Article

A Spline Kernel-Based Approach for Nonlinear System Identification with Dimensionality Reduction

Wanxin Zhang ^{1,*} and Jihong Zhu ²¹ School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China² Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; jhzhu@tsinghua.edu.cn

* Correspondence: zhangwx59@mail.sysu.edu.cn

Received: 18 May 2020; Accepted: 29 May 2020; Published: 5 June 2020

Abstract: This paper proposes a novel approach for identification of nonlinear systems. By transforming the data space into a feature space, kernel methods can be used for modeling nonlinear systems. The spline kernel is adopted to produce a Hilbert space. However, a problem exists as the spline kernel-based identification method cannot deal with data with high dimensions well, resulting in huge computational cost and slow estimation speed. Additionally, owing to the large number of parameters to be estimated, the amount of training data required for accurate identification must be large enough to satisfy the persistence of excitation conditions. To solve the problem, a dimensionality reduction strategy is proposed. Transformation of coordinates is made with the tool of differential geometry. The purpose of the transformation is that no intersection of information with relevance to the output will exist between different new states, while the states with no impact on the output are extracted, which are then abandoned when constructing the model. Then, the dimension of the kernel-based model is reduced, and the number of parameters to be estimated is also reduced. Finally, the proposed identification approach was validated by simulations performed on experimental data from wind tunnel tests. The identification result turns out to be accurate and effective with lower dimensions.

Keywords: differential manifold; dimensionality reduction; identification; nonlinear system; spline kernel

1. Introduction

System identification plays an important role in control systems with applications in many fields, e.g., biology, chemistry, physics, electrical engineering, aeronautical engineering, computer engineering, and marine systems [1–7]. When the model for the dynamic system is unknown, the identification method finds an extract representation with estimated parameters for the system, which is the base for designing the control law. System identification is divided into two steps. The first step is model selection [8], and the second step is parameter estimation [9]. The model selection depends on the model validation techniques [10]. The most commonly used is cross validation [11]. In terms of parameter estimation, the most widely used methods are prediction error-based parametric methods [12,13], such as least squares algorithm and maximum likelihood algorithm.

Recently, system identification attracts many interests from researchers in the machine learning field. In terms of the uncertainty and nonlinearity, machine learning provides a tool for modeling the control system. Machine learning schemes support efficient models in system identification. Many existing modeling frameworks are proposed with elements of artificial intelligence, which can be integrated to the control loop. The use of artificial intelligence combination to solve system identification problems produces effective results.

The artificial intelligence techniques applied for modeling include neural network [14,15], K-nearest neighbors [16], support vector machine [17–19], and kernel methods [20,21]. Each of these artificial techniques has its own characteristics. Neural networks are usually not adopted in the control system. Since a neural network exists as a black box, no physical meaning is performed by both the structure and parameters. Consequently, when mistakes happen in modeling the system, it is hard to find the cause. K-nearest neighbors method is also usually not desired to be used, as the performance degrades heavily when dealing with big data or high dimensions, and the method is sensitive to the noises in the data. Support vector machine can be regarded as one of the kernel methods.

Among the artificial intelligence techniques applied for identification problems, kernel methods are most widely used, with different definitions of kernel functions in different situations. Kernel methods provide a structured framework to transform the data space into a feature space. The transformation is mostly nonlinear. Therefore, kernel methods can be used for modeling nonlinear systems. When kernel methods are used for system identification, the model is constructed in the feature space produced, which is always proved to be a Hilbert space in applications. Pillonetto and De Nicolao [20] defined a Mercer kernel to obtain a new Gaussian prior, and then constructed the identification problem in the unique reproducing kernel Hilbert space. Pillonetto et al. [22] also concluded remarkable results of kernel-based methods for system identification. Chen [23] discussed the continuous-time diagonal correlated kernel by deriving from the first-order spline kernel, for identification of LTI system. Han et al. [21] employed the Gaussian kernel function to resolve the inner product in learning algorithms. Moreno et al. [3] proposed kernel ridge regression which combines the kernel trick with ridge regression.

