

## Variably scaled kernels: an overview

Milvia Rossini<sup>a</sup>

### Abstract

This paper aims to provide an overview of variably scaled kernels introduced in [10] and review their applications highlighting the contribution by Stefano De Marchi and his collaborators, especially in the theoretical study of the so-called variably scaled discontinuous kernels.

## 1 Introduction

Variably scaled kernels (VSKs) are introduced in [10] to provide an alternative technique to handle the problem of choosing the scale or shape parameter in kernel-based interpolation problems. The idea is to define a *scale function*

$$c : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}, \tag{1}$$

and to transform an interpolation problem from data-sites  $x_j, j = 1, \dots, N$  in  $\mathbb{R}^d$  to data locations  $(x_j, c(x_j))$  in  $\mathbb{R}^{d+1}$  and to use a fixed-scale kernel on  $\mathbb{R}^{d+1}$ . The  $(d + 1)$ -variate solution is then evaluated at  $(x, c(x))$  for  $x \in \mathbb{R}^d$  and gives the  $d$ -variate interpolant with a varying scale with  $x$ . Various examples in [10], show how VSK interpolants provide better results both in terms of stability and accuracy than the fixed-scale kernels. They work quite well in cases spoiled by a considerable instability of the standard method and they can significantly improve the recovery quality by preserving shape properties and particular features of the function underlying the given data. In addition, the background theory coincides with fixed-scale interpolation on the sub-manifold of  $\mathbb{R}^{d+1}$  given by the points  $(x, c(x))$  of the graph of  $c$ .

This paper aims to review the applications and further studies on VKSs that the scientific community has been pursuing in recent years. Important protagonists are certainly Stefano de Marchi and his collaborators, especially for the theoretical study of the so-called variably scaled discontinuous kernels (VSDKs).

The outline of the paper is as follows. In Section 2 we briefly review the main theoretical aspects of kernel-based interpolation, the role of the shape parameter for fixed scale kernels, spatially variable scale parameters, and the VSK theory. Section 3 deals with stability issues while Section 4 shows how a proper choice of the scaling function  $c$  allows for a faithful interpolation of the function underlying the given scattered data and highlights how they can be a valuable tool to approximate non-smooth functions. Section 5 is devoted to the special case of variably scaled discontinuous kernels (VSDKs). Finally, concluding remarks are drawn in Section 6.

## 2 Background

### 2.1 Preliminaries

The reader can refer to [37, 67, 75] for more details on the topic. A symmetric *kernel*

$$K : \Omega \times \Omega \rightarrow \mathbb{R}$$

defined on a domain  $\Omega \subseteq \mathbb{R}^d$  is very useful for a variety of purposes going from interpolation or approximation to solving PDE, if certain *nodes* or *centers*  $X = \{x_1, \dots, x_N\} \subset \Omega$  are used to define *kernel translates*  $K(\cdot, x_j)$  as *trial functions*.

If the kernel is positive definite, i.e. the *kernel matrices* with elements  $K(x_i, x_j), 1 \leq i, j \leq N$  are positive definite for all choices of nodes, there is a *native* Hilbert space  $\mathcal{N}_K(\Omega)$  in the background in which the kernel is reproducing, i.e.

$$g(x) = (g, K(\cdot, x))_{\mathcal{N}_K(\Omega)} \quad \forall x \in \Omega, \quad \forall g \in \mathcal{N}_K(\Omega).$$

The use of reproducing kernels in Hilbert spaces leads to various optimality properties and applications. Interpolation of values

$$\{f_1, \dots, f_N\} \quad \text{on} \quad X = \{x_1, \dots, x_N\} \tag{2}$$

proceeds via solving a linear system

$$A_X a = f, \quad f = [f_1, \dots, f_N]^T,$$

with interpolation matrix  $A_X = (K(x_j, x_k))_{1 \leq j, k \leq N}$ , which is positive definite. The coefficient vector  $a \in \mathbb{R}^N$  then allows the interpolant function to be written as

$$s_{X,f}(x) = \sum_{j=1}^N a_j K(x, x_j). \tag{3}$$

<sup>a</sup>Dipartimento di Matematica e Applicazioni, Università degli Studi di Milano - Bicocca, Milano, Italy

By adding a lower degree polynomial to (3), the existence of a solution is ensured also for conditionally positive definite kernel. In the sequel we consider *radial* kernels, i.e.

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

for a scalar function

$$\phi : [0, \infty) \rightarrow \mathbb{R}.$$

The function  $\phi$  is called *radial basis function (RBF)*.

When dealing with scattered data and kernel-based interpolation, two indicators of data regularity have to be considered: *separation* and *fill-distances*.

**Definition 2.1.** The separation-distance is

$$q_X = \frac{1}{2} \min_{i \neq k} \|x_i - x_k\|_2. \tag{4}$$

It represents the radius of the largest ball that can be centered at every point in  $X$  such that no two balls overlap.

**Definition 2.2.** The fill-distance is

$$h_X = \sup_{x \in \Omega} \left( \min_{x_k \in X} \|x - x_k\|_2 \right). \tag{5}$$

The quantity (5) is a measure of data distribution and indicates how well the data fill out the domain  $\Omega$ . It denotes the radius of the largest empty ball that can be placed among the data locations. They are respectively related to the interpolation matrix conditioning and to the accuracy of the solution. It is worthwhile to remember the *uncertainty or trade-off principle*: there is always a conflict between the theoretical accuracy that one can obtain and the numerical stability (see e.g. [68] and [70]). For well-distributed data (i.e.  $q_X \approx h_X$ ) a small error bound usually implies a large condition number of  $A_X$ .

## 2.2 Scaled Kernels

It is well-known that kernels on  $\mathbb{R}^d$  can be *scaled* by a positive factor  $\delta$ . This turns to have a new kernel

$$K(x, y; \delta) = K(x/\delta, y/\delta), \quad \forall x, y \in \mathbb{R}^d. \tag{6}$$

In general, a large  $\delta$  increases the ill-conditioning of kernel matrices, while small  $\delta$  let the translates turn into sharp peaks which approximate functions badly, if separated too far from each other.

The shape or scale parameter can be tuned by the user (according to the applications) and it plays an important role both in the accuracy of the method and its stability. How to choose the scale parameter  $\delta$  is a well-documented problem but not been solved yet. It started with Hardy many years ago with multiquadrics. When working with a fixed scale one has several possibilities. The most popular ones are to pick the parameter by some ad hoc criterion as done by Hardy and later by Franke (see [40] and the reference therein), or to choose the parameter by some optimal criterium based for instance on a variant of the cross-validation approach (leave-one-out) [63] and on its extension applied in the setting of iterated approximate moving least squares [40] or in a more general  $k$ -fold cross validation deterministic and stochastic setting [52, 56]. Further optimization and searching techniques were considered in [20, 21, 72]. A special case of scaling is the *flat limit*  $\delta \rightarrow \infty$  (see e.g. [36, 43, 48, 49, 69]).

Typically, optimal values, i.e. shape parameters providing good accuracy, give instability. As shown by R. Schaback in [70], trade-off Principle between error and stability is unavoidable and "If users have strong reasons to insist on very good accuracy, they have to face serious evaluation instabilities" and it is a challenge to cope with these. The literature on kernel-based methods provides several of such techniques, e.g. Contour-Padé [42], RBF-QR [38, 41], and Hilbert-Schmidt-SVD in [39, Chapter 13]. In [29] it is presented a *rescaled-method* based on a proper selection of the supports of compactly supported basis functions that allows keeping the ill-conditioning under control.

People considered also the possibility to have the scale of a kernel translate varying with the translation. This means working with functions

$$\phi(\|x - x_j\|_2 / \delta_j), \quad 1 \leq j \leq N$$

in the radial case, [11, 44, 47]. In these cases, it is easy to come up with examples that let interpolation fail for certain non-uniform choices of scale parameters. But in [9] sufficient conditions for the unique solvability of such interpolation processes are given.

