Variably scaled kernels

M. Bozzini
jointed with L. Lenarduzzi, M. Rossini, R. Schaback

Maia 2013, Erice, September 25-30, 2013

Native Spaces

See e.g [Buhmann 2003], [Fasshauer 2007], [Wendland 2005]

Let H be an Hilbert space and $\Omega \subset \mathbf{R}^d$.

- The function
$$K \; : \; \Omega \times \Omega \to \mathbf{R}$$
is called reproducing kernel for H if
$$K(x, \cdot) \in H \quad \forall x \in \Omega,$$
and
$$f(x) = (f, K(x, \cdot))_H \text{ for all } x \in \Omega, \; f \in H.$$
- The kernel is positive definite, if for all choices of sets of knots
$$\{x_1, \ldots, x_N \in \Omega\},$$
the *kernel matrices* with elements $K(x_i, x_j)$, $1 \leq i, j \leq N$ are positive definite.

- For K positive definite, we define the *native space*

$$\mathcal{H}(K,\Omega) = \mathrm{span}\{K(x,\cdot),\, x \in \Omega\}.$$

- If the kernel is *radial*, i.e. of the form

$$K(x,y) = \phi(\|x-y\|_2)$$

for a scalar function

$$\phi \;:\; [0,\infty) \to \mathbf{R},$$

the function ϕ is called a *radial basis function*.

Interpolation

Given a set

$$X = \{x_1, \dots, x_N\} \subset \Omega,$$

and the associated values

$$f = [f(x_1), \dots, f(x_n)]^T,$$

the goal is to find a continuous function

$$P_f : \mathbf{R}^d \to \mathbf{R}$$

such that

$$P_f(x_i) = f(x_i), \quad i = 1, \dots, N.$$

In the RBF literature

$$P_f(x) = \sum_{i=1}^{N} a_i \phi(\|x - x_i\|),$$

where the coefficients are the solution of the linear system

$$a = A^{-1} f,$$

and

$$A_{ij} = \phi(\|x_i - x_j\|).$$

Properties

- $$P_f = \operatorname{argmin}\{\|s\|_{\mathcal{H}} : s \in \mathcal{H},\ s(x_i) = f(x_i),\ i = 1, \dots, N\}$$

- $$\|f - P_f\|_{\mathcal{H}} \leq \|f - s\|_{\mathcal{H}},$$

$$s \in \mathcal{H}_X = \{\sum_{i=1}^{N} \alpha_i \phi(\|x - x_i\|),\ x_i \in X\}$$

Scale parameter

Fixed a positive number c

$$K(x, y; c) := K(x/c, y/c) \quad x, y \in \mathbf{R}^d.$$

[Franke 82]

$$c = \frac{0.8\sqrt{N}}{D},$$

D is the diameter of the smallest circle containing all data points.

[Rippa 1999],[Fasshauer et al. 2007]

- $$K\left(\frac{\|x - x_j\|}{c_j}\right)$$

[Hardy 71], [Kansa 1990, 1992, 2000, 2006]
[B., Lenarduzzi, Schaback 2002],
[B., Lenarduzzi, Rossini, Schaback 2004] [Fornberg and Zuev 2007]

- $$K\left(\frac{x_1 - y_1}{c_1}, \ldots, \frac{x_d - y_d}{c_d}\right)$$

[B., Lenarduzzi 2005], [Fasshauer 2012]

Introduction

Given a domain $\Omega \subset \mathbf{R}^d$, we consider a bijective map

$$C \ : \ \Omega \mapsto C(\Omega).$$

Given a kernel

$$K \ : \ \Omega \times \Omega \to \mathbf{R}$$

and the map C, the kernel

$$K_C(C(x), C(y)) := K(x, y) \text{ for all } x, y \in \Omega$$

acts on $C(\Omega)$ and inherits the definiteness properties of K.

This gives rise to two native spaces and their properties, i.e.

- the space $\mathcal{H}(K,\Omega)$

$$\begin{aligned} f(x) &= (f, K(x,\cdot)) \\ K(x,y) &= (K(x,\cdot), K(y,\cdot)) \end{aligned}$$

  for all $x, y \in \Omega$.

- The space $\mathcal{H}_C(K_C, C(\Omega))$

$$\begin{aligned} g(u) &= (g, K_C(u,\cdot))_{\mathcal{H}_C} \\ K_C(u,v) &= (K_C(u,\cdot), K_C(v,\cdot))_{\mathcal{H}_C} \end{aligned}$$

  for all $u, v \in C(\Omega)$.