However, there exists a problem for kernel method that, in cases with large dimensions, the performance of kernel-based identification method is poor due to the many parameters to be estimated, together with the large computational cost and low constructing speed. The accuracy is especially low when few data are obtained for training. The complexity of kernel methods depends on the number of training data and the dimension of the input. Therefore, reducing the dimension of the variables which will be input to the kernel method would contribute to the performance of the identification result. For the purpose, a new coordinate with new state variables is constructed with the tool of differential geometry. The variables including the information of the output are retained, while the rest are abandoned. Then, the spline kernel-based model is constructed with a lower dimension, and the number of parameters to be estimated is reduced. Meanwhile, the generalization ability of the model and the estimation speed are improved.

This paper is organized as follows. The identification problem is described first in Section 2, and then related work on kernel method in the identification process is given. In Section 3, the spline kernel-based identification method is first given, and then we propose a dimension reduction method for the method. In the proposed method, the key step is the coordinate transformation, which is discussed in detail in Section 4. In Section 5, we demonstrate the effectiveness of the proposed method by simulations on experimental data taken from the wind tunnel tests for identification of aerodynamics. Conclusions are finally made.

2. Related Work

In most cases, the intended use for system identification is control design. For the perspective of control systems, the state variable is denoted as the vector x , consisting of several states with $x = [x_1, x_2, x_3, \dots]^T$, while the input and output of the system are denoted as u and y , respectively. The goal of system identification is to construct the relationship between u and y , given the set of pairs $(x_k, y_k) \in (\mathbb{R}^n, \mathbb{R})$, where x_k is the value of the state vector at time k with $x_k = [x_1^k, x_2^k, \dots, x_n^k]^T$, and n is the number of the states.

Then, the relationship between u and y can be given by a state-space model. In the model set, a simple example is the linear state space equations,

$$\begin{aligned}\dot{x} &= Ax + Bu \\ y &= Cx\end{aligned}\quad (1)$$

where A , B , and C are matrices with estimated parameters.

With the given parametric form in Equation (1), the problem of system identification is transformed to be the problem of parameter estimation from sampled data from observations. The idea is extended with the goal to identify nonlinear systems. Consider a nonlinear state-space model [24–26],

$$\begin{aligned}\dot{x} &= f(x) + g(x)u \\ y &= h(x)\end{aligned}\quad (2)$$

where $f(x) = [f_1(x), f_2(x), \dots, f_n(x)]^T$ and $g(x) = [g_1(x), g_2(x), \dots, g_n(x)]^T$ are sets of $\mathbb{R}^n \rightarrow \mathbb{R}$ mappings defined on state space, which are usually obtained according to the physical mechanism and prior knowledge of the system, and $h(x)$ is a nonlinear real-valued function defined on state space.

In terms of the parametric model, the function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is determined by the parameter vector denoted as θ , after the model structure is selected. The identification method is used to find the function $h(x)$ which approximates the output well. In the step of constructing model structures, the function is often assumed to be linear in θ [27], and then the relationship between y and x can be represented as $h(x) = \sum_{i=1}^m \theta_i \phi_i(x)$, where θ_i is the i th element in the parameter vector θ , $\{\phi_1, \phi_2, \dots, \phi_m\}$ is the basis function set, and m is the number of functions in the set. There exist many choices for the form of basis function. Among them, the polynomial model is most widely used [28], with $\phi_i(x) = (\lambda^T x)^i$, where λ is a weight vector. m is determined by the model validation technique. Then, the least squares algorithm can be used to estimate θ by minimizing $\sum_{i=1}^N (y_i - h(x_i))^2$, where N is the number of data.

However, generalization ability cannot be guaranteed when the objective function in the minimization is merely determined by the error between the prediction and the real value. Additionally, overfitting will be caused, which means small noise in the data will have significant effect on the final model. A method usually used for deflecting overfitting is regularization [29]. A regularization term is added to the objective function in the minimization process [22],

$$\min_{\theta} \left(\sum_{i=1}^N (y_i - h(x_i))^2 + \gamma r(h) \right), \quad (3)$$

where $r(\cdot)$, a function of h , represents the regularizer, which can penalize the unexpected features of h , and γ is the weight parameter, which is used to adjust the relative importance between the error term $\sum_{i=1}^N (y_i - h(x_i))^2$ and the regularization term $r(h)$.

