

Application and comparison of kernel functions for linear parameter varying model approximation of nonlinear systems

Faisal Saleem¹ Ahsan Ali²
Inam-ul-hassan Shaikh² Muhammad Wasim³

Abstract. In this paper, a comparative study for kernel-PCA based linear parameter varying (LPV) model approximation of sufficiently nonlinear and reasonably practical systems is carried out. Linear matrix inequalities (LMIs) to be solved in LPV controller design process increase exponentially with the increase in a number of scheduling variables. Fifteen kernel functions are used to obtain the approximate LPV model of highly coupled nonlinear systems. An error to norm ratio of original and approximate LPV models is introduced as a measure of accuracy of the approximate LPV model. Simulation examples conclude the effectiveness of kernel-PCA for LPV model approximation as with the identification of accurate approximate LPV model, computation complexity involved in LPV controller design is decreased exponentially.

§1 Introduction

LPV is a class of systems to approximate the dynamic behavior of nonlinear systems by some linear dynamic relation between inputs and outputs of system. This relation depends on a set of some measurable signals which are called scheduling variables. These signals represent variable operating conditions of the original system. This ability to represent and capture nonlinear dynamics using linear relation, which depends on measurable signals makes the use of linear control theory possible in controller synthesis process.

In linear parameter varying (LPV) models, the nonlinear systems can be described as the parameterized linear systems where the parameters are dependent on the measurable signals

Received: 2019-11-16. Revised: 2020-09-28.

MR Subject Classification: 93Cxx.

Keywords: kernel-PCA, LMIs, LPV, error to norm ratio, computational complexity, and control design.

Digital Object Identifier(DOI): DOI: <https://doi.org/10.1007/s11766-023-3965-8>.

known as scheduling signals [1]. The LPV systems in state-space form, can be represented as:

$$\begin{aligned} \dot{x}(t) &= A(\theta(t))x(t) + B(\theta(t))u(t) \\ y(t) &= C(\theta(t))x(t) + D(\theta(t))u(t), \\ \theta(t) &\in S \end{aligned} \quad (1)$$

here $\theta(t) \in \mathfrak{R}_\theta$ represents the parameter-varying vector in region S , where \mathfrak{R} denotes the set of real numbers, $A(\theta(t)) \in \mathfrak{R}^{n_s \times n_s}$, $B(\theta(t)) \in \mathfrak{R}^{n_s \times n_u}$ are the system matrix and input matrix while n_s is number of states and n_u is the number of inputs of the LPV system. Where $x(t)$, $u(t)$, and $y(t)$ are state, input and output vectors. The scheduling parameters $\theta(t)$ vary as function of scheduling signals $\rho(t)$. The scheduling variables can be external or internal to the system making the model pure-LPV or quasi-LPV respectively. The function relating the parameter variations and scheduling signals is called the scheduling function, i.e., $\theta(t) = p(\rho(t))$. The scheduling function is chosen to reflect the similar behavior as the nonlinear model in the whole scheduling region S . Representation of a nonlinear system as an LPV system can be a nontrivial issue. There are many methods to represent the nonlinear systems as an LPV system including the one where nonlinearities in the systems can be treated as the scheduling variables [1]. This nonlinear embedding approach can be represented as:

$$A(\theta(t))x + B(\theta(t))u = f(x, u), \quad \theta(t) \in S \quad (2)$$

where $f(x, u)$ is the nonlinear model. From Eq. 2 it can be concluded that the LPV representation has the equivalent variation as the nonlinear model in the scheduling region S . One way of looking at LPV system is as the extension of linear-time-invariant (LTI) systems since both have similar properties when the parameters in LPV systems are frozen at some operating point for a given time. The representation of LPV model, when frozen at some given operating points can be viewed as shown in Fig. 1.

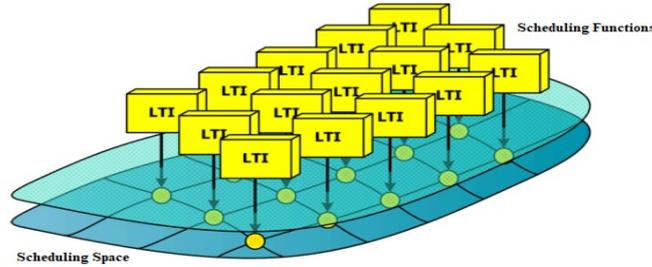


Figure 1. Representation of LPV systems [2].

However, a number of variables in LPV model have significant impact in conservatism, over-bounding and increasing complexity in scheduling regions [3]. For polytopic LPV representation, computational complexity in controller synthesis is increased exponentially with increase in a number of variables, i.e. with 2^n where n is number of scheduling variables. Therefore, a common problem to solve in LPV controller synthesis is the reduction of scheduling variables [4]-[5]. This reduction of scheduling variables can be viewed, in the LPV context, a form

of model reduction. Hence, in LPV sense, model reduction can be thought of as either the reduction of model order or the reduction of scheduling variables or both. Both of these are strongly related from a complexity point of view [5]. In this paper, reduction of a number of variables in LPV models is addressed.

Complexity reduction posed by a number of scheduling variables in LPV controller synthesis of LPV systems is addressed in [3]. Linear principal-component-analysis (PCA) is applied for the reduction of a number of scheduling variables in LPV system. The paper discusses the parameter reduction for a 2-link robot-manipulator which is a sufficiently simple and slower system with only ten scheduling variables. Kernel-PCA is proposed for the same system in [6]. The kernel-PCA was able to capture more dynamics for the same system using the same number of reduced parameters and thus resulting in a faster controller using the approximate LPV model with more accurate tracking compared to that in [3]. Linear-PCA is also applied for LPV model order reduction of selective-catalytic-reduction (SCR) in [7]. SCR is extremely slow industrial process and thus one can not conclude that PCA can be used for the LPV model order reduction of faster and extremely nonlinear systems.

From the literature, several kernel functions can be chosen for LPV model reduction. Choosing a suitable kernel function however still remains an open problem [8]. Testing all, or several of the available kernel functions is obviously a time-consuming task. Neither are there guidelines available from which one can have an idea as to which kernel is likely to do the job. In this paper, an attempt is made to provide answer to this question. It has been shown through simulation experiments, on two coupled nonlinear multivariable systems, that polynomial kernel is the kernel of choice to start with for LPV model reduction.

The main contributions of this paper include the comparative analysis of kernel-PCA based approximated LPV models obtained through fifteen kernel functions for highly coupled, nonlinear, and faster systems. The paper also uses two statistical tools, including error to norm ratio and best fit ratio for the comparison of approximated LPV and full-order LPV model.