This non-constant shape parameters framework is commonly referred to as interpolation with variable shape kernels when  $\delta_j = \delta(x_j)$  for some smooth function  $\delta$  and with random shape kernels when  $\delta_j$  is a positive random variable following some probability distribution. See [22] where various random distributions were numerically examined for Gaussian and anisotropic Gaussian kernels.

## 2.3 Variably Scaled Kernels

So far, it is clear the importance of the scale parameter or scale parameter vectors, and, to make more effective their potential, the vector case can be generalized by introducing a *scale function*. In [10], this is done by letting the scale parameter be an additional coordinate. This continuously allows varying scales without leaving the well-established theory of kernel-based interpolation. It turns out that this approach can be fully understood as the standard fixed-scale method applied to a certain sub-manifold of  $\mathbb{R}^{d+1}$ .

**Definition 2.3.** Let  $K$  be a kernel on  $\mathbb{R}^{d+1}$ . If a *scale function*

$$c : \mathbb{R}^d \rightarrow (0, \infty)$$

is given, we define a *Variably Scaled Kernel (VSK)* on  $\mathbb{R}^d$  by

$$K_c(x, y) = K((x, c(x)), (y, c(y))), \quad \forall x, y \in \mathbb{R}^d. \tag{7}$$

If  $K$  is positive definite on  $\mathbb{R}^{d+1}$ , so is  $K_c$  on  $\mathbb{R}^d$ . Then interpolation of (2) proceeds as usual by solving a linear system  $A_{c,X} a = f$  with a positive definite kernel matrix  $A_{c,X} = (K_c(x_j, x_k))_{1 \leq j, k \leq N}$ . A VSK interpolant is then written as

$$s_{c,X,f}(x) = \sum_{j=1}^N a_j K_c(x, x_j) = \sum_{j=1}^N a_j K((x, c(x)), (x_j, c(x_j))),$$

and if the kernel  $K(x, y) = \phi(\|x - y\|_2^2)$ , the interpolant is

$$s_{c,X,f}(x) = \sum_{j=1}^N a_j \phi(\|x - x_j\|_2^2 + (c(x) - c(x_j))^2).$$

Note that the interpolant is identical to the standard one if the scale function  $c$  is constant.

The scale function  $c(x)$  introduces a map

$$C : x \mapsto (x, c(x)) \tag{8}$$

that maps the space  $\mathbb{R}^d$  into a  $d$ -dimensional sub-manifold  $C(\mathbb{R}^d)$  of  $\mathbb{R}^{d+1}$  and a set of nodes  $X = \{x_1, \dots, x_N\} \subset \Omega \subset \mathbb{R}^d$  goes into  $C(X) \subset C(\Omega) \subset C(\mathbb{R}^d) \subset \mathbb{R}^{d+1}$ . As a consequence, interpolation by the fixed-scaled kernel  $K$  takes place on  $\mathbb{R}^{d+1}$  at the point set

$$C(X) = \{(x_1, c(x_1)), (x_2, c(x_2)), \dots, (x_N, c(x_N))\}.$$

This means that in  $\mathbb{R}^{d+1}$  the kernel  $K(C(x), C(y))$  is used and if we project points  $C(x) = (x, c(x)) \in \mathbb{R}^{d+1}$  back to  $x \in \mathbb{R}^d$ , the projection of the kernel turns into the VSK  $K_c(x, y)$  on  $\mathbb{R}^d$ , and the analysis of error and stability of the variably-scale problem in  $\mathbb{R}^d$  coincides with the analysis of a fixed-scale problem on a sub-manifold in  $\mathbb{R}^{d+1}$ . This gives rise to two native spaces  $\mathcal{N}_K(C(\Omega))$ , and  $\mathcal{N}_{K_c}(\Omega)$  that are isometric [10, Theorem 3.1].

The fill-distance (5) and the separation-distance (4) will transform with  $C$ , and will roughly be multiplied by a factor related to the norm of the gradient of  $c$  or to the Lipschitz constant  $L$  of  $c$ , depending on the regularity of  $c$ . Namely,

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

This shows that distances will blow up with  $C$ , letting separation-distances never decrease, thus enhancing stability. Fill-distances may also blow up, increasing the usual error bounds. But this argument shows that one can successfully use the varying-scale technique on points  $x_j$  and  $x_k$  that have very small separation-distances until one roughly gets that the transformed centers are approximately uniformly distributed, that is

$$q_{C(X)} \approx h_X \approx h_{C(X)},$$

improving in this way the interpolation matrix condition number and preserving the accuracy.

The examples in [10], show how this can be done in practice. Suitable choices of  $C$  make VSKs rather effective when a considerable instability of standard methods spoils the results, or can significantly improve the recovery quality by preserving shape properties and particular features of the underlying function. The following paragraphs discuss these issues through some practical choices of the map  $C$ .

### 3 Stability issues

Let us start with the one-dimensional case with  $\Omega = [-1, +1]$ , and two configurations of data-sites: Chebyshev points of the second kind,  $x_j = -\cos(\pi(j-1)/(N-1))$ ,  $1 \leq j \leq N$ , and scattered (random) points. As test function we consider the Runge function  $f(x) = 1/(1 + 25x^2)$ , and samples of size  $N = 35, 45, 55$ . The data samples will be both exact and affected by a small additive noise from a normal distribution with zero mean and standard deviation  $\sigma$ . As a specific case, we choose the Gaussian kernel at the fixed scale  $0.1 \cdot \sqrt{2}$ .

#### 3.1 Chebyshev points

For Chebyshev points the fill-distance behaves like  $1/N$ , while the separation-distance behaves like  $1/N^2$ , and this leads to a very large condition in the kernel matrices, no matter which kernel is chosen. To handle that, the most natural choice is to map the interval  $\Omega = [-1, +1] \subset \mathbb{R}$  to the semi-circle  $C_1(\Omega) \subset \mathbb{R}^2$  via  $C_1(x) = (x, \sqrt{1-x^2})$ . Then the resulting points are equidistant, and the separation-distance will now behave like the fill-distance (see Table 1), i.e. like  $1/N$ , and we can work with a single-scale kernel in  $\mathbb{R}^2$  for interpolation in  $C_1(X)$ . The condition numbers of the fixed-scale and VSK matrices together with the maximum interpolation errors  $e_\infty$  for exact and noisy data with  $\sigma = 0.001$  are given in Table 2. If there is no noise, the single-scale methods are slightly superior despite the growing ill-conditioning. When small additive noise corrupts the data, the bad condition spoils the results very seriously while, as expected, the significant reduction in the condition numbers of VSK matrices saves the situation.

$N$	$h_X$	$q_X$	$h_{C_1(x)}$	$q_{C_1(x)}$
35	4.560e-02	2.133e-03	4.612e-02	4.618e-02
45	3.533e-02	1.274e-03	3.567e-02	3.569e-02
55	2.852e-02	8.459e-04	2.879e-02	2.908e-02

Table 1: Fill-distances and separation-distances of the given Chebyshev points and of the mapped ones by  $C_1$ .

$N$	cond fixed-scale	fixed-scale $e_\infty$	fixed scaled $e_\infty$ (noise)	cond VSK	VSK $e_\infty$	VSK $e_\infty$ (noise)
35	3.269e+10	6.606e-04	2.000e-01	1.496e+02	7.888e-04	2.959e-03
45	2.004e+15	8.028e-05	7.398e+00	6.569e+03	1.273e-04	2.683e-03
55	9.783e+17	1.093e-05	3.298e+00	7.102e+05	1.871e-05	3.457e-03

Table 2: Interpolation of Runge function by fixed-scale and variably scaled Gaussians on Chebyshev nodes.