We indicate by $\mathcal{C}$ the map

$$\mathcal{C} : f \text{ on } \Omega \to g \text{ on } C(\Omega)$$

such that

$$g(C(x)) = (\mathcal{C}f)(C(x)) := f(x).$$

Furthermore,

$$\begin{aligned} \mathcal{C}(K(\cdot, y))(C(x)) &:= K(x, y) \\ &= K_C(C(x), C(y)). \end{aligned}$$

The map $\mathcal{C}$ is linear.

- The two native spaces $\mathcal{H}$ and $\mathcal{H}_C$ are isometric:

$$\begin{aligned} (K(x, \cdot), K(y, \cdot))_{\mathcal{H}} &= K(x, y) \\ &= K_C(C(x), C(y)) \\ &= (K_C(C(x), \cdot), K_C(C(y), \cdot))_{\mathcal{H}_C} \end{aligned}$$

It follows that

$$(f, g)_{\mathcal{H}} = (\mathcal{C}f, \mathcal{C}g)_{\mathcal{H}_C}.$$

Summarizing

Let

$$C \ : \ \Omega \mapsto C(\Omega).$$

be a bijective map and

$$K \ : \ \Omega \times \Omega \to \mathbf{R}$$

a positive definite kernel.

- We introduce the transformed kernel $K_C(C(x), C(y))$ which ineherits the definitness property of K.
- The native spaces $\mathcal{H}(K, \Omega)$, $\mathcal{H}_C(K_C, C(\Omega))$ are isometric.

Definition and Properties

Let

$$C \ : \ x \in \Omega \subset \mathbf{R}^d \mapsto (x, c(x)) \in C(\Omega) \subset \mathbf{R}^{d+1}$$

where

$$c \ : \ R^d \mapsto (0, \infty).$$

- Let K a positive definite kernel on $\mathbf{R}^{d+1}$

$$K_c(x, y) := K((x, c(x)), (y, c(y))) \quad x, y \in \mathbf{R}^d.$$

K_c is the *Variably scaled kernel*

- Since K is positive definite on the submanifold, so is K_c.
- If K and c are continuous, so is K_c.

Therefore, given the set $X := \{x_1, \dots, x_N\}$ on $\mathbf{R}^d$, the matrix

$$A_{c,X} := (K_c(x_i, x_j))_{1 \le i,j \le N}$$

is non singular and the interpolant is

$$P_f(x) := \sum_{j=1}^{N} a_j K_c(x, x_j) = \sum_{j=1}^{N} a_j K((x, c(x)), (x_j, c(x_j))).$$

If the kernel is radial, i.e. $K(x, y) = \phi(\|x - y\|_2^2)$, the interpolant is

$$P_f := \sum_{j=1}^{N} a_j \phi(\|x - x_j\|_2^2 + (c(x) - c(x_j))^2).$$

Examples

If K is a power kernel $\phi(r) = r^\beta$, the interpolants take the form

$$P_f(x) := \sum_{j=1}^{N} a_j \left(\|x_j - x\|_2^2 + (c(x_j) - c(x))^2\right)^{\beta/2}$$

and are identical to power interpolants if the scale function $c(x)$ is constant, otherwise similar to scaled multiquadrics.

If K is the Gaussian.

$$\begin{aligned} P_f(x) &= \sum_{j=1}^{N} a_j \exp(-\|x_j - x\|_2^2 - (c(x_j) - c(x))^2) \\ &= \sum_{j=1}^{N} a_j \exp(-\|x_j - x\|_2^2) \exp\left(-(c(x_j) - c(x))^2\right) \end{aligned}$$

which can be seen as a superposition of Gaussians of the same scale but with varying amplitudes for evaluation.

We observe that

- the analysis of error and stability of the varying–scale problem in $\mathbf{R}^d$ coincides with the analysis of a fixed–scale problem on a submanifold in $\mathbf{R}^{d+1}$.
- In particular, let Ω be a compact set and C be a diffeomorphism between Ω and $C(\Omega)$, then $C(\Omega)$ is compact.

As usual, we consider the *fill distance*

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

and the *separation distance*

$$q(X) := \min_{X \ni x \neq y \in X} \|x - y\|_2$$

Then

$$q(C(\Omega)) = \min \|C(x) - C(y)\|_2$$

and

$$\begin{aligned} \|C(x) - C(y)\|_2^2 &= \|x - y\|_2^2 + (c(x) - c(y))^2 \\ &\leq \|x - y\|_2^2 (1 + L)^2 \\ \|C(x) - C(y)\|_2^2 &\geq \|x - y\|_2^2 \end{aligned}$$

L is a constant related to the norm of the gradient of c.

