Least Squares Support Vector Machines for Kernel CCA in Nonlinear State-Space Identification

Vincent Verdult*, Johan A. K. Suykens**, Jeroen Boets**, Ivan Goethals**, Bart De Moor**

* Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands
** K. U. Leuven, ESAT-SCD-SISTA
Kasteelpark Arenberg 10, B-3001 Leuven (Heverlee), Belgium
Email: V.Verdult@dcsc.tudelft.nl and Johan.Suykens@esat.kuleuven.ac.be

Abstract

We show that kernel canonical correlation analysis (KCCA) can be used to construct a state sequence of an unknown nonlinear dynamical system from delay vectors of inputs and outputs. In KCCA a feature map transforms the available data into a high dimensional feature space, where classical CCA is applied to find linear relations. The feature map is only implicitly defined through the choice of a kernel function. Using a least squares support vector machine (LS-SVM) approach an appropriate form of regularization can be incorporated within KCCA. The state sequence constructed by KCCA can be used together with input and output data to identify a nonlinear state-space model. The presented identification method can be regarded as a nonlinear extension of the intersection based subspace identification method for linear time-invariant systems.

1 Introduction

Dynamical systems are often represented using state-space representations. These representations are attractive when dealing with multivariable input and output signals. It is of practical interest to be able to determine a state-space description of a nonlinear system from a finite number of input and output measurements. However, only for linear time-invariant systems well-established state-space identification methods are available (Ljung, 1999). In the nonlinear case the recurrent nature of the state-space representation causes several difficulties in the estimation procedure (Nerrand et al., 1994). The recurrent nature of the problem vanishes if the state sequence is known or can be estimated and is used along with the input and output signals to estimate the state-space representation.

In this paper we use kernel canonical correlation analysis (KCCA) to construct a state sequence of an unknown nonlinear dynamical system from delay vectors of inputs and outputs. A nonlinear kernel function is used to map the data into a feature space in which linear canonical correlation analysis (CCA) techniques can be performed. This principle of using nonlinear kernel functions to construct nonlinear variants of linear techniques is commonly used in the field of machine learning (Schölkopf and Smola, 2002). Using a least squares

support vector machine (LS-SVM) approach instead of just the kernel trick with application of Mercer's theorem, an appropriate form of regularization can be incorporated within KCCA. Primal-dual optimization formulations of this regularized KCCA have been proposed by Suykens et al. (2002). The resulting problem to be solved in the dual space is a generalized eigenvalue problem that corresponds to the formulation of a least squares problem in the primal space involving a feature map.

The estimation of the state by means of KCCA can be regarded as a nonlinear extension of the intersection based subspace identification method for linear time-invariant systems described by Moonen et al. (1989) and Van Overschee and De Moor (1996). They used CCA to determine a state sequence as the intersection of the row spaces of two Hankel matrices containing observed inputs and outputs, shifted by a finite number of lags in time. Subspace identification is a well-accepted method for the identification of linear state-space models (Ljung, 1999). Related work on nonlinear extensions of subspace identification methods has been carried out by Larimore (1992) who uses a conditional expectation algorithm for nonlinear CCA. Another approach based on neural networks was presented by Verdult et al. (2000).

This paper is organized as follows. Section 2 briefly reviews CCA. Section 3 presents KCCA and the derivation of the corresponding LS-SVM via a primal-dual formulation of the optimization problem. Section 4 shows how the state of a nonlinear system can be obtained by performing KCCA on vectors of delayed inputs and outputs. An experimental verification of the proposed state estimation procedure is presented in Section 5 using a simulation example.

2 Canonical correlation analysis

CCA is a method from statistics for determining linear relations among several variables (Gittins, 1985). Consider two sets of vectors $x_i \in \mathbb{R}^{n_x}$ and $y_i \in \mathbb{R}^{n_y}$ for $i = 1, 2, , \ldots, N$, with $N \gg n_x + n_y$ stored row-wise in the matrices $X \in \mathbb{R}^{N \times n_x}$ and $Y \in \mathbb{R}^{N \times n_y}$, respectively. CCA determines vectors $v_j \in \mathbb{R}^{n_x}$ and $w_j \in \mathbb{R}^{n_y}$ such that the so-called variates Xv_j and Yw_j are maximally correlated. Usually, this is formulated as the following optimization problem:

$$\max_{v_j, w_j} v_j^T X^T Y w_j \quad \text{subject to} \quad v_j^T X^T X v_j = 1, \quad w_j^T Y^T Y w_j = 1, \tag{1}$$

or equivalently:

$$\min_{v_j, w_j} \|Xv_j - Yw_j\|_2 \quad \text{subject to} \quad v_j^T X^T X v_j = 1, \quad w_j^T Y^T Y w_j = 1. \tag{2}$$

