Error bounds for anisotropic RBF interpolation

Rick Beatson[*], Oleg Davydov[†] and Jeremy Levesley[‡]

August 8, 2009

University of Canterbury, University of Strathclyde, University of Leicester

Abstract

We present error bounds for the interpolation with anisotropically transformed radial basis functions for both function and its partial derivatives. The bounds rely on a growth function and do not contain unknown constants. For polyharmonic basic functions in $\mathbb{R}^2$ we show that the anisotropic estimates predict a significant improvement of the approximation error if both the target function and the placement of the centres are anisotropic, and this improvement is confirmed numerically.

1 Introduction

In applications it is quite common to encounter anisotropic data sets or functions, where the variation in one direction is much larger or faster than that in other directions. The data fitting technique of kriging used by geostatisticians is highly related to radial basis function (RBF) fitting. In that field fitting with anisotropic directionally dependent covariances is a popular method for ore grade estimation; see Chiles and Delfiner [4]. From the approximation theory community, Cascioli et al. [2, 3] have demonstrated numerically the effectiveness of local anisotropic RBF fitting, and also that the condition of the matrix solution process is improved.

In this paper we show that the standard error estimates are improved by the composition of the RBF with a transformation of the parameter plane which stretches the function in the direction where it is steep.

The paper is organised as follows. In Sections 2 and 3 we define the anisotropic RBF interpolation problem and use the generalised Fourier transform to determine corresponding native spaces and power function. (An alternative derivation of native spaces applicable to compact domains is presented in the Appendix at the end of the paper.) Section 4 is devoted to an error bound for the anisotropic RBF interpolant in terms of a growth function, extending the isotropic result given in [5]. In Section 5 we give an error bound for the derivatives of the anisotropic RBF interpolant. This result is new in the

[*]Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand, rick.beatson@canterbury.ac.nz

[†]Department of Mathematics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, Scotland, oleg.davydov@strath.ac.uk

[‡]Department of Mathematics, University of Leicester, Leicester LE1 7RH, England, jl1@le.ac.uk

isotropic setting as well. Section 6 presents an explicit example where the anisotropic estimates predict an improvement of the approximation error, and this improvement is confirmed numerically.

2 Anisotropic RBF interpolation

A natural procedure employed by many for fitting anisotropic data with RBFs is to transform so that the data becomes approximately isotropic, fit in the transformed setting with a radial basis function, usually with underlying basic function Φ radial/isotropic, and then transform back. This paper considers the approximation error arising from such a procedure. We give theory and numerics which show that one can expect to improve the error when the data or function being approximated is anisotropic.

Suppose $\Phi : \mathbb{R}^d \to \mathbb{R}$ is a *basic function*, i.e., a positive definite function or a conditionally positive definite function of order $s = 1, 2, \ldots$ on $\mathbb{R}^d$, see e.g. [1]. If Φ is positive definite, we set $s = 0$. Thus, for any distinct points $\mathbf{v}_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, the matrix

$$[\Phi(\mathbf{v}_i - \mathbf{v}_j)]_{i,j=1}^n,$$

where $\Phi(\mathbf{v}) := \phi(\|\mathbf{v}\|_2)$, is positive definite on the subspace of $\mathbb{R}^n$ of vectors $a \in \mathbb{R}^n$ satisfying

$$\sum_{j=1}^n a_j\, p(\mathbf{u}_j) = 0, \qquad \text{all } p \in \Pi_{s-1}^d,$$

where Π_ℓ^d denotes the space of all d-variate polynomials of total degree at most ℓ, and $\Pi_{-1}^d := \emptyset$. Φ is usually radial but can be non radial.

Given a set of distinct points $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^d$, and real data values f_j, $j = 1, \ldots, N$ associated with the corresponding $\mathbf{x}_j$, the classical RBF interpolant has the form

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

where $m = 0$ if $\ell = -1$, and $m = \binom{d+\ell}{d}$ if $\ell \geq 0$, with $\{p_1, \ldots, p_m\}$ in the latter case being a basis for Π_ℓ^d. The coefficients $\{a_j\}$ and $\{b_j\}$ in (1) are determined from the conditions

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

and

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

This is uniquely solvable (see e.g. [1]) under the assumptions that $N \geq m$ and $\mathbf{X}$ is a unisolvent set for Π_ℓ^d, i.e. for any $p \in \Pi_\ell^d$, $p|_{\mathbf{X}} = 0$ implies $p \equiv 0$. Note that if Φ is radial one can expect that this method will work best when the data or function being approximated is isotropic.

Consider the case when the transformation procedure of the first paragraph is the linear transformation $\mathbf{u} = A\mathbf{x}$, with A invertible. Then in the transformed setting we approximate $g(\mathbf{u}) = f(A^{-1}\mathbf{u})$, with an RBF based on the basic function Φ,

$$r_{\phi,\ell}(\mathbf{u}) = \sum_{j=1}^{N} a_j \Phi(\mathbf{u} - \mathbf{u}_j) + \sum_{j=1}^{m} b_j p_j(\mathbf{u}). \tag{4}$$

This can also be viewed as approximating in the original untransformed setting with an RBF based upon a non radial basic function Φ_A. We replace in (1) the function Φ by $\Phi_A(\cdot) := \Phi(A\cdot)$ and the polynomial p_j by the polynomial $p_j(A\cdot)$, where $A \in \mathbb{R}^{d\times d}$ is a non-singular matrix. Clearly, the interpolation problem

$$r^A_{\phi,\ell}(\mathbf{x}_j) = f_j, \quad j = 1,\dots,N, \qquad \sum_{j=1}^{N} a_j\, p(A\mathbf{x}_j) = 0, \quad \text{for all } p \in \Pi^d_\ell, \tag{5}$$

where

$$r^A_{\phi,\ell}(\cdot) = \sum_{j=1}^{N} a_j \Phi_A(\cdot - \mathbf{x}_j) + \sum_{j=1}^{m} b_j p_j(A\cdot), \tag{6}$$

is equivalent to

$$r_{\phi,\ell}(\mathbf{u}_j) = f_j, \quad j = 1,\dots,N, \qquad \sum_{j=1}^{N} a_j\, p(\mathbf{u}_j) = 0, \quad \text{for all } p \in \Pi^d_\ell,$$

where $r_{\phi,\ell}$ is an RBF of the form (4). Since $\{\mathbf{u}_j\}$ is a set of distinct points unisolvent for Π^d_ℓ, whenever $\mathbf{X} = \{\mathbf{x}_j\}$ is a set of distinct points unisolvent for Π^d_ℓ, we conclude that the *anisotropic RBF interpolation problem* (5) has a unique solution as soon as $\ell \geq s-1$ and $\mathbf{X}$ is a set of distinct points unisolvent for Π^d_ℓ.

