Coherent States and Wavelets in Least Squares-Support Vector Machines using HHK Transform

M. Fanuel* and J.A.K. Suykens†

KU Leuven, Department of Electrical Engineering (ESAT),

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics,

Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

July 8, 2015

Abstract

The so-called coherent states inspired from quantum physics, are introduced in the framework of Support Vector Machines. The formulation in an infinite dimensional Hilbert space is suggested, making explicit the connection between the Hilbert space in the primal formulation and the underlying Reproducing Kernel Hilbert Space associated to the kernel involved in the dual formulation. This connection is performed via a transform, named here HHK transform, which reduces to the wavelet transform, whenever the coherent states are chosen to be the wavelets. The specific role played by the coherent states and their possible definitions are then emphasized. Subsequently, a list of examples is considered in order to illustrate the advantages of the formulation, allowing for instance to recover the well known setting in the case of a Mercer kernel. As an application, the link between coherent states and wavelets is detailed, providing a novel approach using wavelet kernels in Least Squares-Support Vector Machines. Hence, a strategy for adapting the bandwidth to the data density with the help of wavelet kernels is proposed. Furthermore, using the Spectral Graph Wavelets, a semi-supervised learning problem on a graph is also addressed with the same methodology. This work combines ideas from mathematical physics, function approximation and machine learning, in order to put them under a common umbrella.

*michael.fanuel@esat.kuleuven.be

†johan.suykens@esat.kuleuven.be

1 Introduction, Motivations and Generalities

Many problems of interest in signal processing [15, 12, 16, 10], statistics [18, 19, 20], or recently in data analysis and graph theory [11, 22], have been successfully addressed within the realm of wavelet analysis [43]. This multidisciplinarity is particularly impressive. As it is well-known, the wavelets can be viewed as a particular class of the so-called "coherent states"[28, 1], borrowed from quantum physics, where they are an essential tool in many physical situations. Remarkably, quantum physics has already inspired the so-called quantum clustering method described by Horn and Gottlieb [24], while the work of Nadler *et al.*[32] provided further insights. Hence, it is interesting to investigate additional implications of these related ideas to machine learning methodologies.

In this work, we put forward a formulation connecting the framework of Support Vector Machines (SVM) and the theory of coherent states. More specifically, the primal-dual framework of Least Squares-Support Vector Machine (LS-SVM) is considered for a regression problem, while the ideas explained here can apply for generic SVM problems. Let us first give an overview of the motivations and describe the context of this paper. Support Vector Machines were designed in the context of convex optimization, originally defined with inequality constraints, with successful applications in classification problems and beyond [13, 44]. When equality constraints and L^2 loss are considered, the conditions for optimality take a much simpler form. This methodology, called Least Squares Support Vector Machines [42, 41], involves a clear-cut framework allowing to consider extensions to supervised and unsupervised learning problems [40]. Let us review the framework in the case of LS-SVM regression. Given a sample $\{(x_i, y_i)\}_{i=1}^N$ where $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ for $i = 1, \ldots, N$, the primal formulation of the LS-SVM regression reads

$$\min_{\substack{w\in\mathbb{R}^h \\ e_i,b\in\mathbb{R}}} \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^2 \quad \text{subject to} \quad y_i = w^T \varphi(x_i) + b + e_i, \tag{1}$$

with $i = 1, \ldots, N$, and where $\varphi : \mathbb{R}^d \to \mathbb{R}^h$ is the feature map. In the present case, the feature space is of dimension h. This optimization framework is advantageous since it permits to conceive another representation by regarding the Lagrange multipliers as dual variables. The connection between the primal and dual representations has proven itself very fruitful, also in the context of large data sets [17].

The dual model representation [42, 41] is

$$\hat{y} \;=\; \sum_{j=1}^{N} \alpha_j K(x, x_j) + b, \tag{2}$$

with $K(x, x') = \varphi(x)^T \varphi(x')$ for x and $x' \in \mathbb{R}^d$. The purpose of the present work is to formulate a similar problem with an infinite dimensional feature space, along the lines of Signoretto et al [36]. We choose to state the problem in the context of convex optimization in a possibly infinite dimensional Hilbert space. Let $\{\mathcal{H}, \langle\cdot|\cdot\rangle_{\mathcal{H}}\}$ be an infinite dimensional separable Hilbert space. Because of the separability condition, we can choose $\{|\phi_n\rangle\}_{n=0}^{\infty}$ an orthonormal basis of vectors for the inner product $\langle\cdot|\cdot\rangle_{\mathcal{H}}$. In the sequel, the bra-ket formalism will be used. The vectors in $|\phi\rangle \in \mathcal{H}$ will be denoted by "ket's", while their dual will be written as "bra's" $\langle\phi|$. This vector space is conveniently chosen to be a function space with a L^2 inner product. We formulate the following optimization problem

$$\min_{\substack{|w\rangle \in \mathcal{H} \\ e_i, b \in \mathbb{R}}} \frac{1}{2}\langle w|w\rangle_{\mathcal{H}} + \frac{\gamma}{2}\sum_{i=1}^{N} e_i^2 \quad \text{subject to} \quad y_i = \langle w|\eta_{x_i}\rangle_{\mathcal{H}} + b + e_i \tag{3}$$

with $i = 1, \ldots, N$. In order to make sense of the model, a vector $|\eta_x\rangle$ in the Hilbert space $\mathcal{H}$ has to be associated to each $x \in \mathbb{R}^d$. The corresponding map

$$\begin{aligned} \mathbb{R}^d &\to \mathcal{H} \\ x &\mapsto |\eta_x\rangle, \end{aligned} \tag{4}$$

should be continuous so that the out-of-sample extension remains meaningful. This intuitive continuity requirement suggests to build the function $|\eta_x\rangle$ as a coherent state, in a way completely analogous to the quantum mechanical coherent state, as used for instance in quantum optics, and which are continuously labeled quantum states (for a reference work, see for instance [28, 1]). The idea is to provide a way to formalize a generalization of the feature map

$$x \;\mapsto\; \varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \end{pmatrix}, \tag{5}$$

in the case where the feature space is infinite dimensional. In this work, we shall understand the term "coherent state" in a broad (or weak) way, and consider the following definition:

Definition 1. *Let X be a topological space, and $\mathcal{H}$ be a separable Hilbert space. The functions forming a set $\{|\eta_x\rangle \in \mathcal{H}\}_{x\in X}$ will be called coherent states if the map $x \mapsto |\eta_x\rangle$ is continuous, and if their closed linear span is $\mathcal{H}$.*

Often in the literature, the coherent states are also required to satisfy a resolution of the identity[1]. This question will be briefly discussed in the sequel, within the context of the measure-free approach. For an extensive survey of coherent states, the reader can refer to the books [28, 1], while the connection to physics is briefly outlined in the appendix.

Remark 1. *Before pursuing the discussion, it is worth to emphasize at this stage the potential interpretations of the following inner product $\langle w|\eta_{x_i}\rangle_{\mathcal{H}}$ involved in the model of the primal formulation (3). On the one hand, it can be viewed as a basis expansion as in the finite dimensional model (1). On the other hand, it is reminiscent of a wavelet transform, where $|\eta_{x_i}\rangle$ is regarded as a wavelet. More precisely, considering $|\eta_{x_i}\rangle$ as a coherent state in the primal model, one estimates a vector $|w\rangle$ in the infinite dimensional (separable) Hilbert space $\mathcal{H}$, which is the feature space.*

Furthermore, coherent states, and wavelets in particular, are naturally associated to a Reproducing Kernel Hilbert Space (RKHS), denoted here by $\mathcal{H}_K$ with inner product $\langle\cdot,\cdot\rangle_K$. In fact, there is a mapping from feature space to the RKHS, which is called in this work the HHK transform[2]:

Definition 2. *Fixing a set of coherent states $|\eta_x\rangle$, the corresponding HHK transform is the unitary map*

$$\begin{aligned} W_\eta : \mathcal{H} &\to \mathcal{H}_K \\ |w\rangle &\mapsto \ \langle \eta_\cdot|w\rangle_{\mathcal{H}}, \end{aligned} \tag{6}$$

where $(W_\eta|w\rangle)(x) = \langle\eta_x|w\rangle_{\mathcal{H}}$. The name HHK comes from the fact that the transform is a map from the Hilbert space $\mathcal{H}$ to the RKHS $\mathcal{H}_K$.

We notice that $\langle\eta_x|w\rangle_{\mathcal{H}}$ is the function introduced in the primal model in (3).

Remark 2. *A RKHS with a continuous kernel defined on a separable metric space is a separable Hilbert space of continuous functions. Hence, the Hilbert space $\mathcal{H}$ and the RKHS $\mathcal{H}_K$ are assumed to be infinite dimensional and separable. Therefore, each of them is isomorphic to the space of l_2-sequences*

[1] Coherent states without resolution of the identity are called weak coherent states[26].

[2] For simplicity, we introduce this transform as a surjective map in order to avoid technicalities in this introduction.