To reduce the sensitivity to the noises, the high-frequency components in h is reduced by penalizing the energy of high-order derivatives [30], which is given by

$$r(h) = \int (h^{(p)}(x))^2 dx, \quad (4)$$

where $h^{(p)}(x)$ is the p th order of $h(x)$, and p is determined by the properties of noises in practical applications.

3. Proposed Framework

3.1. Problem Formulation

In recent years, the kernel method has been adopted to construct the basis function. With a defined kernel, $h(x)$ is represented as

$$h(\mathbf{x}) = \sum_{i=1}^N \theta_i K_{x_i}(\mathbf{x}), \quad (5)$$

where $K_{x_i}(\cdot)$ denotes the kernel function, which can be also described as $K(\mathbf{x}_i, \mathbf{x})$.

Based on the definition of the kernel, a Hilbert space $\mathcal{H}$ can be constructed by proving that the Cauchy sequences with the induced norm in the space will converge [31,32]. According to the expected properties for the final model, we set the form for the regularizer $r(\cdot)$ in Equation (3) to be

$$r(h) = \|h\|_{\mathcal{H}}^2 = \sqrt{\langle h, h \rangle_{\mathcal{H}}}, \quad (6)$$

from which, it can be easily derived that

$$r(h) = \sum_{i=1}^N \sum_{j=1}^N \theta_i \theta_j K(\mathbf{x}_i, \mathbf{x}_j). \quad (7)$$

3.2. Spline Kernel-Based Identification Method

The most widely used kernels include the linear kernels, radial basis kernels, and spline kernels. As is discussed, the regularizer has the form of Equation (6) within the constructed Hilbert space. It is always desired to reduce the impact of noises in identification. There is great significance to design the $r(h)$ in the form of Equation (4), which penalizes the energy of high-order derivatives. If Equations (4) and (6) can match with each other, the kernel has huge advantages in modeling $h(\mathbf{x})$. The spline kernel is chosen, because it satisfies the matching relation with the form

$$K(\mathbf{x}_1, \mathbf{x}_2) = \int_0^1 G_p(\|\mathbf{x}_1\|, t) G_p(\|\mathbf{x}_2\|, t) dt, \quad (8)$$

where

$$G_p(\|\mathbf{x}\|, t) = \begin{cases} \frac{(\|\mathbf{x}\| - t)^{p-1}}{(p-1)!} & \text{if } \|\mathbf{x}\| \geq t \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

With the tool of least squares support vector machine method [33], the estimation for θ is obtained by

$$\hat{\theta} = (\mathbf{K} + \gamma \mathbf{I})^{-1} \mathbf{y}, \quad (10)$$

where $\mathbf{K}$ is the kernel matrix with $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{I}$ is an $N \times N$ identity matrix, and $\mathbf{y}$ is the output vector with $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$.

3.3. Dimensionality Reduction

From Equation (5), we can find that the dimension of state vector $\mathbf{x}$, i.e., the number of states, has a huge influence on the model complexity, which also determines the computational complexity of estimation according to Equation (10). However, it is often the case that some states may have little impact on the output of the system because the output has no relationship with the part of states or the information in some states related to the output has already been contained in other states. Consequently, the dimension for the output model can be reduced.

Considering that there exists an intersection among the information of different states, which leads to redundancy in modeling, it is hoped that the information with relevance to the output can be concentrated, and the variables, which provide no additional information on the output except the existing variables, should be abandoned in the modeling process in Equation (5). Coordinate transformation is proposed in this paper, which is discussed in detail in next section. The new state vector is denoted as $\mathbf{z}$, with $\mathbf{z} \in \mathbb{R}^n$. The function realizing the transformation is denoted as $\mathbf{z} = \mathbf{t}(\mathbf{x}) = [t_1(\mathbf{x}), t_2(\mathbf{x}), \dots, t_n(\mathbf{x})]^T$, where $t_i(\mathbf{x})$ is a $\mathbb{R}^n \rightarrow \mathbb{R}$ mapping. $\mathbf{z}$ is divided into two part,

$$\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2]^T, \quad (11)$$

where each of the elements in $\mathbf{z}_1$ is related with y , while all the elements in $\mathbf{z}_2$ have no effect on y . Since the definition of the kernel has no dependence on the dimension of the point, the form of the kernel function remains unchanged. The purpose for the transformation is to reduce the dimension of Equation (5)

