Data-driven approximation of differential inclusions and application to detection of transportation modes

Pierre-Cyril Aubin-Frankowski and Nicolas Petit

***Abstract*— This article applies the Support Vector Data Description (SVDD) algorithm to approximate the graph of differential inclusions. It is proven that Gaussian SVDD can recover any compact graph if a large enough dataset is available. This data-driven approach can be used to identify discrete-valued parameters of nonlinear dynamical systems with unknown input signal. For illustration, the presented method is applied here both on real and synthetic data for detection of transportation modes based on linear velocity measurements.**

I. Introduction

Numerous recent studies have envisioned extending Machine Learning (ML) techniques to the automatic control domain ([1], [2]). Profitable pairing has already been achieved for anomaly detection [3] and system identification [4], among others. This article follows this trend, and, for its part, relates one aspect of the vast question of nonlinear dynamical system identification to a ML task in the framework of set-valued analysis.

The studied problem is as follows. Consider $N_0 > 1$ forced dynamical systems, each denoted by (f_j, U_j), for some index $j = 1, ..., N_0$ that can be interpreted as a discrete-valued parameter (or as a *label* in the ML terminology). The systems have as governing equations

$$q'(t) = f_j(q(t), u(t))$$

where the state vector is $q \in \mathbb{R}^n$, with the particularity that the input signal $u(\cdot) \in U_j$ is unknown. Each set U_j is defined as the set of functions containing every input signal that is possible given the value of j. For example, without further restriction, the sets U_j can be subsets of some functional space (e.g. C_0, L_2) with bounded values in $\mathbb{R}^m$. As stated above, the forcing signal $u(\cdot)$ is unknown. Further, despite being unambiguously defined, the sets U_j are unknown as well. To account for this lack of information, the governing equations above are rewritten as differential inclusions:

$$q'(t) \in F_j(q(t)) \triangleq \{f_j(q(t), u(t)) \,|\, u(\cdot) \in U_j\} \quad (1)$$
$$K_j \triangleq \{(q, q') \,|\, q' \in F_j(q)\} \subset \mathbb{R}^n \times \mathbb{R}^n$$

where the set-valued map F_j is identified with its graph[1] K_j.

In this article, it is assumed that for each value of j, some properly *labeled* values (samples) of $(q(\cdot), q'(\cdot))$ are available, forming N_0 datasets. We first build approximations of the graphs K_j from these data. Once these approximations are available, they can be used for parameter estimation when presented with unlabeled recordings of samples of $(q(\cdot), q'(\cdot))$. This central question here boils down to determining the possible values of j.

Classically (see e.g. [5]), most techniques in parameter estimation assume a smooth and known dependence of the measurements on the unknown i, and consider the data as (ordered) time series. In many cases, estimation is treated as an extended-state reconstruction problem, using for instance Kalman filters or state observers with unknown inputs. Numerous references develop techniques of this type (see [6] and references therein), however we do not make these assumptions. Consequently, the proposed approach is not to be mistaken up with extensions of the Kalman perspective to set-valued mappings ([7], [8]).

The framework advocated here considers the data as labeled clouds of points (q, q') [2], each cloud corresponding to an unknown subset of the unknown U_j. Graphically, constructing an approximation of K_j amounts to delineating a subset of $\mathbb{R}^{2n}$ based on the data (*learning step*). Once the learning step is achieved, when (new) unlabeled data become available, identifying j amounts to testing the membership of the new data to the N_0 learned subsets (*testing step*).

To define approximations of the sets $(K_j)_j$, we apply the Support Vector Data Description (SVDD [14]) algorithm. SVDD is a kernel method that computes a minimal enclosing ball around the data. The output of the algorithm is an indicator function, the evaluation of which allows to readily test membership. In this article, it is shown that the output function for the Gaussian kernel can estimate any compact set of $\mathbb{R}^{2n}$ (we refer to this property by "set-consistency"). Two indexes i_0 and i_1 can then be distinguished if the approximations of their K_j differ (which depends on both f_j and U_j). For applications, both learning and testing have to be compu-

Pierre-Cyril Aubin-Frankowski is PhD student from École des Ponts ParisTech, at CAS - Centre Automatique et Systèmes, MINES ParisTech, PSL Research University.

Nicolas Petit is Professor, CAS - Centre Automatique et Systèmes, MINES ParisTech, PSL Research University.

{pierre-cyril.aubin, nicolas.petit}@mines-paristech.fr

[1] The graph is bounded whenever the state and its derivatives are bounded for $u(\cdot) \in U_j$.

[2] If q' is not directly available from measurements, it can be estimated through filtering [9], [10], Kriging [11], numerical differentiation [12] or high gains observers [13] among other possibilities, possibly with a non-negligible level of noise.

tationally tractable and robust to noise. As highlighted in the article, SVDD satisfies both conditions. To show the applicability and relevance of the method, we consider the problem of detecting transportation modes. This topic has recently attracted much attention ([15],[16]), especially using smartphone inertial data, with reference datasets such as Geolife [17].

The paper is organized as follows. In Section II, the problem under consideration is stated, the SVDD algorithm is presented, and illustrated on GPS-measured linear velocity data for transportation mode detection. The main theoretical contribution (Proposition 1 on the set-consistency of Gaussian SVDD) is to be found in Section III. Finally, a simulation is performed in Section IV to discuss the numerical behavior of the algorithm.

II. Approximating discrete sets with SVDD

A. Notations

Let F be a closed set-valued map defining a differential inclusion (1). Its graph K can be represented by an indicator function ϕ, taking non-positive values only in K. We denote K_N by an approximation of K based on a labeled dataset $X_N \triangleq \{(q_i, q_i')\}_{i\leq N}$ composed of N couples $x_i = (q_i, q_i') \in \mathbb{R}^d$ with $d = 2n$.