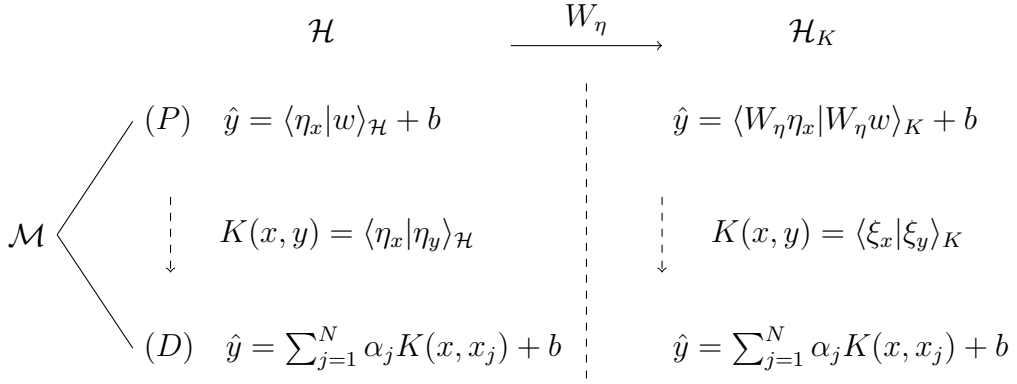


Figure 1: Schematic view of the primal-dual setting and the HHK transform. We denoted $\xi_x = W_\eta \eta_x$.

(for a reference, see for example [8]). Although $\mathcal{H}$ and $\mathcal{H}_K$ are "similar" as Hilbert spaces, they do not share the same properties and therefore they will not be identified. For this reason, we will emphasize their connection in relationship with the primal and dual settings, as schematically illustrated in Figure 1.

1.1 Conditions for optimality and solution to the regression problem

Let us assume that a specific type of coherent state has been chosen. Hence, the following optimization problem has to be solved:

$$\min_{\substack{|w\rangle \in \mathcal{H} \\ e_i, b \in \mathbb{R}}} \frac{1}{2} \langle w | w \rangle_{\mathcal{H}} + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 \quad \text{subject to} \quad y_i = \langle w | \eta_{x_i} \rangle_{\mathcal{H}} + b + e_i, \tag{7}$$

with $i = 1, \ldots, N$, and where $|\eta_x\rangle$ is a coherent state. The presence of equality constraints allows a simple formulation. Let us phrase the problem more accurately. The problem involves a convex functional

$$\mathcal{F}(|w\rangle, \mathbf{e}, b) = \frac{1}{2} \langle w | w \rangle_{\mathcal{H}} + \frac{\gamma}{2} \sum_{i=1}^N e_i^2 \tag{8}$$

and a finite set of linear constraints

$$r_i(|w\rangle, \mathbf{e}, b) = y_i - \langle w | \eta_{x_i} \rangle_{\mathcal{H}} - b - e_i, \tag{9}$$

with $i = 1, \ldots, N$. As a matter of fact, the main difficulty of this optimization problem is that it is formulated in a space of functions. However, the results

obtained in the context of Banach spaces, as detailed for example in [6], can be made more particular in the Hilbert space setting that we use here. Using the concept of Gateaux differential, it is possible to formulate a Lagrange multiplier theorem, as stated in a classification context in the reference [36]. The corresponding Lagrange functional

$$\mathcal{L}(|w\rangle, \mathbf{e}, b, \boldsymbol{\alpha}) = \mathcal{F}(|w\rangle, \mathbf{e}, b) + \sum_{i=1}^{N} \alpha_i r_i(|w\rangle, \mathbf{e}, b) \tag{10}$$

involves a set of Lagrange multipliers α_i, with $i = 1, \ldots, N$. The optimal solution satisfies the KKT conditions

$$\begin{aligned} |w\rangle &= \sum_{i=1}^{N} \alpha_i |\eta_{x_i}\rangle, & (11)\\ \alpha_i &= \gamma e_i, & (12)\\ \sum_{i=1}^{N} \alpha_i &= 0, & (13)\\ y_i &= \langle w|\eta_{x_i}\rangle_{\mathcal{H}} + b + e_i. & (14) \end{aligned}$$

Eliminating $|w\rangle$ from the formulation, one finds a linear relation

$$y_i = \sum_{j=1}^{N} \alpha_j \langle \eta_{x_j}|\eta_{x_i}\rangle_{\mathcal{H}} + b + \alpha_i/\gamma, \tag{15}$$

so that the dual variables can be obtained by merely solving a linear system. Using $\langle \eta_{x_j}|\eta_{x_i}\rangle_{\mathcal{H}} = K(x_j, x_i)$, one recovers the dual model (2) as used also in the finite dimensional case

$$\hat{y} = \sum_{j=1}^{N} \alpha_j K(x_j, x) + b. \tag{16}$$

The major advantage of the formulation is to provide a primal model for a large class of positive semi-definite kernels built from various possible coherent states, the case of multi-dimensional wavelets being of potential practical interest. Before going further, we emphasize that the primal problem as stated in (3) does not explicitly involve a Reproducing Kernel Hilbert Space in its definition. However, defining the set of functions $\{|\eta_x\rangle\}_{x\in\mathbb{R}^d}$ as a set of coherent states will naturally introduce a distinct function space, which is a RKHS. As we shall see in the sequel, a particular choice of coherent state associated to the connected affine group will naturally lead to consider

multidimensional wavelets. In the forthcoming sections, two approaches for the definition of coherent states will be briefly reviewed.

This paper is organized as follows. In order to provide the necessary concepts, Section 2 will present an overview of the various kinds of coherent states which exist in the "physics" literature in view of their application in machine learning. Section 3 will subsequently make explicit the connection between LS-SVM regression and the coherent states. Moreover, a class of kernels based on wavelets, with a bandwidth adapted to the data density, will be suggested. As a second application to learning, a method relying on Spectral Graph Wavelets for semi-supervised learning in undirected graphs will be detailed in Section 4. Some numerical examples and applications are gathered in Section 5. Finally, the last section before the conclusions will be devoted to a discussion of related works.

2 Definition of the coherent states

We shall review the construction of coherent states from two perspectives:

- First of all, the existence of a RKHS will be assumed in order to construct a set of coherent states in the Hilbert space $\mathcal{H}$, while the case of a Mercer kernel associated to an infinite dimensional feature space will be taken as a motivation. Furthermore, the issue of the resolution of the identity will be discussed.

- Secondly, a classic construction of coherent states associated to a group will be explained, and the specific case of the connected affine group will be emphasized in connection with wavelet theory.

Both constructions will be illustrated in the forthcoming sections, whereas the relationship with quantum physics will be succinctly outlined in the Appendix.

2.1 Coherent state in $\mathcal{H}$ from a Reproducing Kernel Hilbert Space

In order to define the continuous map given in (4), one possible manner is to construct the application by means of another distinct Hilbert space, which has the property to be a RKHS. In order to explain the construction, we shall review well-known properties of a RKHS.

Following the references [1, 2], we introduce an abstract Reproducing Kernel Hilbert Space $\{\mathcal{H}_K, \langle\cdot|\cdot\rangle_K\}$. More precisely, a RKHS is defined as the

vector space of functions $f : \mathbb{R}^d \to \mathbb{R}$ with the property that the evaluation functional

$$\begin{aligned} E_x : \mathcal{H}_K &\to \mathbb{R} \\ f &\mapsto f(x) \end{aligned} \tag{17}$$

is continuous. By the representation theorem of continuous linear functionals, there exists $\xi_x \in \mathcal{H}_K$ such that $f(x) = \langle \xi_x | f \rangle_K$. The reproducing kernel is defined by $K(x, y) = \langle \xi_x | \xi_y \rangle_K$. According to the reference [2] that we follow closely, the vector ξ_x can be called the "coherent state" associated to the Reproducing Kernel Hilbert Space $\{\mathcal{H}_K, \langle \cdot | \cdot \rangle_K\}$. The space generated by linear combinations of the vectors ξ_x is dense in $\mathcal{H}_K$.

Introducing the orthonormal basis $\{\psi_n\}_{n=0}^{\infty}$ of $\mathcal{H}_K$, it is straightforward to see that the vector ξ_x admits an expansion

$$\xi_x = \sum_{n=0}^{\infty} c_n \psi_n, \tag{18}$$

where $c_n = \langle \xi_x | \psi_n \rangle_K = \psi_n(x)$. As a consequence of the relation $K(x, y) = \langle \xi_x | \xi_y \rangle_K$, the kernel admits the pointwise convergent expansion

$$K(x, y) = \sum_{n=0}^{\infty} \psi_n(x) \psi_n(y), \tag{19}$$

which is independent of the basis chosen, as it is well known. Let us now define the coherent state $|\eta_x\rangle \in \mathcal{H}$, by introducing an isometric linear map between $\mathcal{H}$ and $\mathcal{H}_K$.

Proposition 1. *[2] Let $\{\psi_n\}_{n=0}^{\infty}$ be an orthonormal basis of $\mathcal{H}_K$ and $\{\phi_n\}_{n=0}^{\infty}$ an orthonormal basis of $\mathcal{H}$. The linear map*

$$\begin{aligned} V : \mathcal{H}_K &\to \mathcal{H} \\ \psi_n &\mapsto |\phi_n\rangle. \end{aligned} \tag{20}$$

is an isometry.