The solution to this problem yields the first pair of canonical vectors v_1 and w_1. Other sets of canonical vectors can be computed by again solving (1) and requiring that the variates Xv_j and Yw_j are orthogonal to the previously computed variates Xv_i and Yw_i with $i < j$. It can be shown that this corresponds to solving the system

$$X^T Y w_j = \lambda_j X^T X v_j, \tag{3}$$
$$Y^T X v_j = \lambda_j Y^T Y w_j, \tag{4}$$

for $j = 1, 2, \ldots, r$ with $r = \min(\text{rank}(X), \text{rank}(Y))$.

CCA can be seen as a natural multidimensional extension of angles between vectors, where the angles in CCA, the so-called canonical angles, are defined as the angles between two corresponding canonical variates. For this reason, CCA is a standard tool for the analysis of linear relations between two sets of data.

3 Kernel CCA and the LS-SVM formulation

A nonlinear extension for CCA is known as kernel CCA or KCCA (Lai and Fyfe, 2000; Bach and Jordan, 2002). Available data are transformed into a high dimensional feature space, where classical CCA is applied. The kernel-aspect in KCCA lies in the fact that the feature map is only implicitly defined through the choice of a kernel, serving as a distance measure between data points. Let $\varphi_x : \mathbb{R}^{n_x} \to \mathbb{R}^{d_x}$ denote the feature map and define

$$\Phi_x := \begin{bmatrix} \varphi_x(x_1) & \varphi_x(x_2) & \cdots & \varphi_x(x_N) \end{bmatrix}^T,$$

and define Φ_y similarly using $\varphi_y : \mathbb{R}^{n_y} \to \mathbb{R}^{d_y}$. Now the CCA problem (2) becomes

$$\min_{v_j, w_j} \|\Phi_x v_j - \Phi_y w_j\|_2 \quad \text{subject to} \quad v_j^T \Phi_x^T \Phi_x v_j = 1, \quad w_j^T \Phi_y^T \Phi_y w_j = 1, \tag{5}$$

where, with a slight abuse of notation, we have $v_j \in \mathbb{R}^{d_x}$ and $w_j \in \mathbb{R}^{d_y}$. A kernel version is obtained by substituting

$$v_j := \sum_{i=1}^N a_{ij} \varphi_x(x_i) = \Phi_x^T a_j, \qquad w_j := \sum_{i=1}^N b_{ij} \varphi_y(y_i) = \Phi_y^T b_j.$$

This leads to

$$\max_{a_j, b_j} a_j^T K_{xx} K_{yy} b_j \quad \text{subject to} \quad a_j^T K_{xx} K_{xx} a_j = 1, \quad b_j^T K_{yy} K_{yy} b_j = 1, \tag{6}$$

or equivalently

$$K_{yy} b_j = \lambda_j K_{xx} a_j, \tag{7}$$
$$K_{xx} a_j = \lambda_j K_{yy} b_j, \tag{8}$$

where $K_{xx} := \Phi_x \Phi_x^T$ and $K_{yy} := \Phi_y \Phi_y^T$ are the kernel Gram matrices.

A disadvantage of KCCA versions proposed by Lai and Fyfe (2000) and Bach and Jordan (2002) is the fact that these kernel versions do not contain regularization (except by an ad hoc and heuristic modification of the algorithm). The KCCA version proposed by Suykens et al. (2002) the problem is formulated using a support vector machine approach (Vapnik, 1998) with primal-dual optimization problems. The regularization is incorporated within the primal formulation in a well established manner leading to better numerically conditioned solutions.

The CCA criterion aims at optimizing the coefficient vectors v_j and w_j, searching for a projection axis in x space that has maximal covariation with a projection axis in y space. Since the coefficients could become arbitrarily large, these are by convention constrained or at least regularized, which has the extra benefit of numerical stability. After mapping the data points into feature space, the least squares optimization formulation results in the following primal form problem:

$$\begin{aligned} \max_{v,w} J_{CCA}(v, w, e, r) &= \gamma \sum_{i=1}^N \left(e_i r_i - \frac{\nu_x}{2} e_i^2 - \frac{\nu_y}{2} r_i^2 \right) - \frac{1}{2} v^T v - \frac{1}{2} w^T w, \\ \text{subject to} \qquad & e_i = v^T \varphi_x(x_i), \quad r_i = w^T \varphi_y(y_i), \quad \text{for } i = 1, \ldots, N, \end{aligned}$$

