

## Article

# Machine Learning Application of Generalized Gaussian Radial Basis Function and Its Reproducing Kernel Theory

Himanshu Singh 

Department of Mathematics, The University of Texas at Tyler, Tyler, TX 75799, USA; hsingh@uttyler.edu

**Abstract:** Gaussian Radial Basis Function Kernels are the most-often-employed kernel function in artificial intelligence for providing the optimal results in contrast to their respective counterparts. However, our understanding surrounding the utilization of the *Generalized Gaussian Radial Basis Function* across different machine learning algorithms, such as kernel regression, support vector machines, and pattern recognition via neural networks is incomplete. The results delivered by the Generalized Gaussian Radial Basis Function Kernel in the previously mentioned applications remarkably outperforms those of the Gaussian Radial Basis Function Kernel, the Sigmoid function, and the ReLU function in terms of accuracy and misclassification. This article provides a concrete illustration of the utilization of the Generalized Gaussian Radial Basis Function Kernel as mentioned earlier. We also provide an explicit description of the reproducing kernel Hilbert space by embedding the Generalized Gaussian Radial Basis Function as an  $L^2$ –measure, which is utilized in implementing the analysis support vector machine. Finally, we provide the conclusion that we draw from the empirical experiments considered in the manuscript along with the possible future directions in terms of spectral decomposition of the Generalized Gaussian Radial Basis Function.

**Keywords:** Gaussian Radial Basis Function; reproducing kernel Hilbert space; support vector machine; neural network; ReLU; Generalized Hypergeometric Function; Pochhammer symbol; Hermite polynomials

**MSC:** 46E22; 68T01; 68T07; 68T09; 68T10; 33C20; 33C90; 28A10; 46E27; 46E30



**Citation:** Singh, H. Machine Learning Application of Generalized Gaussian Radial Basis Function and Its Reproducing Kernel Theory. *Mathematics* **2024**, *12*, 829. <https://doi.org/10.3390/math12060829>

Academic Editor: Florin Leon

Received: 17 February 2024

Revised: 28 February 2024

Accepted: 5 March 2024

Published: 12 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

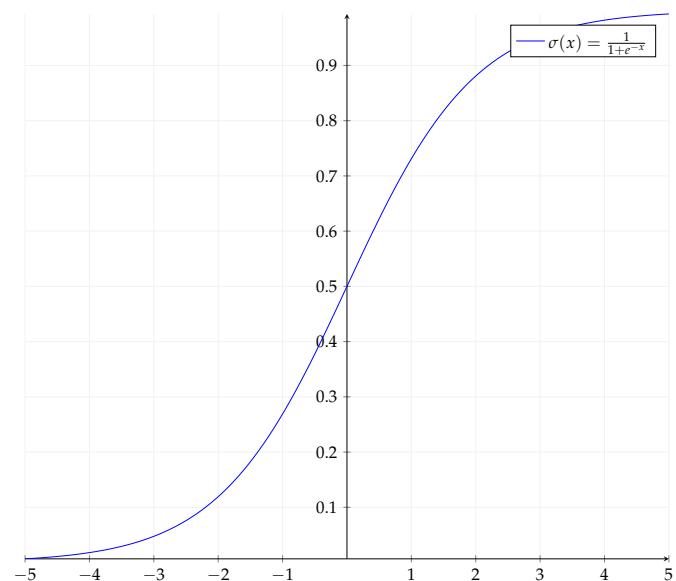
## 1. Introduction

Contemporary data science, which revolves around diverse architectural frameworks of machine learning, signifies crucial potential across a wide range of applications. These applications span from various fields such as physics and engineering [1] to areas like genomic-assisted prediction [2] and predictive modeling for cancer tumors [3]. The central task of (automated) machine learning is to devise an *optimal activation function* or *optimal kernel function*, depending upon the learning architecture, that is best-suited for the specific scientific model. Therefore, it is not hard to see that machine learning applications and many others across various scientific realms duly take advantage of various mathematical functions that arise in basic function theory and incorporate them along with their Hilbert space theories. This symbiotic relationship between advanced machine learning architecture avenues of artificial intelligence and the raw availability of various classes of mathematical functions is aptly captured by pictorial representations produced by the NATIONAL ACADEMY OF SCIENCES in [4]. This does not only simply depict the importance of such functions but also questions and inquires about the possibility of the existence and applicability of other new, unexplored, better math functions, which can advance the horizons of artificial intelligence, for instance in [5].

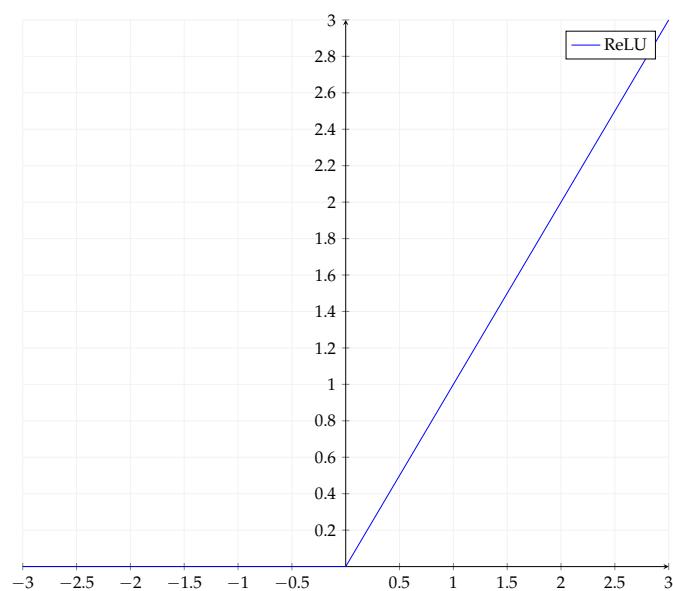
Across the rich and diverse history of machine learning, numerous activation functions, along with their incorporation in a *kernel* sense, have been proposed. Certain important examples of activation functions are the *Sigmoid function* or *S-Shaped curves* in Figure 1, like

the hyperbolic tangent function (given as  $\frac{1}{1+e^{-x}}$  cf. [6]). The class of Sigmoid functions is employed to classify data objects where the output is supposed to be constrained on a limited range (usually on  $[0, 1]$ ).

The other class of important and popular activation functions comprises ReLU activation functions in Figure 2, defined by  $f(x) = x$  if  $x > 0$ , which are employed in machine learning problems of reinforcement learning, quantification, and classification. If one adds a piece of function defined as  $f(x) = \alpha x$  for  $x \leq 0$ , where  $\alpha \neq 0$  to the actual definition of ReLU, we obtain the  $\alpha$ ReLU activation function (see (29) in this article), which is often also referred to as LeakyReLU. It is essential to acknowledge that both ReLU and the  $\alpha$ ReLU activation function have discontinuities at the origin; these lead to the failure of the gradient descent optimizer by producing undefined gradients [7].



**Figure 1.** Graph of Sigmoid activation function.



**Figure 2.** Graph of ReLU activation function.

Furthermore, one of the most important and interesting classes of mathematical functions that leverage crucial platforms of machine learning or deep learning, to name a few, are *radial basis functions*, defined as follows (cf. [8]).

**Definition 1.** A function  $\Phi : \mathbf{R}^d \rightarrow \mathbf{R}$  is referred to as radial if there exists a univariate function  $\phi : \{0\} \cup \mathbf{R}_+ \rightarrow \mathbf{R}$ , such that

$$\Phi(\mathbf{x}) = \phi(r), \quad \text{where } r = \|\mathbf{x}\|.$$

Here,  $\|\bullet\|$  is some norm defined on  $\mathbf{R}^d$ , usually  $\|\bullet\|_2$ , which is the Euclidean norm.

One such special type of mathematical radial basis function is the *Gaussian Radial Basis Function* (GRBF), which is given as:

$$g_{\sigma^2}(r) :=_{\text{def}} \exp(-\sigma^2 r^2),$$

where  $\sigma > 0$ . The Gaussian Radial Basis Function is an activation function that possesses a bell-shaped curve (top in Figure 3); it has impact-full applications in modeling random variables, which are distributed according to the Gaussian probability distribution. For instance, a neural network architecture predicting the speed of a car might employ Gaussian kernel regression, since the speed of the car follows the Gaussian probability distribution.

Moreover, in recent years, learning algorithms of support vector machines (SVMs) have played a key role in data-mining methods; these are workhorse machine learning tools in industry and science. The crucial explanation for this aspect is the unique ability of SVMs to embed in higher-dimensional nonlinear spaces for nonlinear classifiers. The theoretical underpinnings of SVMs revolve around the *kernel function*, which essentially enables one to operate in a high-dimensional space without computing coordinates of data but by simply computing inner products between all pairs of data in the feature space (cf. page 197 in [9]). To that end, we define the most commonly used kernel function, that is, the Gaussian Radial Basis Function Kernel.

Let  $\|\bullet\|_2$  be the usual Euclidean norm on  $\mathbf{R}^d$ ; then, the  $g(\|\bullet - \bullet\|_2)$  function is comfortably synonymous to its kernel notion, which is famously referred to as the Gaussian Radial Basis Function Kernel, given as  $K_\sigma(\mathbf{x}, \mathbf{z}) :=_{\text{def}} g_{\sigma^2}(\|\mathbf{x} - \mathbf{z}\|_2)$  for  $\mathbf{x}, \mathbf{z} \in \mathbf{R}^d$ . Explicitly, that is

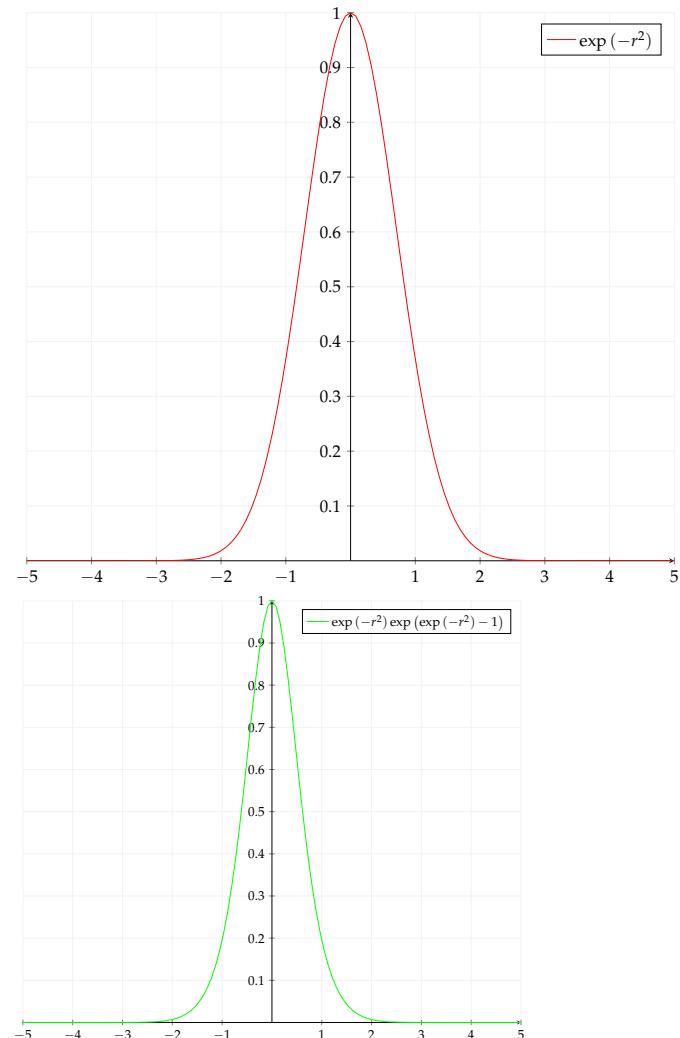
$$K_\sigma(\mathbf{x}, \mathbf{z}) = \exp(-\sigma^2 \|\mathbf{x} - \mathbf{z}\|_2^2).$$

The Gaussian Radial Basis Function Kernel (Figure 4) is a building block for various learning architectures, such as spatial statistics [10], dynamical system identification [11], Gaussian processes for machine learning [12], etc. Additionally, it is also employed for the classification of the existence of objects (cf. [13]). This particular study extends the idea of the Gaussian Radial Basis Function Kernel to what is referred to as the *Generalized Gaussian Radial Basis Function Kernel* (GGRBF Kernel in Figure 5), as introduced in [14], on the avenues of machine learning and deep learning.