B. Theoretical framework of the SVDD algorithm

The graph estimation problem for K can be tackled through the SVDD algorithm, which is a nonlinear version of the minimal enclosing ball problem (MEB [18]). The latter defines the ball of smallest volume of $\mathbb{R}^d$ containing the set X_N. The center of the MEB and its radius are the solutions of the following convex optimization problem:

$$\min_{c\in\mathbb{R}^d, R\in\mathbb{R}} R^2 \ \text{ s.t. } \forall i \leq N,\ \|x_i - c\|_{\mathbb{R}^d} \leq R \qquad (2)$$

The SVDD algorithm offers the flexibility and nonlinearity required to fit general sets (as shown later in Proposition 1). Rather than looking for a ball in $\mathbb{R}^d$ endowed with its usual topology, $\mathbb{R}^d$ is here embedded in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k(\mathbb{R}^d)$ where one seeks a minimal enclosing ball. For clarity, we first sketch RKHS theory before describing further the SVDD problem and its solution.

Definition 1. An RKHS $(\mathcal{H}_k, (\cdot,\cdot)_{\mathcal{H}_k})$ defined on a set X is a Hilbert space of real-valued functions on X such that there exists a reproducing kernel $k : X \times X \to \mathbb{R}$, i.e. a function satisfying:

- $\forall\, x \in X,\ k_x(\cdot) \in \mathcal{H}_k(X)$ where $k_x : \begin{cases} X \to \mathbb{R} \\ y \mapsto k(x,y) \end{cases}$
- $\forall\, x \in X,\ \forall\, f \in \mathcal{H}_k(X),\ f(x) = (f, k_x)_{\mathcal{H}_k}$

The following summarized fundamental characterization allows for all the computations performed below.

Theorem 1 (Aronszajn [19])**.** *If a Hilbert space of functions on X is an RKHS $\mathcal{H}_k(X)$, then $k(\cdot,\cdot)$ is a positive definite kernel, i.e. a kernel being both:*

-positive: $\forall m,\ \forall (a_i, x_i) \in (\mathbb{R}\times X)^m, \sum a_i a_j k(x_i, x_j) \geq 0$

-and symmetrical: $\forall\, x, y \in X,\ k(x,y) = k(y,x)$

Conversely a positive definite kernel k on X is reproducing for a unique $\mathcal{H}_k(X)$. Or, equivalently, there exists a Hilbert space $\mathcal{H}$ of real-valued functions and an embedding $\Phi_k : X \to \mathcal{H}$ s.t. $\forall\, x, x' \in X,\ k(x,x') = (\Phi_k(x), \Phi_k(x'))_{\mathcal{H}}$.

Choosing a positive definite kernel k on X, one can represent any point $x \in X$ as a function $k_x(\cdot) \in \mathcal{H}_k(X)$. Some classical examples of positive definite kernels on $\mathbb{R}^d$ include the Gaussian, Laplacian and linear kernels:

$$k_\sigma(x,y) = \exp\left(-\|x-y\|^2_{\mathbb{R}^d}/(2\sigma^2)\right) \text{ for } \sigma > 0, \qquad (3)$$

$$k_\lambda(x,y) = \exp(-\lambda\|x-y\|_{\mathbb{R}^d}) \text{ for } \lambda \geq 0 \qquad (4)$$

$$k_{lin}(x,y) = (x,y)_{\mathbb{R}^d} \qquad (5)$$

C. The SVDD algorithm

The SVDD algorithm recasts the MEB problem (2) in an RKHS [3] and considers

$$\min_{f\in\mathcal{H}_k, R\in\mathbb{R}} R^2 \ \text{ s.t. } \forall\, i \leq N,\ \|k(x_i,\cdot) - f(\cdot)\|_{\mathcal{H}_k} \leq R \qquad (6)$$

As shown in [14], the solution (f_k, R_k) of (6) is of the form $f_k(\cdot) \triangleq \sum_{i=1}^N \alpha_i k(x_i,\cdot)$, where $(\alpha_i)_i$ solves the dual problem of (6), which can be solved through quadratic programming:

$$\min_{\alpha\in\mathbb{R}^N_+} \alpha^T G\alpha - \alpha^T \mathrm{diag}(G) \ \text{ s.t. } \sum_{i=1}^N \alpha_i = 1 \qquad (7)$$

where the matrix G is the Gram matrix of the $(x_i)_{i\leq N}$ (i.e. $G \triangleq (k(x_i,x_j))_{i,j\leq N}$) and $\mathrm{diag}(G)$ is the diagonal matrix extracted from G. Based on f_k, one can readily define a membership function ϕ_k on $\mathbb{R}^d$:

$$\begin{aligned}\phi_k(x) &\triangleq \|k_x(\cdot) - f_k(\cdot)\|^2_{\mathcal{H}_k} - R_k^2 \qquad (8)\\ &= k(x,x) - R_k^2 + \sum_{i,j\leq N} \alpha_i\alpha_j k(x_i,x_j) - 2\sum_{i\leq N} \alpha_i k(x_i,x)\end{aligned}$$

In turn, the function ϕ_k defines the sought-after closed set K_N:

$$K_N \triangleq \{x \in \mathbb{R}^d \mid \phi_k(x) \leq 0\}$$

Interestingly, the complementarity slackness of the Karush–Kuhn–Tucker (KKT) conditions of (7) ensure that for x_i interior to K_N, we have $\alpha_i = 0$. Thus, except for the so-called *support vectors* x_i on the boundary of K_N, most coefficients are null in (8), which leads to a sparse representation of the set K_N by means of an indicator function that is quick to evaluate. By construction, $X_N \subset K_N$.

