An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels*†

Ingo Steinwart, Don Hush, Clint Scovel
Modeling, Algorithms and Informatics Group, CCS-3
Los Alamos National Laboratory
{ingo,dhush,jcs}@lanl.gov

February 6, 2006

Abstract

Although Gaussian RBF kernels are one of the most often used kernels in modern machine learning methods such as support vector machines (SVMs), little is known about the structure of their reproducing kernel Hilbert spaces (RKHSs). In this work we give two distinct explicit descriptions of the RKHSs corresponding to Gaussian RBF kernels and discuss some consequences. Furthermore, we present an orthonormal system for these spaces. Finally we discuss how our results can be used for analyzing the learning performance of SVMs.

Index Terms: Learning Theory, Support Vector Machines, Gaussian RBF Kernels

1 Introduction

In recent years support vector machines and related kernel-based algorithms (see e.g. [1] for an introduction) have become the state-of-the-art methods for many machine learning problems. The common feature of these methods is that they are based on an optimization problem over a reproducing kernel Hilbert space (RKHS). If the underlying input space X of the machine learning problem has a specific structure, e.g. text strings or DNA sequences, one often uses a RKHS which is suitable to this structure (see e.g. [2] for a recent and thorough overview). If however X is a subset of $\mathbb{R}^d$ then the commonly recommended choice are the RKHSs of the Gaussian RBF kernels (see e.g. [3]). Although there has been substantial progress in understanding these RKHSs and their role in the learning process (see e.g. [4] and [5]) some simple questions are still open. For example, it is still unknown which functions are contained in these RKHSs, how the corresponding norms can be computed, and how the RKHSs for different widths correlate to each other. The aim of this paper is to answer these questions. In addition we discuss how our results can be used to bound the approximation error function of SVMs which plays a crucial role in the analysis of the learning performance of these learning algorithms.

The rest of the paper is organized as follows. In Section 2 we recall the definition and basic facts on kernels and RKHSs. In Section 3 we present our main results and discuss their consequences. Finally, Section 4 contains the proofs of the main theorems.

*Los Alamos Unclassified Report LA-UR 04-8274
†Submitted to *IEEE Transactions on Information Theory* on 12/06/04. Revised version submitted on 02/06/06.

2 Preliminaries

So far, in the machine learning literature only $\mathbb{R}$-valued kernels have been considered. However, to describe the reproducing kernel Hilbert space (RKHS) of Gaussian kernels we will use $\mathbb{C}$-valued kernels and therefore we recall the basic facts on RKHSs for both cases (see e.g. [6], [7], and [8]). To this end let us first recall that for a complex number $z = x + iy \in \mathbb{C}$, $x, y \in \mathbb{R}$, its conjugate is defined by $\bar{z} := x - iy$ and its absolute value is $|z| := \sqrt{z\bar{z}} = \sqrt{x^2 + y^2}$. In particular we have $\bar{x} = x$ and $|x| = \sqrt{x^2}$ for all $x \in \mathbb{R}$. Furthermore, we use the symbol $\mathbb{K}$ whenever we want to treat the real and the complex case simultaneously. For example, a $\mathbb{K}$-Hilbert space is a real Hilbert space when $\mathbb{K} = \mathbb{R}$ and a complex one when $\mathbb{K} = \mathbb{C}$. Recall, that in the latter case the inner product $\langle ., . \rangle$ is sesqui-linear and Hermitian. This fact forces us to be a bit pedantic with the ordering in inner products such as in the following definition.

Definition 2.1 Let X be a non-empty set. Then a function $k : X \times X \to \mathbb{K}$ is called a *kernel* on X if there exists a $\mathbb{K}$-Hilbert space H and a map $\Phi : X \to H$ such that for all $x, x' \in X$ we have

$$k(x, x') = \langle \Phi(x'), \Phi(x) \rangle . \tag{1}$$

We call Φ a *feature map* and H a *feature space* of k.

Note that in the real case condition (1) can be replaced by the well-known equation $k(x, x') = \langle \Phi(x), \Phi(x') \rangle$. In the complex case however, $\langle ., . \rangle$ is Hermitian and hence (1) is equivalent to $k(x, x') = \overline{\langle \Phi(x), \Phi(x') \rangle}$.

Given a kernel neither the feature map nor the feature space are uniquely determined. However, one can always construct a canonical feature space, namely the RKHS. Let us now recall the basic theory of these spaces.

Definition 2.2 Let $X \neq \emptyset$ and H be a Hilbert function space over X, i.e. a Hilbert space which consists of functions mapping from X into $\mathbb{K}$.

i) The space H is called a *reproducing kernel Hilbert space (RKHS)* over X if for all $x \in X$ the *Dirac functional* $\delta_x : H \to \mathbb{K}$ defined by $\delta_x(f) := f(x)$, $f \in H$, is continuous.

ii) A function $k : X \times X \to \mathbb{K}$ is called a *reproducing kernel* of H if we have $k(., x) \in H$ for all $x \in X$ and the *reproducing property*

$$f(x) = \langle f, k(., x) \rangle$$

holds for all $f \in H$ and all $x \in X$.

Recall that reproducing kernel Hilbert spaces have the remarkable and important property that norm convergence implies pointwise convergence. More precisely, let H be a RKHS, $f \in H$, and $(f_n) \subset H$ be a sequence with $\|f_n - f\|_H \to 0$ for $n \to \infty$. Then for all $x \in X$ we have

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \delta_x(f_n) = \delta_x(f) = f(x) . \tag{2}$$

Furthermore, reproducing kernels are actually kernels in the sense of Definition 2.1 since $\Phi : X \to H$ defined by $\Phi(x) := k(., x)$ is a feature map of k. Moreover, the reproducing property says that each Dirac functional can be represented by the reproducing kernel. Consequently, a Hilbert function space H that has a reproducing kernel k is always a RKHS. The following theorem shows that conversely, every RKHS has a (unique) reproducing kernel and that this kernel can be determined by the Dirac functionals.

Theorem 2.3 *Let H be a RKHS over X. Then $k : X \times X \to \mathbb{K}$ defined by $k(x, x') := \langle \delta_x, \delta_{x'} \rangle$, $x, x' \in X$, is the only reproducing kernel of H. Furthermore, if $(e_i)_{i\in I}$ is an orthonormal basis (ONB) of H then for all $x, x' \in X$ we have*

$$k(x, x') = \sum_{i\in I} e_i(x)\overline{e_i(x')}\,, \tag{3}$$

where the convergence is absolute.

For a proof of the above theorem we refer to [9, p. 42ff] and [8, p. 38ff]. Note that the ONB in Theorem 2.3 is not necessarily countable. However, recall that RKHSs over separable metric spaces having a continuous kernel are always separable and hence all their ONBs are countable. In particular, the RKHSs of Gaussian RBF kernels always have countable ONBs.

Theorem 2.3 shows that a RKHS uniquely determines its reproducing kernel. The following theorem (see [8, p. 20–23] for a proof) states that conversely every kernel has a unique RKHS.

Theorem 2.4 *Let $X \neq \emptyset$ and k be a kernel over X with feature space H_0 and feature map $\Phi_0 : X \to H_0$. Then*

$$H := \big\{ \langle w, \Phi_0(.)\rangle_{H_0} : w \in H_0 \big\} \tag{4}$$

equipped with the norm

$$\|f\|_H := \inf\big\{ \|w\|_{H_0} : w \in H_0 \text{ with } f = \langle w, \Phi_0(.)\rangle_{H_0} \big\} \tag{5}$$

is the only RKHS of k. In particular both definitions are independent of the choice of H_0 and Φ_0 and the operator $V : H_0 \to H$ defined by

$$Vw := \langle w, \Phi_0(.)\rangle_{H_0}\,, \qquad\qquad w \in H_0$$

is a metric surjection, i.e. $V\mathring{B}_{H_0} = \mathring{B}_H$, where $\mathring{B}_{H_0}$ and $\mathring{B}_H$ are the open unit balls of H_0 and H, respectively.

Finally, the following result proved in Section 4 relates the $\mathbb{C}$-RKHS with the $\mathbb{R}$-RKHS of a real-valued kernel.

Corollary 2.5 *Let $k : X \times X \to \mathbb{C}$ be a kernel and H its corresponding $\mathbb{C}$-RKHS. If we actually have $k(x, x') \in \mathbb{R}$ for all $x, x' \in X$, then*

$$H_{\mathbb{R}} := \big\{ f : X \to \mathbb{R} \,\big|\, \exists g \in H \text{ with } \operatorname{Re} g = f \big\}$$

equipped with the norm

$$\|f\|_{H_{\mathbb{R}}} := \inf\big\{ \|g\|_H : g \in H \text{ with } \operatorname{Re} g = f \big\}\,, \qquad\qquad f \in H_{\mathbb{R}},$$

is the $\mathbb{R}$-RKHS of the $\mathbb{R}$-valued kernel k.