### 3.2 Scattered points

As shown in [10], even with scattered points  $\{x_1, x_2, \dots, x_N\}$  in  $[-1, 1]$ , it is possible to build a continuous map  $C_2(x) = (x, c_2(x))$  that maps them into equidistant points  $z_j \in \mathbb{R}^2$ . If  $h_X$  is the fill-distance of the given points, one can guarantee

$$\|z_{j+1} - z_j\|_2^2 = (x_{j+1} - x_j)^2 + (c_{j+1} - c_j)^2 = \alpha h_X^2, \quad \alpha \geq 1,$$

by choosing

$$c_{j+1} = c_j + \sqrt{\alpha h^2 - (x_{j+1} - x_j)^2},$$

starting with  $c_1 = 0$ . This gives a monotonic sequence, and by piecewise linear interpolation one obtains  $C_2$  with  $C_2(x_j) = z_j$ ,  $1 \leq j \leq N$ , and  $c_2(x)$  monotone increasing. Experiments with this strategy show a similar behavior as in the previous example. Table 3 displays the fill-distances and separation-distances of the given points and of the mapped ones by  $C_2$ , when  $\alpha = 1.5$ .

$N$	$h_X$	$q_X$	$h_{C_2(x)}$	$q_{C_2(x)}$
35	1.420e-01	9.014e-04	1.752e-01	1.753e-01
45	6.968e-02	9.014e-04	8.830e-02	8.836e-02
55	6.968e-02	9.014e-04	8.811e-02	8.836e-02

Table 3: Fill-distances and separation-distances of the given scattered points and of the mapped ones by  $C_2$ .

For  $N = 55$ , Figure 1 (left) shows the monotonic function  $c_2$  that increases sharply where there are nearby points, while it is flat over holes in the node set.

Unfortunately, the non-differentiability of the map  $C_2$  and its very edgy behavior spoils convergence rates, but thanks to the very low condition number of the VSK matrices, if we have noise with  $\sigma = 0.001$ , the variably scaled method outperforms the fixed-scale one confirming once again to be stable and robust.

Another way to improve both the accuracy and the stability is to find "smoother" maps that reduce separation-distances without increasing fill-distances too much. This can be done with a map  $C_3(x) = (x, c_3(x))$  (see Figure 1, right) with

$$c_3(x) = \sum_{j=1}^N p_j(x), \tag{9}$$

where the functions

$$p_j(x) = 1/\pi (\arctan(\gamma_j |(x - x_j)|) + \beta)$$

increase more steeply at  $x_j$  with a larger value of  $\gamma_j$ , if the local density of the data locations is large at  $x_j$ . The parameter  $\beta$  has been selected via trials and errors. The results obtained with  $\beta = 0.1$  are depicted in Tables 5 and 6. We observe that both the growths of  $h_{C_3(x)}$  and  $q_{C_3(x)}$  are moderate. The growth of  $q_{C_3(x)}$ , although being very moderate, makes sure to lower the condition number to levels that allow the backslash MatLab operator to provide reliable solutions. This and the fact that the fill-distance remains of the same order as that of the given points improves the accuracy too. The scaling function (9) can be extended to the two-dimensional case. Actually the 2D version  $C(x, y) = ((x, y), c(x, y))$  was initially introduced in [10] to enhance the interpolant reproduction quality. Here we consider

$$c(x, y) = \sum_{j=1}^N p_j(x, y), \quad p_j(x, y) = 1/\pi (\arctan(\gamma_j |(x - x_j)|) + \beta) \exp(-5 (y - y_j)^2). \tag{10}$$

Also, in this case, we look for "large" values of  $\gamma_j$  when the point density is large around  $x_j$ . Nevertheless, as observed in [34], this could be computationally expensive and that is why here we fix  $\gamma_j = 5e - 6$  for all  $j = 1, \dots, N$ . As before,  $\beta = 0.01$  has actually been selected via trials and errors. Table 7 provides fill and separation-distances for the mapped points compared with the ones of the original point sets which are Halton points in  $[0, 1]^2$  together with the corresponding kernel matrix condition numbers. It is clear that stability is improved.

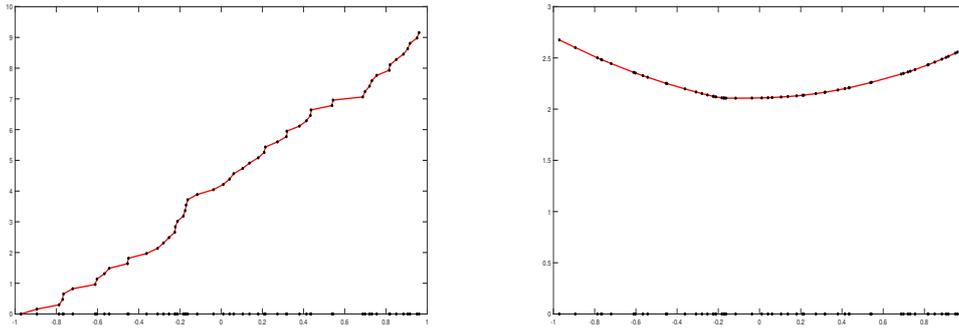


Figure 1: Left: the monotonic scale function  $c_2$  with  $\alpha = 1.5$ . Right: the scale function  $c_3$ .  $N = 55$ .

$N$	cond fixed-scale	fixed-scale $e_\infty$	fixed-scale $e_\infty$ (noise)	cond VSK	VSK $e_\infty$	VSK $e_\infty$ (noise)
35	1.853e+12	6.976e-03	2.438e+01	1.009e+00	5.864e-01	5.865e-01
45	5.980e+17	3.960e-04	3.531e+01	2.430e+00	2.538e-02	2.505e-02
55	2.004e+17	2.741e-03	7.634e+02	2.432e+00	1.727e-02	1.846e-02

Table 4: Interpolation of Runge function by fixed-scale and variably scaled by  $c_2$  Gaussians on scattered points ( $\alpha = 1.5$ ).

### 3.3 Final remarks on stability and related papers

The above examples show how the proper use of the variable scale kernel  $K_c$ , i.e. a proper choice of  $C(x) = (x, c(x))$  always leads to a more stable interpolant and this is crucial for applications. So far, VSKs have been used with success in various articles. In solving magnetohydrodynamic equations, [35] uses an ad hoc  $c(x, y)$  depending on the parameters of the considered problem. In [59] an adaptive learning algorithm for RBF neural networks is implemented on variably scaled kernels with the scaling function  $c_2$ . The algorithm turns out to be accurate and stable especially for noisy data sets. The map  $C_1$  is used in [58] to provide a meshless method to solve Burgers' equation which constructs derivatives discretizations based on variably scaled Newton basis functions for localized sets of nodes.

Stefano De Marchi and his collaborators perform in [33] a local computation via the partition of unity method of rational RBF interpolants. The proposed method allows us to consider large data sets but might suffer from instability due to the ill-conditioning of the local interpolation matrices. To avoid this drawback, they develop a stable computation of the rational RBF local interpolants employing the VSKs, and given that the local domains  $\Omega_i$  are circular patches, they consider the following scale functions

$$c_i(x, y) = 0.5 + \sqrt{v + [(x - \tilde{x}_i)^2 + (y - \tilde{y}_i)^2]}, \quad (x, y) \in \Omega_i, \quad v \geq 1,$$

that maps the circular patches centered in  $(\tilde{x}_i, \tilde{y}_i)$  on a semi-sphere of radius  $v$ . They also remark that in compactly supported RBFs, the scale function enables the user to control the sparsity of the interpolation matrix and the scale function

$$c(x, y) = v \sqrt{x^2 + y^2}, \quad v \geq 1$$

makes the sparsity of the interpolation matrix grows with  $v$ .

As mentioned before, (10) has been used in [34] where we investigate adaptivity issues for the approximation of Poisson equations via radial basis function-based partition of unity collocation. The adaptive residual sub-sampling approach is performed with quasi-uniform node sequences leading to a flexible tool which however might suffer from numerical instability that is overcome by a hybrid method that makes use of VSKs scaled by (10) with  $\gamma_j = 7e - 6$ , and  $\beta = 0$ . VSKs with the same map are used also in [14] to provide a reliable approximation of the solution of elliptic partial differential equation with a singular forcing term, describing a steady state flow determined by a pulse-like extraction at a constant volumetric rate.

