Analytical and Numerical Solutions for Nonlinear Equations
©Available online at https://ansne.du.ac.ir/
Online ISSN: 3060–785X
2024, Volume 9, Issue 1, pp. 34–47
Research article

Enhancing Kernel Ridge Regression Models with Compact Support Wendland Functions

Zahra Behdani[1], Majid Darehmiraki[1,*]

[1] Department of Mathematics and statistics, Behbahan Khatam Alanbia University of Technology, Khouzestan, Iran

* **Corresponding author(s):** darehmiraki@bkatu.ac.ir

Received: 30/11/2024 **Revised:** 27/04/2025 **Accepted:** 12/05/2025 **Published:** 17/05/2025 10.22128/ansne.2025.916.1123

Abstract

Radial Basis Functions (RBFs) have gained significant attention in various machine learning applications, including regression modeling, due to their ability to approximate complex, nonlinear relationships. RBFs offer a flexible approach to capturing intricate dependencies between input features and the target variable, making them particularly useful in high-dimensional and nonparametric settings. This paper investigates the use of a specific class of compactly supported RBFs, known as Wendland functions, within the framework of kernel ridge regression (KRR). We discuss their theoretical advantagessuch as sparsity enforcement and computational efficiency as well as practical challenges, including parameter selection and scalability. A comprehensive overview of RBFs is provided, along with their mathematical formulation and a comparison of different RBF kernels in terms of smoothness and locality. We detail the integration of Wendland functions into KRR models, emphasizing their suitability for problems requiring robustness and interpretability. Through extensive simulation studies, the performance of the proposed approach is evaluated against conventional RBF kernels and other widely used regression techniques. Our results demonstrate that Wendland-based KRR achieves competitive accuracy while offering improved stability in the presence of noise and outliers. Furthermore, real-world case studies illustrate the effectiveness of Wendland functions in handling datasets with high collinearity, where traditional kernels often struggle. The practical implications of our findings are discussed, along with guidelines for implementation and potential extensions to large-scale or sparse data scenarios. This work contributes to the growing body of research on interpretable and efficient kernel methods, providing insights for both theoretical and applied machine learning practitioners.

Keywords: Radial basis function, Wendland functions, Compact support, Regression, Kernel

Mathematics Subject Classification (2020): 62J07, 47B34

1 Introduction

Machine learning (ML) is a robust data-driven technique for identifying patterns in high-dimensional domains via induction. It offers a comprehensive array of tools for data analysis, fitting complex non-linear functions, and dimensional reduction [25]. Machine learning algorithms use induction to predict new data based on a given set of training data. Machine learning methodologies have been formulated within the domains of statistics and computer science and have been used across a vast array of data types, including neurology, image

and text processing, and robotics. Kernel approaches are among the most prevalent models in machine learning. Despite their inherent simplicity, they provide a robust framework that allows for the use of non-linear information while remaining within the confines of convex optimization.

The regression process involves calculating the weighted average of the response variable within a specified neighborhood around a given data point. Kernel regression is a powerful non-parametric technique used for estimating the conditional expectation of a random variable. This paper provides a brief introduction to kernel regression, discussing its underlying principles, advantages, and applications. We explore the concept of kernel functions, the regression process, bandwidth selection, and various extensions of kernel regression methods. Kernel regression is a flexible approach for modeling complex relationships between variables without making strong assumptions about the functional form of the data-generating process. By incorporating a kernel function, kernel regression estimates the local average of the response variable based on nearby data points in the feature space. Kernel functions play a pivotal role in kernel regression by assigning weights to observations based on their distances from the point of interest. Common kernel functions include Gaussian, linear, polynomial, and exponential kernels, each serving different purposes in capturing the underlying patterns in the data. By adjusting the width of the kernel function, known as the bandwidth, the smoothness of the regression curve can be controlled to accommodate different levels of noise in the data.

The ideal theoretical performance of KRR has led to its extensive use for regression purposes. Many different methods, like Nyström regularization [20], distributed KRR [29], localized KRR [17], and boosted KRR [15], have been developed to reduce the computing needs and prevent the saturation [12] of KRR. However, proving that KRR and its variations work well depends on choosing regularization parameters in advance, which is not practical because it's usually hard to get that prior data. However, proving that KRR and its variations are effective depends on choosing regularization parameters beforehand, which is not practical because it's often difficult to obtain that prior information. The authors of [16] concentrated on the challenges of parameter selection in KRR. Because KRR has special spectrum characteristics, it was demonstrated that carefully dividing the range of parameters helps to lessen the difference between two nearby KRR estimates. The authors of [23] introduced a preconditioner-based approach for addressing an issue with kernel ridge regression. Their hybrid methodology effectively decreased the condition number, being both precise and computationally economical, facilitating the processing of extensive datasets with computational complexity proportional to the number of data points.

RBFs have garnered significant attention recently as a sophisticated approach for high-dimensional scattered data approximation, a recognized technique in machine learning, a fundamental component of mesh-free methods, an alternative for developing higher-order methods to solve partial differential equations (PDEs), an innovative strategy for addressing PDEs on surfaces, and a novel solution for mesh repair, among other applications. All these applications are grounded on a singular mathematical principle: high-dimensional approximation/interpolation. Furthermore, we can categorize RBFs into two primary groups: global RBFs and compactly supported RBFs (CSRBFs). Gaussian, multiquadratic, inverse multiquadratic, and polyharmonic splines exemplify global RBF [22]. Conversely, the CSRBF was introduced by [26]. In this context, global RBFs have an infinite range, while CSRBFs are limited to a specific range and are zero outside of it. Regarding approximation, global RBFs exhibit insensitivity to the distribution of data points, but CSRBFs have a heightened sensitivity to such distributions. The resolution of approximation and interpolation using global radial basis functions often results in the solution of a system of linear equations.