A coherent state $|\eta_x\rangle \in \mathcal{H}$ is given by the image of a "coherent state" $\xi_x \in \mathcal{H}_K$ by V. More explicitly, using (18), we can write the definition of a non-normalized coherent state

$$|\eta_x\rangle \triangleq V \xi_x = \sum_{n=0}^{\infty} \psi_n(x) |\phi_n\rangle. \tag{21}$$

Consequently, it is straightforward to verify that the application $x \mapsto |\eta_x\rangle$ is continuous. That is to say, if $x_n \to x$ in $\mathbb{R}^d$, then $\| \ |\eta_x\rangle - |\eta_{x_n}\rangle\|_{\mathcal{H}} \to 0$. Furthermore, as a consequence of the property $K(x,y) = \langle \xi_x | \xi_y \rangle_K$, the overlap of two coherent states in $\mathcal{H}$ gives the reproducing kernel $\langle \eta_x | \eta_y \rangle_{\mathcal{H}} = K(x,y)$, even though $\mathcal{H}$ is not a Reproducing Kernel Hilbert Space.

As explained previously, we defined a linear map $V : \mathcal{H}_K \to \mathcal{H}$ between the two different Hilbert spaces. Similarly, it is possible to define a map going in the inverse direction.

Proposition 2. *The HHK transform, corresponding to the wavelet transform in the context of wavelet analysis[2],*

$$\begin{aligned} W_\eta : \mathcal{H} &\to \ \mathcal{H}_K \\ |\phi\rangle &\mapsto \ \langle \eta_. | \phi \rangle_{\mathcal{H}}, \end{aligned} \tag{22}$$

explicitly given by $(W_\eta |\phi\rangle)(x) = \langle \eta_x | \phi \rangle_{\mathcal{H}}$. *The adjoint of* W_η *defined by*

$$W_\eta^* : \mathcal{H}_K \to \mathcal{H}, \tag{23}$$

is given by $W_\eta^* = V$. *In particular, one has* $(W_\eta |\phi_n\rangle)(x) = \psi_n(x)$, *so that* W_η *is in fact a unitary transformation.*

2.1.1 Resolution of the identity and measure-free coherent states

Finally, let us discuss the question of existence of a resolution of the identity.

- On the one hand, if the inner product in $\mathcal{H}_K$ is an L^2 inner product, that is to say, for f and $g \in \mathcal{H}_K$, their inner product is of the form

$$\langle f | g \rangle_K = \int_X \mathrm{d}\mu(x) f(x) g(x), \tag{24}$$

  with respect to a specified absolutely continuous measure μ, then the coherent states in $\mathcal{H}$ satisfy a resolution of the identity

$$\int_X \mathrm{d}\mu(x) \ |\eta_x\rangle\langle\eta_x| = \mathbb{I}_{\mathcal{H}}, \tag{25}$$

  where X is a subspace of $\mathbb{R}^d$ in the present case. This indicates that the coherent states form a complete set of functions[3]. The relation (25) is understood to converge weakly, i.e., for any $|\phi_1\rangle$ and $|\phi_2\rangle \in \mathcal{H}$, we have

$$\int_X \mathrm{d}\mu(x) \ \langle \phi_1 | \eta_x \rangle \langle \eta_x | \phi_2 \rangle = \langle \phi_1 | \phi_2 \rangle. \tag{26}$$

[3]Actually, they form an overcomplete set.

Using the definition of the inner product (24), the resolution of the identity in the weak sense is viewed as the following property: for any $|\phi_1\rangle$ and $|\phi_2\rangle \in \mathcal{H}$,

$$\langle W_\eta \phi_1 | W_\eta \phi_2 \rangle_K = \langle \phi_1 | \phi_2 \rangle_{\mathcal{H}}. \tag{27}$$

- On the other hand, in the delicate case of a RKHS whose inner product is not of the L^2-type, or if there are ambiguities in determination of the measure μ, one has to analyse the resolution of the identity in a measure-free approach as explained in the paper [25]. The authors of this work argue that, given a set of coherent states without measure, the existence of the unitary transformation $W_\eta : \mathcal{H} \to \mathcal{H}_K$ realizing the relation

$$\langle W_\eta \phi_1 | W_\eta \phi_2 \rangle_K = \langle \phi_1 | \phi_2 \rangle_{\mathcal{H}}, \tag{28}$$

for any $|\phi_1\rangle$ and $|\phi_2\rangle \in \mathcal{H}$, gives the exact analogue of the resolution of the identity (25). The analogy with the integral formula (26) is made explicit if the relation $(W_\eta|\phi\rangle)(x) = \langle \eta_x | \phi \rangle_{\mathcal{H}}$ for $|\phi\rangle \in \mathcal{H}$ is substituted in the measure-free expression (28).

2.2 Coherent state in $\mathcal{H}$ associated to a group

Another typical construction of coherent states, of particular relevance in physics and signal analysis, involves a square integrable representation of a group. These states, coming with a natural unique measure, are often called Gilmore-Perolomov coherent states. Following the references [1, 4, 3], we fix a locally compact group G, and μ the left-invariant Haar measure of G. Let us introduce a strongly continuous, irreducible, unitary representation U of G on the Hilbert space $\mathcal{H}$. Choosing a fiducial vector, called the mother wavelet in signal analysis, $|\eta\rangle \in \mathcal{H}$, we define the set of coherent states as the vectors of the form

$$|\eta_g\rangle \triangleq U(g)|\eta\rangle, \quad \text{with } g \in G. \tag{29}$$

Varying the fiducial state will give rise to different sets of coherent states associated to different RKHS's. There is however a restriction on the choice of $|\eta\rangle$. The admissibility condition of the fiducial vector is

$$c(\eta) = \int_G \mathrm{d}\mu(g)\ |\langle \eta | U(g) \eta \rangle|^2 < \infty. \tag{30}$$

Provided the fiducial state is admissible, the coherent states admit a resolution of the identity

$$\frac{1}{c(\eta)}\int_G \mathrm{d}\mu(g)\ |\eta_g\rangle\langle\eta_g| = \mathbb{I}_{\mathcal{H}}, \tag{31}$$

which is understood in terms of weak convergence. Let us now describe how the Hilbert space $\mathcal{H}$ and the RKHS $\mathcal{H}_K$ are related in the case of group generated coherent states. Fixing a fiducial state, we introduce the HHK transform, analogous to (22),

$$\begin{aligned} W_\eta : \mathcal{H} &\to L^2(G, c(\eta)^{-1}\mathrm{d}\mu) \\ |\phi\rangle &\mapsto \ \langle \eta_{\centerdot}|\phi\rangle_{\mathcal{H}}, \end{aligned} \tag{32}$$

with $W_\eta(|\phi\rangle)(g) = \langle\eta_g|\phi\rangle_{\mathcal{H}}$ for $g \in G$. However, let us remark that, contrary to the definition (22), the map W_η is not surjective. Hence, we define $\mathcal{H}_K$ as the range of W_η, that is to say

$$\mathcal{H}_K = W_\eta \mathcal{H}, \tag{33}$$

which is a RKHS, associated to the reproducing kernel $K(g, g') = \langle\eta_g|\eta_{g'}\rangle_{\mathcal{H}}$. Considering now $\psi \in \mathcal{H}_K$, one can write a reconstruction formula. A vector in the RKHS is easily seen to be of the form $\psi(\centerdot) = \langle\eta_{\centerdot}|\phi\rangle_{\mathcal{H}}$ and can be mapped to a vector in $\mathcal{H}$ by the rule

$$\begin{aligned} V_\eta : \mathcal{H}_K &\to \mathcal{H} \\ \psi &\mapsto \ V_\eta\psi, \end{aligned} \tag{34}$$

which is the analogue of (20), where

$$V_\eta\psi = \frac{1}{c(\eta)}\int_G \mathrm{d}\mu(g)\ \psi(g)|\eta_g\rangle = |\phi\rangle. \tag{35}$$

Let us mention that we consider here $\mathcal{H}_K$ as embedded in $L^2(G, c(\eta)^{-1}\mathrm{d}\mu)$ somehow in the spirit of the book [16], while reference [1] normalizes W_η in order to absorb the factor $c(\eta)^{-1}$ and therefore eliminates it from the integration measure over the group. According to [1], we have the following result:

Theorem 1. *[1] The linear map $W_\eta : \mathcal{H} \to L^2(G, c(\eta)^{-1}\mathrm{d}\mu)$ is an isometry, that is to say $W_\eta^* W_\eta = \mathbb{I}_{\mathcal{H}}$, where*

$$\begin{aligned} W_\eta^* : L^2(G, c(\eta)^{-1}\mathrm{d}\mu) &\to \mathcal{H} \\ \psi \quad &\mapsto \frac{1}{c(\eta)}\int_G \mathrm{d}\mu(g)\ \psi(g)|\eta_g\rangle. \end{aligned} \tag{36}$$

The reconstruction formula is given by the adjoint map restricted to $\mathcal{H}_K$ $W_\eta^*|_{\mathcal{H}_K} = V_\eta$. *Furthermore,* $\mathbb{P}_\eta = W_\eta W_\eta^*$ *is a projector:*

$$\mathbb{P}_\eta : L^2(G, c(\eta)^{-1}\mathrm{d}\mu) \to \mathcal{H}_K. \tag{37}$$

This construction may appear abstract, nevertheless it is at the basis of the various coherent states in physics as well as the wavelet analysis in signal processing, as we will see in the examples below.