3 Native space and power function

Given any distinct $\mathbf{x}_j \in \mathbb{R}^d$, $j = 1,\dots,n$, and coefficients a_j satisfying $\sum_{j=1}^{n} a_j\, p(\mathbf{x}_j) = 0$ for all $p \in \Pi^d_{s-1}$, we have $\Phi_A(\mathbf{x}_i - \mathbf{x}_j) = \Phi(\mathbf{u}_i - \mathbf{u}_j)$, where $\mathbf{u}_j := A\mathbf{x}_j$, $i = 1,\dots,n$, are all different and satisfy $\sum_{j=1}^{n} a_j\, p(\mathbf{u}_j) = 0$ for all $p \in \Pi^d_{s-1}$. This shows that Φ_A is a multivariate *conditionally positive definite function* of order s in the sense of [7, Chapter 8]. Moreover, although Φ_A is usually not radial it is an even function. Therefore, its *native space* $\mathcal{F}_{\phi,A}$ can be described with the help of the *generalised Fourier transform* $\hat{f}$ as

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

where the (semi-) norm is

$$\|f\|_{\phi,A} := (2\pi)^{-d/4} \left\| \hat{f}/\sqrt{\widehat{\Phi_A}} \right\|_{L_2(\mathbb{R}^d)}, \qquad f \in L_2(\mathbb{R}^d),$$

see [7, Theorem 10.21]. Recall that $\hat{f}$ 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 [7, Section 8.2].

Since

$$\widehat{\Phi(A\cdot)} = |\det A|^{-1}\,\hat{\Phi}(A^{-T}\cdot), \qquad \widehat{f(A^{-1}\cdot)} = |\det A|\,\hat{f}(A^{T}\cdot),$$

we have

$$c\|f\|_{\phi,A}^2 = |\det A| \int_{\mathbb{R}^d} \frac{|\hat{f}(\mathbf{t})|^2}{\hat{\Phi}(A^{-T}\mathbf{t})}\,\mathbf{dt} = |\det A|^2 \int_{\mathbb{R}^d} \frac{|\hat{f}(A^T\boldsymbol{\omega})|^2}{\hat{\Phi}(\boldsymbol{\omega})}\,\mathbf{d}\boldsymbol{\omega} = \left\|\widehat{f(A^{-1}\cdot)}/\sqrt{\hat{\Phi}}\right\|_{L_2(\mathbb{R}^d)},$$

where $c := (2\pi)^{d/2}$. This shows that

$$\|f\|_{\phi,A} = \|f(A^{-1}\cdot)\|_{\phi}, \tag{7}$$

where $\|\cdot\|_\phi$ denotes the isotropic (semi-) norm

$$\|f\|_\phi := (2\pi)^{-d/4}\left\|\hat{f}/\sqrt{\hat{\Phi}}\right\|_{L_2(\mathbb{R}^d)}, \qquad f \in L_2(\mathbb{R}^d),$$

generated by ϕ in the native space

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

Remark 1. *We can define a native space on compact domains also, and (7) holds in that case too. Details of this can be found in the Appendix.*

Let $\mathbf{x} \in \mathbb{R}^d \setminus \mathbf{X}$. Under the assumption that $\mathbf{X}$ is a unisolvent set for Π_ℓ^d, the following estimate holds (see [7, Theorems 11.4 and 11.5])

$$|f(\mathbf{x}) - r_{\phi,\ell}^A(\mathbf{x})| \le P(\mathbf{x})\,\|f\|_{\phi,A}, \tag{8}$$

where $P(\mathbf{x})$ is the *power function* that satisfies

$$P(\mathbf{x}) = \min\{\sqrt{F(\mathbf{c})} \,:\, \mathbf{c} \in \mathbb{R}^N,\ p(\mathbf{x}) = \sum_{j=1}^N c_j\, p(\mathbf{x}_j) \text{ for all } p \in \Pi_\ell^d\}, \tag{9}$$

with

$$F(\mathbf{c}) := \Phi_A(\mathbf{0}) - 2\sum_{j=1}^N c_j \Phi_A(\mathbf{x} - \mathbf{x}_j) + \sum_{j,k=1}^N c_j c_k \Phi_A(\mathbf{x}_j - \mathbf{x}_k), \quad \mathbf{c} = (c_1, \ldots, c_N) \in \mathbb{R}^N.$$

4 Error bound in terms of a growth function

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 $\mathbf{Y}$,

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

It is easy to see that $\rho_q(\mathbf{x}, \mathbf{Y})$ is finite for all $\mathbf{x} \in \mathbb{R}^d$ if $\mathbf{Y}$ is a unisolvent set for Π_q^d. For suppose $m = \dim(\Pi_q^d)$ and without loss of generality that $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\} \subset \mathbf{Y}$

is unisolvent for Π_q^d. Then writing p in terms of the corresponding Lagrange basis as $p = \sum_{j=1}^m p(\mathbf{y}_j)p_j$ we see that $\rho_q(\mathbf{x}, \mathbf{Y}) \le \sum_{j=1}^m |p_j(\mathbf{x})| < \infty$. 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 at the centres in $\mathbf{Y}$.

Furthermore, we denote by $E(f, \mathcal{S})_{C(G)}$, where $G \subset \mathbb{R}^d$, the *error of the best uniform approximation* to f from a linear space $\mathcal{S}$ of functions on a compact set 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 $\|f\|_{C(G)} := \max_{x \in G} |f(x)|$.

Theorem 2. *Assume that $f_j = f(\mathbf{x}_j)$, $j = 1, \dots, N$, for a function $f \in \mathcal{F}_{\phi,A}$. Then for any non-empty $\mathbf{Y} \subseteq \mathbf{X}$, and any $q \ge \max\{\ell, 0\}$, we have*

$$|f(\mathbf{x}) - r_{\phi,\ell}^A(\mathbf{x})| \le \Big(1 + \rho_q(\mathbf{x}, \mathbf{Y})\Big) \sqrt{E(\Phi, \Pi_q^d)_{C(B_{\mathbf{x},\mathbf{Y}}^A)}}\, \|f\|_{\phi,A}, \quad \mathbf{x} \in \mathbb{R}^d, \tag{10}$$

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

We will need the following two lemmas, see [5].

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

$$p(\mathbf{x}) = \sum_{j=1}^n c_j p(\mathbf{x}_j) \qquad \textit{for all } p \in \Pi_q^d. \tag{11}$$

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

$$p(\mathbf{0}) - 2\sum_{j=1}^n c_j\, p(\mathbf{x} - \mathbf{x}_j) + \sum_{j,k=1}^n c_j c_k\, p(\mathbf{x}_j - \mathbf{x}_k) = 0. \tag{12}$$

Lemma 4. *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*

$$\max_{\{x \in X \,:\, |\lambda_i(x)| \le 1,\, i=1,\dots,k\}} |\lambda(x)| = \min_{\{c \in \mathbb{R}^k \,:\, \lambda = \sum_{i=1}^k c_i \lambda_i\}} \sum_{i=1}^k |c_i|. \tag{13}$$