Regression modeling is a fundamental technique in machine learning used to predict continuous output variables based on input features. Traditional regression models such as linear regression may struggle to capture nonlinear relationships between variables effectively. Radial Basis Functions provide a powerful alternative by allowing the model to adapt to the data's complex structure without imposing strict assumptions on the underlying relationship. For a training set of input vectors $x_1, \ldots, x_n \in X$, the Gram matrix corresponding to a kernel k is a $n \times n$ matrix K, where the components are defined as $K_{ij} = k(x_i, x_j)$ for $i, j = 1, \ldots, n$. K is symmetric, positive definite, and often dense, lacking any zero entries. A dense Gram matrix may be very ill-conditioned, particularly for extensive data sets, hence presenting substantial computational challenges regarding solution feasibility and stability [28]. The objective of our research is to develop kernels that possess local support and exhibit sparse Gram matrices. The presence of several zero components in Gram matrices may significantly reduce computer storage demands and the quantity of floating-point operations required for calculation. Wendland functions are a kind of compactly supported, piecewise polynomial RBFs that are useful because they employ the minimal degree polynomial for a given smoothness and their compact support results in sparse Gram matrix. Due to the unique features of compact support Wendland functions, our main motivation in this article is to use these functions as kernel functions in KRR. To the best of the author's knowledge, this is the very first work that describes the estimation of KRR models using compact support Wendland kernels.

This paper is organized as follows: Sections 2 and 3 present a discussion on radial basis functions and regression. Section 4 covers the fundamentals of kernel ridge regression. In Section 5, we introduce compact radial basis functions as a kernel for kernel ridge regression. Section 6 details the results of our simulations. Finally, Section 7 summarizes the findings and offers recommendations for future research.

2 Radial Basis Functions: An Overview

Radial Basis Functions are mathematical functions that depend on the distance between a data point and a center point or prototype. The key idea behind RBFs is to transform high-dimensional input features into a higher-dimensional space where the relationships between the data points are easier to capture. The most common form of RBF is the Gaussian RBF, which assigns weights to data points based on their similarity to the center points.

The RBF kernels are solely determined by the spatial separation between the test and trial locations. It results in straightforward implementation in high-dimensional systems. The majority of radial basis functions (RBFs) use a constant parameter, namely the shape parameter, which might impact the stability and precision of the solution. When the form parameter of the multiquadratic (MQ) function [8] is varied, the interpolation matrix becomes non-singular, hence establishing a well-defined interpolation issue. Moreover, this particular RBF kernel exhibits higher accuracy compared to the cubic and thin plate spline kernels. The PHS radial function exhibits conditional positive definiteness (CPD) and has the characteristic of scalability (see references [27, 29]). In general, the approach may be executed in two specific ways, namely the global and local strong forms.

Presented below is a concise overview of local RBFs, which will be used in the next sections. An initial set of dispersed points $P = \{u_1, \ldots, u_N\}$ is selected in the domain $D \subseteq \mathbb{R}^d$. The radial function $\phi(\mathbf{u}) = \phi(\|u\|)$ is universally supported on the ring $\mathbb{R}^d$. An estimate of the function $\Psi(u)$ at every point x in domain D is derived by using the function ϕ built on a collection of nodals P as described in reference [29]

$$\Psi(u) \approx \mathscr{L}\Psi(u) = \sum_{i=1}^{N} c_i \phi(\|u - u_i\|). \tag{1}$$

Through the execution of the interpolation conditions, the coefficients c_j expansion (1) are obtained:

$$\mathscr{L}\Psi(u_j) = v_j,$$

whose matrix form is

$$\mathbf{BC} = \mathbf{U},$$

where

$$U = \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ . \\ v_N \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ 1 & \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 1 & \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{bmatrix}, \qquad C = \begin{bmatrix} c_0 \\ c_1 \\ . \\ . \\ . \\ c_N \end{bmatrix}.$$

An important observation is that if a radial basis function ϕ is strictly positive definite, then the corresponding interpolation matrix B is also positive definite and therefore non-singular. Thus, interpolation issue (1) will be well-posed, meaning it has a precisely defined solution. Gaussians and inverse multiquadratic distributions are radial basis functions that are strictly positive definite on any ring $\mathbb{R}$ [11]. The related interpolation matrix in expansion (1) is guaranteed to be non-singular. Moreover, the interpolation matrix for multiquadratic is non-singular, but not possessing stringent positive definiteness [29]. Some of the RBFs are listed in Table 1.

3 Regression and Ridge Regression:An Overview

Regression is a statistical technique used for modeling and analyzing the relationships between variables. The primary goal of regression is to predict the value of a dependent variable (the target) based on one or more independent variables (the predictors). Simple linear regression, which involves one independent variable and one dependent variable, is one of the most common types of regression.

Table 1. $r = \|u\| and \|.\|$ represents the Euclidean distance.