The rest of the paper is organized as: section 2 briefly discusses the application of kernel-PCA LPV model order reduction. Section 3 is about the brief introduction and quasi-LPV modeling of the nonlinear systems used for the model approximation and comparison of reduced LPV models obtained through different kernel functions. A detailed discussion on results is made in section 4. Finally, conclusions are drawn in section 5.

§2 LPV model approximation using kernel-based PCA

To perform model order reduction, one needs to collect data from simulations or measurements. Consider the LPV model in Eq. 3:

$$\begin{aligned}\dot{x}(t) &= A(\rho(t))x(t) + B(\rho(t))u(t) \\ y(t) &= C(\rho(t))x(t) + D(\rho(t))u(t)\end{aligned}\quad (3)$$

where, $x \in \mathfrak{R}^n$, $u \in \mathfrak{R}^{n_u}$, $y \in \mathfrak{R}^{n_y}$ are state matrix, input matrix and output matrix respectively. $A(\rho(t))$, $B(\rho(t))$, $C(\rho(t))$ and $D(\rho(t))$ are continuous functions of scheduling signals. From here

onwards, $\theta(t)$ and $\rho(t)$ will be denoted as θ_t and ρ_t respectively. The scheduling parameters are computed as:

$$\theta(\rho_t) = [\theta_1 \quad \dots \quad \theta_N] = [\rho(p_1) \quad \dots \quad \rho(p_N)] \in \mathfrak{R}^{n \times N} \quad (4)$$

with N number of samples such that $N \geq n$ and n being total number of scheduling parameters and p represents the nonlinear function, known as scheduling function, to relate scheduling signals with scheduling variables. Because of high dimension of feature space, separation of the data can be easily achieved. The effectiveness of this method is due to famous kernel-trick, which allows to perform linear operations in feature-space without mapping the scheduling variables in feature-space. Kernel PCA is used for dimensionality reduction of scheduling variables in LPV modeling. Assuming that scheduling variables mapped in features pace with $\psi(\theta_1) \dots \psi(\theta_n)$, with n being number of parameters, we are assuming that parameters in feature-space are centered i.e. $\sum_{i=1}^N \psi(\theta_i) = 0$. For performing normal PCA, one needs to find the covariance matrix given by:

$$\bar{C} = \frac{1}{N} \sum_{i=1}^N \psi(\theta_i) \psi^T(\theta_i) \quad (5)$$

To extract principal components, $\lambda v = C\bar{v}$ should be satisfied. Hence, we may deduce:

$$\lambda(v \cdot \psi(\theta)) = C\bar{v} \cdot \psi(\theta_j), \quad \forall j = 1 \dots N \quad (6)$$

and $x.y = x^T y$ is the dot product. For details of the procedure, reader is referred to [6]. It is very important to note that k-PCA does not necessary require the computation of mapping in feature-space, what it does require is the characterization of dot product in feature-space. This characterization can be obtained by defining a kernel function. Now, Gram kernel matrix, $K \in \mathfrak{R}^{N \times N}$ is defined as:

$$K_{ij} = (\psi(\theta_i) \cdot \psi(\theta_j)) \quad (7)$$

here, $k(.,.)$ represents nonlinear kernel function. Finally, to extract the principal components, one needs to find the projections of image of some test point θ_t on eign vector v_{red} in feature-space as:

$$\phi_{red,t} = v_{red} \psi(\theta) = \sum_{j=1}^N \alpha_{red,j} k(\theta_j, p(\rho_t)) \quad (8)$$

In Eq. 8 ϕ_t is projection of $\psi(\theta_t)$ on v_{red} and $\phi_{red,t}$ is r^{th} entry of ϕ_t . Since it is assumed that the data is centered in feature-space, which is unknown for certain due to unavailability of featured-space. So, the centered kernel matrix can be obtained by replacing $\psi(\theta_i)$ by $\hat{\psi}(\theta_i) = \psi(\theta_i) - \frac{1}{n} \sum_{i=1}^n \psi(\theta_i)$. Here, new centered matrix is reproduced as [9]:

$$\hat{K} = K - 1_N K - K 1_N + 1_N K 1_N \quad (9)$$

here, $1_N \in \mathfrak{R}^{n \times n}$ with each element in 1_N being $\frac{1}{N}$. Fraction of total variation is used to calculate the accuracy of reduced model. It is defined as:

$$V(m) = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^m \lambda_i} \quad (10)$$

with m being number of the reduced variables and λ_i the i^{th} eign value of \hat{K} . Using Eq. 10,

one can choose m based on the significant eigen-values. \bar{m} represents total number of scheduling variables which is taken to be n . When compared to simple or linear PCA, kernel-PCA has a limitation that the reduced LPV model approximated through kernel-PCA has no longer affine dependence on new scheduling variable ϕ_t [3]. In order to identify the reduced LPV model with affine dependence on ϕ_t , one needs to minimize the square of Frobenius norm of error between original state-space (the one having n scheduling variables) and the new state-space with reduced (m) number of variables. That is the extra computational cost of kernel-PCA involved in model approximation over that of using PCA. The mathematical formulation of optimization problem is given by:

$$\begin{aligned} \text{Min} \quad & \frac{1}{N} \sum_{i=1}^N \|S(\theta_t) - Q_r(\phi_t)\|_F^2 \\ \text{s.t.} \quad & Q_r(\phi_t) = Q(r_o) + \sum_{i=1}^m Q_{r,i} \phi_{t,i} = \begin{bmatrix} \hat{A}(\phi_t) & \hat{B}(\phi_t) \\ \hat{C}(\phi_t) & \hat{D}(\phi_t) \end{bmatrix} \end{aligned} \quad (11)$$

here, $\|\cdot\|_F$ represents Frobenius norm, N is total number of samples and m is reduced number of scheduling variables. While $S(\theta_t)$ is given by:

$$S(\theta) = \begin{bmatrix} A(\theta_t) & B(\theta_t) \\ C(\theta_t) & D(\theta_t) \end{bmatrix} \quad (12)$$

§3 Kernel Functions

Fifteen kernel function are used in this work for LPV model approximation of nonlinear systems using reduced numbers of scheduling variables (θ_t). The kernel functions used in this work are discussed in subsequent sections.

3.1 Polynomial Kernel Function

Polynomial kernel implementation function can given as [10]

$$k(x, y) = (x^T y)^d \quad (13)$$

It is a non-stationary kernel. It should be applied to the normalized training data. It has a single adjustable parameter d where $d \in \mathfrak{R}$, x and y are the vectors. This parameter can be tuned for desired performance.

3.2 Affine Kernel Function

Affine kernel implementation function is given as [11]

$$k(x, y) = (x^T y + c)^d \quad (14)$$

It is a non-stationary kernel. It should be applied to the normalized training data. It has two adjustable parameters c and d . These parameters are tuned or adjusted according to desired specifications.