3 Results

Before we state our main results we need to recall the definition of the Gaussian RBF kernels. To this end we always denote the j-th component of a complex vector $z \in \mathbb{C}^d$ by z_j. Now let us write

$$k_{\sigma,\mathbb{C}^d}(z, z') := \exp\Big(-\sigma^2 \sum_{j=1}^{d} (z_j - \bar{z}'_j)^2 \Big)$$

for $d \in \mathbb{N}$, $\sigma > 0$, and $z, z' \in \mathbb{C}^d$. Then it can be shown that $k_{\sigma,\mathbb{C}^d}$ is a $\mathbb{C}$-valued kernel on $\mathbb{C}^d$ which we call the *complex Gaussian RBF kernel with width* σ. Furthermore, its restriction $k_\sigma := (k_{\sigma,\mathbb{C}^d})_{|\mathbb{R}^d\times\mathbb{R}^d}$ is an $\mathbb{R}$-valued kernel, which we call the *(real) Gaussian RBF kernel with width* σ. Obviously, this kernel satisfies

$$k_\sigma(x, x') = \exp\bigl(-\sigma^2 \|x - x'\|_2^2\bigr)$$

for all $x, x' \in \mathbb{R}^d$, where $\|.\|_2$ denotes the Euclidian norm on $\mathbb{R}^d$.

Besides the Gaussian RBF kernels we also have to introduce a family of spaces. To this end let $\sigma > 0$ and $d \in \mathbb{N}$. For a given holomorphic function $f : \mathbb{C}^d \to \mathbb{C}$ we define

$$\|f\|_{\sigma,\mathbb{C}^d} := \left(\frac{2^d\sigma^{2d}}{\pi^d} \int_{\mathbb{C}^d} |f(z)|^2 e^{\sigma^2 \sum_{j=1}^d (z_j - \bar{z}_j)^2} dz \right)^{1/2},$$

where dz stands for the complex Lebesgue measure on $\mathbb{C}^d$. Furthermore, we write

$$H_{\sigma,\mathbb{C}^d} := \bigl\{ f : \mathbb{C}^d \to \mathbb{C} \,|\, f \text{ holomorphic and } \|f\|_{\sigma,\mathbb{C}^d} < \infty \bigr\}.$$

Obviously, $H_{\sigma,\mathbb{C}^d}$ is a complex function space with pre-Hilbert norm $\|.\|_{\sigma,\mathbb{C}^d}$. Let us now state a lemma which will help us to show that $H_{\sigma,\mathbb{C}^d}$ is a RKHS. Its proof can be found in Section 4.

Lemma 3.1 *For all $\sigma > 0$ and all compact subsets $K \subset \mathbb{C}^d$ there exists a constant $c_{K,\sigma} > 0$ such that for all $z \in K$ and all $f \in H_{\sigma,\mathbb{C}^d}$ we have*

$$|f(z)| \le c_{K,\sigma} \|f\|_{\sigma,\mathbb{C}^d}.$$

The above lemma shows that convergence in $\|.\|_{\sigma,\mathbb{C}^d}$ implies *compact convergence*, i.e. uniform convergence on every compact subset. Using the well-known fact from complex analysis that a compactly convergent sequence of holomorphic functions has a holomorphic limit (see e.g. [10, Thm. I.1.9]) we then immediately obtain the announced

Corollary 3.2 *The space $H_{\sigma,\mathbb{C}^d}$ equipped with norm $\|.\|_{\sigma,\mathbb{C}^d}$ is a RKHS for every $\sigma > 0$.*

We have seen in Theorem 2.3 that the reproducing kernel of a RKHS is determined by an arbitrary ONB of this RKHS. Therefore, to determine the reproducing kernel of $H_{\sigma,\mathbb{C}^d}$ our next step is to find an orthonormal basis (ONB) of $H_{\sigma,\mathbb{C}^d}$. To this end let us recall that the tensor product $f \otimes g : X \times X \to \mathbb{K}$ of two functions $f, g : X \to \mathbb{K}$ is defined by $f \otimes g(x, x') := f(x)g(x')$, $x, x' \in X$. Furthermore, the d-fold tensor product is defined analogously. Now we can formulate the following theorem whose proof can be found in Section 4.

Theorem 3.3 *For $\sigma > 0$ and $n \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ we define the function $e_n : \mathbb{C} \to \mathbb{C}$ by*

$$e_n(z) := \sqrt{\frac{(2\sigma^2)^n}{n!}}\, z^n e^{-\sigma^2 z^2} \tag{6}$$

for all $z \in \mathbb{C}$. Then the system $(e_{n_1} \otimes \cdots \otimes e_{n_d})_{n_1,\dots,n_d \ge 0}$ is an ONB of $H_{\sigma,\mathbb{C}^d}$.

We have seen in Theorem 2.3 that an ONB of a RKHS can be used to determine the reproducing kernel. In our case this yields the following theorem whose proof can again be found in Section 4.

Theorem 3.4 *Let $\sigma > 0$ and $d \in \mathbb{N}$. Then the complex Gaussian RBF kernel $k_{\sigma,\mathbb{C}^d}$ is the reproducing kernel of $H_{\sigma,\mathbb{C}^d}$.*

With the help of Theorem 3.4 we can now obtain some interesting information on the RKHSs of the *real* Gaussian RBF kernels k_σ. To this end we denote the restriction of a function $g : \mathbb{R}^d \to \mathbb{C}$ to a (not necessarily strict) subset $X \subset \mathbb{R}^d$ by $g_{|X}$. Our first result then describes the RKHS of k_σ restricted to $X \times X$ in terms of $H_{\sigma,\mathbb{C}^d}$:

Corollary 3.5 *For $X \subset \mathbb{R}^d$ and $\sigma > 0$ the RKHS of the real-valued Gaussian RBF kernel k_σ on X is*

$$H_\sigma(X) := \big\{ f : X \to \mathbb{R} \,|\, \exists g \in H_{\sigma,\mathbb{C}^d} \text{ with } \operatorname{Re} g_{|X} = f \big\}$$

and for $f \in H_\sigma(X)$ the norm in $H_\sigma(X)$ is given by

$$\|f\|_\sigma := \inf\big\{ \|g\|_{\sigma,\mathbb{C}^d} : g \in H_{\sigma,\mathbb{C}^d} \text{ with } \operatorname{Re} g_{|X} = f \big\} .$$

The above corollary shows that every function in the RKHS $H_\sigma(X)$ of the Gaussian RBF kernel k_σ originates from the complex RKHS $H_{\sigma,\mathbb{C}^d}$ which consists of entire functions. In particular, it is easy to see that every $f \in H_\sigma(X)$ can be represented by a power series which converges on $\mathbb{R}^d$. This observation suggests, that there may be an intimate relationship between $H_\sigma(X)$ and $H_\sigma(\mathbb{R}^d)$ if X contains an open set.

In order to investigate this conjecture we need some additional notations. For a multi-index $\nu := (n_1, \dots, n_d) \in \mathbb{N}_0^d$ we write $|v| := n_1 + \cdots + n_d$. Furthermore, for $X \subset \mathbb{R}$ and $n \in \mathbb{N}_0$ we define $e_n^X : X \to \mathbb{R}$ by

$$e_n^X(x) := \sqrt{\frac{(2\sigma^2)^n}{n!}}\, x^n e^{-\sigma^2 x^2}\,, \qquad x \in X\,, \tag{7}$$

i.e. we have $e_n^X = (e_n)_{|X} = (\operatorname{Re} e_n)_{|X}$, where $e_n : \mathbb{C} \to \mathbb{C}$ is an element of the ONB of $H_{\sigma,\mathbb{C}}$ defined by (6). Furthermore, for a multi-index $\nu := (n_1, \dots, n_d) \in \mathbb{N}_0^d$ we write $e_\nu^X := e_{n_1}^X \otimes \cdots \otimes e_{n_d}^X$ and $e_\nu := e_{n_1} \otimes \cdots \otimes e_{n_d}$. Given an $x := (x_1, \dots, x_d) \in \mathbb{R}^d$ we also adopt the notation $x^\nu := x_1^{n_1} \cdot \ldots \cdot x_d^{n_d}$. Finally, recall that $\ell_2(\mathbb{N}_0^d)$ denotes the set of all *real*-valued square-summable families, i.e.