It follows that *the separation distance never decreases.*

Numerical examples

We now provide some examples that show the different roles of the variable scale parameter: it may affect both the stability and the accuracy.

- Its appropriate choice enhances stability,
- one can significantly improve the recovery quality, in particular by preserving shape properties in a much better way than for interpolation with constant scale.

Chebyshev Points

We chose the Gaussian kernel at fixed scale $0.1/\sqrt{2}$ and took N=55 Chebyshev points $\Omega = [-1, +1]$ from Runge function

$$f(x) = 1/(1 + 25x^2).$$

We map the interval $\Omega = [-1, +1] \subset \mathbf{R}$ to the semi–circle $C(\Omega) \subset \mathbf{R}^2$ via

$$C(x) = (x, \sqrt{1 - x^2}).$$

The L_∞ errors and condition numbers are

Points and scaling	Condition	no noise	0.001 noise
Chebyshev, single scale	$1 \cdot 10^{16}$	$1.1 \cdot 10^{-5}$	1.4294
Chebyshev, variable scale	$8 \cdot 10^{5}$	$1.3 \cdot 10^{-4}$	0.0012

Table: Interpolation of Runge function by Gaussians

Cluster of data

We take $N = 47$ points $x_i \in [-1, 1]$ so that 41 nodes are equispaced in the interval and 6 close to 0.4, with mutual distance $q = 10^{-4}$.

As $c(x)$, we consider the *skew-Gaussian*

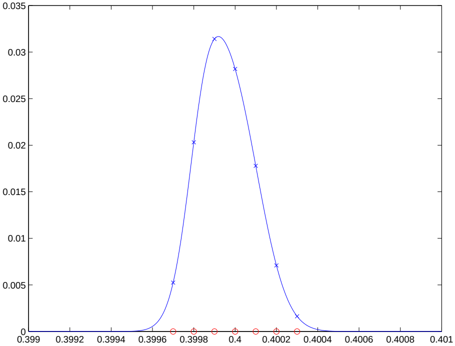

Figure: $C(x) = (x, c(x))$ with some $c(x_i)$ values

In this table, we show the condition numbers and the L_∞ errors obtained interpolating the Runge function by the Gaussian kernel at fixed scale $0.1/\sqrt{2}$ and by the proposed technique (VSK).

Points and scaling	Condition	error	0.001 noise
cluster, single scale	$3.5 \cdot 10^{16}$	$6.02 \cdot 10^{-5}$	$5.4 \cdot 10^{-1}$
cluster, variable scale	$6.9 \cdot 10^{10}$	$9.4 \cdot 10^{-6}$	$3.0 \cdot 10^{-3}$

Table: Interpolation of Runge function by Gaussians, cluster nodes

Plots of the absolute error for the two interpolants

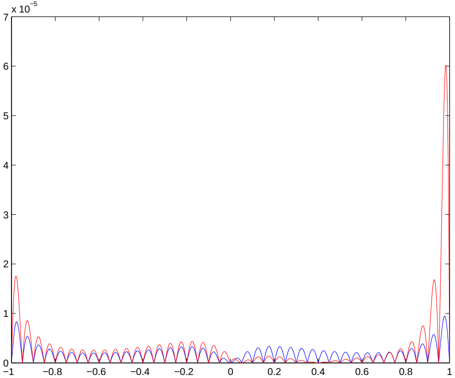

Figure: No noise case; absolute error for the classic case in red; absolute error for the VSK-interpolant in blue

Now, we deal with the problem of obtaining interpolants which reproduce faithfully the underlying functions.
(see e.g. [B., Lenarduzzi 2003], [Casciola et al. 2006]).

In the following examples, we compare

- the classical interpolant provided by the C^2 Wendland kernel with support radius 1
- the VSK interpolation provided by the d-variate C^2 Wendland kernel with support radius $\mu(C(\Omega))^{1/d}$, where μ is the length or area of $C(\Omega)$.

The logistic function

We consider the logistic function

$$f(x) = (1 + 2 \cdot \exp(p(x)))^{-0.5}$$

where $p(x) = -3 \cdot (10\sqrt{2x^2} - 6.7)$. We take $N = 11$ nodes in the interval $\Omega = [0, 1]$.