$$h(\mathbf{x}) = h(\mathbf{z}_1) \quad (12)$$

with

$$\sum_{i=1}^N \theta_i K_{x_i}(\mathbf{x}) = \sum_{i=1}^{N'} \beta_i K_{z_{1i}}(\mathbf{z}_1), \quad (13)$$

where β is the parameter vector to be estimated, $\mathbf{z}_1$ is part of the transformed state vector $\mathbf{z}$, N' is the number of data for modeling after transformation, and $\mathbf{z}_1 \in \mathbb{R}^d$, $d \leq n$.

Figure 1 shows the graphical illustration of the proposed framework. It is easy to complete the coordinate transformation for a linear system. However, for a nonlinear system, the transformation is difficult, especially when the transformation is also a nonlinear change.

Figure 1. Graphical illustration of the proposed framework.

4. Differential Manifold-Based Transformation

As discussed in the previous section, some of the states in $\mathbf{x}$ may present similar information related to y , and some of the states in $\mathbf{x}$ may have little impact on y . Aiming at removing the correlation of different states with y , we adopt coordinate transformation to reduce the dimension. Coordinate transformation has been used to obtain more efficient solutions and accelerate the convergence process in optimization problems [34,35]. However, the linear coordinate transformation, which is most widely used, cannot deal with the problem.

We deal with the problem with the tool of differential geometry [36] to obtain nonlinear coordinate transformation. A smooth distribution of dimension $n - d$ is first defined, denoted as Δ , which represents a differentiable manifold. The manifold Δ is required to be invariant under the vector fields f, g , i.e., $f(\mathbf{x}) \in \Delta, g(\mathbf{x}) \in \Delta$ for all $\mathbf{x}$ in Δ . A vector field on Δ means an assignment of one tangent vector to each $\mathbf{x}$ in Δ . f and g can be regarded as mappings from Δ into the tangent bundles.

To find the transformation $t(\mathbf{x})$, which is a vector of n real-valued functions of n variables, denoted as

$$\mathbf{z} = t(\mathbf{x}) = \begin{pmatrix} t_1(x_1, x_2, \dots, x_n) \\ t_2(x_1, x_2, \dots, x_n) \\ \vdots \\ t_n(x_1, x_2, \dots, x_n) \end{pmatrix}, \quad (14)$$

two assumptions are made:

- The inverse function $t'(\mathbf{z})$ of $t(\mathbf{x})$ can be found. For any $\mathbf{x}$ in Δ , $t'(t(\mathbf{x})) = \mathbf{x}$.
- The functions $t(\mathbf{x})$ and $t'(\mathbf{z})$ are both continuous and analytic on their domain of definition.

Based on the division of $\mathbf{z}$ in Equation (11), we have

$$\frac{\partial h}{\partial z_i} = 0, \quad (15)$$

where $d < i \leq n$.

Then, Δ is contained in the distribution $(\text{span}\{dh\})^\perp$. Ω is denoted as the orthogonal complement distribution of Δ with $\Omega = \Delta^\perp$. For any x in Δ ,

$$\Delta^\perp(x) = \{a^* \in (\mathbb{R}^n)^* : \langle a^*, b \rangle = 0 \text{ for all } b \in \Delta(x)\}, \quad (16)$$

where a^* is the dual vector and $(\mathbb{R}^n)^*$ is the dual space. If b is a column vector,

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}, \quad (17)$$

which is a point in Δ , then a^* is a row vector,

$$a^* = [a_1, a_2, \dots, a_n], \quad (18)$$

which is a point in Ω . The inner product $\langle a^*, b \rangle$ is given by

$$\langle a^*, b \rangle = \sum_{i=1}^n a_i b_i. \quad (19)$$

Then, it can be given that Ω contains the $\text{span}\{dh\}$, and also is invariant under the vector fields f, g .