## 4 Improving reproduction quality

VSKs can significantly improve the recovery quality by preserving shape properties and particular features of the underlying function. In this section we consider data sampled from smooth functions. We start with a simple 1D example. We consider the logistic function

$$f(x) = \frac{1}{\sqrt{1 + 2 \exp - 3(10\sqrt{2}x^2 - 6.7)}}, \tag{11}$$

and we take the  $N = 11$  nodes  $X = \{0, 0.2, 0.35, 0.4, 0.45, 0.5, 0.52, 0.66, 0.8, 1\}$  with higher density where the function changes more quickly and a lower density where it is nearly constant. We interpolate the set  $(X, f(X))$  by using the  $C^2$  Wendland RBF with  $\delta = 10$ . As shown in Figure 2 (left), the fixed-scaled interpolant exhibits an undesired oscillation. If we consider as scale

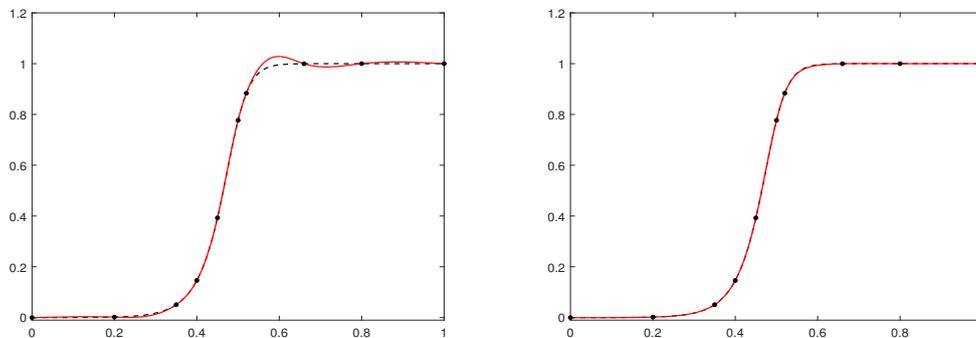
$N$	$h_X$	$q_X$	$h_{C_3(X)}$	$q_{C_3(X)}$
55	6.968e-02	9.014e-04	7.988e-02	9.442e-04
75	5.219e-02	1.470e-05	6.571e-02	1.704e-05
95	4.043e-02	1.470e-05	4.309e-02	1.660e-05

**Table 5:** Fill-distances and separation-distances of the given scattered points and of the mapped points by  $C_3$ .

$N$	cond fixed-scale	$e_\infty$ fixed-scale	cond VSK	$e_\infty$ VSK
55	2.004e+17	2.741e-03	1.250e+12	1.020e-03
75	6.814e+17	1.245e-03	9.787e+13	8.468e-05
95	2.063e+18	6.721e-05	6.085e+14	7.552e-06

**Table 6:** Interpolation of Runge function by fixed-scale and variably scaled by  $c_3$  Gaussians on scattered points.

function for the Wendland VSK  $c(x) = f(x)$ , the interpolant behaviour is definitely improved (see Figure 2, right). This is obvious as we are using complete information about the function we want to recover, but in general, this is not possible. Therefore finding a scale function that incorporates the features of the underlying function is a challenging problem. The examples in [10], reveal that using as scale function a "tight" multiquadric interpolant of the given data provides good results. Indeed, provided that some



**Figure 2:** Dashed black line:  $f(x)$ . Black dots: the given sample. Left: The  $C^2$  Wendland interpolant. Right: The VSK  $C^2$  Wendland interpolant.

a priori knowledge on the data sample is known, we can easily encode such information into the kernel and the VSKs can be also seen in the context of feature augmentation tools. Precisely, and as underlined by [17], many methods based on feature augmentation, e.g., zero padding and feature replication [25, 51, 57] fall into the general VSK setting. In [15] and [16], the authors present a novel procedure for the extrapolation from samples that decay with an exponential or rational trend. In this situation the scaling function is defined as a parametric weighted sum of a finite number of exponential terms, whose parameters are deduced by a non-linear fitting of the samples. In [60] and [61] VSKs are used for interpolating hard X-ray visibilities where the scaling function is defined by exploiting prior information on the flaring source. In this context, VSKs allow remarkable image reconstruction accuracy when the information on the flaring source is encoded by a small set of scattered Fourier data and when the visibility surface is affected by significant oscillations in the frequency domain. The link between VSKs and the field of machine learning is investigated in [17] with a particular focus on support vector machine classifiers and kernel regression networks where they propose to use a probabilistic approach based on the naive Bayes classifier in the first case, and non-linear fitting of the data as an augmented feature in the second case. Under appropriate assumptions, the VSKs turn out to be more expressive and more stable than standard kernels.

It is now well established that VSKs are an effective tool when one needs to interpolate data sampled from a non-regular function. In 1D,  $f$  and/or its first derivative may be discontinuous at some points while in 2D, we assume that  $f$  is smooth on  $n$  disjoint subsets  $\Omega_i$  of  $\Omega$  such that  $\Omega = \cup_{i=1}^n \Omega_i$ , and that  $f$  or its gradient is discontinuous across  $\partial\Omega_i$  which represents respectively the *edges (faults)* or *gradient faults* of  $f$ . It is evident that kernel-based methods are in principle not suitable for non-regular function reconstruction since the obtained interpolant are linear combinations of translates of smooth functions (at least continuous), so undesired artifacts appear in the final solution, i.e. the so-called Gibbs phenomenon occurs near the edges or, in the case of gradient discontinuities, creases are over-smoothed. VSKs contribute to filling the gap between the reconstruction of non-regular functions and kernels methods, and feature augmentation with suitable scaling functions allows us to incorporate non-regular behaviors in a very simple way and make the VSKs truly performing especially when information on singularities in the underlying function is a priori known. This was initially observed in the master thesis [18] and further explored in [66] in the case of functions with discontinuous derivatives. Around the same time, discontinuous scale functions were employed for edge detection [64] while their use in the reconstruction of discontinuous functions began to be studied by De Marchi and his collaborators.

$N$	$h_x$	$q_x$	$h_{C(x)}$	$q_{C(x)}$	cond fixed-scaled	cond VSK
81	1.272e-01	5.667e-04	1.278e-01	5.685e-04	1.202e+17	2.774e+03
256	8.367e-02	2.579e-04	8.512e-02	2.955e-04	2.538e+17	2.212e+04
1024	3.521e-02	1.584e-04	7.112e-02	2.077e-04	9.255e+17	2.826e+05
2048	2.873e-02	8.317e-05	8.478e-02	1.681e-04	1.785e+18	1.131e+06
4096	2.195e-02	2.513e-05	6.519e-02	6.606e-05	3.778e+18	2.310e+08

Table 7: Fill-distances, separation-distances of Halton points, of mapped ones by (10) and the associated kernel matrix condition numbers.

Here are some examples in one and two dimensions that confirm what has been stated above and show a way to define suitable scaling functions. In what follows we use the  $C^2$  Wendland kernel.