[3] The MEB in $\mathbb{R}^d$ corresponds to SVDD with the linear kernel k_{lin}.

D. Extension of SVDD to noisy data

Because outliers can be present in the data [4], not all the points should be included in the estimate K_N. Following [14], problem (6) can be adapted by introducing some slack variables $\xi_i \geq 0$:

$$\min_{\substack{f\in\mathcal{H}_k,\, R\in\mathbb{R},\, \xi\in\mathbb{R}_+^N \\ \forall\, i\leq N, \|k(x_i,\cdot)-f(\cdot)\|^2_{\mathcal{H}_k}\leq R^2+M+\xi_i}} \left(R^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i\right) \quad (9)$$

where the parameter $\nu \in]0,1]$ and the margin $M \in \mathbb{R}$ allow to adjust the level of conservatism. The dual problem of (9) writes in a form similar to (7):

$$\min_{\substack{\alpha\in\mathbb{R}^N \\ \forall i\leq N,\ \frac{1}{\nu N}\geq\alpha_i\geq 0}} \alpha^T G\alpha - \alpha^T\cdot\mathrm{diag}(G) \ \text{ s.t. } \sum_{i=1}^{N}\alpha_i = 1 \quad (10)$$

The value of the parameter ν can be related to the quantiles and minimum volume sets of the distribution of points. For $1 \geq \nu N$, every point is considered as a true point. Moreover the solution α is piecewise linear in ν as shown in [20], and ν is an upper bound on the proportion of outliers [21]. Some sparsity is necessarily lost as the set of support vectors is enlarged compared to the noiseless case.

E. Discussion on application of SVDD and illustration on real data

Among the available set approximation procedures, SVDD has many pros making it a suitable choice for the problem of identification of control systems, among which the detection of transportation modes, further exposed in Section IV. The sparsity of the solution, mentioned in § II-C, holds for all support vector machines and allows for the quick computation of the indicator functions (8), irrespective of the offline training cost. SVDD is non-parametric (with the obvious exception of the choice of kernels) as the model class is a Hilbert space of functions, so it can be applied to compare systems with the same variables but vastly different governing equations.

However, SVDD has some classical pitfalls. Some of them have been circumscribed as follows: limiting the effect of noise on the training step by taking out sheer outliers through an $L0$-penalty [22]; accelerating the computation for online training [23]; performing interpretation of the output f_k by converting the membership scores to probabilities [24]. Interestingly, none of these problems occurs in our context. Although, the considered time-series are ridden with noise, sheer anomalies can be smoothed out through continuity assumptions bearing on the underlying dynamics. In contrast, the low-power additive noise, due to sensors or to numerical differentiation, has little impact on the SVDD output. Below, we used a Savistky-Golay filter to smooth the data $q(\cdot)$ and compute its derivative.

[4] especially if the filtering employed to generate the q' samples are dealing with missing or incorrectly time-stamped data

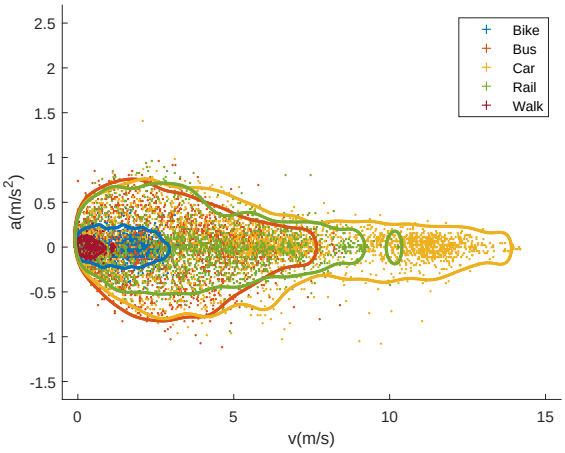

Fig. 1: Graph approximations obtained on a reference dataset [16]. The envelopes of various modes of transportation are computed through SVDD in the (speed-acceleration) phase space (v, a) with parameter $\nu = 5\%$ and various Gaussian kernels of width $\sigma \in [0.1, 0.5]$

Moreover, the typical applications of travel mode detection are in low dimension, and allow for offline computation of the boundaries, taking advantage of the online efficiency. As an illustration, anticipating on further studies conducted in Section IV, SVDD is applied to a dataset provided in [16], where the GPS-measured linear velocity was recorded at 1Hz for a variety of transportation modes. The linear acceleration was computed differentiating the velocity after smoothing the data (to avoid problems related to accelerometer measurements). On Fig.1, the five modes of transportation are represented by their SVDD envelopes in the (speed-acceleration) phase space (v, a). The envelopes overlap, so it is only when a trajectory leaves a set that it is made clear it does not belong to the corresponding class. To resolve ambiguities, one can for instance assign a trajectory to the class of smallest volume in $\mathbb{R}^d$ among the compatible ones, as it is the class that is easiest to leave. Developing such ambiguity resolution techniques is left out of the discussion and is postponed to future work.

In this example, SVDD was computed with Gaussian kernels k_σ, which variances σ^2 were adapted to each dynamic to get sharp estimates. However manipulating multiple σ entails considering functions in several RKHSs, with different norms. Furthermore, for every mode, the SVDD estimates are somewhat ill-behaved: when increasing σ, the boundary does not expand monotonically (in the inclusion sense). This calls for studies on comparison of multiple RKHSs and the increasingness property, which are the topic of the next section.