$$\ell_2(\mathbb{N}_0^d) := \left\{ (a_\nu)_{\nu \in \mathbb{N}_0^d} : a_\nu \in \mathbb{R} \text{ for all } \nu \in \mathbb{N}_0^d \text{ and } \|(a_\nu)\|_2^2 := \sum_{\nu \in \mathbb{N}_0^d} a_\nu^2 < \infty \right\} .$$

With the help of these notations we can now show the following intermediate result:

Proposition 3.6 *Let $\sigma > 0$, $X \subset \mathbb{R}^d$ be a subset with non-empty interior, i.e. $\mathring{X} \neq \emptyset$, and $f \in H_\sigma(X)$. Then there exists a unique $(b_\nu) \in \ell_2(\mathbb{N}_0^d)$ with*

$$f(x) \;=\; \sum_{\nu \in \mathbb{N}_0^d} b_\nu e_\nu^X(x)\,, \qquad x \in X, \tag{8}$$

where the convergence is absolute. Furthermore, for all functions $g : \mathbb{C}^d \to \mathbb{C}$ the following statements are equivalent:

i) We have $g \in H_{\sigma,\mathbb{C}^d}$ and $\operatorname{Re} g_{|X} = f$.

ii) There exists an element $(c_\nu) \in \ell_2(\mathbb{N}_0^d)$ with

$$g \;=\; \sum_{\nu \in \mathbb{N}_0^d} (b_\nu + i c_\nu) e_\nu\,. \tag{9}$$

Finally, we have the identity $\|f\|^2_{H_\sigma(X)} = \sum_{\nu \in \mathbb{N}_0^d} b_\nu^2$.

With the help of the above proposition we can now establish our main result on $H_\sigma(X)$ for input spaces X having non-empty interior:

Theorem 3.7 *Let $\sigma > 0$ and $X \subset \mathbb{R}^d$ be a subset with non-empty interior. Furthermore, for $f \in H_\sigma(X)$ having the representation (8) we define*

$$\hat{f} := \sum_{\nu \in \mathbb{N}_0^d} b_\nu e_\nu \,.$$

Then the extension operator $\hat{}: H_\sigma(X) \to H_{\sigma,\mathbb{C}^d}$ *defined by $f \mapsto \hat{f}$ satisfies*

$$\begin{aligned} \operatorname{Re} \hat{f}_{|X} &= f \\ \|\hat{f}\|_{H_{\sigma,\mathbb{C}^d}} &= \|f\|_{H_\sigma(X)} \end{aligned}$$

for all $f \in H_\sigma(X)$. Moreover, (e_ν^X) is an ONB of $H_\sigma(X)$ and for $f \in H_\sigma(X)$ having the representation (8) we have $b_\nu = \langle f, e_\nu^X \rangle$ for all $\nu \in \mathbb{N}_0^d$.

In the following we present some interesting consequences of the above theorem. We begin with:

Corollary 3.8 *Let $X \subset \mathbb{R}^d$ be a subset with non-empty interior, $\sigma > 0$, and* $\hat{}: H_\sigma(X) \to H_{\sigma,\mathbb{C}^d}$ *be the extension operator defined in Theorem 3.7. Then the extension operator $I : H_\sigma(X) \to H_\sigma(\mathbb{R}^d)$ defined by $If := \operatorname{Re} \hat{f}_{|\mathbb{R}^d}$, $f \in H_\sigma(X)$, is an isometric isomorphism.*

Roughly speaking the above corollary means that $H_\sigma(\mathbb{R}^d)$ does not contain "more" functions than $H_\sigma(X)$ if X has non-empty interior. Moreover, Corollary 3.8 in particular shows that $H_\sigma(X_1)$ and $H_\sigma(X_2)$ are isometrically isomorphic via a simple extension-restriction mapping, whenever both input spaces $X_1, X_2 \subset \mathbb{R}^d$ have non-empty interior. Besides these isometries, Theorem 3.7 also yields the following interesting observation whose implications for learning theory are discussed at the end of this section:

Corollary 3.9 *Let $\sigma > 0$, $X \subset \mathbb{R}^d$ be a subset with non-empty interior, and $f \in H_\sigma(X)$. If f is constant on an open subset A of X then we actually have $f(x) = 0$ for all $x \in X$.*

The above corollary states that the space $H_\sigma(X)$ does not contain non-trivial constant functions for typical input sets X, and consequently we have $\mathbf{1}_A \notin H_\sigma(X)$ for all open subsets $A \subset X$.

Remark 3.10 As observed by Saitoh [8, p. 79] one can also obtain Theorem 3.4 by the so-called Bargmann spaces introduced in [11]. Indeed, [11] shows that these spaces are the RKHSs of the exponential kernels $(z, z') \mapsto \exp(\langle z, \bar{z}' \rangle)$ on $\mathbb{C}^d$, $d \geq 1$, and therefore one can determine the RKHSs of $k_{\sigma,\mathbb{C}^d}$ by using the relation between the exponential and the Gaussian RBF kernels. Using further results of [11] one can then derive Theorem 3.3 which played a key role in our analysis of the real Gaussian RBF kernels k_σ. However, this path requires more knowledge on both RKHS theory and Bargmann spaces and therefore we decided to present more "elementary" proofs for Theorem 3.3 and Theorem 3.4.

It is well known that a kernel has many different feature spaces and feature maps. Let us now present another feature space and feature map for k_σ which add insight into the spaces $H_\sigma(X)$. To this end let $L_2(\mathbb{R}^d)$ be the space of square-integrable functions on $\mathbb{R}^d$ equipped with the usual norm $\|\cdot\|_2$. Our first result shows that $L_2(\mathbb{R}^d)$ is a feature space of k_σ.

Lemma 3.11 *Let $0 < \sigma < \infty$, $X \subset \mathbb{R}^d$. We define $\Phi_\sigma : X \to L_2(\mathbb{R}^d)$ by*

$$\Phi_\sigma(x) \;:=\; \frac{(2\sigma)^{\frac{d}{2}}}{\pi^{\frac{d}{4}}} e^{-2\sigma^2 \|x-\cdot\|_2^2}, \qquad\qquad x \in X.$$

Then $L_2(\mathbb{R}^d)$ is a feature space and $\Phi_\sigma : X \to L_2(\mathbb{R}^d)$ is a feature map of k_σ.

With the help of the above feature space and map we will now present a representation of the inclusion $\mathrm{id} : H_\sigma(X) \to H_\tau(X)$. To this end recall (see e.g. [12]) that for $t > 0$ the Gauss-Weierstraß integral operator $W_t : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ is defined by

$$W_t g(x) \;:=\; (4\pi t)^{-\frac{d}{2}} \int_{\mathbb{R}^d} e^{-\frac{\|x-y\|_2^2}{4t}} g(y) dy$$

for all $g \in L_2(\mathbb{R}^d)$, $x \in \mathbb{R}^d$. Now we can formulate the announced result.

Proposition 3.12 *For $0 < \sigma < \tau < \infty$ we define $\delta := \frac{1}{8}\big(\frac{1}{\sigma^2} - \frac{1}{\tau^2}\big)$. Furthermore, let $X \subset \mathbb{R}^d$ and W_δ be as above. Then we obtain a commutative diagram*

$$\begin{tikzcd}
H_\sigma(X) \arrow[r, "\mathrm{id}"] & H_\tau(X) \\
L_2(\mathbb{R}^d) \arrow[u, "V_\sigma"] \arrow[r, "(\frac{\tau}{\sigma})^{\frac{d}{2}} W_\delta"'] & L_2(\mathbb{R}^d) \arrow[u, "V_\tau"']
\end{tikzcd}$$

where the vertical maps V_σ and V_τ are the metric surjections of Theorem 2.4.

Since V_σ of the above proposition is a metric surjection we obtain $\|\mathrm{id} \circ V_\sigma\| = \|\mathrm{id}\|$, and hence the commutativity of the diagram implies

$$\|\mathrm{id} : H_\sigma(X) \to H_\tau(X)\| = \|\mathrm{id} \circ V_\sigma\| = \Big(\frac{\tau}{\sigma}\Big)^{\frac{d}{2}} \|V_\tau \circ W_\delta\| \le \Big(\frac{\tau}{\sigma}\Big)^{\frac{d}{2}} \|W_\delta\|.$$

Moreover, it is well known (see e.g. [12]) that $\|W_\delta\| \le 1$. Therefore we have established the following corollary.