$$c(x) = 2 \cdot s_{MQ}(x).$$

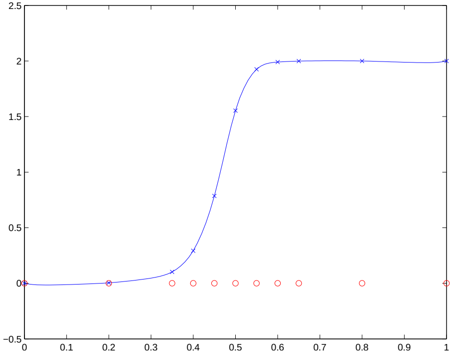

Figure: $C(x) = (x, c(x))$

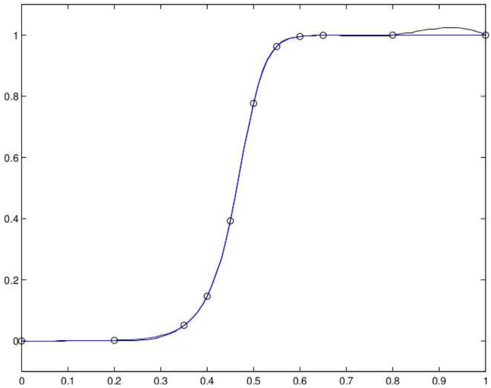

Figure: Classical Wendland interpolant: black line; logistic function: blue line

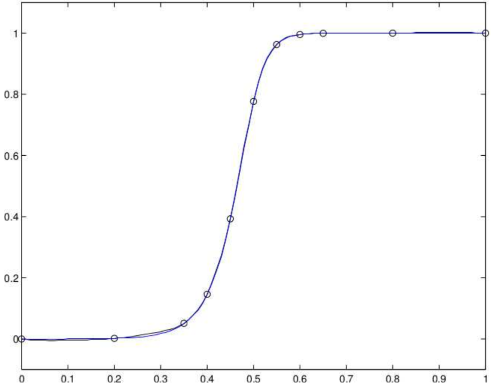

Figure: VSK-interpolant: black line; logistic function: blue line

The L_∞ errors are $2.5 \cdot 10^{-2}$ for the fixed–scale case and $6.4 \cdot 10^{-3}$ for the variable–scale case.

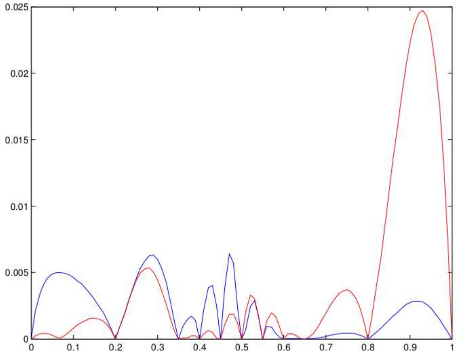

Figure: Absolute error for the classic interpolant: red; absolute error for the VSK interpolant: blue