Aiming at obtaining the maximal dimension reduction of Equation (5), we need to find the smallest distribution which contains Ω and is invariant under the vector fields f, g , denoted by $\langle f, g | \Omega \rangle$. A nondecreasing sequence of distributions is defined

$$\begin{aligned} \Omega_0 &= \Omega = \text{span}\{dh\}, \\ \Omega_1 &= \Omega_0 + [f, \Omega_0] + [g, \Omega_0] = \text{span}\{dh, [f, dh], [g, dh]\}, \\ &\dots \\ \Omega_k &= \Omega_{k-1} + [f, \Omega_{k-1}] + [g, \Omega_{k-1}]. \end{aligned} \quad (20)$$

where $[,]$ is the Lie bracket of vector field, which is an operator to assign a third vector field given two vector fields on a smooth manifold. For any given functions $f_1(x), f_2(x)$, the operation $[f_1, f_2](x)$ is given by

$$[f_1, f_2](x) = \frac{\partial f_2}{\partial x} f_1(x) - \frac{\partial f_1}{\partial x} f_2(x). \quad (21)$$

According to Lemma 1.9.2 in [36], in iteration process of Equation (20), the equation $\Omega_{k^*} = \Omega_{k^*+1}$ finally appears for some integer k^* . Meanwhile, $\langle f, g | \Omega \rangle$ is obtained,

$$\langle f, g | \Omega \rangle = \Omega_{k^*}. \quad (22)$$

Finally, we find the real-valued functions $z_1(x) = [t_1(x), t_2(x), \dots, t_d(x)]^T$ such that

$$\text{span}\{t_1(x), t_2(x), \dots, t_d(x)\} = \Omega_{k^*}. \quad (23)$$

Then, the model of Equation (2) is transformed to

$$\begin{aligned}\dot{z}_1 &= \tilde{f}(z_1) + \tilde{g}(z_1)u, \\ y &= \tilde{h}(z_1),\end{aligned}\quad (24)$$

where

$$\begin{aligned}\tilde{f}(z) &= \left[\frac{\partial f}{\partial x} \right]_{x=l'(z)}, \\ \tilde{g}(z) &= \left[\frac{\partial g}{\partial x} \right]_{x=l'(z)}, \\ \tilde{h}(z) &= [h(x)]_{x=l'(z)}.\end{aligned}\quad (25)$$

Compared with the original description, the dimension of the new description is reduced. The number of data required for accurate identification is reduced, which corresponds to N' in Equation (13). Consequently, the number of parameters to be estimated is highly reduced.

5. Simulation and Evaluation

In this section, the proposed identification approach is evaluated and compared with the commonly used methods. A brief introduction to the experimental settings is first given. Comparison of the evaluation results with other methods is then made.

5.1. Background Information

To validate the proposed method for nonlinear systems, aerodynamic coefficients at large angles of attack were considered, which present significant nonlinear phenomena in the wind tunnel experiments. With the FL-8 wind tunnel, large amplitude forced oscillation tests were made. The model to be tested was chosen to be the fourth-generation fighter configuration [37]. Table 1 gives the main physical parameters for the experimental model. In the tests, the flow speed was 30 m/s. The sampling period was 0.02 s.

Table 1. Physical parameters of the experimental prototype.

Physical Parameter	Value
Weight	≤ 8 kg
Length	1.1825 m
Wing Span	0.8475
Wing Area	0.3047 m ²
Mean Aerodynamic Chord	0.4269 m
Position of Aerodynamic Center	0.6926 m

The state vector for aerodynamic identification was constructed by

$$\mathbf{x} = [\alpha, \dot{\alpha}, \ddot{\alpha}, q, \dot{q}, \delta_e, M] \quad (26)$$

where α is the angle of attack, q is the pitch rate, δ_e is the elevator control surface deflection, and M is the Mach number.

The state equation for $x_1(k) = \alpha(k)$, $x_2(k) = \dot{\alpha}(k)$, $x_3(k) = \ddot{\alpha}$ can be given by

$$\begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_2(k-1) \\ x_3(k-1) \end{bmatrix} + \begin{bmatrix} w_1(k-1) \\ w_2(k-1) \\ w_3(k-1) \end{bmatrix}. \quad (27)$$

where Δt denotes the time period of the measurements, and $w(k)$ is the state noise.

Corresponding to the form of Equation (2), the first three elements of $f(x)$ and $g(x)$ are linear. The construction for the remaining part of $f(x)$ and $g(x)$ can be obtained according to the physical mechanism, e.g., through table interpolation.