Corollary 3.13 *Let $X \subset \mathbb{R}^d$ and $0 < \sigma \le \tau < \infty$. Then we have*

$$\|\mathrm{id} : H_\sigma(X) \to H_\tau(X)\| \le \Big(\frac{\tau}{\sigma}\Big)^{\frac{d}{2}}.$$

Our last result which is proved in Section 4 shows that for sufficiently large X the metric surjections $V_\sigma : L_2(\mathbb{R}^d) \to H_\sigma(X)$ are isometric isomorphisms, and consequently $\mathrm{id} : H_\sigma(X) \to H_\tau(X)$ shares many important properties with W_δ.

Corollary 3.14 *Let $X \subset \mathbb{R}^d$ contain a non-empty open subset. Then $V_\sigma : L_2(\mathbb{R}^d) \to H_\sigma(X)$ is an isometric isomorphism for all $\sigma > 0$. In addition, for all $0 < \sigma < \tau < \infty$ and $\delta := \frac{1}{8}\big(\frac{1}{\sigma^2} - \frac{1}{\tau^2}\big)$ we* *have the following commutative diagram*

$$\begin{tikzcd} H_\sigma(X) \arrow[r, "\mathrm{id}"] \arrow[d, "V_\sigma^{-1}"'] & H_\tau(X) \\ L_2(\mathbb{R}^d) \arrow[r, "(\frac{\tau}{\sigma})^{\frac{d}{2}} W_\delta"'] & L_2(\mathbb{R}^d) \arrow[u, "V_\tau"'] \end{tikzcd}$$

and consequently the following statements are true:

i) $\mathrm{id} : H_\sigma(X) \to H_\tau(X)$ *is not compact.*

ii) $\mathrm{id} : H_\sigma(X) \to H_\tau(X)$ *is not surjective, i.e.* $H_\sigma(X) \subsetneq H_\tau(X)$*.*

iii) The estimate of Corollary 3.13 is exact, i.e. we have

$$\|\mathrm{id} : H_\sigma(X) \to H_\tau(X)\| = \left(\frac{\tau}{\sigma}\right)^{\frac{d}{2}} .$$

Finally, let us briefly discuss how the above result can be used in the analysis of support vector machines (see [1] for these learning algorithms). For the sake of simplicity we only consider the support vector machines (SVMs) with Gaussian RBF kernels and with hinge loss $L(y,t) := \max\{0, 1-yt\}$, $y \in Y := \{-1,1\}$, $t \in \mathbb{R}$, which are used for binary classification problems (see [13] for an introduction to classification). Moreover, let $X \subset \mathbb{R}^d$ be as in the above corollary and P be a probability measure on $X \times Y$. Then for a measurable $f : X \to \mathbb{R}$ we define the *L-risk* by

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

Furthermore, the *minimal L-risk* is denoted by $\mathcal{R}^*_{L,P} := \inf_f \mathcal{R}_{L,P}(f)$, where the infimum runs over all measurable functions. Now, it has recently been discovered that for analyzing the learning performance of SVMs the behaviour of the *approximation error function*

$$a_\sigma(\lambda) := \inf_{f\in H_\sigma(X)} \lambda\|f\|^2_{\sigma,X} + \mathcal{R}_{L,P}(f) - \mathcal{R}^*_{L,P} \tag{10}$$

for $\lambda \to 0$ plays an important role. Indeed, $a_\sigma(\lambda) \to 0$ for $\lambda \to 0$ was used in [14] to show that SVMs can learn in the sense of universal consistency (see [13] for an introduction to this notion of learning). Furthermore, [15], [16] and [5] established small bounds on $a_\sigma(\lambda)$ for certain P, σ and λ which were used for stronger guarantees on the learning performance of SVMs. Unfortunately, the techniques used are rather involved and in particular it is completely open whether the obtained bounds are sharp. Now, observe that Corollary 3.14 shows

$$a_\sigma(\lambda) = \inf_{g\in L_2(\mathbb{R}^d)} \lambda\|g\|^2_{L_2(\mathbb{R}^d)} + \mathcal{R}_{L,P}(V_\sigma g) - \mathcal{R}^*_{L,P} , \tag{11}$$

which may significantly help in understanding the behaviour of $a_\sigma(\lambda)$. Indeed, in order to establish a small bound of $a_\sigma(\lambda)$ via (10) one has to simultaneously control both the shape and the $\|\,.\,\|_{\sigma,X}$-norm of certain $f \in H_\sigma(X)$ which is rather challenging because of the analyticity of these f. In contrast to this, we see that when considering (11) the task is to simultaneously control $\|g\|_{L_2(\mathbb{R}^d)}$

and the shape of $V_\sigma g$ for suitable $g \in L_2(\mathbb{R}^d)$. Obviously, the first term is easy to determine for many g and the second term can be investigated by e.g. the well-established theory of the Gauss-Weierstraß integral operator, or more generally, convolution operators. Remarkably, this approach was already used implicitly in [5], however arising technical difficulties in [5] make it hard to see the simple structure there. We hope that by outlining (11) and its usability the existing bounds on $a_\sigma(\lambda)$ can be further improved.

Moreover, note that the results established in this work also give a negative result on the approximation error function for a large class of distributions and *fixed* σ. Indeed, if we write $\eta(x) := P(y = 1|x)$, $x \in X$, and assume e.g. that the set $\{x : 1/2 < \eta(x) < 1\}$ has a non-empty interior then Corollary 3.9 shows that the infimum of $\mathcal{R}_{L,P}(.)$ over $H_\sigma(X)$ is not attained since every possible minimizer f^* must satisfy $f^*(x) = 1$ for all x with $1/2 < \eta(x) < 1$. With the help of [17] we then see that there exists *no* constant c_σ with $a_\sigma(\lambda) \leq c_\sigma \lambda$ for all (small) $\lambda > 0$. In particular this shows that for such P the recent methods (see e.g. [15, 17]) for establishing learning rates can only yield learning rates converging to 0 slower than the regularization sequence (λ_n).

Finally, it is worth mentioning that the injectivity of the integral operator W_t has been recently used in [18] to establish the relation

$$\inf_{f \in H_\sigma(X)} \mathcal{R}_{L,P}(f) = \mathcal{R}^*_{L,P}$$

for almost all commonly used convex loss functions and all distributions P on $\mathbb{R}^d \times \mathbb{R}$. In particular, this equality allows consistency results in the spirit of [14] for *unbounded* input spaces $X \subset \mathbb{R}^d$ which were previously not possible due to the "non-universality" of k_σ on $\mathbb{R}^d$.

4 Proofs

Proof of Corollary 2.5: It is easy to check that $H_0 := H$ equipped with the inner product

$$\langle f, f' \rangle_{H_0} := \mathrm{Re}\langle f, f' \rangle_H \,, \qquad\qquad f,\, f' \in H_0,$$

is an $\mathbb{R}$-feature space of the $\mathbb{R}$-valued kernel k. Moreover, for $f \in H_0$ and $x \in X$ we have

$$f(x) = \langle f, \Phi(x) \rangle_H = \mathrm{Re}\langle f, \Phi(x) \rangle_H + i\ \mathrm{Im}\langle f, \Phi(x) \rangle_H = \langle f, \Phi(x) \rangle_{H_0} + i\ \mathrm{Im}\, f(x),$$

i.e. we have found $\langle f, \Phi(x) \rangle_{H_0} = \mathrm{Re}\, f(x)$. Now, the assertion follows from Theorem 2.4. ■

For the proof of Lemma 3.1 we need the following technical lemma.

Lemma 4.1 *For all $d \in \mathbb{N}$, all holomorphic functions $f : \mathbb{C}^d \to \mathbb{C}$, all $r_1, \ldots, r_d > 0$, and all $z \in \mathbb{C}^d$ we have*

$$|f(z)|^2 \leq \frac{1}{(2\pi)^d} \int_0^{2\pi} \cdots \int_0^{2\pi} \left| f(z_1 + r_1 e^{i\theta_1}, \ldots, z_d + r_d e^{i\theta_d}) \right|^2 d\theta_1 \cdots d\theta_d \,. \tag{12}$$

Proof: We proceed by induction over d. For $d = 1$ the assertion follows from Hardy's convexity theorem (see e.g. [19, p. 9]) which states that the function

$$r \mapsto \frac{1}{2\pi} \int_0^{2\pi} |f(z + re^{i\theta})|^2 d\theta$$

is non-decreasing on $[0, \infty)$.