Proof. Although a proof for this lemma can be found in [5], we provide a new short proof showing that it is a consequence of the well known duality theorem for the best approximation in normed linear spaces. Without loss of generality we assume that $X = \mathbb{R}^n$, $n := \dim X$. Denote by A the matrix in $\mathbb{R}^{k \times n}$ such that $(\lambda_1(x), \dots, \lambda_k(x))^T = Ax$, $x \in \mathbb{R}^n$. Since $\lambda_1, \dots, \lambda_k$ span X^*, there is a vector $b \in \mathbb{R}^k$ such that $\lambda = \sum_{i=1}^k b_i \lambda_i$, that is $\lambda(x) = b^T A x$, $x \in \mathbb{R}^n$. Hence a vector $c \in \mathbb{R}^k$ satisfies $\lambda = \sum_{i=1}^k c_i \lambda_i$ if and only if $c^T A = b^T A$. It follows that the right hand side of (13) can be formulated as a best approximation problem

$$\min_{u \in \ker A^T} \|b - u\|_1,$$

where $\ker A^T = \{u \in \mathbb{R}^k : A^T u = 0\}$. By a duality theorem (see e.g. [6, Section 1.3]), this is equal to

$$\max_{\{v\in\mathbb{R}^k:\, \|v\|_\infty \le 1,\; v\in(\ker A^T)^\perp\}} v^T b,$$

where $(\ker A^T)^\perp = \{v \in \mathbb{R}^k : v^T u = 0 \text{ for all } u \in \ker A^T\}$. By the Fundamental Theorem of Linear Algebra, $(\ker A^T)^\perp = \operatorname{Im} A = \{v \in \mathbb{R}^k : v = Ax \text{ for some } x \in \mathbb{R}^n\}$. Hence, the expression in the last display is equal to

$$\max_{\{x\in\mathbb{R}^n:\, \|Ax\|_\infty \le 1\}} b^T A x,$$

which is easily recognized as the left hand side of (13). ■

Proof of Theorem 2. Let $\mathbf{x} \in \mathbb{R}^d \backslash \mathbf{X}$. Referring to (8), we aim at producing a bound for the power function $P(\mathbf{x})$ given by (9). Choose a $q \ge \max\{\ell, 0\}$ and any subset $\mathbf{Y} \subseteq \mathbf{X}$ such that $\rho_q(\mathbf{x}, \mathbf{Y}) < \infty$. Assume without loss of generality that $\mathbf{Y} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where $n \le N$. Clearly, the condition $\rho_q(\mathbf{x}, \mathbf{Y}) < \infty$ holds if and only if $\mathbf{Y}$ is a unisolvent 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 $\delta_{\mathbf{x}_j} : \Pi_q^d \to \mathbb{R}$, $j = 1, \ldots, n$, 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 $\delta_{\mathbf{x}_j}$, $j = 1, \ldots, n$. We conclude that there exist vectors $c \in \mathbb{R}^N$ satisfying

$$p(\mathbf{x}) = \sum_{j=1}^{n} c_j\, p(\mathbf{x}_j) \qquad \text{for all } p \in \Pi_q^d \tag{14}$$

and

$$c_j = 0, \qquad \text{for all } j = n+1, \ldots, N. \tag{15}$$

Since $p(A\cdot)$ is a polynomial of the same degree as p, it follows that

$$p(A\mathbf{x}) = \sum_{j=1}^{n} c_j\, p(A\mathbf{x}_j) \qquad \text{for all } p \in \Pi_q^d. \tag{16}$$

Let us fix for a moment a vector $c \in \mathbb{R}^N$ satisfying (16) and (15). Lemma 3 implies that for any $p \in \Pi_q^d$,

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

Since $\Pi_\ell^d \subset \Pi_q^d$, we obtain by taking into account (15),

$$F(c) \;=\; [\Phi(\mathbf{0}) - p(\mathbf{0})] - 2\sum_{j=1}^{n} c_j\, [\Phi(A\mathbf{x} - A\mathbf{x}_j) - p(A\mathbf{x} - A\mathbf{x}_j)]$$

$$+ \sum_{j,k=1}^{n} c_j c_k \left[\Phi(A\mathbf{x}_j - A\mathbf{x}_k) - p(A\mathbf{x}_j - A\mathbf{x}_k)\right]$$

$$\leq \Big(1 + \sum_{j=1}^{n} |c_j|\Big)^2 \|\Phi - p\|_{C(B^A_{\mathbf{x},\mathbf{Y}})}.$$

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

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

for any $c \in \mathbb{R}^N$ such that (14) and (15) hold.

By Lemma 4, where we take $X = \Pi_q^d$, $\lambda = \delta_{\mathbf{x}}$ (point evaluation at $\mathbf{x}$), $\lambda_j = \delta_{\mathbf{x}_j}$, $j = 1, \ldots, n$, there exist $\tilde{c}_1, \ldots, \tilde{c}_n \in \mathbb{R}$ 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

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

and (10) follows by (8) and (9). ■

The proof given for Theorem 2 involves working with the non-radial basic function Φ_A in the untransformed setting. A proof could also be given working with the isotropic basic function Φ in the transformed setting, where [5, Theorem 1] can be used. In this regard note the equality $\rho_q(\mathbf{x}, \mathbf{Y}) = \rho_q(A\mathbf{x}, A\mathbf{Y})$.

5 Error estimates for derivatives

In this section our aim is to obtain an estimate for the error in approximation to derivatives by radial basis functions involving an appropriate polynomial growth function. The analog of Theorem 2 obtained applies to the anisotropic approximations $r^A_{\phi,\ell}$ defined in (6).

With the help of QR factorisation it can be shown that any real invertible matrix can be written as the product of an orthogonal matrix, a diagonal scaling, and an upper triangular matrix with unit diagonal (i.e. a shear matrix). This is sometimes called Iwasawa decomposition.

For the sake of simplicity we assume throughout this section that the $d \times d$ invertible transformation matrix A has a special form. Namely $A = \Gamma Q^T$ where the scaling matrix $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_d)$ is diagonal with positive diagonal entries, and Q^T is a rotation (i.e. an orthogonal matrix with determinant 1). Let $\mathbf{q}_1, \ldots, \mathbf{q}_d$ denote the columns of Q. Then multiplication by Q^T maps the ray in direction $\mathbf{q}_i$ into the i-th coordinate axis, which we call $\mathbf{e}_i$. Define the directional derivatives $D_{\mathbf{q}_i} f = \nabla f \cdot \mathbf{q}_i$, and the iterated directional derivatives

$$D_Q^\alpha f = (D_{\mathbf{q}_1})^{\alpha_1} (D_{\mathbf{q}_2})^{\alpha_2} \dots (D_{\mathbf{q}_d})^{\alpha_d} f.$$

In applications we aim to select $A^T A = Q\Gamma^2 Q^T$, and thus the principal axes of the positive definite quadratic form $\mathbf{x}^T A^T A \mathbf{x}$, to 'undo' the anisotropic behaviour of the function f being approximated; see Figure 1.