Name	RBF	Shape parameter
Multiquadric (MQ)	$\phi(r) = \sqrt{r^2 + c^2}$	$c > 0$
Inverse Multiquadric (IMQ)	$\phi(r) = \frac{1}{\sqrt{r^2+c^2}}$	$c > 0$
Gaussian (GA)	$\phi(r) = exp(-(cr)^2)$	$c > 0$
Thin-plate splines	$(-1)^{1+c2} r^c log\, r$	$c \in 2\mathscr{N}$
Wendland function	$(1-r)_+^4(1+4r)$	$c > 0$

The linear regression model is defined as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{2}$$

where

- Y is the dependent variable.
- X is the independent variable.
- β_0 is the intercept (the value of Y when $X = 0$).
- β_1 is the slope of the line (the change in Y for a one-unit change in X).
- ε is the random error term.

The parameters β_0 and β_1 are estimated using the method of least squares, which aims to minimize the sum of the squared differences between the observed values and the values predicted by the model. The least squares estimates can be calculated using the following formulas:

1. **Estimate of the Slope ($\hat{\beta}_1$)**:

$$\hat{\beta}_1 = \frac{n\sum(XY) - \sum X \sum Y}{n\sum(X^2) - (\sum X)^2}. \tag{3}$$

2. **Estimate of the Intercept ($\hat{\beta}_0$)**:

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}, \tag{4}$$

where $\bar{Y}$ and $\bar{X}$ are the means of the dependent and independent variables, respectively.

In multiple linear regression, we extend the concept to include multiple independent variables. The model can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon, \tag{5}$$

where

- Y is the dependent variable.
- $X_1, X_2, \ldots, X_p$ are the independent variables.
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients for the independent variables.
- ε is the random error term.

The parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are estimated using the least squares method as well. The solution can be expressed in matrix form:

$$\hat{\beta} = (X^T X)^{-1} X^T y, \tag{6}$$

where

- $\hat{\beta}$ is the vector of estimated coefficients.
- X is the matrix of independent variables (including a column of ones for the intercept).
- y is the vector of observed values of the dependent variable.

This matrix approach allows for efficient computation, especially when dealing with multiple predictors.Ridge regression is a type of linear regression designed to address the problem of collinearity among independent variables. In ridge regression, a penalty term is added to the cost function, which helps reduce the magnitude of the regression coefficients. This approach makes the model less sensitive to the training data and improves its generalizability.

The ridge regression formula is expressed as:

$$\hat{\beta} = \arg\min_{\beta} \left(\sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right), \tag{7}$$

where

- $\hat{\beta}$ are the estimated coefficients.
- y_i are the actual values of the dependent variable.
- X_i is the matrix of independent variables.
- λ is the penalty parameter (ridge penalty) that determines how much to shrink the coefficients. Larger values of λ lead to greater shrinkage of the coefficients.

In ridge regression, the estimation of parameters is modified by the addition of the penalty term. The solution can be expressed in a matrix form:

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y, \tag{8}$$

where

- I is the identity matrix.
- This formula indicates that the ridge regression estimates are influenced by both the data matrix X and the penalty term λ, which helps to stabilize the estimates when multicollinearity is present.

Regression and ridge regression are useful methods for analyzing data and making predictions. Simple linear regression helps show how variables relate to each other, while ridge regression is made to deal with problems that come from having too many similar predictors and to make the model more reliable. Using these methods can provide important insights from the data. Estimating parameters in both methods is key to understanding the data relationships and making precise predictions. In multiple linear regression, being able to work with several predictors allows for a better analysis of complicated datasets.

4 Kernel Ridge Regression

A kernel on a domain U, an arbitrary set with no assumed structure, is a symmetric positive semi-definite function of two variables, denoted as $\mathbf{K} : U \times U \to \mathscr{R}$, such that [30]

- for all $u_1, u_2 \in \mathscr{R}$ we have $\mathbf{K}(u_1, u_2) = \mathbf{K}(u_2, u_1)$ and
- For every positive integer N, and for any elements $u_1, u_2, \ldots, u_N \in U$, as well as any real numbers $c_1, c_2, \ldots, c_N \in \mathscr{R}$, it holds that $\sum_{i,j \in N} c_i c_j \mathbf{K}(u_i, u_j) \geq 0$.

Perhaps the most basic kernelized method is ridge regression. Our objective is to develop a linear function that represents the dependent relationships between the continuous covariates u_i and the response variables v_i. For that purpose, the conventional approach is to minimize the quadratic cost. Nevertheless, if we choose to operate in feature space, the replacement of u_i with $\phi(u_i)$ carries a distinct risk of overfitting. Therefore, regularization is essential to prevent overfitting. We will revisit this subject in subsequent educational sessions due to its great significance

$$C(w) = \frac{1}{2}\sum_i (v_i - w^T u_i)^2. \tag{9}$$

An efficient and straightforward method to regularize is to impose a penalty on the norm of w. We alternatively refer to this phenomenon as "weight decay." It is still to be determined how to choose λ. The most often- used strategies are cross-validation and leave-one-out estimations. Consequently, the total cost function is given by,

$$C = \frac{1}{2}\sum_i (v_i - w^T u_i)^2 + \frac{1}{2}\lambda \|w\|^2. \tag{10}$$

The solution to this minimization problem is unique [6]. It need minimization. Calculating derivatives and setting them equal to zero results in,

$$\sum_i (v_i - w^T u_i)u_i = \lambda w \Longrightarrow w = (\lambda \mathbf{I} + \sum_i u_i u_i^T)^{-1}(\sum_j v_j u_j). \tag{11}$$

Departing from the optimization of the cost function mentioned before, we may include Lagrange multipliers into the issue. Such an approach will result in the derivation following a very similar pattern to that of the SVM scenario. We propose the introduction of additional variables, $\xi = y_i - w^T \phi_i$, and rephrase the aim as the restricted quantum probability formula.