Now let us suppose that we have already shown the assertion for $d \in \mathbb{N}$. Let $f : \mathbb{C}^{d+1} \to \mathbb{C}$ be a holomorphic function, and choose $r_1, \ldots, r_{d+1} > 0$. Since for fixed $(z_1, \ldots, z_d) \in \mathbb{C}^d$ the function $z_{d+1} \mapsto f(z_1, \ldots, z_d, z_{d+1})$ is holomorphic by the induction hypothesis, we obtain

$$|f(z_1, \ldots, z_{d+1})|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} |f(z_1, \ldots, z_d, z_{d+1} + r_{d+1} e^{i\theta_{d+1}})|^2 d\theta_{d+1} \, .$$

Now applying the induction hypothesis to the holomorphic function

$$(z_1, \ldots, z_d) \mapsto f(z_1, \ldots, z_d, z_{d+1} + r_{d+1} e^{i\theta_{d+1}})$$

on $\mathbb{C}^d$ gives the assertion for $d+1$. ∎

Proof of Lemma 3.1: Let us define $c := \max\{e^{-\sigma^2 \sum_{j=1}^d (z_j - \bar{z}_j)^2} : (z_1, \ldots, z_d) \in K + (B_{\mathbb{C}})^d\}$, where $B_{\mathbb{C}}$ denotes the closed unit ball of $\mathbb{C}$. Now, by Lemma 4.1 we have

$$2^d r_1 \cdots r_d |f(z)|^2 \leq \frac{r_1 \cdots r_d}{\pi^d} \int_0^{2\pi} \cdots \int_0^{2\pi} \left| f(z_1 + r_1 e^{i\theta_1}, \ldots, z_d + r_d e^{i\theta_d}) \right|^2 d\theta_1 \cdots d\theta_d$$

and integrating this inequality with respect to $r = (r_1, \ldots, r_d)$ over $[0,1]^d$ then yields

$$\begin{aligned} |f(z)|^2 &\leq \frac{1}{\pi^d} \int\limits_{z + (B_{\mathbb{C}})^d} |f(z')|^2 dz' \\ &\leq \frac{c}{\pi^d} \int\limits_{z + (B_{\mathbb{C}})^d} |f(z')|^2 e^{\sigma^2 \sum_{j=1}^d (z_j - \bar{z}_j)^2} dz' \\ &\leq \frac{c}{(2\sigma^2)^d} \|f\|^2_{\sigma, \mathbb{C}^d} \, . \end{aligned}$$

∎

For the proof of Theorem 3.3 we need the following technical lemma.

Lemma 4.2 *For all $n, m \in \mathbb{N}_0$ and all $\sigma > 0$ we have*

$$\int_{\mathbb{C}} z^n (\bar{z})^m e^{-2\sigma^2 z\bar{z}} dz = \begin{cases} \frac{\pi \, n!}{(2\sigma^2)^{n+1}} & \textit{if } n = m \\ 0 & \textit{otherwise.} \end{cases}$$

Proof: Let us first consider the case $n = m$. Then we have

$$\begin{aligned} \int_{\mathbb{C}} z^n (\bar{z})^m e^{-2\sigma^2 z\bar{z}} dz &= \int_0^\infty \int_0^{2\pi} r^{2n} e^{-2\sigma^2 r^2} d\theta \, r \, dr \\ &= 2\pi \int_0^\infty r^{2n+1} e^{-2\sigma^2 r^2} dr \\ &= \frac{\pi}{(2\sigma^2)^{n+1}} \int_0^\infty t^n e^{-t} dt \\ &= \frac{\pi \, n!}{(2\sigma^2)^{n+1}} \, . \end{aligned}$$

Now let us assume $n \neq m$. Then we obtain

$$\int_{\mathbb{C}} z^n (\bar{z})^m e^{-2\sigma^2 z\bar{z}} dz = \int_0^\infty r \int_0^{2\pi} r^{n+m} e^{i(n-m)\theta} e^{-2\sigma^2 r^2} d\theta dr = 0 \, .$$

∎

Proof of Theorem 3.3: In order to avoid cumbersome technical notations that hide the structure of the proof we first consider the case $d = 1$.

Let us show that $(e_n)_{n\geq 0}$ is an orthonormal system. To this end for $n, m \in \mathbb{N}_0$, and $z \in \mathbb{C}$ we observe

$$e_n(z)\overline{e_m(z)}e^{\sigma^2(z-\bar z)^2} = \sqrt{\frac{(2\sigma^2)^{n+m}}{n!\,m!}}z^n(\bar z)^m e^{-\sigma^2 z^2-\sigma^2\bar z^2}e^{\sigma^2(z-\bar z)^2} = \sqrt{\frac{(2\sigma^2)^{n+m}}{n!\,m!}}z^n(\bar z)^m e^{-2\sigma^2 z\bar z}\,.$$

Therefore for $n, m \geq 0$ we obtain

$$\langle e_n, e_m\rangle = \frac{2\sigma^2}{\pi}\int_{\mathbb{C}} e_n(z)\overline{e_m(z)}e^{\sigma^2(z-\bar z)^2}dz = \frac{2\sigma^2}{\pi}\cdot\sqrt{\frac{(2\sigma^2)^{n+m}}{n!\,m!}}\int_{\mathbb{C}} z^n(\bar z)^m e^{-2\sigma^2 z\bar z}dz = \begin{cases}1 & \text{if } n=m\\ 0 & \text{otherwise}\end{cases}$$

by Lemma 4.2. This shows that $(e_n)_{n\geq 0}$ is indeed an orthonormal system.

Now, let us show that this system is also complete. To this end let $f \in H_\sigma(\mathbb{C})$. Then $z \mapsto e^{\sigma^2 z^2} f(z)$ is an entire function, and therefore there exists a sequence $(a_n) \subset \mathbb{C}$ such that

$$f(z) = \sum_{n=0}^{\infty} a_n z^n e^{-\sigma^2 z^2} = \sum_{n=0}^{\infty} a_n \sqrt{\frac{n!}{(2\sigma^2)^n}}\, e_n(z) \tag{13}$$

for all $z \in \mathbb{C}$. Obviously, it suffices to show that the above convergence also holds with respect to $\|.\|_{\sigma,\mathbb{C}}$. To prove this we first recall from complex analysis that the series in (13) converges absolutely and compactly. Therefore for $n \geq 0$ Lemma 4.2 yields

$$\begin{aligned}\langle f, e_n\rangle &= \frac{2\sigma^2}{\pi}\int_{\mathbb{C}} f(z)\overline{e_n(z)}e^{\sigma^2(z-\bar z)^2}dz\\ &= \frac{2\sigma^2}{\pi}\sum_{m=0}^{\infty} a_m \int_{\mathbb{C}} z^m e^{-\sigma^2 z^2}\overline{e_n(z)}e^{\sigma^2(z-\bar z)^2}dz\\ &= \frac{2\sigma^2}{\pi}\sqrt{\frac{(2\sigma^2)^n}{n!}}\sum_{m=0}^{\infty} a_m \int_{\mathbb{C}} z^m(\bar z)^n e^{-2\sigma^2 z\bar z}dz\\ &= a_n\sqrt{\frac{n!}{(2\sigma^2)^n}}\,.\end{aligned} \tag{14}$$

Furthermore, since (e_n) is an orthonormal system we have $(\langle f, e_n\rangle) \in \ell_2$ by Bessel's inequality. Using again that (e_n) is an orthonormal system in $H_\sigma(\mathbb{C})$ we hence find a function $g \in H_\sigma(\mathbb{C})$ with $g = \sum_{n=0}^{\infty}\langle f, e_n\rangle e_n$, where the convergence takes place in $H_\sigma(\mathbb{C})$. Now, using (13), (14), and the fact that norm convergence in RKHSs implies point-wise convergence we find $g = f$, i.e. the series in (13) converges with respect to $\|.\|_{\sigma,\mathbb{C}}$.