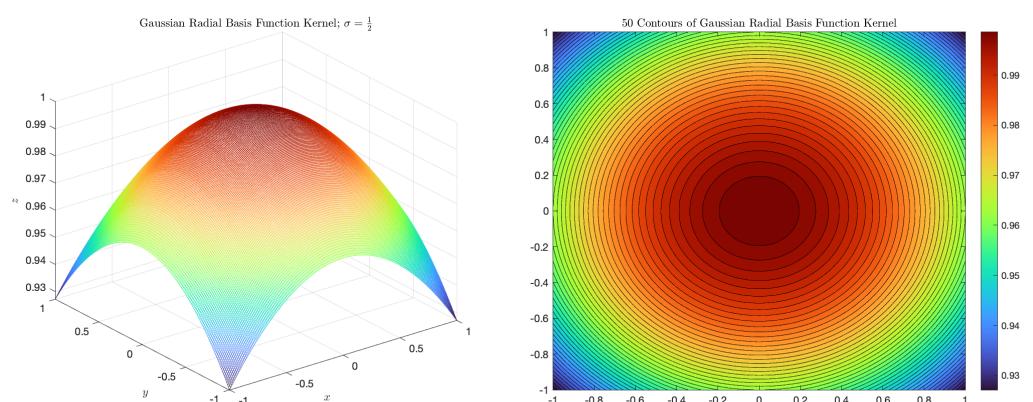
**Definition 2.** Let  $\sigma > 0$  and  $\sigma_0 \geq 0$ . Then, the Generalized Gaussian Radial Basis Function Kernel for  $\mathbf{x}$  and  $\mathbf{z} \in \mathbf{R}^d$  is defined as:

$$\begin{aligned} K_{\sigma, \sigma_0}(\mathbf{x} - \mathbf{z}) &:=_{\text{def}} \left[ g_{\sigma^2}(r) \exp(g_{\sigma_0^2}(r) - 1) \circ (\|\mathbf{x} - \mathbf{z}\|_2) \right] \\ &= g_{\sigma^2}(\|\mathbf{x} - \mathbf{z}\|_2) e^{(g_{\sigma_0^2}(\|\mathbf{x} - \mathbf{z}\|_2) - 1)} \\ &= e^{-\sigma^2 \|\mathbf{x} - \mathbf{z}\|_2^2} e^{-\sigma_0^2 (\|\mathbf{x} - \mathbf{z}\|_2^2) - 1}. \end{aligned} \tag{1}$$

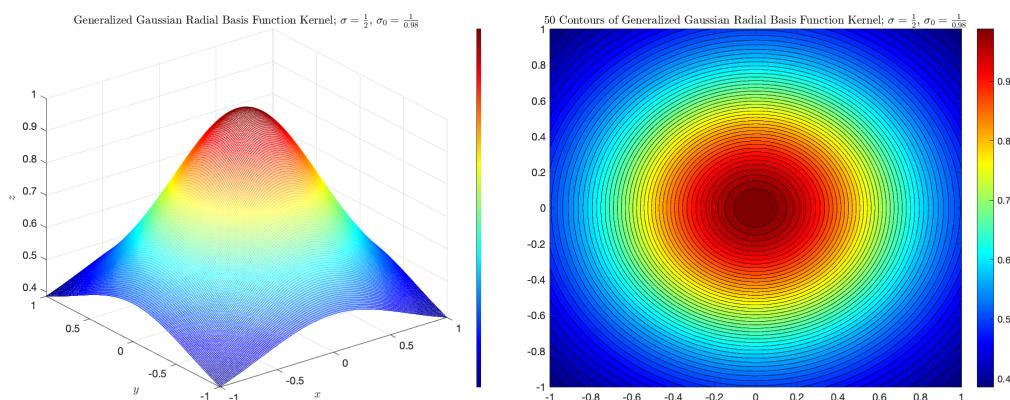
Now that we have introduced the Generalized Gaussian Radial Basis Function Kernel, let us briefly discuss the motivation of the present article with respect to it.



**Figure 3.** Graphs of radial basis functions of exponential types.



**Figure 4.** GRBF kernel.

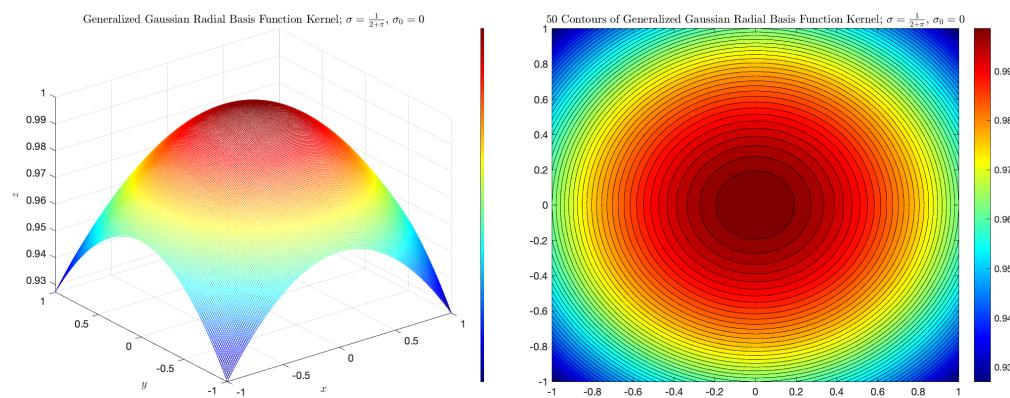


**Figure 5.** GGRBF kernel with  $\sigma = \frac{1}{2}$  and  $\sigma_0 = \frac{1}{0.98}$ .

### 1.1. Motivation

The applicability of the Generalized Gaussian Radial Basis Function was introduced in [15] to provide better results in contrast to the Gaussian Radial Basis Function results; here, we specifically refer to the results of convergence and stability for interpolation problems on Franke's test function and Runge's function or solving the system with Tikhonov regularization and Riley's algorithm. Drawing significant inspiration from [15], the application of the unexplored Generalized Gaussian Radial Basis Function Kernel was documented in [14], specifically in the contexts of SVMs, kernel regression, and pattern recognition, through the activation function in a neural network. These state-of-the-art methods in learning architectures are leveraged by a peculiar topic from the Hilbert function space called the *reproducing kernel Hilbert space* (RKHS) [16]. The analysis from the reproducing kernel theory perspective for the Gaussian Radial Basis Function Kernel has already been established by [17]; in the paper, an investigation related to the norms, orthonormal basis, and feature space was presented. However, with the present empirical evidence supporting better results obtained by employing the Generalized Gaussian Radial Basis Function Kernel, it becomes important to perform the same investigation for the Generalized Gaussian Radial Basis Function Kernel.

Note that, if  $\sigma_0 = 0$ , then we obtain the traditional Gaussian Radial Basis Function Kernel. This is demonstrated in Figure 6.



**Figure 6.** GGRBF kernel with  $\sigma = \frac{1}{2+\pi}$  and  $\sigma_0 = 0$ .

### 1.2. Contributions

This article is motivated by documenting the application of the Generalized Gaussian Radial Basis Function Kernel on various machine learning and deep learning architectures. In the process of executing the Generalized Gaussian Radial Basis Function Kernel over learning architectures, we discover that the results produced by the Generalized Gaussian

Radial Basis Function Kernel excel in terms of comparative accuracy or misclassification; these are achieved in contrast to the results obtained from the existing counterparts. These intriguing and promising findings naturally demand an exploration of the Generalized Gaussian Radial Basis Function from the perspective of mathematical function theory, which starts with the identification of the Hilbert space, generated by it as an  $L^2$ -measure. Furthermore, with the established Hilbert function space theory, we are also able to address the corresponding reproducing kernel theory by virtue of *Aronszajn's theorem* (see [16]).

Having aptly addressed the Hilbert function theory generated by the Generalized Gaussian Radial Basis Function, this study immediately provides its application in learning architectures. The detailed investigation on the reproducing kernel for the Hilbert space of the Generalized Gaussian Radial Basis Function is proved to be *universal* (see in Section 4) on the restrictions over  $\mathbf{R}^d \times \mathbf{R}^d$ ; this makes it ready-to-use for learning architecture problems, such as standard binary classification through SVMs, function approximation through kernel regression, and pattern recognition through deep convolutional neural networks. In particular, we determine that the results related to the Hilbert function theory of the Generalized Gaussian Radial Basis Function Kernel can be leveraged for SVM applications by addressing the hinge loss function  $L$  and minimal  $L$ -risk losses as well.

Finally, this article also provides promising future directions in light of the spectral decomposition of the Generalized Gaussian Radial Basis Function.

### 1.3. Plan of the Article

The present article is organized as follows: we provide the essential preliminaries of reproducing kernel theory in Section 2. Then, we give the results from the Hilbert space function theory, such as the orthonormal basis and the reproducing kernel of the Generalized Gaussian Radial Basis Function Kernel in Section 3. Then, we provide the functionality of the reproducing kernel in a real sense in Section 4; this is followed by compelling empirical comparison results in Section 5. Lastly, we provide our conclusions and results in Section 6 and describe our future research directions in Section 7.

## 2. Notation and Preliminaries

### 2.1. Hypergeometric Function Notation

We begin first by recalling the basic calculus results related to the *Generalized Hypergeometric Function*  ${}_pF_q \left[ \begin{smallmatrix} a_1 & a_2 & \cdots & a_p \\ b_1 & b_2 & \cdots & b_q \end{smallmatrix} ; z \right]$  [18].

**Definition 3.** We define the Pochhammer symbol [19] (rising factorial notation) as

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)} = a(a+1)\cdots(a+k-1).$$

Then, we define Generalized Hypergeometric Function  ${}_pF_q \left[ \begin{smallmatrix} a_1 & a_2 & \cdots & a_p \\ b_1 & b_2 & \cdots & b_q \end{smallmatrix} ; z \right]$ , which is given as

$${}_pF_q \left[ \begin{smallmatrix} a_1 & a_2 & \cdots & a_p \\ b_1 & b_2 & \cdots & b_q \end{smallmatrix} ; z \right] :=_{\text{def}} \sum_{l=0}^{\infty} \frac{\prod_{i=1}^p (a_i)_l}{\prod_{i=1}^p (b_i)_l} \frac{z^l}{l!}. \quad (2)$$

**Example 1.** The summation  $\sum_{l=0}^{\infty} \frac{1}{(l+x)^{n+1}} \frac{1}{l!}$  can be represented as:

$$\begin{aligned} \sum_{l=0}^{\infty} \frac{1}{(l+x)^{n+1}} \frac{1}{l!} &= \sum_{l=0}^{\infty} \left( \frac{\Gamma(l+x)}{\Gamma(l+x+1)} \right)^{n+1} \frac{1}{l!} \\ &= \sum_{l=0}^{\infty} \left( \frac{(x)_l \Gamma(x)}{(x+1)_l \Gamma(x+1)} \right)^{n+1} \frac{1}{l!} \\ &= \sum_{l=0}^{\infty} \left( \frac{(x)_l \Gamma(x)}{(x+1)_l x \Gamma(x)} \right)^{n+1} \frac{1}{l!} \\ &= \frac{1}{x^{n+1}} \sum_{l=0}^{\infty} \frac{(x)_l^{n+1}}{(x+1)_l^{n+1}} \frac{1}{l!} \\ &= \frac{1}{x^{n+1}} {}_{n+1}F_{n+1} \left[ \begin{matrix} x \\ x+1 \end{matrix} ; 1 \right] \quad (\text{use (2)}). \end{aligned}$$

The example presented above will be useful in addressing further details such as the orthonormal basis in the present manuscript. To avoid heavy notation clutter, we write

$${}_{n+1}F_{n+1} \left[ \begin{matrix} x \\ x+1 \end{matrix} ; 1 \right] :=_{\text{notation}} \mathcal{F}_{n,x,1} \quad (3)$$

from now onward where necessary. Note that  $\mathcal{F}_{n,\infty,1} = e$ .

## 2.2. Field and Space Notations

The set of natural numbers in union with 0 is denoted by  $\mathbf{W}$ , that is  $\mathbf{W} := 0, 1, 2, \dots$ . We use Kronecker delta  $\delta_{nm}$  on non-negative integers  $n$  and  $m$  to show that  $\delta_{nm} = 1$ ; this is the case whenever  $n = m$  and  $\delta_{nm} = 0$ , if  $n \neq m$ . We denote a complex number,  $z = x + iy$ , where  $x$  and  $y \in \mathbf{R}$ . With that  $z$ , its conjugate part is given as  $\bar{z} = x - iy$ , along with its absolute value as  $|z|^2 = z \cdot \bar{z} = x^2 + y^2$ . We reserve symbol  $\mathbf{K}$  to treat with a choice of fields on which we will operate; in particular,  $\mathbf{K}$  can either be  $\mathbf{R}$  or  $\mathbf{C}$ .

If  $X$  denotes the space of input values (sometimes a closed subset of  $\mathbf{R}^d$ ), then we write  $C(X)$  as the space of continuous functions  $f : X \rightarrow \mathbf{R}$ ; on the other hand,  $C^m(X)$  represents the space of the  $m$ -times differentiable function, including  $m = \infty$ .

## 2.3. Tensor Product Notation

We recall the *tensor product* between two functions; say that  $f_1, f_2 : X \rightarrow \mathbf{K}$  is given as  $f_1 \otimes f_2 : X \times X \rightarrow \mathbf{K}$ . Then, for all  $x, x' \in X$ , the tensor product  $f_1 \otimes f_2$  is defined as  $f_1 \otimes f_2(x, x') := f_1(x)f_2(x')$ .

## 2.4. Preliminaries

**Definition 4.** Let  $X = \emptyset$ ; then, a function  $k : X \times X \rightarrow \mathbf{K}$  is referred to as the *kernel* on  $X$  if there exists a  $\mathbf{K}$ -Hilbert space  $(H, \langle \cdot, \cdot \rangle_H)$  accompanied by a map  $\Phi : X \rightarrow H$  such that  $\forall x, x' \in X$ , we have

$$k(x, x') = \langle \Phi(x'), \Phi(x) \rangle_H. \quad (4)$$

We regard  $\Phi$  as the *feature map* and  $H$  as the *feature space* of  $k$ .

Now that we have introduced the basic notion from the kernel theory in the definition provided above, we can now comfortably define the building block of this article: *reproducing kernel Hilbert space—RKHS*.

**Definition 5.** Let  $X = \emptyset$  and  $(H, \langle \cdot, \cdot \rangle_H)$  be the Hilbert function space over  $X$ .

1. The space  $H$  is referred to as the reproducing kernel Hilbert space (RKHS) if  $\forall x \in X$ , the evaluation functional  $\mathcal{E}_x : H \rightarrow \mathbf{K}$  defined as  $\mathcal{E}_x(f) := f(x)$ ,  $f \in H$  is continuous.

**Definition 6.** A function  $k : X \times X \rightarrow \mathbf{K}$  is called reproducing kernel of  $H$  if we have:

1.  $k(\cdot, x) \in H \forall x \in X$ , that is  $\|k(\cdot, x)\|_H < \infty$ , and
2.  $k(\cdot, \cdot)$  has the reproducing property; that is

$$f(x) = \langle f, k(\cdot, x) \rangle_H \forall f \in H \text{ and } x \in X.$$

It is worthwhile mentioning that the norm convergence yields the point-wise convergence inside RKHS. This fact can be readily learned due to the continuity of the evaluation function. This is demonstrated as follows for an arbitrary  $f \in H$  and  $\{f_n\}_n \in H$  with  $\|f - f_n\|_H \rightarrow 0$  as  $n \rightarrow \infty$ ; then,

$$\lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \mathcal{E}_x(f_n) =_{(\text{continuity of } \mathcal{E}_x)} \mathcal{E}_x(f) = f(x).$$

Now, we will state an important theorem from [16], which dictates the relationship between the reproducing kernel of the RKHS  $H$  and the orthonormal basis of it.