$$\begin{aligned} &\textit{minimize} \quad \sum_i \xi_i^2, \\ &S.t. \\ &\xi_i = y_i - w^T \phi_i, \quad \forall i \\ &\|w\| \leq B \end{aligned} \tag{12}$$

5 Wendland Functions for KRR

Typically, kernel matrices are inherently dense. Although certain kernels, like the square-exponential kernel $K(t) = e^{-t^2}$, diminish rapidly, they never attain a value of zero. To attain sparsity, we may seek kernel functions with compact support; specifically, those that are zero outside the interval $[-c, c]$, which we will generally consider as $[-1, 1]$. Although any function may be modified to possess compact support by truncation (i.e., defining $\tilde{f}(t)$ as $f(t)$ for $t \in [-1, 1]$ and zero elsewhere), the outcome often does not provide proper kernels [3]. The Wendland functions [26] and the function described in [18] are two examples of compactly supported positive definite kernels that have been found in the literature.

Several types of RBF kernels can be used in regression models, with the Gaussian kernel being the most common. Other popular choices include the Multiquadratic, Inverse Multiquadratic, and Thin Plate Splines kernels, each with its characteristics and applications. The choice of the RBF kernel can significantly impact the model's performance and generalization ability.

To incorporate RBFs into regression models, the model typically combines the RBF outputs with a linear combination of weights. This technique allows the model to capture both linear and nonlinear relationships between the input features and the target variable. Training such models often involves optimizing the weights and tuning the hyperparameters, such as the number of RBF centers and the bandwidth parameter.Let $u \in U = \{u_i, i = 1, \ldots, N\}$ be an independent variable, and $v \in V = \{v_i, i = 1, \ldots, N\}$ be a dependent variable. The regression model consists in finding a function Psuch that $P(u_i) = v_i$, $i = 1, 2 \ldots, N$. In order to do this, we consider $P \in span\{\phi(., u_i), u_i \in U\}$ where $\Omega \times \Omega \Rightarrow \mathscr{R}$ is a strictly positive definite and symmetric kernel. This presumption is not limiting; in fact, there exists an associated normalized positive definite kernel for each conditionally positive definite kernel. This selection results in the interpolate taking on the form

$$P(x) = \sum_{k=1}^{N} \alpha_k \Phi(u, u_k), \quad u \in \Omega, \tag{13}$$

We find the coefficients $\alpha = (\alpha_1, \ldots, \alpha_N)^T$ in equation (13) by solving the equation $\mathbf{A}\alpha = \mathbf{Y}$, where the matrix $\mathbf{A} \in \mathscr{R}^{N\times N}$ has entries $(\mathbf{A})_{ik} = \Phi(u_i, u_k)$ for $i, k = 1, \ldots, N$; and $V = (v_1, \ldots, v_N)^T$. The solution is unique because the kernel is both positive definite and symmetric. The uniqueness of the solution is guaranteed by the stringent positive definiteness and symmetry of the kernel. Specifically, we consider RBFs and assume the existence of a function ϕ in the interval $[0;\infty)$ and a shape parameter ε greater than 0 such that for each $u, v \in \Omega$, the equation $\Phi(u,v) = \phi_\varepsilon(\|u-v\|_2) := \phi(r)$ holds.

Wendland functions were first introduced in [26], with further instances shown in [21]. We shall designate the Wendland functions from [26] as the original Wendland functions and those from [21] as the missing Wendland functions. The first exhibition of the Wendland functions took place in 1947 [10]:

$$\phi l,k(r) = pk(r)(1-r)_+^{l+k}, \tag{14}$$

where $(x)_+ = max(x,0)$ and is a polynomial of degree k. The index l indicates that these functions are positive definite in $\mathscr{R}^l$. The index k denotes the quantity of continuous derivatives. The Wendland functions are defined as the polynomial of the lowest degree that fulfills these criteria [19].Schaback [21] demonstrated that the missing Wendland functions have the form

$$\phi l,k+\frac{1}{2}(r) = pl,k\frac{r^2}{2}L(r) + q_{l,k}\frac{r^2}{2}S(r), \tag{15}$$

where

$$L(r) = log(\frac{r}{1+\sqrt{1-r^2}}), \quad \text{and} \quad S(r) = \sqrt{1-r^2},$$

and $p_{l,k}$ and $q_{l,k}$ are polynomials. The missing Wendland functions, although they continue to be supported on the interval $[0,1]$, now contain logarithmic and square-root multipliers with polynomial components. The result is a significant difference between the original Wendland functions and the missing Wendland functions [9]. We provide explicit formulas for the missing Wendland functions for $d = 2$ and $k = \frac{1}{2}, \frac{3}{2}$ and $\frac{5}{2}$ in Table 2.

Table 2. The missing Wendland functions $\phi_{l,k}(r)$ for $d = 2$ and $k = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$.

k	Missing Wendland function
$\frac{1}{2}$	$\phi_{2,\frac{1}{2}}(r) = 3r^2L(r) + (2r^2+1)S(r)$
$\frac{3}{2}$	$\phi_{3,\frac{3}{2}}(r) = -15r^4(6+r^2)L(r) - (18r^428r^2-4)S(r)$
$\frac{5}{2}$	$\phi_{4,\frac{5}{2}}(r) = (945r^8+252r^6)L(r) + (256r^8+2639r^6+690r^4-136r^2+16)S(r)$

The CSRBFs of Wendland, which have continuity of C^0, C^2, C^4, and C^6 correspondingly, are shown in equation (16).

$$\begin{aligned}
\phi_{3,0}(r) &= (1-r)_+^2, \\
\phi_{3,1}(r) &= (1-r)_+^4(4r+1), \\
\phi_{3,2}(r) &= (1-r)_+^6(35r^2+18r+3), \\
\phi_{3,3}(r) &= (1-r)_+^8(32r^3+25r^2+8r+1).
\end{aligned} \tag{16}$$

Compactly supported RBF kernels provide sparse Gram matrices, offering significant benefits for certain applications involving large data sets. Sparse matrices provide significant savings in storage and computational memory by omitting several zero entries. Secondly, sparse matrix methods need much less computational time than conventional algorithms by circumventing arithmetic operations on zero members.