5.2. Comparison Results and Discussions

The nonlinear aerodynamic coefficients in the large amplitude pitch oscillation tests for three frequencies are shown in Figure 2. The test is conducted at the balanced angle of 40° with amplitude of 40° , and the sideslip angle is -30° .

Figure 2. Example of longitudinal oscillations: (a) lift coefficients C_L for 0.4, 0.6, and 0.8 Hz; and (b) Drag coefficients C_D for 0.4, 0.6, and 0.8 Hz.

When the test data are not sufficient, the data-driven models including step response model and neural networks cannot be identified. We compared the proposed identification approach with the standard spline kernel-based method. We chose the experimental condition with frequency of 0.8 Hz. The estimation results are shown in Figure 3. In terms of the lift coefficient with slow time-varying characteristics, the proposed approach and the standard spline kernel method demonstrate almost the same tracking ability. The improvement is not obvious because of the locally linear features in the curve. In terms of the drag coefficient, which contains more complex nonlinear characteristics, the performance of the proposed method is obviously better. Given the same amount of training data, more features are caught by the proposed approach. The comparison results of objective function with the form of Equation (3) is given in Table 2, which proves that the proposed algorithm can track the hysteresis with a higher accuracy. As discussed above, in addition to the accuracy, the construction of Equation (3) also takes into account of the generalization ability of the model. Therefore, the proposed method has better generalization ability and is less sensitive to the noises in observations.

Additionally, with the strategy of dimensionality reduction, the new state vector z transformed from x is divided into z_1 and z_2 . The dimension of the state vector z_1 finally input to the spline kernel-based model is 3. Therefore, the dimension of the final output model $h(z_1)$ is 3. Compared with the standard spline kernel-based model, the proposed approach constructs a model with fewer parameters. Estimation is made with lower computational cost, while the information in the data is used more efficiently.

Table 2. Comparison results of values of objective function in Equation (3).

Algorithm	C_L	C_D
The Proposed Approach	1.2996	0.0032
The Standard Kernel Method	1.6458	0.0102

Figure 3. Real-time estimation results of the longitudinal oscillation, where the dotted line shows the real values from the wind tunnel test and the solid line is the estimation result: (a) Lift coefficients C_L with the proposed approach; (b) lift coefficients C_L with the standard spline kernel method; (c) drag coefficients C_D with the proposed approach; and (d) drag coefficients C_D with the standard spline kernel method.

6. Conclusions

In this work, a novel spine kernel-based approach with dimensionality reduction is designed for nonlinear system identification. A regularization term based on the produced Hilbert space is added into the objective function to guarantee the generalization ability of the model. Dimensionality reduction is proposed to solve the problem of kernel methods when applied to data with high dimensions. By reducing the dimension of the state vector, the number of parameters in the estimation is significantly reduced. With the tool of Lie bracket, nonlinear coordinate transformation is made based on the defined distribution of the differentiable manifold. Clear physical meaning is given that the transformation concentrates the information on the states related to the output, and abandons the redundant states. After the transformation, the new state vector reduces the dimensions of the new description. Evaluation on the nonlinear aerodynamic coefficients under the wind tunnel tests was made. The experimental results show that the proposed method successfully approximates the data in the nonlinear system and achieves higher accuracy and better generalization ability through dimensionality reduction. The future work will include analyzing the effects of the regularization parameter γ on the performance of the identification method, and subsequently linking the estimation of γ into the identification framework.

Author Contributions: Conceptualization, W.Z. and J.Z.; methodology, W.Z. and J.Z.; software, W.Z.; validation, W.Z.; resources, J.Z.; and data curation, W.Z. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 61673240 and the National Defense Innovation Special Zone of Science and Technology Project under Grant 18-163-00-TS-006-038-01.