3 Coherent states, multidimensional wavelets and HHK transform in Least Squares Support Vector Machines

Once a coherent state of some type $|\eta_x\rangle \in \mathcal{H}$ is specified, one proceed with the solving of the optimization problem

$$\min_{\substack{|w\rangle\in\mathcal{H}\\ e_i,b\in\mathbb{R}}} \frac{1}{2}\langle w|w\rangle_{\mathcal{H}} + \frac{\gamma}{2}\sum_{i=1}^{N} e_i^2 \quad \text{subject to} \quad y_i = \langle w|\eta_{x_i}\rangle_{\mathcal{H}} + b + e_i, \tag{38}$$

for $i = 1, \ldots, N$. Let us review a list of possible choices of $|\eta_x\rangle$ and the corresponding transforms.

3.1 The case of a Mercer kernel and its associated RKHS

We shall relate in this section our construction to the way the RKHS is used in statistical data analysis. The Hilbert space $\mathcal{H}_K$ can be defined starting from a continuous, symmetric and positive semi-definite kernel (Mercer)[31, 14], which is used in order to define an integral operator on square integrable functions. Important instances of such kernels are the exponential and RBF kernels.

Definition 3. *Let* $X \subset \mathbb{R}^d$. *A symmetric kernel* $K : X \times X \to \mathbb{R}$ *is positive semi-definite, if for any finite set* $\{x_i \in X,\ i = 1, \ldots, n\}$ *and any finite sequence* $\{\alpha_i \in \mathbb{R},\ i = 1, \ldots, n\}$,

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0. \tag{39}$$

This definition [14] is equivalent to the one of Aronszajn in reference [5].

For the sake of completeness, let us mention that we shall only consider here kernels which are genuinely pointwise defined functions in the sense of Aronszajn. As explained for instance in [9], this situation is a particular case of a more abstract framework introduced by Schwartz in [35].

A typical example of a kernel associated to an infinite dimensional feature space is given by the RBF kernel $k^{\sigma}_{RBF}(x';x) = \exp(-\|x-x'\|^2/4\sigma^2)$. Let us relate the usual construction [23] to the one of the coherent states. The real-valued kernel $K(x,x')$ is considered as a linear integral operator on a L^2-space of real valued functions. In order to define the inner product, the eigenfunctions and eigenvalues of the integral operator are calculated as follows

$$\int K(x,x')f_n(x')\mathrm{d}\rho(x') = \lambda_n f_n(x), \tag{40}$$

where the integration is taken according to an absolutely continuous positive measure ρ. Hence, the absolutely convergent spectral representation of the kernel reads

$$K(x,x') = \sum_{n=0}^{\infty} \lambda_n f_n(x) f_n(x'), \tag{41}$$

where the eigenfunctions f_n are orthonormal. Therefore, the inner product in $\mathcal{H}_K$ is defined by

$$\langle f|g\rangle_K = \sum_{n=0}^{\infty} \frac{1}{\lambda_n} \langle f|f_n\rangle_{L^2} \langle f_n|g\rangle_{L^2}. \tag{42}$$

As a result the Hilbert space $\mathcal{H}_K$ is considered as the set of L^2-functions verifying $\langle f|g\rangle_K < \infty$.

In this setting, the coefficients $\psi_n(x)$ in the definition (21) are seen as the components $\varphi_n(x)$ of the feature map in the equation (5), with the property: $\psi_n(x) = \sqrt{\lambda_n} f_n(x)$, where f_n is the eigenfunction with eigenvalue λ_n of the kernel integral operator.

Remark 3. *As a result, it is worth emphasizing that the feature map (5) in the case of an infinite dimensional feature space can indeed be understood as mapping a data point to a specific coherent state as stated in (4).*

Finally, since the inner product in this case is not an L^2 inner product, the vectors $|\eta_{x_i}\rangle \in \mathcal{H}$ do not satisfy a resolution of the identity with the integral formula (25), as explained previously. These coherent states can then be considered as measure-free and satisfy the property given in (28). The authors of reference [1, 2] explain that only a "symbolic" resolution of the identity with an integral symbol can still be introduced.

3.2 The case of multidimensional canonical coherent states and the windowed Fourier transform

An elementary example of coherent states is given by the d-dimensional canonical coherent states, belonging to the category of group generated coherent states, the corresponding group being the Heisenberg group (for more details see [28, 1]). Let us define the complex valued coherent state $|\eta^{\sigma}_{p,x}\rangle \in \mathcal{H} = L^2(\mathbb{R}^n, \mathbb{C})$, as follows

$$\eta^{\sigma}_{p,x}(q) = (\pi\sigma^2)^{-d/4} e^{\mathrm{i}(q-x/2)\cdot p} e^{-\frac{1}{2}\|x-q\|^2/\sigma^2}, \tag{43}$$

for x and $p \in \mathbb{R}^d$. Among their properties, they verify in particular the properties of group generated coherent states summarized in the previous section. If we fix $|w\rangle \in \mathcal{H}$, the following inner product

$$\langle w|\eta^{\sigma}_{p,x}\rangle_{\mathcal{H}} = (\pi\sigma^2)^{-d/4} e^{-\mathrm{i}p\cdot x/2} \int_{\mathbb{R}^d} \mathrm{d}q \; w^*(q) e^{\mathrm{i}p\cdot q} e^{-\frac{1}{2}\|q-x\|^2/\sigma^2}, \tag{44}$$

can be interpreted as a d-dimensional windowed Fourier transform (or Gabor transform), whose relation to canonical coherent states is well-known. The reproducing kernel is given by

$$K^{\sigma}(p', x'; p, x) = e^{\mathrm{i}(x'\cdot p - p'\cdot x)/2} e^{-\sigma^2\|p-p'\|^2/4} e^{-\|x-x'\|^2/4\sigma^2}. \tag{45}$$

Noticeably, choosing to compute the windowed Fourier transform for the "zero frequency", i.e. if we choose to define the optimization problem (38) with

$$y_i = \langle w|\eta^{\sigma}_{0,x_i}\rangle_{\mathcal{H}} + b + e_i, \tag{46}$$

in the primal formulation will give rise to an RBF kernel

$$K^{\sigma}(0, x'; 0, x) = e^{-\|x-x'\|^2/4\sigma^2} = k^{\sigma}_{RBF}(x'; x), \tag{47}$$

in the dual formulation. Actually, the RBF kernel $k_{RBF}(x'; x)$ is considered here in a restrictive sense, as the function entering the solution of the optimization problem (38). Actually, $k_{RBF}(x'; x)$ is not understood here as the reproducing kernel of a RKHS of real valued functions, as it can be used in the framework of "estimation of function in a RKHS". This is a major difference in terms of interpretation. Actually, in the present example, the Figure 1 is valid if the coherent state in the primal formulation is chosen to be $|\eta^{\sigma}_{0,x}\rangle$.

It is worth mentioning that Steinwart *et al* [38] have studied in details the RKHS's associated to the complex RBF kernel, with a particular emphasis on the connection between the complex and real RBF kernels with the Fock-Bargmann-Segal space. The relation with the coherent states of the harmonic oscillator and the Fock-Bargmann-Segal space is outlined in the appendix.

3.3 The case of multidimensional wavelets and the continuous wavelet transform

The example that we are going to describe involves the coherent states in $\mathcal{H} = L^2(\mathbb{R}^d)$, associated with a group. The natural RKHS structure of the wavelet transform will be exploited. We suggest to consider d-dimensional spherically symmetric wavelets, following the reference [16]. In the sequel, we shall denote the mother wavelet $|\eta\rangle$ by the spherically symmetric function $\eta(\|x\|)$, corresponding to the position space representation of $|\eta\rangle$ in the physical jargon[4]. The mother wavelet $|\eta\rangle \in L^2(\mathbb{R}^d, \mathrm{d}x)$, has to satisfy the admissibility condition

$$c(\eta) = \int_0^\infty \frac{\mathrm{d}r}{r} |\hat{\eta}(r)|^2 < \infty \tag{48}$$

where $\hat{\eta}(\|p\|)$ is the spherically symmetric Fourier transform of $\eta(\|y\|)$.

The unitary representation of the group of translations and dilations [27] of $\mathbb{R}^d$ is given by

$$U(b,a)\eta(x) = a^{-d/2}\eta(\frac{x-b}{a}) = \eta_{a,b}(x), \tag{49}$$

with $b \in \mathbb{R}^d$ and $a > 0$. Naturally, the law of the connected affine group is represented by $U(b,a)U(b',a') = U(b+ab', aa')$. Let us write $|\eta_{a,b}\rangle = \eta_{a,b}(\cdot)$. The resolution of the identity is

$$\frac{1}{c(\eta)} \int_0^\infty \frac{\mathrm{d}a}{a^{d+1}} \int_{\mathbb{R}^d} \mathrm{d}b \; |\eta_{a,b}\rangle\langle\eta_{a,b}| = \mathbb{I}_{\mathcal{H}}. \tag{50}$$

As a consequence of the theorem for the case of group generated coherent states (69), we have the following corollary,

Corollary 1. *The wavelet transform*

$$\begin{aligned} W_\eta : \mathcal{H} \to \; & L^2(\mathbb{R}_0^+ \times \mathbb{R}^d, c(\eta)^{-1} a^{-d-1} \mathrm{d}a\mathrm{d}b) \\ |\phi\rangle \mapsto \; & \langle \eta_. | \phi \rangle_{\mathcal{H}}, \end{aligned} \tag{51}$$

is an isometry.