6 Numerical Simulation

We conduct numerical analyses to validate the theoretical findings and evaluate the effectiveness of the kernel ridge regression estimator with limited sample sizes. Additionally, we present several case studies that illustrate the Wendland functions in regression models across various domains. These case studies highlight how Wendland functions can enhance the predictive accuracy of models, particularly when addressing complex and nonlinear relationships within the data.

Mean Squared Error (MSE) is a widely used metric for evaluating the accuracy of a predictive model. It measures the average of the squares of the errors, that is, the differences between the predicted values and the actual values. The formula for MSE is given by:

$$MSE(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}(u_i) - f(u_i))^2, \quad i = 1,2,\ldots,n, \tag{17}$$

where $f(u_i)$ represents the actual values, $\hat{f}(u_i)$ represents the predicted values, and n is the number of observations.

MSE is particularly useful because it penalizes larger errors more severely than smaller ones, making it sensitive to outliers. This property helps in identifying models that consistently produce accurate predictions.

In this paper, we utilize MSE as a criterion for evaluating and comparing the performance of different models. By analyzing the MSE values, we can determine which model provides the best fit to the data, thereby allowing for a more informed selection of the most effective modeling approach.

Example 1. *In this example, we explore the application of different modeling techniques to analyze a dataset. The primary objective is to evaluate how well these models can predict outcomes based on the available data. We specifically compare two methods: Least Squares Regression and the Wendland function. We use the least squares method, a traditional approach, to minimize the sum of squares of the differences between observed and predicted values. Despite its widespread use, it may not always effectively capture complex relationships. On the other hand, the Wendland function is a type of radial basis function that offers greater flexibility in modeling non-linear relationships. By examining the mean squared error (MSE) of both methods, we aim to determine which model provides a better fit for the data. This comparison not only highlights the importance of selecting the appropriate modeling technique but also demonstrates how different approaches can yield varying levels of accuracy in predictions.Initially, we generated two random variables, x_1 and x_2, from a uniform distribution over the interval* $[-3,3]$. *Next, we computed the variable y using Equation*

$$y = \sin(2*(x_1+x_2)) + 0.5x_1x_2 + \varepsilon,$$

that incorporates a normal error term. This error term is characterized by a mean of zero and a standard deviation of 0.2*, reflecting a small degree of variability around the expected values. The inclusion of this normal error is crucial for simulating realistic data, as it mimics the inherent uncertainties present in real-world measurements. Figure 1 provides a visual representation of the generated data and the simulated model from two different perspectives. This figure is particularly useful for gaining a more profound understanding of the models behavior and the distribution of data in a two-dimensional space. We can observe the distribution of data points over the model's surface in the figure on the left. This initial perspective allows us to see how the actual data (generated from a uniform distribution) aligns with the simulated model. The model's surface clearly positions the data points, demonstrating the simulation's accuracy.The observation changes in the figure on the right help us better understand the depth and overall shape of the model. This perspective may reveal additional details about how the models surface varies and the distances between the data points and the model. Examining these two angles enables us to identify the strengths and weaknesses of the model and gain a clearer understanding of its predictive accuracy.This figure not only aids in comprehending the data and the model but also allows us to observe the impact of random errors (introduced through the normal error with a mean of zero and a standard deviation of* 0.2*) on the simulation results. This information can assist us in optimizing the model and improving the accuracy of predictions. In our analysis, we compared the mean squared error (MSE) of two modeling methods. In the Least Squares Regression, the MSE is approximately* 2*, indicating a relatively poor fit between the predicted values and the actual data points. In the Wendland function, the MSE is significantly lower at* 0.4*, suggesting a much better fit and indicating that the Wendland function captures the underlying data structure more accurately. This comparison highlights the Wendland function as a more effective modeling choice for this dataset.*

General information about the dataset, parameter settings, and evaluation criteria used in the data used in Example 1 at a glance

- ***Dataset****: Two variables* x_1, x_2 *uniformly sampled from* $[-3,3]$*. Target* $y = \sin(2(x_1+x_2)) + 0.5x_1x_2 + \varepsilon$*,* $\varepsilon \sim \mathcal{N}(0, 0.2^2)$*.*
- ***Parameters****: Wendland kernel* $\phi_{3,1}(r) = (1-r)_+^4(4r+1)$ *with* $c = 1$*;* λ *optimized via 5-fold CV.*
- ***Metrics****: MSE for OLS (2.3) vs. Wendland-KRR (0.4).*

To evaluate the accuracy of our model, we calculated the amount of error using two distinct methods: the least squares approach and Wendlands function. The results indicated that the error associated with the least squares method was 2.3, while the error calculated using