Acknowledgments: The authors would like to thank AVIC Aerodynamics Research Institute for their great help on the wind tunnel tests and the anonymous reviews for their helpful suggestions to improve the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, M.; Feng, Z.; Zhou, X.; Xiao, R.; Qi, Y.; Zhang, X. Specific emitter identification based on synchrosqueezing transform for civil radar. *Electronics* **2020**, *9*, 658. [\[CrossRef\]](#)
2. Guo, M.; Su, Y.; Gu, D. System identification of the quadrotor with inner loop stabilisation system. *Int. J. Model. Identif. Control* **2017**, *28*, 245–255. [\[CrossRef\]](#)
3. Moreno, R.; Moreno-Salinas, D.; Aranda, J. Black-box marine vehicle identification with regression techniques for random manoeuvres. *Electronics* **2019**, *8*, 492. [\[CrossRef\]](#)
4. Baldini, A.; Ciabattoni, L.; Felicetti, R.; Ferracuti, F.; Freddi, A.; Monteriù, A. Dynamic surface fault tolerant control for underwater remotely operated vehicles. *ISA Trans.* **2018**, *78*, 10–20. [\[CrossRef\]](#)
5. He, W.; Ge, W.; Li, Y.; Liu, Y.J.; Yang, C.; Sun, C. Model identification and control design for a humanoid robot. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 45–57. [\[CrossRef\]](#)
6. Veksler, A.; Johansen, T.A.; Borrelli, F.; Realfsen, B. Dynamic positioning with model predictive control. *IEEE Trans. Control. Syst. Technol.* **2016**, *24*, 1340–1353. [\[CrossRef\]](#)
7. Haidegger, T.; Kovács, L.; Preitl, S.; Precup, R.E.; Benyo, B.; Benyó, Z. Controller design solutions for long distance telesurgical applications. *Int. J. Artif. Intell.* **2011**, *6*, 48–71.
8. Ljung, L.; Wahlberg, B. Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Adv. Appl. Probab.* **1992**, *24*, 412–440. [\[CrossRef\]](#)
9. Yang, C.; Jiang, Y.; He, W.; Na, J.; Li, Z.; Xu, B. Adaptive parameter estimation and control design for robot manipulators with finite-time convergence. *IEEE Trans. Ind. Electron.* **2018**, *65*, 8112–8123. [\[CrossRef\]](#)
10. Jasim, W.; Gu, D. Robust path tracking control for quadrotors with experimental validation. *Int. J. Model. Identif. Control* **2018**, *29*, 1–13. [\[CrossRef\]](#)
11. Pillonetto, G.; De Nicolao, G. Pitfalls of the parametric approaches exploiting cross-validation for model order selection. *IFAC Proc. Vol.* **2012**, *45*, 215–220. [\[CrossRef\]](#)
12. Fanesi, M.; Scaradozzi, D. Adaptive control for non-linear test bench dynamometer systems. In Proceedings of the 2019 IEEE 23rd International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 9–11 October 2019; pp. 768–773.
13. Sun, J.; Yang, Q. Research on least means squares adaptive control for automotive active suspension. In Proceedings of the 2008 IEEE International Conference on Industrial Technology, Chengdu, China, 21–24 April 2008; pp. 1–4.
14. Yang, E.; Gu, D.; Hu, H. Performance improvement for formation-keeping control using a neural network HJI approach. In *Trends in Neural Computation*; Springer: Berlin, Germany, 2007; pp. 419–442.
15. He, W.; Yan, Z.; Sun, Y.; Ou, Y.; Sun, C. Neural-learning-based control for a constrained robotic manipulator with flexible joints. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5993–6003. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Zhang, Z.; Jiang, T.; Li, S.; Yang, Y. Automated feature learning for nonlinear process monitoring—An approach using stacked denoising autoencoder and k-nearest neighbor rule. *J. Process. Control* **2018**, *64*, 49–61. [\[CrossRef\]](#)
17. Zhou, X.; Jiang, P.; Wang, X. Recognition of control chart patterns using fuzzy SVM with a hybrid kernel function. *J. Intell. Manuf.* **2018**, *29*, 51–67. [\[CrossRef\]](#)
18. Ai, Q.; Wang, A.; Zhang, A.; Wang, W.; Wang, Y. An effective multiclass twin hypersphere support vector machine and its practical engineering applications. *Electronics* **2019**, *8*, 1195. [\[CrossRef\]](#)
19. Chodunaj, M.; Szcześniak, P.; Kaniewski, J. Mathematical modeling of current source matrix converter with Venturini and SVM. *Electronics* **2020**, *9*, 558. [\[CrossRef\]](#)
20. Pillonetto, G.; De Nicolao, G. A new kernel-based approach for linear system identification. *Automatica* **2010**, *46*, 81–93. [\[CrossRef\]](#)
21. Han, T.; Wang, L.; Wen, B. The kernel based multiple instances learning algorithm for object tracking. *Electronics* **2018**, *7*, 97. [\[CrossRef\]](#)