with hyper-parameters $\gamma, \nu_x, \nu_y \in \mathbb{R}_0^+$. Note that for ease of notation we have dropped the subscript j from all variables. Introducing a_i, b_i as Lagrange multiplier parameters, the Lagrangian is written as

$$\begin{aligned} L(v,w,e,r;a,b) &= \gamma \sum_{i=1}^{N} \Big(e_i r_i - \frac{\nu_x}{2} e_i^2 - \frac{\nu_y}{2} r_i^2 \Big) - \frac{1}{2} v^T v - \frac{1}{2} w^T w \\ &\quad - \sum_{i=1}^{N} a_i \Big(e_i - v^T \varphi_x(x_i) \Big) - \sum_{i=1}^{N} b_i \Big(r_i - w^T \varphi_y(y_i) \Big). \end{aligned}$$

The optimality conditions for the corresponding dual optimization problem are given by:

$$\begin{aligned} \frac{\partial L}{\partial v} &= 0 \quad \Rightarrow \quad v = \sum_{i=1}^{N} a_i \varphi_x(x_i) = \Phi_x^T a, \\ \frac{\partial L}{\partial w} &= 0 \quad \Rightarrow \quad w = \sum_{i=1}^{N} b_i \varphi_y(y_i) = \Phi_y^T b, \\ \frac{\partial L}{\partial e_i} &= 0 \quad \Rightarrow \quad \gamma r_i = \nu_x e_i + a_i, \quad i = 1, 2, \ldots, N, \\ \frac{\partial L}{\partial r_i} &= 0 \quad \Rightarrow \quad \gamma e_i = \nu_y r_i + b_i, \quad i = 1, 2, \ldots, N, \\ \frac{\partial L}{\partial a_i} &= 0 \quad \Rightarrow \quad e_i = v^T \varphi_x(x_i), \quad i = 1, 2, \ldots, N, \\ \frac{\partial L}{\partial b_i} &= 0 \quad \Rightarrow \quad r_i = w^T \varphi_y(y_i), \quad i = 1, 2, \ldots, N. \end{aligned}$$

By elimination of the variables e, r, v, w, and defining $\lambda := 1/\gamma$, we can simplify the dual problem further, which results into the regularized system of equations

$$\begin{aligned} K_{yy} b &= \lambda(\nu_x K_{xx} + I) a, \\ K_{xx} a &= \lambda(\nu_y K_{yy} + I) b. \end{aligned}$$

This regularized KCCA variant was proposed within the context of primal-dual LS-SVM formulations by Suykens et al. (2002). Using LS-SVMs, the process of finding canonical variates and correlations in feature space can be dealt with, by solving a generalized eigenvalue problem in the dual space.

4 Nonlinear state-space identification

We now turn to the state-space identification problem. Consider the nonlinear state-space system

$$\begin{aligned} x_{k+1} &= f(x_k, u_k), \quad (9) \\ y_k &= h(x_k), \quad (10) \end{aligned}$$

with $x_k \in \mathbb{R}^n$ the state, $u_k \in \mathbb{R}^m$ the input and $y_k \in \mathbb{R}^\ell$ the output and where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^\ell$ are smooth functions. The goal is to identify from a finite number of measurements of the input u_k and the output y_k a nonlinear state-space system that has

the same dynamic behavior as the system (9)–(10). It is important to realize that a unique solution to this identification problem does not exist. This is due to the fact that the state x_k can only be determined up to a smooth (nonlinear) invertible map $T : \mathbb{R}^n \to \mathbb{R}^n$. This map can be regarded as a (nonlinear) state transformation, because the state space system

$$\begin{aligned} \xi_{k+1} = \quad & T\Big(f\big(T^{-1}(\xi_k), u_k\big)\Big) && = \overline{f}(\xi_k, u_k), && (11)\\ y_k = \quad & h\big(T^{-1}(\xi_k)\big) && = \overline{h}(\xi_k), && (12) \end{aligned}$$

with $\xi_k = T(x_k)$ has the same input-output dynamic behavior as the system (9)–(10).