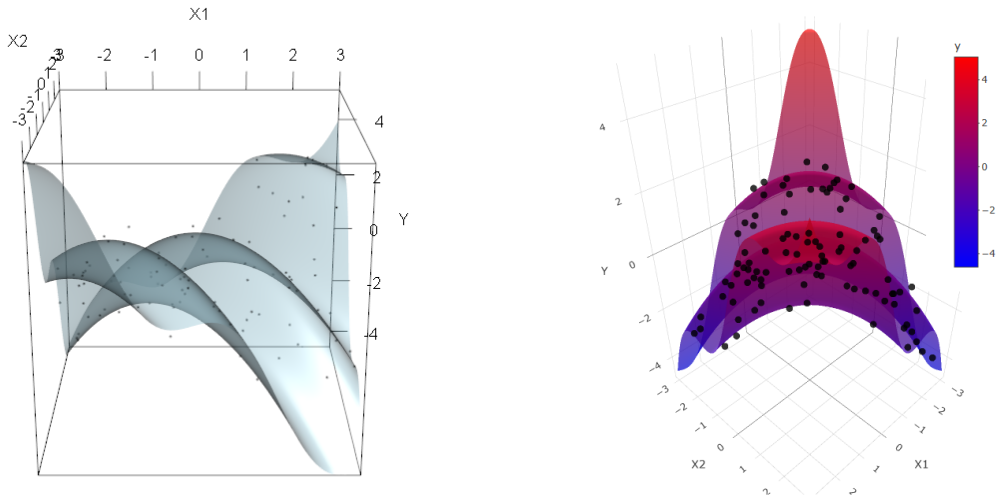


Figure 1. Two Perspectives of Data and Simulated Model in Example 1

Wendlands function was 5.4. These findings highlight the differences in error estimation between the two methods, suggesting that the least squares approach may provide a more accurate fit for our simulated data in this context.

Example 2. *In the second example, we employed the Monte Carlo method to assess the accuracy of the studied approach. We generated 100 random samples with two independent variables from a uniform distribution in 100 replications. The dependent variable was then produced based on models 18 to 22 For $d = 3$, we examine three instances of smooth regression functions exhibiting varying complexity [2].*

$$f(u) = -4 + 3u1u2 + 0.5u3^2 4u1u2u3, \tag{18}$$

$$f(u) = sin(2\pi(u_1 + u_2)) + u_3^3 + 0.5u_1^2 u_2, \tag{19}$$

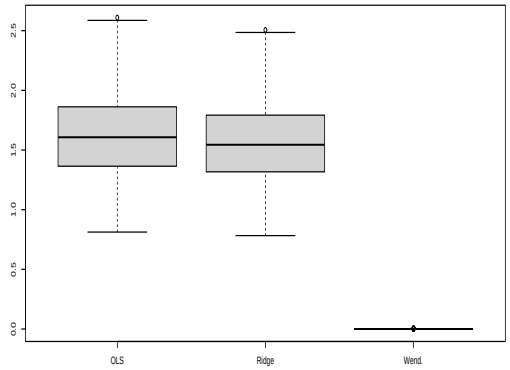
$$f(u) = sin^2(2\pi u_1) + cos(2\pi u_2^2) + 3u_2 u_3^3, \tag{20}$$

$$f(u) = e^{u_1^2 u_2} sin(2\pi u_2 u_3^2). \tag{21}$$

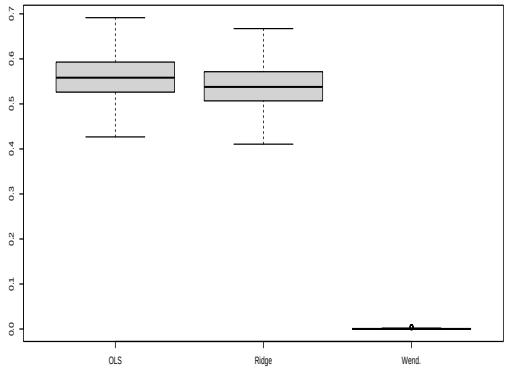
For $d = 5$

$$f(x) = e^{x_5} sin(2\pi(x_1 + x_2))cos(2\pi x_3^2 x_4)x_5^3. \tag{22}$$

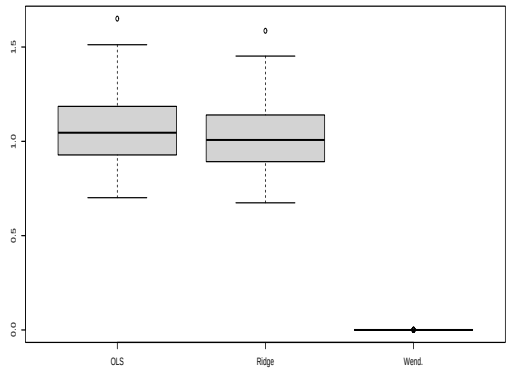
Subsequently, we estimated the parameters and calculated the model accuracy using three methods: Least Squares Regression, Ridge Regression, and the Wendland function. The results of this analysis are presented in Table 3.
To assess the accuracy of the three modeling methodsLeast Squares Regression, Ridge Regression, and the Wendland functionbox plots of the error rates from 100 experiments were created in Figure2. The results clearly indicate that the Wendland function exhibits higher accuracy compared to the other methods. This conclusion is evident from the box plots, which show that the error distribution for the Wendland method is significantly lower. The reduced error range and tighter clustering around the lower error values highlight its effectiveness in predicting outcomes more accurately than the other approaches.