**Theorem 1** (Aronszajn's Theorem [16]). Let  $H$  be an RKHS over an nonempty set  $X$ ; then,  $k : X \times X \rightarrow \mathbf{K}$ , defined as  $k(x, x') := \langle \mathcal{E}_x, \mathcal{E}_{x'} \rangle_H$  for  $x, x' \in X$  is the only reproducing kernel of  $H$ . Additionally, for some index set,  $\mathcal{I}$ , if we have  $\{\mathbf{e}_i\}_{i \in \mathcal{I}}$  as an orthonormal basis, then for all  $x, x' \in X$ , we have

$$k(x, x') = \sum_{i \in \mathcal{I}} \mathbf{e}_i(x) \overline{\mathbf{e}_i(x')}, \quad (5)$$

with an absolute convergence.

### 3. Function Space of Generalized Gaussian Radial Basis Measure

Let  $d \in \mathbf{N}$ ,  $\sigma > 0$  and  $\sigma_0 \geq 0$  and  $f$  be holomorphic functions  $f : \mathbf{C}^d \rightarrow \mathbf{C}$ ; we first write the measure of our interest:

$$d\mu_{\sigma, \sigma_0, d}(z) := e^{-\sigma^2|z|^2} e^{-\sigma_0^2|z|^2 - 1} dV_{\mathbf{C}^d}(z). \quad (6)$$

Here, ' $dV_{\mathbf{C}^d}(z)$ ' is the usual Lebesgue measure on the entire  $\mathbf{C}^d$ . For  $d = 1$ , we write simply  $d\mu_{\sigma, \sigma_0}(z)$  to denote the typical Lebesgue area measure on  $\mathbf{C}$ . We now provide the inner product associated with this measure as:

$$\langle f, g \rangle_{\sigma, \sigma_0, \mathbf{C}^d} := \mathcal{N}_{\sigma, \sigma_0, d} \int_{\mathbf{C}^d} f(z) \overline{g(z)} d\mu_{\sigma, \sigma_0, d}(z). \quad (7)$$

Here, ' $\mathcal{N}_{\sigma, \sigma_0, d}$ ' is the normalization constant whose value is explicitly given as  $\left(\frac{e\sigma^2}{2\pi}\right)^d$ . Once we have define the inner product for the space, the norm for holomorphic function  $f : \mathbf{C}^d \rightarrow \mathbf{C}$  is:

$$\|f\|_{\sigma, \sigma_0, \mathbf{C}^d}^2 := \left(\frac{e\sigma^2}{2\pi}\right)^d \int_{\mathbf{C}^d} |f(z)|^2 d\mu_{\sigma, \sigma_0, d}(z). \quad (8)$$

We write the following for the collection of holomorphic functions  $f : \mathbf{C}^d \rightarrow \mathbf{C}$ , whose  $\|\bullet\|_{\sigma, \sigma_0, \mathbf{C}^d}$  is finite, that is:

$$H_{\sigma, \sigma_0, \mathbf{C}^d} := \left\{ \text{holomorphic function } f : \mathbf{C}^d \rightarrow \mathbf{C} : \|f\|_{\sigma, \sigma_0, \mathbf{C}^d} < \infty \right\}. \quad (9)$$

Once we have defined norm in (8) and the associated Hilbert space in (9), we can provide the following formulation, which makes the Hilbert space  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  as an RKHS.

**Theorem 2.** For all  $\sigma > 0$ ,  $\sigma_0 \geq 0$  and all compact sets  $K \subset \mathbf{C}^d$ , there exists a positive constant  $c_{\sigma, \sigma_0, d}$  such that for all  $z \in K$  and  $f \in H_{\sigma, \sigma_0, \mathbf{C}^d}$ , we have

$$|f(z)| \leq c_{\sigma, \sigma_0, d} \|f\|_{\sigma, \sigma_0, \mathbf{C}^d}. \quad (10)$$

**Proof.** Denote  $\mathbb{B}_{(0,1)}$  as the complex unit ball in  $\mathbf{C}$ . Define

$$c_{\sigma, \sigma_0, d} := \sup_{z \in K + \mathbb{B}_{(0,1)}^d} \left\{ e^{-\sigma^2 |z|^2} e^{e^{-\sigma_0^2 |z|^2} - 1} \right\}.$$

In the spirit of ([17], Lemma-3, Page 4639), we have

$$\prod_{j=1}^d r_j |f(z)|^2 \leq \frac{1}{(2\pi)^d} \prod_{j=1}^d r_j \int_{[0, 2\pi]^d} \left| f(z_1 + r_1 e^{i\theta_1}, \dots, z_d + r_d e^{i\theta_d}) \right|^2 d\theta.$$

Integration of the above with respect to  $(r_1, \dots, r_d) \in [0, 1]^d$  yields:

$$\begin{aligned} |f(z)|^2 &\leq \frac{1}{(2\pi)^d} \int_{z + \mathbb{B}_{(0,1)}^d} |f(z')|^2 dV(z') \\ &\leq \frac{c_{\sigma, \sigma_0, d}}{(2\pi)^d} \int_{z + \mathbb{B}_{(0,1)}^d} |f(z')|^2 e^{-\sigma^2 |z'|^2} e^{e^{-\sigma_0^2 |z'|^2} - 1} dV(z') \\ &\leq \frac{c_{\sigma, \sigma_0, d}}{(e\sigma^2)^d} \|f\|_{\sigma, \sigma_0, d}^2. \end{aligned}$$

In particular, we have  $c_{\sigma, \sigma_0, d} = \sqrt{\frac{c_{\sigma, \sigma_0, d}}{(e\sigma^2)^d}}$ . Hence, the result is established.  $\square$

Establishing Theorem 2 immediately yields that  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  is indeed an RKHS and we state here as an important corollary.

**Corollary 1.** The Hilbert space  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  identified by the inner product  $\langle \cdot, \cdot \rangle_{\sigma, \sigma_0, \mathbf{C}^d}$  is an RKHS for all  $\sigma > 0$  and  $\sigma_0 \geq 0$ .

#### Orthonormal Basis

We will need following technical result to establish the orthonormal basis for the RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$ .

**Theorem 3.** For every  $\sigma > 0$ ,  $\sigma_0 \geq 0$  and  $n, m \in \mathbf{W}$ , we have

$$\int_{\mathbf{C}} z^n \overline{z^m} d\mu_{\sigma, \sigma_0}(z) = \left( \sqrt{\frac{2\pi n!}{e\sigma^{2n+2}}} \mathcal{F}_{n, \hat{\sigma}, 1} \sqrt{\frac{2\pi m!}{e\sigma^{2m+2}}} \mathcal{F}_{m, \hat{\sigma}, 1} \right) \delta_{nm} \quad (11)$$

where  $\hat{\sigma} = \frac{\sigma_0^2}{\sigma^2}$  and  $\mathcal{F}_{n, \hat{\sigma}, 1}$  are defined in (3).

**Proof.** Employ the polar coordinate of  $z = re^{i\theta}$  to have:

$$\int_{\mathbf{C}} z^n \overline{z^m} d\mu_{\sigma, \sigma_0}(z) = \int_0^\infty r^{n+m} e^{-\sigma^2 r^2} e^{e^{-\sigma_0^2 r^2} - 1} r dr \int_0^{2\pi} e^{i(n-m)\theta} d\theta. \quad (12)$$

The quantity  $\int_0^\infty r^{n+m} e^{-\sigma^2 r^2} e^{-(\sigma_0^2 r^2 - 1)} r dr \int_0^{2\pi} e^{i(n-m)\theta} d\theta$  is 0 when  $n \neq m$ . Now, assume that  $n = m$  in (12), then:

$$\begin{aligned} \int_{\mathbb{C}} z^n \overline{z^m} d\mu_{\sigma, \sigma_0}(z) &= 2\pi \int_0^\infty r^{2n} e^{-\sigma^2 r^2} e^{-(\sigma_0^2 r^2 - 1)} r dr \\ &= \frac{2\pi}{e} \int_0^\infty t^n e^{-\sigma^2 t} e^{-(\sigma_0^2 t)} dt \quad (\text{put } t = r^2) \\ &= \frac{2\pi}{e} \int_0^\infty \left(\frac{s}{\sigma^2}\right)^n e^{-s} e^{-(\sigma_0^2/\sigma^2 s)} \frac{ds}{\sigma^2} \quad (\text{put } s = \sigma^2 t) \\ &= \frac{2\pi}{e(\sigma^2)^{n+1}} \int_0^\infty s^n e^{-s} e^{-(\sigma_0^2/\sigma^2 s)} ds \\ &= \frac{2\pi}{e(\sigma^2)^{n+1}} \int_1^0 \left(\log \frac{1}{y}\right)^n e^{y(\sigma_0^2/\sigma^2)} (-dy) \quad (\text{put } e^{-s} = y) \\ &= \frac{2\pi}{e(\sigma^2)^{n+1}} \int_0^1 \left(\log \frac{1}{y}\right)^n \left( \sum_{l=0}^{\infty} \frac{(y(\sigma_0^2/\sigma^2))^l}{l!} \right) dy \end{aligned} \tag{13}$$

$$= \frac{2\pi}{e(\sigma^2)^{n+1}} \sum_{l=0}^{\infty} \frac{1}{l!} \int_0^1 \left(\log \frac{1}{y}\right)^n y^{l\hat{\sigma}} dy \quad (\text{use } \hat{\sigma} = \frac{\sigma_0^2}{\sigma^2}) \tag{14}$$

$$= \frac{2\pi}{e(\sigma^2)^{n+1}} \sum_{l=0}^{\infty} \frac{1}{l!} \frac{\Gamma(n+1)}{(l\hat{\sigma} + 1)^{n+1}} \tag{15}$$

$$= \frac{2\pi n!}{e\sigma^{2n+2}} \mathcal{F}_{n, \hat{\sigma}, 1} \quad (\text{use (3)}).$$

We used the entry 4.272(6)), Page 551 in [20] for the result in (15). The summation over  $l$  in (13) is interchanged with integral with respect to  $dy$  for  $0 \leq y \leq 1$  by virtue of Fubini–Tonelli's theorem as  $\int_0^1 \left(\log \frac{1}{y}\right)^n e^{y(\sigma_0^2/\sigma^2)} dy$  is finite which is demonstrated as:

$$\left| \int_0^1 \left(\log \frac{1}{y}\right)^n e^{y(\sigma_0^2/\sigma^2)} dy \right| \leq \int_0^1 \left| \left(\log \frac{1}{y}\right)^n e^{y(\sigma_0^2/\sigma^2)} \right| dy \leq e \int_0^1 \left| \left(\log \frac{1}{y}\right)^n \right| dy = e\Gamma(n+1) < \infty.$$

Thus, the result prevails.  $\square$

We will employ the notation  $\hat{\sigma} = \frac{\sigma_0^2}{\sigma^2}$  for  $\sigma > 0$  and  $\sigma_0 \geq 0$  wherever it is required. In the light of Theorem 1, we have to determine the orthonormal basis of  $H_{\sigma, \sigma_0, \mathbb{C}^d}$ .

**Theorem 4.** Let  $\sigma > 0$ ,  $\sigma_0 \geq 0$  and

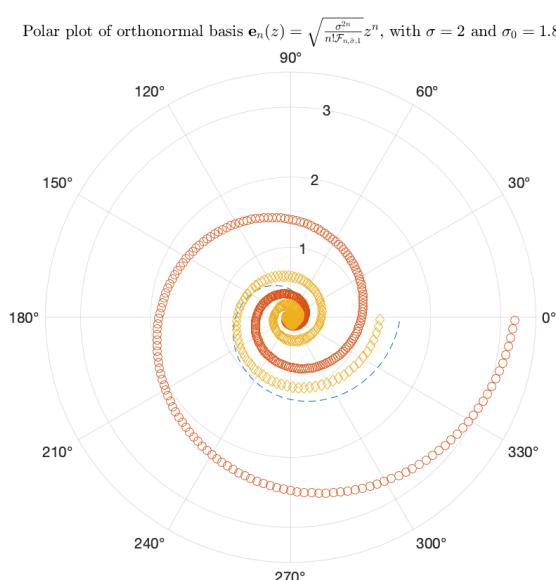
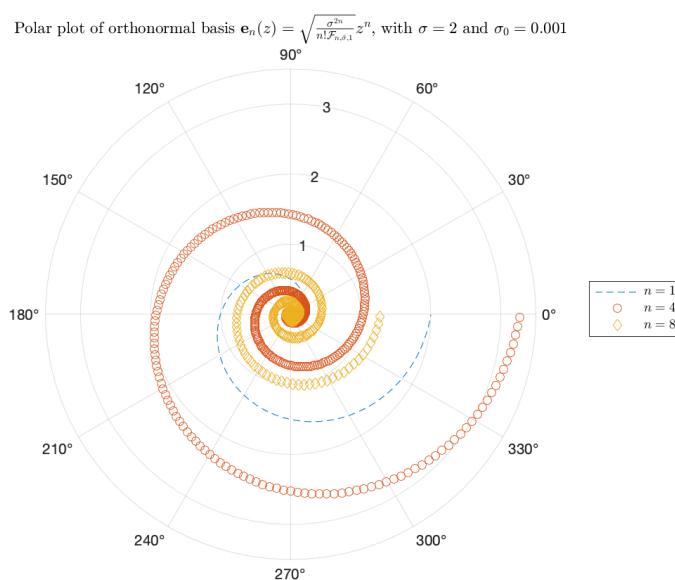
$$\mathbf{e}_n(z) := \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} z^n \tag{16}$$

for some  $z \in \mathbb{C}$ . Then, the tensor–product system  $(\mathbf{e}_{n_1} \otimes \cdots \otimes \mathbf{e}_{n_d})_{n_1, \dots, n_d \geq 0}$  forms the orthonormal basis of  $H_{\sigma, \sigma_0, \mathbb{C}^d}$ .

