^N$, we arrive at

$$F(\tilde{c}) \le \Big(1 + \rho_q(\mathbf{x}, \mathbf{Y})\Big)^2 E(\Phi, \Pi_q^d)_{C(B_{\mathbf{x},\mathbf{Y}})},$$

and (4) follows by (11) and (12). ■

References

[1] M. D. Buhmann, Radial Basis Functions, Cambridge University Press, 2003.

[2] O. Davydov, On the approximation power of local least squares polynomials, *in* "Algorithms for Approximation IV," (J. Levesley, I. J. Anderson and J. C. Mason, Eds.), pp. 346–353, University of Huddersfield, UK, 2002.

[3] O. Davydov, R. Morandi and A. Sestini, Local hybrid approximation for scattered data fitting with bivarite splines, *Comput. Aided Geom. Design* **23** (2006), 703–721.

[4] O. Davydov, A. Sestini and R. Morandi, Local RBF approximation for scattered data fitting with bivarite splines, *in* "Trends and Applications in Constructive Approximation," (M. G. de Bruin, D. H. Mache, and J. Szabados, Eds.), ISNM Vol. 151, Birkhäuser, 2005, pp. 91–102.

[5] O. Davydov and L. L. Schumaker, Scattered data fitting on surfaces using projected Powell-Sabin splines, manuscript, 2007.

[6] O. Davydov and F. Zeilfelder, Scattered data fitting by direct extension of local polynomials to bivariate splines, *Adv. Comp. Math.* **21** (2004), 223–271.

[7] O. Davydov and F. Zeilfelder, TSFIT: A Software Package for Two-Stage Scattered Data Fitting, 2005, available under GPL from `http://www.maths.strath.ac.uk/∼aas04108/tsfit/`

[8] R. Franke, Smooth interpolation of scattered data by local thin plate splines, Comp. & Maths. with Appls. 8, 1982, 273–281.

[9] K. Jetter, J. Stöckler and J. D. Ward, Error estimates for scattered data interpolation on spheres, *Math. Comp.* 68, 1999, 733–747.

[10] D. Lazzaro, L. B. Montefusco, Radial basis functions for the multivariate interpolation of large scattered data sets, J. Comput. Appl. Math. 140, 2002, 521–536.

[11] W. R. Madych and S. A. Nelson, Bounds on multivariate polynomials and exponential error estimates for multiquadric interpolation, *J. Approx. Theory* **70** (1992), 94–114.

[12] H. N. Mhaskar, F. J. Narcowich and J. D. Ward, Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature, *Math. Comp.* 70 (2000), 1113–1130.

[13] M. J. D. Powell, The uniform convergence of thin plate spline interpolation in two dimensions, Numer. Math. 68 (1994), 107–128.

[14] R. Schaback, Reconstruction of Multivariate Functions from Scattered Data, manuscript, 1997. (available online from `http://www.num.math.uni-goettingen.de/schaback/`)

[15] L. L. Schumaker, Two-stage methods for fitting surfaces to scattered data, in *Quantitative Approximation*, R. Schaback and K. Scherer (eds.), Lecture Notes 556, Springer, Berlin, 1976, pp. 378–389.

[16] H. Wendland, Fast evaluation of radial basis functions: methods based on partition of unity, in Approximation Theory X, C. K. Chui, L. L. Schumaker, J. Stöckler (eds), Vanderbilt Univ. Press, Nashville, TN, 2002, 473–483.

[17] H. Wendland, Scattered Data Approximation, Cambridge University Press, 2005.

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