3.3 Sigmoid Kernel Function

Sigmoid kernel implementation function can be written as [10]

$$k(x, y) = \tanh(kx^T y + b) \quad (15)$$

It is also known as hyperbolic tangent function or Multilayer Perceptron (MLP) kernel. It comes from Artificial Neural Network (ANN) field. In ANN field bipolar sigmoid function is often used as an activation function of artificial neuron. As its origin is from neural network field so it is quite popular in Support Vector Machine (SVM) [12]. Surprisingly SVM model with sigmoid activation function is equivalent to two-layer perceptron. Although it is positive definite, but it performs good in actual. Sigmoid kernel has two adjustable parameters, k and b . Usually value of k is taken $1/N$ where N is dimension of data.

3.4 Gaussian Kernel Function

Gaussian kernel is implemented using following equation [13]

$$k(x, y) = \exp\left(\frac{-\|x - y\|^2}{(2\sigma^2)}\right) \quad (16)$$

It is an example of radial basis function kernel [11]. Alternatively, it could be implemented using following function

$$k(x, y) = \exp(-\gamma\|x - y\|^2) \quad (17)$$

In the first equation σ is an adjustable parameter. It plays a crucial role in kernel performance. It should be properly tuned for problem at hand. If a high value of σ is chosen, then the exponential will start to behave linearly, and nonlinear power of its higher dimensional projection will start to deteriorate. A small value of σ on the other hand would let sensitivities of decision boundary propagate and function will start losing data.

3.5 Laplacian Kernel Function

Laplacian kernel can be implemented using the following equation

$$k(x, y) = \exp\left(\frac{-\|x - y\|}{\sigma}\right) \quad (18)$$

It can be termed as a type of radial basis function kernel. It is equivalent to exponential kernel, but it is less sensitive to the variation of σ parameter. The consequences of changing sigma parameter to the Gaussian kernel also can be held for the Laplacian kernel [13].

3.6 Log Kernel Function

Log kernel can be represented by the following equation

$$k(x, y) = -\log(\|x - y\|^d + 1) \quad (19)$$

It is conditionally positive definite and well suited for images.

3.7 Cosine Kernel Function

The cosine kernel function can be written as:

$$k(x, y) = \frac{(x^T y)}{\|x\| \|y\|} \quad (20)$$

here, $\|\cdot\|$ represents the L_2 norm.

3.8 ANOVA Kernel Function

The function that expresses ANOVA kernel can be written as [14]

$$k(x, y) = \sum_{k=1}^n \exp(-\sigma(x^k - y^k)^2)^d \quad (21)$$

It works well in regression problems with multiple dimensions. It is also a type of radial basis function kernel just like gaussian, Laplacian, and exponential kernels.

3.9 Multiquadric Kernel Function

The expression defining multiquadric kernel is given as [15]

$$k(x, y) = \sqrt{\|x - y\|^2 + c^2} \quad (22)$$

It is a well-known example of non-positive definite kernel. It can be used in same circumstances as rational quadratic kernel.

3.10 Inverse Multiquadric Kernel Function

It is represented by the following equation [15]

$$k(x, y) = \frac{1}{\sqrt{\|x - y\|^2 + c^2}} \quad (23)$$

It results in an infinite dimensional feature space just like the Gaussian kernel because it results into matrix with full rank.

3.11 Power Kernel Function

The equation describing power kernel is given as

$$k(x, y) = -\|x - y\|^d \quad (24)$$

It is a triangular kernel. It can also be called as scale-invariant kernel [16]. It is only conditionally positive definite.

3.12 Rational Quadratic Kernel Function

The defining equation of rational quadratic kernel is

$$k(x, y) = 1 - \frac{\|x - y\|^2}{(\|x - y\|^2 + c)} \quad (25)$$

It is less computationally intensive and can be used in place of Gaussian kernel when Gaussian kernel become too massive.

3.13 Exponential Kernel Function

It is described by equation

$$k(x, y) = \exp\left(\frac{-\|x - y\|}{2\sigma^2}\right) \quad (26)$$

It is closely related to the Gaussian kernel where only power of norm is removed. Like Gaussian kernel it is also a type of radial basis function kernel [13].

3.14 Spherical Kernel Function

It can be described by the following equation

$$k(x, y) = 1 - \frac{3\|x - y\|}{2\sigma} + \frac{1}{2}\left(\frac{\|x - y\|}{\sigma}\right)^3 \quad (27)$$

$$if\|x - y\| < \sigma \text{ zerootherwise} \quad (28)$$

It is similar to the circular kernel function but is positive definite in R^3 .

3.15 Cauchy Kernel Function

It is described by following formula [15]:

$$k(x, y) = \frac{1}{(1 + \frac{\|x - y\|^2}{\sigma^2})} \quad (29)$$

Over the high dimension space, it can be used to give sensitivity and long-range influence. Basically, it comes from Cauchy distribution and it is long tailed kernel.

§4 Nonlinear systems consideration for model approximation

In this work, two highly coupled nonlinear systems are used for the verification of kPCA based LPV model approximation using fifteen different kernel functions. The nonlinear systems considered as case studies in this work include 4-DOF control moment gyroscope (CMG) and twin-rotor aerodynamic systems. Both of these systems are highly coupled nonlinear and naturally unstable. Considering the complexities posed by the dynamics of sated systems, the effectiveness kernel-based PCA for LPV model approximation can be accurately validated. The quasi-LPV modeling of both systems is briefly discussed in following sections.

4.1 4-DOF Control Moment Gyroscope (CMG)

CMG has four bodies named as; A, B, C and D providing the 4 angular degrees of freedom. Coordinate frames of a 4-DOF CMG is shown in Fig 2. Bodies B, C and D are referred to as gimbals while, A is the rotor in the gimbal B . Two DC motors provide torques τ_1 and τ_2 for spinning the rotor and rotating the gimbal B . Gimbals C and gimbal D rotate about the axes

3 and axes 4 [17]. No active torque is applied to gimbal C and D . q_1 is the angular position of body A , q_2 represents relative angle between gimbal B and gimbal C , relative angle between gimbal C and gimbal D is q_3 and q_4 is the relative angle between the gimbal D and inertial frame of reference. $\omega_1, \omega_2, \omega_3$ and ω_4 denote the angular speeds of rotor, gimbal B , gimbal C and gimbal D respectively.



Figure 2. ECP Model 750 - The Control Moment Gyroscope [18].