III. Theoretical guarantees on set estimation

Let $X_N \triangleq \{x_i\}_{i\leq N}$ be a finite set of distinct points of X, k be a positive definite kernel on X and $\mathcal{H}_k(X)$ be its associated RKHS. In studies on RKHSs, the kernel

is usually kept fixed. However, we seek some robustness of the estimate K_N with respect to the kernel. This is the path we explore below considering multiple kernels.

Notations: We denote by $B_{\mathbb{R}^d}(x,R)$ the closed ball in $\mathbb{R}^d$ of center x and radius R and by $B_k(f,R)$ the closed ball in $\mathcal{H}_k(X)$. For any subset K, the set $\mathrm{co}(K)$ designates its convex hull and ∂K stands for its boundary. We denote by $B_k^{SVDD} \triangleq B_k(f_k,R_k)$ the minimal enclosing ball of $\{k_{x_i}\}_i = \Phi_k(X_N)$ in $\mathcal{H}_k(X)$. Let $K_k^{SVDD} \triangleq \Phi_k^{-1}(B_k^{SVDD}) \subset X$.

We make two assumptions on the kernel k:

Assumption 1. $0 \notin \mathrm{co}(\{k_{x_i}(\cdot)\}_{i\leq N})$

Assumption 2. $\exists \kappa_k \in \mathbb{R},\ \forall\, x \in X,\ k(x,x) = \kappa_k$

Assumption 1 is verified for $X = \mathbb{R}^d$ by the Gaussian (3) and Laplacian kernels (4), due to the positivity of the functions, but not by the linear kernel (5).

Assumption 2 is verified for $X = \mathbb{R}^d$ by all the translation-invariant continuous positive definite kernels, which, owing to Bochner's theorem [25], are of the form $h(\|x-y\|_{\mathbb{R}^d})$ with h being the Fourier transform of a finite positive measure on $\mathbb{R}^d$.

A. Main results

We first obtain the following result for noiseless SVDD.

Lemma 1 (SVDD as an orthogonal projection)**.** *Under Assumptions 1 and 2, the center f_k of B_k^{SVDD} is the orthogonal projection of 0 for the norm $\|\cdot\|_k$ onto $\mathrm{co}(\{k_{x_i}\}_i)$. The solution (f_k,R_k) of the noiseless SVDD problem (6) satisfies:*

$$f_k \triangleq \sum_{i=1}^N \overline{\alpha}_i k_{x_i} = \underset{f\in\mathrm{co}(\{k_{x_i}\}_i)}{\arg\min} \|f\|_k^2\ ;\ R_k = \sqrt{\kappa_k - \|f_k\|_k^2} \tag{11}$$

Furthermore:

$$\forall\, x \in X,\ x \in K_k^{SVDD} \Leftrightarrow \|f_k\|_k^2 \leq f_k(x) \tag{12}$$

Proof. The Lagrangian $\mathcal{L}$ corresponding to (6) is

$$\mathcal{L}(f,\alpha,R) \triangleq R^2 + \sum_{i=1}^N \alpha_i \left(\|k_{x_i} - f\|_k^2 - R^2\right) \tag{13}$$

The KKT conditions require the solution $(\overline{f},\overline{R},\overline{\alpha})$ to satisfy:

$$\begin{aligned}
0 &= \frac{\partial \mathcal{L}}{\partial R}(\overline{f},\overline{R},\overline{\alpha}) = 2R\left(1 - \sum_{i=1}^N \overline{\alpha}_i\right)\\
0 &= \frac{\partial \mathcal{L}}{\partial f}(\overline{f},\overline{R},\overline{\alpha}) = 2\left(\overline{f} - \sum_{i=1}^N \overline{\alpha}_i k_{x_i}\right)\\
&0 \leq \overline{\alpha}_i \text{ and } 0 \leq \overline{R}\\
0 &= \overline{\alpha}_i\left(\|k_{x_i} - \overline{f}\|_k^2 - \overline{R}^2\right)
\end{aligned}$$

Therefore, the solution $\overline{f}$ belongs to $\mathrm{co}(\{k_{x_i}\}_i) \subset \mathcal{H}_k$. For the $p \in [\![1,N]\!]$ points $(k_{x_{i_j}})_{j\leq p} \in \partial B(\overline{f},\overline{R})$ having a non-null coefficient α, we have:

$$\forall j \leq p,\ \|k_{x_{i_j}} - \overline{f}\|_k^2 = \overline{R}^2 = \kappa_k + \|\overline{f}\|_k^2 - 2(k_{x_{i_j}},\overline{f})_k \tag{14}$$

Hence, the scalar product $(k_{x_{i_j}},\overline{f})_k$ is constant w.r.t. j.

$$\forall j \leq p,\ (k_{x_{i_j}},\overline{f})_k = (\sum_{j=1}^p \overline{\alpha}_j k_{x_{i_j}},\overline{f})_k = \|\overline{f}\|_k^2$$

So $\overline{R}^2 = \kappa_k - \|\overline{f}\|_k^2$, and the constraints of (6) become:

$$\|k_{x_i} - \overline{f}\|_k^2 \leq \kappa_k - \|\overline{f}\|_k^2 \Leftrightarrow 0 \leq (k_{x_i} - \overline{f},\overline{f})_k \tag{15}$$

Owing to Assumption 1 and as $\overline{f}$ lies in $\mathrm{co}(\{k_{x_i}\}_i)$, we can apply the Best Approximation Theorem: the constraints require $\overline{f}$ to be the orthogonal projection of 0 for $\|\cdot\|_k$ onto $\mathrm{co}(\{k_{x_i}\}_i)$. Conversely, $\overline{f}$ satisfies the constraints. Finally, (12) stems from the same calculus as (15). □