We introduce the delay vectors

$$\overline{u}_k^d := \begin{bmatrix} u_k \\ u_{k+1} \\ \vdots \\ u_{k+d-1} \end{bmatrix}, \qquad \overline{y}_k^d := \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+d-1} \end{bmatrix},$$

and define the iterated map: $f^i(x_k, \overline{u}_k^i) := f_{u_{k+i-1}} \circ f_{u_{k+i-2}} \circ \cdots \circ f_{u_{k+1}} \circ f_{u_k}(x_k)$, where $f_{u_k}(x_k)$ is used to denote $f(x_k, u_k)$ evaluated for a fixed u_k. Now we can derive a dynamical system in terms of the delay vectors

$$\begin{aligned} x_{k+d} &= f^d(x_k, \overline{u}_k^d), \\ \overline{y}_k^d &= h^d(x_k, \overline{u}_k^d), \end{aligned}$$

where

$$h^d(x_k, \overline{u}_k^d) := \begin{bmatrix} h(x_k) \\ h \circ f^1(x_k, \overline{u}_k^1) \\ \vdots \\ h \circ f^{d-1}(x_k, \overline{u}_k^{d-1}) \end{bmatrix}.$$

We assume the existence of an equilibrium point for the system (9)–(10): there exist constant input u^0, state x^0 and output y^0, such that $x^0 = f(x^0, u^0)$ and $y^0 = h(x^0)$. Without loss of generality it is assumed that $x^0 = u^0 = y^0 = 0$. This assumption can be satisfied by simply shifting the input, state and output axes.

4.1 Constructing states from input and output data

The system (9)–(10) is assumed to be strongly locally observable at the equilibrium point. Following Nijmeijer (1982) we adopt the following definition:

Definition 1 *The system (9)–(10) is strongly locally observable at (x_k, u_k) if*

$$rank\left(\frac{\partial h^n(x_k, \overline{u}_k^n)}{\partial x_k}\right) = n.$$

By the observability assumption $\partial h^d(0,0)/\partial x_k$ has full column rank n. Application of the implicit function theorem to the map $h^d : \mathcal{X} \times \mathcal{U}^d \to \mathcal{Y}^d$ shows that any state in the neighborhood of the equilibrium point can be uniquely determined from a finite sequence of input and output delay vectors.

Lemma 1 *If the system (9)–(10) is strongly locally observable at the origin, then there exist open neighborhoods $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{U} \subset \mathbb{R}^m$, $\mathcal{Y} \subset \mathbb{R}^\ell$ of the origin and a smooth function $\Upsilon : \mathcal{U}^d \times \mathcal{Y}^d \to \mathcal{X}$, for all $d \geq n$, such that if $\overline{y}_k^d = h^d(x_k, \overline{u}_k^d)$ then $x_k = \Upsilon(\overline{u}_k^d, \overline{y}_k^d)$ for all $x_k \in \mathcal{X}$ and $\overline{u}_k^d \in \mathcal{U}^d$.*

The state of a system summarizes all the information from the past behavior needed to determine the future behavior. The past inputs and outputs given by

$$\overline{z}_k^d := \begin{bmatrix} \overline{u}_{k-d}^d \\ \overline{y}_{k-d}^d \end{bmatrix} \tag{13}$$

form in fact a (nonminimal) state of the system. To prove that $\overline{z}_k^d$ is indeed a state, we have to show that $\overline{z}_{k+1}^d$ is a function of $\overline{z}_k^d$ and u_k and does not depend on other values of the input.

Lemma 2 *If the system (9)–(10) is strongly locally observable at the origin, then there exist open neighborhoods $\mathcal{U} \subset \mathbb{R}^m$, $\mathcal{Y} \subset \mathbb{R}^\ell$ of the origin and a smooth function $F : \mathcal{U}^d \times \mathcal{Y}^d \times \mathcal{U} \to \mathcal{U}^d \times \mathcal{Y}^d$ for all $d \geq n$ such that the state candidate $\overline{z}_k^d$ given by (13) satisfies the dynamical equation $\overline{z}_{k+1}^d = F(\overline{z}_k^d, u_k)$ for all $\overline{z}_k^d \in \mathcal{U}^d \times \mathcal{Y}^d$ and $u_k \in \mathcal{U}$.*

Proof: Since the system (9)–(10) is assumed to be strongly locally observable, Lemma 1 shows that there exist open neighborhoods $\mathcal{X} \subset \mathbb{R}^n$, $\mathcal{U} \subset \mathbb{R}^m$, $\mathcal{Y} \subset \mathbb{R}^\ell$ of the origin and a smooth function $\Upsilon : \mathcal{U}^d \times \mathcal{Y}^d \to \mathcal{X}$ such that if $\overline{y}_k^d = h^d(x_k, \overline{u}_k^d)$ then $x_k = \Upsilon(\overline{u}_k^d, \overline{y}_k^d) = \Upsilon(\overline{z}_{k+d}^d)$. Therefore,

$$\begin{aligned} \overline{y}_{k-d+1}^d &= h^d(x_{k-d+1}, \overline{u}_{k-d+1}^d) = h^d\Big(f(x_{k-d}, u_{k-d}), \overline{u}_{k-d+1}^d\Big) \\ &= h^d\Big(f(\Upsilon(\overline{z}_k^d), u_{k-d}), \overline{u}_{k-d+1}^d\Big). \end{aligned}$$