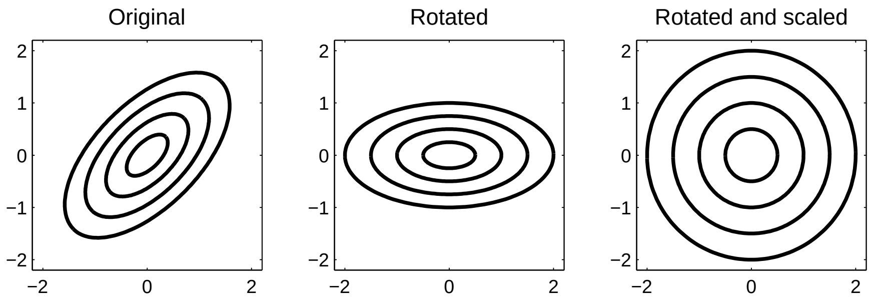

Figure 1: Undoing the anisotropy

We note that for any differentiable function g and a diagonal scaling Γ^{-1}

$$\left[\frac{\partial}{\partial \mathbf{e}_i} g\left(\Gamma^{-1}\cdot\right)\right](\mathbf{u}) = \gamma_i^{-1} \left[\frac{\partial g}{\partial \mathbf{e}_i}\right] \left(\Gamma^{-1}\mathbf{u}\right),$$

and that for Q a rotation

$$\left[\frac{\partial}{\partial \mathbf{e}_i} f(Q\cdot)\right](\mathbf{u}) = \left[D_{\mathbf{q}_i} f(\cdot)\right](Q\mathbf{u}).$$

Hence

$$\begin{aligned}\left[\frac{\partial}{\partial \mathbf{e}_i} f\left(Q\Gamma^{-1}\cdot\right)\right](\mathbf{u}) &= \gamma_i^{-1}\left[D_{\mathbf{q}_i} f\right]\left(Q\Gamma^{-1}\mathbf{u}\right)\\ &= \gamma_i^{-1}\left[D_{\mathbf{q}_i} f\right](\mathbf{x}), \quad \text{where } \mathbf{u} = A^{-1}\mathbf{x} \quad \text{and } A = \Gamma Q^T.\end{aligned}$$

More generally

$$\left[D^\alpha g\right](\mathbf{u}) = \gamma^{-\alpha}\left[D_Q^\alpha f\right](\mathbf{x}), \quad \text{where } x = A^{-1}\mathbf{u},\ g(\mathbf{u}) := f(A^{-1}\mathbf{u}). \tag{17}$$

We will need a more general definition of the growth function. Given a nonempty subset $\mathbf{Y} \subset \mathbb{R}^d$ we set

$$\rho_{q,\alpha}(\mathbf{x}, \mathbf{Y}) := \max\{|\left(D^\alpha p\right)(\mathbf{x})| \;:\; p \in \Pi_q^d,\ \|p|_{\mathbf{Y}}\|_\infty \le 1\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

This definition only has content when $|\alpha| \le q$.

Since $p(A\cdot) \in \Pi_q^d$ as soon as $p \in \Pi_q^d$, it is easy to see that

$$\rho_{q,\alpha}(A\mathbf{x}, A\mathbf{Y}) = \gamma^{-\alpha}\rho_{q,\alpha,Q}(\mathbf{x}, \mathbf{Y}), \tag{18}$$

where the additional index Q in the right hand side indicates that the coordinate derivatives D^α are replaced there by the directional derivatives D^α_Q.

Given a closed ball about the origin B, and a number $\mu > 0$, define the space $V^\mu_\alpha(B)$ as the set of all functions in $g \in C^\alpha(B)$ with a defined 2α-th derivative at zero, and norm

$$\|g\|_{V^\mu_\alpha(B)} := \max\{\mu^{2|\alpha|}|\left(D^{2\alpha}g\right)(\mathbf{0})|, \mu^{|\alpha|}\|D^\alpha g\|_{C(B)}, \|g\|_{C(B)}\}.$$

Then we have the following extension of Theorem 2, which gives derivative estimates when the invertible matrix A has special form.

Theorem 5. *Let the transformation matrix $A = \Gamma Q^T$, where the scaling matrix $\Gamma =$ diag(γ) is diagonal with positive diagonal entries, and Q^T is a rotation. Assume that $\Phi \in C^{2k}(\mathbb{R}^d)$, $f_j = f(\mathbf{x}_j)$, $j = 1, \ldots, N$, with $f \in \mathcal{F}_{\phi,A}$. Then, for $|\alpha| \le k$, $\mathbf{x} \in \mathbb{R}^d$, any $\mu > 0$, non-empty $\mathbf{Y} \subseteq \mathbf{X}$, and $q \ge \max\{|\alpha|, \ell\}$,*

$$|\left(D^\alpha_Q f\right)(\mathbf{x}) - \left(D^\alpha_Q r^A_{\phi,\ell}\right)(\mathbf{x})| \le \gamma^\alpha \left(\mu^{-|\alpha|} + \rho_{q,\alpha}(A\mathbf{x}, A\mathbf{Y})\right)\sqrt{E(\Phi, \Pi^d_q)_{V^\mu_\alpha(B^A_{\mathbf{x},\mathbf{Y}})}}\, \|f\|_{\phi,A}, \tag{19}$$

where $B^A_{\mathbf{x},\mathbf{Y}}$ denotes the ball in $\mathbb{R}^d$ with center $\mathbf{0}$ and radius $\operatorname{diam}(\{A\mathbf{x}\} \cup A\mathbf{Y})$*, and $E(\Phi, \Pi^d_q)_{V^\mu_\alpha(B^A_{\mathbf{x},\mathbf{Y}})}$ is the error in best approximation of Φ from Π^d_q in the space $V^\mu_\alpha(B^A_{\mathbf{x},\mathbf{Y}})$ defined above.*

A variant of Lemma 3 appropriate for the estimation of derivative error as in Theorem 5 is

Lemma 6. *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*

$$\left(D^\alpha p\right)(\mathbf{x}) = \sum_{j=1}^n c_j p(\mathbf{x}_j) \qquad \text{for all } p \in \Pi^d_q. \tag{20}$$

Then, for all $p \in \Pi^d_q$, we have

$$(-1)^{|\alpha|}\left(D^{2\alpha}p\right)(\mathbf{0}) - 2\sum_{j=1}^n c_j \left(D^\alpha p\right)(\mathbf{x} - \mathbf{x}_j) + \sum_{j,k=1}^n c_j c_k\, p(\mathbf{x}_j - \mathbf{x}_k) = 0. \tag{21}$$

Proof: If $p \in \Pi^d_q$ and $\mathbf{y} \in \mathbb{R}^d$ then both $p(\cdot - \mathbf{y})$ and $p(\mathbf{y} - \cdot)$ belong to Π^d_q. It follows from (20) that

$$\left(D^\alpha p\right)(\mathbf{x} - \mathbf{y}) = \sum_{j=1}^n c_j p(\mathbf{x}_j - \mathbf{y}),$$

and

$$(-1)^{|\alpha|}\left(D^\alpha p\right)(\mathbf{y} - \mathbf{x}) = \sum_{j=1}^n c_j p(\mathbf{y} - \mathbf{x}_j).$$

Taking $\mathbf{y} = \mathbf{x}_k$ in the first identity

$$\left(D^\alpha p\right)(\mathbf{x} - \mathbf{x}_k) = \sum_{j=1}^n c_j p(\mathbf{x}_j - \mathbf{x}_k),$$

and taking $\mathbf{y} = \mathbf{x}$ in the second identity

$$(-1)^{|\alpha|} \left(D^\alpha p\right)(\mathbf{0}) = \sum_{j=1}^{n} c_j p(\mathbf{x} - \mathbf{x}_j).$$

Therefore

$$\begin{aligned}
&(-1)^{|\alpha|} \left(D^{2\alpha} p\right)(\mathbf{0}) - 2\sum_{j=1}^{n} c_j \ \left(D^\alpha p\right)(\mathbf{x} - \mathbf{x}_j) + \sum_{k}^{n} c_k \sum_{j=1}^{n} c_j\, p(\mathbf{x}_j - \mathbf{x}_k) \\
&= (-1)^{|\alpha|} \left(D^{2\alpha} p\right)(\mathbf{0}) - 2(-1)^{|\alpha|} \left(D^{2\alpha} p\right)(\mathbf{0}) + \sum_{k=1}^{n} c_k \left(D^\alpha p\right)(\mathbf{x} - \mathbf{x}_k) \\
&= 0.
\end{aligned}$$

as required. ■

Proof of Theorem 5. Let $\mathbf{x} \in \mathbb{R}^d \setminus \mathbf{X}$. We will apply (24) in the transformed setting for approximation with the radial function Φ. Choose a $q \geq \max\{|\alpha|, \ell\}$ and any subset $\mathbf{Y} \subseteq \mathbf{X}$ unisolvent for Π_q^d. Then $A\mathbf{Y}$ is also unisolvent for Π_q^d. Write $\mathbf{u}$ for Ax and $\mathbf{u}_j$ for $A\mathbf{x}_j$. Assume without loss of generality that $A\mathbf{Y} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, where $n \leq N$.