The reproducing kernel Hilbert space $\mathcal{H}_K$ is given by the square integrable functions of the type

$$\psi(a,b) = \langle \eta_{a,b} | \phi \rangle_{\mathcal{H}} \in \mathcal{H}_K \subset L^2(\mathbb{R}_0^+ \times \mathbb{R}^d, c(\eta)^{-1} a^{-d-1} \mathrm{d}a\mathrm{d}b), \tag{52}$$

[4]Often in the physical literature, one writes $\eta(y) = \langle y|\eta\rangle$, which may appear as confusing in the present context.

with the reproducing kernel $K(a,b;a',b') = \langle \eta_{a,b} | \eta_{a',b'} \rangle_{\mathcal{H}}$. Explicitly, let ψ_1 and $\psi_2 \in \mathcal{H}_K$, then their inner product is given by

$$\langle \psi_2 | \psi_1 \rangle_K = \frac{1}{c(\eta)} \int_0^\infty \frac{\mathrm{d}a}{a^{d+1}} \int_{\mathbb{R}^d} \mathrm{d}b \; \psi_2(a,b) \psi_1(a,b). \tag{53}$$

Finally, the inverse of the wavelet transform restricted on $\mathcal{H}_K$, i.e. the reconstruction formula (34), reads

$$\frac{1}{c(\eta)} \int_0^\infty \frac{\mathrm{d}a}{a^{d+1}} \int_{\mathbb{R}^d} \mathrm{d}b \; |\eta_{a,b}\rangle \langle \eta_{a,b} | \phi \rangle_{\mathcal{H}} = |\phi\rangle. \tag{54}$$

As a consequence, the kernel function associated to a choice of mother wavelet will be given by

$$K(a',x';a,x) = \langle \eta_{a',x'} | \eta_{a,x} \rangle_{\mathcal{H}}, \tag{55}$$

with the scales $a > 0$ and $a' > 0$. The construction guarantees that the wavelet kernels obtained are positive semi-definite, as a consequence of the RKHS theory [5]. A continuous, symmetric and positive semi-definite kernel is called a Mercer kernel [14]. More precisely, for any $\{(a_i, x_i) \in \mathbb{R}_0^+ \times \mathbb{R}^d, \quad i = 1, \ldots, n\}$ and for any $\{c_i \in \mathbb{R}, \quad i = 1, \ldots, n\}$, we have then the condition on the Gramian matrix of the kernel

$$\sum_{i,j=1}^{n} c_i c_j K(a_i, x_i; a_j, x_j) = \langle \sum_{j=1}^{n} c_j K(\cdot; a_j, x_j) | \sum_{i=1}^{n} c_i K(\cdot; a_i, x_i) \rangle_K \geq 0, \tag{56}$$

thanks to the reproducing property of the kernel. Thus, in order to formulate our LS-SVM regression, we can choose to take a fixed scale $a > 0$, so that we have the continuous map,

$$\begin{aligned} \mathbb{R}^d &\to \mathcal{H} \\ x &\mapsto |\eta_{a,x}\rangle. \end{aligned} \tag{57}$$

Actually, this example provides a new possible way to introduce wavelets in Support Vector Machines, by considering the wavelets already from the primal formulation. In this example, the Figure 1 illustrates the role of the HHK transform in the primal-dual framework, where the coherent state in the primal is taken to be $|\eta_{a,x}\rangle$ for a fixed scale $a > 0$. It is worth emphasizing that a whole family of positive definite kernels could be constructed by considering various possible mother wavelets.

Example 1. *An explicit form of the kernel (55) can be found for instance in the case of the Marr wavelet (or Mexican hat) in d-dimensional space*

$$\eta(\|x\|) = (d - \|x\|^2)e^{-\|x\|^2/2} = -\Delta_d(e^{-\|x\|^2/2}), \tag{58}$$

giving, after integration, the symmetric positive definite kernel

$$\tilde{K}_a(x', x) = \pi^{d/2} p(\frac{\|x - x'\|}{a}) \exp(-\frac{\|x - x'\|^2}{4a^2}), \tag{59}$$

where, $a > 0$ is interpreted as a scale analogous to the one in the RBF kernel and with the polynomial

$$p(\zeta) = \frac{\zeta^4}{16} - \frac{d}{2}\zeta^2 + \frac{d}{2}(\frac{d}{2} + 1). \tag{60}$$

Beyond the analogy with the RBF kernel, the Marr wavelet kernel, as illustrated in Figure 2, also involves the dimension of the input space $\mathbb{R}^d$. In this sense, it is possible to choose the kernel in order to adapt it according to the d-dimensional data points $x \in \mathbb{R}^d$.

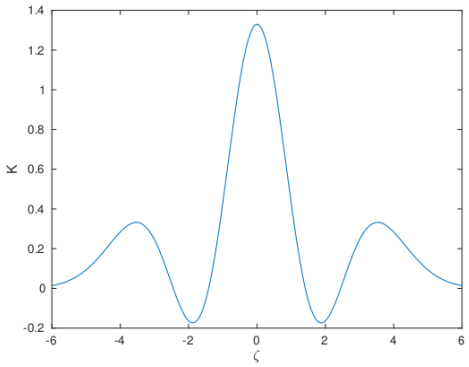


Figure 2: The Marr kernel of (59) for $a = 1$ and $d = 1$.

Example 2. *Considering again the Marr wavelet, a potential kernel choice is given by the overlap (55) where $a \neq a'$. Hence, in this picture, a scale has to be associated to each data point. Fixing a set $\{(a_i, x_i)\}_{i=1}^N$, the wavelet kernel is then given by*

$$K(a', x'; a, x) = (2\pi)^{d/2}\Big(\frac{aa'}{a^2 + a'^2}\Big)^{2+d/2} q(\frac{\|x - x'\|}{\sqrt{a^2 + a'^2}}) \exp(-\frac{\|x - x'\|^2}{2(a^2 + a'^2)}), \tag{61}$$

with the polynomial

$$q(\zeta) = \zeta^4 - 4d\zeta^2 + d(d + 2). \tag{62}$$

This example potentially provides an insight to improve the kernel in a data dependent way, as it is for example suggested in reference [45] within a classification context. Let us develop a heuristic argument. In a practical perspective, we suggest to take for the scale $a_i > 0$ associated to the data point x_i, the "average distance" from the M-th nearest neighbours,

$$a_i = \frac{1}{M} \sum_{j=1}^{M} \|x_i - x_j\|. \tag{63}$$

This means that, in regions where data points are very dense, the scale (or width) of the wavelet would be smaller than is sparser regions. This proposition will be illustrated in a following section.

4 Learning on a graph with spectral graph wavelets

We shall here consider a connected undirected graph $G = \{V, E\}$, where V is the set of $|V|$ vertices and E the set of edges. Each edge is also weighted and let w_{uv} be the weight corresponding to the edge between the vertices u and v. It is customary to introduce the matrix of the weights is given by $(W)_{uv} = w_{uv}$, while the diagonal degree matrix is $D = \mathrm{diag}(d(u))$ with $d(u) = \sum_{\{v|(u,v)\in E\}} w_{uv} \geq 0$. We shall denote $\mathcal{H}$ the Hilbert space of complex valued functions on V with the inner product $\langle f|g\rangle_{\mathcal{H}} = \sum_{v\in V} f^*(v)g(v)$. From a computational perspective, this finite dimensional Hilbert space can be viewed as equivalent to the vector space of complex valued vectors. For simplicity, the unnormalized Laplacian $L = D - W$ will be considered, while the case of the normalized Laplacian will not be studied here.

The eigenvectors of the Laplacians are denoted by $\{|\chi_\ell\rangle \in \mathcal{H}\}_{\ell=0}^{|V|-1}$ and satisfy

$$\mathcal{L}|\chi_\ell\rangle = \lambda_\ell|\chi_\ell\rangle, \quad \text{for} \quad \ell = 0, \ldots, |V| - 1, \tag{64}$$

whereas we assume that the eigenspaces are not degenerate, i.e. the eigenvalues' multiplicity is one. Following the reference [22] and adapting slightly our notations, we introduce the Spectral Graph Wavelet at the scale $t > 0$ localized at $u \in V$,

$$|\psi_{t,u}^{(g)}\rangle = g(t\mathcal{L})|\delta_u\rangle = \sum_{\ell=0}^{|V|-1} g(t\lambda_\ell)\chi_\ell^*(u)|\chi_\ell\rangle \in \mathcal{H}. \tag{65}$$

The role of the mother wavelet is played here by the admissible function g verifying $g(0) = 0$ and $\int_0^{+\infty} \mathrm{d}t\, \frac{g^2(t)}{t} = c_g < \infty$.

Proposition 3. *[22] The spectral graph wavelets $|\psi_{t,v}\rangle$ satisfy the "resolution of the identity"*

$$\frac{1}{c_g}\int_0^{+\infty}\frac{\mathrm{d}t}{t}\sum_{v\in V}|\psi_{t,v}\rangle\langle\psi_{t,v}| = \mathbb{I}_{\mathcal{H}} - \mathbb{P}_0, \tag{66}$$

where $\mathbb{P}_0$ is the projector on the null space of the Laplacian L.