For the Gaussian kernel (3) we write for conciseness $\mathcal{H}_\sigma \triangleq \mathcal{H}_{k_\sigma}$ and $K_\sigma^{SVDD} \triangleq K_{k_\sigma}^{SVDD}$ (same for B_σ and Φ_σ). We recall the result of [26, Corollary 3.14];

$$\begin{cases} \forall \sigma_2 \geq \sigma_1 > 0,\ \mathcal{H}_{\sigma_2}(\mathbb{R}^d) \subset \mathcal{H}_{\sigma_1}(\mathbb{R}^d) \\ \forall f \in \mathcal{H}_{\sigma_2}(\mathbb{R}^d),\ \|f\|_{\sigma_1} \leq \left(\frac{\sigma_2}{\sigma_1}\right)^{d/2} \|f\|_{\sigma_2} \end{cases} \tag{16}$$

from which we derive the following result.

Lemma 2 (Norm inequalities for Gaussian kernels)**.** *Let $\sigma_2 \geq \sigma_1 > 0$. The optimal f_{σ_1}, f_{σ_2} defined in Lemma 1 satisfy the inequality:*

$$\left(\frac{\sigma_2}{\sigma_1}\right)^d \|f_{\sigma_1}\|_{\sigma_1}^2 \geq \|f_{\sigma_2}\|_{\sigma_2}^2 \geq \frac{1}{N} \tag{17}$$

Proof. Let $\sigma_2 \geq \sigma_1 > 0$, and $f_{\sigma_2} = \sum_{j=1}^N \alpha_{i,\sigma_2} k_{\sigma_2,x_i}$ be as in (11). Define analogously $(\alpha_{i,\sigma_1})_i$. Below, we use the classical result that

$$\frac{1}{N} = \min_{\{\alpha_i\geq 0\,|\,\sum_{i\leq N}\alpha_i = 1\}} \sum_{i\leq N} \alpha_i^2$$

and, as the Gaussian kernel has positive values and $\alpha_{i,\sigma_2} \geq 0$, we lower bound $\|f_{\sigma_2}\|_{\sigma_2}^2$ by the diagonal terms:

$$\|f_{\sigma_2}\|_{\sigma_2}^2 = \sum_{i,j\leq N} \alpha_{i,\sigma_2}\alpha_{j,\sigma_2} k_{\sigma_2}(x_i,x_j) \geq \sum_{i\leq N} \alpha_{i,\sigma_2}^2 \geq \frac{1}{N}$$

Set $\tilde{f}_{\sigma_1} = \sum_{j=1}^N \alpha_{i,\sigma_1} k_{\sigma_2,x_i}$. Owing to Lemma 1, f_{σ_2} is the element of smallest norm $\|\cdot\|_{\sigma_2}$ in $\mathrm{co}(\{k_{\sigma_2,x_i}\}_i)$. Hence

$$\|\tilde{f}_{\sigma_1}\|_{\sigma_2}^2 \geq \|f_{\sigma_2}\|_{\sigma_2}^2$$

Then, after some calculus, and after introducing the dual RKHS space which we omit for brevity, one gets

$$\left(\frac{\sigma_2}{\sigma_1}\right)^d \|f_{\sigma_1}\|_{\sigma_1}^2 \geq \|\tilde{f}_{\sigma_1}\|_{\sigma_2}^2$$

This concludes the proof. □

Relation (16) states that the Gaussian RKHSs form an increasing sequence of embeddings when σ decreases. This is a first step to compare the indicator functions defined by (8). Furthermore the SVDD algorithm with the Gaussian kernel is "set-consistent", in the sense that with enough data it can recover any compact set of $\mathbb{R}^d$:

Proposition 1 (Main result: Set-consistency of Gaussian SVDD)**.** *The estimate K_σ^{SVDD} of X_N by the SVDD algorithm for Gaussian kernels satisfies the following two properties*

$$i): \forall\, \epsilon > 0,\ \exists\, \sigma_0 > 0,\ s.t.\ \forall\, 0 < \sigma \leq \sigma_0,\ K_\sigma^{SVDD} \subset X_N + B_{\mathbb{R}^d}(0, \epsilon) \tag{18}$$

$$ii): \exists\, M > 0,\ \forall\, \sigma > 0,\ K_\sigma^{SVDD} \subset X_N + B_{\mathbb{R}^d}(0, M) \tag{19}$$

Proof. Let $\epsilon > 0$, $\sigma > 0$ and $x \in K_\sigma^{SVDD}$. We combine (12) and (17):

$$\frac{1}{N} \leq \|f_\sigma\|_\sigma^2 \leq f_\sigma(x) = \sum_{i \leq N} \alpha_{i,\sigma} e^{-\|x - x_i\|_{\mathbb{R}^d}^2/(2\sigma^2)} \leq e^{-\min_i \|x - x_i\|_{\mathbb{R}^d}^2/(2\sigma^2)}$$

Thus, $\min_i \|x - x_i\|_{\mathbb{R}^d}^2 \leq 2\sigma^2 \ln N$. Set $\sigma_0 \triangleq \epsilon/\sqrt{2 \ln N}$. Therefore, for any $\sigma \in]0, \sigma_0]$, we have that $K_\sigma^{SVDD} \subset X_N + B_{\mathbb{R}^d}(0, \epsilon)$.