Example 4.1. Let us consider the test function

$$f(x) = \begin{cases} \sin 2x, & -1 \leq x \leq -0.3 \\ 0.2 \exp(-x^2) + 0.1, & -0.3 < x \leq 0.1 \\ |x^2 - 0.49|, & 0.1 < x \leq 1, \end{cases}$$

sampled at  $N = 55$  scattered points in  $[-1, 1]$ . We encode in the  $C^2$  Wendland kernel, as additional coordinates, three scaling functions  $\psi_1, \psi_2, \psi_3$ . The first two are step functions that mimic the jumps at  $x_1 = -0.3$ , and  $x_2 = 0.1$ , while the third mimics the corner at  $x_3 = 0.7$  (see Figure 3, left). For a generic interval  $[a, b]$  and discontinuity point  $x^*$  for  $f'$ ,  $\psi_3$  is defined as (see [66])

$$\psi_3(x) = \begin{cases} 1 - 3/2|x - x^*|/R + 1/2|x - x^*|^3/R^3, & |x - x^*| < R, \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

where  $2R$  is the support of  $\psi_3(x)$  which goes to zero smoothly. Extensive experiments have shown that  $R$  had to be chosen less than  $(b - a)/2$ . Actually, we obtain fully equivalent results with the scale function  $\psi$  in Figure 3 (right). Figure 4 shows the

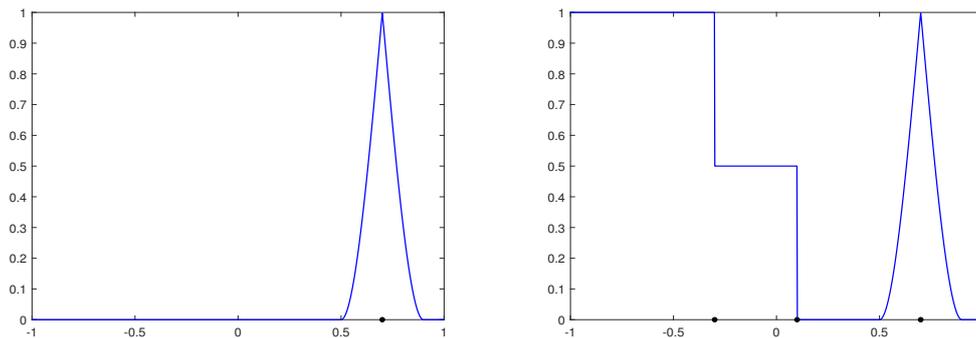


Figure 3: Left:  $\psi_3$ . Right:  $\psi$ . ( $R = 2$ ).

standard fixed-scaled ( $\delta = 1$ )  $C^2$  Wendland interpolant (right) and the VSK one (left) that outperforms the standard method as also shown by the root mean square error  $e_2$  and maximum error  $e_\infty$  in Table 8.

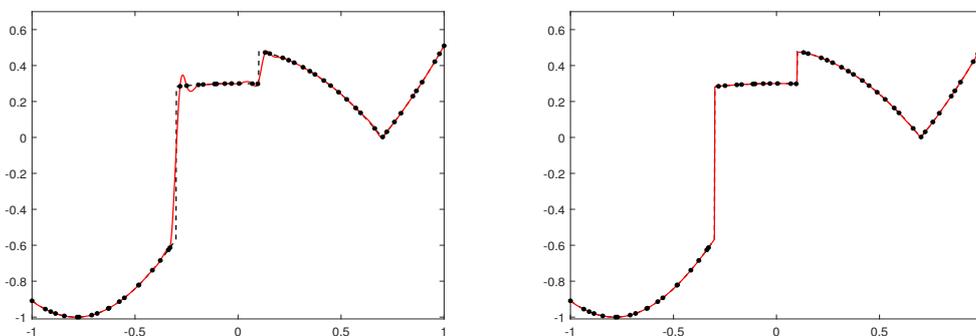


Figure 4: Dashed black line:  $f(x)$ . Left: fixed scale interpolant. Right: VSK interpolant.

error	fixed-scale	VSK
$e_\infty$	0.4739e+00	5.853e-03
$e_2$	0.1212e-02	1.695e-05

Table 8: Example 4.1: Maximum and root mean square errors.

error	fixed-scale	VSK
$e_\infty$	0.60881e+00	6.704e-03
$e_2$	0.67831e-03	5.716e-06

Table 9: Example 4.2: Maximum and root mean square errors.

**Example 4.2.** We consider  $N = 625$  Halton points in  $\Omega = [0, 1]^2$  and the corresponding values sampled from the test function depicted in Figure 5 (left)

$$f(x, y) = |y - 0.75 - 0.1 \sin(2\pi x)| + \begin{cases} x^2, & \text{if } (x - 0.3)^2 + (y - 0.3)^2 < (0.15)^2 \\ y^2 - 0.5, & \text{otherwise,} \end{cases} \quad (x, y) \in \Omega = [0, 1]^2,$$

which is discontinuous across  $\partial\Omega_1 = (x - 0.3)^2 + (y - 0.3)^2 = (0.15)^2$  and has first partial derivatives discontinuous across  $\partial\Omega_2 = y - 0.75 - 0.1 \sin(2\pi x) = 0$ . In this case, we can define the scale function that mimics both the jumps and the creases as

$$\psi(x, y) = \chi_{(x-0.3)^2+(y-0.3)^2 < (0.15)^2} + \begin{cases} 1 - 3/2|y - \Gamma(x)|/R + 1/2|y - \Gamma(x)|^3/R^3, & |y - \Gamma(x)| < R, \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

where  $\Gamma(x) = 0.75 + 0.1 \sin(2\pi x)$ , and  $R = 0.5$  (see Figure 5, right). As in the previous example, we could have encoded the different features by two scaling functions. Figure 6 shows the standard fixed-scaled ( $\delta = 10$ )  $C^2$  Wendland interpolant (left) and the VSK one (right) that, as expected, reproduces the features of the underlying function. The root mean square error  $e_2$  and maximum error  $e_\infty$  are given in Table 9.

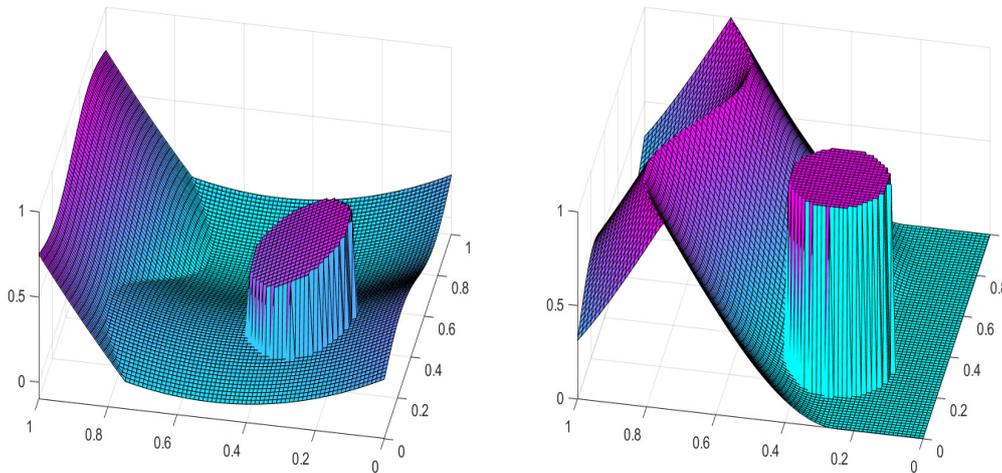


Figure 5: Left: the test function  $f(x, y)$ . Right: The scaling function (13).

## 5 Variably scaled discontinuous kernels

The examples 4.1 and 4.2 help us understand that the construction of a discontinuous kernel via a VSK is particularly simple, indeed it is enough to use a piecewise constant scale function. Although a discontinuous kernel based on a Mercer decomposition is given in [71], not much effort has been addressed to construct robust approximants for functions with jumps. This is done by Stefano De Marchi and his collaborators in [31] where they define this new class of variably scaled discontinuous kernels, and provide very general error bounds in terms of the power function. More precisely they consider the following setting:

*Assumption 5.1.* (i) The bounded set  $\Omega \subset \mathbb{R}^d$  is the union of  $n$  pairwise disjoint sets  $\Omega_i$ ,  $i \in \{1, \dots, n\}$ .

(ii) The subsets  $\Omega_i$  have a Lipschitz boundary  $\partial\Omega_i$ .

(iii) Let  $\Sigma = \{\alpha_1, \dots, \alpha_n\}$ ,  $\alpha_i \in \mathbb{R}$ . The function  $\psi : \Omega \rightarrow \Sigma$  is piecewise constant so that  $\psi(x) = \alpha_i$  for all  $x \in \Omega_i$ . In particular, the jumps of  $\psi$  appear only at the boundaries of the subsets  $\Omega_i$ . We assume that  $\alpha_i \neq \alpha_j$  if  $\Omega_i$  and  $\Omega_j$  are neighboring sets.