**Proof.** We establish our result for  $d = 1$  for an initial basic understanding. For this, let us show that  $\{\mathbf{e}_n\}_{n \in \mathbb{W}}$  forms an orthonormal system. So, consider  $z \in \mathbb{C}$  and let  $m, n \in \mathbb{W}$ . Then,

$$\begin{aligned}\langle \mathbf{e}_n, \mathbf{e}_m \rangle_{\sigma, \sigma_0} &= \frac{e\sigma^2}{2\pi} \int_{\mathbb{C}} \mathbf{e}_n(z) \overline{\mathbf{e}_m(z)} d\mu_{\sigma, \sigma_0}(z) \\ &= \frac{e\sigma^2}{2\pi} \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} \sqrt{\frac{\sigma^{2m}}{m! \mathcal{F}_{m, \hat{\sigma}, 1}}} \int_{\mathbb{C}} z^n \overline{z^m} d\mu_{\sigma, \sigma_0}(z) \\ &= \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases} \quad (\text{use Theorem 3}).\end{aligned}$$

We use the ‘polarplot’ function available in MATLAB R2023b to construct the polar plot of  $\mathbf{e}_n(z)$ . These results are given in Figure 7, where the numbers 1, 2, and 3 are the radius values in the polar plot.



**Figure 7.** Polar plot of orthonormal basis  $\mathbf{e}_n(z) := \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} z^n$  given in (16).

The above result concludes that  $\{\mathbf{e}_n\}_{n \in \mathbf{W}}$  is actually an orthonormal system. To this end, we have to establish that it is also complete. So, for this, pick a holomorphic function  $f \in H_{\sigma, \sigma_0, \mathbf{C}}$  with  $f(z) = \sum_{l=0}^{\infty} a_l z^l$  and observe that

$$\begin{aligned}\langle f, \mathbf{e}_n \rangle_{\sigma, \sigma_0} &= \frac{e\sigma^2}{2\pi} \int_{\mathbf{C}} f(z) \overline{\mathbf{e}_n(z)} d\mu_{\sigma, \sigma_0}(z) \\ &= \frac{e\sigma^2}{2\pi} \sum_{l=0}^{\infty} a_l \int_{\mathbf{C}} z^l \overline{\mathbf{e}_n(z)} d\mu_{\sigma, \sigma_0}(z) \\ &= \frac{e\sigma^2}{2\pi} \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} \sum_{l=0}^{\infty} a_l \int_{\mathbf{C}} z^l \overline{z^n} d\mu_{\sigma, \sigma_0}(z) \\ &= \frac{e\sigma^2}{2\pi} \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} \sum_{l=0}^{\infty} a_l \left( \sqrt{\frac{2\pi l!}{e\sigma^{2l+2}} \mathcal{F}_{n, \hat{\sigma}, 1}} \sqrt{\frac{2\pi n!}{e\sigma^{2n+2}} \mathcal{F}_{n, \hat{\sigma}, 1}} \right) \delta_{ln} \\ &= \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} a_n \frac{n! \mathcal{F}_{n, \hat{\sigma}, 1}}{\sigma^{2n}} \\ &= \left[ \sqrt{\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}}} \right]^{-1} a_n.\end{aligned}$$

Since the constant is  $\frac{\sigma^{2n}}{n! \mathcal{F}_{n, \hat{\sigma}, 1}} \neq 0$  for any choice of  $n$ , the condition that  $\langle f, \mathbf{e}_n \rangle = 0$  for all  $n \in \mathbf{W}$  yields that  $a_n = 0$  for all  $n \in \mathbf{W}$ , which results in the conclusion that  $f \equiv 0$ . Therefore,  $\{\mathbf{e}_n\}_{n \in \mathbf{W}}$  is complete. Now, we establish these results in a  $d$ -dimensional situation by employing the tensor product notation of Section 2.3. To this end, we see that

$$\langle \mathbf{e}_{n_1} \otimes \cdots \otimes \mathbf{e}_{n_d}, \mathbf{e}_{m_1} \otimes \cdots \otimes \mathbf{e}_{m_d} \rangle_{\sigma, \sigma_0, d} = \prod_{j=1}^d \langle \mathbf{e}_{n_j}, \mathbf{e}_{m_j} \rangle_{\sigma, \sigma_0}.$$

Hence, the orthonormality of  $\{\mathbf{e}_{n_1} \otimes \cdots \otimes \mathbf{e}_{n_d}\}_{n_1, \dots, n_d \in \mathbf{W}^d}$  is established due to the orthonormality of each  $\langle \mathbf{e}_{n_j}, \mathbf{e}_{m_j} \rangle_{\sigma, \sigma_0}$ . We still need to ensure that this  $d$ -dimensional orthonormal system is complete. Now, observe

$$\begin{aligned}\langle f, \mathbf{e}_{n_1} \otimes \cdots \otimes \mathbf{e}_{n_d} \rangle_{\sigma, \sigma_0, d} &= \left( \frac{e\sigma^2}{2\pi} \right)^d \int_{\mathbf{C}^d} f(z) \overline{\mathbf{e}_{n_1} \otimes \cdots \otimes \mathbf{e}_{n_d}(z)} d\mu_{\sigma, \sigma_0, \mathbf{C}^d}(z) \\ &= \left( \frac{e\sigma^2}{2\pi} \right)^d \sum_{l_1, \dots, l_d}^{\infty} a_{l_1, \dots, l_d} \mathbb{I}_{l, d},\end{aligned}$$

where  $\mathbb{I}_{l, d} = \int_{\mathbf{C}^d} z^l (\mathbf{e}_{n_1} \otimes \cdots \otimes \mathbf{e}_{n_d})(\bar{z}) d\mu_{\sigma, \sigma_0, \mathbf{C}^d}(z)$ . We further can simplify  $\mathbb{I}_d$ , as follows:

$$\begin{aligned}\mathbb{I}_{l, d} &= \int_{\mathbf{C}^d} z^l \mathbf{e}_{n_1}(\bar{z}_1) \wedge \cdots \wedge \mathbf{e}_{n_d}(\bar{z}_d) d\mu_{\sigma, \sigma_0}(z_1) \wedge \cdots \wedge d\mu_{\sigma, \sigma_0}(z_d) \\ &= \prod_{j=1}^d \left( \int_{\mathbf{C}} z_j^{l_j} \mathbf{e}_{n_j}(\bar{z}_j) d\mu_{\sigma, \sigma_0}(z_j) \right) \\ &= \prod_{j=1}^d \left( \int_{\mathbf{C}} z_j^{l_j} \bar{z}_j^{n_j} d\mu_{\sigma, \sigma_0}(z_j) \right) \\ &= \prod_{j=1}^d \left( \sqrt{\frac{2\pi l_j!}{e\sigma^{2l_j+2}} \mathcal{F}_{l_j, \hat{\sigma}, 1}} \sqrt{\frac{2\pi n_j!}{e\sigma^{2n_j+2}} \mathcal{F}_{n_j, \hat{\sigma}, 1}} \right) \delta_{l_j n_j} a_{l_1, \dots, l_d}.\end{aligned}$$

Finally,

$$\left(\frac{e\sigma^2}{2\pi}\right)^d \sum_{l_1, \dots, l_d}^{\infty} a_{l_1, \dots, l_d} \mathbb{I}_{l,d} = \left( \prod_{j=1}^d \left[ \sqrt{\frac{\sigma^{2n_j}}{n_j! \mathcal{F}_{n_j, \hat{\sigma}, 1}}} \right]^{-1} \right) a_{n_1, \dots, n_d}.$$

The further result for completeness in  $d$ -dimension follows a routine procedure from single-dimension case, as already discussed before.  $\square$

The following theorem provides the reproducing kernel for the RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  defined in (9).

**Theorem 5.** For  $\sigma > 0, \sigma_0 \geq 0$  and  $\hat{\sigma} = \frac{\sigma^2}{\sigma_0^2}$ , the reproducing kernel for the RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  is indeed (because Theorem 1) given as

$$K(\mathbf{z}, \mathbf{w}) := \sum_{n_1, \dots, n_d=0}^{\infty} \lambda_n (\mathbf{z} \bar{\mathbf{w}})^n, \quad (17)$$

where multi-index notation is employed:  $n = (n_1, \dots, n_d)$  and  $\lambda_n = \prod_{i=1}^d \frac{\sigma^{2n_i}}{n_i! \mathcal{F}_{n_i, \hat{\sigma}, 1}}$ .

**Proof.** We will demonstrate the desired proof as follows:

1. For  $\mathbf{w} \in \mathbf{C}^d$ , we will show that  $\|K(\cdot, \mathbf{w})\|_{\sigma, \sigma_0, d} < \infty$ . For this, consider that

$$\begin{aligned} \|K(\cdot, \mathbf{w})\|_{\sigma, \sigma_0, d}^2 &= \left( \frac{e\sigma^2}{2\pi} \right)^d \int_{\mathbf{C}^d} |K(\mathbf{z}, \mathbf{w})|^2 d\mu_{\sigma, \sigma_0, d}(\mathbf{z}) \\ &= \left( \frac{e\sigma^2}{2\pi} \right)^d \sum_{n_1, \dots, n_d}^{\infty} \sum_{m_1, \dots, m_d}^{\infty} \lambda_n \lambda_m \mathbf{w}^n \bar{\mathbf{w}}^m \int_{\mathbf{C}^d} \mathbf{z}^n \bar{\mathbf{z}}^m d\mu_{\sigma, \sigma_0, d}(\mathbf{z}) \end{aligned} \quad (18)$$

$$= \left( \frac{e\sigma^2}{2\pi} \right)^d \sum_{n_1, \dots, n_d}^{\infty} \lambda_n^2 |\mathbf{w}|^{2n} \left( \prod_{i=1}^d \int_{\mathbf{C}} z_i^{n_i} \bar{z}_i^{m_i} d\mu_{\sigma, \sigma_0}(z_i) \right) \quad (19)$$

$$\begin{aligned} &= \left( \frac{e\sigma^2}{2\pi} \right)^d \sum_{n_1, \dots, n_d}^{\infty} \lambda_n^2 |\mathbf{w}|^{2n} \left( \prod_{i=1}^d \frac{2\pi n_i!}{\sigma^{2n_i+2}} \mathcal{F}_{n_i, \hat{\sigma}, 1} \right) \\ &= \sum_{n_1, \dots, n_d}^{\infty} \frac{|\mathbf{w}|^{2n}}{\prod_{i=1}^d \frac{n_i!}{\sigma^{2n_i}} \mathcal{F}_{n_i, \hat{\sigma}, 1}}. \end{aligned}$$

We used the result of Theorem 4 from (18) to (19). For all  $\mathbf{w} \in \mathbf{C}^d$ , the quantity  $\sum_{n_1, \dots, n_d}^{\infty} \frac{|\mathbf{w}|^{2n}}{\prod_{i=1}^d \left( \frac{n_i!}{\sigma^{2n_i}} \right) \mathcal{F}_{n_i, \hat{\sigma}, 1}}$  achieves convergence. This implies that  $\|K(\cdot, \mathbf{w})\|_{\sigma, \sigma_0, d} < \infty$  for all  $\mathbf{w} \in \mathbf{C}^d$ . Therefore, the kernel function is  $K(\cdot, \mathbf{w}) \in H_{\sigma, \sigma_0, \mathbf{C}^d}$ .

2. In order to establish the reproducing property of  $K(\cdot, \mathbf{w})$ , pick an arbitrary  $f = \sum_{n_1, \dots, n_d}^{\infty} a_{n_1, \dots, n_d} \mathbf{w}^n \in H_{\sigma, \sigma_0, \mathbf{C}^d}$ . Then, consider the inner product of  $f$  with  $K(\cdot, \mathbf{w})$ , as follows:

$$\begin{aligned}\langle f, K(\cdot, \mathbf{w}) \rangle_{\sigma, \sigma_0, \mathbf{C}^d} &= \left( \frac{e\sigma^2}{2\pi} \right)^d \int_{\mathbf{C}^d} f(z) \overline{K(z, \mathbf{w})} d\mu_{\sigma, \sigma_0, \mathbf{C}^d}(z) \\ &= \left( \frac{e\sigma^2}{2\pi} \right)^d \sum_{n_1, \dots, n_d=0}^{\infty} \sum_{l_1, \dots, l_d}^{\infty} a_{n_1, \dots, n_d} \lambda_{l_1, \dots, l_d} \mathbf{w}^n \left( \prod_{i=1}^d \frac{2\pi}{e\sigma^2} \frac{\sigma^{2n_i}}{n_i! \mathcal{F}_{n_i, \hat{\sigma}, 1}} \right) \\ &= \sum_{n_1, \dots, n_d=0}^{\infty} a_{n_1, \dots, n_d} \mathbf{w}^n \\ &= f(\mathbf{w}).\end{aligned}$$

Hence, the desired result is achieved.  $\square$

The proof in the preceding theorem to demonstrate reproducing kernel nature of  $K(z, \mathbf{w})$  of RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  utilizes the basic machinery borrowed from the two-part definition for the reproducing kernel given in Definition 6. Now that we have determined the reproducing kernel theory for the Generalized Gaussian Radial Basis Function Kernel, we will investigate their functionality in terms of *universality*, which makes them applicable in addressing the learning architecture problems, such as standard binary classification through SVMs, function approximation through kernel regression, and pattern recognition through deep convolutional neural networks.

#### 4. Restriction and Universality of Reproducing Kernel of RKHS $H_{\sigma, \sigma_0, \mathbf{C}^d}$

It is crucial to address the fact that the restriction of the reproducing kernel  $K(z, \mathbf{w})$  on  $\mathbf{R}^d \times \mathbf{R}^d$  is also a kernel function, which eventually will assist us in establishing the *universal* nature of the  $K(z, \mathbf{w})$ .