Observe that u_{k-d} is an entry of the vector $\overline{u}_{k-d}^d$ and that $\overline{u}_{k-d+1}^d$ can be replaced by $\overline{u}_{k-d}^d$ if we add u_k to the arguments of the function h^d. Therefore, we can define a smooth function $F : \mathcal{U}^d \times \mathcal{Y}^d \times \mathcal{U} \to \mathcal{U}^d \times \mathcal{Y}^d$ such that the equation $\overline{z}_{k+1}^d = F(\overline{z}_k^d, u_k)$ holds. □

4.2 State reconstruction by KCCA

The conceptual idea behind the following approach for state reconstruction is that the state is the minimal interface between past and future input-output data. From Lemma 1 it follows that the state can be obtained from future values of the inputs and outputs

$$x_k = \Upsilon(u_k^d, y_k^d) =: \Psi_f(\overline{z}_{k+d}^d), \tag{14}$$

with the smooth map $\Psi_f : \mathcal{U}^d \times \mathcal{Y}^d \to \mathcal{X}$ where the subscript f is used to denote 'future'. In addition we have

$$x_k = f^d(x_{k-d}, \overline{u}_{k-d}^d) = f^d\Big(\Upsilon(\overline{z}_k^d), \overline{u}_{k-d}^d\Big) =: \Psi_p(\overline{z}_k^d), \tag{15}$$

with the smooth map $\Psi_p : \mathcal{U}^d \times \mathcal{Y}^d \to \mathcal{X}$, which shows that the states can be obtained from past values of the inputs and outputs. Thus, the state x_k satisfies

$$x_k = \Psi_f(\overline{z}_{k+d}^d) = \Psi_p(\overline{z}_k^d).$$

Since we do not know the system, we do not know the maps Ψ_f and Ψ_p. However, we can determine the state up to a nonlinear state transformation T by searching for maps $\Psi_{f,T} : \mathcal{U}^d \times \mathcal{Y}^d \to \mathbb{R}^n$ and $\Psi_{p,T} : \mathcal{U}^d \times \mathcal{Y}^d \to \mathbb{R}^n$ such that $\Psi_{f,T}(\overline{z}^d_{k+d}) = \Psi_{p,T}(\overline{z}^d_k)$ for all k. The transformed state is then given by $\xi_k = \Psi_{f,T}(\overline{z}^d_{k+d}) = \Psi_{p,T}(\overline{z}^d_k) = T(x_k)$.

The functions $\Psi_{f,T}$ and $\Psi_{p,T}$ can be determined using the KCCA algorithm of Section 3 as follows: let $\psi_{\cdot,T,j} : \mathcal{U}^d \times \mathcal{Y}^d \to \mathbb{R}$ be the jth entry of the function $\Psi_{\cdot,T}$ and parameterize it as follows:

$$\begin{aligned} \psi_{f,T,j}(\overline{z}^d_{k+d}) &= \varphi^T_{f,T}(\overline{z}^d_{k+d})v_j, \\ \psi_{p,T,j}(\overline{z}^d_k) &= \varphi^T_{p,T}(\overline{z}^d_k)w_j, \end{aligned}$$

with $\varphi_{f,T}$ and $\varphi_{p,T}$ playing the roles of φ_x and φ_y in the KCCA problem (5). Thus, the entries of the state vector ξ_k are obtained by determining the maximum correlation between $\varphi_f(\overline{z}^d_{k+d})$ and $\varphi_p(\overline{z}^d_k)$ in the CCA sense.

In this way the state ξ_k is obtained as a nonlinear transformation of a low dimensional intersection of the linear feature maps of the past data $\overline{z}^d_k$ and future data $\overline{z}^d_{k+d}$. From the linear case it is known that the dimension of the intersection depends on the persistency of excitation conditions of the input (Moonen et al., 1989). Generalizing this for the nonlinear case is a topic for future research.

Once the state sequence is constructed up to a nonlinear state transformation, we can determine the corresponding system description given by $\overline{f}$ and $\overline{h}$ in (11)–(12). The functions $\overline{f}$ and $\overline{h}$ can be approximated using a suitable approximation method, like for example: a sigmoidal neural network, piece-wise affine function, or LS-SVM.