Fix $M \triangleq \sup_{i,j \leq N} \|x_i - x_j\|_{\mathbb{R}^d}$ and $f_\sigma = \sum_{i \leq N} \alpha_i k_{x_i}$ as in (11). Let $y \in \mathbb{R}^d \backslash (X_N + B_{\mathbb{R}^d}(0, M))$, then:

$$\begin{aligned} \forall\, i, j \leq N,\ k_\sigma(y, x_i) &= e^{-\|y - x_i\|_{\mathbb{R}^d}^2/(2\sigma^2)} \\ &< e^{-M^2/(2\sigma^2)} \\ &\leq e^{-\|x_i - x_j\|_{\mathbb{R}^d}^2/(2\sigma^2)} = k_\sigma(x_i, x_j) \end{aligned}$$

$$\begin{aligned} \forall\, i \leq N,\ \sum_{i \leq N} \alpha_i k_\sigma(y, x_i) &= (\sum_{j \leq N} \alpha_j)(\sum_{i \leq N} \alpha_i k_\sigma(y, x_i)) \\ &< \sum_{i,j \leq N} \alpha_i \alpha_j k_\sigma(x_i, x_j) = \|f_\sigma\|_\sigma^2 \end{aligned}$$

We conclude from (12) that $y \notin K_\sigma^{SVDD}$. This yields $K_\sigma^{SVDD} \subset X_N + B_{\mathbb{R}^d}(0, M)$. □

Relations (18) and (19) show that the sequence $(K_\sigma^{SVDD})_{\sigma > 0}$ is bounded and, for σ small enough, lies in a neighborhood of X_N for the norm of $\mathbb{R}^d$. In the limit case, when N tends to infinity, if X_∞ is dense in a given compact $K \subset \mathbb{R}^d$, then K_σ^{SVDD} is dense as well and lies in a neighborhood of K.

B. Further results

Numerical experiments confirm that the sequence $(K_\sigma^{SVDD})_{\sigma > 0}$ of sets produced by the SVDD algorithm for Gaussian kernels is not increasing with σ w.r.t the inclusion. The increasingness property of a sequence of sets is akin to a "stability property" in the sense that the predictions drawn for a "large" kernel should contain the predictions obtained for "narrower" kernels.

As an extension, for a given kernel k_σ, we may look for a modified ball $B_\sigma(\tilde{f}_\sigma, \tilde{R}_\sigma)$, that should at least contain the set $\{k_{x_i}\}_i$. We may also wish for the modified ball to include $\Phi_k(K_\sigma^{SVDD})$, i.e. the set of all the k_x that are in the minimal enclosing ball of $\{k_{x_i}\}_i$ in $\mathcal{H}_\sigma$. One possibility is to keep the center fixed and to expand the radius accordingly. We first state a proposition for general kernels.

Proposition 2. *Let k_1 and k_2 be two positive definite kernels on X, satisfying Assumptions 1 and 2, and such that for some $\gamma > 0$ the kernel $\gamma^2 k_1 - k_2$ is positive definite (or equivalently that $\mathcal{H}_{k_2}(X) \subset \mathcal{H}_{k_1}(X)$). Then:* $\text{co}(\{k_2(x_i, \cdot)\}_i) \subset B_{k_2}(f_{k_2}, R_{k_2}) \subset B_{k_1}(f_{k_2}, \gamma R_{k_2}) \subset \mathcal{H}_{k_1}(X)$.

Proof. From [27, Th 2.17], we deduce that the existence of $\gamma > 0$ s.t. the kernel $\gamma^2 k_1 - k_2$ is positive definite is equivalent to the inclusion $\mathcal{H}_{k_2}(X) \subset \mathcal{H}_{k_1}(X)$, the identity being continuous, of norm smaller than γ. Therefore we have $B_{k_2}(f_{k_2}, R_{k_2}) \subset B_{k_1}(f_{k_2}, \gamma R_{k_2}) \subset \mathcal{H}_{k_1}(X)$ and, by definition of the minimal enclosing ball coupled with the triangular inequality, we obtain $\text{co}(\{k_2(x_i, \cdot)\}_i) \subset B_{k_2}(f_{k_2}, R_{k_2})$. This completes the proof. □

Corollary 1 (Increasingness of σ-concentric SVDD)**.** *Let $\sigma_0 > 0$. The sequence $(\Phi_\sigma^{-1}(B_\sigma(f_{\sigma_0}, (\sigma/\sigma_0)^{d/2} R_{\sigma_0})))_{\sigma \geq \sigma_0}$ is increasing w.r.t. the inclusion when σ increases.*

Proof. Inequality (17) implies that the Gaussian kernel satisfies the assumptions of Proposition 2 with $\gamma^2 = (\sigma/\sigma_0)^d$. This proof requires as well introducing dual RKHSs. □

IV. Application to detection of transportation mode on simulated data

In this section, we illustrate the practicability of SVDD for approximation of differential inclusions. We consider a general model (20) representing two types of vehicles: cars and bikes. In this model, the input exerted onto the vehicles is such that it produces asymptotic tracking of a reference velocity signal $v_{req}(\cdot)$ stemming from an urban-part of the NEDC cycle (New European Driving Cycle) generating a small set ($N \approx 500$) and a large set ($N \approx 15000$) of points. The chosen parameters are listed in Table I along with the magnitude of the speed.

$$m\dot{v}(t) = -kv^2(t) + u(t) \tag{20}$$

$$\text{where } u(t) \triangleq \begin{cases} -F_{max} \text{ if } k_p(v_{req}(t) - v(t)) < -F_{max} \\ F_{max} \text{ if } k_p(v_{req}(t) - v(t)) > F_{max} \\ k_p(v_{req}(t) - v(t)) \text{ otherwise} \end{cases}$$