**Lemma 1** (Restriction of kernels [21]). *Let  $k$  be a kernel on  $X$ , let  $\tilde{X}$  be a set, and let  $A : \tilde{X} \rightarrow X$  be a map. Then,  $\tilde{k}$  defined by  $\tilde{k}(x, x') := k(A(x), A(\tilde{x}))$ ,  $x, \tilde{x} \in \tilde{X}$ , which is a kernel on  $\tilde{X}$ . In particular, if  $\tilde{X} \subset X$ , then  $k|_{\tilde{X} \times \tilde{X}}$  is a kernel.*

In other words, the conclusion of the above lemma is given for all  $x, x' \in \mathbf{R}^d$ ; if  $k(x, x') \in \mathbf{R}$ , then  $k$  is also a kernel in the real sense (cf. [21]). This notion of restricting the kernel from  $\mathbf{C}$  to  $\mathbf{R}$  can allow the restriction of  $K(z, \mathbf{w})$  for the RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  on  $\mathbf{R}^d \times \mathbf{R}^d$  to have kernel in a *real* sense. Note that the  $\lambda'_n$ 's present in the reproducing kernel  $K(z, \mathbf{w})$  in (17) are strictly positive real numbers for all  $n \in \mathbf{W}$ ; also, observe that for all  $x, x' \in \mathbf{R}^d$ , we have

$$K|_{\mathbf{R}^d \times \mathbf{R}^d}(x, x') = \sum_{n_1, \dots, n_d=0}^{\infty} \lambda_n (xx')^n = \left[ \sum_{n_1, \dots, n_d=0}^{\infty} \lambda_n (xx')^n \right] \in \mathbf{R}, \quad (20)$$

which implies that the restriction of  $K(\cdot, \cdot)$  on  $\mathbf{R}^d \times \mathbf{R}^d$  is a kernel. Now, we shall recall the notion of *universal kernel*.

**Definition 7** (Universal Kernels [21]). *A continuous kernel  $k$  on a compact metric space  $X$  is called universal if the RKHS  $H$  of  $k$  is dense in  $C(X)$ . This means that, for every function  $g \in C(X)$  and all  $\epsilon > 0$ , there exists  $f \in H$ , such that  $\|f - g\|_{\infty} \leq \epsilon$ .*

Now that we have defined the notion of universal kernel function in the above definition, we provide an example of the same.

**Example 2** (Universal Taylor Kernels). For a fixed  $R \in (0, \infty]$  and a  $C^\infty$ -function  $f : (-R, R) \rightarrow \mathbf{R}$  that can be expanded into its Taylor series at 0, i.e.,

$$f(t) = \sum_{n=0}^{\infty} a_n t^n \quad t \in (-R, R). \quad (21)$$

Let  $X := \{x \in \mathbf{R}^d : \|x\|_2 < \sqrt{R}\}$ . If we have  $a_n > 0$  for all  $n \geq 0$  in (21), then  $k$  given by

$$k(x, x') := f(\langle x, x' \rangle), \quad x, x' \in X, \quad (22)$$

is a universal kernel on every compact subset of  $X$ .

With the help of universal Taylor kernels, as given above, we can see that the reproducing kernel of the RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  is a universal Taylor kernel, followed by its restriction on  $\mathbf{R}^d \times \mathbf{R}^d$ . This motivating result is captured in the following important theorem.

**Theorem 6.** Let  $X$  be a compact subset of  $\mathbf{R}^d$ ,  $\sigma > 0$  and  $\sigma \geq 0$ . Then, the kernel  $K|_{\mathbf{R}^d \times \mathbf{R}^d}$  defined in (20) is universal.

**Proof.** Let  $\sigma > 0$  and  $\sigma_0 \geq 0$ . Recall  $a_n$  from (21) and  $\lambda_n$  from (17) and set both of them equal to each other, that is,  $a_n = \lambda_n$  for all  $n \in \mathbf{W}$ . With these choices of  $a_n$  for all  $n \in \mathbf{W}$ , we now define a  $C^\infty$ -function  $f : (-R, R) \rightarrow \mathbf{R}$  as

$$f(t) = \sum_{n=0}^{\infty} \lambda_n t^n \quad t \in (-R, R). \quad (23)$$

We learn that, for  $x, x' \in \mathbf{R}^d$ , whose Euclidean norm is bounded by  $\sqrt{R}$ , we can express  $K|_{\mathbf{R}^d \times \mathbf{R}^d}(x, x')$  in terms of  $f$  as defined in (23) as:

$$K|_{\mathbf{R}^d \times \mathbf{R}^d}(x, x') = \sum_{n=0}^{\infty} \lambda_n (xx')^n = \sum_{n=0}^{\infty} \lambda_n (\langle x, x' \rangle)^n = f(\langle x, x' \rangle) \quad (24)$$

The establishment of  $K|_{\mathbf{R}^d \times \mathbf{R}^d}(x, x')$  for  $x, x' \in \mathbf{R}^d$  in terms of  $C^\infty$ -function  $f$  and in combination of  $\lambda_n > 0$  for all  $n \in \mathbf{W}$  makes  $K|_{\mathbf{R}^d \times \mathbf{R}^d}$  a universal kernel by the virtue of Example 2.  $\square$

The establishment of the universal nature of the reproducing kernel  $K(\cdot, \cdot)$  of the RKHS  $H_{\sigma, \sigma_0, \mathbf{C}^d}$  with restriction on  $\mathbf{R}^d \times \mathbf{R}^d$  in Theorem 6 allows one to approximate arbitrary continuous functions on compact subsets of  $\mathbf{R}^d$  with those functions that can be represented against the reproducing kernel  $K(\cdot, \cdot)$  to the inner product, as defined in (7) or consequently whose norm given in (8) is finite.

## 5. Empirical Evidence and Results Comparison

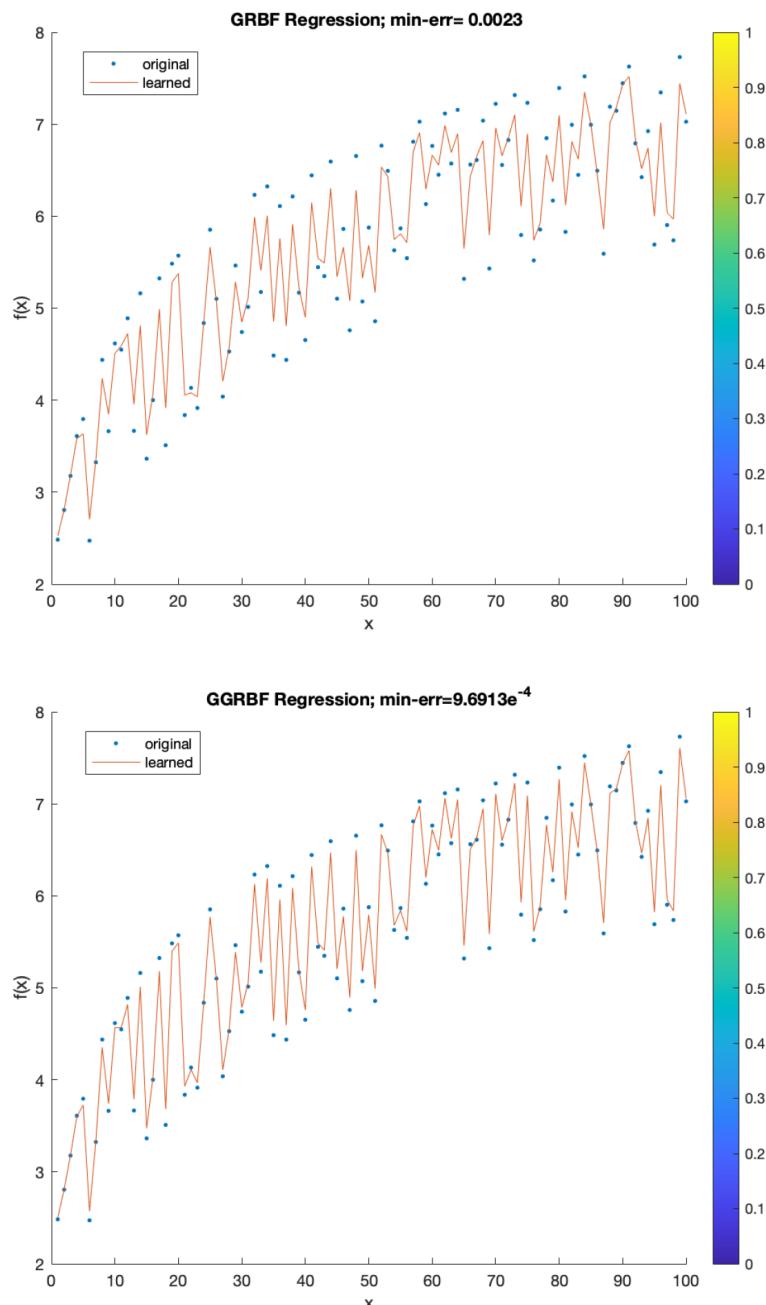
We present the empirical evidence for the application of the Generalized Gaussian Radial Basis Function Kernel on various basic yet important learning architecture routines. The technical details of every experiment presented here are provided in Section 6.

### 5.1. Kernel Regression

We know that kernel regression aims to approximate or interpolate arbitrary continuous functions through the available choices of universal kernel functions on compact subsets of  $\mathbf{R}^d$ . Usually, these functions are radial basis functions that are employed in the kernel sense to achieve the approximation. In the current example given as follows, we leverage the universal nature of both Gaussian Radial Basis Function Kernel (cf. Corollary 4.58 in [21]) and Generalized Gaussian Radial Basis Function Kernel from Theorem 6.

### 5.1.1. Example 1

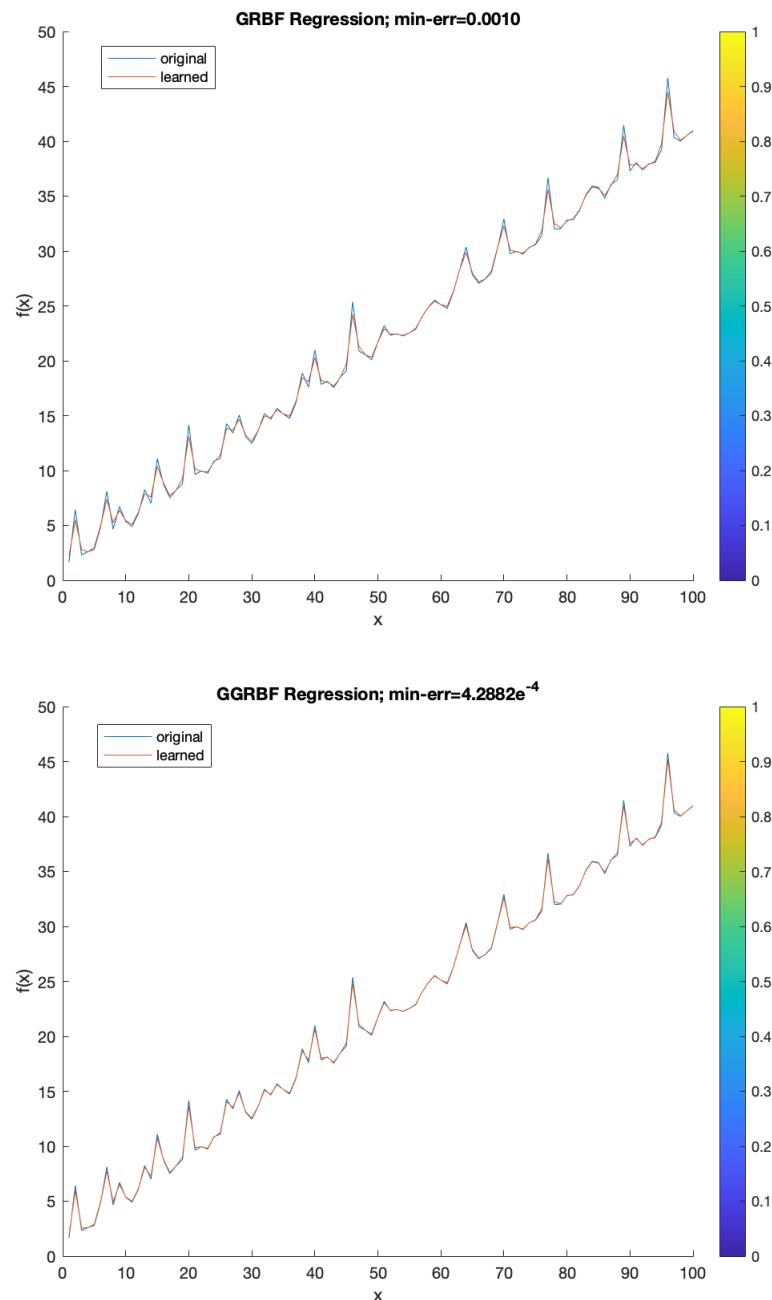
Kernel regression of  $f(x) = e^{\frac{(1-9x^2)}{4}} + \tan x + x^{\frac{1}{6}} + \sin x^{\vartheta(n)}$  is performed via both GRBF and (1) and the respective results are given below in Figure 8. Here,  $\vartheta(n)$  is uniform random distributed number in  $(0, 1)$ ,  $x = n \in \mathbf{Z}_{100}$ .



**Figure 8.** Kernel regression of  $f(x) = e^{\frac{(1-9x^2)}{4}} + \tan x + x^{\frac{1}{6}} + \sin x^{\vartheta(n)}$  with GRBF (L) and GGRBF (R).

### 5.1.2. Example 2

Kernel regression of  $f(x) = e^{\sin x - \sin x^2} + \sqrt{2\pi}|x + \cos \vartheta(n)|$  is performed via both GRBF and (1) and the respective results are given below in Figure 9. Here,  $\vartheta(n)$  is uniform random distributed number in  $(0, 1)$ ,  $x = n \in \mathbf{Z}_{100}$ .