5 Simulation example

In this section, we illustrate the use of KCCA for nonlinear state-space identification on the following example:

$$\begin{aligned} \frac{d}{dt}x_1(t) &= x_2(t) - 0.1\cos\Big(x_1(t)\Big)\Big(5x_1(t) - 4x_1^3(t) + x_1^5(t)\Big) - 0.5\cos\Big(x_1(t)\Big)u(t), \\ \frac{d}{dt}x_2(t) &= -65x_1(t) + 50x_1^3(t) - 15x_1^5(t) - x_2(t) - 100u(t), \\ y(t) &= x_1(t). \end{aligned}$$

This system was simulated using a 4th and 5th order Runge-Kutta method with a sampling time of 0.05 s. The input was a zero-order hold white noise signal, uniformly distributed between -0.5 and 0.5.

As a first step, a reconstruction of the state was obtained by applying KCCA on a set of 600 data points, using past and future delay vectors of the input-output data with dimension $d = 40$ each. As a kernel function we used the radial basis function (RBF): $K(x_i, x_j) = \exp(-\|x_i - x_j\|_2^2/\sigma^2)$. The parameters to be tuned for the KCCA procedure are thus the width of the kernel σ and the regularization parameters ν_x and ν_y which we chose to be equal: $\nu_x = \nu_y$.

Using the reconstructed state sequence, the functions $\overline{f}$ and $\overline{h}$ in (11)–(12) were approximated by means of a feedforward neural network with one hidden layer of fifteen neurons. The

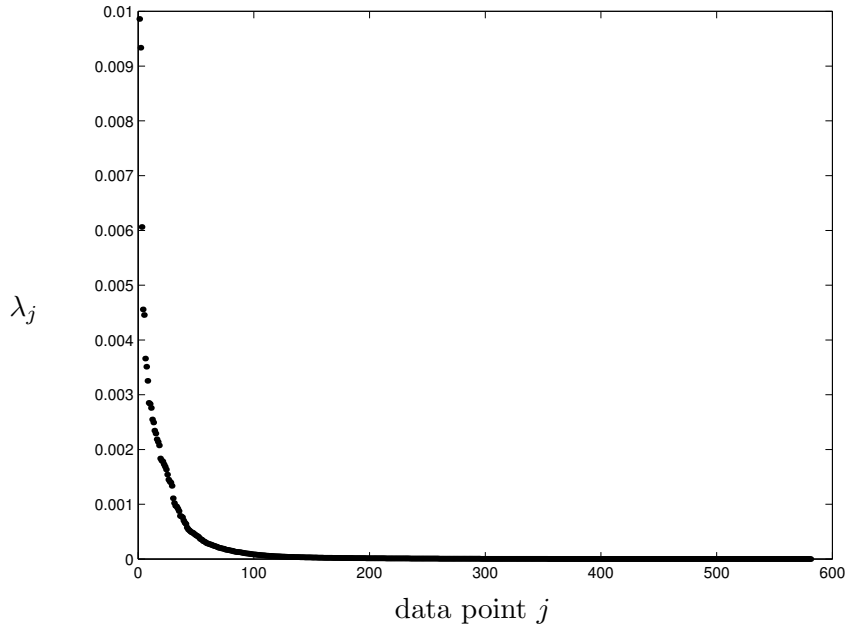


Figure 1: The kernel canonical correlations λ_j between the past and future data.

network was trained using the Levenberg-Marquardt algorithm in combination with Bayesian regularization. The resulting nonlinear state space model was then validated using a fresh data set of 400 points. The linear correlation coefficient between the predicted and the simulated output was calculated as a performance measure.

This whole procedure was repeated for a grid of parameter values σ and ν, out of which we chose the pair with the highest correlation value: $\sigma = 10^3$ and $\nu_x = \nu_y = 10^2$. For this pair the canonical correlations following from the KCCA procedure are shown in Figure 1. Figure 2 shows the corresponding reconstructed state trajectory ξ_k together with the original state trajectory x_k. They correspond quite well.

As a final step, the obtained nonlinear state-space model was tested on an entirely independent set of 400 data points. Figure 3 shows the free run simulation results of the model together with the results of a linear state space model obtained via subspace identification (N4SID). The correlation coefficient between the predicted and the simulated output was 0.95 for the nonlinear model, versus 0.88 for the linear model.