TABLE I: List of parameters of the NEDC simulation

	m	k	k_p	F_{max}	$\max(v_{req})$
Car	1 T	0.27	20	2 kN	80 km/h
Bike	100 kg	0.5	20	30 N	30 km/h

We apply the SVDD algorithm to the pairs $(v, a) \triangleq (v, \dot{v})$ with a Gaussian kernel with a diagonal co-variance matrix $\sigma_v = 1.8$, $\sigma_a = 0.185$. There are between 10 and

30 support vectors in all cases, the learning step computation time ranging from 0.5 sec. to 25 sec. Fig.2a and 2b show the tightness of our estimate for noiseless data. As soon as enough points have been treated (Fig.2b), the SVDD outputs approximate well the theoretical boundaries, i.e. the graph of the differential inclusion (1) (which can be derived from (20) using a very general v_{req}). Furthermore the estimate is robust to 10 % outliers drawn in a uniform fashion (which results in a relatively strongly corrupted dataset compared to the typical noise present in velocity measurements and the subsequent filtered derivatives), by setting $\nu = 10$ % in (9) (Fig.2d) according to the guidelines recalled in §II-D. As shown in Corollary 1, the σ-concentric SVDD balls generate an increasing sequence that rapidly expands (Fig.2e). To test the membership of a trajectory to either class (Fig.2c), one can check the signs of the functions (8). When a function takes a positive value, the trajectory has violated the empirical boundary and thus does not belong to the corresponding class (Fig.2f). Using a representative test trajectory, the car can be easily identified after approximately 1-2 min of measurement.

V. Conclusion and perspectives

The example of detection of transportation modes highlights the relevance for parameter estimation of a set-valued framework involving SVDD, a kernel method. Our investigation has shown the consistency of the algorithm for Gaussian kernels, as it allows to recover any set-valued map with bounded graph. The dependence of SVDD on its parameters was also discussed. Future work will concern identification of continuous parameters, such as mass estimation as is considered in weigh-in-motion applications, a key technology for improving ground transportation safety [28]. This problem is more complex that the one treated here. We believe it may require to incorporate the sequentiality of time series into the above framework, which would speed up mode detection as well.

Acknowledgments

The authors express their gratitude to Stéphane Canu for the enlightning discussions and for the code provided to compute efficiently the SVDD algorithm.

References

[1] F. Lamnabhi-Lagarrigue, A. Annaswamy, S. Engell, A. Isaksson, P. Khargonekar, R. M. Murray, H. Nijmeijer, T. Samad, D. Tilbury, and P. Van den Hof. Systems & control for the future of humanity, research agenda: Current and future roles, impact and grand challenges. *Annual Reviews in Control*, 43:1–64, 2017.

[2] P. P. Khargonekar and M. A. Dahleh. Advancing systems and control research in the era of ML and AI. *Annual Reviews in Control*, 45:1–4, 2018.

[3] N. Laouti, S. Othman, M. Alamir, and N. Sheibat-Othman. Combination of model-based observer and support vector machines for fault detection of wind turbines. *International Journal of Automation and Computing*, 11(3):274–287, jun 2014.

[4] F. Bagge Carlson. *Machine learning and system identification for estimation in physical systems*. PhD thesis, Department of Automatic Control, 12 2018.

[5] E. Walter and L. Pronzato. *Identification of Parametric Models: from Experimental Data)*. Springer, 1997.

[6] G. Besançon. *Nonlinear Observers and Applications, volume 363, chapter An Overview on Observer Tools for Nonlinear Systems*. Springer, 2007.

[7] B. Brogliato and M. Heemels. Observer design for lure systems with multivalued mappings: A passivity approach. *IEEE Transactions on Automatic Control*, 54(8):1996–2001, 2009.

[8] A. Doris, A. L. Juloski, N. Mihajlovic, M. Heemels, N. van de Wouw, and H. Nijmeijer. Observer designs for experimental non-smooth and discontinuous systems. *IEEE Transactions on Control Systems Technology*, 16(6):1323–1332, 2008.

[9] A. Levant. Robust exact differentiation via sliding mode technique. *Automatica*, 34(3):379 – 384, 1998.

[10] J. Luo, K. Ying, P. He, and J. Bai. Properties of Savitzky-Golay digital differentiators. *Digital Signal Processing*, 15(2):122–136, 2005.

[11] E. Vazquez and E. Walter. Estimating derivatives and integrals with Kriging. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 8156–8161, 2005.

[12] S. Diop, J. W. Grizzle, and F. Chaplais. On numerical differentiation algorithms for nonlinear estimation. In *IEEE Conference on Decision and Control*, volume 2, pages 1133–1138. IEEE; 1998, 2000.

[13] A. M. Dabroom and H. K. Khalil. Discrete-time implementation of high-gain observers for numerical differentiation. *International Journal of Control*, 72(17):1523–1537, 1999.

[14] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Machine Learning*, 54(1):45–66, jan 2004.

[15] A. Bolbol, T. Cheng, I. Tsapakis, and J. Haworth. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems*, 36(6):526–537, nov 2012.

[16] B. Martin, V. Addona, J. Wolfson, G. Adomavicius, and Y. Fan. Methods for real-time prediction of the mode of travel using smartphone-based GPS and accelerometer data. *Sensors*, 17(9):2058, sep 2017.

[17] Y. Zheng, X. Xie, and W.-Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Engineering Bulletin*, 33(2):32–39, 2010.

[18] D. J . Elzinga and D. W. Hearn. The minimum covering sphere problem. *Management Science*, 19(1):96–104, sep 1972.

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

[20] K. Sjöstrand and R. Larsen. The entire regularization path for the support vector domain description. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9:241–248, 2006.

[21] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, jul 2001.