**Figure 9.** Kernel regression of  $f(x) = e^{\sin x - \sin x^2} + \sqrt{2\pi} |x + \cos \theta(n)|$  with GRBF (L) and GGRBF (R).

### 5.2. Support Vector Machine

Let us briefly address how the results derived in Section 3 can be used to understand the analysis of SVM. To that end, we define an important notion in SVMs referred to as *hinge loss function*, denoted by  $L(\cdot, \cdot)$  as:

$$L(y, t) := \max_{y \in Y := \{-1, 1\}} \{0, 1 - yt\} \quad t \in \mathbf{R}. \quad (25)$$

Let  $X \subset \mathbf{R}^d$  be a nonempty open subset and  $P$  be a probability measure on  $X \times Y$ ; then, the risk induced by the hinge loss function is *L-risk*, which is defined with help of  $f : X \rightarrow \mathbf{R}$ -measurable function as:

$$\mathcal{R}_{L,P}(f) := \int_{X \times Y} L(y, f(x)) dP(x, y). \quad (26)$$

Additionally, the *minimal L-risk* is given as

$$\mathcal{R}_{L,P}^* := \inf_{f:X \rightarrow \mathbf{R}} \mathcal{R}_{L,P}(f). \quad (27)$$

It is important to recall here that the RKHSs of universal kernels approximate the minimal *L-risk* losses. It is performed in the following corollary (cf. Corollary 5.29 in [21]).

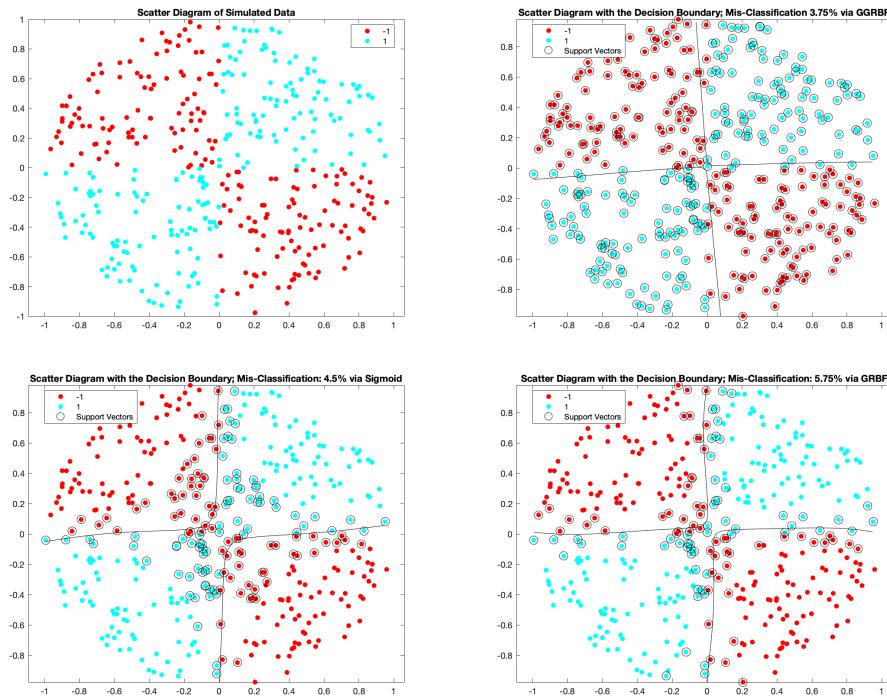
**Corollary 2.** *Let  $X$  be a compact metric space,  $H$  be the RKHS of a universal kernel on  $X$ ,  $P$  be a distribution on  $X \times Y$ , and  $L : X \times Y \times \mathbf{R} \rightarrow [0, \infty)$  be a continuous  $P$ -integrable loss. Then, we have  $\inf_{f \in H} [\mathcal{R}_{L,P}(f)] = \mathcal{R}_{L,P}^*$ .*

Now, consider  $f \in H_{\sigma,\sigma_0,\mathbf{C}^d}$ ; then, let the restriction of  $f$  on  $\mathbf{R}^d$  be denoted by  $f|_{\mathbf{R}^d}$ ; then, the aforementioned result forces us to establish the relation

$$\inf_{f|_{\mathbf{R}^d} \in H_{\sigma,\sigma_0,\mathbf{C}^d}} [\mathcal{R}_{L,P}(f)] = \mathcal{R}_{L,P}^*. \quad (28)$$

The setup presented in (28) enables to have consistent results concerning the classifiers of SVMs, which simply means that the classifiers that are based on minimization of a regularized risk can asymptotically learn in every classification task [22]. With this theory of SVM around the RKHS  $H_{\sigma,\sigma_0,\mathbf{C}^d}$ , we are now ready to ensure the application of the Generalized Gaussian Radial Basis Function Kernel in SVMs.

SVMs are implemented for the standard binary data classification (cf. [23]) data generated in **top left** of Figure 10; this is performed via a different choices of kernels; these kernels include the traditional *Sigmoid* function, the Gaussian Radial Basis Function Kernel, and the Generalized Gaussian Radial Basis Function Kernel in (1).



**Figure 10.** SVM classifier (standard binary classification) via different kernels, namely (respectively) GGRBF (in **top right**), Sigmoid (in **bottom left**) and GRBF (in **bottom right**).

Of the available three kernels—the Generalized Gaussian Radial Basis Function in (1), the Sigmoid, and the Gaussian Radial Basis Function Kernel—the Generalized Gaussian Radial Basis Function Kernel yields the lowest misclassification for the data of 101 sampled points.

### 5.3. Neural Network

The work in this section revolves around the application of *universal approximation theorem* (adapted from [24]) of Cybenko-1989 [25], Hornik-1991 [26], and Pinkus-1999 [27], as follows:

**Theorem 7** (Universal Approximation Theorem in Neural Network). *Let  $\rho : \mathbf{R} \rightarrow \mathbf{R}$  be any continuous function. Let  $\mathcal{N}_d^\rho$  represent the class of feed-forward neural networks with activation function  $\rho$  with  $d$ -neurons in the input layer, one neuron in the output layer, and one hidden layer with an arbitrary number of neurons. Let  $X \subseteq \mathbf{R}^d$  be compact. Then,  $\mathcal{N}_d^\rho$  is dense in  $C(X)$  if and only if  $\rho$  is non-polynomial.*

Now, under the application of the universal approximation theorem, one can have  $\rho(r) = g_{\sigma^2}(r) \exp(g_{\sigma_0^2}(r) - 1)$  in Theorem 7; then, we realize that the Generalized Gaussian Radial Basis Function is an activation function, as it is a non-polynomial function. In the pursuit of an application of the Generalized Gaussian Radial Basis Function, we have following examples that provide the evidence of the optimal results that are yielded by the Generalized Gaussian Radial Basis Function in (1) employment both as in the usage of the *activation function* and the neural net layer in *deep convolution neural nets* (DCNNs).

#### 5.3.1. Activation Function

Consider an  $\alpha$ ReLU activation function, a simple modification of the powerful traditional ReLU activation function from NN (AF-NN), defined as

$$f(x) := \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{if } x \leq 0. \end{cases} \quad (29)$$

In the present experiment, two seven-layered NNs are constructed: one with the activation function defined by (29) and one with Generalized Gaussian Radial Basis Function in (1). Thereafter, the following performance Tables 1 and 2 yield the respective results and comparison for the NNs.

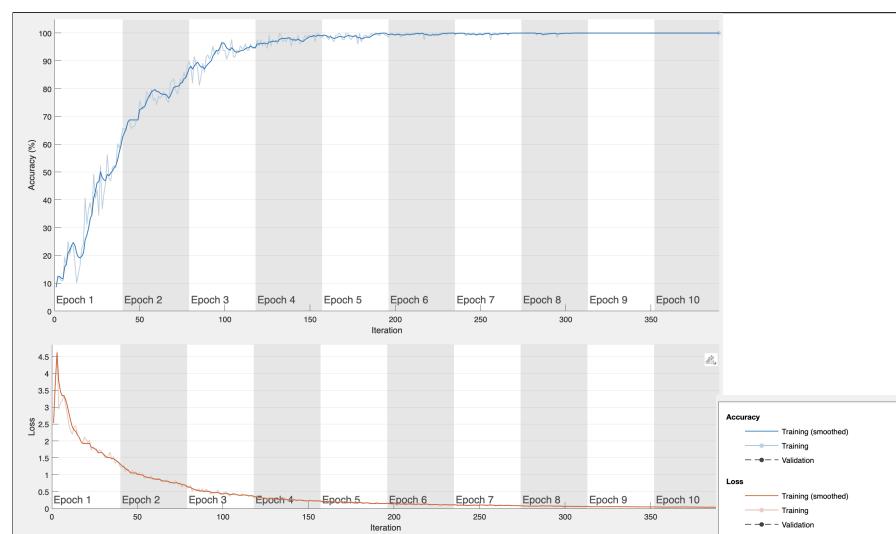
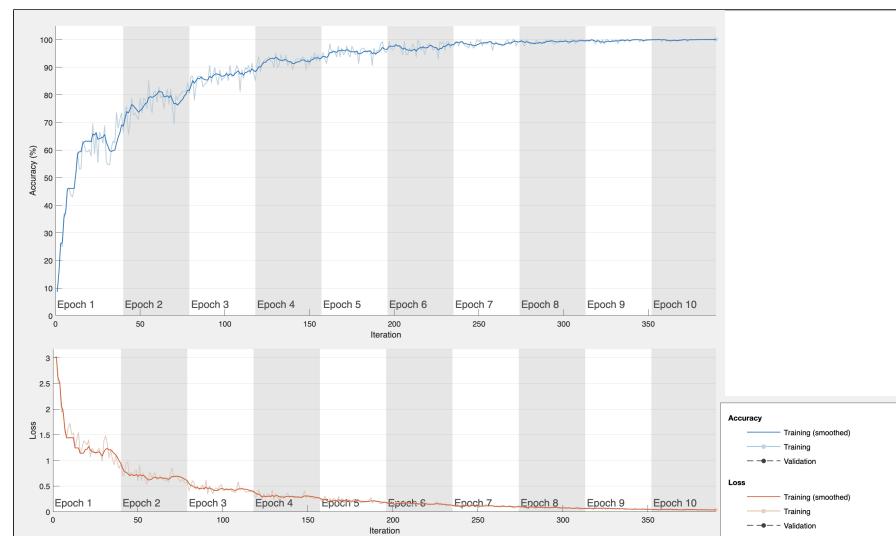
**Table 1.** Performance summary for GGRBF with an accuracy of 97.88%.

Training with GGRBF (on Single CPU)					
Epoch	Iteration	Runtime (hh:mm:ss)	Accuracy (%)	Batch Loss	Learning Rate
1	1	00:00:04	8.59	2.5329	0.0010
2	50	00:00:10	75.78	1.0087	0.0010
3	100	00:00:15	93.75	0.4377	0.0010
4	150	00:00:21	99.22	0.2452	0.0010
6	200	00:00:26	99.22	0.1790	0.0010
7	250	00:00:31	100.00	0.1028	0.0010
8	300	00:00:36	100.00	0.0681	0.0010
9	350	00:00:41	100.00	0.0458	0.0010
10	390	00:00:45	100.00	0.0382	0.0010

Furthermore, the training progress data for the individual neural network architectures are presented next. The legends for *accuracy* and *loss* are same for both of the neural network architectures and their respective information is present in between the Figures 11 and 12.

**Table 2.** Performance summary for (29) with accuracy of 93.48%.

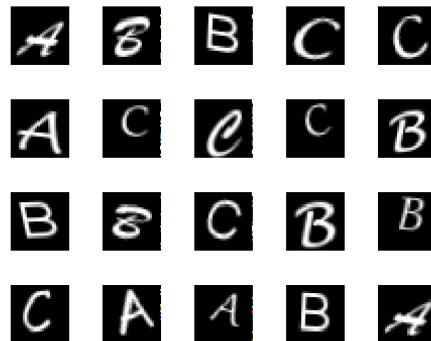
Training with (29) (on Single CPU)					
Epoch	Iteration	Runtime (hh:mm:ss)	Accuracy (%)	Batch Loss	Learning Rate
1	1	00:00:08	8.59	2.9976	0.0010
2	50	00:00:15	78.91	0.6053	0.0010
3	100	00:00:19	86.72	0.4108	0.0010
4	150	00:00:25	90.62	0.3122	0.0010
6	200	00:00:30	96.88	0.1632	0.0010
7	250	00:00:36	99.22	0.0695	0.0010
8	300	00:00:41	99.22	0.0778	0.0010
<b>9</b>	<b>350</b>	<b>00:00:46</b>	<b>100.00</b>	<b>0.0390</b>	<b>0.0010</b>
<b>10</b>	<b>390</b>	<b>00:00:49</b>	<b>100.00</b>	<b>0.0355</b>	<b>0.0010</b>

**Figure 11.** Training progress map for NN architecture with GGRBF as AF.**Figure 12.** Training progress map for NN architecture with (29) as AF.

### 5.3.2. DCNN

The above experiments demonstrate the concrete functionality of the Generalized Gaussian Radial Basis Function in (1) as an activation function; this finding further motivates the

performance of the deep convolution neural net. Therefore, in pursuit of this, a typical 7-layered deep convolution neural net is constructed; it has an activation function defined by (29) and one with (1). In the following experiment given in Figures 13 and 14, the training data contain  $1500-28 \times 28$  gray-scale letter images of A, B, and C in a 4-D array.