This result provides a reconstruction formula

$$f = \frac{1}{c_g}\int_0^{+\infty}\frac{\mathrm{d}t}{t}\sum_{v\in V}W_f(t,v)|\psi_{t,v}\rangle + \mathbb{P}_0 f. \tag{67}$$

The spectral graph wavelet (65) is exactly a finite dimensional coherent state of the form (21), associated to the RKHS of reproducing kernel $K(t',v';t,v) = \langle\psi_{t',u'}|\psi_{t,u}\rangle_{\mathcal{H}}$, whose spectral representation is

$$K(t',v';t,v) = \sum_{\ell=0}^{|V|-1} g(t'\lambda_\ell)g(t\lambda_\ell)\chi_\ell^*(v)\chi_\ell(v'). \tag{68}$$

The Spectral Graph Wavelet Transform is a particular instance of the HHK transform

$$\begin{aligned} W_g : \mathcal{H} &\to L^2(\mathbb{R}_0^+\times V, \mathrm{d}\rho) \\ |f\rangle &\mapsto \ \langle\psi_\cdot^{(g)}|f\rangle_{\mathcal{H}}, \end{aligned} \tag{69}$$

where the transform is explicitly given by $(W_g|f\rangle)(t,v) = \langle\psi_{t,v}^{(g)}|f\rangle_{\mathcal{H}}$. The span of W_g is a RKHS $\mathcal{H}_K$ whose reproducing kernel is given by (68). The inner product for ψ_2 and $\psi_1 \in L^2(\mathbb{R}_0^+\times V, \mathrm{d}\rho)$ is naturally defined by

$$\langle\psi_2|\psi_1\rangle_{L^2} = \frac{1}{c_g}\int_0^{+\infty}\frac{\mathrm{d}t}{t}\sum_{v\in V}\psi_2(t,v)\psi_1(t,v). \tag{70}$$

Noticeably the transform is not onto.

The LS-SVM semi-supervised problem on a graph can be formulated as a regression problem, although the model is used to classify vertices as follows. There are N data points $\{(v_i,y_i)\}_{i=1}^N$, where $v_i \in V$ is a vertex of the graph, to which $y_i = \pm 1$ is associated for each $i = 1,\dots,N$. The problem is to find the unknown values y_i associated to the vertices $v_i \in V$ for $i = N+1,\dots,|V|$. Fixing a scale $t > 0$, the optimization framework is given by (3),

$$\min_{\substack{|w\rangle\in\mathcal{H}\\ e_i,b\in\mathbb{R}}} \frac{1}{2}\langle w|w\rangle_{\mathcal{H}} + \frac{\gamma}{2}\sum_{i=1}^N e_i^2 \quad \text{subject to} \quad y_i = \langle\psi_{t,v_i}|w\rangle_{\mathcal{H}} + b + e_i, \tag{71}$$

with $i = 1, \ldots, N$.

Solving the linear system

$$y_n = \sum_{i=1}^{N} \alpha_i K(t, v_i; t, v_n) + b + \alpha_i/\gamma, \quad \text{for} \quad n = 1, \ldots, N, \tag{72}$$

allows to determine dual variables α_i for $i = 1, \ldots, N$. The solution in the primal reads $|w\rangle = \sum_{i=1}^{N} \alpha_i |\psi_{t,v_i}\rangle$, while the model predicts in practice the unknown $y_i \in \{-1, 1\}$ for $i = N+1, \ldots, |V|$ thanks to $\hat{y}_i = \text{sign}(\langle w|\psi_{t,v_i}\rangle + b)$. Here, we train the model on the labeled vertices and use the out-of-sample extension of the model to predict on the remaining vertices. However, we can notice that the feature map contains information about all vertices.

Remark 4. *Noticeably, there is an apparent similarity between the kernel constructed by the overlap of the graph wavelets*[5]

$$K(t, u; t, v) = \sum_{\ell=0}^{|V|-1} g^2(t\lambda_\ell) \chi_\ell(u) \chi_\ell^*(v), \tag{73}$$

and the spectral representation of the graph Laplacian

$$\langle \delta_u | \Delta | \delta_v \rangle_{\mathcal{H}} = \sum_{\ell=0}^{|V|-1} \lambda_\ell \ \chi_\ell(u) \chi_\ell^*(v), \tag{74}$$

where $\lambda_0 = 0$. As a matter of fact, the kernel obtained thanks to the spectral graph wavelets belongs to the family of kernels studied in the reference [37, 29]. Actually, the method presented here is different from the semi-supervised learning problem of graphs based on a Laplacian smoothing functional [7]. More precisely, the latter method consists in learning $f \in \mathcal{H}(V)$ by solving

$$\min_{f \in \mathcal{H}(V)} \ \frac{1}{2} \langle f | \Delta f \rangle_{\mathcal{H}} + \frac{\gamma}{2} ||f - y||^2, \tag{75}$$

where we have of course $\langle f|\Delta f\rangle = \sum_{u,v \in V} f(u) \langle \delta_u|\Delta|\delta_u\rangle_{\mathcal{H}} f(v)$. The function $y \in \mathcal{H}(V)$ is chosen to satisfy $y(v_i) = 0$ for the unlabeled vertices $v_i \in V$ with $i = N+1, \ldots, |V|$. This assumption, although "reasonable", is not needed in our method since no values are assigned to the unlabeled points.

From a practical point of view, the computation of the eigenvectors of the graph Laplacian becomes excessive for large graphs, i.e. involving more

[5] This kernel is the reproducing kernel in the space of wavelet transforms.

than thousand nodes. Hence, although the Spectral Graph Wavelets[6] are defined in Fourier space, the authors of reference [22] suggest to compute the wavelets directly in the vertex space with the help of a Chebychev polynomial approximation of the function g in (65). Hence, the numerical diagonalisation of the Laplacian matrix is avoided, while the so-called "fast Spectral Graph Wavelet Transform" only applies polynomials of the Laplacian to input vectors. For the details of the construction of the wavelets, we refer to the paper [22].

5 Numerical examples

In this section, we provide numerical illustrations of the LS-SVM problems with kernels obtained from wavelets.

5.1 Marr wavelet kernel

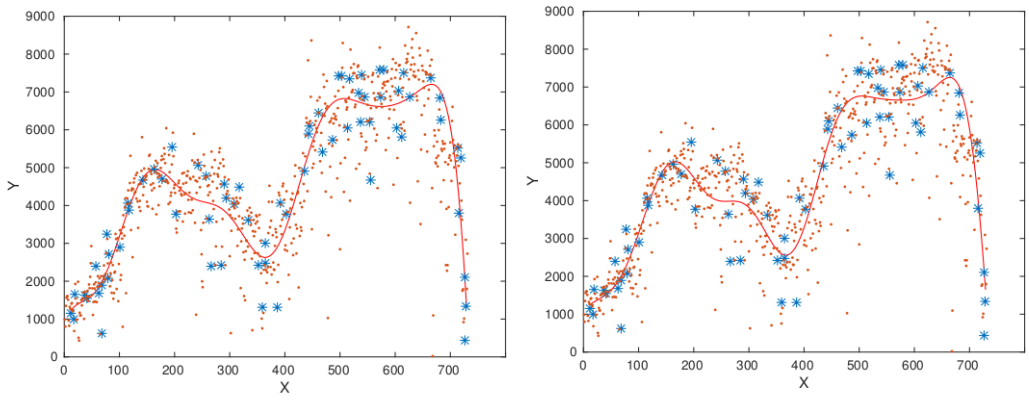
In a first stage, the performance of the Marr kernel, given in (59), is compared to the Gaussian RBF kernel. In order to illustrate the technique, two data sets with only one feature are considered. Firstly, a sample corresponding to a function $y = \mathrm{sinc}(x)$ is constructed by adding to the value of the function a zero mean white Gaussian noise. Secondly, a data set from the Machine Learning Repository UCI is treated. The result concerning the UCI dataset is illustrated in Figure 3.

In any case, a training set and test set are randomly selected. The regularization parameter γ in (3) determining the trade-off between the minimization of the training error and the smoothness, as well as the bandwidth parameter in (59), are determined by means of 10-fold cross-validation. Coupled Simulated Annealing first determines starting points for a simplex algorithm which is used subsequently. After this tuning step, the α_i's and b parameters are computed by solving a linear system on the training sample. The results are displayed in Table 1, reporting the results on test data.

Data set	Marr kernel		RBF kernel	
	MAE	± std	MAE	± std
Bike sharing	745.9	± 11.7	746.69	± 13
Sinc	0.0878	± 0.0024	0.0897	± 0.0034

Table 1: The numerical experiments were performed 50 times with calculation of the mean value of the error and the standard deviation.

[6] A matlab toolbox by D. Hammond is available at http://wiki.epfl.ch/sgwt.



(a) RBF kernel regression.

(b) Marr wavelet kernel regression.

Figure 3: A regression problem was solved with the "Bike sharing" data set from UCI machine learning repository, representing the number of bikes shared each day. The problem was considered as one-dimensional. There are 70 randomly selected training points (blue stars) and 661 test points (red dots).

The differences in performance are not very significant however it seems from Table 1 that the standard error is smaller when the Marr kernel is used.