Then the mapping $\delta_{A\mathbf{Y}} : \Pi_q^d \to \mathbb{R}^n$ defined by $\delta_{A\mathbf{Y}}(p) = p|_{A\mathbf{Y}}$ is injective, and its image has dimension $\binom{d+q}{d} = \dim \Pi_q^d$. This implies that among the point evaluation functionals $\delta_{\mathbf{u}_j} : \Pi_q^d \to \mathbb{R}$, $j = 1, \ldots, n$, that form the components of $\delta_{A\mathbf{Y}}$, there are $\binom{d+q}{d}$ that are linearly independent over Π_q^d. Therefore, $\{\delta_{\mathbf{u}_j}\}_{j=1}^n$ span the dual space $(\Pi_q^d)^*$. Now, the linear functional $\delta_{\mathbf{u}}^{(\alpha)}$ defined by $\delta_{\mathbf{u}}^{(\alpha)}(p) = \left(D^\alpha p\right)(\mathbf{u})$ is also in $(\Pi_q^d)^*$, and hence it can be written as a linear combination of $\delta_{\tilde{x}_j}$, $j = 1, \ldots, n$. We conclude that there exist vectors $\mathbf{c} \in \mathbb{R}^N$ satisfying

$$\left(D^\alpha p\right)(\mathbf{u}) = \sum_{j=1}^{n} c_j\, p(\mathbf{u}_j) \qquad \text{for all } p \in \Pi_q^d \tag{22}$$

and

$$c_j = 0, \qquad \text{for all } j = n+1, \ldots, N. \tag{23}$$

A power function error estimate result appropriate for derivative estimation [7, Theorem 11.4, Theorem 11.5] is as follows. Let $r_{\phi,\ell}$ be the RBF interpolant for form (4) interpolating to g at nodes $\{\mathbf{u}_j\}$. Then, if $\Phi \in C^{2k}(\mathbb{R}^d)$, $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a unisolvent set for Π_ℓ^d, and $|\alpha| \leq k$,

$$\left|\left(D^\alpha g\right)(\mathbf{u}) - \left(D^\alpha r_{\phi,\ell}\right)(\mathbf{u})\right| \leq P_\alpha(\mathbf{x})\, \|g\|_\phi. \tag{24}$$

Here, $P_\alpha(\mathbf{u})$ is the *power function* that satisfies

$$P_\alpha(\mathbf{u}) = \min\left\{ \sqrt{F_\alpha(\mathbf{c})} \ : \ \mathbf{c} \in \mathbb{R}^N,\ \left(D^\alpha p\right)(\mathbf{u}) = \sum_{j=1}^{N} c_j\, p(\mathbf{u}_j) \text{ for all } p \in \Pi_\ell^d \right\}, \tag{25}$$

with

$$F_\alpha(\mathbf{c}) := (-1)^{|\alpha|} \left(D^{2\alpha}\Phi\right)(\mathbf{0}) - 2\sum_{j=1}^{N} c_j \left(D^\alpha \Phi\right)(\mathbf{u}-\mathbf{u}_j) + \sum_{j,k=1}^{N} c_j c_k \Phi(\mathbf{u}_j - \mathbf{u}_k). \quad (26)$$

Note here that

$$F_\alpha(\mathbf{c}) = [1 \;\; -\mathbf{c}^T] \begin{bmatrix} (-1)^{|\alpha|}\left(D^{2\alpha}\Phi\right)(\mathbf{0}) & \mathbf{b}^T \\ \mathbf{b} & B \end{bmatrix} \begin{bmatrix} 1 \\ -\mathbf{c} \end{bmatrix},$$

where $\mathbf{b} = [(D^\alpha\Phi)(\mathbf{u}-\mathbf{u}_1), \ldots, (D^\alpha\Phi_A)(\mathbf{u}-\mathbf{u}_N)]^T \in \mathbb{R}^N$, and B is $N \times N$ with jk entry $b_{jk} = \Phi(\mathbf{u}_j - \mathbf{u}_k)$.

Let us fix for a moment a vector $\mathbf{c} \in \mathbb{R}^N$ satisfying (22) and (23). Lemma 6 implies that for any $p \in \Pi_q^d$,

$$(-1)^{|\alpha|}\left(D^{2\alpha}p\right)(\mathbf{0}) - 2\sum_{j=1}^{n} c_j \; (D^\alpha p)(\mathbf{u}-\mathbf{u}_j) + \sum_{j,k=1}^{n} c_j c_k \, p(\mathbf{u}_j - \mathbf{u}_k) = 0.$$

Since $\Pi_\ell^d \subset \Pi_q^d$, we obtain by taking into account (23),

$$F_\alpha(\mathbf{c})\mu^{2|\alpha|} = (-1)^{|\alpha|}\mu^{2|\alpha|}\left(D^{2\alpha}t\right)(\mathbf{0}) - 2\sum_{j=1}^{N} c_j \mu^{2|\alpha|}(D^\alpha t)(\mathbf{u}-\mathbf{u}_j) + \sum_{j,k=1}^{N} \mu^{2|\alpha|} c_j c_k \, t(\mathbf{u}_j - \mathbf{u}_k),$$

where $t = \Phi - p$. Since $p \in \Pi_q^d$ is arbitrary, it follows that

$$F_\alpha(\mathbf{c}) \le \mu^{-2|\alpha|}\Big(1 + \sum_{j=1}^{n}|c_j|\mu^{|\alpha|}\Big)^2 E(\Phi, \Pi_q^d)_{V_\alpha^\mu(B^A_{\mathbf{x},\mathbf{Y}})} \quad (27)$$

for any $\mathbf{c} \in \mathbb{R}^N$ such that (22) and (23) hold.