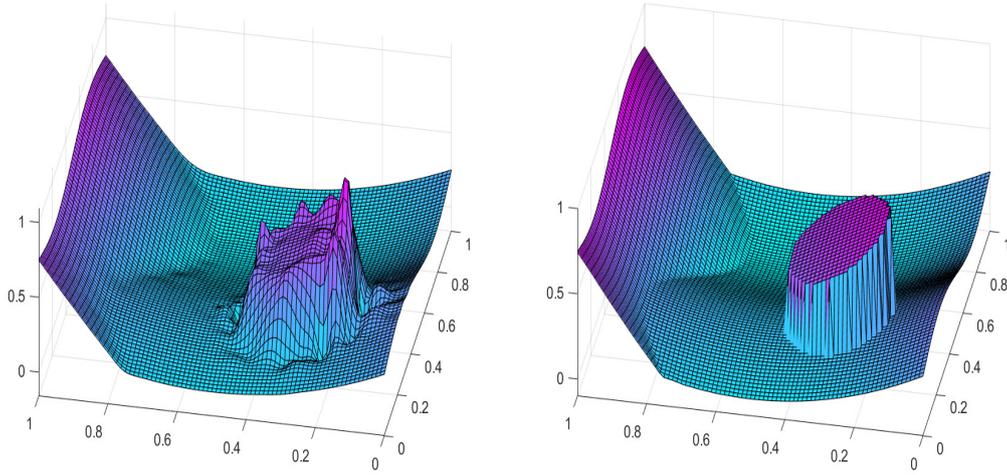


Figure 6: Left: fixed scale interpolant. Right: VSK interpolant.

The associated VSK  $K_\psi$  is only piecewise continuous and denoted as variably scaled discontinuous kernel (VSDK). The scale function  $\psi$  can be seen as the limit for  $k \rightarrow +\infty$  of a suitable sequence of continuous scaling functions  $\{\psi_k\}_{k \in \mathbb{N}}$  and this allows them to prove the following error bound.

**Proposition 5.1** ([31]). *Let  $X = \{x_1, \dots, x_N\}$  be a set of distinct points in  $\Omega$  and  $\psi$  be defined as in (5.1). Then for all  $f$  in the native space of  $K_\psi$ ,  $\mathcal{N}_{K_\psi}(\Omega)$*

$$|f(x) - s_{\psi, X, f}(x)| \leq P_{X, K_\psi}(x) \|f\|_{\mathcal{N}_{K_\psi}(\Omega)}, \quad x \in \Omega$$

where  $P_{X, K_\psi}$  is the power function.

As in the classical case, the error is bounded in terms of the power function and consequently takes into account both the VSK and the given centers. The native space for VSDKs and error estimate in terms of *global fill-distance* for VSDK interpolation are characterized by S. De Marchi and his collaborators in [28]. Based on the decomposition of the domain  $\Omega$ , we define for  $s \geq 0$  and  $1 \leq p \leq \infty$  the following spaces of piecewise smooth functions on  $\Omega$ :

$$\text{WP}_p^s(\Omega) = \left\{ f : \Omega \rightarrow \mathbb{R} \mid f_{\Omega_i} \in W_p^s(\Omega_i), \quad i \in \{1, \dots, n\} \right\}.$$

Here,  $f_{\Omega_i}$  denotes the restriction of  $f$  to the subdomain  $\Omega_i$  and  $W_p^s(\Omega_i)$  denote the standard Sobolev spaces on  $\Omega_i$ . As norm on  $\text{WP}_p^s(\Omega)$  we set

$$\|f\|_{\text{WP}_p^s(\Omega)}^p = \sum_{i=1}^n \|f_{\Omega_i}\|_{W_p^s(\Omega_i)}^p.$$

We also introduce the *regional fill-distance*  $h_i$  on the subset  $\Omega_i$ :

$$h_i = \sup_{x \in \Omega_i} \inf_{x_i \in \mathcal{X} \cap \Omega_i} \|x - x_i\|_2,$$

and the *global fill-distance*

$$h = \max_{i \in \{1, \dots, n\}} h_i.$$

The piecewise Sobolev space  $\text{WP}_p^s(\Omega)$  and the corresponding norm strongly depend on the decomposition of the domain  $\Omega$ . However, for any decomposition of  $\Omega$  in Assumption 5.1, the standard Sobolev space  $W_p^s(\Omega)$  is contained in  $\text{WP}_p^s(\Omega)$  and the native space associated with  $K_\psi$  is, under the assumptions of Theorem 5.2, the piecewise Sobolev space  $\text{WP}_2^s(\Omega)$ . In fact we can prove [28]:

**Theorem 5.2.** *Let Assumption 5.1 hold true, and assume that the continuous strictly positive definite kernel  $K : \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  based on the radial basis function  $\phi$  has the Fourier decay*

$$\widehat{\phi(\|\cdot\|)}(\omega) \sim (1 + \|\omega\|_2^2)^{-s-\frac{1}{2}}, \quad s > \frac{d-1}{2}. \tag{14}$$

Then, for the discontinuous kernel  $K_\psi$ , we have

$$\mathcal{N}_{K_\psi}(\Omega) = \text{WP}_2^s(\Omega),$$

with the norms of the two Hilbert spaces being equivalent.

Sobolev–type error estimates, based on the global fill–distance  $h$ , are given in the following theorem.

**Theorem 5.3.** *Let Assumption 5.1 be satisfied. Further, let  $s > 0$ ,  $1 \leq q \leq \infty$  and  $m \in \mathbb{N}_0$  such that  $[s] > m + \frac{d}{2}$ . Additionally, suppose that the RBF  $\phi$  satisfies the Fourier decay (14). Then, for  $f \in \text{WP}_2^s(\Omega)$ , and for all  $h \leq h_0$ , we have that*

$$\|f - s_{\psi, X, f}\|_{\text{WP}_q^m(\Omega)} \leq Ch^{s-m-d(1/2-1/q)_+} \|f\|_{\text{WP}_2^s(\Omega)}. \quad (15)$$

The constant  $C > 0$  is independent of  $h$ .

The numerical experiments in [28, Section 4] confirm the theoretical error estimates given in Theorem 5.3 and, as already observed above, show that if the discontinuities of a function are a priori known, the interpolation model based on the discontinuous kernels is significantly better than fixed-scaled interpolation. On the other hand, the outcome of the variably scaled kernel interpolation depends sensitively on the scaling function  $\psi$  i.e. on a *correct knowledge* of the subdomain  $\Omega_i$  boundaries. Therefore, if the VSDK interpolation scheme is applied in a setting in which the edges are not known, a robust edge estimator is needed. In [28, Section 4] some possible choices for such an estimator are discussed. They are essentially based on support vector machine algorithms from machine learning (see e.g. [74] for a survey on the topic) that can be applied directly to scattered data and provide segmentation of the domain  $\Omega$ . Among the possible labeling strategies, one can use thresholds based on function values or RBF coefficients [64]. Of course, other techniques could be used. There is extensive literature for gridded data, and in addition to classical methods such as the Prewitt and Sobel algorithms, as well as the well-known Canny algorithm (see, e.g., [45]), a variety of edge detectors has been proposed [7, 73], and various multiscale transforms can be applied to capture edge information. Typical examples include wavelets and other "let" transforms (see e.g., [55, 62, 65, 76]), but we can find alternative approaches that provide also an approximation of the edge profiles as, for instance, those in [2, 4, 5, 23]. For scattered data, we find a number of papers developing algorithms for both edge detection and its profile approximation (see e.g. [1, 6, 12, 13, 24, 54]). There are also strategies that combine the fault detection with the function reconstruction (see e.g. [3, 8, 46, 53]).