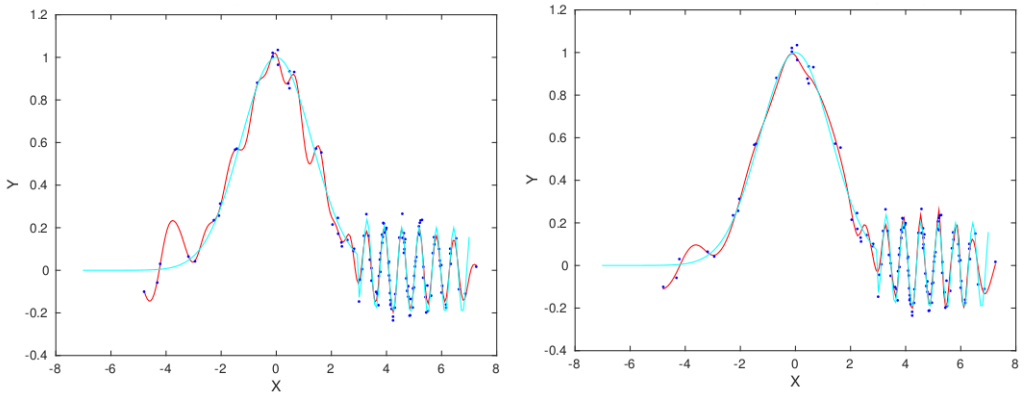
5.2 Marr kernel with adapted bandwidth

It is also possible to use the kernel (61) where a scale is associated to each data point using the nearest neighbour formula (76). We choose to modify this prescription by defining the local scale

$$a_i = \frac{\sigma}{M} \sum_{j=1}^{M} \|x_i - x_j\|. \tag{76}$$

up to an overall scale factor $\sigma > 0$, which is considered as a hyperparameter. These regularization and bandwidth parameters are tuned with the same procedure as described in Section 5.1. As an illustration, two examples are considered.

Firstly, as a proof of concept, a regression problem for a synthetic data sample combining a region of rapid oscillation and a dense sampling with a smooth region with a sparse sampling is pictured in Figure 4. This simple one-dimensional example allows to compare the difference in behaviour between the method with an adapted bandwidth and the more standard RBF kernel. The corresponding test errors are given in Table 2.



(a) RBF kernel.

(b) Adapted Marr kernel.

Figure 4: Comparison of the adapted Marr kernel with the RBF kernel. The blue dots are the 120 training points, while the blue and red curves are the original function and its estimation, respectively. The 1200 test points are not drawn on the figure.

Adapted Marr	RBF
MAE $\pm$ std	MAE $\pm$ std
0.056 ± 0.012	0.061 ± 0.015

Table 2: We compare here the Adapted Wavelet kernel regression to the RBF kernel regression. The local average distance is computed over the 10 nearest neighbours. The computation was performed 50 times.

Secondly, the same method is applied to the regression problem of the Boston Housing dataset from the UCI machine learning repository. This higher dimensional regression problem consists in estimating the price of houses considering data points with 11 features. The results are displayed in Table 3, where two different training sets were considered. The test errors are compared for different numbers of nearest neighbours entering the calculation of the local scale in (76).

6 Discussion of related works

As a matter of fact, the method described in this work differs from the previous applications of wavelets in Support Vector Machines, or in the mathematical framework for the estimation of functions in a Reproducing Kernel Hilbert Space. We shall emphasize here the main differences between these

Training	100 points	80 points
Test error	MAE ± std	MAE ± std
RBF	2.83 ± 0.01	2.85 ± 0.34
Adapted Marr M=16	2.8 ± 0.17	2.73 ± 0.14
Adapted Marr M=18	2.81 ± 0.18	2.78 ± 0.34
Adapted Marr M=20	2.83 ± 0.13	2.72 ± 0.09
Adapted Marr M=22	2.78 ± 0.11	2.69 ± 0.08

Table 3: Comparison of the test errors for the regression problem on the Boston Housing dataset from the UCI repository. Various values of integers M in the formula (76) were compared. The 506 data points were randomly divided in a training and test set. The numerical calculations were repeated 50 times.

approaches.

As it is well known, wavelets were already introduced in SVM in the literature, as for instance in [30], where wavelets are used in order to define the kernel function directly in the dual formulation. The work [30] does not really take advantage of the RKHS structure of the wavelet transform space. More precisely, the authors of reference [30] choose an admissible mother wavelet $h : \mathbb{R} \to \mathbb{R}$, and then two types of kernels are defined. The dot-product wavelet kernel

$$K(x, x') = \prod_{i=1}^{N} h(\frac{x_{(i)} - b_{(i)}}{a}) h(\frac{x'_{(i)} - b_{(i)}}{a}), \tag{77}$$

$$\tag{78}$$

is introduced, with $b \in \mathbb{R}^d$ and $a > 0$, while the conditions under which the translation invariant kernel

$$K(x - x') = \prod_{i=1}^{N} h(\frac{x_{(i)} - x'_{(i)}}{a}), \tag{79}$$

is positive definite, are discussed.

Noticeably, as explained in the references [21, 14, 23], a typical Hilbert space setting consists of the estimation of functions in a Reproducing Kernel Hilbert Space $\{\mathcal{H}_K, \langle \cdot | \cdot \rangle_K\}$, which shares similarities with the setting discussed here. Actually, in this picture, the optimization problem

$$\min_{f \in \mathcal{H}_K} \frac{1}{N} \sum_{i=1}^{N} (f(x_i) - y_i)^2 + \lambda \|f\|_K^2, \tag{80}$$

has to be solved, where the cost function was chosen to be L^2, and where $\lambda > 0$ is a regularization parameter controlling the bias-variance trade-off. We shall briefly compare the approach chosen in the present work. In the reference [34], considering the framework given by (80), a method is presented to construct a RKHS from a Hilbert space $\mathcal{B}$ with the help of a linear map similar to the wavelet transform. Choosing a set of functions $\Gamma_x(\cdot) \in \mathcal{B}$, with $x \in X \subset \mathbb{R}^d$, allows to define the map:

$$\begin{aligned} T : \mathcal{B} &\to R^X \\ f &\mapsto Tf(\cdot), \end{aligned} \tag{81}$$

with $Tf(x) = \langle \Gamma_x(\cdot), f(\cdot) \rangle_{\mathcal{B}}$, for $x \in X$, and where $R^X = \{g : X \to \mathbb{R}^d\}$. The set of functions of the form $Tf(\cdot) \in \mathcal{B}_K = \mathrm{Im}(T)$, endowed with the inner product

$$\langle Tf_1, Tf_2 \rangle_{\mathcal{B}_K} \triangleq \langle f_1, f_2 \rangle_{\mathcal{B}} \tag{82}$$

is shown to be a reproducing kernel Hilbert space, with kernel $K(s,t) = \langle \Gamma_s, \Gamma_t \rangle_{\mathcal{B}}$. The authors of [34] then suggest to use wavelets in the following way. Choosing a discrete orthonormal basis $\{e_i\}$ of $L^2(\mathbb{R})$, and a discrete wavelet basis $\{\phi_i\}$ of $L^2(\mathbb{R})$, the expansion

$$\Gamma_x(\cdot) = \sum_{i,j} \alpha_{i,j} \phi_i(x) e_j(\cdot) \tag{83}$$

is introduced, with the parameters[7] $\alpha_{i,j} \in \mathbb{R}$ so that $\Gamma_x(\cdot) \in L^2(\mathbb{R})$. As a consequence, the reproducing kernel can be written

$$K(s,t) = \sum_{i,j,n} \alpha_{i,j} \alpha_{i,n} \phi_j(s) \phi_n(t), \tag{84}$$

for any s and $t \in \mathbb{R}$. From the last result, we stress that the formulation presented in this paper, though also involving a Reproducing Kernel Hilbert Space, is significantly different for the following reasons. First of all, the model considered here can be expressed both in its primal and dual formulation. Starting from the primal formulation in a generic separable Hilbert space, the coherent states $\{|\eta_x\rangle \in \mathcal{H}\}_{x \in \mathbb{R}^d}$ are introduced through the continuous map $x \mapsto |\eta_x\rangle$. These states can be defined either by considering a

[7] Actually, the definition (83) as put forward by reference [34] is analogous to the construction of the coherent state (21), with the significant difference that the n-th coefficient $\psi_n(x)$ in (21) are explicitly chosen to be the value at $x \in \mathbb{R}^d$ of the basis function ψ_n in a specified RKHS.

distinct RKHS, which can be defined for instance from a Mercer kernel, or by considering group generated coherent states. The multidimensional wavelets are possible example of the latter case. As a matter of fact, our paper treats wavelets as particular coherent states, while the RHKS structure is the one naturally carried by wavelets. On the other hand, the reference [34] uses a discrete basis of wavelets to expand a L^2 function, thanks to the formula (83), so that the reproducing kernel (84) is not the reproducing kernel associated to the wavelets. In the same spirit of (84), let us mention that it is also possible to use a frame and its dual [33], instead of a basis, in order to expand a L^2 function in a way analogous to (83).

7 Conclusions

From the theoretical point of view, the optimization framework of Support Vector Machines was considered from a different perspective, allowing to use tools from quantum physics and signal processing. Two Hilbert spaces, a L^2-space in the primal formulation and a RKHS in the dual formulation, were shown to be related by the HHK transform. The role of this transform in the primal-dual framework of Support Vector Machines has been emphasized in connection with the coherent states. The feature map obtained thanks to the Mercer theorem is then also interpreted as a coherent state. Hence, various other examples of coherent states, wavelets and their associated transforms can be considered in this picture. We have provided a unifying framework for the Support Vector Machines in the case of infinite dimensional feature space in terms of coherent states and HHK transform.