By Lemma 4, where we take $X = \Pi_q^d$, $\lambda = \delta_{\mathbf{u}}^{(\alpha)}$ (derivative evaluation at $\mathbf{u}$), $\lambda_j = \delta_{\mathbf{u}_j}$, $j = 1, \ldots, n$, there exist $\tilde{c}_1, \ldots, \tilde{c}_n \in \mathbb{R}$ such that $(D^\alpha p)(\mathbf{u}) = \sum_{j=1}^n \tilde{c}_j \, p(\mathbf{u}_j)$ for all $p \in \Pi_q^d$, and

$$\rho_{q,\alpha}(A\mathbf{x}, A\mathbf{Y}) = \max\{|\,(D^\alpha p)(\mathbf{u})| \;:\; p \in \Pi_q^d, \; \|p|_{A\mathbf{Y}}\|_\infty \le 1\} = \sum_{j=1}^{n} |\tilde{c}_j|.$$

Thus, by setting $\tilde{c} = (\tilde{c}_1, \ldots, \tilde{c}_n, 0, \ldots, 0) \in \mathbb{R}^N$ in (27), we arrive at

$$F_\alpha(\tilde{c}) \le \mu^{-2|\alpha|}\Big(1 + \mu^{|\alpha|}\rho_{q,\alpha}(A\mathbf{x}, A\mathbf{Y})\Big)^2 E(\Phi, \Pi_q^d)_{V_\alpha^\mu(B^A_{\mathbf{x},\mathbf{Y}})}.$$

Then, from (24) and (25),

$$|(D^\alpha g)(\mathbf{u}) - (D^\alpha r_{\phi,\ell})(\mathbf{u})| \le \mu^{-|\alpha|}\Big(1 + \mu^{|\alpha|}\rho_{q,\alpha}(A\mathbf{x}, A\mathbf{Y})\Big)\sqrt{E(\Phi, \Pi_q^d)_{V_\alpha^\mu(B^A_{\mathbf{x},\mathbf{Y}})}}\;\|g\|_\phi.$$

Applying formulas (17) and (7)

$$\begin{aligned} \left|\left(D_Q^\alpha f\right)(\mathbf{x}) - \left(D_Q^\alpha r_{\phi,\ell}^A\right)(\mathbf{x})\right| &= \gamma^\alpha \, |(D^\alpha g)(\mathbf{u}) - (D^\alpha r_{\phi,\ell})(\mathbf{u})| \\ &\le \gamma^\alpha\Big(\mu^{-|\alpha|} + \rho_{q,\alpha}(A\mathbf{x}, A\mathbf{Y})\Big)\sqrt{E(\Phi, \Pi_q^d)_{V_\alpha^\mu(B^A_{\mathbf{x},\mathbf{Y}})}}\;\|f\|_{\phi,A}. \end{aligned}$$

■

Remark 7. *It is possible for* $\rho_{q,\alpha}(\mathbf{x}, \mathbf{Y}) < \infty$ *to hold without* $\mathbf{Y}$ *being a unisolvent set for* Π_q^d. *For example in* $\mathbb{R}^2$ *and with* $q = 1$ *take* Y *as lots of points along the* x *axis and no other points. Then the* x*–partial* $\partial p/\partial x$ *is certainly controlled, but the set is not unisolvent for* Π_1^2, *and the* y*–partial* $\partial p/\partial y$ *is completely uncontrolled, as is the function value at points off the* x *axis. If instead of just trying to control* $D^\alpha p$ *we want all derivatives* $D^\beta p$ *with* $\mathbf{0} \le \beta \le \alpha$ *controlled, then we want function values controlled, and immediately we require* $\mathbf{Y}$ *unisolvent for* Π_q^d.

Now we need $F_\alpha(\mathbf{c})$ to be small in order to achieve a small error bound for our RBF approximation. From the form of (27) this requires at least the ability to drive $E(\Phi, \Pi_q^d)_{V_\alpha^\mu(B^A_{\mathbf{x},\mathbf{Y}})}$ toward zero. Therefore, in the case where $D^{2\alpha}\Phi(\mathbf{0}) \ne 0$, we will need to require $q \ge 2|\alpha|$. In the case that $D^{2\alpha}\Phi(\mathbf{0}) = 0$ we need to be able to approximate $D^\alpha\Phi$ well by polynomials, so may get useful estimates when $|\alpha| \le q < 2|\alpha|$. These remarks underly the requirement that $|\alpha| \le q$ in the statement of the theorem.

6 Examples

Consider the case of approximation with the polyharmonic basic functions in $\mathbb{R}^2$, given by $\Phi(\mathbf{x}) = \|\mathbf{x}\|^{2k-2}\log\|x\|$, $k \ge 2$, supplemented with polynomials of degree $k-1$, in $\mathbb{R}^2$. The corresponding native space is the Beppo-Levi space

$$\mathrm{BL}_k(\mathbb{R}^d) = \{f \in C(\mathbb{R}^d) \,:\, D^\alpha f \in L_2(\mathbb{R}^d),\ \forall\, |\alpha| = k\},$$

with

$$|f|^2_{\mathrm{BL}_k(\mathbb{R}^d)} = \sum_{|\alpha|=k} \frac{k!}{\alpha!}\|D^\alpha f\|^2_{L_2(\mathbb{R}^d)}.$$

In the case $k = 2$, $\Phi(x)$ is the classical thin plate spline.

In applying the bounds of Theorems 2 and 5 we will make the additional assumption that $q \ge 2k-2$ rather than just $q \ge k-1$. The reason is that then a known trick estimates the error in uniform approximation of Φ in the disk of radius h by polynomials of degree $2k-2$ as $\mathcal{O}(h^{2k-2})$. The details are as follows. It is easy to see that $\Phi(x) = h^{2k-2}\Phi(x/h) + \|x\|^{2k-2}\log h$. Hence, for any $p \in \Pi^2_{2k-2}$, we have $\Phi(x) - p(x) = h^{2k-2}(\Phi(x/h) - \tilde{p}(x/h))$, where $\tilde{p}(x) := h^{2-2k}p(hx) - \|x\|^{2k-2}\log h$. Since $\tilde{p}(\cdot/h)$ also belongs to Π^2_{2k-2}, we have $E(\Phi, \Pi^2_{2k-2})_{C(hB_1)} = h^{2k-2}E(\Phi, \Pi^2_{2k-2})_{C(B_1)}$ for any $h > 0$, where B_1 is the unit disk. Therefore $E(\Phi, \Pi^2_{2k-2})_{C(hB_1)} = \mathcal{O}(h^{2k-2})$, and this rate of convergence as $h \to 0$ cannot be improved. Moreover, if $k \ge 3$ then $E(\Phi, \Pi^2_{2k-2})_{V_\alpha^h(hB_1)} = \mathcal{O}(h^{2k-2})$ for any α with $|\alpha| \le k-2$. Indeed, $D^\nu(\Phi(x) - p(x)) = h^{2k-2-|\nu|}(D^\nu\Phi(x/h) - D^\nu\tilde{p}(x/h))$ for $\nu = \alpha, 2\alpha$, which implies $\|\Phi - p\|_{V_\alpha^h(hB_1)} = h^{2k-2}\|\Phi - \tilde{p}\|_{V_\alpha^1(B_1)}$ and hence $E(\Phi, \Pi^2_{2k-2})_{V_\alpha^h(hB_1)} = h^{2k-2}E(\Phi, \Pi^2_{2k-2})_{V_\alpha^1(B_1)}$.