22. Pillonetto, G.; Dinuzzo, F.; Chen, T.; De Nicolao, G.; Ljung, L. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica* **2014**, *50*, 657–682. [\[CrossRef\]](#)
23. Chen, T. Continuous-time DC kernel-a stable generalized first-order spline kernel. *IEEE Trans. Autom. Control* **2018**, *63*, 4442–4447. [\[CrossRef\]](#)
24. Dreano, D.; Tando, P.; Pulido, M.; Ait-El-Fquih, B.; Chonavel, T.; Hoteit, I. Estimating model-error covariances in nonlinear state-space models using Kalman smoothing and the expectation—Maximization algorithm. *Q. J. R. Meteorol. Soc.* **2017**, *143*, 1877–1885. [\[CrossRef\]](#)
25. Zhang, J.X.; Yang, G.H. Prescribed performance fault-tolerant control of uncertain nonlinear systems with unknown control directions. *IEEE Trans. Autom. Control* **2017**, *62*, 6529–6535. [\[CrossRef\]](#)
26. Zhang, K.; Zhang, H.; Xiao, G.; Su, H. Tracking control optimization scheme of continuous-time nonlinear system via online single network adaptive critic design method. *Neurocomputing* **2017**, *251*, 127–135. [\[CrossRef\]](#)
27. Ding, F.; Xu, L.; Alsaadi, F.E.; Hayat, T. Iterative parameter identification for pseudo-linear systems with ARMA noise using the filtering technique. *IET Control Theory Appl.* **2018**, *12*, 892–899. [\[CrossRef\]](#)
28. Esfahani, A.F.; Dreesen, P.; Tiels, K.; Noël, J.P.; Schoukens, J. Parameter reduction in nonlinear state-space identification of hysteresis. *Mech. Syst. Signal Process.* **2018**, *104*, 884–895. [\[CrossRef\]](#)
29. Zanotti, C.; Rotiroti, M.; Sterlacchini, S.; Cappellini, G.; Fumagalli, L.; Stefania, G.A.; Nannucci, M.S.; Leoni, B.; Bonomi, T. Choosing between linear and nonlinear models and avoiding overfitting for short and long term groundwater level forecasting in a linear system. *J. Hydrol.* **2019**, *578*, 124015. [\[CrossRef\]](#)
30. Schaefer, I.; Kosloff, R. Optimization of high-order harmonic generation by optimal control theory: Ascending a functional landscape in extreme conditions. *Phys. Rev. A* **2020**, *101*, 023407. [\[CrossRef\]](#)
31. Mu, B.; Chen, T.; Ljung, L. On asymptotic properties of hyperparameter estimators for kernel-based regularization methods. *Automatica* **2018**, *94*, 381–395. [\[CrossRef\]](#)
32. Steinwart, I.; Hush, D.; Scovel, C. An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *IEEE Trans. Inf. Theory* **2006**, *52*, 4635–4643. [\[CrossRef\]](#)
33. Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [\[CrossRef\]](#)
34. Fang, W.; Zhang, L.; Yang, S.; Sun, J.; Wu, X. A multiobjective evolutionary algorithm based on coordinate transformation. *IEEE Trans. Cybern.* **2019**, *49*, 2732–2743. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Yu, Z.; Qin, M.; Chen, X.; Meng, L.; Huang, Q.; Fu, C. Computationally efficient coordinate transformation for field-oriented control using phase shift of linear hall-effect sensor Signals. *IEEE Trans. Ind. Electron.* **2019**, *67*. [\[CrossRef\]](#)
36. Isidori, A. *Nonlinear Control Systems*; Springer Science & Business Media: Berlin, Germany, 2013.
37. Liu, C.; Zhao, Z.; Bu, C.; Wang, J.; Mu, W. Double degree-of-freedom large amplitude oscillation test technology in low speed wind tunnel. *Acta Aeronaut. Astronaut. Sin.* **2016**, *37*, 2417–2425.