All the angular positions are measured using optical encoders. Angular momentum of fixed magnitude is generated through the spin of rotor at a constant speed. By applying torque τ_2 on gimbal B , the orientation of rotor is changed which redirects the angular momentum of the rotor. This change in direction of angular momentum results in generating gyroscopic torque which is used to rotate body D [19]. The quasi-LPV model based on Eq. 1 can be designed for CMG. The state and input matrices for the CMG can be represented as:

$$A(\rho) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \theta_1(\rho) & \theta_2(\rho) & \theta_3(\rho) \\ 0 & 0 & \theta_4(\rho) & \theta_5(\rho) & \theta_6(\rho) \\ 0 & 0 & \theta_7(\rho) & \theta_8(\rho) & \theta_9(\rho) \end{bmatrix}$$

$$B(\rho) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \theta_{10}(\rho) & \theta_{11}(\rho) \\ \theta_{12}(\rho) & \theta_{13}(\rho) \\ \theta_{14}(\rho) & \theta_{15}(\rho) \end{bmatrix} \quad (30)$$

The inputs, outputs, scheduling signal and state of the designed quasi-LPV model are given as:

$$u = \begin{bmatrix} \tau_1 & \tau_2 \end{bmatrix}^T, y = \begin{bmatrix} q_3 & q_4 \end{bmatrix}^T, \rho = \begin{bmatrix} q_2 & q_3 \end{bmatrix}^T$$

$$x = \begin{bmatrix} q_3 & q_4 & \omega_1 & \omega_2 & \omega_3 & \omega_4 \end{bmatrix}^T \quad (31)$$

here, $(\theta_1(\rho) \dots \theta_{15}(\rho))$ are scheduling parameters and function of scheduling signal, $\rho_t \in \{q_2, q_3\}$,

which can be measured in real time or simulation. In this work, ρ_t are obtained through simulations. It can be seen from Eq. 30 that the quasi-LPV model of 4-DOF CMG consists of fifteen scheduling variables ($\theta_1(\rho) \dots \theta_{15}(\rho)$) requiring a massive computational effort, as one needs to solve 2^{15} linear-matrix-inequalities (LMIs) to design the LPV controller. Same scheduling variables are used in [21].

4.2 Twin-rotor Aerodynamic System (TRAS)

TRAS is a multi-input and multi output laboratory system. It is an approximate model of helicopter. It can perform motion in two directions, lateral and longitudinal. It consists of a beam fixed at its center on vertical rod. Two propellers are mounted on it, a main propeller and tail propeller. Main propeller is responsible for pitch motion and tail propeller is responsible for azimuth motion. TRAS is shown in Fig. 3.



Figure 3. Twin-rotor Aerodynamic (TRAS) system [20].

In the control study of TRAS, mainly the interest lies in controlling azimuth and pitch angles by calculating controlled input for main and tail propellers. Like CMG, TRAS is also highly coupled nonlinear and unstable systems. Nonlinear model of TRAS is given in following equations [22].

$$J_\psi \ddot{\psi} = u_\psi + J\dot{\psi}\dot{\phi} \sin(2\phi) - K_\psi \psi - C_\psi \dot{\psi} \quad (32)$$

$$J_\phi \ddot{\phi} = u_\phi - J \frac{\dot{\psi}^2}{2} \sin(2\phi) - C_\phi \dot{\phi} - K_g g \quad (33)$$

where,

$$J_\psi = (m_m l_m^2 + m_t l_t^2) \cos^2(\phi) + 2m_{c\omega} l_{c\omega} \sin^2(\phi) + J_z$$

$$J_\phi = m_m l_m^2 + m_t l_t^2 + 2m_{c\omega} l_{c\omega} + J_x$$

$$J = m_m l_m^2 + m_t l_t^2 - 2m_{c\omega} l_{c\omega}$$

$$K_g = (m_m l_m - m_t l_t) \cos(\phi) + 2m_{c\omega} l_{c\omega} \sin(\phi)$$

The distance of tail and main rotors from the origin is denoted by l_t and l_m respectively. Whereas m_m and m_t represent the counterweights at the main and tail of the beam. Moreover,

$m_{c\omega}$ and $l_{c\omega}$ represent the masses at the ends of levers and the relevant lengths of the levers respectively. Jacobean linearization is applied to the nonlinear model to linearize it around moving operating points to obtain quasi-LPV representation. Parameter values can be found in [22]. TRAS model consists of four states azimuth angle (ϕ), azimuth velocity ($\dot{\psi}$), pitch angle ($\dot{\phi}$) and pitch velocity ($\dot{\psi}$). Performing jacobian linearization of nonlinear model, the quasi-LPV model of the TRAS is given by:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \theta_1(\rho) & \theta_2(\rho) & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \theta_3(\rho) & \theta_4(\rho) & -0.4233 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ \theta_5(\rho) & 0 \\ 0 & 0 \\ 0 & 6.73 \end{bmatrix} \quad (34)$$

where $x \in Re^{n \times 1}$, $A \in \mathfrak{R}^{n \times n}$, $B \in \mathfrak{R}^{n \times p}$, $\theta(t) \in \{\phi, \psi\}$. The state vector for TRAS is given as:

$$x = [\phi \quad \dot{\phi} \quad \psi \quad \dot{\psi}]^T \quad (35)$$

The quasi-LPV model of the twin-rotor designed in this paper consists of five scheduling variables required to solve 2^5 LMIs in LPV polytopic control design process. The quasi-LPV system is comparatively simpler than that of CMG as it includes one third number of scheduling variables compared to CMG. The scheduling variables of the TRAS are given as follows.

$$\begin{aligned} \theta_1(\rho) &= \frac{-0.005889}{0.0238 \cos^2(\psi) + 0.00301} \\ \theta_2(\rho) &= \frac{(0.0238 \cos^2(\psi) + 0.00301)(-0.216 \sin(\psi)) - (-0.216 \cos(\psi))}{(0.0238 \cos^2(\psi) + 0.00301)^2} \\ &\quad - \frac{(0.005889)(0.0476 \sin(\psi) \cos(\psi))}{(0.0238 \cos^2(\psi) + 0.00301)^2} \dot{\phi} \\ \theta_3(\rho) &= \frac{-0.021248}{0.03} \sin(\psi) \cos(\psi) \times 2\dot{\phi} \\ \theta_4(\rho) &= \frac{-1.7143}{0.03} (-0.0292 \sin(\psi) + 0.054 \cos(\psi)) \\ &\quad - \frac{0.021248}{0.03} \dot{\phi}^2 (\cos^2(\psi) \sin^2(\psi)) \\ \theta_5(\rho) &= \frac{0.216 \cos(\psi)}{0.0238 \cos^2(\psi) + 0.00301} \end{aligned}$$

§5 Results and discussion

Fifteen kernel functions of the support vector machines, discussed in section 3, are chosen for the LPV model approximation of above mentioned highly coupled nonlinear systems. The approximated LPV model needs to be compared with the full order LPV model. For that purpose, two statistical tools, i.e. error to norm ratio and best fit ratio are chosen. Error to norm ratio can be defined by the following equation.