Now assume that the function to be approximated is anisotropic with

$$\|D^\alpha f\|_{L_2(\mathbb{R}^2)} = (1, \beta)^\alpha, \quad \text{for all } |\alpha| = k,$$

where $\beta \gg 1$. Then $\|f\|_\Phi = (1+\beta^2)^{k/2}$. Using the transformation

$$g(\mathbf{u}) = f(\mathbf{x}), \quad \mathbf{u} = A\mathbf{x}, \qquad A = \begin{bmatrix} 1 & 0 \\ 0 & \beta \end{bmatrix},$$

we find, using (17), that

$$\begin{aligned} \|D^\alpha g\|^2_{L_2(\mathbb{R}^2)} &= \int_{\mathbb{R}^2} |D^\alpha g(\mathbf{u})|^2 \, d\mathbf{u} \\ &= \int_{\mathbb{R}^2} (1,\beta)^{-2\alpha} \, |D^\alpha f(\mathbf{x})|^2 \, |\det A| \, d\mathbf{x} \;=\; \beta, \quad \text{for all } |\alpha| = k. \end{aligned}$$

Hence $\|g\|_\Phi = 2^{k/2}\beta^{1/2}$.

Consider approximating f on the unit square in the untransformed setting. Distribute N nodes approximately uniformly on the unit square so that the points are $h \approx N^{-1/2}$ apart. In this mesh the growth function $\rho_{2k-2}(\mathbf{x},\mathbf{Y})$ for the data sites $\mathbf{Y}$ taken from disks centered at $\mathbf{x}$ and of a suitable small multiple of h in radius, is uniformly bounded. Applying Theorem 2, and the above remark concerning the approximation of Φ from Π^2_{2k-2}, we get

$$\|f - r_{\phi,k-1}\|_{L_\infty([0,1]^2)} \le CN^{-(k-1)/2}\,(1+\beta^2)^{k/2}, \tag{28}$$

for some constant C depending only on k. In view of (18), we have $\rho_{q,\alpha}(\mathbf{x},\mathbf{Y}) = h^{-|\alpha|}\rho_{q,\alpha}(\mathbf{x}/h,\mathbf{Y}/h)$. Hence, Theorem 5 with $\mu = h$, $\mathbf{Y}$ as above and $q = 2k-2$, gives

$$\|D^\alpha(f - r_{\phi,k-1})\|_{L_\infty([0,1]^2)} \le CN^{-(k-1-|\alpha|)/2}\,(1+\beta^2)^{k/2}, \quad k \ge 3, \quad |\alpha| \le k-2, \tag{29}$$

where C is a different constant.

Now apply Theorems 2 and 5 to approximation with transformation matrix A. The unit square transforms into the rectangle $[0,1]\times[0,\beta]$ with area β. Distributing N nodes approximately uniformly apart the nodes are now $\tilde{h} \approx (\beta/N)^{1/2}$ apart. Then using the bounds of Theorems 2 and 5 we get

$$\left\|f - r^A_{\phi,k-1}\right\|_{L_\infty([0,1]^2)} \le C\tilde{h}^{k-1}2^{k/2}\beta^{1/2} = CN^{-(k-1)/2}(2\beta)^{k/2}, \tag{30}$$

and, for $k \ge 3$, $|\alpha| \le k-2$,

$$\|D^\alpha(f - r^A_{\phi,k-1})\|_{L_\infty([0,1]^2)} \le C\tilde{h}^{k-1-|\alpha|}(1,\beta)^\alpha 2^{k/2}\beta^{1/2}.$$

Equivalently,

$$\left\|D^\alpha(f - r^A_{\phi,k-1})\right\|_{L_\infty([0,1]^2)} \le CN^{-(k-1-|\alpha|)/2}2^{k/2}(1,\beta)^\alpha\beta^{(k-|\alpha|)/2}. \tag{31}$$

Thus the error bounds for the anisotropic method, (30) and (31), are much smaller than those for ordinary RBF interpolation, (28) and (29), respectively, when $\beta \gg 1$. In fact, these expressions suggest an improvement for the function error by about the factor $(\beta/2)^{k/2}$, and between $(\beta/2)^{k/2}\beta^{-|\alpha|/2}$ and $(\beta/2)^{k/2}\beta^{|\alpha|/2}$ for the errors in various partial derivatives.

The comparison of upper bounds above in no way guarantees that using the anisotropic approximation method will actually improve the error. However it does indicate that this may well be the case.

We now present a numerical example approximating the function

$$f(x,y) = f_\beta(x,y) = \frac{\sqrt{2}\sqrt{\beta}}{\sqrt{\pi}}e^{-(x^2+\beta^2y^2)}. \tag{32}$$

Calculations show that for small $|\alpha|$, $\|D^\alpha f_\beta\|_{L_2(\mathbb{R}^2)}$ has order of magnitude $(1, \beta)^\alpha$, the order constant depending on the particular partial derivative.

We performed numerical computations with f_β with $\beta = 9$. The function was approximated using polyharmonic splines with $k = 3$, both using a uniform mesh on $[0, 1]^2$, and using the strategy of transforming, approximating and transforming back. The transformed domain was $[0, 1] \times [0, 9]$, the mesh there being $nx \times ny$ where $ny = 9nx$. The error in function approximation was estimated by evaluating it on a 27 times finer uniform grid in the untransformed domain. The results in Table 1 show that adapting to the anisotropy of f significantly improves the error in approximation of both f and its derivatives. Note that both approximation processes are displaying convergence of order considerably faster than the order $1/N$ predicted by our bound for approximation to f. It is indeed known that the order can be improved for sufficiently smooth functions.

Isotropic				Anisotropic			
Mesh	f error	f_x error	f_y error	Mesh	f error	f_x error	f_y error
9×9	3.787(−1)	3.918(−1)	1.757(1)	3×27	6.803(−2)	5.854(−1)	2.271(0)
18×18	4.479(−2)	8.837(−1)	4.151(0)	6×54	3.615(−3)	7.859(−2)	5.589(−1)
36×36	4.203(−3)	4.024(−1)	8.470(−1)	12×108	4.687(−4)	2.161(−2)	1.702(−1)
72×72	5.499(−4)	1.586(−1)	2.303(−1)	24×216	7.013(−5)	6.647(−3)	5.511(−2)
144×144	8.448(−5)	5.032(−2)	7.034(−2)	48×432	1.1273(−5)	2.169(−3)	1.844(−2)

Table 1: Improvement in the approximation error for the function (32) and its derivatives using anisotropically transformed polyharmonic splines

Acknowledgement. Financial support from EPSRC (grant EP/F009615) is gratefully acknowledged.

References

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

[2] G. Casciola, D. Lazzaro, L. B. Montefusco, and S. Morigi, Shape preserving surface reconstruction using locally anisotropic radial basis function interpolants, *Computers and Mathematics with Applications* **51** (2006), 1185–1198.

[3] G. Casciola, L. B. Montefusco, and S. Morigi, The regularising properties of anisotropic radial basis functions, *Applied Mathematics and Computation* **190** (2007), 1050–1062.

[4] J.-P., Chilès and P. Delfiner, Goestatistics, Modeling Spatial Uncertainty, Wiley, 1999.

[5] O. Davydov, Error bound for radial basis interpolation in terms of a growth function, *in* "Curve and Surface Fitting: Avignon 2006," (A. Cohen, J.-L. Merrien and L. L. Schumaker, Eds.), Nashboro Press, Brentwood, 2007, pp. 121–130.

[6] N. P. Korneichuk, Exact Constants in Approximation Theory, Cambridge University Press, 1991.