6 Conclusions

Using CCA states can be recovered from a time-invariant linear system. This process is known as the subspace intersection algorithm. In this paper, we show that the intersection algorithm has a natural extension to nonlinear systems by applying the concept of KCCA. Through an implicitly defined feature map, the columns of Hankel matrices containing past and future data are transferred to feature space and their relations are studied using the KCCA algorithm. A state vector estimate is obtained as a nonlinear transformation of the intersection in feature space by using an LS-SVM approach to the KCCA problem which boils down to solving a generalized eigenvalue problem involving kernel matrices with regularization. The conceptual

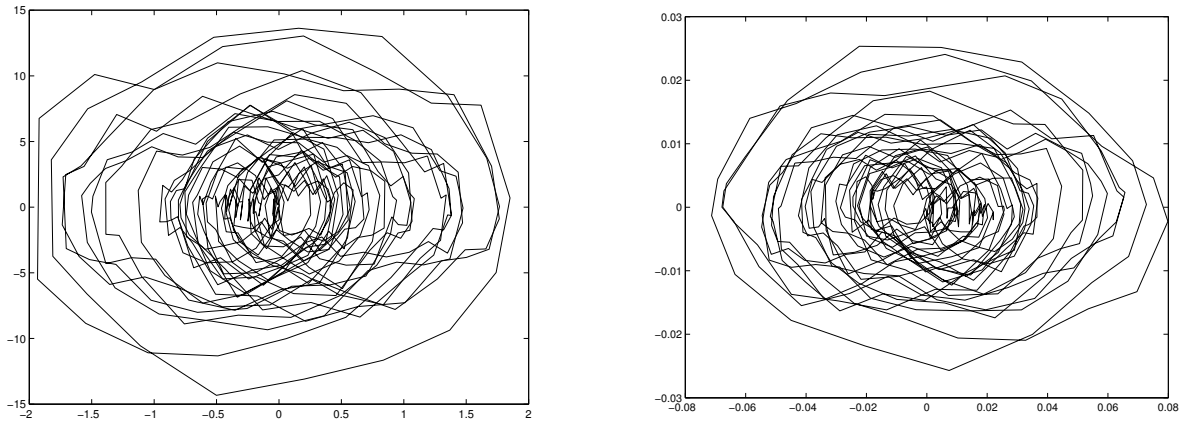


Figure 2: The original state trajectory x_k *(left)* and the reconstructed state trajectory ξ_k *(right)* plotted in the state space.

idea behind this approach for state reconstruction is that the state is the minimal interface between past and future input-output data. Once the state sequence is constructed up to a nonlinear state transformation, we can determine the corresponding system; in general a nonlinear map describing the internal state dynamics and the inclusion of new inputs into the system, and a nonlinear map relating states to outputs.

Acknowledgments

The authors would like to thank Luc Hoegaerts for useful discussions during the writing of this paper. Part of this research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven. It is supported by grants from several funding agencies and sources: Research Council KU Leuven: Concerted Research Action GOA-Mefisto 666 (Mathematical Engineering), IDO (IOTA Oncology, Genetic networks), several PhD/postdoc & fellow grants; Flemish Government: Fund for Scientific Research Flanders (several PhD/postdoc grants, projects G.0407.02 (support vector machines), G.0256.97 (subspace), G.0115.01 (bio-i and microarrays), G.0240.99 (multilinear algebra), G.0197.02 (power islands), research communities ICCoS, ANMMM), AWI (Bil. Int. Collaboration Hungary/ Poland), IWT (Soft4s (softsensors), STWW-Genprom (gene promotor prediction), GBOU-McKnow (Knowledge management algorithms), Eureka-Impact (MPC-control), Eureka-FLiTE (flutter modeling), several PhD grants); Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001) and IUAP V-10-29 (2002-2006) (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling), Program Sustainable Development PODO-II (CP/40: Sustainability effects of Traffic Management Systems); Direct contract research: Verhaert, Electrabel, Elia, Data4s, IPCOS.

References

Bach, F. R. and M. I. Jordan (2002). Kernel independent component analysis. *Journal of Machine Learning Research* **3**, 1–48.

Gittins, R. (1985). *Canonical Analysis – A Review with Applications in Ecology.* Berlin: Springer-Verlag. ISBN 0-387-13617-7.