$$E_{nx} = \frac{\|x - \hat{x}\|_2}{\|x\|_2} \quad (36)$$

$\|\cdot\|_2$ represents the L_2 norm, x and \hat{x} denote the states of full-order and approximated LPV models of the system. The smaller is the value of E_{nx} , the better is performance of the ap-

proximated LPV model. Ideally the value of E_{nx} is zero i.e. when the approximated model is same as the original one. The significance of error to norm ratio lies in its ability to provide comparison of each state of the approximated LPV model with the corresponding state of the full-order LPV model. A comparison of complete approximate LPV model with the full-order LPV model is also made using best-fit-ratio (BFR). Best fit ratio is a statistical tool used to check the performance of the approximate LPV model. The best fit ratio can be defined as [6]:

$$\%BFR = 100\% \times \max\left\{\frac{\|x - \hat{x}\|_2}{\|x - \bar{x}\|_2}, 0\right\} \quad (37)$$

$\|\cdot\|_2$ shows the L_2 norm, x represents the states of the original model, \hat{x} represents the states of the approximated model and \bar{x} is the mean of the states of the original model. BFR is taken in percentage to see how much percent the dynamics of the approximated model match with the original one. It should be noted that the approximation errors, and the criteria based on which the approximations are good and poor is given in Eq. 36 (error to norm ratio) and Eq. 37 (percent best fit ratio). The model approximation is good if the error to norm ratio is minimum and percent best fit ratio is maximum, i.e. greater than 90% while the approximation will be poor if vice versa and hence the approximation error would be large. The error to norm ratio criterion is applied for validation of accuracy of approximation of each state while the percent best fit ratio is applied to the reduced model to validate the accuracy of complete model.

The results obtain through the application of kernel-PCA based LPV model approximation for 4-DOF CMG and TRAS are discussed as follows.

5.1 LPV Model Approximation for 4-DOF CMG

The scheduling signals generated through the simulation of nonlinear CMG model are used to create the scheduling variables θ_t . The kernel PCA-based LPV model approximation discussed in section 2 is applied on scheduling variables. Based on fraction of total variation V_m , the number of scheduling variables for reduced order LPV model are selected. The fraction of total variation for all fifteen kernel functions applied on CMG is shown in Table 1. The table shows V_m for fifteen kernel functions when used for the model approximation by using up to 7 scheduling variables. The V_m up to 15 variables is not shown due to limitation of the space. It can be seen from the table that polynomial and affine kernel functions offer more flexibility in model approximation using lesser number of scheduling variables when compared with any other function. Also, one can note from the table, V_m using lesser number of scheduling variables is increased for both affine and polynomial kernel functions by increasing their degree d . Increasing the tuning parameter σ for gaussian and exponential kernel functions also causes increase in V_m using lesser number of scheduling variables. Both, gaussian and exponential kernel functions have same V_m . Similarly, Laplacian and log functions have same V_m . Approximate model for CMG is then identified using one, two and three number of scheduling variables for each function. Table 2 shows the performance of approximated LPV models of CMG with full-order LPV model.

Table 1. Fraction of total variation for CMG.

| Kernel Function | | V_m | | | | | | |
|------------------------|---------------|---------|---------|---------|---------|---------|---------|---------|
| | | $m = 1$ | $m = 2$ | $m = 3$ | $m = 4$ | $m = 5$ | $m = 6$ | $m = 7$ |
| Polynomial & Affine | d=1 | 0.45 | 0.68 | 0.78 | 0.85 | 0.91 | 0.94 | 0.97 |
| | d=3 | 0.55 | 0.76 | 0.85 | 0.92 | 0.96 | 0.98 | 0.99 |
| | d=5 | 0.63 | 0.81 | 0.89 | 0.95 | 0.98 | 0.99 | 1.00 |
| | d=6 | 0.68 | 0.86 | 0.92 | 0.96 | 0.99 | 1.00 | 1.00 |
| | d=7 | 0.71 | 0.88 | 0.94 | 0.97 | 0.99 | 1.00 | 1.00 |
| | d=8 | 0.75 | 0.92 | 0.97 | 0.98 | 0.99 | 1.00 | 1.00 |
| Sigmoid | k=1, b=1 | 0.21 | 0.35 | 0.46 | 0.54 | 0.62 | 0.69 | 0.74 |
| | k=1, b=2 | 0.34 | 0.52 | 0.67 | 0.75 | 0.82 | 0.86 | 0.91 |
| | k=1, b=3 | 0.38 | 0.58 | 0.72 | 0.80 | 0.87 | 0.91 | 0.94 |
| | k=2, b=3 | 0.45 | 0.64 | 0.78 | 0.86 | 0.91 | 0.94 | 0.96 |
| Gaussian & Expo. | $\sigma = 5$ | 0.25 | 0.32 | 0.47 | 0.62 | 0.69 | 0.72 | 0.74 |
| | $\sigma = 10$ | 0.41 | 0.57 | 0.69 | 0.80 | 0.83 | 0.85 | 0.89 |
| | $\sigma = 5$ | 0.75 | 0.92 | 0.97 | 0.98 | 0.99 | 1.00 | 1.00 |
| Laplacian & Log | | 0.13 | 0.23 | 0.32 | 0.41 | 0.48 | 0.54 | 0.59 |
| Cosine | | 0.50 | 0.53 | 0.61 | 0.63 | 0.77 | 0.79 | 0.84 |
| ANOVA | | 0.35 | 0.43 | 0.48 | 0.52 | 0.61 | 0.72 | 0.77 |
| MQ | | 0.17 | 0.19 | 0.28 | 0.29 | 0.34 | 0.43 | 0.56 |
| IMQ | | 0.34 | 0.43 | 0.49 | 0.58 | 0.69 | 0.78 | 0.82 |
| Power | | 0.51 | 0.54 | 0.59 | 0.62 | 0.64 | 0.68 | 0.71 |
| RQ | | 0.23 | 0.28 | 0.32 | 0.38 | 0.46 | 0.49 | 0.55 |
| Spherical | | 0.11 | 0.19 | 0.22 | 0.28 | 0.34 | 0.39 | 0.45 |

Table 2. Comparison of original and approximated models of CMG.

| Kernel Function | | E_{nx} | | | | | | |
|--------------------|-----|----------|-------|-------|------------|------------|------------|------|
| | | q_2 | q_3 | q_4 | ω_2 | ω_3 | ω_4 | %BFR |
| Polynomial | m=1 | 0.012 | 0.009 | 0.039 | 0.219 | 0.022 | 0.913 | 91.2 |
| | m=2 | 0.003 | 0.002 | 0.014 | 0.142 | 0.001 | 0.282 | 92.3 |
| | m=3 | 0.004 | 0.002 | 0.005 | 0.062 | 0.009 | 0.030 | 95.1 |
| Affine | m=1 | 1.151 | 1.670 | 0.985 | 1.409 | 2.471 | 0.957 | 83.4 |
| | m=2 | 1.027 | 1.670 | 0.932 | 1.501 | 2.273 | 0.901 | 87.0 |
| | m=3 | 0.241 | 1.070 | 0.725 | 1.106 | 1.813 | 0.724 | 91.0 |