The valley

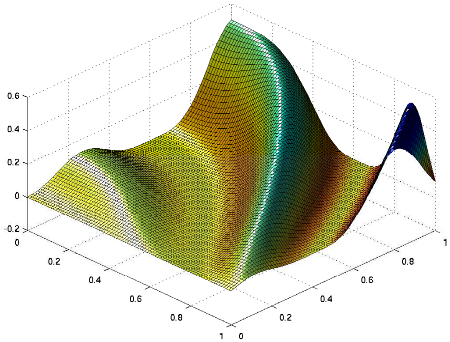

Figure: Test function

We have considered $N = 257$ points

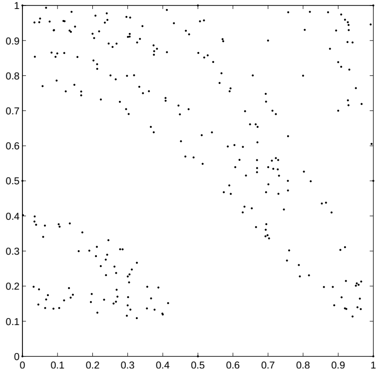

Figure: Locations of the data

$$c(x, y) = 0.5 s_{MQ}(x, y).$$

The VSK interpolant

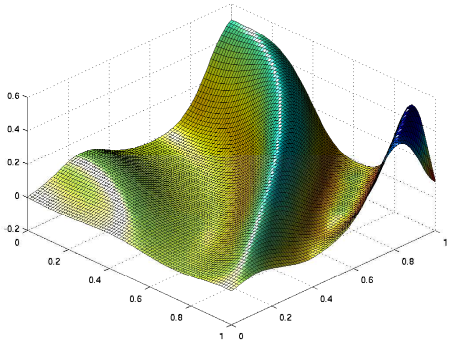

Figure: VSK interpolant

Errors for the VSK and classic interpolants

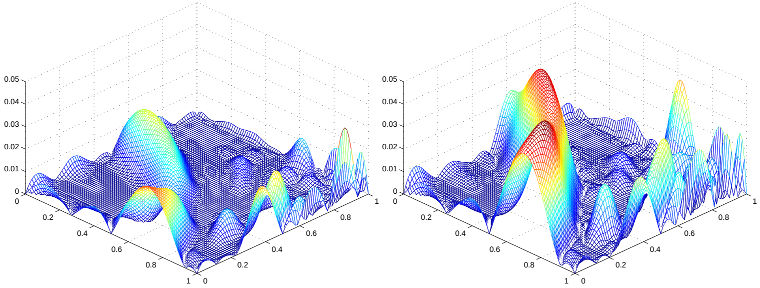

Figure: Left: VSK interpolant error. Right:Classic interpolant error

The L_∞ errors and condition numbers are $3.4e-2$, $1.0e+8$ the variable–scale case and $4.9e-2$, $9e+7$ for the classic Wendland's interpolant.

THANK YOU FOR YOUR ATTENTION!

Aronszajn N. (1950)
Theory of reproducing kernels.
Trans. Amer. Math. Soc., **68**, 337–404.

Bozzini M., Lenarduzzi L. (2005)
Reconstruction of surfaces from a not large data set by interpolation.
Rendiconti di Matematica, Serie VII, **25**, 223–239.

Bozzini M., Lenarduzzi L. (2003)
Bivariate knot selection procedure and interpolation perturbed in scale and shape.
Curve and Surface Design,
T. Lyche, M. L. Mazure and L. Schumaker eds, 1–10.

Bozzini M., Lenarduzzi L., Rossini M., Schaback R. (2004)
Interpolation by basis functions of different scales and shapes.
Calcolo, **41** (2), 77–87.

Bozzini M., Lenarduzzi L., Schaback R. (2002)
Adaptive interpolation by scaled multiquadrics.
Advances in Computational Mathematics, **16** 375–387.

Buhmann M. D. (2003)

Radial Basis Functions, Theory and Implementations.
Cambridge University Press.

Casciola G., Lazzaro D., Montefusco L. B., Morigi S.(2006)
Shape preserving reconstruction using locally anisotropic radial basis function interpolants.
Computer and Mathematics with Applications, **1**, 1185–1198.

De Marchi S., Schaback R., Wendland H. (2005)
Near-optimal data-independent point locations for radial basis function interpolation.
Adv. Comput. Math., **23**(3), 317–330.

Driscoll T. A., Fornberg B. (2002)
Interpolation in the limit of increasingly flat radial basis functions.
Comput. Math. Appl., **43**, 413–422.

Fasshauer G., Zhang J.G. (2007)
On choosing optimal shape parameters for rbf approximation.
Numerical Algorithms, **45**, 345–368.

Fasshauer G. (2007)
Meshfree Approximation Methods with MATLAB, volume 6 of *Interdisciplinary Mathematical Sciences*.
World Scientific Publishers, Singapore.

Fornberg B., Wright G., Larsson L. (2004)
Some observations regarding interpolants in the limit of flat radial basis functions.
Computers & Mathematics with Applications, **47**, 37–55.

Fornberg B., Zuev J. (2007)
The Runge phenomenon and spatially variable shape parameters in RBF interpolation.
Comput. Math. Appl., **54** (3), 379–398.

www.math.nps.navy.mil/ rfranke.

Kansa E. J., Carlson R.E. (1992)
Improved accuracy of multiquadric interpolation using variable shape parameters.
Comput. Math. Appl., **24**, 99–120.

Iske A. (2004)
Multiresolution Methods in Scattered Data Modelling, volume 37 of
Lecture Notes in Computational Science and Engineering.
Springer.

Jamshidi A. A., Kirby M. J. (2009)

Skew-radial basis function expansions for empirical modeling.
SIAM J. Sci. Comp. **31**, 4715–4743.

Larsson E., Fornberg B. (2005)
Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions.
Comput. Math. Appl., **49**,103–130.

Lee Y.J., Yoon G.J., Yoon J. (2007)
Convergence property of increasingly flat radial basis function interpolation to polynomial interpolation.
SIAM J. Mathematical Analysis, **39**, 470-488.

Rippa S. (1999)
An algorithm for selecting a good value for the parameter c in radial basis function interpolation.
Adv. Comput. Math., **11** (2-3), 193–210.

Schaback R. (2005)
Multivariate interpolation by polynomials and radial basis functions.
Constructive Approximation, **21**, 293–317.

Schaback R. (2008)
Limit problems for interpolation by analytic radial basis functions.

J. Comp. Appl. Math., **212**, 127–149.

Schaback R., Wendland H. (2006)
Kernel techniques: from machine learning to meshless methods.
Acta Numerica, **15**, 543–639.

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

Wendland H. (2005)
Scattered Data Approximation.
Cambridge University Press.