Beyond its theoretical interest, the introduction of d-dimensional wavelets in Support Vector Machines also brings additional insights concerning the determination of the bandwidth parameter of the kernel function. These ideas were illustrated on a series of artificial, as well as real data samples. Further consequences deserve to be investigated, in particular in the case of Spectral Graph Wavelets.

Acknowledgements

EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC AdG A-DATADRIVE-B (290923). This paper reflects only the authors' views, the Union is not liable for any use that may be made of the contained information. Research Council KUL:

GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC), BIL12/11T; PhD/Postdoc grants Flemish Government: FWO: projects: G.0377.12 (Structured systems), G.088114N (Tensor based data similarity); PhD/Postdoc grants IWT: projects: SBO POM (100031); PhD/Postdoc grants iMinds Medical Information Technologies SBO 2014 Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017).

Appendix

In quantum physics, the coherent states are introduced in a different way, that we shall outline shortly with an illustrative purpose. Since the $\mathcal{H}$ is here a L^2-space of complex-valued functions, this example may not be of direct relevance for an application such as SVM regression, however a similar methodology could still be applied in a quantum physics context [39].

The Reproducing Kernel Hilbert Space has an inner product of the L^2 type, i.e., if f and g are complex valued functions in $\mathcal{H}_K$, their inner product is

$$\langle f|g\rangle_K = \int_X \mathrm{d}\mu(x)\bar{f}(x)g(x), \tag{85}$$

where X is a measurable topological space and $\bar{f}$ denotes the complex conjugate of f. Each function f has also to satisfy the integral equation,

$$\langle \xi_x|f\rangle_K = \int_X \mathrm{d}\mu(x')\langle \xi_x|\xi_{x'}\rangle_K\langle \xi_{x'}|f\rangle_K, \tag{86}$$

where $|\xi_x\rangle \in \mathcal{H}_K$ represents the evaluation map by $\langle \xi_x|f\rangle_K = f(x)$. Equivalently, the last integral equation can be rephrased as

$$f(x) = \int_X \mathrm{d}\mu(x')K(x,x')f(x'). \tag{87}$$

In this case, the coherent states (21) admit a resolution of the identity

$$\int_X \mathrm{d}\mu(x)\ |\eta_x\rangle\langle\eta_x| = \mathbb{I}_{\mathcal{H}}. \tag{88}$$

Among a series of physically relevant coherent states, a simple example is given by the Fock-Bargman-Segal space [1, 28], defined as the set of functions $f(z) = h(z)\exp(-|z|^2/2)$ in $L^2(\mathbb{C}, \mathrm{d}^2z/\pi)$, where $h(z)$ is entire. The coherent states are given by (21), and named (non normalized) canonical coherent states

$$|\eta_z\rangle \triangleq \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}}|\phi_n\rangle. \tag{89}$$

The kernel is $K(z,\bar{z}) = \exp(z\bar{z})$, while the reproducing property

$$\int_{\mathbb{C}} \frac{\mathrm{d}z}{\pi} K(z,\bar{z}) f(z) = f(\bar{z}) \tag{90}$$

is satisfied. In a quantum physics context, the basis $\{|\phi_n\rangle\}$ is constructed as a Fock basis, by considering the representation of the Fock algebra $[a, a^\dagger] \triangleq aa^\dagger - a^\dagger a = 1$. Then, the basis functions are $|\phi_n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|\Omega\rangle$, where the "vacuum" state verifies $a|\Omega\rangle = 0$. In this framework, we have a direct relation to the coherent states of the harmonic oscillator, and the coherent state (89) can be written

$$|\eta_z\rangle = \exp(za^\dagger)|0\rangle. \tag{91}$$

References

[1] S. T. Ali, J-P. Antoine, and J-P. Gazeau. *Coherent States, Wavelets and their generalizations*. Springer, 2014.

[2] S. T. Ali and F. Bagarello. Some physical appearances of vector coherent states and CS related to degenerate Hamiltonians. *J.Math.Phys.*, 46:053518–1–053518–28, 2005.

[3] S.T. Ali, J-P. Antoine, J-P. Gazeau, and U. A. Mueller. Coherent states and their generalizations: A mathematical overview. *Reviews in Mathematical Physics*, 7:1013–1104, 1995.

[4] J-P. Antoine. Square integrable representations, coherent states, wavelets. In J. Govaerts, M.N. Hounkonnou, and A.Z. Msezane, editors, *Proc. Sixth Int. Workshop*, Contemporary Problems in Mathematical Physics, pages 3–47, Cotonou, Bénin, Nov. 2009.

[5] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[6] V. Barbu and T. Precupanu. *Convexity and optimization in Banach spaces*. Springer, 1986.

[7] M. Belkin, I. Matveeva, and P. Niyogi. Regularization and semi-supervised learning on large graphs. In *COLT*, pages 624–638. Springer, 2004.

[8] S. K. Berberian. *Introduction to Hilbert Spaces*. Oxford University Press, 1961.

[9] S. Canu, X. Mary, and A. Rakotomamonjy. Functional learning through kernel. In *Advances in Learning Theory: Methods, Models and Applications NATO Science Series III: Computer and Systems Sciences*, pages 89–110. IOS Press, 2002.

[10] C. K. Chui. *An Introduction to Wavelets*. Academic Press, 1992.

[11] R Coifman and M. Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, 2006.

[12] R. R. Coifman, Y. Meyer, and V. Wickerhauser. Wavelet analysis and signal processing. In *Wavelets and their Applications*, pages 153–178, 1992.

[13] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, 20:273–297, 1995.

[14] F. Cucker and D. X. Zhou. *Learning theory: An approximation Theory Viewpoint*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2007.

[15] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.*, 41:906–996, 1988.

[16] I. Daubechies. *Ten lectures on wavelets*. CBMS-NSF regional conference series in applied mathematics. Society for industrial and applied mathematics, 1992.

[17] K. De Brabanter, J. De Brabanter, J. A. K. Suykens, and B. De Moor. Optimized fixed-size kernel models for large data sets. *Computational Statistics & Data Analysis*, 54(6):1484–1504, 2010.

[18] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[19] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the Amercian Statistical Association*, 90:1200–1224, 1995.

[20] D. L. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26:879–921, 1998.

[21] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural network architectures. *Neural Computation*, 7:219–269, 1995.

[22] D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

[23] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*. Springer, 2009.

[24] D. Horn and A. Gottlieb. Algorithm for data clustering in pattern recognition problems based on quantum mechanics. *Phys. Rev. Lett.*, 88:018702–1–018702–4, 2002.

[25] A. Horzela and F. H. Szafraniec. A measure-free approach to coherent states. *J.Phys.A*, 45:244018–1 – 244018–9, 2012.

[26] J. R. Klauder. Coherent state path integrals without resolutions of unity. *Foundations of Physics*, 31(1):57–67, 2001.

[27] J. R. Klauder and E. W. Aslaksen. Unitary representations of the affine group. *J.Math.Phys.*, 9:206–211, 1968.

[28] J. R. Klauder and B. S. Skagerstam. *Coherent States: Applications in Physics and Mathematical Physics*. World Scientific, 1985.

[29] R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *ICML*, pages 315–322, 2002.

[30] Z. Li, Z. Weida, and J. Licheng. Wavelet support vector machine. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 34(1):34–39, 2004.

[31] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society*, A 209(441-458):415–458, 1909.

[32] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidi. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, 2006.

[33] A. Rakotomamonjy and S. Canu. Frames, reproducing kernels, regularization and learning. *Journal of Machine Learning Research*, 6:1485–1515, 2005.

[34] A. Rakotomamonjy, X. Mary, and S. Canu. Non parametric regression with wavelet kernels. *Applied Stochastics Model for Business and Industry*, 21:153–163, 2005.

[35] L. Schwartz. Sous espaces Hilbertiens d'espaces vectoriels topologiques et noyaux associés. *Journal d'analyse mathématique*, pages 115–256, 1964.

[36] M. Signoretto, L. De Lathauwer, and J. A. K. Suykens. A kernel-based framework to tensorial data analysis. *Neural Networks*, 24(8):861–874, 2011.

[37] A. J. Smola and R. Kondor. Kernels and regularization on graphs. In *COLT*, pages 144–158. Springer, 2003.

[38] I. Steinwart, D. Hush, and C. Scovel. An Explicit Description of the Reproducing Kernel Hilbert Spaces of Gaussian RBF Kernels. *IEEE Transactions on Information Theory*, 52(10):4635–4643, 2006.

[39] J. A. K. Suykens. Generating quantum-measurement probabilities from an optimality principle. *Phys. Rev.*, A87:052134–1 – 052134–8, 2013.

[40] J. A. K. Suykens, C. Alzate, and K. Pelckmans. Primal and dual model representations in kernel-based learning. *Statistical Surveys*, 4:148–183, 2010.

[41] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, 2002.

[42] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.

[43] N. Tremblay and P. Borgnat. Graph Wavelets for Multiscale Community Mining. *IEEE Transactions on Signal Processing*, 62(20):5227–5239, 2014.

[44] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[45] Si Wu and Shun-Ichi Amari. Conformal transformation of kernel functions: A data-dependent way to improve support vector machine classifiers. *Neural Processing Letters*, 15(1):59–67, 2002.