Continued on next page

Table 2 – Continued from previous page

| Kernel | | q_2 | q_3 | q_4 | ω_2 | ω_3 | ω_4 | %BFR |
|-------------|-----|--------|--------|-------|------------|------------|------------|------|
| Sigmoid | m=1 | 1.823 | 2.429 | 0.845 | 2.955 | 3.865 | 0.646 | 53.1 |
| | m=2 | 1.518 | 1.955 | 0.839 | 2.010 | 2.789 | 0.636 | 60.2 |
| | m=3 | 1.338 | 1.678 | 0.858 | 1.649 | 2.215 | 0.674 | 90.0 |
| Gaussian | m=1 | 1.259 | 1.548 | 0.855 | 1.413 | 2.035 | 0.686 | 54.4 |
| | m=2 | 1.183 | 1.377 | 0.848 | 1.240 | 1.739 | 0.682 | 62.1 |
| | m=3 | 1.135 | 1.265 | 0.836 | 1.156 | 1.492 | 0.658 | 83.5 |
| Laplacian | m=1 | 1.293 | 1.622 | 0.848 | 1.481 | 2.111 | 0.668 | 64.1 |
| | m=2 | 1.284 | 1.597 | 0.854 | 1.478 | 2.099 | 0.679 | 72.5 |
| | m=3 | 1.279 | 1.584 | 0.857 | 1.476 | 2.093 | 0.685 | 88.1 |
| Log | m=1 | 10.343 | 29.606 | 0.947 | 22.158 | 57.623 | 0.879 | 53.1 |
| | m=2 | 10.343 | 1.066 | 0.947 | 22.158 | 1.107 | 0.879 | 60.2 |
| | m=3 | 1.441 | 1.659 | 0.947 | 1.711 | 2.196 | 0.879 | 90.0 |
| Cosine | m=1 | 1.234 | 1.498 | 0.853 | 1.349 | 1.952 | 0.687 | 47.4 |
| | m=2 | 1.208 | 1.439 | 0.851 | 1.288 | 1.856 | 0.686 | 52.3 |
| | m=3 | 1.134 | 1.265 | 0.836 | 1.156 | 1.492 | 0.658 | 73.6 |
| ANOVA | m=1 | 3.407 | 6.488 | 0.829 | 6.025 | 10.021 | 0.696 | 57.4 |
| | m=2 | 3.346 | 6.316 | 0.835 | 6.052 | 9.834 | 0.704 | 72.1 |
| | m=3 | 3.347 | 6.316 | 0.835 | 6.051 | 9.833 | 0.704 | 81.3 |
| MQ | m=1 | 1.092 | 1.691 | 0.990 | 1.284 | 2.498 | 0.971 | 47.1 |
| | m=2 | 1.064 | 1.483 | 0.990 | 1.149 | 2.028 | 0.971 | 52.8 |
| | m=3 | 1.019 | 2.430 | 0.990 | 1.009 | 4.156 | 0.971 | 73.9 |
| IMQ | m=1 | 6.684 | 7.953 | 0.990 | 47.729 | 17.091 | 0.971 | 67.1 |
| | m=2 | 6.683 | 2.386 | 0.991 | 47.792 | 4.097 | 0.971 | 77.1 |
| | m=3 | 1.558 | 7.946 | 0.990 | 4.898 | 17.087 | 0.971 | 89.1 |
| Power | m=1 | 13.092 | 1.261 | 0.979 | 136.295 | 1.582 | 0.941 | 69.9 |
| | m=2 | 5.180 | 1.013 | 0.979 | 47.789 | 1.026 | 0.941 | 81.0 |
| | m=3 | 1.185 | 1.131 | 0.979 | 2.586 | 1.280 | 0.941 | 89.9 |
| RQ | m=1 | 0.979 | 1.989 | 1.856 | 1.508 | 3.388 | 3.456 | 56.2 |
| | m=2 | 0.950 | 2.902 | 1.622 | 3.379 | 6.021 | 3.053 | 72.3 |
| | m=3 | 0.952 | 2.871 | 1.611 | 3.373 | 5.918 | 3.056 | 87.5 |
| Exponential | m=1 | 1.259 | 1.548 | 0.855 | 1.413 | 2.035 | 0.686 | 54.4 |
| | m=2 | 1.183 | 1.377 | 0.848 | 1.240 | 1.740 | 0.682 | 62.1 |
| | m=3 | 1.134 | 1.265 | 0.836 | 1.156 | 1.492 | 0.658 | 93.5 |
| | m=1 | 1.234 | 1.498 | 0.853 | 1.349 | 1.952 | 0.687 | 68.9 |

Continued on next page

Table 2 – Continued from previous page

| Kernel | | q_2 | q_3 | q_4 | ω_2 | ω_3 | ω_4 | %BFR |
|-----------|-----|--------|--------|-------|------------|------------|------------|------|
| Spherical | m=2 | 1.208 | 1.439 | 0.851 | 1.288 | 1.856 | 0.686 | 69.6 |
| | m=3 | 1.134 | 1.265 | 0.836 | 1.156 | 1.492 | 0.658 | 79.3 |
| Cauchy | m=1 | 26.684 | 12.252 | 3.991 | 47.029 | 27.091 | 2.372 | 69.0 |
| | m=2 | 8.683 | 2.386 | 1.250 | 57.792 | 24.041 | 1.974 | 76.0 |
| | m=3 | 3.041 | 0.946 | 1.091 | 34.898 | 17.088 | 1.902 | 78.4 |

One can conclude from Table 2 that polynomial kernel function provides the best approximated LPV model for CMG. It should be noted that MQ, IMQ and RQ in above Tables 1-2 denote multiquadratic, inverse multiquadratic, and rational quadratic kernel functions respectively. Error to norm ratio (E_{nx}) is a more relevant tool as it can provide the comparison of each state of the approximated model with the full-order LPV model. If the approximated model is not close enough to the original LPV model based on best fit ratio, E_{nx} can provide the insight by comparison of the states. The comparison of the best approximated LPV and full-order LPV models of the CMG are shown in Fig. 4 and Fig. 5. It can be viewed graphically from Fig.