[22] M. El Azami, C. Lartizien, and S. Canu. Robust outlier detection with L0-SVDD. In *22th European Symposium on Artificial Neural Networks, ESANN 2014, Bruges, Belgium, April 23-25, 2014*, 2014.

[23] P. Laskov, Ch. Gehl, S. Krüger, and K.-R. Müller. Incremental support vector learning: Analysis, implementation and applications. *Journal of Machine Learning Research*, 7:1909–1936, 2006.

[24] M. El Azami, C. Lartizien, and S. Canu. Converting SVDD scores into probability estimates: Application to outlier detection. *Neurocomputing*, 268:64–75, dec 2017.

[25] S. Bochner. Monotone funktionen, stieltjessche integrale und harmonische analyse. *Mathematische Annalen*, 108(1):378–410, dec 1933.

[26] 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, oct 2006.

[27] S. Saitoh and Y. Sawano. *Theory of Reproducing Kernels and Applications*. Springer Singapore, 2016.

[28] B. Jacob and V. Feypell-de La Beaumelle. Improving truck safety: Potential of weigh-in-motion technology. *IATSS Research*, 34(1):9–15, jul 2010.

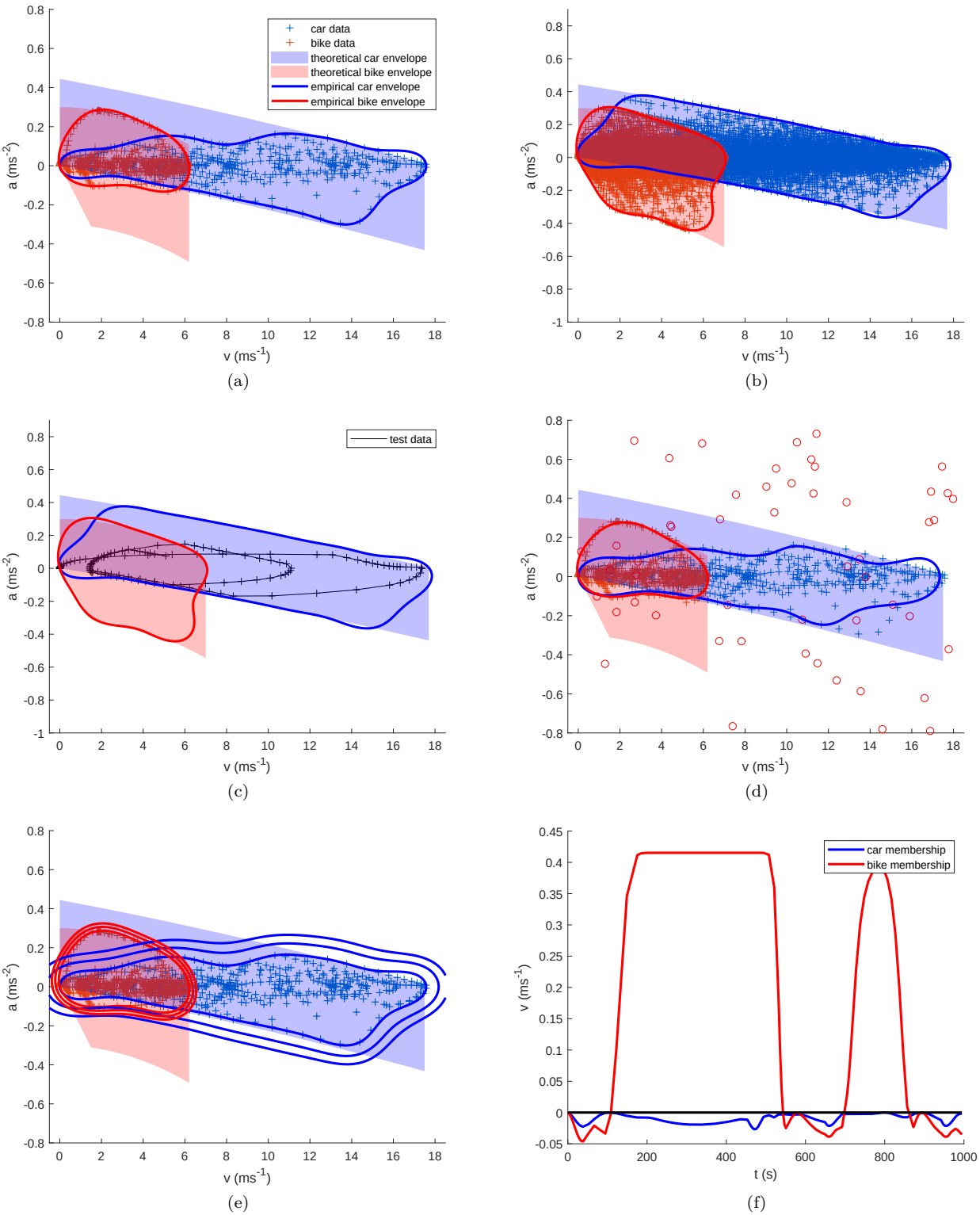

Fig. 2: Estimate sets K_σ^{SVDD} of theoretical dynamical limits (filled areas) by the SVDD algorithm on simulation data. (a): noiseless case with a small number of data (∼500 pts). (b): noiseless case with a large number of data (∼15k pts) (c): a test trajectory is compared to the SVDD boundaries. (d): case with noise on the small data and 10% uniformly distributed outliers. The boundaries are mildly altered. (e): applying σ-concentric SVDD on the noiseless small data with varying σ (100%,102%,104% of the previous σ).(f): the indicator functions show that the test trajectory of (c) is a car as the car values are negative while some bike values are positive.