It is worth stressing that it is not enough to have a good approximation of the discontinuity curves, i.e. of  $\partial\Omega_i$  to avoid artifacts, but you must ensure that each data site  $x_j$  is assigned to the subdomain to which it belongs as done in [50], and more recently in [19].

## 6 Concluding remarks

In this article, the properties and applications of VSKs have been reviewed. The scaling function plays a decisive role both in the construction of the kernel and in its performance. Scaling functions improving stability of the interpolation process have been designed and others have been studied to improve the quality of the reconstruction. VSKDs are particularly effective in the interpolation of scattered data sampled from discontinuous functions, and have been successfully applied in magnetic particle imaging which is a recent non-invasive tomographic technique that detects super-paramagnetic nano-particle tracers and finds applications in diagnostic imaging and material science (see [26, 28] and reference therein). It is worth mentioning that in this area the *fake node* approach [27, 32] is very effective too. Finally, very recently, VSKs have been interpreted in the framework of persistent homology to deal with persistence diagrams in supervised learning approaches [30]. Variably scaled persistence kernels (VSPKs) have been successfully tested in different classification experiments, and the obtained results show that they can improve the performance and efficiency of existing standard kernels.

Although a lot has been done, there is still room for further applications and investigations on scaling function design that are the core of VSKs.

**Acknowledgements:** This research has been accomplished within the Research ITalian network on Approximation (RITA) and UMI-TAA. The author is a member of the INdAM research group GNCS.

## References

- [1] G. Allasia, R. Besenghi, and R. Cavoretto. Adaptive detection and approximation of unknown surface discontinuities from scattered data. *Simulation Modelling Practice and Theory*, 17(6):1059–1070, 2009.
- [2] S. Amat, D. Levin, J. Ruiz-Alvarez, and D. F. Yanez. Global and explicit approximation of piecewise-smooth two-dimensional functions from cell-average data. *IMA Journal of Numerical Analysis*, August 2022.
- [3] A. Amir and D. Levin. High order approximation to non-smooth multivariate functions. *Computer Aided Geometric Design*, 63:31–65, 2018.
- [4] F. Arandiga, A. Cohen, R. Donat, N. Dyn, and B. Matei. Approximation of piecewise smooth functions and images by edge-adapted (ENO-EA) nonlinear multiresolution techniques. *Applied and Computational Harmonic Analysis*, 24(2):225–250, 2008.
- [5] F. Arandiga, A. Cohen, R. Donat, and B. Matei. Edge detection insensitive to changes of illumination in the image. *Image and Vision Computing*, 28(4):553–552, 2010.
- [6] R. Archibald, A. Gelb, and J. Yoon. Polynomial fitting for edge detection in irregularly sampled signals and images. *SIAM Journal on Numerical Analysis*, 43(1):259–279, 2005.
- [7] S. Bhardwaj and A. Mittal. A survey on various edge detector techniques. *Procedia Technology*, 4:220–226, 2012.
- [8] M. Bozzini, L. Lenarduzzi, and M. Rossini. Non-regular surface approximation. In *Mathematical methods for curves and surfaces*, volume 8177 of *Lecture Notes in Computer Science*. Springer, Heidelberg, 68–87, 2014.
- [9] M. Bozzini, L. Lenarduzzi, M. Rossini, and R. Schaback. Interpolation by basis functions of different scales and shapes. *Calcolo*, 41(2):77–87, 2004.

- [10] M. Bozzini, L. Lenarduzzi, M. Rossini, and R. Schaback. Interpolation with variably scaled kernels. *IMA Journal of Numerical Analysis*, 35(1):199–219, 2015.
- [11] M. Bozzini, L. Lenarduzzi, and R. Schaback. Adaptive interpolation by scaled multiquadrics. *Advances in Computational Mathematics*, 16(4):375–387, 2002.
- [12] M. Bozzini and M. Rossini. The detection and recovery of discontinuity curves from scattered data. *Journal of Computational and Applied Mathematics*, 240:148–162, 2013.
- [13] C. Bracco, O. Davydov, C. Giannelli, and A. Sestini. Fault and gradient fault detection and reconstruction from scattered data. *Computer Aided Geometric Design*, 75:101786, 2019.
- [14] R. Campagna, S. Cuomo, S. De Marchi, E. Perracchione, and G. Severino. A stable meshfree PDE solver for source-type flows in porous media. *Applied Numerical Mathematics*, 149:30–42, 2020.
- [15] R. Campagna and E. Perracchione. Data-driven extrapolation via feature augmentation based on variably scaled thin plate splines. *Journal of Scientific Computing*, 88(1):15, 2021.
- [16] R. Campagna and E. Perracchione. Feature augmentation for numerical inversion of multi-exponential decay curves. *AIP Conference Proceedings*, 2425:050004, 2022.
- [17] C. Campi, F. Marchetti, and E. Perracchione. Learning via variably scaled kernels. *Advances in Computational Mathematics*, 47(4):51, 2021.
- [18] F. Carnevali. Interpolazione con basi radiali a scala variabile, 2014. Master Thesis in Mathematics (supervisor: M. Rossini).
- [19] R. Cavoretto, A. De Rossi, W. Erb, and M. Rossini. Adaptive and accurate approximation of discontinuous functions with variably scaled kernels, 2022. Preprint.
- [20] R. Cavoretto, A. De Rossi, M. S. Mukhametzhanov, and Y. D. Sergeyev. On the search of the shape parameter in radial basis functions using univariate global optimization methods. *Journal of Global Optimization*, 79(2):305–327, 2021.
- [21] R. Cavoretto, A. De Rossi, and E. Perracchione. Optimal selection of local approximants in RBF-PU interpolation. *Journal of Scientific Computing*, 74(1):1–22, 2018.
- [22] S.N. Chiu, L. Ling, and M. McCourt. On variable and random shape Gaussian interpolations. *Applied Mathematics and Computation*, 377:125159, 2020.
- [23] C. Conti, L. Romani, and D. Schenone. Semi-automatic spline fitting of planar curvilinear profiles in digital images using the Hough transform. *Pattern Recognition*, 74:64–76, 2018.
- [24] A. Crampton and J. C. Mason. Detecting and approximating fault lines from randomly scattered data. *Numerical Algorithms*, 39(1-3):115–130, 2005.
- [25] H. Daumé III. Frustratingly easy domain adaptation. *ACL 2007—Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 256–263, 2007.
- [26] S. De Marchi. Mapped polynomials and discontinuous kernels for Runge and Gibbs phenomena. *SEMA SIMAI Springer Series*, 29:3–43, 2022.
- [27] S. De Marchi, W. Erb, E. Francomano, F. Marchetti, E. Perracchione, and D. Poggiali. Fake nodes approximation for magnetic particle imaging. *20th IEEE Mediterranean Electrotechnical Conference, (MELECON)*, 434–438, 2020.
- [28] S. De Marchi, W. Erb, F. Marchetti, E. Perracchione, and M. Rossini. Shape-driven interpolation with discontinuous kernels: Error analysis, edge extraction, and applications in magnetic particle imaging. *SIAM Journal on Scientific Computing*, 42(2):B472–B491, 2020.
- [29] S. De Marchi, A. Idda, and G. Santin. A rescaled method for RBF approximation. In *Approximation theory XV: San Antonio 2016*, volume 201 of *Springer Proc. Math. Stat.*, pages 39–59. Springer, Cham, 2017.
- [30] S. De Marchi, F. Lot, F. Marchetti, and D. Poggiali. Variably scaled persistence kernels (VSPKs) for persistent homology applications. *Journal of Computational Mathematics and Data Science*, 4:100050, 2022.
- [31] S. De Marchi, F. Marchetti, and E. Perracchione. Jumping with variably scaled discontinuous kernels (VSDKs). *BIT Numerical Mathematics*, 60(2):441–463, 2020.
- [32] S. De Marchi, F. Marchetti, E. Perracchione, and D. Poggiali. Multivariate approximation at fake nodes. *Applied Mathematics and Computation*, 391:125628, 2021.
- [33] S. De Marchi, A. Martínez, and E. Perracchione. Fast and stable rational RBF-based partition of unity interpolation. *Journal of Computational and Applied Mathematics*, 349:331–343, 2019.
- [34] S. De Marchi, A. Martinez, E. Perracchione, and M. Rossini. RBF-based partition of unity methods for elliptic PDEs: Adaptivity and stability issues via variably scaled kernels. *Journal of Scientific Computing*, 79(1):321–344, 2019.
- [35] M. Dehghan and V. Mohammadi. The method of variably scaled radial kernels for solving two-dimensional magnetohydrodynamic (MHD) equations using two discretizations: The Crank-Nicolson scheme and the method of lines (MOL). *Computers and Mathematics with Applications*, 70(10):2292–2315, 2015.
- [36] T. A. Driscoll and B. Fornberg. Interpolation in the limit of increasingly flat radial basis functions. *Computers & Mathematics with Applications*, 43(3-5):413–422, 2002.
- [37] G. E. Fasshauer. *Meshfree approximation methods with MATLAB*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.
- [38] G. E. Fasshauer and M. J. McCourt. Stable evaluation of Gaussian radial basis function interpolants. *SIAM Journal on Scientific Computing*, 34(2):737–762, 2012.
- [39] G. E. Fasshauer and M. J. McCourt. *Kernel-based approximation methods using Matlab*. World Scientific Publishing Company, 2015.
- [40] G. E. Fasshauer and J. G. Zhang. On choosing “optimal” shape parameters for RBF approximation. *Numerical Algorithms*, 45(1-4):345–368, 2007.
- [41] B. Fornberg, E. Larsson, and N. Flyer. Stable computations with Gaussian radial basis functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011.