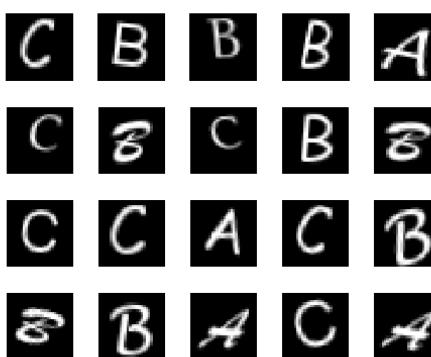


```

Training on single CPU.
Initializing input data normalization.
=====
| Epoch | Iteration | Time Elapsed | Mini-batch | Mini-batch | Base Learning |
|       |           | (hh:mm:ss) | Accuracy | Loss       | Rate          |
|-----|
|   1   |      1   | 00:00:03 | 36.72% | 1.1980 | 0.0100 |
|   5   |     50   | 00:00:09 | 98.44% | 0.0645 | 0.0100 |
|  10   |    100   | 00:00:13 | 100.00% | 0.0043 | 0.0100 |
|  14   |    150   | 00:00:18 | 100.00% | 0.0019 | 0.0100 |
|  19   |    200   | 00:00:23 | 100.00% | 0.0028 | 0.0100 |
|  23   |    250   | 00:00:27 | 100.00% | 0.0012 | 0.0100 |
|  28   |    300   | 00:00:32 | 100.00% | 0.0013 | 0.0100 |
|  30   |    330   | 00:00:35 | 100.00% | 0.0009 | 0.0100 |
|-----|
accuracy =
0.9633

```

**Figure 13.** Accuracy of 96.33% registered with GGRBF in (1) as a DCNN neural layer.



```

Training on single CPU.
Initializing input data normalization.
=====
| Epoch | Iteration | Time Elapsed | Mini-batch | Mini-batch | Base Learning |
|       |           | (hh:mm:ss) | Accuracy | Loss       | Rate          |
|-----|
|   1   |      1   | 00:00:00 | 37.50% | 1.0878 | 0.0100 |
|   5   |     50   | 00:00:01 | 100.00% | 0.0588 | 0.0100 |
|  10   |    100   | 00:00:03 | 100.00% | 0.0145 | 0.0100 |
|  14   |    150   | 00:00:05 | 100.00% | 0.0077 | 0.0100 |
|  19   |    200   | 00:00:06 | 100.00% | 0.0067 | 0.0100 |
|  23   |    250   | 00:00:08 | 100.00% | 0.0043 | 0.0100 |
|  28   |    300   | 00:00:09 | 100.00% | 0.0039 | 0.0100 |
|  30   |    330   | 00:00:10 | 100.00% | 0.0030 | 0.0100 |
|-----|
accuracy =
0.9187

```

**Figure 14.** Accuracy of 91.87% registered with (29) as a DCNN neural layer.

## 6. Results and Technical Details

Now that we have successfully constructed the Hilbert function space theory around the Generalized Gaussian Radial Basis Function, when it is viewed as the  $L^2$ –measure, we immediately employ it in various basic artificial intelligence learning architectures. The learning architectures, namely *kernel regression*, *SVM*, and *DCNN*, that are considered here are quite basic yet well-versed in the field of machine learning and artificial intelligence. These can be regarded as the building blocks of several other artificial intelligence routines. Following Table 3 is essentially a compilation of the results documented for the various experiments we have discussed so far in comparison with the Generalized Gaussian Radial Basis Function Kernel as opposed to its available counterpart(s), where none of hyperparameters were optimized for the available kernel functions.

**Table 3.** Compilation of technical results.

AI Learning Architecture	Activation Function	Reference	Min. Error	Misclass. %	Accuracy %	Runtime (s)
Kernel Regression	GRBF	Figure 8	0.00230	-	-	-
Kernel Regression	<b>GGRBF</b>	Figure 8	<b>0.00097</b>	-	-	-
Kernel Regression	GRBF	Figure 9	0.00100	-	-	-
Kernel Regression	<b>GGRBF</b>	Figure 9	<b>0.00043</b>	-	-	-
SVM	GRBF	Figure 10	-	5.75	94.25	-
SVM	Sigmoid	Figure 10	-	4.50	95.50	-
SVM	<b>GGRBF</b>	Figure 10	-	<b>3.75</b>	<b>96.25</b>	-
AF in NN	<b>GGRBF</b>	Table 1	-	<b>2.12</b>	<b>97.88</b>	<b>0.43077</b>
AF in NN	$\alpha$ ReLU	Table 2	-	6.52	93.48	1.8935
DCNN	<b>GGRBF</b>	Figure 13	-	<b>3.67</b>	<b>96.33</b>	-
DCNN	$\alpha$ ReLU	Figure 14	-	8.13	91.87	-

The empirical evidence recorded here simply directs us to conclude that the results prompted by the Generalized Gaussian Radial Basis Function outperform every considered learning architecture routine. A detailed discussion of the empirical evidence collected is presented here:

1. We considered two experiments for the *kernel regression* of the functions mentioned in Figures 8 and 9. These functions, as easily can be seen, are not well-behaved functions, since they involve transcendental behavior along with function discontinuities (due to  $\tan x$  or  $x^{1/6}$ , etc.) followed by uniform randomness due to the  $\vartheta(n)$  of 101–random points. In both of the experiments of this nature, we record the minimum error obtained by employing either the Gaussian Radial Basis Function Kernel or the Generalized Gaussian Radial Basis Function Kernel. From this experiment, we learn that the minimum error recorded in both of them is remarkably small, in the order of  $\sim 10^{-4}$ , if we use the Generalized Gaussian Radial Basis Function Kernel in contrast to the Gaussian Radial Basis Function Kernel, which provides a minimum error in the order of  $\sim 10^{-3}$ .
2. Next, we consider the experiment of *support vector machines for binary classification*, in which we generated a random set of points (one can think of them as data) within the unit circle. We identified the random points spread across the circle with the positive class if it lies either in the first or the third quadrant; otherwise, we consider it to be within the negative class. After we generate and identify these datasets, we then identify the support vectors and the decision boundary by employing three (custom) kernel functions over MATLAB R2023b: *Gaussian Radial Basis Function Kernel*, *Sigmoid Kernel*, and *Generalized Gaussian Radial Basis Function Kernel*. We then register the out-of-sample misclassification rate by using *10-fold cross-validation* from the aforementioned kernel functions. Here, we immediately learn that the misclassification yielded by Gaussian Radial Basis Function Kernel is comparatively high among all

three employed kernel functions; the Generalized Gaussian Radial Basis Function Kernel yields the lowest misclassification rate: *upward of 96%*.

3. For the *activation function* application in the domain of neural networks, we consider a practical example of pattern recognition for the synthetic hand-written English letter dataset in the form of a  $4 - D$  array (`lettersTrainSet` in MATLAB). As already mentioned in the respective subsection, two 7-layered neural nets were constructed whose layer names are given as follows:

- (a) Image input;
- (b) 2D convolution;
- (c) Batch normalization;
- (d) (29) vs. Generalized Gaussian Radial Basis Function;
- (e) Ten fully connected layers;
- (f) Softmax;
- (g) Classification layer.

It should be noted that, from the above 7-layered neural network constructed, both of the neural network training experiments came to a complete stop when the maximum epoch of 30 amounting to 330 iterations were finished. Additionally, the sequence of inserting the neural net layer was the same as that which was given above from (a)–(g). Once the desired neural net is set with both of the activation functions, we immediately learn that the neural net containing the activation function of the Generalized Gaussian Radial Basis Function Kernel as one of its layers simply outperforms its opponent, i.e., the neural net layer containing the activation function of (29).

4. With the establishment of the basics of the neural network architecture in the previous listing point, we train and execute a deep convolution neural net for the representative images of the alphabet characters A, B, and C. A list of the seven layers involved in the construction of a deep convolution neural network is presented as follows:

- (a) Image input;
- (b) 2D convolution;
- (c) (29) vs. Generalized Gaussian Radial Basis Function;
- (d) A  $2 \times 2$  max pooling with stride [2 2] and padding [0 0 0];
- (e) Three fully connected layers;
- (f) Softmax;
- (g) Classification layer.

Again, the sequence of inserting the neural net layers was the same as what is given above from (a)–(g) to construct the present DCNN. Similarly, the result for the successful classification performances by the DCNN which contains the activation layer of Generalized Gaussian Radial Basis Function Kernel achieves an accuracy upward of 96%. However, we fail to see the matchable competitiveness by the DCNN, which contains the (29) as the activation layer.

## 7. Future Directions

### *Eigen Function Expansion of GGRBF*

We recall the *Mercer's theorem* from ([12], Theorem 4.2, Page 96).

**Theorem 8** (Mercer's Theorem). *Let  $(\mathcal{X}, \mu)$  be a finite measure space and  $k \in L_\infty(\mathcal{X}^2, \mu^2)$  be a kernel, such that  $T_k : L_2(\mathcal{X}, \mu) \rightarrow L_2(\mathcal{X}, \mu)$  is a definite positive. Let  $\{\phi_i\}_i \in L_2(\mathcal{X}, \mu)$  be the normalized eigenfunctions of  $T_k$  associated with the eigenvalues  $\{\Lambda_i\}_i$ . Then,*

1. *the eigenvalues  $\{\Lambda\}_i$  are absolutely summable*
2.  *$k(\mathbf{x}, \mathbf{x}') = \sum_{i=0}^{\infty} \Lambda_i \phi_i(\mathbf{x}) \phi_i(\mathbf{x}')^*$  holds  $\mu^2$  almost everywhere, where the series converges absolutely and uniformly  $\mu^2$  almost everywhere.*

**Example 3.** With the application of Theorem 8, we can provide the eigen function decomposition of  $K_\sigma(x, z) = e^{-\sigma^2(x-z)^2}$  for  $x, z \in \mathbf{R}$ , that is:

$$e^{-\sigma^2(x-z)^2} = \sum_{i=0}^{\infty} \Lambda_i \phi_i(x) \phi_i(z), \quad (30)$$

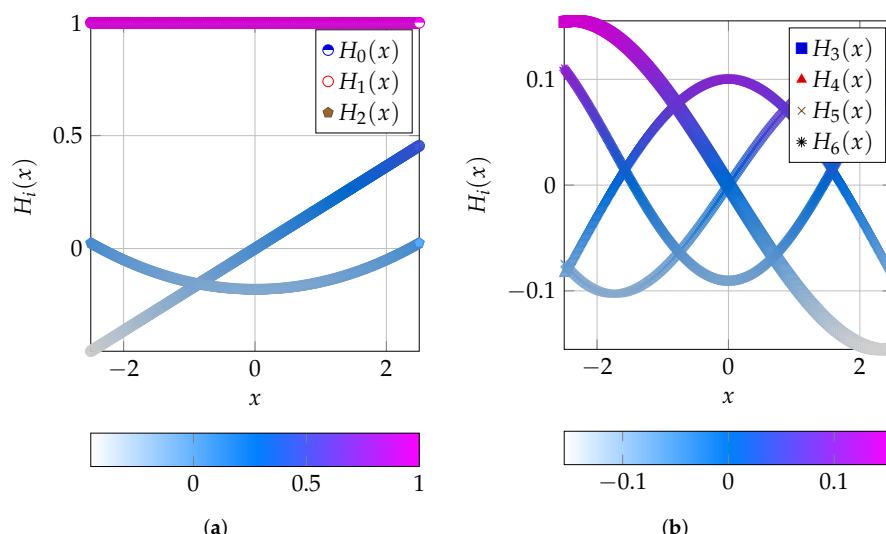
$$\Lambda_i = \frac{\alpha \sigma^{2i}}{\left( \frac{\alpha^2}{2} \left( 1 + \sqrt{1 + \left( \frac{2\sigma}{\alpha} \right)^2} \right) + \frac{\sigma^2}{2} \right)^{i+1/2}}, \quad (31)$$

$$\phi_i(x) = \frac{\sqrt[8]{1 + \left( 1 + \frac{2\sigma}{\alpha} \right)^2}}{\sqrt{2^i i!}} e^{-\left( \sqrt{1 + \left( \frac{2\sigma}{\alpha} \right)^2} - 1 \right) \frac{\alpha^2 x^2}{2}} H_i \left( \sqrt[4]{1 + \left( \frac{2\sigma}{\alpha} \right)^2} \alpha x \right). \quad (32)$$

The expression  $\{H_i(\bullet)\}_i$  in (32) are the Hermite polynomials which are  $L_2$  orthonormal against the weight  $\frac{\alpha}{\pi} e^{-\alpha^2 x^2}$ , that is:

$$\int_{\mathbf{R}} \phi_n(x) \phi_m(x) \frac{\alpha}{\pi} e^{-\alpha^2 x^2} dx = \delta_{nm}. \quad (33)$$

We have the graphical representation of first seven Hermite polynomials in Figure 15a,b.



**Figure 15.** First seven Hermite polynomials. (a) Hermite polynomials  $H_i(x)$  for  $i = 0, 1, 2$ ; (b) Hermite polynomials  $H_i(x)$  for  $i = 3, 4, 5, 6$ .

Based on the Hermite polynomials [28] and as its application for the eigen decomposition analysis for the Gaussian Radial Basis Function Kernel, we have the following promising future research direction. In the spirit of the application of Mercer's theorem, we know the eigen function decomposition of the Gaussian Radial Basis Function Kernel. However, presently, we are not fortunate enough to have such a decomposition for the Generalized Gaussian Radial Basis Function Kernel. A preliminary investigation towards the desired eigen function decomposition of the Generalized Gaussian Radial Basis Function Kernel (from [29–31]) directs us toward the incorporation of a new variety of functions, as defined in (34).