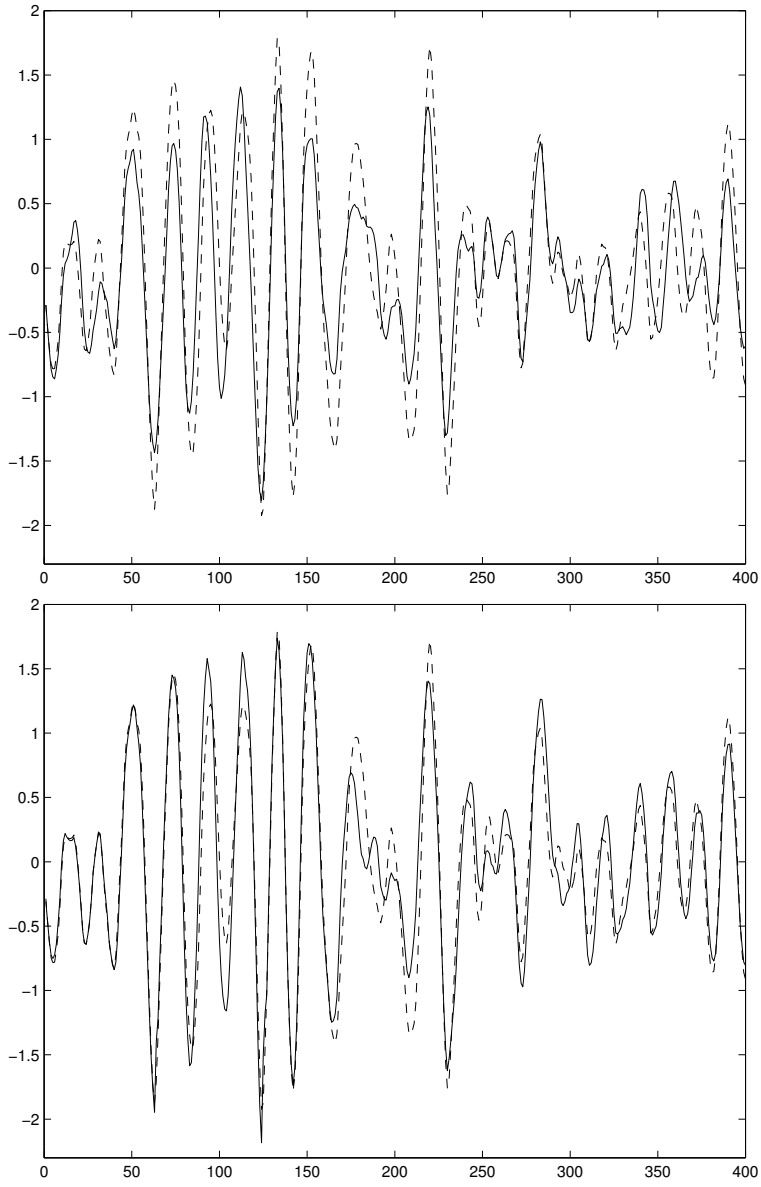


Figure 3: Free run simulation of the output: the linear state-space model *(top figure)*, and the nonlinear state-space model *(bottom figure)*. The dashed lines represent the real output, the full lines the output of the models.

Lai, P. L. and C. Fyfe (2000). Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems* **10**(5), 365–377.

Larimore, W. E. (1992). Identification and filtering of nonlinear systems using canonical variate analysis. In M. Casdagli and S. Eubank (Eds.), *Nonlinear Modelling and Forecasting*, SFI Studies in the Science of Complexity, pp. 263–303. Addison-Wesley.

Ljung, L. (1999). *System Identification: Theory for the User* (second ed.). Upper Saddle River, New Jersey: Prentice-Hall. ISBN 0-13-656695-2.

Moonen, M., B. De Moor, L. Vandenberghe and J. Vandewalle (1989). On- and off-line identification of linear state-space models. *International Journal of Control* **49**(1), 219–232.

Nerrand, O., P. Roussel-Ragot, D. Urbani, L. Personnaz and G. Dreyfus (1994). Training recurrent neural networks: Why and how? An illustration in dynamical process modeling. *IEEE Transactions on Neural Networks* **5**(2), 178–184.

Nijmeijer, H. (1982). Observability of autonomous discrete time non-linear systems: a geometric approach. *International Journal of Control* **36**(5), 867–874.

Schölkopf, B. and A. J. Smola (2002). *Learning with Kernels*. Cambridge, Massachusetts: MIT press. ISBN 0-262-19475-9.

Suykens, J. A. K., T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle (2002). *Least Squares Support Vector Machines*. Singapore: World Scientific. ISBN 981-238-151-1.

Van Overschee, P. and B. De Moor (1996). *Subspace Identification for Linear Systems; Theory, Implementation, Applications*. Dordrecht, The Netherlands: Kluwer Academic Publishers. ISBN 0-7923-9717-7.

Vapnik, V. (1998). *Statistical Learning Theory*. New York: John Wiley & Sons. ISBN 0-471-03003-1.

Verdult, V., M. Verhaegen and J. Scherpen (2000). Identification of nonlinear nonautonomous state space systems from input-output measurements. In *Proceedings of the 2000 IEEE International Conference on Industrial Technology*, Goa, India (January), pp. 410–414.