Now, let us briefly treat the general, d-dimensional case. In this case a simple calculation shows

$$\langle e_{n_1}\otimes\cdots\otimes e_{n_d}, e_{m_1}\otimes\cdots\otimes e_{m_d}\rangle_{H_{\sigma,\mathbb{C}^d}} = \prod_{j=1}^{d}\langle e_{n_j}, e_{m_j}\rangle_{H_\sigma(\mathbb{C})}\,,$$

and hence we find the orthonormality of $(e_{n_1}\otimes\cdots\otimes e_{n_d})_{n_1,\ldots,n_d\geq 0}$. In order to check that this orthonormal system is complete let us fix an $f \in H_{\sigma,\mathbb{C}^d}$. Then $z \mapsto f(z)\exp(\sigma^2\sum_{i=1}^{d} z_i^2)$ is an entire function, and hence [10, Thm. I.1.18] shows there there exist $a_{n_1,\dots,n_d} \in \mathbb{C}$, $(n_1,\dots,n_d) \in \mathbb{N}_0^d$, such that

$$f(z) = \sum_{(n_1,\dots,n_d)\in\mathbb{N}_0^d} a_{n_1,\dots,n_d} \prod_{i=1}^d z_i^{n_i} e^{-\sigma^2 z_i^2} = \sum_{(n_1,\dots,n_d)\in\mathbb{N}_0^d} a_{n_1,\dots,n_d} \prod_{i=1}^d \sqrt{\frac{n_i!}{(2\sigma^2)^{n_i}}}\, e_{n_i}(z)$$

for all $z = (z_1,\dots,z_d) \in \mathbb{C}^d$. From this we easily derive $\langle f, e_{n_1} \otimes \cdots \otimes e_{n_d}\rangle = a_{n_1,\dots,n_d} \prod_{i=1}^d \sqrt{\frac{n_i!}{(2\sigma^2)^{n_i}}}$, and hence we obtain the completeness as in the 1-dimensional case. ■

Proof of Theorem 3.4: Let k be the reproducing kernel of $H_{\sigma,\mathbb{C}^d}$. Then using the ONB of Theorem 3.3 and the Taylor series expansion of the exponential function we obtain

$$\begin{aligned}
k(z,z') &= \sum_{n_1,\dots,n_d=0}^{\infty} e_{n_1} \otimes \cdots \otimes e_{n_d}(z)\overline{e_{n_1} \otimes \cdots \otimes e_{n_d}(z')} \\
&= \sum_{n_1,\dots,n_d=0}^{\infty} \prod_{j=1}^d \frac{(2\sigma^2)^{n_j}}{n_j!} (z\bar z')^{n_j} e^{-\sigma^2 z_j^2 - \sigma^2 (\bar z_j')^2} \\
&= \prod_{j=1}^d \sum_{n_j=0}^{\infty} \frac{(2\sigma^2)^{n_j}}{n_j!} (z\bar z')^{n_j} e^{-\sigma^2 z_j^2 - \sigma^2 (\bar z_j')^2} \\
&= \prod_{j=1}^d e^{-\sigma^2 z_j^2 - \sigma^2 (\bar z_j')^2 + 2\sigma^2 z_j \bar z_j'} \\
&= e^{-\sigma^2 \sum_{j=1}^d (z_j - \bar z_j')^2},
\end{aligned}$$

which shows the assertion. ■

Proof of Corollary 3.5: The assertion directly follows from Theorem 3.4, the definition of $k_{\sigma,\mathbb{C}^d}$, and Corollary 2.5. ■

Proof of Proposition 3.6: *i)* $\Rightarrow$ *ii)*. Let us fix a $g \in H_{\sigma,\mathbb{C}^d}$ with $\operatorname{Re} g_{|X} = f$. Since (e_ν) is an ONB of $H_{\sigma,\mathbb{C}^d}$ we then have

$$g = \sum_{\nu\in\mathbb{N}_0^d} \langle g, e_\nu\rangle\, e_\nu\,,$$

where the convergence is with respect to $H_{\sigma,\mathbb{C}^d}$. In addition, recall that the family of Fourier coefficients is square-summable and satisfies Parseval's identity

$$\|g\|^2_{H_{\sigma,\mathbb{C}^d}} = \sum_{\nu\in\mathbb{N}_0^d} \big|\langle g, e_\nu\rangle\big|^2 .$$

Since convergence in $H_{\sigma,\mathbb{C}^d}$ implies pointwise convergence we then obtain

$$f(x) = \operatorname{Re} g_{|X}(x) = \operatorname{Re}\Big(\sum_{\nu\in\mathbb{N}_0^d} \langle g, e_\nu\rangle e_\nu(x)\Big) = \sum_{\nu\in\mathbb{N}_0^d} \operatorname{Re}\big(\langle g, e_\nu\rangle\big) e_\nu^X(x)\,, \qquad x \in X,$$

where in the last step we used $e_\nu(x) \in \mathbb{R}$ for $x \in X$. In order to show *ii)* it consequently remains to show that $b_\nu := \operatorname{Re}\langle g, e_\nu\rangle$ only depends on f but not on g. To this end let $\tilde g \in H_{\sigma,\mathbb{C}^d}$ be another function with $\operatorname{Re} \tilde g_{|X} = f$. By repeating the above argument for $\tilde g$ we then find

$$f(x) = \sum_{\nu\in\mathbb{N}_0^d} \operatorname{Re}\big(\langle \tilde g, e_\nu\rangle\big) e_\nu^X(x)\,, \qquad x \in X.$$

Using the definition (7) we then obtain

$$\sum_{\nu\in\mathbb{N}_0^d} \mathrm{Re}\big(\langle \tilde g, e_\nu\rangle\big)\, a_\nu\, x^\nu \;=\; \sum_{\nu\in\mathbb{N}_0^d} \mathrm{Re}\big(\langle g, e_\nu\rangle\big)\, a_\nu\, x^\nu\,, \qquad x\in X,$$

where $a_\nu := a_{n_1}\cdot \ldots \cdot a_{n_d}$ and $a_n := \big(\frac{2^n\sigma^{2n}}{n!}\big)^{1/2}$. Since X has non-empty interior the identity theorem for power series and $a_\nu \neq 0$ then give $\mathrm{Re}\langle \tilde g, e_\nu\rangle = \mathrm{Re}\langle g, e_\nu\rangle$ for all $\nu\in\mathbb{N}_0^d$. This shows both (8) and (9). Finally, Corollary 3.5 and Parseval's identity give

$$\|f\|^2_{H_\sigma(X)} = \inf\big\{\|g\|_{\sigma,\mathbb{C}^d} : g\in H_{\sigma,\mathbb{C}^d} \text{ with } \mathrm{Re}\, g_{|X} = f\big\} \;=\; \inf\Big\{\sum_{\nu\in\mathbb{N}_0^d} b_\nu^2 + c_\nu^2 : (c_\nu)\in \ell_2(\mathbb{N}_0^d)\Big\} \;=\; \sum_{\nu\in\mathbb{N}_0^d} b_\nu^2\,.$$

ii) $\Rightarrow$ *i)*. Since $(b_\nu)\in\ell_2(\mathbb{N}_0^d)$ and $(c_\nu)\in\ell_2(\mathbb{N}_0^d)$ imply $\big(|b_\nu + ic_\nu|\big)\in\ell_2(\mathbb{N}_0^d)$ we have $g\in H_{\sigma,\mathbb{C}^d}$. Furthermore, $\mathrm{Re}\, g_{|X} = f$ follows from

$$\mathrm{Re}\, g(x) = \mathrm{Re}\sum_{\nu\in\mathbb{N}_0^d}(b_\nu + ic_\nu)e_\nu(x) = \sum_{\nu\in\mathbb{N}_0^d} b_\nu e_\nu^X(x) = f(x)\,, \qquad x\in X.$$

■

Proof of Theorem 3.7: By (8) the extension operator is well-defined. The identities then follow from Proposition 3.6 and Parseval's identity. Moreover, the extension operator is obviously $\mathbb{R}$-linear and satisfies $\hat e_\nu^X = e_\nu$ for all $\nu\in\mathbb{N}_0^d$. Consequently, we obtain

$$\|e^X_{\nu_1}\pm e^X_{\nu_2}\|_{H_\sigma(X)} \;=\; \|\hat e^X_{\nu_1}\pm \hat e^X_{\nu_2}\|_{H_{\sigma,\mathbb{C}^d}} \;=\; \|e_{\nu_1}\pm e_{\nu_2}\|_{H_{\sigma,\mathbb{C}^d}}$$

for $\nu_1,\nu_2\in\mathbb{N}_0^d$. Using the polarization identity we then see that (e_ν^X) is an ONS in $H_\sigma(X)$. To see that it actually is an ONB we fix an $f\in H_\sigma(X)$. Furthermore, let $(b_\nu)\in\ell_2(\mathbb{N}_0^d)$ be the family that satisfies (8). Then

$$\tilde f := \sum_{\nu\in\mathbb{N}_0^d} b_\nu e_\nu^X$$

converges in $H_\sigma(X)$. Since convergence in $H_\sigma(X)$ implies pointwise convergence, (8) then yields $\tilde f(x) = f(x)$ for all $x\in X$. Consequently, (e_ν^X) is an ONB of $H_\sigma(X)$. Finally, the identity $b_\nu = \langle f, e_\nu^X\rangle$, $\nu\in\mathbb{N}_0^d$, follows from the fact that the representation of f by (e_ν^X) is unique. ■