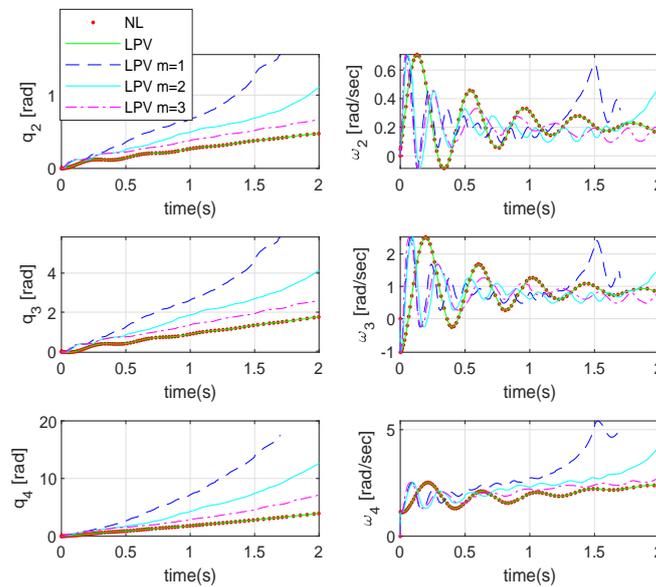


Figure 4. Approximated and original LPV model of CMG using polynomial kernel function.

4 that approximated model using three scheduling variables is more realistic as its response to excitation is very close to the full-order LPV and nonlinear models of the CMG. Here, m in the

Table 3. Fraction of total variation for TRAS.

| Kernel Function | V_m | | | | |
|------------------------|---------|---------|---------|---------|---------|
| | $m = 1$ | $m = 2$ | $m = 3$ | $m = 4$ | $m = 5$ |
| Polynomial& Affine | 0.94 | 0.96 | 0.97 | 1.00 | 1.00 |
| Sigmoid | 0.89 | 0.90 | 0.94 | 0.99 | 1.00 |
| Gaussian & Exponential | 0.90 | 0.91 | 0.95 | 0.98 | 1.00 |
| Laplacian& Log | 0.81 | 0.83 | 0.87 | 0.90 | 0.99 |
| Multiquadratic | 0.79 | 0.81 | 0.83 | 0.90 | 0.98 |
| Inverse Multiquadric | 0.74 | 0.79 | 0.88 | 0.92 | 0.99 |
| Power | 0.70 | 0.77 | 0.84 | 0.91 | 1.00 |
| Rational Quadratic | 0.69 | 0.72 | 0.78 | 0.85 | 0.95 |
| Spherical | 0.71 | 0.79 | 0.88 | 0.95 | 0.99 |
| Cauchy | 0.86 | 0.89 | 0.95 | 0.98 | 1.00 |

figure represents the number of reduced scheduling variables. The comparison of original and approximated models using affine kernel function is shown in Fig. 5. From Fig. 4 and Fig. 5 it is clear that the approximate LPV model using polynomial kernel function has more accurate response. The same can be concluded from Table 2.

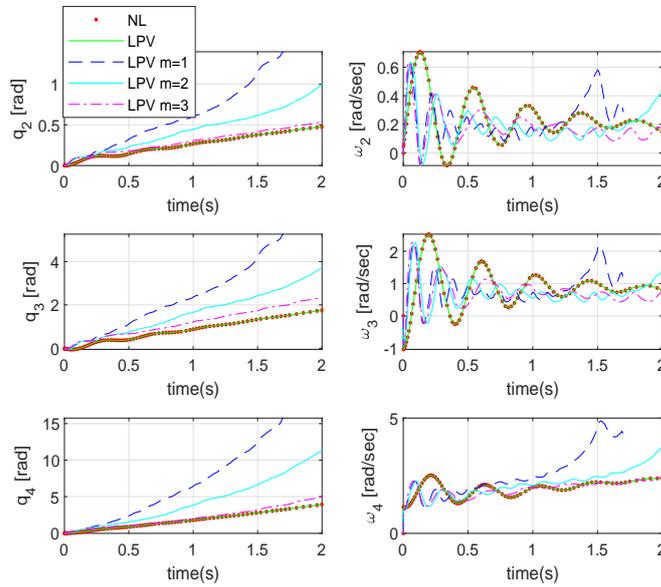


Figure 5. Approximated and original LPV model of CMG using affine kernel function.

5.2 LPV Model Approximation for TRAS

TRAS has lesser complexity in terms of number of scheduling variables as compared to CMG. TRAS has 5 scheduling variables causing one to solve 2^5 LMIs during LPV controller design process. We applied the same fifteen kernel functions and compared the approximate LPV models obtained through each kernel. The fraction of total variation to select the number of reduced variables for model approximation is shown in Table 3.

From Table 3, one can conclude that affine and polynomial kernel functions have highest V_m using single scheduling variable. The comparison of approximate and original LPV models is shown in Table 4.

Table 4. Comparison of original and approximated models of TRAS.

| Kernel Function | E_{nx} | | | | %BFR |
|----------------------|----------|--------|--------------|--------------|-------|
| | ϕ | ψ | $\dot{\phi}$ | $\dot{\psi}$ | |
| Polynomial | 0.971 | 0.0721 | 0.000 | 0.000 | 96.94 |
| Affine | 31.23 | 13.514 | 0.000 | 0.000 | 97.19 |
| Sigmoid | 1.825 | 2.0242 | 0.000 | 0.000 | 94.18 |
| Gaussian | 7.225 | 8.1435 | 2.460 | 51.254 | 83.06 |
| Laplacian | 7.225 | 8.1435 | 2.460 | 3.254 | 91.09 |
| Log | 27.55 | 98.93 | 2.460 | 25.150 | 83.91 |
| Cosine | 1.825 | 2.024 | 0.209 | 0.837 | 89.33 |
| ANOVA | 2.172 | 4.852 | 1.000 | 1.000 | 86.16 |
| Multiquadric | 21.235 | 49.465 | 6.225 | 2.573 | 78.72 |
| Inverse Multiquadric | 21.238 | 43.589 | 6.225 | 3.517 | 80.9 |
| Power | 31.231 | 13.514 | 1.073 | 0.870 | 83.84 |
| Rational Quadratic | 20.142 | 38.199 | 2.301 | 2.148 | 86.02 |
| Exponential | 7.225 | 8.143 | 2.460 | 35.254 | 88.91 |
| Spherical | 13.01 | 1.261 | 0.978 | 1.294 | 91.42 |
| Cauchy | 7.745 | 8.362 | 2.021 | 4.001 | 89.72 |

From Table 4, it can be seen that three kernel functions including affine, polynomial, and sigmoid kernel functions provide the approximated LPV model with closest dynamics to the original LPV model as the error to norm ratio of states for these functions is very small. The above-mentioned kernel functions are able to provide the model with the best-fit-ratio of more than 96% by using a single scheduling variable. The states of approximated and original LPV models of TRAS using polynomial and sigmoid kernel functions are shown in Fig. 6 and Fig. 7 respectively.