- [42] B. Fornberg and G. Wright. Stable computation of multiquadric interpolants for all values of the shape parameter. *Computers & Mathematics with Applications*, 48(5-6):853–867, 2004.
- [43] B. Fornberg, G. Wright, and E. Larsson. Some observations regarding interpolants in the limit of flat radial basis functions. *Computers & Mathematics with Applications*, 47(1):37–55, 2004.
- [44] B. Fornberg and J. Zuev. The Runge phenomenon and spatially variable shape parameters in RBF interpolation. *Computers & Mathematics with Applications*, 54(3):379–398, 2007.
- [45] R.C. Gonzalez and R. E. Woods. *Digital Image Processing (2nd Ed)*. Prentice Hall, 2002.
- [46] C. Gout, C. Le Guyader, L. Romani, and A.-G. Saint-Guirons. Approximation of surfaces with fault(s) and/or rapidly varying data, using a segmentation process,  $D^m$ -splines and the finite element method. *Numerical Algorithms*, 48(1-3):67–92, 2008.
- [47] E.J. Kansa and R.E. Carlson. Improved accuracy of multiquadric interpolation using variable shape parameters. *Computers & Mathematics with Applications*, 24(12):99–120, 1992.
- [48] E. Larsson and B. Fornberg. Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. *Computers & Mathematics with Applications*, 49(1):103–130, 2005.
- [49] Y.J. Lee, G.J. Yoon, and J. Yoon. Convergence of increasingly flat radial basis interpolants to polynomial interpolants. *SIAM Journal on Mathematical Analysis*, 39(2):537–553, 2007.
- [50] L. Lenarduzzi and R. Schaback. Kernel-based adaptive approximation of functions with discontinuities. *Applied Mathematics and Computation*, 307:113–123, 2017.
- [51] W. Li, L. Duan, D. Xu, and I. W. Tsang. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1134–1148, 2014.
- [52] L. Ling and F. Marchetti. A stochastic extended Rippa’s algorithm for LpOCV. *Applied Mathematics Letters*, 129:107955, 2022.
- [53] M. C. López de Silanes, M. C. Parra, M. Pasadas, and J. J. Torrens. Spline approximation of discontinuous multivariate functions from scattered data. *Journal of Computational and Applied Mathematics*, 131(1-2):281–298, 2001.
- [54] M. C. López de Silanes, M. C. Parra, and J.J. Torrens. Vertical and oblique fault detection in explicit surfaces. *Journal of Computational and Applied Mathematics*, 140(1-2), 2002.
- [55] S. Mallat. *A wavelet tour of signal processing*. Elsevier Academic Press, Amsterdam, 2009.
- [56] F. Marchetti. The extension of Rippa’s algorithm beyond LOOCV. *Applied Mathematics Letters*, 120:107262, 2021.
- [57] B. Muquet, Z. Wang, G. B. Giannakis, M. De Courville, and P. Duhamel. Cyclic prefixing or zero padding for wireless multicarrier transmissions? *IEEE Transactions on Communications*, 50(12):2136–2148, 2002.
- [58] H. Nojavan, S. Abbasbandy, and M. Mohammadi. Local variably scaled Newton basis functions collocation method for solving Burgers’ equation. *Applied Mathematics and Computation*, 330:23–41, 2018.
- [59] M. Pazouki, S.S. Allaei, M. Hossein Pazouki, and D.P.F. Moller. Adaptive learning algorithm for RBF neural networks in kernel spaces. In *Proceedings of the International Joint Conference on Neural Networks*. 4811–4818, 2016.
- [60] E. Perracchione, P. Massa, A.M. Massone, and M. Piana. Visibility interpolation in solar hard X-ray imaging: Application to RHESSI and STIX. *Astrophysical Journal*, 919(2):133, 2021.
- [61] E. Perracchione, A.M. Massone, and M. Piana. Feature augmentation for the inversion of the Fourier transform with limited data. *Inverse Problems*, 37(10):105001, 2021.
- [62] D. D. Y. Po and M. N. Do. Directional multiscale modeling of images using the contourlet transform. *IEEE Transactions on Image Processing*, 15(6):1610–1620, 2006.
- [63] S. Rippa. An algorithm for selecting a good value for the parameter  $c$  in radial basis function interpolation. *Advances in Computational Mathematics*, 11(2-3):193–210, 1999.
- [64] L. Romani, M. Rossini, and D. Schenone. Edge detection methods based on RBF interpolation. *Journal of Computational and Applied Mathematics*, 349:532–547, 2019.
- [65] M. Rossini. Detecting discontinuities in two-dimensional signals sampled on a grid. *Journal of Numerical Analysis, Industrial and Applied Mathematics*, 4(3-4):203–215, 2009.
- [66] M. Rossini. Interpolating functions with gradient discontinuities via variably scaled kernels. *Dolomites Research Notes on Approximation*, 11:3–14, 2018.
- [67] R. Schaback. Reconstruction of multivariate functions from scattered data, 1997. Manuscript.
- [68] R. Schaback. Multivariate interpolation by polynomials and radial basis functions. *Constructive Approximation*, 21(3):293–317, 2005.
- [69] R. Schaback. Limit problems for interpolation by analytic radial basis functions. *Journal of Computational and Applied Mathematics*, 212(2):127–149, 2008.
- [70] R. Schaback. Small errors imply large instabilities, March 2022. Preprint.
- [71] R. Schaback and H. Wendland. Approximation by positive definite kernels. In Martin D. Buhmann and Detlef H. Mache, editors, *Advanced Problems in Constructive Approximation*, 203–222, 2003. Birkhäuser Basel.
- [72] M. Scheuerer. An alternative procedure for selecting a good value for the parameter  $c$  in RBF-interpolation. *Advances in Computational Mathematics*, 34(1):105–126, 2011.
- [73] S. Singh and R. Singh. Comparison of various edge detection techniques. In *2015 2nd International Conference on Computing for Sustainable Global Development*, 2015. 393–396.
- [74] A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2004.
- [75] H. Wendland. *Scattered data approximation*. Cambridge University Press, Cambridge, 2005.
- [76] S. Yi, D. Labate, G. R. Easley, and H. Krim. A shearlet approach to edge analysis and detection. *IEEE Transactions on Image Processing*, 18(5):929–941, 2009.