Proof of Corollary 3.8: For $f\in H_\sigma(X)$ we have $(\langle f, e_\nu^X\rangle)\in\ell_2(\mathbb{N}_0^d)$ and hence

$$\tilde f := \sum_{\nu\in\mathbb{N}_0^d}\langle f, e_\nu^X\rangle e_\nu^{\mathbb{R}^d}$$

is a well-defined element in $H_\sigma(\mathbb{R}^d)$. Moreover, for $\nu\in\mathbb{N}_0^d$ we have $(\mathrm{Re}\, e_\nu)_{|\mathbb{R}^d} = e_\nu^{\mathbb{R}^d}$ and $\langle f, e_\nu^X\rangle\in\mathbb{R}$, and hence we find $If = \tilde f$. Furthermore, $\|f\|_{H_\sigma(X)} = \|If\|_{H_\sigma(\mathbb{R}^d)}$ immediately follows from Parseval's identity. Consequently, I is isometric, linear, and injective. The surjectivity finally follows from the fact that given an $\tilde f\in H_\sigma(\mathbb{R}^d)$ the function

$$f := \sum_{\nu\in\mathbb{N}_0^d}\big\langle f, e_\nu^{\mathbb{R}^d}\big\rangle\, e_\nu^X$$

obviously satisfies $f\in H_\sigma(X)$ and $If = \tilde f$. ■

Proof of Corollary 3.9: Let $c \in \mathbb{R}$ be a constant with $f(x) = c$ for all $x \in A$. Let us define $a_n := (\frac{(2\sigma^2)^n}{n!})^{1/2}$ for all $n \in \mathbb{N}_0$. Furthermore, for $\nu := (n_1, \ldots, n_d) \in \mathbb{N}_0^d$ we write $b_\nu := \langle f, e_\nu^X \rangle$ and $a_\nu := a_{n_1} \cdot \ldots \cdot a_{n_d}$. For $x := (x_1, \ldots, x_d) \in A$ the definition (7) and the representation (8) then yield

$$c \exp\Bigl(\sigma^2 \sum_{j=1}^d x_j^2\Bigr) = f(x) \exp\Bigl(\sigma^2 \sum_{j=1}^d x_j^2\Bigr) = \sum_{\nu \in \mathbb{N}_0^d} b_\nu a_\nu x^\nu . \tag{15}$$

Moreover, for $x \in \mathbb{R}^d$ a simple calculation shows

$$\exp\Bigl(\sigma^2 \sum_{j=1}^d x_j^2\Bigr) = \prod_{j=1}^d e^{\sigma^2 x_j^2} = \prod_{j=1}^d \Bigl(\sum_{n_j=0}^\infty \frac{\sigma^{2n_j} x_j^{2n_j}}{n_j!}\Bigr) = \sum_{n_1,\ldots,n_d=0}^\infty \ \prod_{j=1}^d \frac{\sigma^{2n_j} x_j^{2n_j}}{n_j!} .$$

Using (15) and the identity theorem for power series we hence obtain

$$b_\nu = \begin{cases} c \prod_{j=1}^d \frac{\sqrt{(2n_j)!}}{n_j!} 2^{-n_j} & \text{if } \nu = (2n_1, \ldots, 2n_d) \text{ for some } (n_1, \ldots, n_d) \in \mathbb{N}_0^d \\ 0 & \text{otherwise} . \end{cases}$$

Consequently, Parseval's identy yields

$$\|f\|^2_{H_\sigma(X)} \ = \ \sum_{\nu \in \mathbb{N}_0^d} b_\nu^2 = \sum_{n_1,\ldots,n_d=0}^\infty c^2 \prod_{j=1}^d \frac{(2n_j)!}{(n_j!)^2} 2^{-2n_j} = \Bigl(\sum_{n=0}^\infty c^{2/d} \frac{(2n)!}{(n!)^2} 2^{-2n}\Bigr)^d .$$

Let us write $\alpha_n := \frac{(2n)!}{(n!)^2} 2^{-2n}$ for $n \in \mathbb{N}_0$. By an easy calculation we then obtain

$$\frac{\alpha_{n+1}}{\alpha_n} = \frac{\bigl(2(n+1)\bigr)! \ (n!)^2 \ 2^n}{(2n)! \ \bigl((n+1)!\bigr)^2 \ 2^{2(n+1)}} = \frac{(2n+1)(2n+2)}{4(n+1)^2} = \frac{2n+1}{2n+2} \geq \frac{n}{n+1}$$

for all $n \geq 1$. In other words, $(n\alpha_n)$ is a increasing, positive sequence. Consequently we have $\alpha_n \geq \frac{\alpha_1}{n}$ for all $n \geq 1$, and hence we find $\sum_{n=0}^\infty \alpha_n = \infty$. Therefore, $\|f\|^2_{H_\sigma(X)} < \infty$ implies $c = 0$, and thus we have $b_\nu = 0$ for all $\nu \in \mathbb{N}_0^d$. This shows $f = 0$. ■

Proof of Lemma 3.11: We begin by collecting some well known facts about manipulating Gaussians that are useful in proving Lemma 3.11, Theorem 4.3, and Corollary 3.14. First it is well known that for all $t > 0$ and $x \in \mathbb{R}^d$ we have

$$\int_{\mathbb{R}^d} e^{-\frac{\|y-x\|_2^2}{t}} dy = (\pi t)^{\frac{d}{2}} . \tag{16}$$

Second, an elementary calculation shows

$$\|y - x\|_2^2 + \alpha \|y - x'\|_2^2 = \frac{\alpha}{1+\alpha} \|x - x'\|_2^2 + (1+\alpha) \Bigl\| y - \frac{x + \alpha x'}{1+\alpha} \Bigr\|_2^2 \tag{17}$$

for all $\alpha \geq 0$ and all $y, x, x' \in \mathbb{R}^d$. Now by using (16) and setting $\alpha := 1$ in (17) we obtain

$$\begin{aligned} \langle \Phi_\sigma(x), \Phi_\sigma(x') \rangle_{L_2(\mathbb{R}^d)} \ &= \ \frac{(2\sigma)^d}{\pi^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-2\sigma^2 \|x-z\|_2^2} e^{-2\sigma^2 \|x'-z\|_2^2} dz \\ &= \ \frac{(2\sigma)^d}{\pi^{\frac{d}{2}}} e^{-\sigma^2 \|x-x'\|_2^2} \int_{\mathbb{R}^d} e^{-4\sigma^2 \|z - \frac{x+x'}{2}\|_2^2} dz \\ &= \ \frac{(2\sigma)^d}{\pi^{\frac{d}{2}}} \cdot e^{-\sigma^2 \|x-x'\|_2^2} \Bigl(\frac{\pi}{4\sigma^2}\Bigr)^{\frac{d}{2}} \\ &= \ k_\sigma(x, x'). \end{aligned}$$

Therefore Φ_σ is a feature map and $L_2(\mathbb{R}^d)$ is a feature space of k_σ. ■

Proof of Proposition 3.12: Theorem 2.4 shows that we can compute the metric surjection $V_\sigma : L_2(\mathbb{R}^d) \to H_\sigma(X)$ by

$$V_\sigma g(x) = \langle g, \Phi_\sigma(x)\rangle_{L_2(\mathbb{R}^d)} = \frac{(2\sigma)^{\frac{d}{2}}}{\pi^{\frac{d}{4}}} \int_{\mathbb{R}^d} e^{-2\sigma^2\|x-y\|_2^2} g(y)dy\,, \qquad g \in L_2(\mathbb{R}^d),\, x \in X\,,$$

where Φ_σ is the feature map defined in Lemma 3.11. Note, that in this formula the *computation* of V_σ is independent of the chosen domain X. Therefore let us first consider the case $X = \mathbb{R}^d$. Then the relationships

$$V_\sigma = \left(\frac{\pi}{\sigma^2}\right)^{\frac{d}{4}} W_{\frac{1}{8\sigma^2}} \qquad \text{and} \qquad W_{\frac{1}{8\tau^2}} = \left(\frac{\tau^2}{\pi}\right)^{\frac{d}{4}} V_\tau$$

are easily derived. Furthermore it is well known (see e.g. Hille and Phillips [12]) that the Gauss-Weierstraß integral operator corresponds to a solution of the heat equation and so satisfies the semigroup identity

$$W_s = W_t W_{s-t}$$

for all $0 < t < s$. Combining this with the relations between the operators W_t and V_σ we obtain

$$V_\sigma = \left(\frac{\pi}{\sigma^2}\right)^{\frac{d}{4}} W_{\frac{1}{8\sigma^2}} = \left(\frac{\pi}{\sigma^2}\right)^{\frac{d}{4}} W_{\frac{1}{8\tau^2}} W_{\frac{1}{8}\left(\frac{1}{\sigma^2}-\frac{1}{\tau^2}\right)} = \left(\frac{\tau}{\sigma}\right)^{\frac{d}{2}} V_\tau W_{\frac{1}{8}\left(\frac{1}{\sigma^2}-\frac{1}{\tau^2}\right)} \tag{18}$$

for all $0 < \sigma < \tau$, and thus the diagram commutes in the case of $X = \mathbb{R}^d$. The general case $X \subset \mathbb{R}^d$ follows from (18) using the fact that the *computation* of V_σ is independent of X. ■

For the proof of Corollary 3.14 we have to recall the following important theorem which for completeness is proved below.