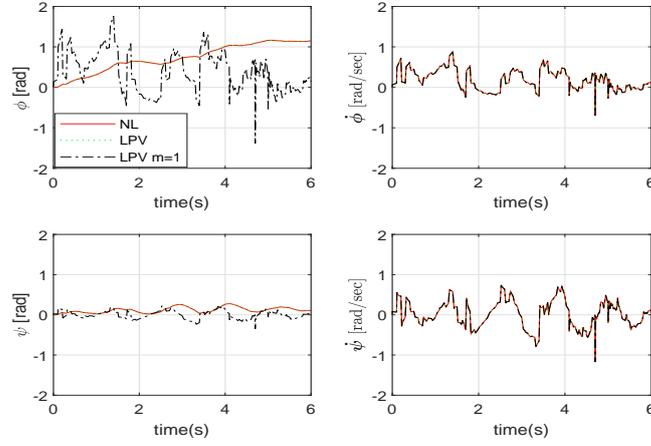


Figure 6. Approximated and original LPV model of TRAS using polynomial kernel function.

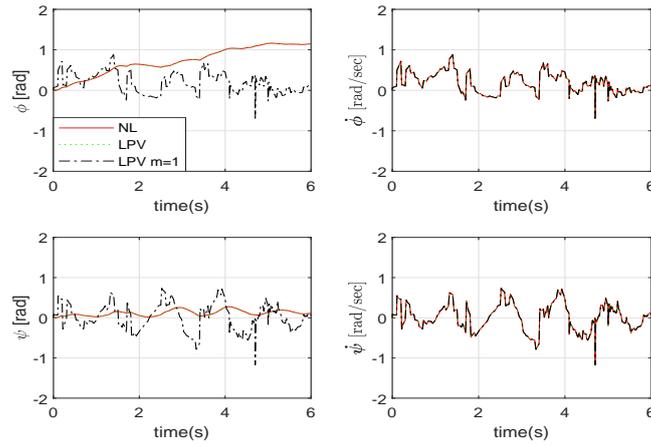


Figure 7. Approximated and original LPV model of CMG using affine kernel function.

Fig. 6 and Fig. 7 represent the states of original and approximated LPV model of TRAS using polynomial and sigmoid kernel functions respectively. One can conclude that polynomial kernel function provides better approximate LPV model using single scheduling signal compared to sigmoid kernel function.

§6 Conclusion

Comparative analysis of kernel-PCA based model approximation is carried out using fifteen kernel functions available in literature. The analysis is applied to two nonlinear systems of sufficient high complexity. Choice of kernel function ensuring accurate approximated LPV

model varies from system to system. However, the analytical study carried out in this paper shows that the polynomial kernel function is common in both cases. Based on the findings of this research, one is suggested to try polynomial kernel function before moving to other functions available in literature. In future by following the suggested method for selection of kernel function, the parameter dependent Lyapunov function-based controllers can be designed for complex nonlinear systems and can be applied for experimental validation.

References

- [1] W J Rugh. *Research on gain scheduling*, 1998.
- [2] A Bachnas. *Linear Parameter Varying Modeling of a High-Purity Distillation Column*, 2012.
- [3] A Kwiatkowski, H Werner. *PCA-based parameter set mappings for LPV models with fewer parameters and less overbounding*, IEEE Trans Control Syst Technol, 2008, 16: 781-788.
- [4] M M Siraj, R Tóth, S Weiland. *Joint order and dependency reduction for LPV state-space models*, in Decision and Control (CDC) IEEE 51st Annual Conference, 2012, 6291-6296.
- [5] R Toth, H S Abbas, H Werner. *On the state-space realization of LPV input-output models: Practical approaches*, IEEE Trans Control Syst Technol, 2012, 20: 139-153.
- [6] S Rizvi, J Mohammadpour, R Tóth. *A kernel-based PCA approach to model reduction of linear parameter-varying systems*, IEEE Trans, 2016.
- [7] M Meisami-Azad, J Mohammadpour, K M Grigoriadis, M P Harold, M A Franchek. *PCA-based linear parameter varying control of SCR aftertreatment systems*, American Control Conference (ACC), 2011, 1543-1548.
- [8] B Schölkopf, A Smola, K R Müller. *Kernel principal component analysis*, 1997, 583-588.
- [9] B Schölkopf, A Smola, K R Müller. *Nonlinear component analysis as a kernel eigenvalue problem*, Neural Comput, 1998, 10: 1299-1319.
- [10] A J Smola, B Schölkopf. *Learning with kernels*, 1998, 4.
- [11] B Schölkopf, A J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, 2001.
- [12] H T Lin, C J Lin. *A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods*, Submitt to Neural Comput, 2003, 3: 1-32.
- [13] C E Rasmussen. *Gaussian processes in machine learning*, Advanced lectures on machine learning Springer, 2004, 63-71.
- [14] T Hofmann, B Schölkopf, A J Smola. *Kernel methods in machine learning*, Ann Stat, 2008, 1171-1220.
- [15] N Cristianini, J Shawe-Taylor, et al. *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.
- [16] F Fleuret, H Sahbi. *Scale-invariance of support vector machines based on the triangular kernel*, in 3rd International Workshop on Statistical and Computational Theories of Vision, 2003, 1-13.

- [17] H Abbas, A Ali, S Hashemi, H Werner. *LPV state-feedback control of a control moment gyroscope*, Control Eng Pract, 2014.
- [18] T Parks. *Manual For Model 750: Control Moment Gyroscope*, Educ Control Prod Bell Canyon, CA, 1999.
- [19] W Bong. *Space vehicle dynamics and control*, Washington, DC AIAA Educ Ser AIAA, 1998.
- [20] P Chalupa, J Novák, J Příkryl. *Design and verification of a robust controller for the twin rotor MIMO system*, SYSTEMS AND SIGNAL PROCESSING, 2016, 10.
- [21] F Saleem, A Ali, M Wasim, I Shaikh. *An Iterative Kernel-PCA based LPV Control Approach for a 4-DOF Control Moment Gyroscope*, IEEE Access, 2019, 7: 164000-164008.
- [22] *Two Rotor Aero-Dynamical System User's Manual*, INTECO, 2013.

¹Department of Measurements and Control Systems, Silesian University of Technology, Gliwice, Poland.
Email: faisal.saleem.14@outlook.com

²Department of Electrical Engineering, University of Engineering and Technology, Taxila, Pakistan.
Email: ahsan.ali@uettaxila.edu.pk, inam.hassan@uettaxila.edu.pk

³Department of Aeronautics and Astronautics Engineering, Institute of Space Technology, Islamabad, Pakistan.

Email: m.wasim@mail.ist.edu.pk