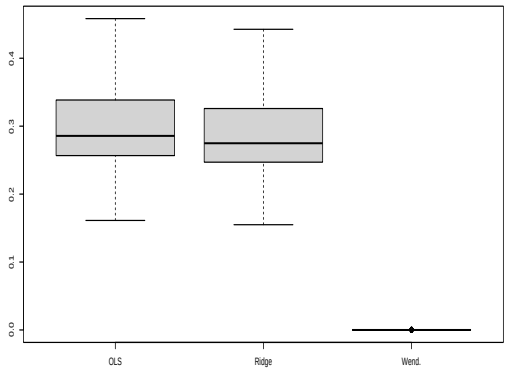
(a) The box plot in 1000 replications for MSE in Model 18



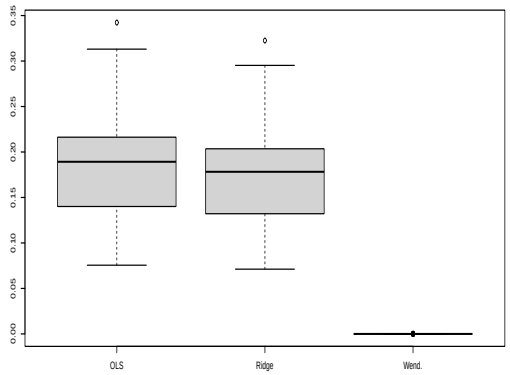
(b) The box plot in 1000 replications for MSE in Model 19



(c) The box plot in 1000 replications for MSE in Model 20



(d) The box plot in 1000 replications for MSE in Model 21



(e) The box plot in 1000 replications for MSE in Model 22

Figure 2. Box plot for comparing three methods of estimation of regression parameters for different models.

General information about the dataset, parameter settings, and evaluation criteria used in the data used in Example 2 at a glance

Table 3. Ridge fitting error and least square regression also using kernel method with Wendland function for simulated data.

Model	OLS	Ridge regression	Wendland
18	1.667207	1.604626	0.0005825475
19	0.5579917	0.5378446	0.0006072424
20	1.058185	1.017672	0.0001295131
21	0.2966672	0.2857586	0.00003615231
22	0.1853804	0.1748771	$1.292559e-25$

- ***Datasets***: *100 Monte Carlo replications for Models 6.26.6 (Equations 6.26.6) with* $d = 3, 5$.
- ***Parameters***: *Wendland* $\phi_{3,2}(r)$; λ *via 10-fold CV.*
- ***Results***: *Wendland-KRR outperformed OLS and Ridge (Table 3, MSE* $\approx 10^{-25}$*).*

Example 3. *In this example, we use the seatpos data available in the R software. Everyone likes to find their perfect driving posture by adjusting the seat. It would be useful for car designers to know how various-sized and aged drivers would place the seat. The HuMoSim lab at Michigan State University gathered information from 38 drivers. Here are the variables that are part of the dataset: There is a center of gravity for the hip, leg, age, weight, shoes, and sitting.The independent variables in these data have a strong correlation, or, in other words, there is collinearity. By examining the eigenvalues, it is clear that the ratio of the largest to the smallest eigenvalue is equal to 500625, which indicates strong collinearity in the data. We fitted ridge regression and ordinary least squares regression for these data and calculated the error values, which equal* 1164.453 *and* 1085.836*, respectively. Table 1 contains this information in full. Using the Wendland functions, we performed the Wendland transformation on the data and re-estimated the parameters using the ridge and least squares method (columns 3 and 4 of Table 4). As you can see, the Wendland transformation significantly reduced the amount of error. On the other hand, the error using ridge regression is lower than that of least squares regression, meaning that its accuracy is higher than that of normal least squares regression.*

Table 4. Ridge fit error and least squares regression also using kernel method with Wendland function for seatpos data set.

	OLS	Ridge regression	KOLS	KRR
MSE	1164.453	1085.836	1.317509	$1.699972e-06$

General information about the dataset, parameter settings, and evaluation criteria used in the data used in Example 3 at a glance

- ***Dataset***: `seatpos` *Publicly available in R, the dataset contains 38 observations with 8 variables (e.g., hip center, age, weight). Independent variables exhibit high collinearity (largest-to-smallest eigenvalue ratio = 500,625).*
- ***Transformation***: *Adaptive Wendland* $\phi_{3,2}(r)$ *with adaptive support radius c tuned to minimize collinearity. reduced collinearity.*
- ***Results***: *Wendland-KRR achieved MSE* $= 1.7 \times 10^{-6}$ *(Table 4).*

7 Conclusion

Kernel methods offer a clear and organized way to train learning machines, and the good performance they achieve can be easily supported by statistical learning theory or Bayesian reasoning. RBFs have been extensively researched because of their effective generalization and universal approximation capabilities via the collocation approach. This study outlines a novel technique for creating regression models using a compact support kernel approach. This approach includes many benefits. The technique is systematic and conceptually well-motivated. The learning engine is intentionally designed using the most informative patterns within the data. The reliance on the data is evident,

facilitating the explanation and interpretation of the model, although data cleaning might be used to enhance performance.
In conclusion, RBFs offer a versatile and powerful tool for enhancing regression models' predictive capabilities. By leveraging the flexibility and adaptability of RBFs, researchers and practitioners can build regression models that effectively capture complex relationships in the data. However, challenges such as tuning hyperparameters and selecting appropriate kernel functions remain critical aspects of using RBFs in regression modeling. Future research could focus on developing more efficient algorithms for training RBF-based regression models and exploring their applications in emerging domains. In high-dimensional settings, the proposed method may encounter challenges such as distance concentration (indistinguishable Euclidean distances), increased computational complexity, and sensitivity to parameter tuning (e.g., support radius). These factors could reduce model accuracy and demand heavier computational resources. In future work, we will address these challenges by integrating dimensionality reduction techniques (e.g., PCA), distributed algorithms, and adaptive parameter tuning to enhance the methods scalability and applicability to complex, large-scale datasets.