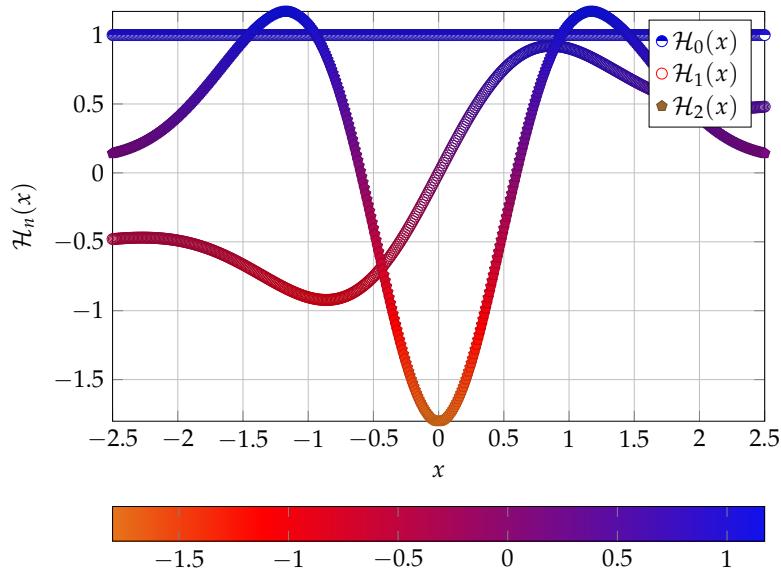
$$\mathcal{H}_n(x) := (-1)^n e^{ax^2} e^{-e^{-bx^2} + 1} \frac{d^n}{dx^n} \left( e^{-ax^2} e^{e^{-bx^2} - 1} \right) \quad (34)$$

into our desired eigen function decomposition analysis. Here,  $a > 0$  and  $b \geq 0$ ; therefore, if  $a = 1$  &  $b = 0$ , then  $\mathcal{H}_i = H_i$ . The first two expressions for  $\mathcal{H}_n(x)$  are explicitly given as:

$$\mathcal{H}_1(x) = 2ax + 2bxe^{-bx^2} \quad (35)$$

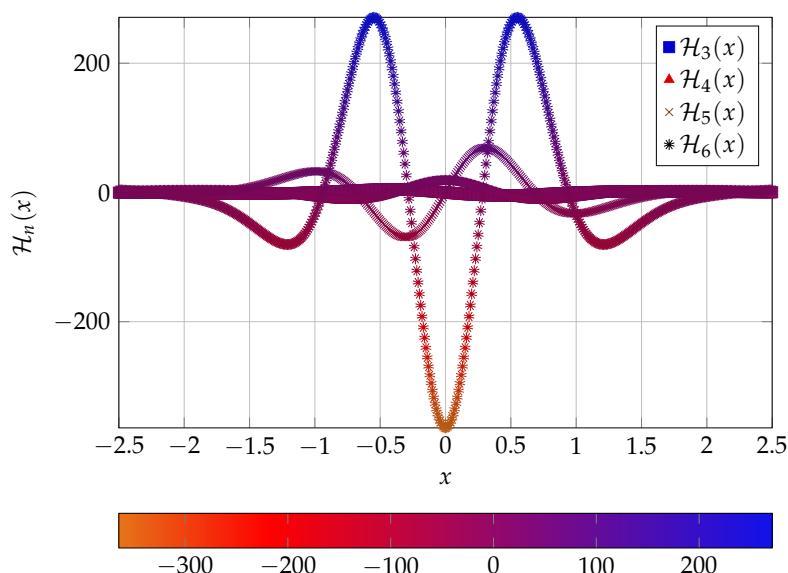
$$\mathcal{H}_2(x) = -2a + 4b^2x^2e^{-bx^2} - 2be^{-bx^2} + (2ax + 2bxe^{-bx^2})^2. \quad (36)$$

The real constants  $a$  and  $b$  present in (34) corresponds to the respective constants present in (1) as  $a = \sigma^2$  and  $b = \sigma_0^2$ . Next, see the graphical presentation of the first seven functions from the family defined in (34) in Figures 16 and 17.



**Figure 16.** Graph of  $\mathcal{H}_n(x)$  in (34) for  $n = 0, 1, 2$  with  $a = 0.091$  &  $b = 0.81$ .

**Future Direction 1.** We remain in a position lacking in knowledge of eigenvalues for the Generalized Gaussian Radial Basis Function Kernel. Additionally, we still need to investigate whether the function introduced in (34) is an orthonormal in the sense that the traditional Hermite polynomials are. Therefore, it will be interesting to understand the analysis of the function given in (34).



**Figure 17.** Graph of  $\mathcal{H}_n(x)$  in (34) for  $n = 3, 4, 5, 6$  with  $a = 0.091$  &  $b = 0.81$ .

## 8. Discussion

In the present manuscript, we explore and introduce the Hilbert function space theory; in particular, the RKHS was generated by the Generalized Gaussian Radial Basis Function  $g_{\sigma^2}(r) \exp(g_{\sigma_0^2}(r) - 1)$  as an  $L^2$ -measure, followed by an interesting application of it in the learning architectures of machine learning and deep learning. Throughout this article, we have learned from practical results that it is indeed beneficial to employ the Generalized Gaussian Radial Basis Function Kernel; here, one might achieve phenomenal results compared to those derived using its corresponding counterparts. Even though it is quite advantageous to use the Generalized Gaussian Radial Basis Function Kernel in learning architectures, we also encounter a few shortcomings in this endeavor. Oftentimes, the result yielded by the Generalized Gaussian Radial Basis Function is similar to the results that one might obtain using the preexisting and well-studied activation functions, such as *Gaussian error linear unit* (GELU) (cf. [32]), whose operation on  $x$  is given as

$$\text{GELU}(x) := \frac{x}{2} \left( 1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right). \quad (37)$$

For instance, the comparison of the results based on choice of either Generalized Gaussian Radial Function Kernel or Gaussian error linear unit as an activation function is given in Tables 4 and 5, respectively. We see that the results from both of the cases are comparable in terms of acquiring an accuracy rate. The key point to note here is that the Generalized Gaussian Radial Basis Function Kernel is a brand-new novel mathematical function to operate on; hence, it demands a detailed, in-depth investigation, which is initiated by its RKHS in this article, before it can be exploited to its full potential on the venues of artificial intelligence. Out of the other few major concerns in connection with the Generalized Gaussian Radial Basis Function Kernel is its mathematical structure, i.e., the presence of an extra exponential of an exponential of an argument. Hence, the simple presence of this can often slow the training process of learning architectures in which the Generalized Gaussian Radial Basis Function Kernel is employed; therefore, this might pose a challenge in attaining the desired results.

**Table 4.** Performance summary for GGRBF in comparison with (37) with accuracy of **98.18%**.

Training with GGRBF (on a Single CPU)					
Epoch	Iteration	Runtime (hh:mm:ss)	Accuracy (%)	Batch Loss	Learning Rate
1	1	00:00:00	11.72	2.5329	0.0010
2	50	00:00:04	72.66	1.0087	0.0010
3	100	00:00:09	84.38	0.4377	0.0010
4	150	00:00:13	89.06	0.2452	0.0010
6	200	00:00:18	98.44	0.1790	0.0010
7	250	00:00:23	100.00	0.1028	0.0010
8	300	00:00:27	100.00	0.0681	0.0010
9	350	00:00:32	100.00	0.0458	0.0010
11	400	00:00:36	100.00	0.0382	0.0010
12	450	00:00:41	100.00	0.0382	0.0010
13	500	00:00:45	100.00	0.0382	0.0010
14	546	00:00:50	100.00	0.0382	0.0010

**Table 5.** Performance summary for (37) in comparison with GGRBF with accuracy of 98.06%.

Training with (37) (on a Single CPU)					
Epoch	Iteration	Runtime (hh:mm:ss)	Accuracy (%)	Batch Loss	Learning Rate
1	1	00:00:00	10.94	2.7556	0.0010
2	50	00:00:04	82.81	0.5065	0.0010
3	100	00:00:07	97.66	0.1923	0.0010
4	150	00:00:11	100.00	0.0738	0.0010
6	200	00:00:14	100.00	0.0446	0.0010
7	250	00:00:28	100.00	0.0262	0.0010
8	300	00:00:21	100.00	0.0211	0.0010
9	350	00:00:25	100.00	0.0173	0.0010
11	400	00:00:29	100.00	0.0099	0.0010
12	450	00:00:32	100.00	0.0085	0.0010
13	500	00:00:36	100.00	0.0085	0.0010
14	546	00:00:39	100.00	0.0068	0.0010

**Funding:** This research was funded by The Office of Dean, the Office of Research, Scholarship, and Sponsored Programs, Robert R. Muntz Library at The University of Texas at Tyler, TX-75799, USA.

**Data Availability Statement:** The coding framework presented in this article was executed on MATLAB R2023bat (<https://www.mathworks.com/products/matlab.html>, accessed on 15 February 2024).

**Acknowledgments:** The author would like to thank the reviewer for their precious time in reviewing the present manuscript. The author acknowledges the support of *Drishy Singh, 4th Year MSc, Department of Mathematics, Babasaheb Bhimrao Ambedkar University, Lucknow, Uttar Pradesh, India*. She provided the explicit expression of the first two functions in (35) and (36) from (34). Further analysis of function in (34) was extended due to these important results. The author wishes to thank the referees for their valuable comments.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Baddoo, P.J.; Herrmann, B.; McKeon, B.J.; Brunton, S.L. Kernel learning for robust Dynamic Mode Decomposition: Linear and Nonlinear disambiguation optimization. *Proc. R. Soc. A* **2022**, *478*, 20210830. [[CrossRef](#)] [[PubMed](#)]
2. Gianola, D.; Van Kaam, J.B. Reproducing Kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **2008**, *178*, 2289–2303. [[CrossRef](#)] [[PubMed](#)]
3. Attia, N.; Akgül, A.; Seba, D.; Nour, A. Reproducing kernel Hilbert space method for the numerical solutions of fractional cancer tumor models. *Math. Methods Appl. Sci.* **2023**, *46*, 7632–7653. [[CrossRef](#)]
4. Mathematical Functions Power Artificial Intelligence. Available online: <https://nap.nationalacademies.org/resource/other/deps/illustrating-math/interactive/mathematical-functions-power-ai.html> (accessed on 15 February 2024).
5. Mathematics and Statistics of Weather Forecasting. Available online: <https://nap.nationalacademies.org/resource/other/deps/illustrating-math/interactive/mathematics-and-statistics-of-weather-forecasting.html> (accessed on 15 February 2024).
6. Kalman, B.L.; Kwasny, S.C. Why tanh: Choosing a sigmoidal function. In Proceedings of the 1992 IJCNN International Joint Conference on Neural Networks, Baltimore, MD, USA, 7–11 June 1992; Volume 4, pp. 578–581.
7. Lu, L.; Shin, Y.; Su, Y.; Karniadakis, G.E. Dying ReLU and initialization: Theory and numerical examples. *arXiv* **2019**, arXiv:1903.06733.
8. Fasshauer, G.E. *Meshfree Approximation Methods with MATLAB*; World Scientific: Singapore, 2007; Volume 6.
9. Brunton, S.L.; Kutz, J.N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2022.
10. Stein, M.L. *Interpolation of Spatial Data: Some Theory for Kriging*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1999.
11. Rosenfeld, J.A.; Russo, B.; Kamalapurkar, R.; Johnson, T.T. The occupation kernel method for nonlinear system identification. *arXiv* **2019**, arXiv:1909.11792.
12. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
13. Park, J.; Sandberg, I.W. Universal approximation using radial-basis-function networks. *Neural Comput.* **1991**, *3*, 246–257. [[CrossRef](#)] [[PubMed](#)]
14. Singh, H. A new kernel function for better AI methods. In Proceedings of the 2023 Spring Eastern Sectional Meeting, Virtual, 1–2 April 2023.

15. Karimi, N.; Kazem, S.; Ahmadian, D.; Adibi, H.; Ballestra, L. On a generalized Gaussian radial basis function: Analysis and applications. *Eng. Anal. Bound. Elem.* **2020**, *112*, 46–57. [[CrossRef](#)]
16. Aronszajn, N. Theory of reproducing kernels. *Trans. Am. Math. Soc.* **1950**, *68*, 337–404. [[CrossRef](#)]
17. Steinwart, I.; Hush, D.; Scovel, C. An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Trans. Inf. Theory* **2006**, *52*, 4635–4643. [[CrossRef](#)]
18. Barnes, E.W.V. The asymptotic expansion of integral functions defined by Taylor’s series. *Philos. Trans. R. Soc. London. Ser. A Contain. Pap. A Math. Phys. Character* **1906**, *206*, 249–297.
19. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; US Government Printing Office: Washington, DC, USA, 1968; Volume 55.
20. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*, 7th ed.; Academic Press: Cambridge MA, USA; Elsevier: Amsterdam, The Netherlands, 2007.
21. Christmann, A.; Steinwart, I. *Support Vector Machines*; Springer: Berlin/Heidelberg, Germany, 2008.
22. Steinwart, I. Consistency of support vector machines and other regularized kernel classifiers. *IEEE Trans. Inf. Theory* **2005**, *51*, 128–142. [[CrossRef](#)]
23. Devroye, L.; Györfi, L.; Lugosi, G. *A Probabilistic Theory of Pattern Recognition*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 31.
24. Kidger, P.; Lyons, T. Universal approximation with deep narrow networks. In Proceedings of the Conference on Learning Theory, Graz, Austria, 9–12 July 2020; Proceedings of Machine Learning Research; pp. 2306–2327.
25. Cybenko, G. Approximation by Superpositions of a Sigmoidal Function. *Math. Control. Signals Syst.* **1989**, *2*, 303–314. [[CrossRef](#)]
26. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257. [[CrossRef](#)]
27. Pinkus, A. Approximation theory of the MLP model in neural networks. *Acta Numer.* **1999**, *8*, 143–195.
28. Hermite, M. *Sur un Nouveau Développement en Série des Fonctions*; Imprimerie de Gauthier-Villars: 1864; Cambridge University Press: Cambridge, UK, 2011.
29. Fasshauer, G.E.; McCourt, M.J. Stable evaluation of Gaussian radial basis function interpolants. *Siam J. Sci. Comput.* **2012**, *34*, A737–A762. [[CrossRef](#)]
30. Rasmussen, C.E.; Williams, C.K. *Gaussian Processes for Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 1.
31. Zhu, H.; Williams, C.K.; Rohwer, R.; Morciniec, M. *Gaussian Regression and Optimal Finite Dimensional Linear Models*; Neural Networks and Machine Learning; Springer: Berlin/Heidelberg, Germany, 1997.
32. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (Gelus). *arXiv* **2016**, arXiv:1606.08415.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.