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

Appendix: Native spaces on compact domains

Suppose we are given a conditionally positive definite function ψ and a domain $\mathcal{D}$. Following Wendland [7, Chapter 10] we define a pre-Hilbert space

$$F_\psi(\mathcal{D}) = \left\{ \sum_{i=1}^{n} \alpha_i \psi(\cdot - \mathbf{x}_i) + p : \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subset \mathcal{D},\ n \in \mathbb{N}, \quad p \in \Pi_{s-1}^d \right\}$$

where

$$\sum_{i=1}^{n} \alpha_i q(\mathbf{x}_i) = 0, \quad q \in \Pi_{s-1}^d.$$

Let $Q = \dim(\Pi_{s-1}^d)$, $\Xi_{\mathcal{D}} = \{\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_Q\} \subset \mathcal{D}$ be unisolvent with respect to Π_{s-1}^d, and $\ell_1, \cdots, \ell_Q$, be a Lagrange basis for Π_{s-1}^d on $\Xi_{\mathcal{D}}$. An inner product for $f, g \in F_\psi(\mathcal{D})$, with

$$f(x) = \sum_{i=1}^{n} \alpha_i \psi(\mathbf{x} - \mathbf{x}_i) + p(\mathbf{x}) \quad \text{and} \quad g(x) = \sum_{i=1}^{n} \beta_i \psi(\mathbf{x} - \mathbf{x}_i) + q(\mathbf{x}),$$

is

$$(f, g)_{\psi,\mathcal{D}} = \sum_{i,j=1}^{n} \alpha_i \beta_j \psi(\mathbf{x}_i - \mathbf{x}_j) + \sum_{i=1}^{Q} f(\boldsymbol{\xi}_i) g(\boldsymbol{\xi}_i).$$

The reproducing kernel for $F_\psi(\mathcal{D})$

$$\begin{aligned} G(\mathbf{x}, \mathbf{y}) \quad &= \quad \psi(\mathbf{x} - \mathbf{y}) - \sum_{k=1}^{Q} \ell_k(\mathbf{x}) \psi(\boldsymbol{\xi}_k - \mathbf{y}) - \sum_{k=1}^{Q} \ell_k(\mathbf{y}) \psi(\mathbf{x} - \boldsymbol{\xi}_k) \\ &\quad + \sum_{i,k=1}^{Q} \ell_k(\mathbf{x}) \ell_i(\mathbf{y}) \psi(\boldsymbol{\xi}_k - \boldsymbol{\xi}_i) + \sum_{i,k=1}^{Q} \ell_k(\mathbf{x}) \ell_i(\mathbf{y}). \end{aligned}$$

It may easily be checked that, for fixed $\mathbf{x}$, $G(\mathbf{x}, \cdot) \in F_\psi(\mathcal{D})$, and, for $f \in F_\psi(\mathcal{D})$, $f(\mathbf{x}) = (f, G(\mathbf{x}, \cdot))_{\psi,\mathcal{D}}$. The native space $\mathcal{N}_\psi(\mathcal{D})$ is the completion of $F_\psi(\mathcal{D})$ with respect to the norm $\| \cdot \|_{\psi,\mathcal{D}} = (\cdot, \cdot)_{\psi,\mathcal{D}}^{1/2}$.

We can view a function

$$s(\mathbf{x}) = \sum_{y \in Y} \alpha_y \phi(A(\mathbf{x} - \mathbf{y})) + p(\mathbf{x}), \quad \mathbf{x} \in \Omega$$

as a function in $F_{\phi_A}(\Omega)$, where $\phi_A(\mathbf{x}) = \phi(A\mathbf{x})$, $x \in \mathbf{R}^d$. Alternatively, setting $\mathbf{u} = A\mathbf{x}$, and $\mathbf{u}_i = A\mathbf{x}_i$, $i = 1, \cdots, n$, we have

$$s(A^{-1}\mathbf{u}) = \sum_{i=1}^{n} \alpha_i \phi(\mathbf{u} - \mathbf{u}_i) + p(A^{-1}\mathbf{u}), \mathbf{u} \in A\Omega,$$

so that $s(A^{-1}\cdot) \in F_\phi(A\Omega)$. Letting $\Xi_{A\Omega} = A\Xi_\Omega$ we have

$$\begin{aligned}
\|s(A^{-1}\cdot)\|_{\phi,A\Omega} &= \sum_{i,j=1}^{n} \alpha_i\alpha_j\phi(\mathbf{u}_i - \mathbf{u}_j) + \sum_{i=1}^{Q} p^2(A^{-1}A\boldsymbol{\xi}_i) \\
&= \sum_{i,j=1}^{n} \alpha_i\alpha_j\phi(A\mathbf{x}_i - A\mathbf{x}_j) + \sum_{i=1}^{Q} p^2(\boldsymbol{\xi}_i) \\
&= \|s\|_{\phi_A,\Omega}.
\end{aligned}$$

Since these sets are dense in their respective native spaces we have, exactly as in the Euclidean case,

Theorem 8. *Let Ω be a compact domain and $A : \mathbf{R}^d \to \mathbf{R}^d$ be an invertible linear map. Suppose $f \in \mathcal{N}_{\phi_A}(\Omega)$, for some conditionally positive definite function ϕ. Then $f(A^{-1}\cdot) \in \mathcal{N}_\phi(A\Omega)$, and*

$$\|f(A^{-1}\cdot)\|_{\phi,A\Omega} = \|f\|_{\phi_A,\Omega}.$$

If $\phi \in \mathbf{C}^{2k}(\mathbb{R})$ is conditionally positive definite of order k with smoothness k say, then the native space $\mathcal{N}_\phi(\Omega)$ is continuously embedded in $C^k(\Omega)$ (see [7, 10.6]). Thus, for $|\alpha| \le k$, due to the linearity and continuity of the inner product $(\cdot,\cdot)_{\psi,\mathcal{D}}$ we have,

$$D^\alpha f(\mathbf{x}) = (f, D_2^\alpha G(\cdot,\mathbf{x}))_{\psi,\mathcal{D}},$$

where $D_2^\alpha G$ means differentiation of G with respect to its second variable.

Thus, to compute the derivatives of the error of interpolation we compute

$$\begin{aligned}
D^\alpha(f(x) - r^A_{\phi,\ell}(\mathbf{x})) &= \left(f - r^A_{\phi,\ell}, D_2^\alpha G(\cdot,\mathbf{x})\right)_{\psi,\mathcal{D}}, \\
&= \left(f - r^A_{\phi,\ell}, D_2^\alpha G(\cdot,\mathbf{x}) - \sum_{i=1}^{n} c_i^\alpha G(\cdot,\mathbf{x}_i)\right)_{\psi,\mathcal{D}},
\end{aligned}$$

for any choice of the c_i^α. If we now select these coefficients to reproduce the αth derivatives of polynomials of degree $k-1$,

$$D^\alpha p(\mathbf{x}) = \sum_{i=1}^{n} c_i^\alpha p(\mathbf{x}_i), \quad p \in \Pi^d_{k-1},$$

then this may be simplified, as in [7, Lemma 11.3], to

$$D^\alpha(f(x) - r^A_{\phi,\ell}(\mathbf{x})) = \left(f - r^A_{\phi,\ell}, D^\alpha\phi(\cdot - \mathbf{x}) - \sum_{i=1}^{n} c_i^\alpha \phi(\cdot - \mathbf{x}_i)\right)_{\psi,\mathcal{D}},$$

exactly as in the positive definite case.