Authors' Contributions

All authors have the same contribution.

Data Availability

The manuscript has no associated data or the data will not be deposited.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

The authors have diligently addressed ethical concerns, such as informed consent, plagiarism, data fabrication, misconduct, falsification, double publication, redundancy, submission, and other related matters.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

References

[1] M. Antillon, J. Bilcke, A. D. Paltiel, and V. E. Pitzer, Cost-effectiveness analysis of typhoid conjugate vaccines in five endemic low-and middle-income settings. Vaccine, 35(27), 3506–3514, (2017).

[2] K.Y. Bak, and W. Lee, Effect of dimensionality on convergence rates of kernel ridge regression estimator. Journal of Statistical Planning and Inference, 106228, (2024).

[3] J. Barber, Sparse gaussian processes via parametric families of compactly-supported kernels. arXiv preprint arXiv:2006.03673, (2020).

[4] M.D. Buhmann, Radial Basis Functions: Theory and Implementations. Vol. 12. Cambridge University Press, 2003.

[5] C. Campbell, An introduction to kernel methods. Studies in Fuzziness and Soft Computing, 66, 155–192, (2001).

[6] A. Caponnetto, and E.D. Vito, Optimal rates for the regularized least-squares algorithm. Foundations of Computational Mathematics, 7, 331–368, (2007).

[7] J. Chachi, S.M. Taheri, and P. DUrso, Fuzzy regression analysis based on M-estimates. Expert Systems with Applications, 187, 115891, (2022).

[8] J. M. F. Chamayou, and M.E.A. El Tom, On the approximate solution of the delay integral equation of the statistical theory of radiation damage. Computer Physics Communications, 9(3), 131–140, (1975).

[9] A. Chernih, I.H. Sloan, R.S. Womersley, Wendland functions with increasing smoothness converge to a Gaussian. Advances in Computational Mathematics, 40, 185–200, (2014).

[10] A. Chernih, Multiscale Wendland radial basis functions and applications to solving partial differential equations. PhD Thesis, University of New South Wales (2013).

[11] G.E. Fasshauer, Meshfree methods. Handbook of theoretical and computational nanotechnology, 27, 33–97, (2005).

[12] L.L. Gerfo, L. Rosasco, F. Odone, E.D. Vito, and A. Verri, Spectral algorithms for supervised learning. Neural Computation, 20(7), 1873–1897, (2008).

[13] A. Kouibia Krichi, P. González Rodelas, M. Pasadas Fernández, B. Mustafa, H.O. Yakhlef, and L. Omri, Approximation of Bivariate Functions by Generalized Wendland Radial Basis Functions, (2024).

[14] K.J. Liew, K.H. Tee, A. Ramli, and W.E. Ong, Integrating clustering method in compactly supported radial basis function for surface approximation. IAENG International Journal of Computer Science, 46(1), (2019).

[15] S.B. Lin, Y. Lei, and D.X. Zhou, Boosted kernel ridge regression: Optimal learning rates and early stopping. Journal of Machine Learning Research, 20(46), 1–36, (2019).

[16] S.B. Lin, Adaptive Parameter Selection for Kernel Ridge Regression. Applied and Computational Harmonic Analysis, 101671, (2024).

[17] M. Meister, and I. Steinwart, Optimal learning rates for localized SVMs. Journal of Machine Learning Research, 17(194), 1–44, (2016).

[18] A. Melkumyan, and F.T. Ramos, A sparse covariance function for exact Gaussian process inference in large datasets. In Twenty-first international joint conference on artificial intelligence, (2009).

[19] R.B. Platte, C^{∞} Compactly Supported and Positive Definite Radial Kernels. SIAM Journal on Scientific Computing, 37(4), A1934-A1956, (2015).

[20] A. Rudi, R. Camoriano, and L. Rosasco, Less is more: Nyström computational regularization. Advances in neural information processing systems, 28, (2015).

[21] R. Schaback, The missing Wendland functions. Advances in Computational Mathematics, 34(1), 67–81, (2011).

[22] W. Schempp, and K. Zeller, Constructive Theory of Functions of Several Variables [electronic resource]: Proceedings of a Conference Held at Oberwolfach April 25May 1, (1976).

[23] G. Shabat, E. Choshen, and N. Carmel, Fast and accurate Gaussian kernel ridge regression using matrix decompositions for preconditioning. SIAM Journal on Matrix Analysis and Applications, 42(3), 1073–1095, (2021).

[24] Z. Shengxin, Compactly supported radial basis functions: how and why? OCCAM Preprint Number 12/57, (2012).

[25] K. Vu, J.C. Snyder, L. Li, M. Rupp, B.F. Chen, T. Khelif, T., ... and K. Burke, Understanding kernel ridge regression: Common behaviors from simple functions to density functionals. International Journal of Quantum Chemistry, 115(16), 1115–1128, (2015).

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

[27] H. Wendland, Scattered data approximation, 2005.

[28] H.H. Zhang, M.G. Genton, and P. Liu, Compactly supported radial basis function kernels. North Carolina State University. Dept. of Statistics, (2004).

[29] Y. Zhang, J. Duchi, and M. Wainwright, Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. The Journal of Machine Learning Research, 16(1), 3299–3340, (2015).

[30] F. Zhdanov and Y. Kalnishkan, An identity for kernel ridge regression. Theoretical Computer Science, 473, 157–178, (2013).