Theorem 4.3 *The Gauss-Weierstraß integral operator $W_t : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ is not compact for all $t > 0$.*

Proof: Let $\mathbb{Z}^d$ be the lattice of integral vectors in $\mathbb{R}^d$. For $n = (n_1, ..., n_d) \in \mathbb{Z}^d$ and $s > 0$ we define

$$g_n^{(s)}(x) := (2\pi s)^{-\frac{d}{4}} e^{-\frac{\|x-n\|^2}{4s}}\,, \qquad x \in \mathbb{R}^d.$$

Then (16) shows $\|g_n^{(s)}\|_2^2 = 1$, i.e. $g_n^{(s)}$ is contained in the closed unit ball $B_{L_2(\mathbb{R}^d)}$ of $L_2(\mathbb{R}^d)$. Furthermore from (16) and (17) we infer

$$W_t g_n^{(s)}(x) = (4\pi t)^{-\frac{d}{2}} (2\pi s)^{-\frac{d}{4}} \int_{\mathbb{R}^d} e^{-\frac{\|x-y\|_2^2}{4t}} e^{-\frac{\|y-n\|_2^2}{4s}} dy = (2\pi s)^{-\frac{d}{4}} \left(\frac{s}{s+t}\right)^{\frac{d}{2}} e^{-\frac{\|x-n\|_2^2}{4(s+t)}}.$$

Consequently, by utilizing (16) and (17) yet again, we obtain for $n, m \in \mathbb{Z}^d$ that

$$\langle W_t g_n^{(s)}, W_t g_m^{(s)}\rangle = (2\pi s)^{-\frac{d}{2}} \left(\frac{s}{s+t}\right)^{d} \int_{\mathbb{R}^d} e^{-\frac{\|y-n\|_2^2}{4(s+t)}} e^{-\frac{\|y-m\|_2^2}{4(s+t)}} dy = \left(\frac{s}{s+t}\right)^{\frac{d}{2}} e^{-\frac{\|m-n\|_2^2}{8(s+t)}}. \tag{19}$$

Therefore for $n \neq m \in \mathbb{Z}^d$ and $s := t$ we have

$$\begin{aligned}\|W_t g_n^{(t)} - W_t g_m^{(t)}\|_2^2 \;=\; \|W_t g_n^{(t)}\|_2^2 + \|W_t g_m^{(t)}\|_2^2 - 2\langle W_t g_n^{(t)}, W_t g_m^{(t)}\rangle &= 2^{1-\frac{d}{2}}\left(1 - e^{-\frac{\|m-n\|_2^2}{16t}}\right)\\ &\geq 2^{1-\frac{d}{2}}\left(1 - e^{-\frac{1}{16t}}\right),\end{aligned}$$

and hence $\{W_t g_n^{(t)} : n \in \mathbb{Z}^d\} \subset W_t B_{L_2(\mathbb{R}^d)}$ is not precompact. This implies the assertion. ■

Proof of Corollary 3.14: Let us first show that V_σ is an isometric isomorphism. In view of Theorem 2.4 it suffices to prove that V_σ is injective. To this end let $g \in L_2(\mathbb{R}^d)$ with $V_\sigma g = 0$. Since X contains an open subset the analytic extension $\hat{V}_\sigma g : \mathbb{R}^d \to \mathbb{R}$ of $V_\sigma g$ also satisfies $\hat{V}_\sigma g = 0$. Now, the unique continuation property of Itô and Yamabe [20] for the heat equation implies $g = 0$, and hence V_σ is injective by its linearity. Obviously, the asserted diagram is an immediate consequence of the injectivity of V_σ and the diagram in Proposition 3.12.

Now the remaining assertions can be shown by the established diagram. Indeed, *i*) follows from Theorem 4.3. Beckner's [21] work on sharp Young's inequalities implies $\|W_\delta\| = 1$ which establishes *iii*) but the result can be easily obtained from (19). Indeed, the latter implies $\|W_\delta g_n\|_2 = (\frac{s}{s+t})^{\frac{d}{4}}$ and we obtain *iii*) by letting $s \to \infty$. Finally, by considering the case $X = \mathbb{R}^d$ we note that for $t > 0$ and $\tau := \frac{1}{\sqrt{8t}}$ we have $W_t = (8\pi t)^{-\frac{d}{4}} V_\tau$, i.e. we have the following commutative diagram

$$\begin{tikzcd}
L_2(\mathbb{R}^d) \arrow[rr, "W_t"] \arrow[dr, "V_\tau"'] & & L_2(\mathbb{R}^d) \\
& H_\tau(\mathbb{R}^d) \arrow[ur, "(8\pi t)^{-\frac{d}{4}}\mathrm{id}"'] &
\end{tikzcd}$$

Now, since $H_\tau(\mathbb{R}^d)$ consists of analytic functions we obviously have $H_\tau(\mathbb{R}^d) \subsetneq L_2(\mathbb{R}^d)$ and hence W_t is not surjective. ■

References

[1] B. Schölkopf and A.J. Smola. *Learning with Kernels.* MIT Press, 2002.

[2] N. Cristianini and J. Shawe-Taylor. *Kernel Methods for Pattern Analysis.* Cambridge University Press, 2004.

[3] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, National Taiwan University, 2003.

[4] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, 2:67–93, 2001.

[5] I. Steinwart and C. Scovel. Fast rates for support vector machines using Gaussian kernels. *Ann. Statist.*, submitted, 2004. `http://www.c3.lanl.gov/~ingo/publications/ann-04a.pdf`.

[6] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[7] E. Hille. Introduction to general theory of reproducing kernels. *Rocky Mt. J. Math.*, 2:321–368, 1972.

[8] S. Saitoh. *Integral Transforms, Reproducing Kernels and their Applications.* Longman Scientific & Technical, Harlow, 1997.

[9] H. Meschkowski. *Hilbertsche Räume mit Kernfunktion.* Springer, Berlin, 1962.

[10] R.M. Range. *Holomorphic Functions and Integral Representations in Several Complex Variables.* Springer, 1986.

[11] V. Bargmann. On a Hilbert space of analytic functions and an associated integral transform, part 1. *Comm. Pure Appl. Math.*, 14:187–214, 1961.

[12] E. Hille and R. S. Phillips. *Functional Analysis and Semi-groups.* American Mathematical Society Colloquium Publications Vol. XXXI, Providence, revised edition, 1957.

[13] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer, New York, 1996.

[14] I. Steinwart. Consistency of support vector machines and other regularized kernel machines. *IEEE Trans. Inform. Theory*, 51:128–142, 2005.

[15] Q. Wu and D.-X. Zhou. Analysis of support vector machine classification. Tech. Report, City University of Hong Kong, 2003.

[16] S. Smale and D.-X. Zhou. Estimating the approximation error in learning theory. *Anal. Appl.*, 1:17–41, 2003.

[17] I. Steinwart and C. Scovel. Fast rates for support vector machines. In *Proceedings of the 18th Annual Conference on Learning Theory, COLT 2005*, pages 279–294. Springer, 2005.

[18] I. Steinwart, D. Hush, and C. Scovel. Function classes that approximate the Bayes risk. Technical report, Los Alamos National Laboratory, 2006. `http://www.c3.lanl.gov/~ingo/pubs.shtml`.

[19] P.L. Duren. *Theory of H^p spaces.* Academic Press, 1970.

[20] S. Itô and H. Yamabe. A unique continuation theorem for solutions of a parabolic differential equation. *J. Math. Soc. Japan*, 10:314–321, 1958.

[21] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math.*, 102:159–182, 1975.